ALIBABA CLOUD

阿里云

金融分布式架构

运维管理

文档版本：20220609

阿里云

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。

2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。

3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。

4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。

5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。

6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
危险	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	危险 重置操作将丢失用户配置数据。
警告	该类警示信息可能会导致系统重大变更甚至故障，或者导致人身伤害等结果。	警告 重启操作将导致业务中断，恢复业务时间约十分钟。
注意	用于警示信息、补充说明等，是用户必须了解的内容。	注意 权重设置为0，该服务器不会再接受新请求。
说明	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	说明 您也可以通过按Ctrl+A选中全部文件。
>	多级菜单递进。	单击**设置**> **网络**> **设置网络类型**。
粗体	表示按键、菜单、页面名称等UI元素。	在**结果确认**页面，单击**确定**。
Courier字体	命令或代码。	执行 `cd /d C:/window` 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	`bae log list --instanceid` *Instance_ID*
[] 或者 [a|b]	表示可选项，至多选择一个。	`ipconfig [-all|-t]`
{} 或者 {a|b}	表示必选项，至多选择一个。	`switch {active|stand}`

目录

1.运维管理产品总览 17
2.概述 19
3.蚂蚁 PaaS 平台核心领域模型介绍 20
4.快速入门 31
5.应用管理 37
5.1. 什么是应用管理 37
5.2. 快速开始 38
5.3. 使用指南 38
5.3.1. 管理应用 38
5.3.2. 管理应用分组 42
5.3.3. 管理技术栈 43
6.资源管理 46
6.1. 什么是资源管理 46
6.2. 工作空间 46
6.2.1. 什么是工作空间 46
6.2.2. 管理工作空间 47
6.2.3. 设置专有网络 49
6.3. 计算与网络 50
6.3.1. 云服务器 ECS 50
6.3.1.1. 添加 ECS 50
6.3.1.2. 分配与释放 ECS 51
6.3.1.3. VPC 网络下导入 ECS 实例 52
6.3.1.4. 经典网络下导入 ECS 实例 53
6.3.1.5. 删除 ECS 54
6.3.1.6. 查看磁盘详情 55
6.3.2. 负载均衡 SLB 55

6.3.3. 安全组 61
6.3.4. 专有网络 VPC 63
6.4. 任务记录 64
7.经典应用服务 66
7.1. 概述 66
7.1.1. 什么是经典应用服务 66
7.1.2. 基础术语 67
7.2. 快速入门 67
7.3. 应用部署 71
7.4. 应用运维 76
7.5. 发布包管理 80
7.6. 应用服务实例 81
7.7. 部署单元 84
7.8. 自定义技术栈 85
7.9. 经典应用服务常见问题 85
8.容器应用服务 88
8.1. 什么是容器应用服务 88
8.1.1. 概述 88
8.1.2. 优势 88
8.1.3. 功能特性 88
8.1.4. 应用场景 90
8.1.5. 架构 90
8.1.6. 基础术语 91
8.1.7. 限制说明 93
8.2. 快速入门 93
8.3. 集群 102
8.3.1. 对接 ACK 产品变更说明 102
8.3.2. 创建集群 104

8.3.3. 集群扩容 108
8.3.4. 删除集群 110
8.3.5. 导入 ACK 集群 111
8.3.6. 查看集群详情 112
8.3.7. 节点管理 115
8.3.8. 组件管理 118
8.3.9. 命名空间 120
8.4. 镜像中心 121
8.4.1. 镜像构建 121
8.4.2. 镜像仓库 124
8.4.3. 向金区镜像仓库上传镜像 125
8.4.4. 镜像技术栈 127
8.4.5. 示例镜像 128
8.5. 应用服务 129
8.5.1. 创建应用服务 129
8.5.2. 高级配置说明 136
8.5.3. 更新应用服务 140
8.5.4. 删除应用服务 141
8.5.5. 立即发布服务 141
8.5.6. 发布单 142
8.5.7. 应用服务详情 144
8.6. 网络 148
8.6.1. 创建服务 148
8.6.2. 创建负载均衡 150
8.6.3. 编辑与删除服务 150
8.6.4. 创建路由 151
8.6.5. 编辑与删除路由 152
8.7. 存储 152

8.7.1. 创建存储类型（StorageClass） 152
8.7.2. 创建存储卷（PV） 157
8.7.3. 创建存储卷声明（PVC） 158
8.7.4. 使用 PVC 158
8.7.5. 使用 PVC 挂载动态存储卷 159
8.7.6. 使用 PVC 挂载静态存储卷 162
8.7.7. StatefulSet 挂载动态存储卷 165
8.8. 配置 166
8.8.1. 配置项 166
8.8.2. 保密字典 169
8.9. 应用商店 172
8.9.1. 概述 172
8.9.2. 产品模板 172
8.9.3. 产品安装 173
8.9.4. 自定义 OAM 模板 175
8.9.5. OAM 产品模板 181
8.10. 工作负载 182
8.10.1. 部署（Deployments） 182
8.10.2. 有状态副本集（Statefulsets） 184
8.10.3. 守护进程集（Daemonsets） 185
8.10.4. 容器组（Pods） 187
8.11. 常见问题 188
8.12. 最佳实践 190
8.12.1. 容器镜像服务使用规范 190
8.12.2. 容器网络与节点规划 191
8.12.3. 选择存储卷类型 192
9.单元化应用服务 194
9.1. 什么是单元化应用服务 194

9.1.1. 单元化介绍 194
9.1.2. 概述 197
9.1.3. 产品优势 197
9.1.4. 产品架构 198
9.1.5. 功能特性 198
9.1.6. 应用场景 199
9.1.7. 基础术语 200
9.2. 快速入门 201
9.3. 准备环境 209
9.3.1. 创建单元化工作空间 209
9.3.2. 创建逻辑单元 213
9.3.3. 创建部署单元 215
9.3.4. 删除工作空间 217
9.4. 集群 220
9.4.1. 创建集群 220
9.4.2. CCM 角色授权 224
9.4.3. 导入集群 225
9.4.4. 扩容集群 226
9.4.5. 查看集群详情 228
9.4.6. 解绑集群 231
9.4.7. 节点管理 232
9.4.7.1. 节点调度 232
9.4.7.2. 管理节点标签 233
9.4.7.3. 管理 Taint 234
9.4.7.4. 查看节点详情 234
9.4.7.5. 删除节点 235
9.4.8. 命名空间 236
9.4.8.1. 创建命名空间 236

9.4.8.2. 查看命名空间详情 237
9.4.8.3. 删除命名空间 239
9.4.8.4. 设置资源配额 239
9.4.9. 管理组件 241
9.5. 应用服务 243
9.5.1. 创建应用服务 243
9.5.2. 高级配置 254
9.5.3. 发布应用服务 257
9.5.3.1. 立即发布 257
9.5.3.2. 发布单 260
9.5.3.3. 蓝绿发布 262
9.5.3.4. 多集群联邦发布容灾 265
9.5.4. 更新应用服务 266
9.5.5. 导入与导出应用服务 267
9.5.6. 查看应用服务详情 268
9.5.7. 删除应用服务 270
9.5.8. 系统环境变量 271
9.5.9. 容器水平伸缩 HPA 272
9.5.10. 运维单 275
9.6. 流量管理 277
9.6.1. 应用层流量管理 277
9.7. 网络 281
9.7.1. 打通多集群容器网络 282
9.7.2. 创建统一接入集群 284
9.7.3. 统一接入节点性能压测 286
9.7.4. 网关集群状态说明 292
9.7.5. 创建统一接入实例 293
9.7.6. 联邦负载均衡 295

9.7.6.1. 创建联邦负载均衡 295
9.7.6.2. 配置联邦负载均衡 296
9.7.6.3. 删除联邦负载均衡 297
9.8. 配置管理 297
9.8.1. 管理配置项 297
9.8.2. 使用配置项 298
9.8.3. 管理保密字典 300
9.8.4. 使用保密字典 301
9.9. 工作负载 303
9.9.1. 部署（Deployments） 303
9.9.2. 有状态副本集（Statefulsets） 305
9.9.3. 守护进程集（Daemonsets） 306
9.9.4. 容器组（Pods） 307
9.9.5. 定时任务（CronJobs） 309
9.9.6. 一次性任务（Jobs） 311
9.9.7. 自定义资源对象 313
9.9.7.1. NetworkPolicy 概述 313
9.9.7.2. 管理 NetworkPolicy 314
9.9.7.3. CafeDeployment 317
9.9.7.4. InPlaceSet 319
9.10. 存储 320
9.10.1. 管理存储类型（StorageClass） 320
9.10.2. 管理存储卷（PersistentVolume） 324
9.10.3. 管理存储卷声明（PVC） 325
9.10.4. 使用存储卷声明（PVC） 326
9.10.5. 使用 PVC 挂载动态存储卷 327
9.10.6. 使用 PVC 挂载静态存储卷 329
9.10.7. StatefulSet 挂载动态存储卷 332

9.11. 镜像中心 333
9.11.1. 镜像仓库 333
9.11.2. 构建镜像 334
9.11.3. 镜像技术栈 337
9.12. 应用商店 338
9.12.1. 应用商店概述 338
9.12.2. 产品模板 339
9.12.3. 产品安装 341
9.13. 基础元数据 343
9.13.1. 逻辑单元 343
9.13.2. 部署单元 344
9.14. 常见问题 345
9.15. 最佳实践 348
9.15.1. LHC Pod 域名配置 348
10.实时监控（RMS 3.0） 352
10.1. 产品简介 352
10.1.1. 什么是实时监控 352
10.1.2. 功能架构 353
10.1.3. 功能概览 355
10.1.4. 实现原理 356
10.1.5. 应用场景 359
10.1.6. 基本概念 361
10.2. 产品计费 362
10.2.1. 计费项概述 362
10.2.2. 按量付费（后付费） 363
10.2.3. 资源包抵扣（预付费） 365
10.2.4. 计费案例 365
10.2.5. 欠费停服说明 368

10.3. 快速开始 368
10.3.1. 创建自定义监控 368
10.3.2. 创建告警规则 374
10.3.3. 订阅告警 382
10.4. 总览 385
10.5. 大盘 392
10.6. 自定义监控 392
10.6.1. 自定义监控概述 392
10.6.2. 监控类型概述 393
10.6.3. 创建大盘 395
10.6.4. 分钟统计/无 Key 409
10.6.5. 分钟统计/多 Key 414
10.6.6. 配置常用服务指标 420
10.6.7. 单笔数据 Top 422
10.6.8. 匹配查找告警 426
10.6.9. 归档统计 430
10.6.10. 文件夹操作 432
10.7. 应用监控 434
10.7.1. 应用监控概述 434
10.7.2. 应用总览 437
10.7.3. 服务指标 443
10.7.4. 系统指标 455
10.7.5. JVM 指标 457
10.7.6. 端口探测 460
10.7.7. MOSN 总览 460
10.8. 分布式链路 465
10.8.1. 分布式链路概述 465
10.8.2. 应用拓扑 466

10.8.2.1. 概述 466
10.8.2.2. 查看应用拓扑图 466
10.8.2.3. 查看应用列表 470
10.8.2.4. 查看应用性能详情 470
10.8.2.5. 查看节点调用信息 475
10.8.3. 链路查询 476
10.8.3.1. 查询调用链路 476
10.8.3.2. 查看链路详情 478
10.8.3.3. 链路上报 480
10.8.4. 分布式链路配置 481
10.9. 告警管理 481
10.9.1. 通知组管理 481
10.9.2. webhook 管理 482
10.9.3. 告警历史 484
10.9.4. 订阅管理 485
10.9.5. 告警联系人 486
10.9.6. 设置钉钉报警机器人 488
10.10. 个人中心 490
10.10.1. 收藏夹 490
10.10.2. 告警和订阅 492
10.10.3. 个人设置 494
10.11. 设置 495
10.11.1. Agent 管理 495
10.12. 常见问题 497
11.资金安全监控 500
11.1. 产品简介 500
11.1.1. 概述 500
11.1.2. 功能特性 501

11.1.3. 应用场景 502
11.1.4. 基础术语 503
11.2. 快速入门 503
11.2.1. 离线核对 503
11.2.2. 文件核对 508
11.3. 概览 515
11.4. 核对点 515
11.4.1. 管理核对点 516
11.4.2. 创建核对规则 520
11.4.3. 编辑核对规则 522
11.4.4. SQL 全屏调试 522
11.4.5. 高级配置 523
11.5. 异常列表 525
11.6. 核对模板 527
11.7. 数据节点 528
11.8. 数据源 531
11.8.1. 数据资源配置 531
11.8.2. 创建数据源 532
11.9. 核对场景 534
11.10. 订阅 535
11.11. 常见问题 536
12.高可用管理 538
12.1. 产品简介 538
12.1.1. 什么是高可用管理平台 538
12.1.2. 产品优势 538
12.1.3. 产品架构 538
12.1.4. 功能特性 539
12.1.5. 应用场景 542

12.1.6. 基础术语 542
12.2. 总览 543
12.3. 风险管理 544
12.3.1. 风险事件 544
12.3.2. 风险场景 546
12.4. 日常巡检 547
12.4.1. 日常巡检概览 548
12.4.2. 巡检插件 548
12.4.3. 巡检规则 552
12.4.4. 巡检任务 554
12.4.5. 巡检报告 556
12.5. 故障诊断 558
12.5.1. 诊断规则 558
12.5.2. 诊断节点 559
12.5.3. 诊断决策树 561
12.5.4. 诊断报告 562
12.6. 应急预案 562
12.6.1. 预案管理 562
12.6.2. 预案执行 567
12.6.3. 预案原子操作 569
12.7. 故障演练 570
12.7.1. 演练场景 570
12.7.2. 演练方案 573
12.7.3. 演练执行 574
12.7.4. 演练原子操作 574
12.8. 应急管理 575
12.9. 配置管理 577
12.9.1. 审批管理 577

12.9.2. 基线管理 ---------- 578
12.9.3. 钉钉机器人 ---------- 579

1.运维管理产品总览

运维管理模块提供以下功能：



应用管理

- 创建应用
- 管理应用分组
- 管理技术栈



资源管理

- 统一管理和运维各工作空间内的基础资源
- 绑定基础资源与工作空间以实现精细化运维



经典应用服务

- 提供强大的发布部署和运维管道能力
- 提供应用全生命周期的 DevOps 自动化支持
- 提供可定制的自动化运维



容器应用服务

- 金融级发布策略、一站落地云原生
- 提供一键从 0 到 1 搭建集群的能力
- 提供丰富的发布策略



单元化应用服务

- 提供应用管理、发布运维、流量调拨、配置同步等能力
- 提供跨集群资源管理



实时监控 RMS 3.0

- 实现全方位实时监控
- 灵活配置报警规则
- 提供丰富的可视化大盘
- 提供应用拓扑和链路查询功能



资金安全监控

- 全面覆盖业务资金链路，为核心业务链路保驾护航
- 及时发现异常数据，避免资损



高可用管理

- 提供覆盖技术风险事件前、中、后的业务连续性全链路管理能力
- 通过规则任务报告+巡检插件的体系实现“去人为化”操作

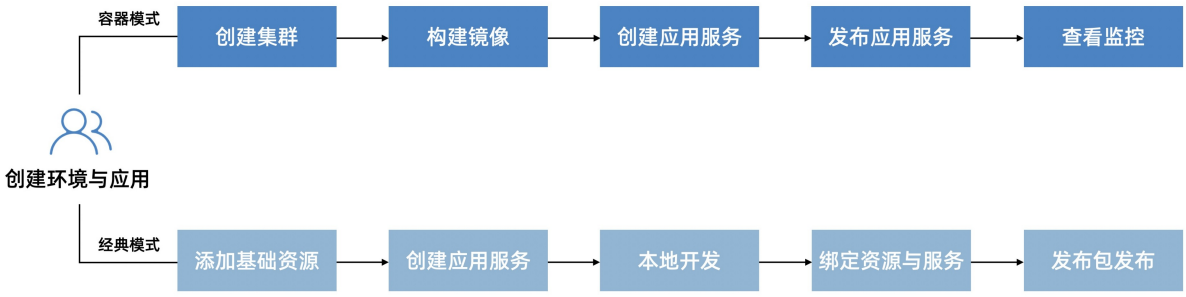
2.概述

组件概述

运维管理模块提供以下功能：

- **应用管理**：创建应用，管理应用基本信息。
- **资源管理**：创建与管理基础云服务产品，例如：云服务器 ECS、云数据库 RDS、对象存储 OSS、云数据库 Memcache、日志服务等。
- **经典应用服务**：以应用为核心视图，对应用的版本、发布包、资源等进行可视化、自动化管理。
- **单元化应用服务**：创建与部署单元化应用服务，提供完整的集群管控、认证授权、容器网络、持久卷存储等方面的平台能力。
- **实时监控（RMS 3.0）**：以日志、REST 接口、Shell 脚本、云产品接口等作为数据采集来源，提供资源、应用、业务等各种视角的监控能力。

使用流程



使用运维管理的完整流程如下：

1. 创建工作空间：在 **全局设置** 中 添加工作空间。
2. 创建应用：在 **应用管理** 中 创建应用。
3. 根据以下不同的模式，构建与发布应用服务。
   - 经典模式：
     a. 添加资源：在 **资源管理** 中添加资源，如 ECS、SLB、安全组、VPC。
     b. 创建应用服务：在 **经典应用服务** 中 创建应用服务。
     c. 本地开发：本地使用 SOFABoot 框架进行开发，参见 SOFABoot 快速开始。
     d. 绑定资源与服务：在 **经典应用服务** 中绑定资源与应用服务，参见 快速开始。
     e. 上传报发布包发布应用：在 **经典应用服务** 中上传发布包并发布应用，参见 快速开始。
   - 容器模式：
     a. 添加资源：在 **单元化应用服务** 中创建集群、交换机、安全组等资源。
     b. 准备镜像：使用 **单元化应用服务** 提供的平台样例镜像或私有镜像仓库中的镜像。
     c. 创建应用服务：在 **单元化应用服务** 中创建应用服务，进行应用服务镜像配置、弹性伸缩配置、访问配置、部署和调度配置。在高级配置中可填写 日志服务配置 对应用进行日志分析。
     d. 发布部署：在 **单元化应用服务** 中 立即发布 刚创建的应用服务，或者创建 发布单 同时分布多个应用服务。
     e. 查看日志与监控：在 **单元化应用服务** 中 查看容器标准日志。如需查看监控，请前往 RMS 3.0 控制台在 RMS 3.0 中 查看应用监控。

3.蚂蚁 PaaS 平台核心领域模型介绍

本篇介绍了蚂蚁 PaaS 平台的基本概念与核心领域模型，同时也对阿里云的 IaaS 核心模型进行了阐述，希望对想要初步了解蚂蚁 PaaS 平台技术架构的用户提供入门指引。

蚂蚁 PaaS 平台介绍

蚂蚁 PaaS 平台是蚂蚁业务全球化输出的技术支撑平台，PaaS 平台的领域模型定义，是 PaaS 层产品技术最核心的架构规范。蚂蚁 PaaS 平台作为金融科技商业化输出的底盘（商业化品牌名为 SOFAStack CAFE）提供了以应用为核心的研发、运行、运维全生命周期能力支撑。

IaaS 层核心领域模型与概念

地域（Region）与可用区（Availability Zone，AZ）

阿里云上的基础物理拓扑包括地域（Region）和可用区（Availability Zone），它们是阿里云的两个核心领域模型。



- **地域 Region**：物理的数据中心。地域通常以城市为单位，如杭州、上海、青岛、北京等，每个地域之间是互相物理独立和隔离的，如要互通则需要通过物理专线的方式打通。不同的 Region 提供的产品能力、安全与稳定性水位、资源价格有所不同。
- **可用区 Availability Zone/AZ**：在同一地域内，电力和网络互相独立的物理区域。可用区之间网络互通，故障隔离。是否将实例放在同一可用区内，主要取决于对容灾能力和网络延时的要求。可用区通常也代指机房，即 IDC，两者概念相同。

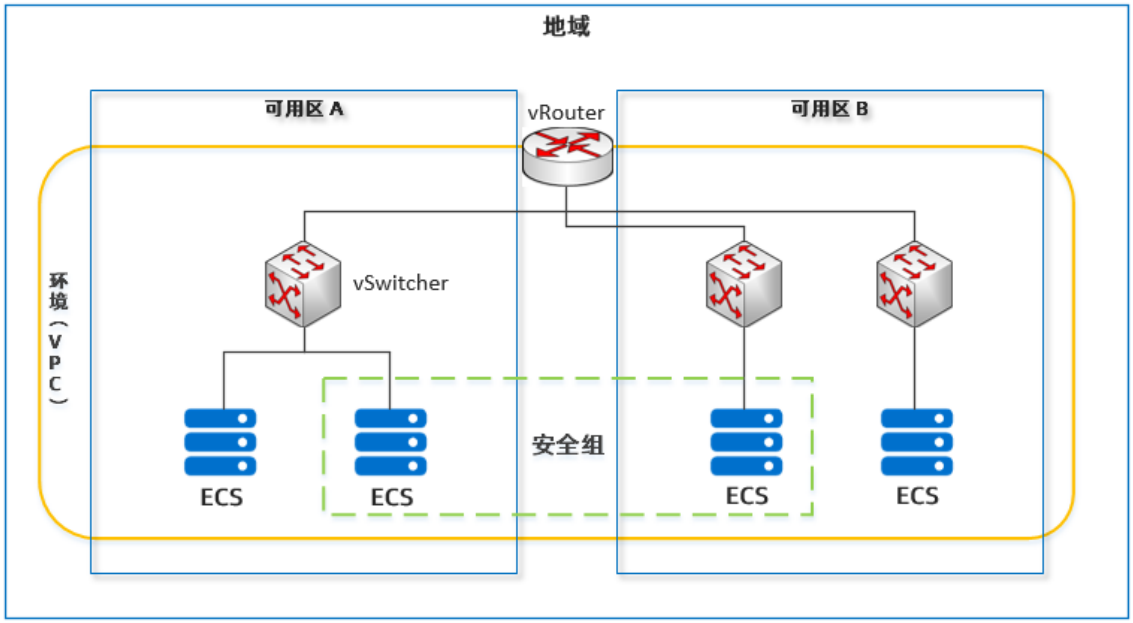
  > 说明
  > 阿里云划分地域和可用区的原因是考虑网络延时。相同地域的不同可用区之间的网络延时不超过 2ms。

- **金融地域（金区）**：针对金融行业客户对安全水位有更高要求的场景，会提供完全隔离，并专门部署更高安全水位的云产品组合的地域，称之为金融地域（或金融区）。例如公有云有三个金融 Region 分别是：杭州金区（cn-hangzhou-finance）、上海金区（cn-shanghai-finance-1）和深圳金区（cn-shenzhen-finance-1）。
  更多介绍，请参见 金融云特性、金融云产品列表。

专用网络（VPC）

专有网络 VPC（Virtual Private Cloud）是您自己独有的云上私有网络。您可以完全掌控自己的专有网络，例如选择IP地址范围、配置路由表和网关等，您可以在自己定义的专有网络中使用阿里云资源如云服务器 ECS、云数据库 Redis 版和负载均衡 SLB 等。

VPC 定义了网络边界， 您可以更加方便地在云上自定义网络的规划和管理：您可以自定义虚拟路由器（vRouter）网段，配置路由表规则，也可以自定义虚拟交换机（vSwitcher），配置子网。



阿里云对 VPC 模型的限制如下：

- 一个 VPC 不可以跨 Region。
- 一个 VPC 只有一个 vRouter。
- 一个 vSwitcher 只能属于一个 AZ。
- 一个安全组可以包括不同 vSwitcher 下的 ECS。

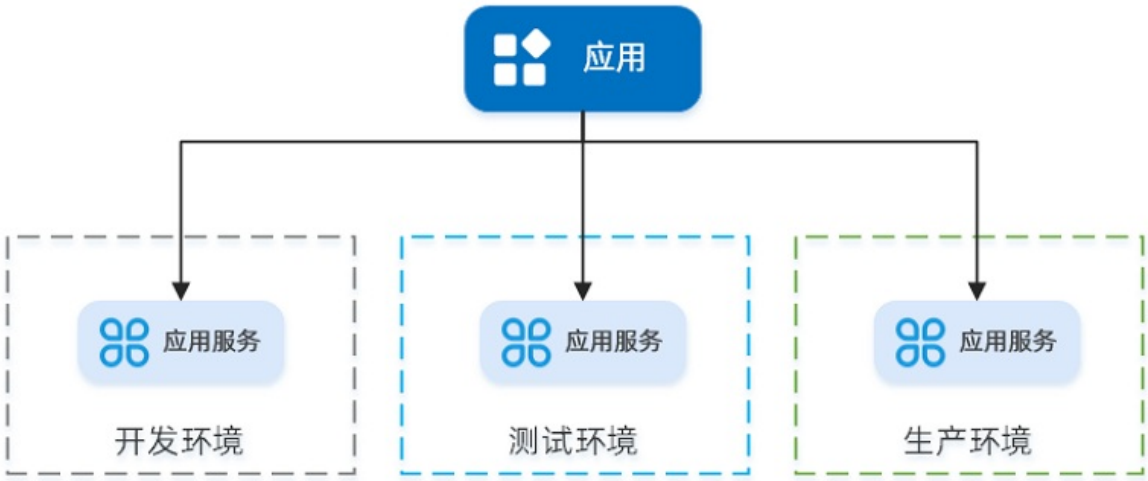
相关操作请参见：专有网络 VPC、网络规划。

PaaS 层核心领域模型

应用（Application）与应用服务（AppService）

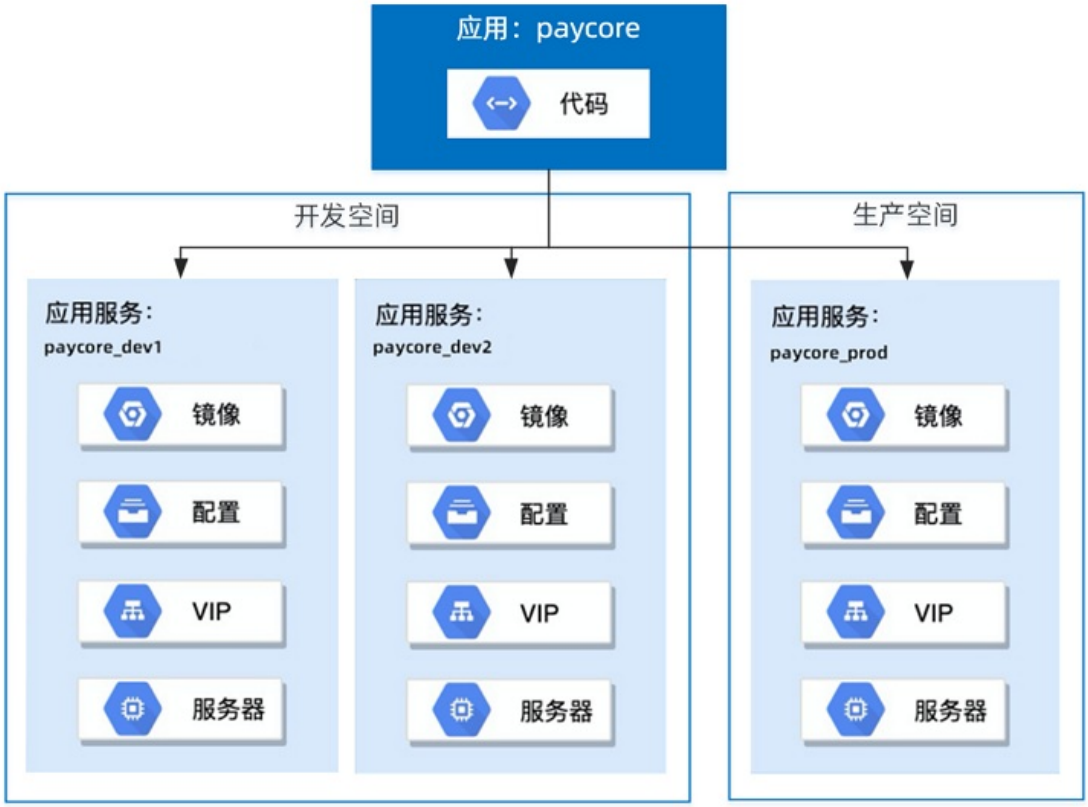
为了方便金融级大规模分布式系统的研发运维，蚂蚁 PaaS 平台在阿里云基础资源之上抽象了“应用（Application）”和 “应用服务（AppService）” 的概念。

一个租户内的一套代码对应一个应用，应用有对应的技术栈、分组、分级等基础元信息。应用是 PaaS 平台的核心概念，蚂蚁研发运维体系都围绕着应用来展开，研发应用、发布应用、运维应用、监控应用等。



应用服务（AppService）是 CAFE 发布运维体系的核心模型，一个业务应用常常由多个应用服务组成。应用服务代表一个应用在一个工作空间的一次部署，部署时需要有应用镜像/发布包、应用配置、以及部署的资源。

例如针对多人开发同一个应用的场景，每个开发同学会拉出自己的代码分支部署在开发环境去部署，即在开发工作空间（Workspace：Dev）中，有三个同学开发测试部署同一个应用的不同分支（App：paycore），会产生三个不同的应用服务（AppService：paycore_dev1、paycore_dev2、paycore_prod）。每个应用服务可以部署在不同的服务器上，关联不同的基线配置。但对于生产环境，不允许一个应用在生产环境有多个不同的代码版本，安全隐患极大。

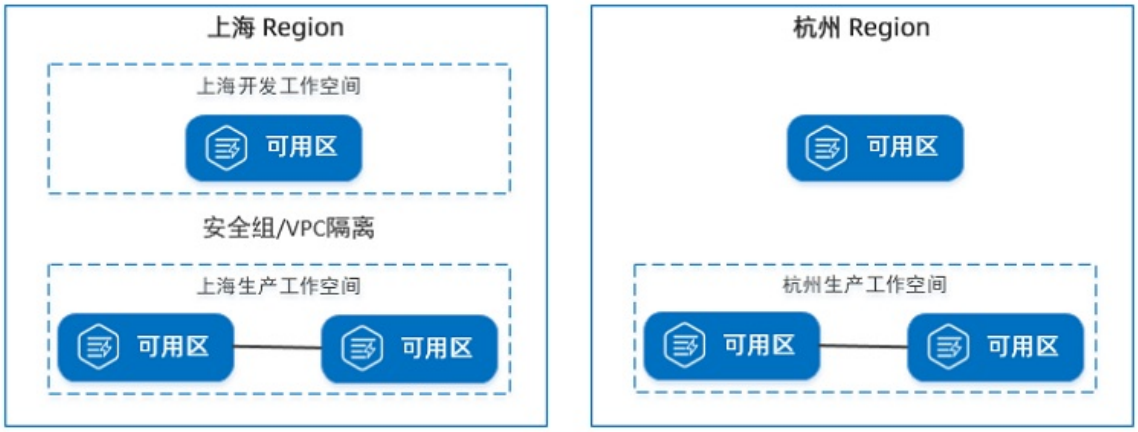


相关操作请参见：管理应用、创建应用服务。

工作空间（Workspace）、地域（Region）与可用区（Availability Zone）

工作空间（Workspace）是 SOFAStack 提供的一种组织机制，用于将服务于不同目标、阶段的资源分组隔离管理。典型的使用场景是根据研发运维流程，为每一个阶段分配工作空间，例如单机房（即单可用区）的开发工作空间、双机房（即包括两个可用区）的生产工作空间。

> 注意
> Workspace 可以跨可用区，实现容灾能力，但必须在同一个地域内，跨地域的多个工作空间可以组成工作空间组（通过单元化架构支持）。



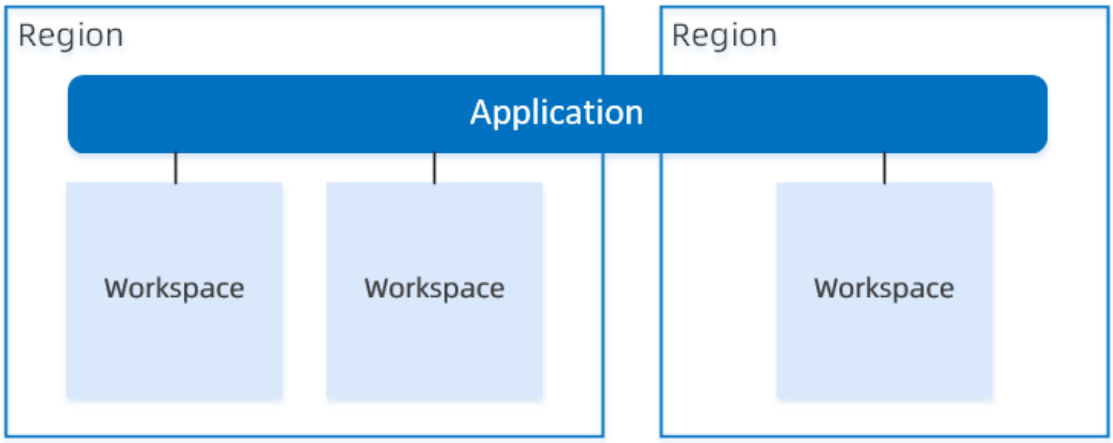
工作空间（Workspace）的核心是一组不受网络连通性束缚的资源边界。在阿里云，地域之间的网络默认不通，同时云产品也基本上是以地域为边界设置产品实例。例如：SLB 不可能跨地域挂载 ECS、VPC 不可以跨地域等。因此，我们设立工作空间模型的原因是希望在这个范围内的资源之间的关系不用考虑网络连通性所带来的复杂性问题。

在实现过程中，工作空间通过 RAM 实现访问控制的隔离，通过 VPC 和安全组实现网络隔离，或者通过分属不同的 Kubernetes 集群实现物理隔离，亦或是通过在同一个 Kubernetes 集群内的不同资源标签实现筛选管理。

相关操作请参见：管理工作空间。

应用（Application）与工作空间（Workspace）

应用是租户级别的概念，一个租户内的一套代码对应一个应用。应用不仅跨 Workspace 存在，也会跨 Region 存在。



应用服务（AppService）与工作空间（Workspace）

应用研发的生命周期中，研发人员经常会接触多个环境。我们一般可以认为，一个 Workspace 就是一个环境。但是，实际上一个环境，也可以由多个 Workspace 组成。

> 说明
> 在 PaaS 的核心领域模型中，不再尝试定义“环境”这个模型，因为环境的概念在不同组织的研发协同流程中定义是不一致的，所以环境可以根据当前的 PaaS 领域模型组合搭建。

一个应用服务，并不受工作空间范围的约束，因为一个应用服务往往可能同时部署在多个工作空间，比如一个“生产环境”可能由多个工作空间组成。

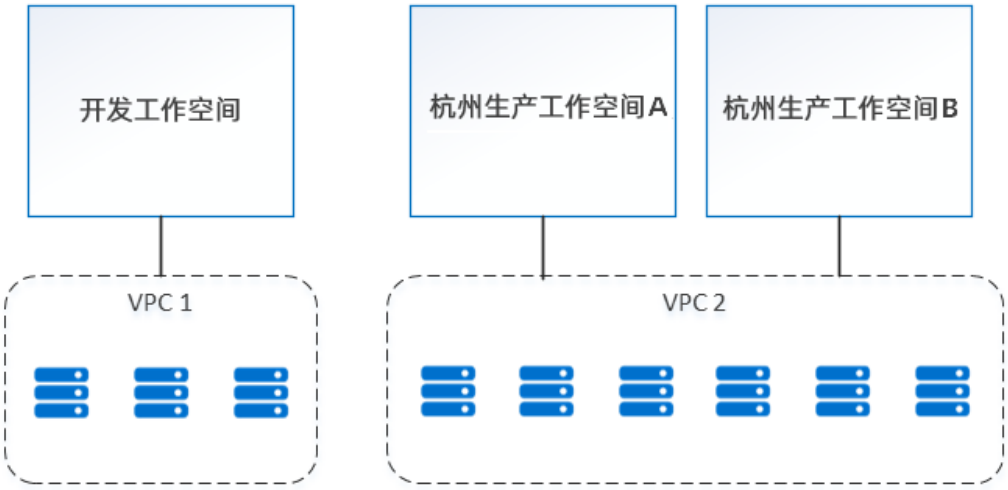


但是，应用服务下的资源一定是属于某一个工作空间的，同一资源不能同时属于两个工作空间。

相关操作请参见：创建应用服务。

工作空间（Workspace）与专用网络（VPC）

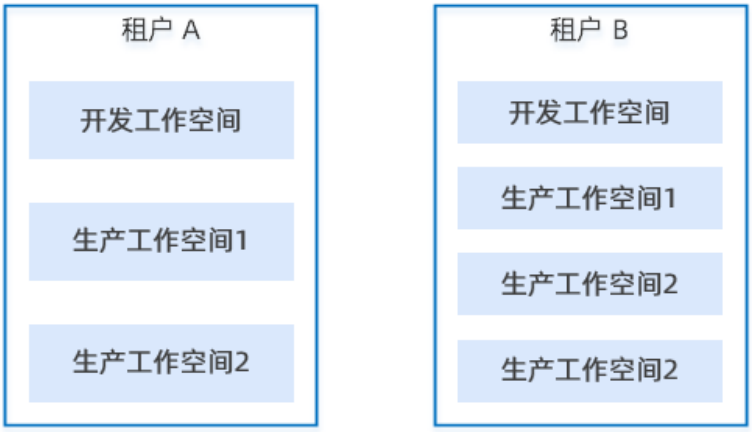
每个工作空间只会对应到唯一一个 VPC，这是因为工作空间的核心表达是一组不受网络连通性束缚的资源边界的定义。



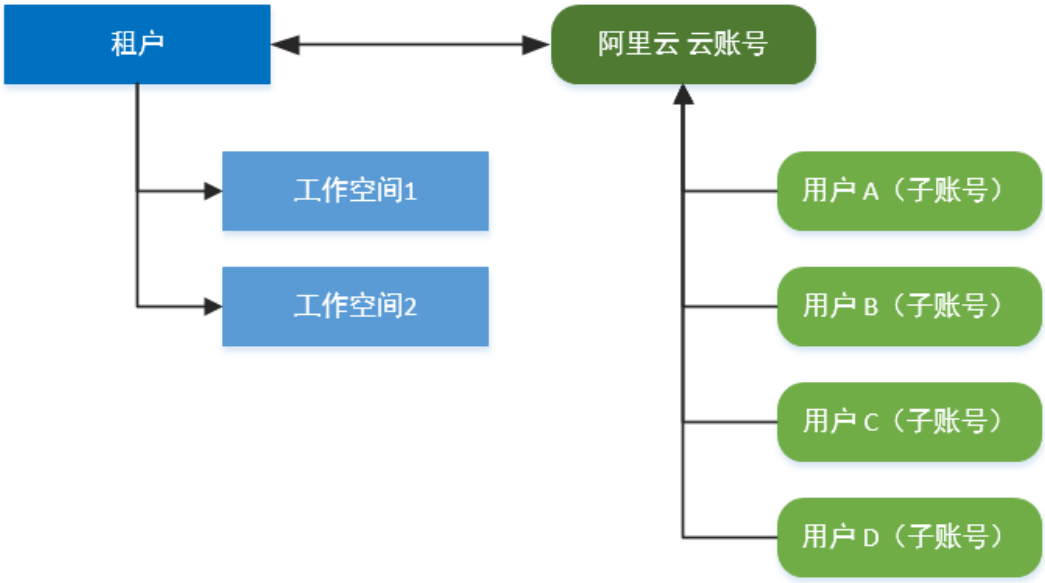
> 说明
> - 同一个 VPC 可以绑定到不同的工作空间中，这时 VPC 中的资源需要手动导入到工作空间中。
> - 可以通过高速通道进行不同工作空间的 VPC 之间的打通。

多租户（Multi-tenant）

公有云 SOFAStack 是多租户（Multi-tenant）架构，各租户可以按照需要创建自己的工作空间，并且定义人员可以进行的操作权限。



SOFAStack 与阿里云的账号体系进行了深层次的融合：一个租户映射到一个阿里云的云账号（即主账户）上。每个用户，都是属于这个云账号下的子账号。



- **主账号**：即阿里云云账号，可用于登录阿里云平台，具有 Root 权限，是阿里云独立的结算个体，通常由企业持有，独立负责阿里云侧的结算。
- **子账号**：主账号下虚拟的子账号，可以创建任意多个，通常对应企业内的人或者系统。

相关操作请参见：RAM 主子账号授权、跨云账号授权。

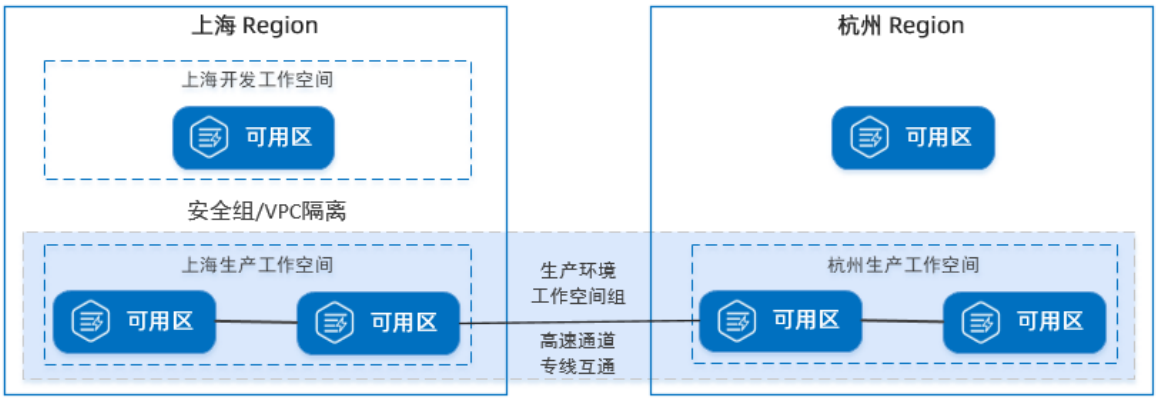
PaaS 层扩展领域模型

为了支撑更加复杂的应用场景，金融云 PaaS 还定义了以下更高级的模型。

- 工作空间组（WorkspaceGroup ）
- 应用服务实例组（AppService Instance Group，AIG）
- 部署单元（Cell）

工作空间组（WorkspaceGroup ）

工作空间组（WorkspaceGroup）是工作空间（Workspace）在多地域的扩展，在多地域内对资源进行分组隔离管理。多个地域的网络可以通过专线高速通道实现互通。如生产工作空间组（WorkspaceGroup：prod），包含两个工作空间：上海生产工作空间（Workspace：prod_shanghai）、杭州生产工作空间（Workspace：prod_hangzhou）。



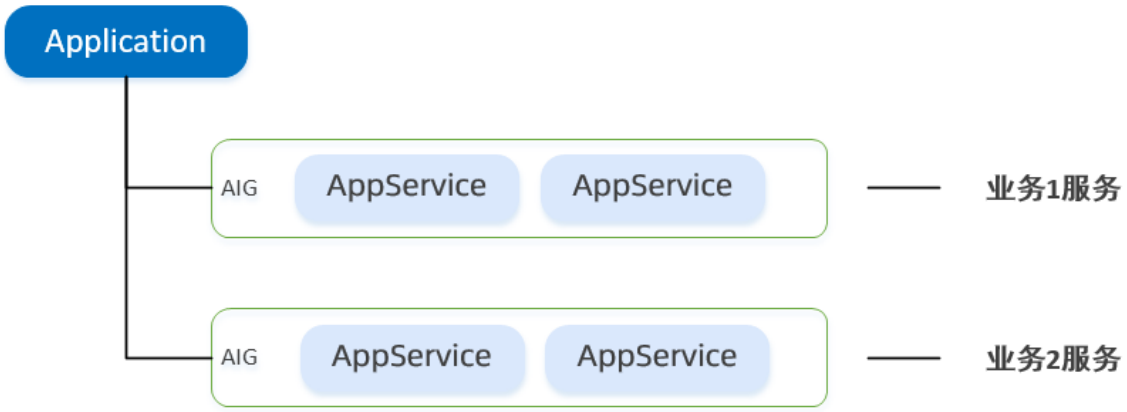
在 SOFAStack 控制台上，有些产品会支持以 WorkspaceGroup 视角进行操作：比如单元化应用服务 LHC、异地容灾产品、监控产品等。

相关操作请参见：创建工作空间组。

应用服务组（AppServiceInstance Group，AIG）

应用服务组（AIG） 用于圈定一组应用服务下的资源，是 Application 的一个更加细粒度的部署分片。

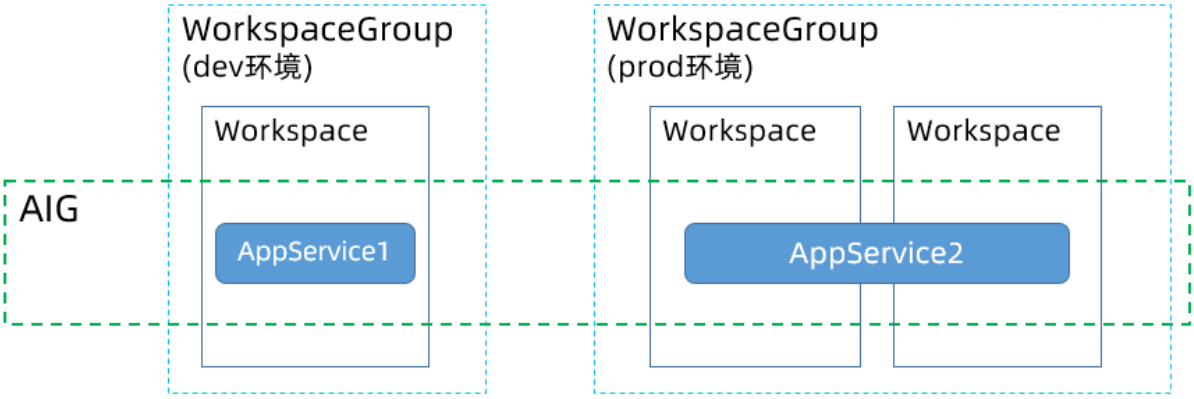
> 注意
> 一个应用服务（AppService）不一定必须属于一个 AIG，但一定属于一个 Application。



有些业务场景下，我们需要对同一个应用（Application）从业务层面进行更细粒度的划分。例如，不同的业务都依赖同一个 Application，如何从一个业务的视角，对这个应用下的部署实例进行更细粒度的控制，比如部署隔离、流量分配等。因此在 Application 模型下，增加了 AIG 这个模型。

应用服务组（AIG）与工作空间组（WSG）

AIG 可以跨 WSG。按照蚂蚁集团的经验，可以将一个 WSG 定义为一个环境。

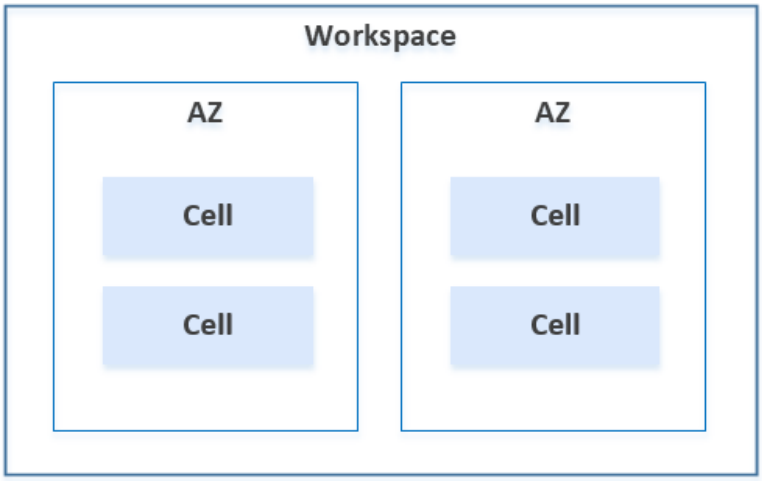


AppService 可以跨 WS，但不可以跨 WSG。因为通常来说，一个 AppService 不可能包括两个“环境”下的资源，两个“环境”下的 AppService 往往意味着两种不同的基线配置。

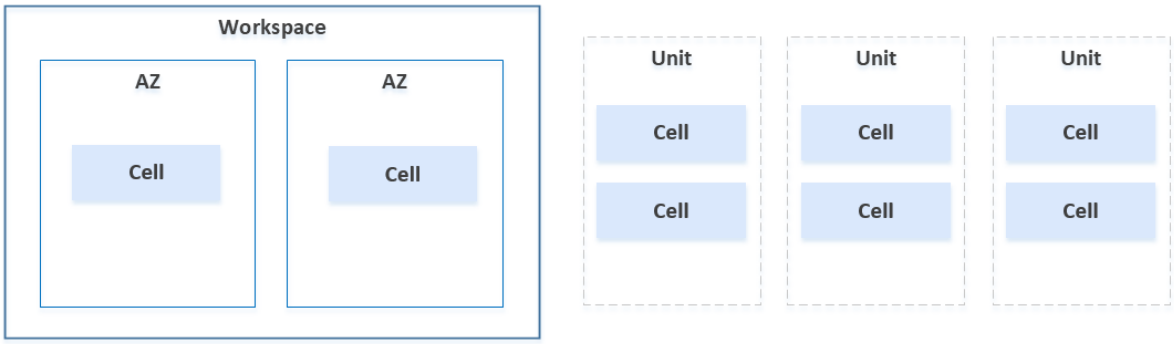
部署单元（Cell）

Cell 是发布部署要感知的最小部署单元。一个 Cell 一般会对应一个中间件配置中心。

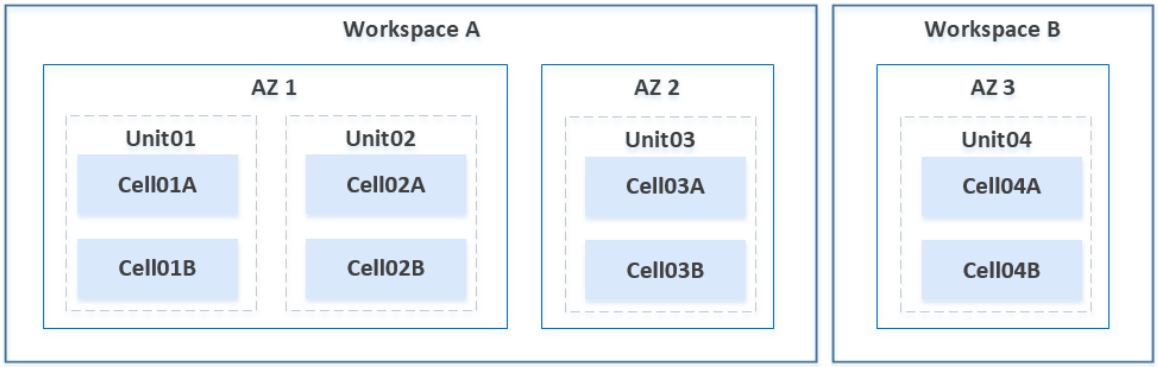
> 注意
> Cell 不可以跨 AZ。



Cell 是实现同城双活、蓝绿发布、单元化架构的基础。每个 Cell 内可以理解为有一个逻辑独立的配置中心。



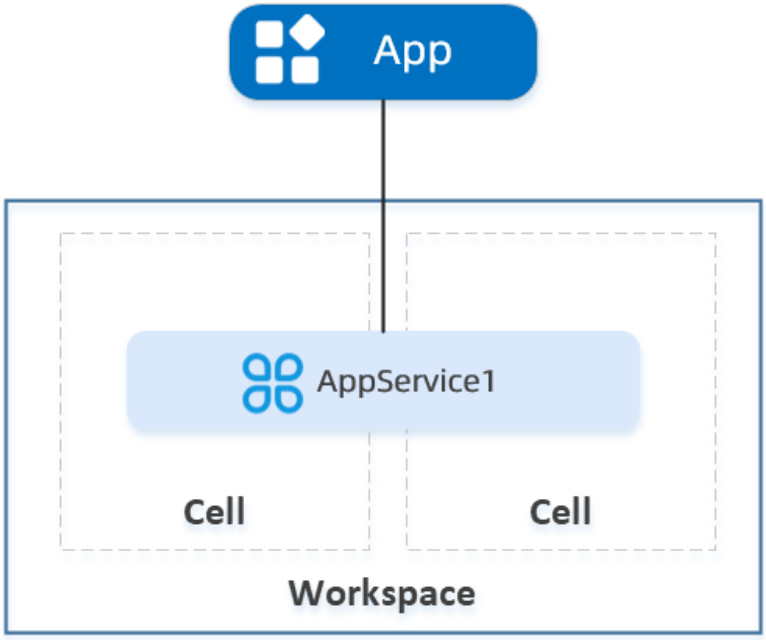
Unit 是 LHC 产品封装的模型，不是 PaaS 层核心领域模型。LHC 产品利用 PaaS 层的 Cell 模型对单元化架构中的逻辑 Zone 进行划分。



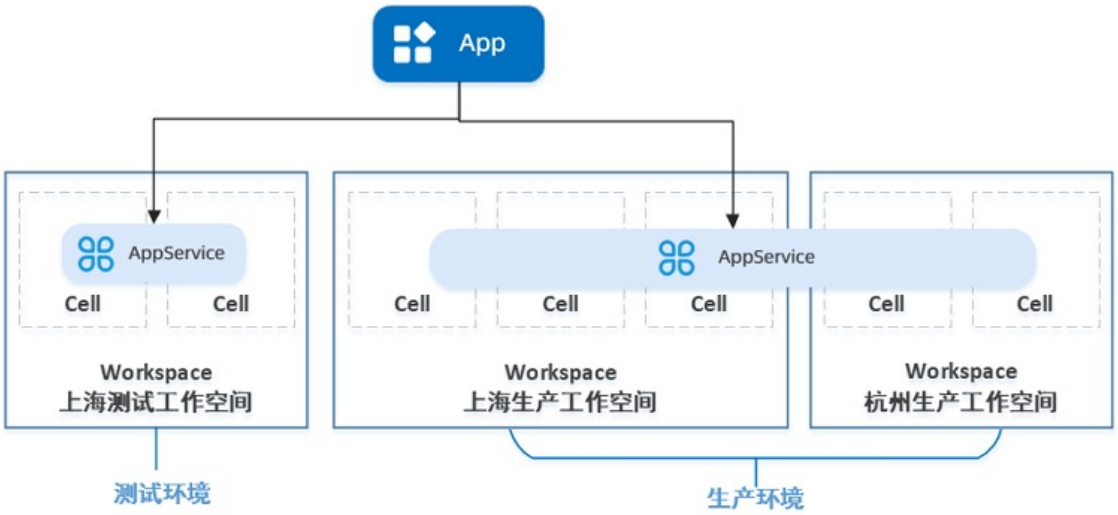
相关操作请参见：管理部署单元。

应用服务（AppService）与部署单元（Cell ）

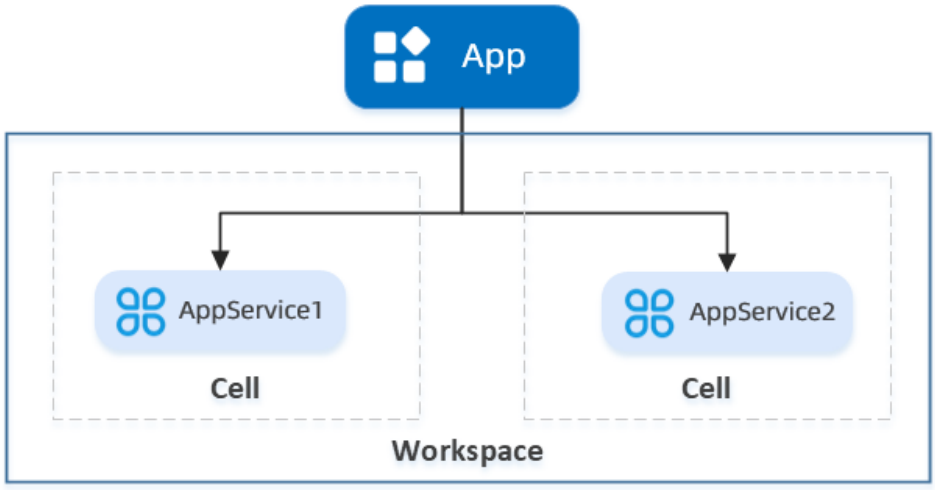
- 一个 App 的 AppService 可以跨 Cell。适用于双机房双活及蓝绿发布场景。



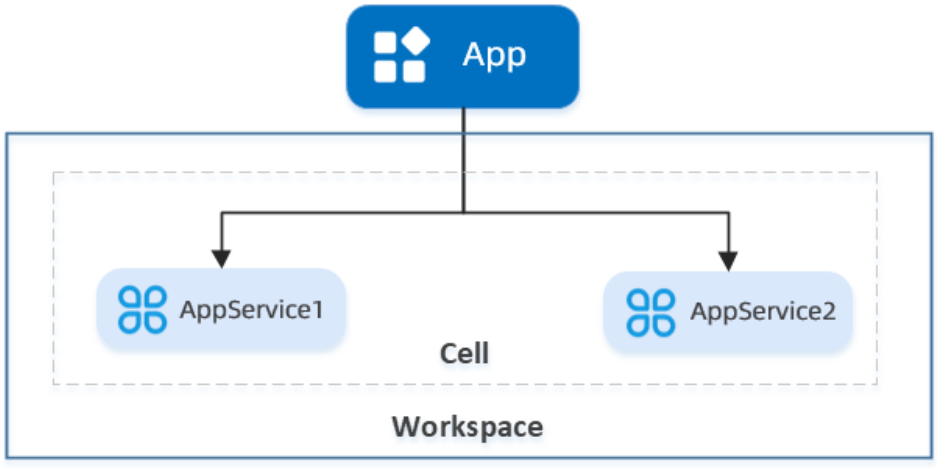
- 一个 App 的 AppService 可以跨 Cell，也可以跨 Workspace，适用于异地多活单元化场景。



- 一个 App 可以创建多个 AppService 放到多个 Cell 内。适用于多分支独立研发场景及单元化场景（假设一个 unit 内只有一个 cell）。



- 同一 App 创建多个 AppService 部署在同一个 Cell。适用于使用 sofaRouter 的联调场景，或者不依赖配置中心的普通 App 场景。



PaaS 层领域模型与 Kubernetes 领域模型映射

PaaS 发展到现阶段，与 Kubernetes 的关系越来越紧密，Kubernetes 是一个很好的技术平台，但是 Kubernetes 不等于 PaaS，PaaS 层的领域模型本质是为了 PaaS 产品提供的业务语义服务。

为了让基于 Kubernetes 实现的 SOFAStack 平台能够继承前面所述 PaaS 层领域模型支撑的复杂场景和架构。我们有必要对 Kubernetes 与 PaaS 层领域模型进行一定的规约和映射。

集群（Cluster）与租户（Tenant）

公有云与专有云上，Tenant 对一个 Kubernetes Cluster 有不同的约束：

- 公有云上的一个 Cluster 只能属于唯一一个租户，而一个租户可以有多个 Cluster。
- 专有云的一个 Cluster 往往是多租户共享的。

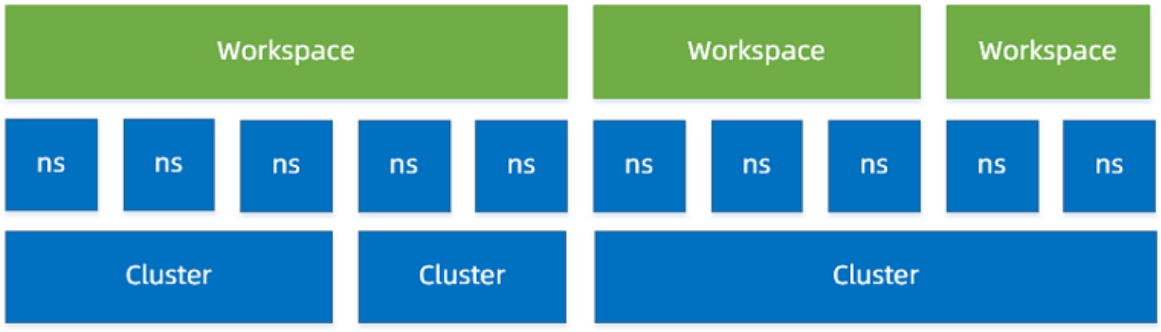


相关操作请参见：创建集群。

集群（Cluster），命名空间（Namespace）与工作空间（Workspace）

Cluster 与 Workspace 为多对多关系：一个 Cluster 可以被多个 Workspace 共享，一个 Workspace 也可以拥有多个 Cluster。

- **公有云**：一个 Workspace 可以包含多个 Cluster。且因为一个 Workspace 只能对应到一个 VPC， 所以多个 Cluster 必须属于相同的 VPC。一个 Cluster 的集群管理，比如创建和初始化，都在 Workspace 内完成。
- **专有云**：一个 Cluster 的集群管理，并不在 Workspace 内部完成，会由全局统一管理，将一个 Cluster 分配给一个 Workspace。

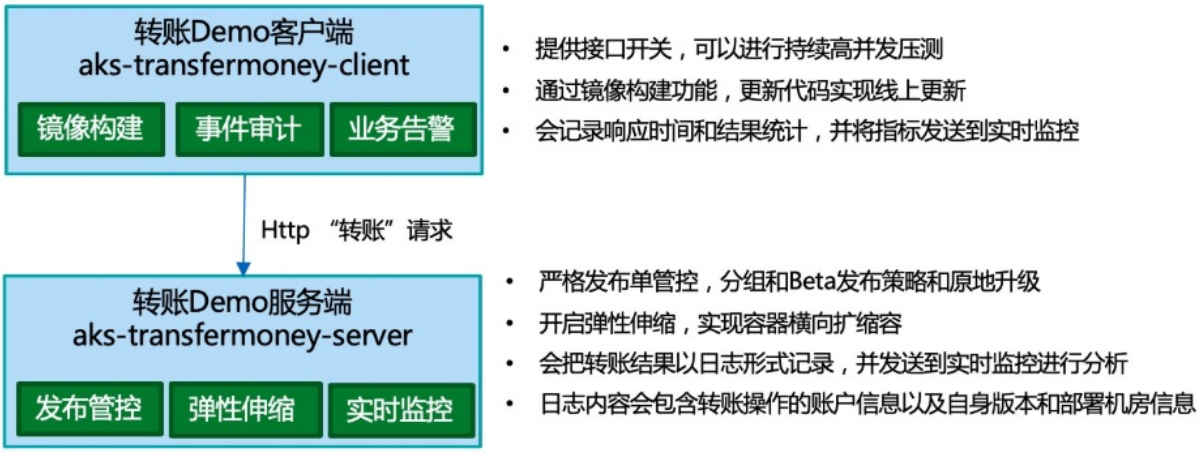


命名空间（Namespace）是对一组资源和对象的抽象整合。在同一个集群内可创建不同的命名空间，不同命名空间中的数据彼此隔离，使得它们既可以共享同一个集群的服务，也能够互不干扰。一个 Namespace 只能关联到一个 Workspace 中。

相关操作请参见：创建集群。

4.快速入门

本文以转账应用的发布、变更及监控为例，介绍运维管理模块中的各个组件的功能。



前提条件

- 已创建 VPC 环境。
- 已创建集群。

步骤一：创建转账应用

在 **应用管理** 中创建名为 transfer money 的应用。技术栈选择 SOFABoot 2.0 版本。具体步骤参见 创建应用。

步骤二：创建应用服务

在 **容器应用服务** 中创建以下 2 个应用服务，参见 创建应用服务。

1. 创建转账服务端的应用服务，名为 **aks-transfermoney-server**。
   涉及到的配置信息如下：
   - 基本信息和 Pod 模版配置：
     - CPU：200m Core - 500m Core，Memory：512 MB - 1 GB；
     - image：
       使用示例镜像 aks-transfermoney-server，镜像地址：
       ```
       registry-vpc.cn-shanghai.aliyuncs.com/sofa-samples/aks-transfermoney-server:20190726095608708-master.40215e87
       ```
       。
     - 环境变量配置：`APPSVC_VERSION = V1`，`ANTCLOUD_SOFA_PROFILE = prod`。
     - 健康检查配置：Readiness Probe 配置检查方法为 **HTTP GET 方法**，访问路径为 `/health/readiness`，访问端口 `8080`。其余项按照输入框提示配置。
     - 日志服务配置：选择已有或新建日志库，日志路径
       `/home/admin/release/run/logs/aks-transfermoney-server/digest/*.log`。
     - 业务实时监控：监控目录 `/home/admin/release/run/logs`、
       `/home/admin/release/run/logs/aks-transfermoney-client/digest/transfermoney.log`。

- 弹性配置：Server 开启副本弹性扩缩容功能，最小副本数 1，最大副本数 4。根据 CPU 使用率 50% 进行弹性扩缩容。即在副本数允许的情况下，确保 CPU 使用率在 50% （实际使用 / 请求资源量） 水位之下。可添加多个指标，如匹配内存共同使用，弹性规则 replica count 取最大期望值进行下发操作。
- 访问配置：Server 使用 ClusterIP Service，允许集群内访问。端口映射设置为 TCP 协议，访问端口 80，容器端口 8341。需记录下这个 Service Name，需要以环境变量形式配置到客户端应用。
- 发布配置：都保持默认设置。

> **说明**
> 该应用服务配置编辑完成提交后，进行版本备注，方便区分版本。其他设置参见 创建应用服务。

2. 创建转账客户端的应用服务，名称为 **aks-transfermoney-client**。
   涉及到的配置信息如下：
   - 基本信息和 Pod 模版配置：
     - CPU：200m Core - 500m Core，Memory：512 MB - 1 GB；
     - image：使用示例镜像 aks-transfermoney-client，镜像地址：

       ```
       registry-vpc.cn-shanghai.aliyuncs.com/sofa-samples/aks-transfermoney-client:20190815134744553-master.5ea6278b
       ```

       。
     - 环境变量配置：变量名称为 `SERVER_TM_ADDRESS`，值为 Server 端 Service Name。
     - 业务实时监控：监控目录 `/home/admin/release/run/logs`。
   - 弹性配置：Client 使用固定副本数 1。每个 client 能够压测并发上限为 1000。
   - 访问配置：Client 绑定公网 SLB Service，前端端口 80，后端端口 8341。其余为默认值。
   - 发布配置：保持默认设置。

> **说明**
> 该应用服务配置编辑完成提交后，进行版本备注，方便区分版本。其他设置参见 创建应用服务。

步骤三：创建发布单发布应用

创建发布单。

> **说明**
> 选取两个应用服务并进行串行发布。在高级配置中设置 aks-transfermoney-client 依赖于 aks-transfermoney-server。

步骤四：客户端触发批量操作

1. 在 client 应用服务详情页面的 **服务** 页签中获取到刚刚创建的 SLB Service 公网 IP 地址。
2. 在本地执行以下 curl 命令，以 1000 并发启动客户端进行批量调用服务。

> **说明**
> 若客户端实例有多个，则多调用几次， SLB 默认是轮询配置。本示例为 1 个客户端实例。

```
# 启动,最高并发1000
# 转账client因高并发 (max 1000)导致实例宕机，系统会自动重启容器应用
 curl -X POST "http://$Money_Client_SERVICE_IP/webapi/transferclient/start?concurrency=100
0&minamount=100&maxamount=100"

# 统计
 curl http://$Money_Client_SERVICE_IP/webapi/transferclient/gauges

# 停止
 curl -X PUT "http://$Money_Client_SERVICE_IP/webapi/transferclient/stop"
```

```
C:\Users\ >curl -X POST "http:// /webapi/transferclient/start?concurrency=1000&minamount=100&max
amount=100"
{"data":"Start the transfer money Successfully!","success":true}
C:\Users\ >curl http://139.224.138.203/webapi/transferclient/gauges
{"data":"{\n\t\"504\":9629\n}","success":true}
C:\Users\ >curl -X PUT "http://139.224.138.203/webapi/transferclient/stop"
{"data":"Successfully stop the money transfer process.","success":true}
```

步骤五：创建业务监控大盘

1. 在 **实时监控** 的 **自定义监控** 页面，创建如下各种维度的监控项：

   > **说明**
   > 监控项配置完成后，约 2 至 3 分钟后可以查看监控数据。单击左侧导航栏 **自定义监控**，然后单击您新增的监控项，可查看统计值。具体步骤请参见 创建自定义监控。

   - **服务端响应请求的容器所在 Cell**：

     > **说明**
     > 具体操作步骤请参见 秒级统计_多 Key。

     - 监控类型：秒级统计/多 Key。
     - 监控配置名：输入监控项名称，例如：transferserver_cell。
     - 所属应用：选择本教程中创建的 Server 应用服务。
     - 采集日志文件：选择 Server 应用服务中指定的日志路径。例如：`/home/admin/release/run/logs/aks-transfermoney-server/digest/transfermoney.log` 。
     - 列值分组：指定分组列为 Cell。
     - 统计方式： 求值方式为求行数。

   - **应用版本**：

     > **说明**
     > 具体操作步骤请参见 秒级统计_多 Key。

     - 监控类型：秒级统计/多 Key。
     - 监控配置名：输入监控项名称，例如：transferserver_Version。
     - 所属应用：选择本教程中创建的 Server 应用服务。
     - 采集日志文件：选择 Server 应用服务中指定的日志路径。例如：`/home/admin/release/run/logs/aks-transfermoney-server/digest/transfermoney.log` 。
     - 列值分组：指定分组列为 Version。

- 统计方式： 求值方式为求行数。

- **客户端记录的 HTTP 访问结果码**：

> 说明
> 具体操作步骤请参见 秒级统计_多 Key。

  - 监控类型：秒级统计/多 Key。
  - 监控配置名：输入监控项名称，例如：transferclient_httpstatus。
  - 所属应用：选择本教程中创建的 client 应用服务。
  - 采集日志文件：选择 client 应用服务的日志路径。例如：`/home/admin/release/run/logs/aks-transfermoney-client/digest/transfermoney.log` 。
  - 列值分组：指定分组列为 HttpStatus。
  - 统计方式： 求值方式为求行数。

- **转出账户名称按分钟的 Top 统计**：

> 说明
> 具体操作步骤请参见 分钟统计 Top。

  - 监控类型：分钟统计 Top。
  - 监控配置名：输入监控项名称，例如：transfer_top_From。
  - 所属应用：选择本教程中创建的 Server 应用服务。
  - 采集日志文件：选择 Server 应用服务中指定的日志路径。例如：`/home/admin/release/run/logs/aks-transfermoney-server/digest/transfermoney.log` 。
  - 列值分组：指定分组列为 From。
  - 统计方式： 最大排名数为 10，求值方式为求行数。

- **转入账户名称按分钟的 Top 统计**：

> 说明
> 具体操作步骤请参见 分钟统计 Top。

  - 监控类型：分钟统计 Top。
  - 监控配置名：输入监控项名称，例如：transfer_top_To。
  - 所属应用：选择本教程中创建的 Server 应用服务。
  - 采集日志文件：选择 Server 应用服务中指定的日志路径。例如：`/home/admin/release/run/logs/aks-transfermoney-server/digest/transfermoney.log` 。
  - 列值分组：指定分组列为 To。
  - 统计方式： 最大排名数为 10，求值方式为求行数。

- **分钟级客户端调用统计**：

> 说明
> 具体操作步骤请参见 分钟统计_多 Key。

- 监控类型：分钟统计/多 Key。
- 监控配置名：输入监控项名称，例如：transferclient_status_minutes。
- 所属应用：选择本教程中创建的 client 应用服务。
- 采集日志文件：选择 client 应用服务的日志路径。例如：`/home/admin/release/run/logs/aks-transfermoney-client/digest/transfermoney.log` 。
- 列值分组：指定分组列为 status。
- 求值方式：对列值求平均。
- 求平均列：latency。

2. 在 **自定义监控** 页面，新建 TransferMoney 大盘，配置如下：
   - 名称：TransferMoney 大盘。
   - 大盘布局：单击新增报表，添加上一步中创建的各类监控项。具体配置参数参见 大盘。
     可以将自定义的 TransferMoney 大盘添加到 关键业务大盘 中，方便查看各个维度进行统计的信息：



步骤六：监控告警

针对某个监控项或者所有监控项进行报警规则设置和报警订阅。

1. 设置通知组。
2. 配置报警规则。
   例如：以下配置对分钟级客户端调用统计 （transferclient_status_minutes）监控项，配置报警规则，即一旦延迟 （所有 HTTP code）或错误（非 HTTP 200）在 1 分钟内 latency 时延指标到达一定程度后就报警。



3. 订阅自定义监控报警。

步骤七：验证应用服务弹性能力

1. 在本地执行以下 curl 命令，以 1500 并发启动客户端进行批量调用服务。

```
 curl -X POST "http://$Money_Client_SERVICE_IP/webapi/transferclient/start?concurrency=150
0&minamount=100&maxamount=100"
```

2. 查看 Server 应用服务的实例数扩容情况。
   之前对 Server 应用服务的弹性配置为 1 ~ 4。client 端压力启动后，Server 实例数很快会达到最大实例数 4。实时监控大盘亦开始展示相关数据趋势。上限并发到一定程度会引发 OOM（内存不足）。
3. 在 Server 应用服务详情页，单击 **快捷操作** > **扩缩容**，扩容到 10 个实例，单击 **提交**。提交后会自动生成发布单，单击 **整体发布** 进行发布。
   可以看到此次发布单因为只进行弹性配置变更，会很快生效。
4. 在本地执行以下 curl 命令，停止 client 端调用。

```
curl -X PUT "http://$Money_Client_SERVICE_IP/webapi/transferclient/stop"
```

5. 查看 Server 应用服务的实例数缩容情况。
   Server 实例数很快会跌回最小实例数 1。实时监控大盘亦开始展示相关数据趋势。

5.应用管理

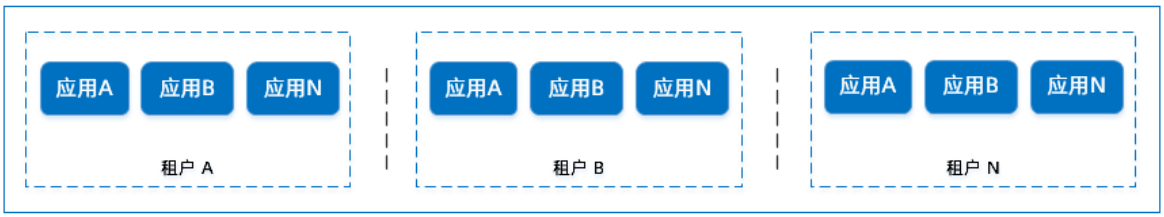
5.1. 什么是应用管理

应用管理为您提供了应用基本信息的管理以及基于 Git 的代码库托管服务。
其关键功能如下：

- 应用管理：应用管理是 SOFAStack 的基础服务之一，为用户提供了应用基本信息的管理以及基于 Git 的代码库托管服务。
- 应用分组管理：应用分组管理提供了垂直维度的应用区分功能，例如有业务关联的应用可以设置为同一分组。
- 应用分级管理：应用分级管理提供了水平维度的应用区分功能，例如从系统对业务的重要性来区分不同的应用。
- 元数据管理：在创建或者编辑应用时，给应用添加自定义的元数据，便于应用分类管理。
- 技术栈管理：技术栈定义了应用发布部署和运维时应用的依赖信息。例如，对典型的 Web 应用系统来说，常见的依赖信息包括服务器系统和版本、Web 服务器类型和版本、应用启动脚本等。

租户隔离

应用、应用分组、应用分级和技术栈均以租户的维度进行隔离。



工作空间共享

应用、应用分组、应用分级和技术栈均在租户内所有 工作空间 共享。



5.2. 快速开始

本文中将创建一个技术栈为 SOFABoot 类型的应用 demo。

前提条件

已有应用管理的权限。

创建应用

1. 登录应用管理控制台。
2. 在左侧导航栏单击 **应用列表** 进入应用列表页面。
3. 单击右上角 **创建应用**，填写如下应用信息。
   - **应用名称**：应用名称允许包含字母、数字、连字符、下划线，且必须以字母开头、以字母或者数字结尾，应用名称长度不能超过 60 个字符。此处填写 `demo` 。

     > 说明
     > - 应用名称在同一租户内必须是唯一的。
     > - 应用负责人默认为当前操作用户。

   - **技术栈**：选择 SOFA Boot。可根据技术栈名称以及版本名称搜索。
   - **应用分组**：选择应用所属分组。若尚未自定义分组，可以选择系统默认分组。
   - **应用标签**：应用分级标签，便于进行应用快速检索，支持添加系统标签和自定义标签。
   - **应用描述**：应用描述的长度不能超过 500 个字符。选填，如 `这是一个示例应用` 。
4. 单击 **提交**。
   在提示创建成功的弹出窗口中，单击 **查看应用详情** 即可进入应用详情页面查看刚创建的应用。

5.3. 使用指南

5.3.1. 管理应用

新版应用管理为用户提供以应用为中心的全局视角，您可以在新版应用管理中查看和管理应用相关资源，比如服务器资源、部署情况、发布记录等。
管理应用包括以下任务：

- 创建应用
- 添加应用服务实例
- 管理应用发布包
- 发布应用
- 查看应用详情
- 编辑应用信息
- 搜索应用
- 收藏应用
- 删除应用

前提条件

已有应用管理的权限。

创建应用

本文中将创建一个技术栈为 SOFABoot 类型的应用 demo。

1. 登录应用管理控制台。
2. 在左侧导航栏单击 **应用列表** 进入应用列表页面。
3. 单击右上角 **创建应用**，填写如下应用信息。
   - **应用名称**：应用名称允许包含字母、数字、连字符、下划线，且必须以字母开头、以字母或者数字结尾，应用名称长度不能超过 60 个字符。此处填写 `demo` 。

     > 说明
     > - 应用名称在同一租户内必须是唯一的。
     > - 应用负责人默认为当前操作用户。

   - **技术栈**：选择 `SOFA Boot` 。可根据技术栈名称以及版本名称搜索。
   - **应用分组**：选择应用所属分组。若尚未自定义分组，可以选择系统默认分组。
   - **应用标签**：应用分级标签，便于进行应用快速检索，支持添加系统标签和自定义标签。
   - **应用描述**：应用描述的长度不能超过 500 个字符。选填，如 `这是一个示例应用` 。
4. 单击 **提交**，在提示创建成功的弹出窗口中单击 **查看应用详情** 即可进入应用详情页面查看刚创建的应用。

添加应用服务实例

> 说明
> 目前不支持在应用列表中直接删除应用服务。发布应用服务前，可单击应用服务名称进入对应的详情页面进行删除。

创建应用后，您可以在 **应用详情** > **部署信息** 页面，直接为应用添加不同类型的应用服务实例。

- **经典应用服务**：单击 **添加应用服务**，跳转到经典应用服务产品的创建应用服务实例页面，创建步骤参见 应用服务实例。
- **容器应用服务**：单击 **添加应用服务**，在弹出页面中填写以下信息后单击 **确认**，通过应用商店提供的模板快速创建应用服务。
  - **工作空间**：选择应用服务所属的工作空间。
  - **命名空间**：选择应用服务所属的命名空间。
  - **应用服务名称**：选填，系统会自动生成。
  - **应用服务模版**：选择应用商店提供的应用服务模板。
  - **应用服务参数**：以下参数选填，系统会自动注入参数值。
    - app：应用名称。选填。
    - loginName：登录名。选填。
    - access：中间件的访问控制 key。由系统自动注入。
    - secret：中间件的访问控制 secret。由系统自动注入。
    - endpoint：AntVIP 地址，中间件基于 AntVIP 寻址。由系统自动注入。
    - instanceid：中间件实例唯一标识，由系统自动注入。
- **单元化应用服务**：单击 **添加应用服务**，在弹出页面中填写信息后单击 **提交**，快速创建应用服务。创建步骤参见 创建应用服务。

管理应用发布包

若要选择经典发布方式，创建应用后，您可以在 **部署信息** > **经典应用服务** 页单击右侧的 **发布包管理**，跳转到对应页面，进行发布包的上传。具体操作参见 发布包管理。

发布应用

在为应用选择创建了应用服务之后，您可以立即进行应用发布部署。应用列表页面为发布容器应用服务和经典应用服务均提供了发布的入口。

- 若发布经典应用服务，在 **部署信息** > **经典应用服务** 页单击右侧的 **发布包管理**，跳转到对应页面，进行发布包的上传，然后单击右上角的 **发布部署**，在弹出框中选择要发布的服务和发布包版本，单击 **立即发布** 或选择 **高级选项** 进行发布单配置，详情参见 应用部署。
- 若发布容器应用服务，在 **部署信息** > **容器应用服务** 页单击右侧的列表视图图标，在对应的应用服务右侧的操作栏中单击 **发布**，跳转到 **发布预览** 页，单击 **开始发布**，在弹出框中单击 **确定**，跳转到 **发布单详情** 页，单击 **整体发布**。更多信息参见 发布单。
- 若发布单元化应用服务，在 **部署信息** > **单元化应用服务** 页单击右侧的列表视图图标，在对应的应用服务右侧的操作栏中单击 **发布**，跳转到 **发布预览** 页，单击 **开始发布**，在弹出框中单击 **确定**，跳转到 **发布单详情** 页，单击 **整体发布**。更多信息参见 发布单。

若要同时发布多个应用，可从 查看应用详情 中的 **发布记录**，进入发布单页面。

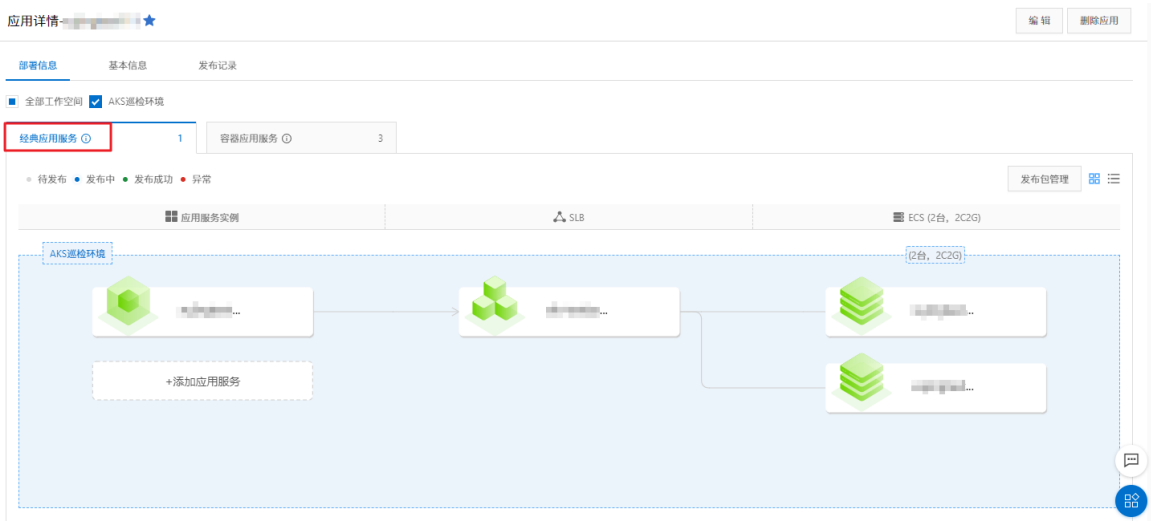
查看应用详情

在 **应用详情** 页面，您可以了解应用当前的详细信息。
应用详情分为三块内容：**部署信息**、**基本信息**、**发布记录**。

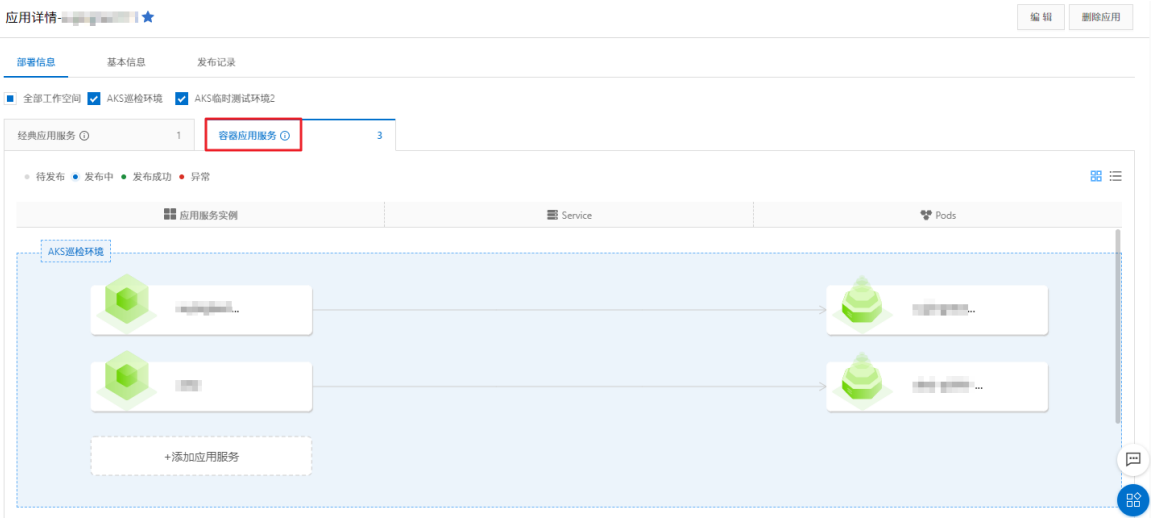
- **部署信息**：查看应用的部署情况，根据不同的应用发布方式，分为经典应用服务、容器应用服务和单元化应用服务。选择相应的页签去查看应用的部署详情。

> **说明**
> 您可以勾选不同的工作空间展示对应工作空间下的应用部署情况。

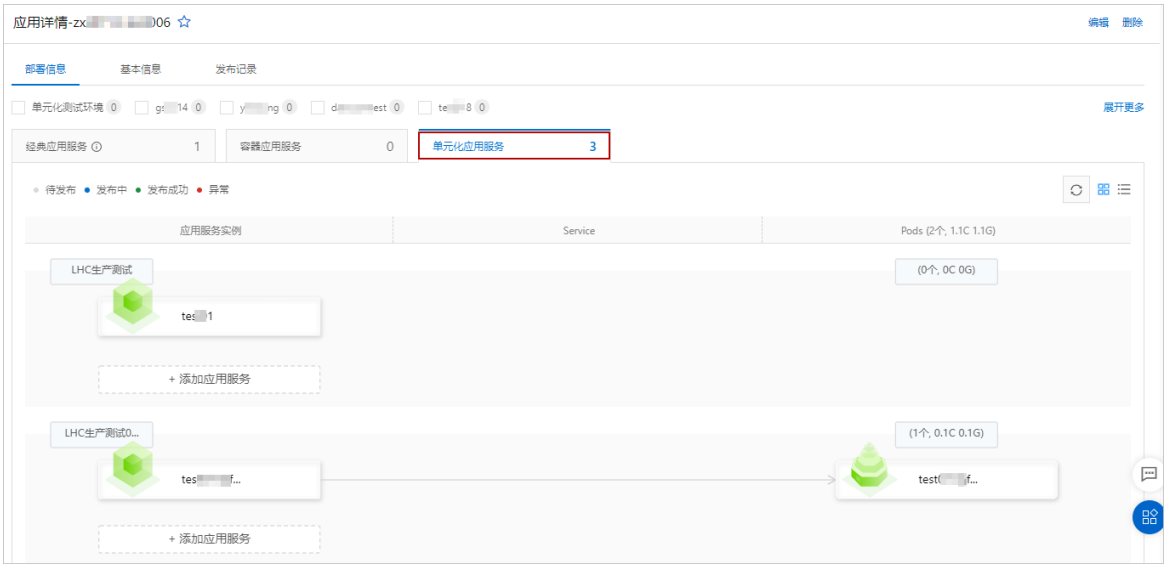
  - 通过传统虚拟机发布的经典应用服务部署信息，示例如下：



- 通过容器镜像发布的容器应用服务部署信息，示例如下：



- 通过容器镜像发布的单元化应用服务部署信息，示例如下：



- **基本信息**：查看应用的基本元数据信息，包含应用名称、技术栈、负责人、描述、创建时间、标签。
- **发布记录**：查看应用的发布记录。根据不同的应用发布部署方式，分为经典应用服务和容器应用服务。也可以单击 **创建发布单**，跳转到相应的创建发布单页面，直接开始发布应用。发布经典应用服务，参见 应用部署；发布容器应用服务，参见 发布单；发布单元化应用服务，参见 发布单。

编辑应用信息

在 **应用详情** 页面，单击右上角的 **编辑**，在 **编辑应用** 页面，您可以修改以下信息：

- **应用分组**：选择应用所属分组。若尚未自定义分组，可以选择系统默认分组。
- **应用标签**：应用分级标签，便于进行应用快速检索，支持添加系统标签和自定义标签。
- **应用描述**：应用描述的长度不能超过 500 个字符。选填，如 `这是一个示例应用` 。

搜索应用

在新版应用管理页面，您可以在左侧应用结构树上方的搜索框内输入应用名称，然后单击查找图标搜索应用。



收藏应用

在新版应用管理首页，您可以在全部应用下搜索到目标应用，将鼠标悬停在目标应用上，然后单击出现的收藏图标（☆）。

删除应用

在 **应用详情** 页面，单击右上角的 **删除应用**，在弹出的对话框中单击 **确认**。

> 说明
> 若应用下还存在关联的应用服务，则无法删除。

5.3.2. 管理应用分组

应用分组提供了垂直维度的应用区分功能，例如有业务关联的应用可以设置为同一分组。 **应用分组** 以树形结构直观地展示分组之间的层级关系。

> 说明
> 在新版应用管理中，必须先创建应用分组，并且在创建应用时设置应用的分组信息。为了兼容旧版的应用管理，系统提供默认的 **未分组应用** 列表将未设置分组信息的应用统一展示。

创建应用分组

可以通过设置一级分组指定应用分组之间的层级关系，每个分组只能指定一个一级分组，一个一级分组下可以有多个二级分组。

> 说明
> 同一层级的应用（指具有相同父分组的应用分组，没有父分组的全部属于同一层级）分组不允许重名。

创建一级应用分组

1. 登录应用管理控制台，单击左侧导航栏中的 **应用列表** 进入应用列表页面。
2. 在左侧的应用分组树中单击 **添加分组** ，弹出的新建分组对话框中输入以下信息后，单击 **确定**。

   > 说明
   > 应用分组负责人默认为当前操作用户。

   - **应用分组名称**：必填， 请确保同一层级应用分组不重名，否则会更新失败。

- **应用分组描述**：选填。

创建二级应用分组

1. 登录应用管理控制台，单击左侧导航栏中的 **应用列表** 进入应用列表页面。
2. 在左侧的应用分组树中，找到要添加二级分组的一级分组，单击右侧的添加图标，弹出的新建分组对话框中输入以下信息后，单击 **确定**。



- **应用分组名称**：必填， 请确保同一层级应用分组不重名，否则会更新失败。
- **应用分组描述**：选填。

编辑应用依赖关系

您可以在应用分组中，设置组内所有应用的组内和跨组依赖关系。

1. 登录应用管理控制台，单击左侧导航栏中的 **应用列表** 进入应用列表页面。
2. 在左侧的应用分组树中，单击要编辑依赖关系的分组，单击右侧的**编辑依赖关系**，在弹出的对话框中选择组内各应用要依赖的其他应用，单击 **确定**。

> 说明
> 对于跨组的依赖和被依赖关系，只显示一级。依赖关系图中的组内应用用蓝色图标表示，组外应用用灰色图标表示。



删除应用分组

您可以删除不再使用的应用分组。删除分组前，请先移除该分组下的应用或其它分组。

1. 登录应用管理控制台，单击左侧导航栏中的 **应用列表** 进入应用列表页面。
2. 在左侧的应用分组树中，单击要删除的分组，单击右侧的 **删除分组**。
3. 在弹出的对话框中单击 **确认**。

搜索应用分组

您可以在输入框中输入要搜索的应用分组名称，单击搜索图标进行搜索。



5.3.3. 管理技术栈

管理技术栈提供了灵活强大的自定义技术栈能力，支持部署任意的自定义技术栈及对已有技术栈的灵活扩展，以满足您的定制需求。应用管理通过技术栈来管理应用在发布部署时需要安装的软件包以及要执行的部署脚本。目前已经提供了 SOFA Lite、Spring 等常用的系统技术栈供您使用。当然您也可以通过自定义技术栈来定制更多个性化的需求。
本文档将以创建 Spring Boot 技术栈为例，介绍如何自定义技术栈。更多内容，参见 技术栈使用指南。
您可以在 **技术栈管理** 页面进行以下操作：

> **说明**
> 仅可对自定义的技术栈实例进行以上所有操作，系统提供的技术栈实例仅支持克隆与查看。

- 创建技术栈实例
- 发布技术栈实例
- 废弃技术栈实例
- 克隆技术栈实例
- 删除技术栈实例
- 查看技术栈实例

创建自定义技术栈实例

1. 进入应用管理控制台，单击左侧导航栏中的 **技术栈管理**，进入技术栈管理页面。
2. 在技术栈管理页面中，单击左上方的 **创建** 按钮创建技术栈实例。
3. 填写以下技术栈信息后，单击 **确定**。
   - **技术栈**：选择所要创建的技术栈框架。
   - **技术栈版本**：输入版本前缀，后缀默认为租户名。
   - **支持操作系统**：支持 CentOS 6.5 64bit、CentOS 7.2 64bit、CentOS 7.6 64bit、CentOS 7 64bit 操作系统。
   - **调试模式**：勾选调试模式意味着在发布技术栈后，仍可以对该技术栈进行任意修改，从而方便您对技术栈在云上进行集成测试，但调试模式的技术栈无法共享给任何租户。发布之后取消调试模式则无法再次开启调试模式，取消调试模式之后，技术栈可重新分享给其他租户。
   - **描述**：输入关于技术栈的描述，长度不能超过 500 个字符。
   - **技术栈包**：单击 **上传文件** 上传本地技术栈包。
   - **环境参数**：单击 **添加** 输入环境参数。

   创建成功的技术栈处于 **未发布** 状态。

发布技术栈实例

1. 进入应用管理控制台，单击左侧导航栏中的 **技术栈管理**，进入技术栈管理页面。
2. 在技术栈管理页面中，单击 **未发布** 状态的技术栈实例右侧的 **更多 > 发布**。
3. 在弹出框中，单击 **确定**。
   发布成功的技术栈处于 **已发布** 状态。

废弃自定义技术栈实例

1. 进入应用管理控制台，单击左侧导航栏中的 **技术栈管理**，进入技术栈管理页面。
2. 在技术栈管理页面中，单击技术栈实例右侧的 **更多 > 废弃**。
3. 在弹出框中，单击 **确定**。
   发布成功的技术栈处于 **已废弃** 状态。

克隆技术栈实例

1. 进入应用管理控制台，单击左侧导航栏中的 **技术栈管理**，进入技术栈管理页面。
2. 在技术栈管理页面中，单击技术栈实例右侧的 **更多** > **克隆**。
3. 在弹出框中修改技术栈的信息后，单击 **确定**。
   发布成功的技术栈处于 **未发布** 状态。

删除技术栈实例

您仅可删除未使用的自定义技术栈实例。

1. 进入应用管理控制台，单击左侧导航栏中的 **技术栈管理**，进入技术栈管理页面。
2. 在技术栈管理页面中，单击技术栈实例右侧的 **更多** > **删除**。
3. 在弹出框中，单击 **确定**。

查看技术栈实例

1. 进入应用管理控制台，单击左侧导航栏中的 **技术栈管理**，进入技术栈管理页面。
2. 在技术栈管理页面中，单击技术栈实例右侧的 **更多** > **查看**。
3. 进入技术栈实例详情的只读页面，查看具体信息。

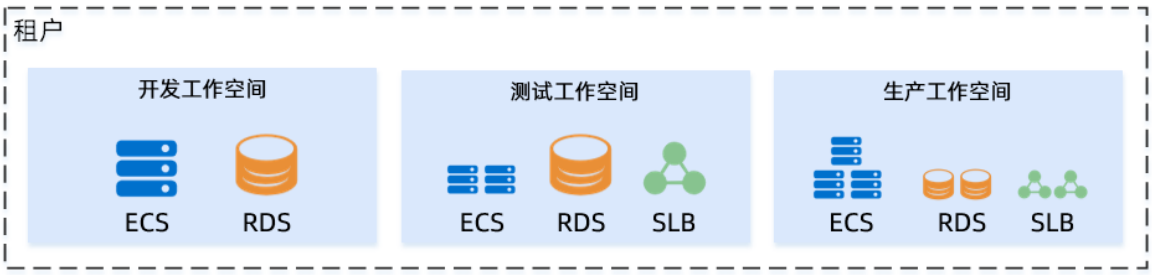
系统技术栈版本说明

SOFAStack 提供了 SOFABoot 技术栈（2.0 架构）及 NodeJS 等常用的系统技术栈供您使用。相关信息，请参见 技术栈版本说明。

6.资源管理

6.1. 什么是资源管理

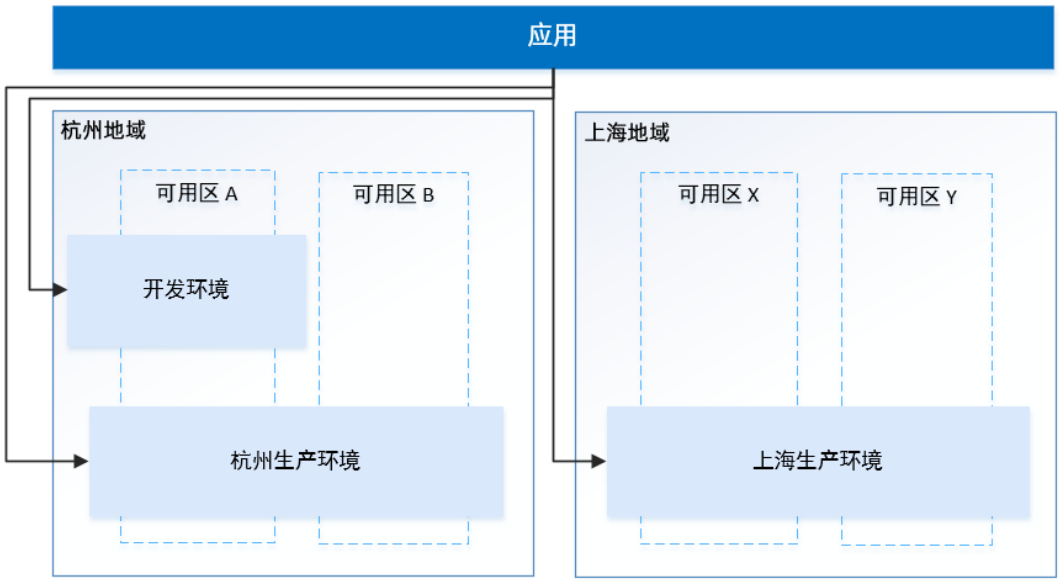
资源管理（Resource Management Center，简称 RMC）通过将基础资源（如：云服务器 ECS、云数据库 RDS、对象存储 OSS、云数据库 Memcache、日志服务等）与工作空间进行绑定，帮助您按照工作空间来管理资源，以满足在复杂的大规模研发及生产的工作场景之下，更加细粒度的运维管理要求。
租户、工作空间、资源的关系，如下图所示：



6.2. 工作空间

6.2.1. 什么是工作空间

SOFAStack 在阿里云基础资源上，增加了工作空间（Workspace）的概念，一个工作空间的本质，是指“网络互通、安全策略一致、访问延时极小”的一组资源。您可以通过工作空间方便地将资源进行分组管理，例如：根据不同的研发交付需求，将工作空间划分为开发工作空间、测试工作空间、生产工作空间等。同时，不同工作空间中的资源互相隔离，可以为每个工作空间分配单独的操作员权限进行管理。
为支持高可用（High Availability），工作空间可以跨可用区（Zone），但不能跨地域（Region）。每个工作空间下都需要为应用分配相应的资源，不同工作空间可以发布不同的应用版本。应用、工作空间、地域、可用区的关系如下图所示：



完成工作空间创建后，您可以对相应工作空间中的基础资源进行统一管理和运维。另外，可以在各工作空间中使用更多的 SOFAStack 产品，如：经典应用服务、实时监控、密钥管理服务等。

工作空间隔离机制

不同工作空间中的资源互相隔离，专有网络工作空间中的资源通过专有网络（VPC）来进行隔离，每一个工作空间对应一个 VPC，相当于每一个工作空间都是一个独立的私有网络，逻辑上彻底隔离。

工作空间资源绑定

工作空间资源绑定是指将购买、创建的 IaaS 资源导入至目标工作空间，按工作空间进行分组的过程。

专有网络下资源的导入方式：绑定专有网络（VPC）至当前工作空间，VPC 下的 ECS 实例、磁盘、快照、安全组、私网 SLB 会自动导入至当前工作空间。

> **警告**
> 对于在仅支持专有网络类型（VPC）的地域（Region）创建的工作空间，必须绑定专有网络后方可使用。

除上述资源外，其他资源包括 RDS、OTS、OCS、OSS 等，都可以进入工作空间资源管理中各个资源的列表页面，进行手动导入。

也可以将已导入的资源从目标工作空间中移出，移出后资源和工作空间解除关系。此时资源并没有被删除，您还可以在阿里云基础资源控制台管理被移出的资源，也可以再次将其导入至某个工作空间。

资源元数据同步

基础资源导入至目标工作空间后，阿里云基础资源的元数据信息会被定时同步至 SOFAStack 平台，例如在阿里云控制台修改某一 ECS 实例规格，资源管理中已导入的该 ECS 实例规格会在定时同步后发生变化，目前的数据同步周期在 5 分钟左右。

推荐的做法是资源的准备，包括购买、修改规格等，提前在阿里云控制台完成操作，并在工作空间资源管理中统一导入至目标工作空间，完成导入后在工作空间资源管理中进行常规的运维操作。需要避免两边同时操作，减少因为数据同步延迟等不一致情况下产生的非预期结果。

6.2.2. 管理工作空间

您可以在 **全局设置** 中进行当前租户的所有工作空间的管理，包括进行查看、添加、编辑、删除工作空间等操作。



创建工作空间

> **说明**
> 本文以创建标准工作空间为例，单元化工作空间的创建步骤参考 创建单元化工作空间。

前提条件

您拥有工作空间管理员权限。

操作步骤

1. 登录控制台，在左侧导航栏单击左下角的 **全局设置** 图标，进入工作空间页面。

2. 单击 **创建工作空间**，选择 **标准工作空间** 类型，单击 **创建**。
   - **标准工作空间**：标准工作空间（Workspace）是 SOFAStack 提供的一种组织机制，用于将服务于不同目标、阶段的资源分组隔离管理。您可以根据研发运维流程，为每个阶段分配工作空间，例如单机房（即单可用区）的开发工作空间、双机房（即包括两个可用区）的生产工作空间。
   - **单元化工作空间**：在标准工作空间的基础上提供了单元化能力，可用于同城双活及异地容灾场景，本质是一组标准工作空间的集合。您可以通过单元化工作空间组对用户资源进行隔离，不同工作空间组下的集群彼此隔离。输入以下基本信息。
3. 在 **创建工作空间** 页面，输入以下基本信息。
   - **工作空间标识**：工作空间的英文标识，全局唯一，一经确定无法修改，例如：dev、test、prod 等。
   - **工作空间名称**：工作空间的显示名称，例如：开发工作空间、测试工作空间、生产工作空间。
   - **地域**：工作空间所在的地域（Region），一个工作空间一定属于某个地域。
   - **网络类型**：默认为当前工作空间的网络类型。仅支持专有网络类型 VPC。
   - **导入 VPC**：开启后，您可以复用已有的 VPC 资源，将工作空间与已创建的 VPC 绑定。
   - **可用区**：工作空间配置的可用区（Available Zone），最多不能超过两个，工作空间只能导入已配置可用区内的资源，且其他产品如经典应用服务、容灾等，基于工作空间的可用区来做分组发布与容灾切换等操作。推荐配置两个可用区，为支持双机房高可用等架构做准备。
4. 单击 **下一步**，进入 **创建 VPC** 页面。
   - 若您在上一步中开启了 **导入 VPC**，则无需在该步骤创建 VPC，该页面仅展示将导入的 VPC 信息。
   - 若您没有在上一步中开启 **导入 VPC**，则需要在该步骤创建 VPC，并填写以下信息：
     - **专有网络名称**：由 [2, 128] 个英文或中文字符组成，必须以大小字母或中文开头，可包含数字、下划线（_）或连字符（-），不能以 http:// 和 https:// 开头。推荐与工作空间名称相同。
     - **专有网络网段**：专有网络的网段，一旦选择便无法更改，专有网络内的所有资源，如 ECS、RDS、SLB 的私网 IP 都在该网段内。可选网段如下：
       - 10.0.0.0/9
       - 172.16.0.0/12
       - 192.168.0.0/16
     - **交换机**：需为每个可用区创建一个交换机，同一 VPC 内的不同交换机之间内网互通。单击 **添加交换机**，并填写以下信息：
       - **名称**：输入交换机名称。
       - **可用区**：选择交换机所在可用区。同一 VPC 内不同可用区的交换机内网互通。建议您将交换机部署在不同可用区内，这样可以实现跨可用区容灾。
       - **自定义网段**：默认关闭。开启后，需填写网段地址。交换机的网段可以和其所属的 VPC 网段相同或者是其 VPC 网段的子网。
       - **子网掩码**：**自定义网段** 关闭时，需分别选择子网掩码和网段地址。默认专有的网段掩码是 16 位，例如 172.31.0.0/16，最多可提供 65536 个私网 IP 地址。范围为 16~29 位之间，可提供 4~65532 个地址。
       - **描述**：可选。输入交换机的描述信息。
5. 单击 **下一步**，进入 **创建安全组** 页面。单击 **添加安全组**，填写以下信息：
   - **安全组名称**：输入 2-128 个中英文字符，只能包含字母、数字、短横线以及“.”，必须以字母开头。
   - **描述**：可以为空，或输入 2-256 个中英文字符，不能以 `http://` 和 `https://` 开头。
6. 单击 **确定** 后返回 **创建安全组** 页面，可以继续 添加安全组规则，也可以直接单击 **提交**。

切换当前工作空间

在控制台的任何一个页面，页面右上方显示的名称（如 default）为当前工作空间。将鼠标移至这里，可以从工作空间列表中选择一个工作空间将其切换为当前工作空间。

删除工作空间

> 说明
> - 确保当前工作空间不是您要删除的目标工作空间。否则，切换至其它任一工作空间。
> - 更多信息，请参考 删除工作空间。

删除资源

删除工作空间前，需要删除工作空间中的所有资源与应用发布包。

1. 切换至目标工作空间。
2. 进入 **资源管理 > 工作空间** 页面，查看该工作空间中的资源信息总览。
3. 依次删除该工作空间中的资源。

删除应用发布包

1. 切换至目标工作空间。
2. 进入 **经典应用服务> 应用发布 > 发布包管理** 页面，查看该工作空间中的发布包列表。
3. 依次删除该工作空间中的发布包。

删除工作空间

1. 进入 **全局配置** 页面，选择要删除的工作空间，单击 **删除**。
2. 在弹出框中单击 **确定**。

6.2.3. 设置专有网络

工作空间资源需要至少与一个专有网络进行绑定，完成绑定后，该专有网络内的所有资源自动属于该工作空间。

您可以在添加工作空间时，为工作空间添加并绑定专有网络，也可以在工作空间添加完成后再进行绑定。

- 添加工作空间（专有网络）时设置专有网络
- 添加工作空间后设置专有网络

添加工作空间（专有网络）时设置专有网络

1. 录入工作空间基本信息 时选择 **专有网络**，单击 **下一步**。
2. 在 **工作空间绑定** 页面，选择 **添加专有网络并绑定至当前工作空间**。
3. 在 **设置专有网络** 页面，输入 **专有网络名称** 与 **专有网络网段**，单击 **确定**。
   - **专有网络名称**：由 [2, 128] 个英文或中文字符组成，必须以大小字母或中文开头，可包含数字、下划线（_）或连字符（-），不能以 `http://` 和 `https://` 开头。推荐与工作空间名称相同。
   - **专有网络网段**：专有网络的网段，一旦选择便无法更改，专有网络内的所有资源，如 ECS、SLB 的私网 IP 都在该网段内。可选网段如下：
     - 10.0.0.0/9
     - 172.16.0.0/12
     - 192.168.0.0/16

> 说明
> 此步骤中添加的是一个新的专有网络，里面尚未有任何资源。您需要在购买资源时，选择此专有网络，或将已有的资源导入到这个专有网络。参考 创建交换机 和 导入云服务器 ECS。

添加工作空间后设置专有网络

选择进入 **工作空间** 页面，在 **工作空间绑定专有网络信息** 区域：

- 单击 **添加并绑定专有网络**，输入 **专有网络网段**，单击 **确定**。
  专有网络名称 默认设定与工作空间名称一致。
- 单击 **绑定已有专有网络**，选择已有专有网络，单击 **绑定**。
  默认勾选 **用当前工作空间名称重命名绑定的专有网络**。

6.3. 计算与网络

6.3.1. 云服务器 ECS

6.3.1.1. 添加 ECS

云服务器 ECS（Elastic Compute Service，简称 ECS）是一种弹性可伸缩的计算服务，助您降低 IT 成本，提升运维效率，使您更专注于核心业务创新。使用 ECS 服务，您必须拥有一个 ECS 实例。本文介绍如何通过控制台添加云服务器 ECS。

前提条件

具备开发者权限。

操作步骤

1. 登录资源管理控制台。
2. 在左侧导航栏单击 **计算和网络** > **云服务器 ECS**，单击 **添加**。
3. 在 **云服务器 ECS 创建** 页面，完成以下配置：
   - **基本配置**：
     - **可用区**：从下拉菜单中选择可用区。
   - **网络**：
     - **网络类型**：提示当前的网络类型。
     - **专有网络**：提示当前的专有网络名称。仅专有网络环境中出现。
     - **交换机**：选择当前可用区下的交换机来分配私网 IP，也可以到管理控制台新建交换机。仅专有网络环境中出现。
     - **安全组**：选择加入当前可用区下的安全组。您也可以到管理控制台新建安全组。
   - **规格配置**：
     - **系列**：选择系列。
     - **规格**：选择规格。
   - **镜像**：
     - **镜像类型**：提示当前的镜像类型。
     - **操作系统**：选择操作系统。
   - **存储**：
     - **系统盘**：选择存储的系统盘和容量大小。

- **数据盘**：添加数据盘，可以添加 4 个数据盘。

- **实例信息**：
  - **实例数量**：填写实例数量。
  - **实例名称前缀**：输入实例名称。 您可以选择是否勾选 **实例名称前缀后自动添加序号，生成唯一实例名**：
    - 当您不勾选此选项时，当您建多个 ECS 实例，实例名称前缀后自动从 0 开始追加。
    - 当您勾选此选项时，后台会查询当前可用 ID 并追加，确保实例名唯一。
  - **实例描述**：输入实例描述。

- **安全设置**：
  - **设置 Root 密码**：如果您要设置 Root 密码，勾选 **请牢记您所设置的密码，如遗忘可登录 ECS 控制台重置密码**，并在下方输入 Root 密码并二次确认。

4. 单击 **创建**。

更多常用操作

新添加或导入的 ECS 会出现在 ECS 列表中。在 ECS 列表页，还提供了一些简单的运维功能：

- 单击 **操作** 栏中的 **编辑** 修改 ECS 实例的名称与描述。
- 单击 **操作** 栏中的 **远程连接**，您可以使用远程管理终端登录 ECS 服务器。
- 单击 **操作** 栏中的 **更多**，选择以下运维功能：
  - **分配给应用**：将 ECS 分配给某个应用，参考 分配与释放 ECS。
  - **从应用中释放**：选择要释放的 ECS 实例，参考 分配与释放 ECS。
  - **关机**：关机 ECS。
  - **重启**：重启 ECS。
  - **重置密码**：修改登录 ECS 的密码。
  - **重置 hostname**：将 ECS 的 hostname 重置为 ECS 名称。重启后生效。
  - **删除**：删除 ECS 实例。
  - **安全组设置**：将 ECS 添加到安全组。资源必须至少属于一个安全组。
  - **任务日志**：查看实例的操作日志。

6.3.1.2. 分配与释放 ECS

在购买或导入 ECS 实例后，您还可以将 ECS 实例分配至应用，或从应用中释放。本文介绍如何通过控制台分配或释放云服务器 ECS。

前提条件

已创建应用。

分配 ECS 实例

1. 登录资源管理控制台。
2. 在左侧导航栏单击 **计算和网络** > **云服务器 ECS**，在 **云服务器 ECS** 标签页中，找到需要分配至应用的 ECS 实例。
3. 您可以分配单个 ECS 实例至应用，也可批量分配多个 ECS 实例至同一个应用。
   - 单个分配：单击实例右侧的 **更多** > **分配给应用**，在弹出的 **分配给应用** 窗口中输入 **应用** 以及 **应用服务实例**。
   - 批量分配：在左侧勾选多个实例，并单击实例列表上方的 **更多** > **分配给应用**，在弹出的统一分配确认窗口中单击 **确定**，随后在 **分配给应用** 窗口中输入 **应用** 以及 **应用服务实例**。

> 说明
>
> - 如果应用没有服务实例，您可以直接分配到应用，此时会创建一个与应用技术栈最新版本关联的新服务实例。
> - 如果应用有服务实例，则您必须指定要分配的服务实例。
>
> 关于应用与应用服务实例之间的关系，以及如何创建应用服务实例，参考 应用服务实例。

4. 单击 **确定**。
   完成分配后，该 ECS 实例分配至的应用名称会显示在 **服务实例绑定** 列中。

释放 ECS 实例

1. 登录资源管理控制台。
2. 在左侧导航栏单击 **计算和网络** > **云服务器 ECS**，在 **云服务器 ECS** 标签页中，找到需要从应用中释放的 ECS 实例。
3. 单击该实例右侧的 **更多** > **从应用中释放**，系统会弹出释放操作确认窗口，以避免误操作。
4. 在确认窗口中，单击 **确定**。
   释放完成后，在该ECS 实例的 **服务实例绑定** 列中，应用名也将消失。

6.3.1.3. VPC 网络下导入 ECS 实例

ECS 实例支持导入功能。您可以将相关 ECS 实例导入到工作空间中，快速创建或恢复 ECS 实例。本文介绍如何通过控制台基于 VPC 网络导入云服务器 ECS。

前提条件

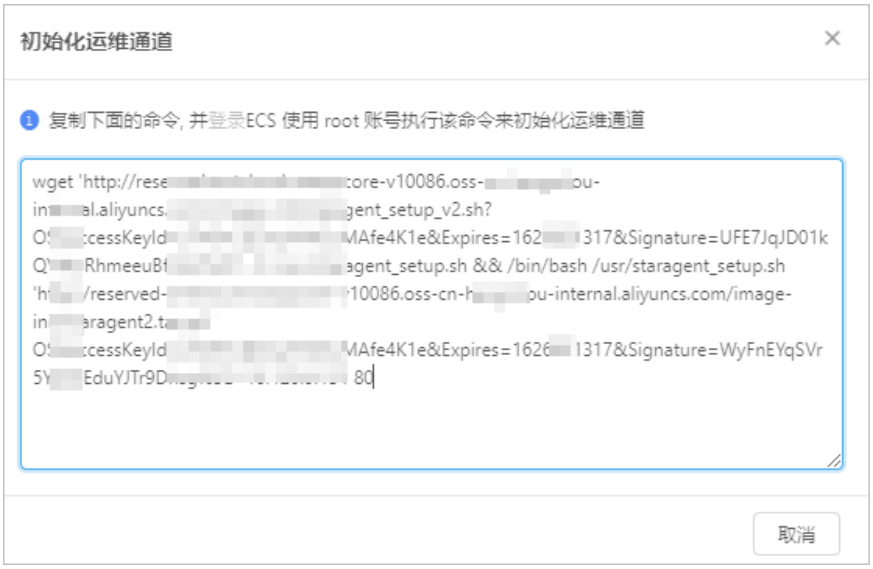
已在同一 VPC 下创建机器。

操作步骤

1. 登录资源管理控制台。
2. 在左侧导航栏单击 **计算和网络** > **云服务器 ECS**。
3. 在 ECS 实例列表页，单击 **导入**。
4. 在 **VPC 网络** 页签下，选择需要导入的实例，单击 **初始化运维通道** 列中的 **检验状态**。



5. 单击 **初始化运维通道** 列中显示的 🔧，将显示 **初始化运维通道** 弹框。



6. 复制弹框内的命令，前往虚拟机终端执行该命令。
7. 勾选要导入的 ECS 实例，单击 **导入**。

> **说明**
> 可通过选择可用区快速筛选出可导入的实例，或者直接在搜索框中输入实例 ID 或实例名称进行查询。

完成导入后，该实例会显示在实例列表中。

6.3.1.4. 经典网络下导入 ECS 实例

ECS 实例支持导入功能。您可以将相关 ECS 实例导入到工作空间中，快速创建或恢复 ECS 实例。本文介绍如何通过控制台基于经典网络导入云服务器 ECS。

前提条件

已创建工作空间，且设置的网络类型为经典网络。

操作步骤

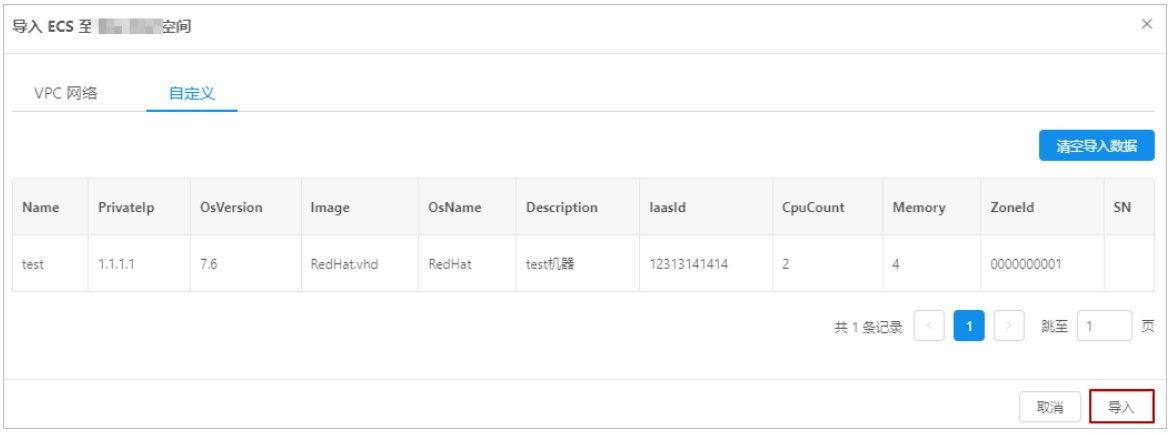
1. 登录资源管理控制台。
2. 在左侧导航栏单击 **计算和网络** > **云服务器 ECS**。
3. 在 ECS 实例列表页，单击 **导入**。
4. 在 **自定义** 页签下，单击右上角 **下载模板**，将模板文件下载到本地。



5. 根据模板填写数据并保存。
   需要填写的信息如下。

	A	B	C	D	E	F	G	H	I	J
1	name	privateIp	osVersion	image	osName	description	iaasId	cpuCount	memory	zoneId
2	test	1.1.1.1	7.6	RedHat.vhd	RedHat	test	12313141414	2	4	0000000001
3										

6. 在 **自定义** 页签下，单击此区域或将上述步骤中保存的 CSV 模板文件拖动到此区域，在弹框中单击 **导入**。



结果验证

导入完成后，新导入的实例会显示在实例列表页，可参考下图。



6.3.1.5. 删除 ECS

在购买或导入 ECS 实例后，您还可以将 ECS 实例删除或从工作空间中移出。本文介绍如何通过控制台删除云服务器 ECS。

前提条件

待删除的 ECS 必须不能被系统占用。

操作步骤

1. 登录资源管理控制台。
2. 在左侧导航栏，单击 **计算和网络** > **云服务器 ECS**。
3. 根据实际情况，删除 ECS 实例。
   - 删除单个实例：在 **云服务器 ECS** 标签页中找到待删除的实例，在 **操作** 列中，单击 **更多** > **删除**，在弹出的确认窗口中单击 **确定**。



- 批量删除实例：在 **云服务器 ECS** 标签页中选中多台待删除的实例，单击上方的 **删除**，在弹出的确认窗口中单击 **确定**。



完成删除后，该 ECS 实例会从实例列表中消失，其计费也将停止。

6.3.1.6. 查看磁盘详情

您可以在 **磁盘** 页查看已有的系统盘或数据盘信息。

前提条件

已添加云服务器 ECS。

操作步骤

1. 登录资源管理控制台。
2. 在左侧导航栏单击 **计算和网络 > 云服务器 ECS**。
3. 在 **磁盘** 标签页，查看以下信息：
   - 基本信息：包括名称/ID、描述、服务器（挂载点）、类型、容量、可用区及创建时间。
   - 操作：
     - **重新初始化**：对选择的磁盘执行初始化操作。
     - **修改策略**：根据需要修改 **自动快照** 和 **计算资源** 策略。
     - **任务日志**：查看磁盘的最近一次任务日志。

6.3.2. 负载均衡 SLB

负载均衡（Server Load Balancer，简称 SLB）是对多台云服务器进行流量分发的负载均衡服务，将访问流量根据转发策略分发到多台云服务器（ECS 实例）的流量分发控制服务。这样不仅扩展了应用的服务能力，而且还增强了应用的可用性。

购买与导入负载均衡实例

使用负载均衡服务，您必须拥有一个负载均衡实例。

- 如果已有一个负载均衡实例，您可以直接将其导入至当前工作空间。

> 说明
> 导入 SLB 有两种情况：
> - 在开通资源管理之前，通过阿里云控制台购买的 SLB。
> - 在资源管理中添加之后又移出绑定工作空间的 SLB。

- 如果没有，您就需要购买一个负载均衡实例。

本文将详细介绍如何购买与导入一个负载均衡实例。

购买负载均衡实例

登录控制台，在 **产品与服务** > **计算和网络** > **负载均衡** 来购买负载均衡实例。详情参考 购买负载均衡。

> 注意
> 若购买负载均衡之后修改了其计费方式，则修改之后的变配操作（适用所有参数的变配，包括带宽）需要到次日 0 点才能生效，生效前无法再做其他变配操作。

导入负载均衡实例

导入一个已有的负载均衡实例至当前工作空间。

前提条件

具备开发者权限。

操作步骤

1. 登录资源管理控制台。
2. 在左侧导航栏单击 **计算和网络** > **负载均衡 SLB**。
3. 单击左上方的 **导入** 按钮，弹出框中展示该 Region 下所有的 SLB 列表，选择未导入的 SLB 实例，单击 **确定** 完成 SLB 导入。
   导入列表中包含以下内容：
   - **名称/（ID）**：SLB 的名称与 ID。
   - **导入信息**：该 SLB 是否已导入，以及已导入的工作空间信息。
   - **网络类型**：经典网络或专有网络，专有网络的 SLB 会自动导入至其所属 VPC 绑定的工作空间中。
   - **服务地址**：该 SLB 提供的服务地址以及类型：公网、内网。
   - **创建时间**：该 SLB 的创建时间。
4. 导入后，新导入的 SLB 出现在列表中。单击 SLB 的名字可以配置负载均衡监听器、修改挂载的 ECS 权重等。
   同时，在这个列表的 **操作** 栏中，提供了一些简单的运维功能：
   - **停止**：表示停止当前 SLB 实例，流量无法进入 SLB 再转发到后端 ECS；
   - **更多** > **分配给应用**：将 SLB 分配给某个应用（依赖于您已 创建应用）。

分配与释放负载均衡实例

在购买或导入负载均衡实例后，您还可以将负载均衡实例分配至应用，或从应用中释放。

前提条件

已创建应用。

分配负载均衡实例

1. 登录资源管理控制台。
2. 在左侧导航栏单击 **计算和网络** > **负载均衡 SLB**，在 **负载均衡 SLB** 标签页中，找到需要分配至应用的负载均衡实例。
3. 您可以分配单个负载均衡实例至应用，也可批量分配多个负载均衡实例至同一个应用。

- 单个分配：单击实例右侧的 **更多 > 分配给应用**，在弹出的 **分配给应用** 窗口中输入 **应用** 以及 **应用服务实例**。
- 批量分配：在左侧勾选多个实例，并单击实例列表上方的 **更多 > 分配给应用**，在弹出的统一分配确认窗口中单击 **确定**，随后在 **分配给应用** 窗口中输入 **应用** 以及 **应用服务实例**。

> 说明
> - 如果应用没有服务实例，您可以直接分配到应用，此时会创建一个与应用技术栈最新版本关联的新服务实例。
> - 如果应用有服务实例，则您必须指定要分配的服务实例。

4. 单击 **确定** 完成分配后，该负载均衡实例分配至的应用名称会显示在 **服务实例绑定** 列中。

释放负载均衡实例

1. 登录资源管理控制台。
2. 在左侧导航栏单击 **计算和网络 > 负载均衡 SLB**，在 **负载均衡 SLB** 标签页中，找到需要从应用中释放的负载均衡实例。
3. 单击该实例右侧的 **更多 > 从应用中释放**，系统会弹出释放操作确认窗口，以避免误操作。
4. 在确认窗口中单击 **确定** 完成释放。释放完成后，在该负载均衡实例的 **服务实例绑定** 列中，应用名也将消失。

删除与移出负载均衡实例

在购买或导入负载均衡实例后，您还可以将负载均衡实例删除或从工作空间中移出。

删除负载均衡实例

1. 登录资源管理控制台。
2. 在左侧导航栏单击 **计算和网络 > 负载均衡 SLB**，在 **负载均衡 SLB** 标签页中，找到需要删除的负载均衡实例。
3. 您可以删除单个负载均衡实例，也可批量删除多个负载均衡实例。
   - 单个删除：单击负载均衡实例右侧的 **更多 > 删除**，在弹出的删除操作确认窗口中单击 **确定**。
   - 批量删除：在左侧勾选多个要删除的负载均衡实例，并单击实例列表上方的 **删除**。
     完成删除后，该负载均衡实例会从实例列表中消失，其计费也将停止。

移出负载均衡实例

1. 进入资源管理控制台，单击左侧导航栏中的 **计算和网络 > 负载均衡 SLB**，在 **负载均衡 SLB** 标签页中，找到需要移出的负载均衡实例。
2. 在左侧勾选一个或多个要移出的负载均衡实例，并单击实例列表上方的 **移出**。
   完成移出后，该负载均衡实例会从实例列表中消失。

> 说明
> 移出操作仅将负载均衡实例从当前工作空间中移出，而并未真正的删除，所以被移出的负载均衡实例仍会按照计费标准进行计费。

添加证书

1. 进入资源管理控制台，单击左侧导航栏中的 **计算与网络 > 负载均衡 SLB**，进入 SLB 列表页面。
2. 在列表上方单击 **证书** 标签页，进入证书列表页面。
3. 在列表上方单击 **添加**，并输入证书相关内容。
   证书上传后，在配置 SLB 时，就可以使用 HTTPS 协议并且选择证书了。

证书格式要求

- 证书必须是 Linux 环境下 pem 格式的证书。负载均衡不支持其他格式的证书，如是其它格式的证书需要首先转换证书格式。
- 如果是通过 root CA 机构颁发的证书，您拿到的证书为唯一的一份，不需要额外的证书，配置的站点即可被浏览器等访问设备认为可信。
- 如果是通过中级 CA 机构颁发的证书，您拿到的证书文件包含多份证书，需要人为的将服务器证书与中间证书合并在一起上传。
- 拼接规则为：服务器证书放第一份，中间证书放第二份，中间不要有空行。

> **说明**
> 一般情况下，机构在颁发证书的时候会有对应说明， 请注意规则说明。

以下为证书格式和证书链格式范例，请确认格式正确后上传：

- root CA 机构颁发的证书（格式为 Linux 环境下 pem 格式）：

```
-----BEGIN CERTIFICATE-----
MIIE+TCCA+GgAwIBAgIQU30GHIX4KsioTW1s2A2krTANBgkqhkiG9w0BAQUFADCB
tTELMAkGA1UEBhMCVVMxFzAVBgNVBAoTDlZlcmlTaWduLCBJbmMuMR8wHQYDVQQL
ExZWZXJpU2lnbiBUcnVzdCBOZXR3b3JrMTswOQYDVQQLEzJUZXJtcyBvZiB1c2Ug
YXQgaHR0cHM6Ly93d3cudmVyaXNpZ24uY29tL3JwYSoAYykwOTEvMC0GA1UEAxMm
VmVyaVNpZ24gQ2xhc3MgMyBTZWN1cmUgU2VydmVyIENBIC0gRzIwHhcNMTAxMDA4
MDAwMDAwWhcNMTMxMDA3MjM1OTU5WjBqMQswCQYDVQQGEwJVUzETMBEGA1UECBMK
V2FzaGluZ3RvbjEQMA4GA1UEBxQHU2VhdHRsZTEYMBYGA1UEChQPQW1hem9uLmNv
bSBJbmMuMRowGAYDVQQDFBFpYW0uYW1hem9uYXdzLmNvbTCBnzANBgkqhkiG9w0B
AQEFAAOBjQAwgYkCgYEA3Xb0EGea2dB8QGEUwLcEpwvGawEkUdLZmGL1rQJZdeeN
3vaF+ZTm8Qw5Adk2Gr/RwYXtpx04xvQXmNm+9YmksHmCZdruCrW1eN/P9wBfqMMZ
X964CjVov3NrF5AuxU8jgtw0yu//C3hWnOuIVGdg76626ggOoJSaj48R2n0MnVcC
AwEAAaOCAdEwggHNMAkGA1UdEwQCMAAwCwYDVR0PBAQDAgWgMEUGA1UdHwQ+MDww
OqA4oDaGNGh0dHA6Ly9TVlJTZWN1cmUtRzItY3JsLnZlcmlzaWduLmNvbS9TVlJT
ZWN1cmVHMi5jcmwwRAYDVR0gBD0wOzA5BgtghkgBhvhFAQcXAzAqMCgGCCsGAQUF
BwIBFhxodHRwczovL3d3dy52ZXJpc2lnbi5jb20vcnBhMB0GA1UdJQQWMBQGCCsG
AQUFBwMBBggrBgEFBQcDAjAfBgNVHSMEGDAWgBSl7wsRzsBBA6NKZZBIshzgVy19
RzB2BggrBgEFBQcBAQRqMGgwJAYIKwYBBQUHMAGGGGh0dHA6Ly9vY3NwLnZlcmlz
aWduLmNvbTBABggrBgEFBQcwAoY0aHR0cDovL1NWUlNlY3VyZS1HMi1haWEudmVy
aXNpZ24uY29tL1NWUlNlY3VyZUcyLmNlcjBuBggrBgEFBQcBDARiMGChXqBcMFow
WDBWFglpbWFnZS9naWYwITAfMAcGBSsOAwIaBBRLa7kolgYMu9BSOJsprEsHiyEF
GDAmFiRodHRwOi8vbG9nby52ZXJpc2lnbi5jb20vdnNsb2dvMS5naWYwDQYJKoZI
hvcNAQEFBQADggEBALpFBXeG782QsTtGwEE9zBcVCuKjrsl3dWK1dFiq3OP4y/Bi
ZBYEywBt8zNuYFUE25Ub/zmvmpe7p0G76tmQ8bRp/4qkJoiSesHJvFgJ1mksr3IQ
3gaE1aN2BSUIHxGLn9N4F09hYwwbeEZaCxfgBiLdEIodNwzcvGJ+2LlDWGJOGrNI
NM8S6xjqhJCPxYzk9buuCl1B4Kzu0CTbexz/iEgYV+DiuTxcfA4uhwMDSe0nynbn
1qiwRk450mCOnqH4ly4P4lXo02t4A/DI1I8ZNct/Qfl69a2Lf6vc9rF7BELT0e5Y
R7CKx7fc5xRaeQdyGj/dJevm9BF/mSdnclS5vas=
-----END CERTIFICATE-----
```

证书规则：

- 以 `-----BEGIN CERTIFICATE----------END CERTIFICATE-----` 做为开头和结尾。
- 请将这些内容一并上传。
- 每行 64 字符，最后一行不超过 64 字符。

- 中级机构颁发的证书链：

```
——BEGIN CERTIFICATE——
——END CERTIFICATE——
——BEGIN CERTIFICATE——
——END CERTIFICATE——
——BEGIN CERTIFICATE——
——END CERTIFICATE——
```

证书链规则：

- 证书之间不能有空行；

- 每一份证书遵守第一点关于证书的格式说明。

RSA 私钥格式要求



RSA 私钥规则：

- 以 `-----BEGIN RSA PRIVATE KEY-----` 和 `-----END RSA PRIVATE KEY-----` 作为开头和结尾。
- 请将这些内容一并上传。
- 每行 64 字符，最后一行长度可以不足 64 字符。

若通过上述方法您没有得到 `-----BEGIN RSA PRIVATE KEY-----` 和 `-----END RSA PRIVATE KEY-----` 这种样式的私钥，可以按照如下方式转换：

```
openssl rsa -in old_server_key.pem -out new_server_key.pem
```

然后将 `new_server_key.pem` 的内容与证书一起上传。

配置监听器

在资源管理中，您可以配置负载均衡实例的监听器，从而实现对每个连接按端口进行检查，再将请求流量分发至后端服务器。本节介绍如何通过资源管理来配置负载均衡实例的监听器。

配置监听器

在监听器中配置前后端协议与端口，若连接符合监听条件，则请求流量将被分发至后端服务器。具体添加步骤如下：

1. 登录资源管理控制台。
2. 在左侧导航栏单击 **计算和网络 > 负载均衡 SLB**，在 **负载均衡 SLB** 标签页中，单击需要创建监听器的负载均衡实例名，进入该实例的详情页。
3. 在 **监听器** 标签页中，单击 **添加**，在弹出的 **创建监听器** 窗口中输入以下信息：
   - **前端协议[端口]**：在下拉列表中选择 **HTTP**、**HTTPS** 或 **TCP**，并填入相应端口。
     - 一般网站选择 HTTP 协议，填入 80 端口。
     - 如果是 HTTPS 协议的网站，则填入 443 端口，并选择证书。
     - 如果是自定义协议，可选择 TCP 协议，自定义端口允许的范围是：80、443、2800-3300、5000-10000、13000-14000。

- **后端协议[端口]**：后端协议与默认前端协议一致，需要输入后端服务的监听端口。
- **转发规则**：支持 **轮询模式**、**最小连接数**。
- **峰值带宽**：默认关闭。启用后，可设置带宽峰值。
- **X-Forwarded-For**：默认开启。获取真实 IP。
- **会话保持**：默认关闭。启用后，负载均衡会把来自同一客户端的访问请求分发到同一台后端服务器上进行处理。
- **健康检查**：开启健康检查功能后，当后端某台后端服务器出现异常时，负载均衡会自动将新的请求分发到其它健康检查正常的后端服务器上；而当该后端服务器恢复正常运行时，负载均衡会将其自动恢复到负载均衡服务中。具体配置方式，参考 配置健康检查。

4. 单击 **确定**，完成创建。
该监听器会出现在列表中，并且状态为 **使用中**。

监听器访问控制

> 警告
> **访问控制设置默认关闭，一旦开启只有白名单中的 IP 才可以访问 SLB 实例的监听，请谨慎使用。如您在阿里云 SLB 产品控制台操作过访问控制权限，请直接前往阿里云 SLB 控制台管理访问控制。**

1. 登录资源管理控制台。
2. 在左侧导航栏单击 **计算和网络** > **负载均衡 SLB**，在 **负载均衡 SLB** 标签页中，单击需要创建监听器的负载均衡实例名，进入该实例的详情页。
3. 在 **监听器** 标签页中，在 **操作** 栏中单击 **访问控制** > **设置**，在弹出的窗口中打开访问控制的开关，单击 **确定**。
4. 在 **操作** 栏中单击 **访问控制** > **编辑**，在弹出的窗口中输入 IP 地址，单击 **确定**。
IP 地址之间请用英文逗号隔开，最多可输入 300 个 IP。

更多相关操作

监听器创建完成后，您可以在 **监听器** 页签对其进行 **停止**、**启动**、**编辑**、**删除**、以及 **查看任务日志** 等操作。

相关链接

什么是监听

管理后端服务器

在资源管理中，负载均衡实例的后端服务器组指一个负载均衡实例可使用的所有后端服务器。一般情况下，请求会被默认分发至后端服务器组。
您可以对负载均衡实例的后端服务器进行挂载、卸载、分配权重等操作。
进入资源管理控制台，单击左侧导航栏中的 **计算和网络** > **负载均衡 SLB**，在 **负载均衡 SLB** 标签页中，单击需要配置后端服务器的负载均衡实例名，进入该实例的详情页，然后再进行以下各操作：

挂载后端服务器

1. 在 **后端服务器组** 标签页中，选择 **待挂载**，找到需要挂载的云服务器实例，单击右侧的 **挂载**。
2. 在弹出框中设置服务器的 **流量权重**，单击 **确认**。在 **已挂载** 中可查看挂载成功的云服务器实例。

卸载后端服务器

1. 在 **后端服务器组** 标签页中，选择 **已挂载**，通过以下方式单个卸载某台服务器或者批量卸载服务器。
   - 批量卸载：勾选需要卸载的云服务器实例，单击上方的 **卸载**。
   - 单个卸载：找到需要卸载的云服务器实例，单击右侧的 **卸载**。
2. 在弹出框中单击 **确认**。在 **待挂载** 中可查看卸载成功的云服务器实例。

> 说明
> 卸载后的云服务器实例并未删除，依然可以重新挂载或在云服务器控制台中进行管理。

分配权重

在 **后端服务器组** 标签页中，可以对单台或多台已挂载的后端服务器分配权重。

修改单台权重

1. 在 **后端服务器组** 标签页中，选择 **已挂载**。
2. 找到目标服务器，单击其右侧的 **修改权重**。
3. 在 **流量权重** 中输入所需权重。
4. 单击 **确定** 完成修改。

批量修改权重

1. 勾选多台目标服务器，单击上方的 **修改权重**。
2. 在 **批量设置流量权重** 中输入所需权重，或对各台服务器分别设置 **流量权重**。
3. 单击 **确定** 完成修改。

相关链接

后端服务器概述

6.3.3. 安全组

安全组是一种虚拟防火墙，具备状态检测和数据包过滤功能，用于在云端划分安全域。您可以通过配置安全组规则，允许或禁止安全组内的 ECS 实例对公网或私网的访问。

概述

安全组是一个逻辑上的分组，这个分组是由同一个地域（Region）内具有相同安全保护需求并相互信任的实例组成。每个 ECS 实例至少属于一个安全组，在创建的时就需要指定。同一安全组内的实例之间网络互通，不同安全组的实例之间默认内网不通。可以授权两个安全组之间互访。

更多关于安全组的介绍，参见 安全组。

导入与添加安全组

您可以通过添加或导入安全组将安全组与工作空间绑定，绑定后该安全组内的所有资源自动属于该工作空间。

前提条件

如果您要创建专有网络 VPC 类型安全组，请确认您已经有可用的专有网络 VPC 和虚拟交换机。更多详情，请参见 创建专有网络。

导入安全组

1. 进入资源管理控制台，单击左侧导航栏中的 **计算和网络 > 安全组**。
2. 单击 **导入**。
3. 在弹出的对话框中，勾选要导入的安全组，单击 **确定**。

> 说明
> 一个安全组只能属于一个工作空间。弹出框中列出了该租户下所有的安全组列表，已经属于某个工作空间的安全组将不能勾选和导入。 VPC 网络下，只能导入同一工作空间中创建的安全组，不同工作空间中的安全组无法导入。

添加安全组

1. 进入资源管理控制台，单击左侧导航栏中的 **计算和网络 > 安全组**。
2. 单击 **添加**。

3. 在弹出的 **添加安全组** 对话框中，按页面提示要求输入 **名称** 与 **描述**（可选），单击 **确定**。

编辑安全组

创建完安全组后，您可以编辑安全组的名称和描述。

操作步骤

1. 进入资源管理控制台，单击左侧导航栏中的 **计算和网络** > **安全组**。
2. 找到目标安全组，单击右侧 **操作** 栏中的 **编辑**。
3. 安全组的名称和描述进入可修改状态，修改后单击 **保存**。

删除与移出安全组

若您的业务已经不再需要一个或多个安全组，您可以删除或移出安全组。

前提条件

待删除或移出的安全组内不存在 ECS 实例。如果安全组内有 ECS 实例，您需要将实例移出安全组。

删除安全组

1. 进入资源管理控制台，单击左侧导航栏中的 **计算和网络** > **安全组**。
2. 找到要删除的安全组，单击右侧 **操作** 栏中的 **删除**。在弹出框中单击 **确定**。
   安全组删除后，组内所有安全组规则同时被删除。

移出安全组

1. 进入资源管理控制台，单击左侧导航栏中的 **计算和网络** > **安全组**。
2. 找到要删除的安全组，单击右侧 **操作** 栏中的 **移出**。在弹出框中单击 **确定**。
   移出的安全组不会被删除，还可以通过导入的方式加入到工作空间中。

添加安全组规则

您可以通过添加安全组规则，允许或禁止安全组内的 ECS 实例对公网或私网的访问。

前提条件

已创建安全组。

操作步骤

1. 进入资源管理控制台，单击左侧导航栏中的 **计算和网络** > **安全组**。
2. 找到目标的安全组，单击右侧 **操作** 栏中的 **规则管理**。
3. 单击 **添加**，在弹出对话框中进行规则配置后，单击 **确定**。
   - **网卡类型**：VPC 网络下的安全组，仅支持 **内网** 网卡，经典网络下的安全组，可以设置 **内网** 或 **公网** 网卡。
   - **规则方向**：
     - 出方向：是指 ECS 实例访问内网中其他 ECS 实例或者公网上的资源。
     - 入方向：是指内网中的其他 ECS 实例或公网上的资源访问 ECS 实例。
   - **授权策略**：
     - 允许：放行该端口相应的访问请求。
     - 拒绝：直接丢弃数据包，不会返回任何回应信息。如果两个安全组规则其他都相同只有授权策略不同，则拒绝授权生效，允许策略不生效。
   - **协议类型**：TCP、UDP、GRE、ICMP、全部。端口范围和协议类型的关系，参考 添加安全组规则。
   - **端口范围**：-1/-1，表示不限制端口。
   - **授权类型**：安全组访问、地址段访问。授权类型和授权对象的关系，参考 添加安全组规则。

- **授权对象**：支持格式如 `10.15.6.8/12` 或 `10.15.6.18` 。多个用 `,` 隔开。 `0.0.0.0/0` 表示所有 IP，请谨慎设置。
- **优先级**：取值范围为 1 ~ 100。优先级数值越小，优先级越高。

删除安全组规则

如果您不再需要某个安全组规则，可以删除安全组规则。

前提条件

- 已创建安全组。
- 已 添加安全组规则。

操作步骤

1. 进入资源管理控制台，单击左侧导航栏中的 **计算和网络** > **安全组**。
2. 找到目标的安全组，单击右侧 **操作** 栏中的 **规则管理**。
3. 选择对应的规则类型，找到要删除的规则，单击右侧 **操作** 栏中的 **删除**， 在弹出框中单击 **确定**。

6.3.4. 专有网络 VPC

管理交换机

若您的工作空间属于专有网络（VPC）类型，您可以在专有网络详情页面管理交换机。同一专有网络内的不同交换机之间内网互通。

创建交换机

您可以通过创建交换机为专有网络划分一个或多个子网。

前提条件

已 创建专有网络。

操作步骤

1. 进入资源管理控制台，单击左侧导航栏中的 **计算和网络** > **专有网络 VPC**，进入 **专有网络详情** 页。
2. 在 **交换机** 页签，单击 **添加**。
3. 在弹出的 **创建交换机名称。** 对话框中，填写以下信息后，单击 **确定**。
   - **名称**：输入交换机名称。
   - **可用区**：选择交换机所在可用区。同一 VPC 内不同可用区的交换机内网互通。建议您将交换机部署在不同可用区内，这样可以实现跨可用区容灾。
   - **自定义网段**：默认关闭。开启后，需填写网段地址。交换机的网段可以和其所属的 VPC 网段相同或者是其 VPC 网段的子网。
   - **子网掩码**：**自定义网段** 关闭时，需分别选择子网掩码和网段地址。默认专有的网段掩码是 16 位，例如 `172.31.0.0/16` ，最多可提供 65536 个私网 IP 地址。范围为 16~29 位之间，可提供 4~65532 个地址。
   - **描述**：可选。输入交换机的描述信息。

编辑交换机

创建完交换机后，您可以编辑交换机的名称和描述信息。

前提条件

已创建交换机。

操作步骤

1. 进入资源管理控制台，单击左侧导航栏中的 **计算和网络** > **专有网络 VPC**，进入 **专有网络详情** 页。

2. 在 **交换机** 页签，找到要标记的交换机，单击右侧操作栏中的 **编辑**，交换机的名称和描述列进入可编辑状态，修改后单击 **保存**。

删除交换机

您可以删除不再使用的交换机。

前提条件

- 已创建交换机。
- 删除交换机前需要确保该交换机 **关联的资源实例数** 为 0，否则无法删除。

操作步骤

1. 进入资源管理控制台，单击左侧导航栏中的 **计算和网络** > **专有网络 VPC**，进入 **专有网络详情** 页。
2. 在 **交换机** 页签，找到要删除的交换机，单击右侧操作栏中的 **删除**，在弹出框中单击 **确定**。

管理自定义路由表条目

创建专有网络和交换机后，系统会自动为您创建系统路由条目来管理专有网络的流量。

添加自定义路由条目

您不能创建系统路由条目，但您可以创建自定义路由条目将指定目标网段的流量路由至指定的目的地。

前提条件

- 已 创建专有网络。
- 若要路由至其他 VPC 环境，需先创建路由器接口。

操作步骤

1. 进入资源管理控制台，单击左侧导航栏中的 **计算和网络** > **专有网络 VPC**，进入 **专有网络详情** 页。
2. 选择 **路由表条目** 页签，单击 **添加**。
3. 在弹出对话框中，填写以下信息后，单击 **确定**。
   - **目标网段**：要转发到的目标网段。必须是一个合法的 CIDR 或 IP 地址，例如：`192.168.0.0/24` 或 `192.168.0.1`。
   - **下一跳类型**：选择 **ECS 实例** 将目的地址在目标网段范围内的流量路由至选择的 ECS 实例。 选择 **路由器接口** 将目的地址在目标网段范围内的流量路由至选择的 VPC。
   - 根据 **下一跳类型** 的选择，选择以下信息：
     - **下一跳 ECS 实例**：选择目标 ECS 实例。
     - **下一跳路由器接口**：选择目标路由器接口。

删除自定义路由条目

您可以删除自定义路由条目，但无法删除系统路由条目。

前提条件

已添加自定义路由条目。

操作步骤

1. 进入资源管理控制台，单击左侧导航栏中的 **计算和网络** > **专有网络 VPC**，进入 **专有网络详情** 页。
2. 选择 **路由表条目** 页签，找到要删除的自定义路由条目，单击右侧操作栏的 **删除**，在弹出框中单击 **确定**。

6.4. 任务记录

进入 SOFAStack 控制台，点击左侧导航栏中的 **资源管理** > **任务记录** 中，可通过 简单查询 和 高级查询 快速查找对应的操作记录。

在操作记录中，您可以查看每个操作的任务日志，对执行失败的操作进行重试。

简单查询

可通过以下条件进行搜索：

- **资源类型**：选择某一资源类型，如云服务器 ECS、负载均衡 SLB、云数据库 RDS 等。
- **资源实例**：输入资源序列号。
- **操作来源**：输入操作员姓名。

高级查询

可通过以下条件进行搜索：

- **状态**：勾选 **查询异常请求**，筛选所有 **执行失败** 的操作。
- **请求时间**：选择操作时间段。
- **唯一查询**：根据 **请求序列号** 或 **幂等token** 进行精确查询。

7.经典应用服务

7.1. 概述

7.1.1. 什么是经典应用服务

经典应用服务（CAS）是以应用为核心视图，对应用的版本、发布包、资源等进行可视化、自动化管理。经典应用服务提供自动化、智能化的应用全生命周期 DevOps 支持，提升效率、降低成本、减少人为错误，让开发人员专注于业务逻辑本身。

简介

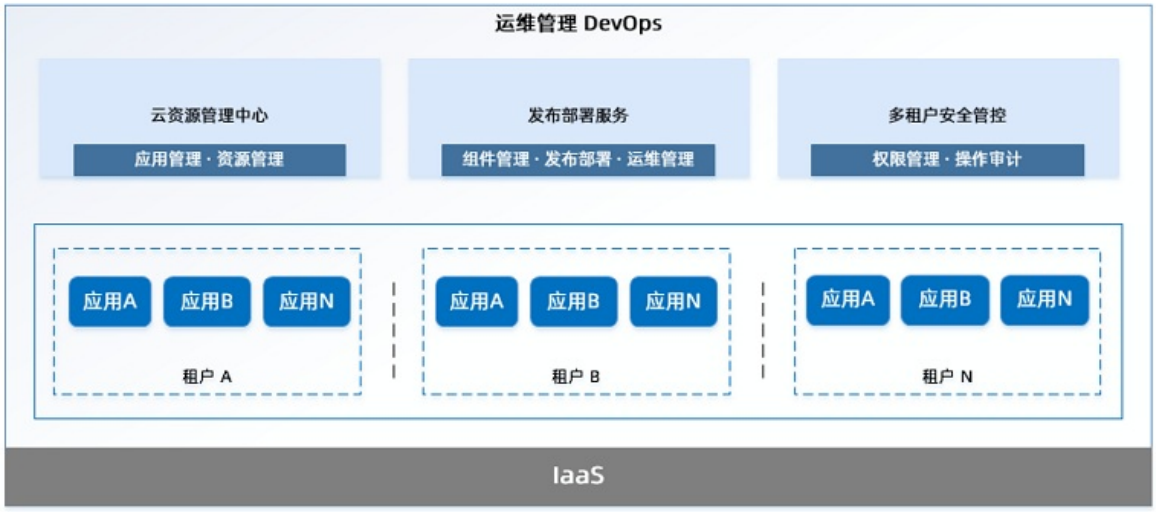
经典应用服务能够直观地展现整个应用的拓扑结构、资源状态及健康度，可以依据资源使用情况对资源进行弹性伸缩，保障底层基础设施具备更高的稳定性、可用性及性能；同时，增加了金融级安全保障和故障切换功能。

核心优势

- 以应用为核心视角的 DevOps：提供应用全生命周期的 DevOps 自动化支持，将传统的以 IT 资源为核心的管理视角转换成以应用、业务为核心视角，使用户可以专注于业务价值的同时，提升研发效率、降低人为出错的可能。
- 可定制的自动化运维：以自定义的技术栈方案，为用户提供可定制的自动化运维，提升了云平台的灵活性和对用户存量系统的兼容支持性，方便用户在平台上使用自己熟悉的、非 SOFAStack 原生提供的技术框架。
- 强大的发布部署能力：提供分组发布、Beta 发布、灰度发布、单机房发布、蓝绿发布等多种灵活的部署策略，从各种需求层面，支持可视化、自动化、可重试、可回滚的发布部署。
- 灵活的运维管道能力：提供录入，执行用户自定义的运维命令和脚本通道，方便用户做自定义的运维指令操作。

总体框架

运维管理的总体框架如下：



至顶而下，共分为三个层次：

1. 基于应用视角的管理层：
   - 虚拟化、弹性化：将底层物理资源抽象、封装成各种虚拟资源，提供可视化、自助式管理门户；实时监控资源使用情况，一旦某项资源达到设定阈值，即可根据预设规则进行弹性扩/缩容。
   - 自动化：提供应用全生命周期的 DevOps 支持，涵盖应用创建、组件安装、发布部署及运维管理的自动化服务。

- 访问控制：基于应用域、工作空间、用户角色等维度，采用逻辑隔离、安全组、白名单和认证登录等技术进行访问控制，所有操作行为可追溯。

2. 租户应用层：所有租户共享底层资源池，租户与租户间、相同租户的不同应用间实现逻辑隔离，为企业提供安全的私有空间，保障其应用不被非法访问。
3. IaaS 层：为底层基础资源层，主要包括：计算和网络，数据库和存储。具体介绍可参见 资源管理。

7.1.2. 基础术语

中文	英文	说明
工作空间	Workspace	一个工作空间的本质是指“网络互通、安全策略一致、访问延时极小”的一组资源。您可以通过环境方便地将资源进行分组管理，例如：根据不同的研发交付需求，将环境划分为开发环境、测试环境、生产环境等。同时，不同环境中的资源互相隔离，可以为每个环境分配单独的操作员权限进行管理。
应用	Application	是 SOFAStack 在阿里云基础资源之上抽象出的一个概念。一个租户内的一套代码对应一个应用，应用有对应的技术栈、分组、分级等基础元信息。应用是 PaaS 平台的核心概念，蚂蚁研发运维体系都围绕着应用来展开，研发应用、发布应用、运维应用、监控应用等。
应用服务	Application Service	又称应用服务实例，是应用（Application）在一个环境中的服务实例，包括技术栈、部署单元的信息，还包括所关联的计算资源、负载均衡以及存储资源。一个应用在一个环境中可以有多个应用服务。
资源	Resource	指代 ECS、SLB、OCS、Memcache、RDS 等。
租户	Tenant	最新的 Tenant 概念，曾用名包括 project、空间。
部署单元	Cell	使用独立的中间件服务，属于一个独立的物理可用区或者跨多个物理可用区。
技术栈	Buildpack	在 SOFAStack 平台上，是指一个应用程序所依赖的全部框架及附属资源的集合，它定义了应用发布部署和运维时的依赖信息，主要包括操作系统和版本、开发框架类型和版本、应用启动和部署脚本以及环境参数。

7.2. 快速入门

本文以 SOFABoot 技术栈 2.0 为例描述如何创建并在 SOFAStack 平台上发布一个 SOFABoot 应用。

> 说明
> SOFABoot 应用的开发、发布过程与基于 Spring 技术栈的过程类似。

步骤一：创建标准工作空间

工作空间（Workspace）本质上是“网络互通、安全策略一致、访问延时极小”的一组资源。您可以通过工作空间方便地将资源进行分组管理，例如：根据不同的研发交付需求，将工作空间划分为开发工作空间、测试工作空间、生产工作空间等。不同工作空间中的资源互相隔离。

1. 进入 SOFAStack 控制台。

2. 在左侧导航栏，单击下方的 **全局设置**，进入工作空间列表页面。
3. 单击 **创建工作空间**，选择 **标准工作空间** 类型，单击 **创建**。
4. 在 **创建工作空间** 页面，输入以下基本信息。
   - **工作空间标识**：工作空间的英文标识，全局唯一，一经确定无法修改，例如：dev、test、prod 等。
   - **工作空间名称**：工作空间的显示名称，例如：开发工作空间、测试工作空间、生产工作空间。
   - **地域**：工作空间所在的地域（Region），一个工作空间一定属于某个地域。

     > 说明
     > 目前开放的地域为 **华东 2（上海）** 和 **华东 1（杭州）**（金融云）。

   - **网络类型**：只支持专有网络 VPC。
   - **可用区**：工作空间配置的可用区（Available Zone），最多不能超过两个，工作空间只能导入已配置可用区内的资源，且发布部署、容灾等其他功能都是基于工作空间的可用区来做分组发布与容灾切换等操作。推荐配置两个可用区，为支持双机房高可用等架构做准备。
5. 单击 **下一步**，在 **创建 VPC** 页面，输入以下配置信息：
   - **专有网络名称**：由 2-128 个英文或中文字符组成，必须以大小写字母或中文开头，可包含数字、下划线（_）或连字符（-），不能以 `http://` 或 `https://` 开头。推荐与工作空间名称相同。
   - **专有网络网段**：专有网络的网段，一旦选择便无法更改，专有网络内的所有资源，如 ECS、RDS、SLB 的私网 IP 都在该网段内。可选网段如下：
     - 10.0.0.0/9
     - 172.16.0.0/12
     - 192.168.0.0/16
   - **交换机**：单击 **添加交换机**，在弹出的 **创建交换机** 窗口，填写以下信息，单击 **提交**。
     - **名称**：交换机名称。长度为 2-128 个字符，以英文字母或中文开头，可包含数字、下划线（_）和短横线（-）。系统自动生成 `工作空间标识-vsw` 组成的默认名称。
     - **可用区**：交换机的可用区。同一 VPC 内不同可用区的交换机内网互通。您需要为每个可用区创建一台交换机。
     - **自定义网段**：默认关闭。开启后，需填写网段地址。交换机的网段可以和其所属的 VPC 网段相同或者是其 VPC 网段的子网。
     - **子网掩码**：**自定义网段** 关闭时，需分别选择子网掩码和网段地址。默认专有的网段掩码是 24 位，例如 172.31.0.0/24，最多可提供 65536 个私网 IP 地址。范围为 16~29 位之间，可提供 4~65532 个地址。
     - **描述**：输入交换机的描述信息。可包含 2-256 个中英文字符，不能以 `http://` 和 `https://` 开头。
6. 单击 **下一步**，在 **创建安全组** 页面，单击 **添加安全组**，在弹出的 **添加安全组** 窗口，填写以下信息，单击 **提交**。
   - **安全组名称**：长度为 2-128 个英文或中文字符，必须以大小字母或中文开头，不能以 `http://` 和 `https://` 开头。可以包含数字、半角冒号（:）、下划线（_）或者连字符（-）。系统自动生成 `工作空间标识-sg` 组成的默认名称。

- **描述**：可包含 2-256 个中英文字符，不能以 `http://` 和 `https://` 开头。
- **规则**：保持默认设置，接受所有的出入访问。

步骤二： 创建应用

在该示例中，我们创建一个 SOFABoot 应用。具体步骤如下：

1. 进入应用管理控制台，单击左侧导航栏中的 **应用列表** 进入应用列表页面。
2. 单击 **创建应用**。
3. 在 **创建应用** 页，填写以下信息。
   - **应用名称**：允许包含字母、数字、连字符、下划线，且必须以字母开头、以字母或者数字结尾。本例中输入 **sofabootdemo**。

     > 说明
     > 应用名称在同一租户内必须唯一。

   - **技术栈**：选择 **SOFA Boot**。
   - **应用分组**：选择应用所属分组。若尚未自定义分组，可以选择系统默认分组。
   - **应用标签**：应用分级标签，便于进行应用快速检索，支持添加系统标签和自定义标签。
   - **应用描述**：输入应用的描述信息。
4. 单击 **提交**，在提示创建成功的弹出窗口中单击 **查看应用详情** 即可进入应用详情页面查看刚创建的应用。

步骤三：创建应用服务

1. 进入经典应用服务控制台，单击左侧导航栏中的 **应用发布 > 应用服务** ，单击 **创建服务实例**。
2. 在跳转的页面上，输入以下信息。
   - **应用**：选择上一步中创建的应用。
   - **服务实例名称**：允许包含字母、数字、连字符、下划线，且必须以字母开头、以字母或者数字结尾。
   - **发布包**：选择应用的发布包。可在 **应用发布 > 发布包管理** 页面为应用上传发布包。
   - **描述**：输入服务实例描述。
   - **技术栈**：根据所属应用自动展示。
   - **技术栈版本**：选择所需版本。
   - **部署单元**：勾选应用要发布到的部署单元。
3. 单击 **确定**。

步骤四：本地开发

本地使用 SOFABoot 框架进行开发详见 SOFABoot 快速开始。在这里，您可以先忽略具体的开发过程，直接 下载样例 Jar 包 并解压到本地。样例工程中的应用都已经在本地完成编译打包，生成的包在 `target` 目录下。

步骤五：购买资源并绑定资源到应用服务

本步骤将为前述创建的应用服务 sofabootdemo 购买一台 ECS 和一个公网 SLB，并将 ECS 绑定到该 SLB。

购买 ECS

1. 进入经典应用服务控制台。
2. 展开左侧 **应用发布 > 应用服务** 。
3. 找到之前创建的应用服务 **sofabootdemo**，单击该服务实例名称或单击 **详情**，进入 **应用服务实例信息** 页。
4. 单击 **云服务器 ECS** 页签中的 **添加** 按钮，根据提示录入相应信息，单击 **创建** 完成 ECS 创建。

> 注意
>
> - SOFAStack 平台定义的技术栈 2.0，如 SOFABoot，目前支持的操作系统为 Linux，镜像为 SOFAStack 平台定义的 CentOS 7.x 和 CentOS 6.x 镜像。
> - 对于技术栈 1.0，如 sofa-lite2、Spring、sofa-lite-core、sofa-lite-web，目前支持的操作系统为 Linux，镜像为 SOFAStack 平台定义的 CentOS 6.x 镜像。

购买公网 SLB

1. 在 **应用服务实例信息** 页面中，选择 **负载均衡 SLB** 页签，单击 **添加** 按钮。
2. 在创建负载均衡页面，输入相应信息，单击 **确定** > **创建**，完成 SLB 创建。
   - **实例类型**：选择 **公网**。
   - **监听器信息**：添加以下监听器：
     - **前端协议**：**HTTP**
     - **端口**：**80**
     - **后端协议端口**：**8341**
     - 其余属性保持默认即可

挂载 ECS 到 SLB

当 ECS 和 SLB 都创建成功后，根据以下步骤进行挂载：

1. 在 **负载均衡 SLB** 页签，单击刚创建的 SLB 名称进入详情页。
2. 在 **后端服务器组** 中选择 **待挂载** 标签，勾选 **同一 App**，选择之前创建好的 ECS，单击 **挂载**。
3. 在 **挂载 ECS** 窗口，设置 **流量权重** 为 **100**，单击 **确定**。
   完成挂载后，可在 **已挂载** 中查看挂载的 ECS。

步骤六：发布应用

1. 在左侧导航栏，选择 **应用服务** > **发布包管理**。
2. 找到 **sofabootdemo**，单击 **管理** 进入包管理界面，单击 **上传发布包**。
3. 在弹出框中，使用默认版本号，文件选择本地下载的样例 `.jar` 包，然后单击 **确定**。
   待进度条完成后可以看到上传的版本。
4. 单击页面右上方的 **发布部署**，然后在 **发布部署** 窗口中选择服务实例：sofabootdemo，以及刚才上传的版本，单击 **立即发布**，跳转至发布单页面，单击 **整体发布**。
   在发布单详情看到状态 **执行成功**，即表示该应用已经发布成功。

步骤七： 验证结果

您可以通过以下步骤来验证应用是否发布成功。

1. 在左侧导航栏，选择 **应用服务** > **服务实例列表**，单击列表中的 **sofabootdemo** 进入实例详情页。
2. 在 **负载均衡 SLB** 标签中复制服务地址。
3. 在本地浏览器的地址栏中访问 `http://{负载均衡器服务地址}/webapi/users/tester`，如 `http://120.27.186.140/webapi/users/tester`。
   若页面显示以下内容表示发布成功：

```
{"data":{"realName":"Real tester","userId":0,"userName":"tester"},"success":true}
```

小结

至此，您已经了解如何通过技术栈 2.0 创建并部署 SOFABoot 应用。基于其他技术栈的操作类似。本文在操作过程中没有详细讲述原理，您可以参考 应用部署 了解更多。

FAQ

Q：为什么根据文档操作最后会发布失败？

A：请检查 挂载 ECS 到 SLB 步骤中的 **流量权重** 设置，确保流量权重不为 0。

7.3. 应用部署

经典应用服务提供应用的发布、回滚功能，您可以对需要发布的应用编辑依赖关系、调整服务器分组，并选择部署策略等。

前提条件

- 应用和服务器绑定，且服务器经过初始化。
- 应用存在发布包。
- 灰度引流功能需要应用绑定 SLB 或添加至 SLB 下的虚拟服务器组。
- 通知功能要求用户在侧栏的 **设置** > **通知** 中配置钉钉接收群。

操作步骤

经典应用服务的操作分为以下几大步骤：

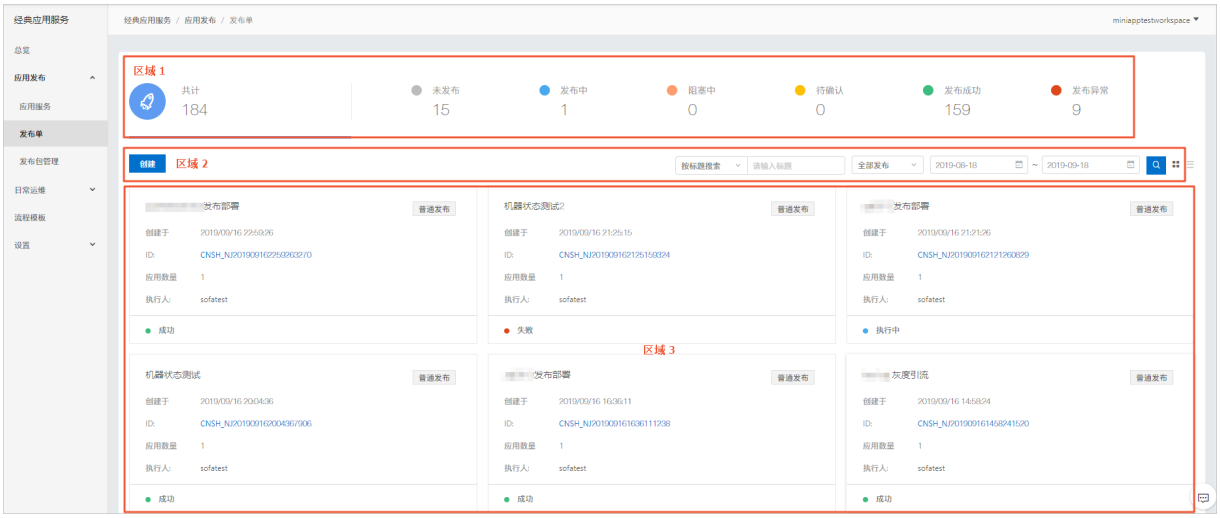
1. 进入发布部署大盘
2. 查看发布部署大盘
3. 新建并执行发布单
4. 操作发布单
5. 回滚

进入发布部署大盘

进入经典应用服务控制台，单击左侧导航栏中的 **应用发布** > **发布单** 进入发布部署大盘。

查看发布部署大盘

发布部署大盘提供总览功能，展示了 **未发布**、**发布中**、**阻塞中**、**待确认**、**发布成功** 和 **发布异常** 等状态的发布单，如下图所示。



- 区域 1：发布单大盘总览，显示发布单总体信息。单击某个状态信息，区域 3 即展示处于相应状态的发布单，否则区域 3 展示所有发布单；
- 区域 2：提供筛选功能，用户可以按照输入的条件筛选所有发布单，并在区域 3 展示符合条件的发布单；
- 区域 3：发布单信息显示区域。

新建并执行发布单

在 **发布部署大盘** 页面，单击 **创建** 按钮即进入新建发布流程。新建发布分为三个步骤：编辑基础信息、编辑高级信息、预览提交。

1. 编辑以下基础信息。
   - **标题**：自定义发布标题。
   - **类型**：选择发布方式。

     > 说明
     > 只有双机房工作空间下，才支持蓝绿发布。

   - **分组策略**：决定该发布单下所有应用服务器的分组规则，目前支持快速分组、按部署单元分组、每台一组、共分一组。
   - **使用 Beta 分组**：开启 Beta 分组后，发布时，给服务器分组设置一个特殊的 Beta 分组，即系统从每个数据中心获取一台机器，加入 Beta 分组，Beta 分组会在第一组（组 1）之前发布。
     Beta 分组发布完成后系统会自动暂停应用发布，此时系统负责人或者 PE 需要对应用的发布情况进行确认。若应用发布正常，则单击 Beta 分组确认，使应用继续分组发布。Beta 分组可以与所有分组策略共同决定分组。创建新的发布申请时，默认 **使用 Beta 分组**，此时同一个发布单上的所有应用都设置 Beta 分组。如果不需要使用 Beta 分组，则取消该选项。
   - **使用灰度引流**：针对开启 **使用灰度引流** 的发布单，系统在分组发布完成后 **不执行 SLB 引流操作**，用户可自行调整服务器的 SLB 流量权重。在发布单中的所有机器分组，都具备 Beta 分组的特性，需要用户发布完成后单击 **确认**，才真正结束当前分组的发布流程。

     > 注意
     > 只有成功完成发布的服务器才可进行 SLB 流量权重的调整，对于所有非成功态的机器实例，只能查看机器的 SLB 流量权重数据，但不得进行修改。在应用发布过程中，调整流量权重的入口对所有已完成发布的机器组开放，在整个应用发布完成后关闭。

   - **设置 SLB 引流权重**：完成发布后不再使用快照权重进行引流，而使用当前设置的权重。
   - **开启钉钉消息**：如果需要对发布单进行实时跟进，可通过开启 **开启钉钉消息** 来订阅应用及 Beta 分组的发布情况。订阅前需提前注册钉钉机器人账号，并从左侧菜单栏中选择 **设置** > **通知** 添加钉钉消息来完成配置，否则该功能无效。
   - **部署单元**：选择服务实例发布的部署单位，至少选择一个部署单元。
   - **应用列表**：用户可以在 **待选应用列表** 中单击选择需要的应用，单击 **添加** 将应用添加到 **已选应用列表** 中。

     > 说明
     > 仅那些有发布包的应用可供选择。如果要删除已选择的应用，则可以在 **已选应用列表** 中单击选择所需的应用，再单击 **删除** 即可。

     如果不需要编辑部署策略、部署参数以及多个应用之间的依赖关系，则直接单击 **快速创建** 执行应用发布。
2. 编辑高级信息。
   - **从缓存中安装技术栈**：选择是否从缓存中安装技术栈。

     > 说明
     > 仅使用技术栈 2.0 的相关应用支持该功能。

- **升级/降级技术栈**：若开启该功能，可以更改技术栈的版本（但不能更改技术栈）。
- **部署参数**：支持以下方式的参数添加方式：
  - **整体添加**：若选择整体添加，则对所有应用添加相同的参数。
  - **指定添加**：若只是在应用对应的输入框中添加，则只添加到指定的应用。
  - **批量添加**：若选择批量添加，批量添加参数支持文本格式为 `[参数名]=[参数值]`，参数名只能包含 `[A~Z、a~z、0~9、_、-、.]`，多个参数使用空格或换行隔开，例如 `k1=v1 k2=v2`。
- **依赖关系**：当选择了 2 个及以上应用时，可能需要填写应用之间的依赖关系。如果应用之间不存在依赖关系，则直接跳过。
  添加依赖关系方式分为：
  - **逐条添加**：在 **依赖于** 列中输入某个应用所依赖的应用名称（可以同时选择多个）。若应用 tmp 依赖应用 testapp，则在 tmp 对应的 **依赖于** 列中输入 testapp 即可，依赖于多个应用时同理。
  - **快速添加**：不推荐使用。用户可通过类似于 `应用名 1 > 应用名 2,应用名 3 > 应用名 4` 这个表达式编辑应用之间的依赖关系。

    > **说明**
    > 表达式中的应用名称不可重复，> 符号左边的应用依赖于右边的应用，而且 > 两边不可有空格，两个表达式之间使用半角 , 隔开。

    应用之间不可循环依赖，即不可出现“应用 A 依赖于应用 B，同时应用 B 又依赖于应用 A”这种情况。如果存在循环依赖，则单击 **下一步** 按钮时，屏幕上会报错 `存在循环依赖，请重新编辑依赖关系。`

  单击 **下一步** 按钮，进入 **预览提交** 页面。

3. 预览提交。
   进入 **预览提交** 页面，用户可以在该页面上确认所有发布申请的信息。
   如果在 **编辑高级信息** 中添加了依赖关系，那么系统会自动根据依赖关系对应用进行分组，其中不需要依赖任何其它应用的应用会被分入 A1 分组，依赖于 A1 分组中的应用会被分入 A2 分组，依次类推。在发布应用时，系统会先发布 A1 分组的应用，再发布 A2 分组的应用，依次类推。
   若用户需要修改发布顺序，则单击 **上一步** 按钮，重新编辑应用之间的依赖关系。若确认无误，则单击 **创建** 按钮，创建发布单，系统会自动跳转到发布单详情页面。

操作发布单

进入 **发布单详情** 页面，如下图所示。



- 区域 1：发布单基本信息，包括发布单标题、ID、发布状态、发布进度、应用数量等信息。

- 区域 2：发布单操作区，根据实际的发布情况，用户可以进行重新初始化、整体发布、取消发布单、开始发布、编辑发布顺序等操作。取消发布单后，不能再对该发布单进行任何操作。开始发布时，系统会根据依赖层级先发布 A1 或 B1 分组中的应用（即不依赖于其它应用的应用），再发布其它分组中的应用。
- 区域 3：根据应用发布状态筛选应用，单击每个状态，即可筛选出相应状态的应用并显示在区域 4。该区域还显示处于各个状态的应用数量。
- 区域 4：应用展示区，展示所有应用的发布进度及状态。用户可以单击 **分组发布**，执行指定的分组（只有无前置应用，或者前置应用全部发布完成的分组才能单击 **分组发布**）。

编辑发布顺序

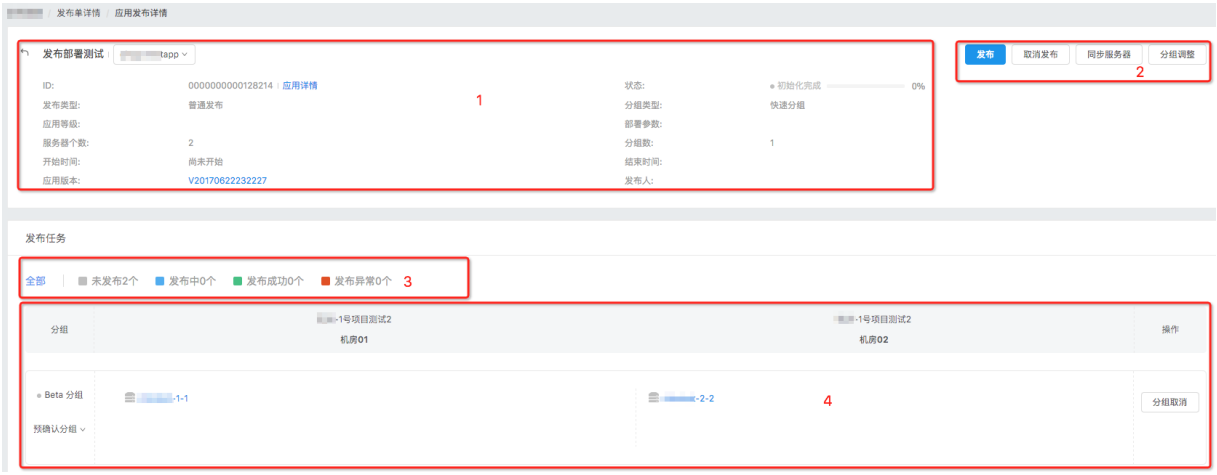
在 **发布单详情** 页面的区域 2 中单击 **编辑发布顺序**，进入 **编辑发布顺序** 页面。

- 移动应用：编辑发布顺序时，根据应用分组、串行分组，通过拖动目标应用至目标分组即可编辑应用的发布顺序。
- 添加应用分组：每个应用分组之间并行发布。添加应用分组后，即可新增一个并行发布的分组，然后再移动应用到该分组。例如：单击 **添加应用分组** 新增一个 B1 分组，然后移动应用至 B1 分组。
- 添加串行分组：单击 **+ 添加串行分组**，即可在该应用分组追加一个空的串行分组，然后再移动应用至该分组。
- 删除分组：单击每个分组右边的 **删除**，即可删除该串行分组。若一个应用分组下所有的串行分组都被删除，则该应用分组被删除。

编辑完成之后，单击 **完成并返回** 按钮，即可返回 **发布单详情** 页面。

应用发布详情

在 **发布单详情** 页面的区域 4 中单击应用名称，进入 **应用发布详情** 页面，如下图所示。



- 区域 1：显示应用基本信息。该区域会显示应用名称、ID、发布包版本、发布状态、执行进度、发布类型、服务器分组策略、服务器数量、分组数量等信息。如果有多个应用，可以直接在不同应用之间切换。
- 区域 2：应用操作区。用户可以执行发布或取消发布、同步服务器、分组调整等操作。
  - 发布：仅发布当前应用；
  - 取消发布：若取消发布，则本次发布部署中，用户不能再对该应用进行任何操作，且整个发布单的状态不受该应用的影响；
  - 分组调整：可以对服务器进行新增分组、删除分组和组间移动服务器等操作。控制同一个应用下的发布顺序；
  - 同步服务器：当需要同步服务器信息时，会显示该按钮。比如，创建发布单之后，如果有新增、删除、上线、下线机器的情况即需要同步服务器信息。
- 区域 3：服务器筛选及服务器状态信息，单击每个状态，即可根据该状态进行过滤，在区域 4 显示相应状态的服务器信息，同时该区域还显示处于各个状态的服务器数量。

- 区域 4：服务器分组区，用于展示服务器分组、服务器发布状态、发布详情，并且可以进行查看详情，分组取消等操作：
  - 详情：发布完成、正在发布、发布失败的分组，用户可以查看该分组中应用发布部署的详情，包括进度、状态、执行时间、失败原因等重要信息。

分组调整

在 **应用发布详情** 页面的区域 2 中，单击 **分组调整** 按钮，进入 **分组调整** 页面。
其中包括：重新分组、删除分组、分组调整三个功能。

- 重新分组：重新选择服务器分组策略对 ECS 进行分组。
- 删除分组：在某个分组的最右端单击 **删除** 即可删除当前分组，并将该分组的 ECS 移动到 **待分配** 组。**待分配** 组中的 ECS 不进行部署。
- 移动分组：用户可以单击某台 ECS，移动其到指定的目标分组。

完成编辑之后，单击 **完成并返回**，即回到 **应用发布详情** 页面。

SLB 灰度引流

如果应用在发布前绑定了 SLB 或绑定了 SLB 下虚拟服务器组，则在创建发布单时，可勾选使用灰度引流。
相对于不打开灰度引流开关的发布单，使用灰度引流后，系统在完成单机发布后不再自动根据快照流量恢复 SLB（及虚拟服务器组） 权重，取而代之，通过人为调整 SLB（及虚拟服务器组） 下的机器权重，实现对已成功发布的机器进行灰度引流。
单击 **调整流量权重** 后，弹出以下界面，用户可根据操作提示进行流量权重操作。需要注意的是，对机器权重的调整仅对状态为 **成功** 的机器生效，对状态为 **失败** 的机器不允许调整流量并保持为 0。

钉钉消息

在应用发布中，可配置 **钉钉消息** 订阅发布的实时通知。在应用发布开始及应用发布完成这两个阶段，接收钉钉群通知。
在左侧栏 **设置** 中选择 **通知**，进行钉钉群配置。
在创建发布单的时候，可再次确认是否勾选 **开启钉钉消息** 来订阅实时通知。若没有设置钉钉消息，勾选无效。

回滚

对于正在发布、发布失败的应用，用户可以在 **应用发布详情** 页面进行回滚。
单击 **回滚**，弹出如下对话框。



回滚条件：只有当前应用有其它成功发布的发布包（除当前版本外）时，用户才可以回滚应用。若无成功发布的其它发布包，对话框中即出现如上图所示提示。
若含有成功发布的历史版本，则界面如下。



在以上对话框中，可以填写 **回滚原因**，选择分组策略，之后再单击 **确定** 按钮，页面自动跳转至 **应用发布详情** 页面，并自动回滚应用。
在 **回滚详情** 页签，查看当前回滚进度、详细信息。
在 **上次部署** 页签，查看回滚前发布的信息。

7.4. 应用运维

应用运维提供以 **应用** 的视角进行运维的能力，包含 **应用** 和 **服务器** 两个维度，目前支持 **重启**、**上线**、**下线**、**静态资源部署**、**初始化服务器** 五种运维操作。

前提条件

- 服务器和应用绑定，且服务器状态合法。
- 对于 SOFAStack 定义的技术栈，如 **SOFA Lite Web** 、**SOFA Lite Core**、**Spring**，目前支持的操作系统仅为 **Linux**，并且是 SOFAStack 定义的 **CentOS 7.x** 镜像。

具体步骤

应用运维的操作包括以下 5 个步骤：

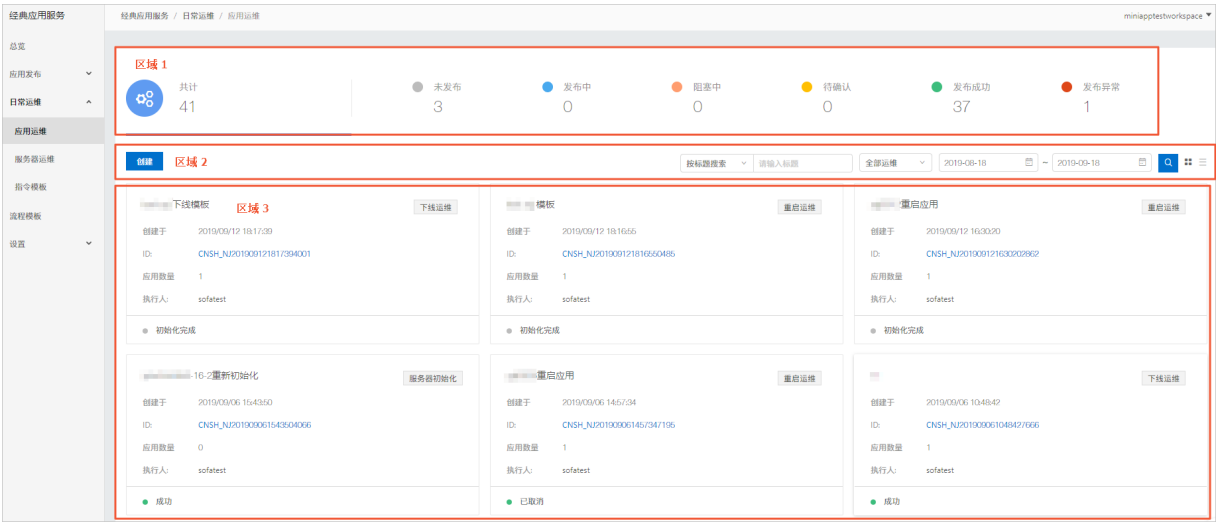
1. 进入应用运维服务
2. 查看应用运维大盘
3. 新建应用运维单并执行
4. 操作应用运维单
5. 分组调整

进入应用运维服务

进入经典应用服务控制台，单击左侧导航栏中的 **日常运维** > **应用运维** 进入。

查看应用运维大盘

应用运维大盘提供总览功能，展示了 **未发布**、**发布中**、**阻塞中**、**待确认**、**发布成功**、**发布异常** 等状态的运维单，如下图所示。



- 区域 1：运维单大盘总览，显示运维单总体信息。单击某个状态信息，区域 3 即展示处于相应状态的运维单，否则区域 3 展示所有运维单。
- 区域 2：提供筛选功能，用户可以按照输入的条件进行筛选所有应用运维单，并在区域 3 显示符合条件的运维单。
- 区域 3：运维单信息显示区域。

新建应用运维单并执行

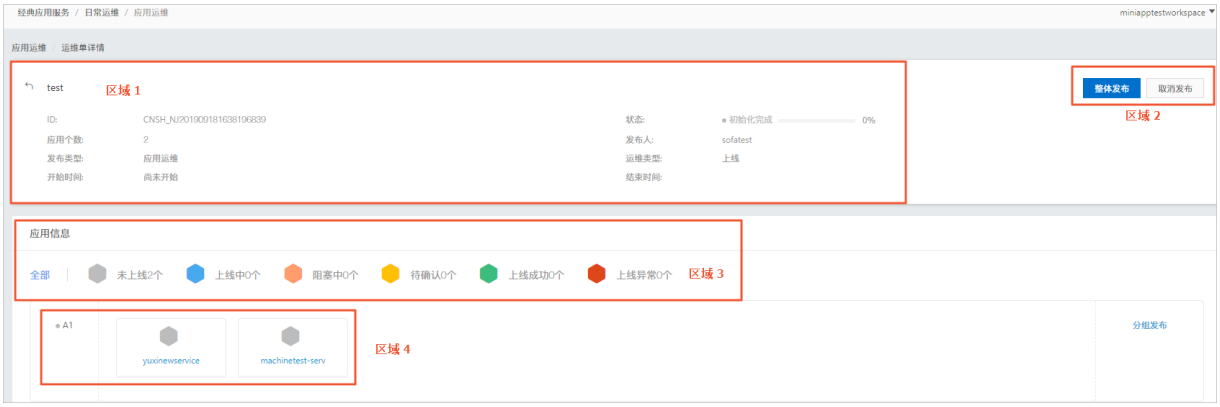
在 **应用运维大盘** 页面，单击 **创建** 进入新建运维流程。完成以下应用服务运维单配置后，单击 **创建**。

配置项	说明
标题	输入运维操作标题。可包含 1~50 个中英文字符。
类型	支持以下运维操作： • **重启**：支持重启机器上发布完成的应用。该操作是对机器上部署的应用做重启而非机器重启。 • **上线**：支持上线状态为 **已下线** 的应用，状态为 **已下线** 或 **待发布** 的服务器，或者已上线应用中 **发布失败** 的服务器。并且该应用在历史上必须有已经成功发布过的版本。 • **下线**：支持下线已上线的应用，和状态为 **发布完成** 或 **发布失败** 的服务器。 • **静态资源部署**：将静态文件部署至发布成功的服务器指定目录，目前仅支持 tar.gz 压缩文件。 • **初始化服务器**：重新初始化应用下的服务器，支持初始化数据盘和系统盘。
使用流程模板	说明 需要先创建流程模板才能使用该功能，具体步骤请参见 创建模板。 默认关闭。开启后，需从已创建的流程模板中选择流程模板。选择对应的流程模板后，将替换现有的固定发布部署流程。 

配置项	说明
分组策略	支持 **快速分组**、**共分一组**、**每台一组**、**按部署单元分组**。
使用 Beta 分组	默认关闭。开启后，将以应用为维度，从各个数据中心获取一台机器作为 Beta 分组机器成员。 在发布过程中，作为首个分组进行发布。完成后需手动确认方可执行后续发布。
运维维度	支持 **应用服务**、**服务器**。 • **应用服务**：选择应用服务，单击 **添加** 后，系统会自动获取所选应用服务下的所有服务器，并提示不合法的服务器。 • **服务器**：对选择的服务器执行运维操作，并提示不合法的服务器。
SLB 引流权重	当 **类型** 选择 **上线** 时需设置。 对挂载 SLB 的机器，设定上线时初始权重，取值范围为 0~100。 **说明** 该选项对未挂载 SLB 的机器无效。
初始化选项	当 **类型** 选择 **上线** 时需设置。 可选择 **初始化系统盘** 或 **初始化数据盘**。
部署单元	当 **运维维度** 选择 **应用服务** 时需设置。 按需勾选对应的部署单元。
应用服务列表	当 **运维维度** 选择 **应用服务** 时需设置。 您可以在 **待选应用服务列表** 中选择需要的应用服务，单击 > 图标将应用服务添加到 **已选应用服务列表** 中。
服务器列表	当 **运维维度** 选择 **服务器** 时需设置。 可选择对服务器做运维操作或对应用服务下对应机房的所有合法服务器做运维操作。您可以在 **待选机器列表** 中选择需要的机器，单击 **>** 图标将机器添加到 **已选机器列表** 中。
资源地址	当 **类型** 选择 **静态资源部署** 时需设置。 填写静态资源下载地址。 **说明** 当前仅支持 tar.gz 压缩包格式。
存储位置	当 **类型** 选择 **静态资源部署** 时需设置。 填写静态资源存储位置，如 `/home/admin/test/` **注意** 必须以 `/` 开头，以 `/` 结尾。
系统用户	当 **类型** 选择 **静态资源部署** 时需设置。 默认为 **root**。单击 **修改** 可进行更改。

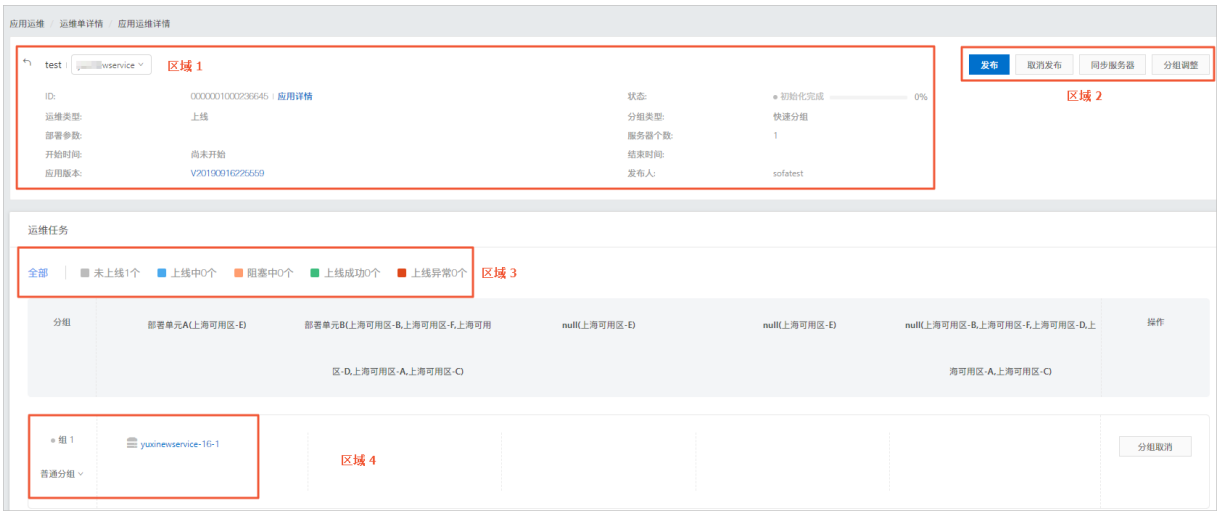
操作应用运维单

提交创建运维单申请之后，自动跳转至 **运维单详情** 页面，如下图所示。



- 区域 1：运维单基本信息，包括运维单状态、进度、运维类型、应用数量等信息。
- 区域 2：运维单操作区，可以取消运维单、执行运维单。执行运维单时，所有应用同时执行；取消运维单后，不能再对该运维单进行任何操作。
- 区域 3：应用筛选及应用发布状态信息，单击每个状态，即可根据该状态进行过滤，在区域 4 显示相应状态的应用。并且该区域显示处于各个状态的应用数量。
- 区域 4：应用展示区，展示该运维单下所有应用的发布进度和状态。

在 **运维单详情** 页面上，单击某个应用名称，进入该应用的 **应用运维详情** 页面，如下图所示。



- 区域 1：显示应用基本信息。该区域会显示应用名称、ID、状态、执行进度、服务器分组策略、服务器数量、分组数量等信息。如果有多个应用，可以直接在不同应用之间切换。
- 区域 2：应用操作区，用户可以执行、取消执行任务、分组调整等操作。
  - 执行：仅执行当前应用。
  - 取消执行：若取消执行，则本次运维中，用户不能再对该应用进行任何操作，但是整个运维单的状态不受该应用的影响。
  - 分组调整：可以对服务器进行新增分组、删除分组和组间移动服务器等操作。控制同一个应用下的执行顺序。
  - 失败重试：当该应用下的服务器进行运维操作失败时，可以进行重试。该操作会忽略已经执行成功的服务器，而只对失败的服务器进行操作。
- 区域 3：服务器筛选及服务器状态信息，单击每个状态，即可根据该状态进行过滤，在区域 4 显示相应状态下的服务器。并且该区域显示处于各个状态的服务器数量。

- 区域 4：服务器分组区，用于展示服务器分组、服务器执行状态、执行详情，并且可以进行加锁、解锁等操作：
  - 详情：执行完成、执行中、执行失败的分组，用户可以查看该分组中应用执行的详情，包括进度、状态、执行时间、失败原因等重要信息。

另外，在单台服务器上，如果其中的一个任务执行失败，可以在任务级别上进行 **重试** 或者 **忽略** 。如果 **忽略** 则进入下一个任务进行执行。

分组调整

在 **应用运维详情** 页面的区域 2 中，单击 **分组调整**，进入 **分组调整** 页面。
其中包括：重新分组、删除分组、分组调整三个功能。

- 重新分组：重新选择服务器分组策略，对 ECS 进行分组。
- 删除分组：在某个分组的最右端单击 **删除** 即可删除当前分组，并将该分组的 ECS 移动到 **待分配**。**待分配** 组中的 ECS 不进行运维。
- 移动分组：用户可以选中某台 ECS，将其拖动到指定的目标分组。

完成编辑之后，单击 **完成并返回**，即回到 **应用运维详情** 页面。

后续操作

用户完成初始化服务器之后，可以到 **发布部署** 页面进行应用的发布。

7.5. 发布包管理

发布包管理页面展示当前应用的发布包相关信息，包括应用名、发布包个数、最新发布包版本、最新发布包创建时间、技术栈/版本和负责人。在该页面，您可以管理应用发布包。

关于此任务

- **发布包管理** 列表分页展示您当前工作空间中所有应用的发布包相关信息，默认每页显示 10 条，每页最大显示条数为 20 条，同时该列表支持根据应用名称搜索应用以及根据发布包个数过滤显示有发布包的应用。
- 在 **发布包管理** 页面，您可以进行以下操作：
  - 搜索应用
  - 筛选有发布包的应用
  - 管理应用发布包：查看应用相关信息、上传发布包、发布应用、删除发布包、删除发布包记录。

搜索应用

1. 登录经典应用服务控制台。
2. 单击 **应用发布** > **发布包管理** 查看当前所有应用的发布包列表页面。
3. 在页面右上方的搜索框中输入待查询应用名称（不是别名或者描述），单击搜索图标开始查询。

筛选有发布包的应用

1. 单击左侧 **发布包管理** 菜单。
2. 单击页面 **发布包个数** 列旁的过滤器图标（ ），选中 **只显示有发布包的应用** 选项，单击 **确定** 按钮开始查询。

管理应用发布包

应用发布包管理详情页面展示应用的详细信息和当前的发布包信息。您可以在该页面上查看应用相关信息、上传发布包、发布应用、删除发布包、删除发布包记录。

查看详情

1. 单击左侧 **发布包管理** 菜单。
2. 单击 **发布包管理** 页面上应用右侧的 **管理** 进入该应用的发布包详情页面。

3. 在详情页面查看当前的应用信息以及发布包和发布包记录信息。

应用发布包管理详情页面包括以下部分：

- 应用基本信息区域：包括应用基本信息、**应用详情** 按钮和 **发布部署** 按钮。
  - 应用基本信息包括应用名称、应用别名、应用技术栈、应用负责人、应用描述。
  - 单击 **应用详情** 按钮后，您可以进入 **应用管理** 查看应用的详细信息。
  - 单击 **发布部署** 按钮后，您可以选择 **立即发布** 和 **高级选项** 两种模式。
    - 单击 **立即发布** 按钮会按照默认配置创建发布单并执行发布。
    - 单击 **高级选项** 会进入发布单创建页面，需要您手动创建发布单，详情参见 应用部署。
- 应用的发布包和发布包记录以及操作按钮。单击 **发布包** 和 **发布包记录** 可以在发布包和发布包记录展示列表间进行切换。发布包和发布包记录列表均支持根据版本号的简单搜索。

上传发布包

应用发布包的版本号在同一个工作空间中要求唯一。如果上传失败则不会生成新的发布包，但是会生成新的发布包记录。

1. 单击左侧 **发布包管理** 菜单。
2. 单击 **发布包管理** 页面上应用右侧的 **管理** 按钮后进入该应用的发布包详情页面。
3. 单击 **上传发布包** 按钮，弹出 **添加新版本** 对话框。
4. 参照如下信息设置发布包信息：
   - **版本号**：必填，确保不与当前工作空间的已有版本号重复。
   - **文件**：上传发布包文件，一般不允许超过 200 MB。在某些特殊工作空间下，文件大小不允许超过 300MB。
   - **备注**：选填，发布包的备注信息。

发布应用

您可以选择某个发布包快速发布应用服务。要发布特定版本的发布包，完成以下操作：

1. 单击左侧 **发布包管理** 菜单。
2. 单击 **发布包管理** 页面上应用右侧的 **管理** 按钮后进入该应用的发布包详情页面。
3. 在列表中选择要发布的应用发布包，单击 **发布** 按钮，弹出 **发布部署** 对话框。
4. 根据您的业务需求，选择以下支持的发布部署模式发布应用服务。如果需要取消发布，请单击 **取消** 按钮关闭对话框：
   - 单击 **立即发布** 按钮会按照默认配置创建发布单并执行发布。
   - 单击 **高级选项** 会进入发布单创建页面，需要您手动创建发布单，详情参见 应用部署。

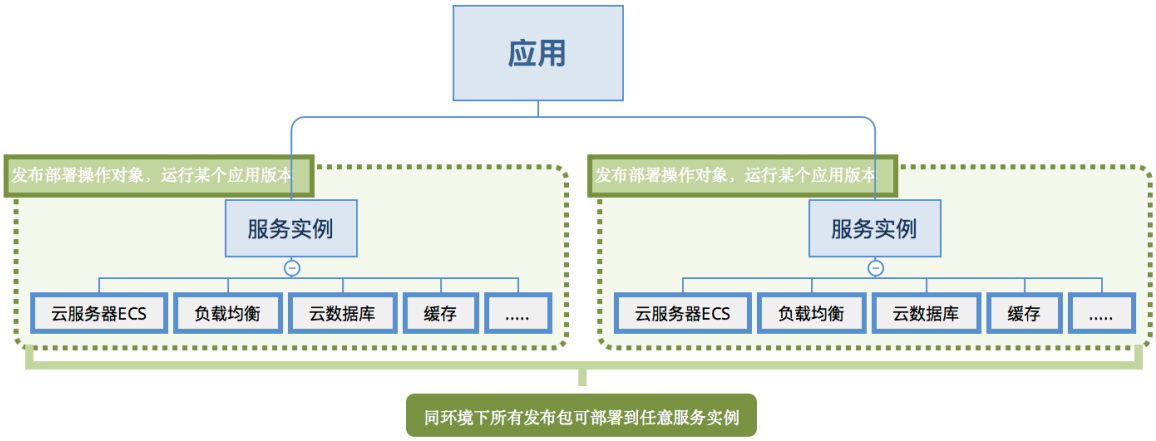
删除发布包

您既可以单个删除应用发布包，也可以进行批量删除。要删除应用发布包，完成以下操作：

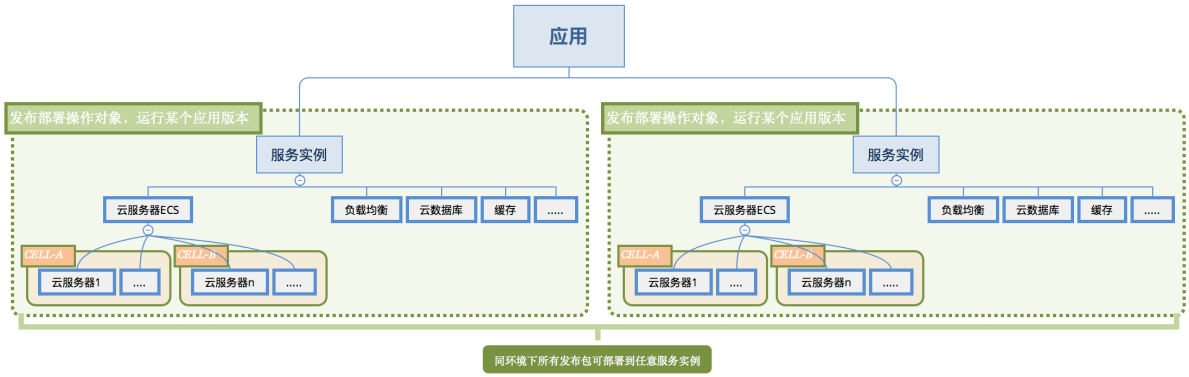
1. 单击左侧 **发布包管理** 菜单。
2. 单击 **发布包管理** 页面上应用右侧的 **管理** 按钮后进入该应用的发布包详情页面。
3. 在列表中选择要删除的应用发布包，单击 **删除** 按钮，弹出应用发布包删除确认对话框。如果要删除多个应用发布包，则选中所有要删除的发布包，然后单击 **批量删除** 按钮，后续操作不变。
4. 如果确认删除，单击 **确定** 按钮并输入短信校验码删除应用发布包，否则单击 **取消** 按钮取消删除应用发布包。

7.6. 应用服务实例

应用服务实例是介于应用和资源之间的新模型，应用在某个工作空间中可以有多个服务实例，支持在单个工作空间中同时运行多个服务实例并对外提供服务，不同服务实例之间依然会有安全组、白名单、VPC 网络等技术实现的隔离。
原先应用与资源直接发生关联，应用服务实例引入后，应用将不再直接关联资源，代替的是应用下的服务实例与资源直接关联。当您做发布部署时，操作对象也有原来的应用变更为了服务实例。除此之外，服务实例还与隔离单元（cell）关联，为您提供了更灵活的发布策略。
单个工作空间中应用、服务实例与资源的关系：



单个工作空间中应用、服务实例与资源以及 cell 的关系：



在 **服务实例列表** 页面，您可以进行以下管理操作：

- 创建应用服务实例
- 简易搜索
- 高级搜索
- 查看应用服务实例详情

关于此任务

- 应用服务实例列表分页展示您当前工作空间中所有的应用服务实例信息，默认每页显示 10 条，每页最大显示条数为 20 条，同时该列表支持根据应用名称的简单搜索或者根据多个条件组合的高级搜索以及运维操作。
- **运维操作** 下拉按钮为您提供对列表中已选的应用服务实例执行 **发布部署**、**重启**、**上线应用**、**下线应用** 操作的快捷入口：
  - **发布部署**：支持 **立即发布** 和 **高级选项** 两种模式。单击 **立即发布** 按钮会按照默认配置创建发布单并执行发布；单击 **高级选项** 会进入发布单创建页面，需要您手动创建发布单。详情参见 应用部署。
  - **重启**、**上线** 和 **下线**：参见 日常运维。

创建应用服务实例

1. 进入经典应用服务控制台，单击左侧导航栏中的 **应用发布 > 应用服务** ，单击 **创建服务实例**。
2. 填写以下配置信息后，单击 **确定**。
   - **应用**：选择之前创建的应用。
   - **服务实例名称**：允许包含字母、数字、连字符、下划线，且必须以字母开头、以字母或者数字结尾，长度不能超过 60 个字符。例如 `sofabootdemo` 。
   - **发布包**：选择应用的发布包。可在 **应用发布 > 发布包管理** 页面为应用上传发布包。
   - **描述**：不能超过 500 个字符。
   - **技术栈**：根据所属应用自动展示。
   - **技术栈版本**：选择所需版本。
   - **部署单元**：勾选应用要发布到的部署单元。

简易搜索

1. 进入经典应用服务控制台，单击左侧导航栏中的 **应用发布 > 应用服务** 查看当前所有应用的服务实例列表。
2. 在页面右上方的搜索框中，输入待查询应用服务实例名称（不是别名或者描述），单击搜索图标开始查询。

高级搜索

应用高级搜索支持根据应用名称、应用负责人、技术栈、技术栈版本号的单个或者多个条件的组合搜索应用。

1. 单击 **应用发布 > 应用服务** 查看当前所有应用的服务实例列表。
2. 单击页面右上方的 **高级搜索** 按钮显示高级搜索视图。
3. 在各个搜索框指定搜索条件后，单击搜索按钮。
4. 单击 **清空** 按钮清空所有的搜索条件。
5. 单击 **简易搜索** 按钮隐藏高级搜索视图。

查看应用服务实例详情

应用服务实例详情页面展示了应用服务实例当前的详细信息以及资源信息。

1. 单击 **应用发布 > 应用服务** 查看当前所有应用的服务实例列表。
2. 单击 **服务实例列表** 页面的服务实例名或者 **详情** 按钮后进入应用服务实例详情页面。
3. 在详情页面查看当前的应用服务实例详细信息以及资源信息。

应用服务实例详情页面包括以下信息：

- 第一部分：应用服务实例基本信息区域，包括应用服务实例基本信息、**应用详情** 按钮和 **运维** 按钮。
  - 应用服务实例基本信息包括实例名称、应用、状态、创建时间、负责人、发布包版本、别名、技术栈、描述。
  - 单击 **详情** 按钮后，您可以进入 **应用管理** 查看应用的详细信息。
  - **运维** 下拉按钮为您提供对当前应用服务实例执行 **发布部署**、**重启应用**、**上线应用**、**下线应用** 操作的快捷入口。
- 第二部分：展示资源使用情况以及环境参数。
  - 应用服务实例的资源包括 **云服务器 ECS**、**负载均衡**、**云数据库 RDS**、**云数据库 Memcache（新）** 和 **云数据库 Memcache**。点击相应资源类型名称如 **云服务器 ECS**、**负载均衡**，可以在不同资源的展示列表间进行切换。各个资源的搜索、添加等功能使用参见 资源管理。

- 环境参数：包括系统参数与自定义参数。您可以添加自定义的环境参数，绑定该应用服务实例。系统参数包含以下参数：
  - bp.-Drpc_enabled_ip_range：SOFA 应用的 RPC 网卡配置。默认为 `10:11,172.16,192.168` 。
  - ac.is_dsr_dependent：应用是否依赖独占配置中心。默认为 false
  - ac.is_bg_deployment_supported：应用是否支持蓝绿发布。默认为 true。
  - ac.is_ldc_supported：应用是否支持 LDC 部署模式。默认为 true。

7.7. 部署单元

部署单元（Cell）定义了一个服务通信的边界，用以规整通信的链路，为发布部署时切流做无损发布、高可用应用架构（包括同城双活、异地多活），以及容灾切换和恢复打下基础。
部署域（Deployment Zone）对可用区（Available Zone）进行了划分，用以规整计算资源，为发布部署、容灾场景打好基础。SOFAStack 规定一个工作空间（Workspace）在一个地域（Region）中最多有两个部署域（仅做同城双活；对于同城，做多活意义不大）。

> 说明
> - 对于工作空间里有不超过两个可用区的情况，初始状态下，会有两个部署域，每个部署域包含一个部署单元，每个部署单元关联一个可用区。
> - 对于工作空间里有超过两个可用区的情况，也仅有两个部署域，这两个部署域的两个部署单元会分别关联最早创建的两个可用区，其余的可用区此时没有关联到任何布署单元；此时会导致没有关联到部署单元的可用区里的计算资源无法用于发布部署，所以必须通过发布部署产品关联到相应的部署单元中。

在 SOFAStack 平台中，一个部署单元里的一组服务优先被发现和访问（可以通过配置中心这类产品实现）。对于一个应用服务（Application Service）会关联一个或者多个部署单元，同时由多个计算节点（ECS）组成，每个节点都发布多个功能相同的服务。在把计算节点分配给应用服务时，会同时把计算节点分配到相应的部署单元里，这样就使得在同一个部署单元里的计算节点上发布的服务能够被优先访问，达到相同部署单元里的服务优先的目的。
在应用服务做发布部署时，现在就可以部署单元的维度进行操作，比如提供按照部署单元实施分组策略，做蓝绿发布等等。所以，在 SOFAStack 平台的发布部署和日常运维中会加强部署单元的透出，尽量屏蔽底层可用区（Available Zone）的概念。
部署单元、可用区和部署域的关系（同一个工作空间内）如下：

- 一个部署单元在公有云下通常只会对应一个可用区；而由于会出现一个可用区资源售罄的场景，故此时，允许一个部署单元关联多个可用区。
- 一个部署单元只能属于一个部署域。
- 一个部署域可以有多个部署单元。
- 一个部署域通常只会有一个可用区，但可以有多个可用区，原因和 1 一致。
- 一个可用区不能同时属于多个部署域。

> 说明
> 在 SOFAStack 平台中，服务通常对应于 SOFA 的 RPC 服务。

创建部署单元

1. 进入经典应用服务控制台，单击左侧导航栏中的 **设置** > **环境参数** 进入创建部署单元页面。
2. 单击 **+ 新建部署单元**，输入标识和描述。

> 说明
> 部署单元创建后不能删除，部署单元标识不能修改。

编辑部署单元

单击已有部署单元右侧的 **编辑** 图标，可对该部署单元进行编辑。
若工作空间添加了一个可用区，进入 **经典应用服务** 时会提示需要将可用区添加到部署域。

> 说明
> 目前需要刷新页面，才能看到提示信息。

7.8. 自定义技术栈

应用管理通过技术栈来管理应用在发布部署时需要安装的软件包以及要执行的部署脚本。目前已经提供了 SOFA Lite、Spring 等常用的系统技术栈供您使用。当然您也可以通过自定义技术栈来定制更多个性化的需求。
本文档将以创建 Spring Boot 技术栈为例，介绍如何自定义技术栈。更多内容，参考 技术栈使用指南。

创建技术栈实例

1. 进入应用管理控制台，单击左侧导航栏中的 **技术栈管理**，进入技术栈管理页面。
2. 在技术栈管理页面中，单击左上方的 **创建** 按钮创建技术栈实例。
3. 填写以下技术栈信息后，单击 **创建**。
   - 技术栈：选择所要创建的技术栈框架。
   - 技术栈版本：输入版本前缀，后缀默认为租户名。
   - 支持操作系统：支持 CentOS 6.5 64bit、CentOS 7.2 64bit、CentOS 7.6 64bit 操作系统。
   - 调试模式：勾选调试模式意味着在发布技术栈后，仍可以对该技术栈进行任意修改，从而方便您对技术栈在云上进行集成测试，但调试模式的技术栈无法共享给任何租户。发布之后取消调试模式则无法再次开启调试模式，取消调试模式之后，技术栈可重新分享给其他租户。
   - 技术栈包：单击 **上传文件** 上传本地技术栈包。
   - 环境参数：单击 **添加** 输入环境参数。

7.9. 经典应用服务常见问题

经典应用服务支持哪些操作系统？

技术栈 1.0 架构应用仅支持 Centos 6.5 版本操作系统的服务器自动化发布。
技术栈 2.0 架构应用可支持 Centos 6.5 与 7.2 版本操作系统的服务器自动化发布。
用户可通过自定义技术栈来支持其他操作系统发布。

发布单或运维单为什么初始化失败？

初始化发布单会检测应用的服务器状态，要求服务器状态是 **待发布**、**发布完成**、**发布中**、**发布失败**。 初始化失败可能的原因如下：

- 应用下未创建服务器。
- 服务器状态不是允许发布的状态，如 **创建中**、**已关机**、**重启中** 等。

如何部署应用？

经典应用服务提供应用发布及回滚的功能，您可以编辑待发布应用的依赖关系及服务器分组，选择发布策略。
在部署大盘页面可新建发布单。

1. 在发布大盘界面，单击 **新建发布**，进入 **发布申请** 页面。
2. 填写发布基本信息后，单击 **快速创建**，则直接创建应用的发布单，且应用之间无依赖关系。
3. 若需编辑 **部署策略**、**部署参数**、**依赖关系** 等信息，单击 **编辑高级信息**，进入高级信息编辑页面。

4. 单击 **下一步**，进入发布单预览界面，确认无误后单击 **创建** 按钮，创建发布单。
5. 单击应用，进入应用详情页，可进行分组调整、回滚操作。

云上应用发布完成后如何对外提供访问？

推荐您使用 SLB 挂载多台 ECS 提供对外访问。外部用户访问流量经 SLB IP 转发到后端应用服务器上。

云上应用发布后不是最新代码版本如何处理？

此种情况可能由以下原因导致：

- 应用发布的版本选择错误，可通过查看应用详情中的代码版本确认是否错误。
- 用户文件夹权限问题。发布部署服务使用的是 admin 用户权限操作。若您的 `app-run` 文件夹非 admin 账户所有则会出现此种情况。

排查方法：查看 `/home/admin/app-run` 目录，使用 `ll app-run` 命令确认目录是否归 admin 账户所有。

解决方案：

- 直接初始化服务器重新发布。
- 登录服务器将 `app-ru` 目录改成 admin 用户权限，再重试发布流程即可。

发布部署过程中为什么分组被自动加锁？

分组被自动加锁，是由于有其他发布单正在发布该分组内的服务器，应用页面顶部会出现提示 `该应用下有服务器正在其他发布单XXX发布`，单击发布单链接，可查看正在发布的应用。

当正在发布的服务器发布完成后，已加锁的分组可进行解锁继续发布。

如何升级应用的技术栈版本？

在应用详情页面可查看应用的技术栈及技术栈版本。应用的技术栈版本可在部署的过程中升级。具体的升级步骤如下：

1. 进行新建发布操作，进入发布申请的 **编辑高级信息** 页。
2. 在 **编辑高级信息** 页，在 **部署策略** 选项中勾选 **升级/降级技术栈**。选择要升降级的目标版本，进行发布部署。

如何在创建应用后修改技术栈？

发布部署服务暂不提供直接修改技术栈的功能。如确定需修改，则必须将应用上绑定的各种资源从该应用上解绑，重新创建新的应用，选择目标技术栈，然后再把资源重新分配到新的应用上。

如何排查发布部署执行失败的原因？

排查思路如下：

1. 查看报错日志，单击分组右侧的 **详情**，弹出部署各个步骤的详情页面。单击失败步骤右侧的 **详情**，可展开详细日志。
2. 应用部署失败的常见原因如下：
   - 应用系统的配置问题，可通过错误日志具体分析。
   - 应用服务检测失败：通常是应用启动失败，可登录到机器上查看具体的部署日志。部署日志的目录为 `/home/admin/logs`。
   - 下载应用失败：通常会出现在上线、回滚、重启的操作中。通常是由于手动修改过服务器的应用包，而上线、回滚、重启操作会对应用包作 md5 校验。处理方法是重新进行部署操作。
   - 关流量、下载脚本包、环境安装、环境配置、开流量的步骤出现失败，请先单击分组右侧的 **分组重试**。若重试仍然失败，请 提交工单 获取帮助。

如何在金融科技 gitlab 的 project 中增加 user?

使用管理员所在租户的成员的金融科技账号添加。因此，如需新增用户必须先有一个该租户下的金融科技账号。

如何自定义技术栈？

自定义技术栈支持 Java、C/C++、PHP 等主流开发语言，您可以自定义技术栈实例，给用户提供常用软件、自定义软件、自定义脚本以及创建代码库、编译打包等功能。
操作步骤如下：

1. 在 **运维管理** > **应用** > **技术栈** 模块，单击 **添加**。
2. 填写技术栈名称，选择技术栈框架，单击 **确定**，进入 **技术栈实例详情** 页面。
3. 在 **技术栈实例详情** 页面，您可进行通用软件配置、自定义软件配置、自定义脚本上传等。
4. 编辑完成之后，单击 **提交**，技术栈状态变成 **已提交** 状态。
   创建应用时即可选择该技术栈实例进行创建。

经典发布失败，但是应用健康检查成功

现象

- 通过经典应用服务发布，发布单显示应用启动失败，但在相关机器执行健康检查脚本，健康检查成功。
- 报错信息：HealthCheck Remaining Time 120 seconds.

原因

由于健康检查脚本默认检查 120s，如果应用启动时间超过 120s，发布部署平台认为健康检查超时，返回部署失败。

解决方案

1. 在 logs 目录下搜索关键字 `Started XXX in` （ `XXX` 替换为应用的启动类），即可在日志中获取应用的启动耗时信息，例如：

   ```
   2019-01-2511:32:38,550[-/// - ] INFO  bootstrap.SOFABootWebApplication - Started SOFABootW
   ebApplication in 10.848 seconds (JVM running for 12.085)
   ```

2. 在应用根目录下的 `conf/bin/healthcheck.sh` 健康检查脚本中，根据应用实际情况，修改健康检查时间。

8.容器应用服务

8.1. 什么是容器应用服务

8.1.1. 概述

业界普遍认为，云原生架构能加快需求交付、降低运营成本、支持容量伸缩、保证业务连续，从而使组织能更从容地接入创新技术、促进渠道触达。金融分布式架构 SOFAStack 致力于提供一整套帮助广大金融场景落地云原生、分布式架构的产品和解决方案，而其中的应用 PaaS 平台，融合金融科技多年在大规模分布式系统和容器平台的实践经验，使用户在专注于业务价值的同时，提升研发效率和自动化水平，降低成本和业务技术风险。

SOFAStack 平台下的容器应用服务（Application Kubernetes Service，简称 AKS），全面集成 Kubernetes，提供完整的集群管控、认证授权、容器网络、持久卷存储等方面的平台能力。在兼顾标准化一致性的 Kubernetes 能力的同时，亦将源自实践的应用全生命周期的发布部署能力通过产品化的形式交付。针对金融级场景下大规模分布式系统的特点，提供了丰富的发布策略以满足不同的场景，如：分组发布、Beta 发布、灰度发布等，帮助传统架构平滑过渡，适应金融技术风险保障需求，实现大规模金融级运维场景下的容器服务落地。

容器应用服务致力于通过成熟的技术和最佳实践经验的支撑，使金融场景亦能从容地应对云原生开发、运维、架构的难题，解决金融系统应用容器化转型的需求，使容器技术真正的大规模应用于金融行业生产环境里，帮助传统应用以更高效、低成本的方式迁移到云原生微服务、容器化体系架构。

8.1.2. 优势

金融级发布

- 发布过程安全可靠，可重试，可灰度，可回滚，可溯源。
- 支持虚拟机和容器的混合发布，提供从虚拟机到容器的过渡方案。

运行时监控

- 业务自定义大盘随时关注业务动态。
- 实时监控应用基础指标，如 PV、Service（应用服务被调用）、SAL（调用外部服务）等。
- 全面收集基础资源指标，如 CPU、内存、IO 流量等。

微服务框架

- 深度集成蚂蚁 SOFA Mesh 做异构语言的服务注册发现及通信。
- 支持原生部署 Istio 做 Service Mesh 提供微服务能力。

网络模式

- 支持 VPC 和经典网络两种网络模式。

高可用和容灾

- 支持经典的同城双活、两地三中心容灾方案。
- 支持阿里云沉淀多年的单元化高可用容灾方案。

8.1.3. 功能特性

金融级发布策略

Kubernetes（简称 K8s）的 Deployment 部署已经趋于稳定，然而对于许多企业级的复杂部署模型，仍旧依赖高阶脚本和运维工具，滚动升级亦无法满足实际金融场景的运维需求。因此，基于对于金融级 PaaS 平台的理解和实践，SOFAStack 基于 K8s 的标准，通过 CRD 扩展了 K8s 的部署模型，并按照实际金融级发布的需求对产品界面流程做了定制，使金融级发布部署的最佳实践在容器应用服务中得以体现。

应用服务

基于发布单形式的应用服务部署，提供丰富发布策略，涵盖从 Pod 和容器定义、存储、网络、发布策略、日志监控等一站式发布部署模型，支持版本快照溯源回退，确保您能够通过可视化、强管控的方式配置应用服务。

发布单

有效管控单个或一组应用的发布状态和策略，检查发布变更。确保应用服务的任何配置变动，都会生成一次版本快照。在发布准备时，您可以配置应用依赖信息，自动实现串行和并行发布策略。在发布过程中，确保前置、发布、后置任务事件完整记录，方便追踪发布历程。

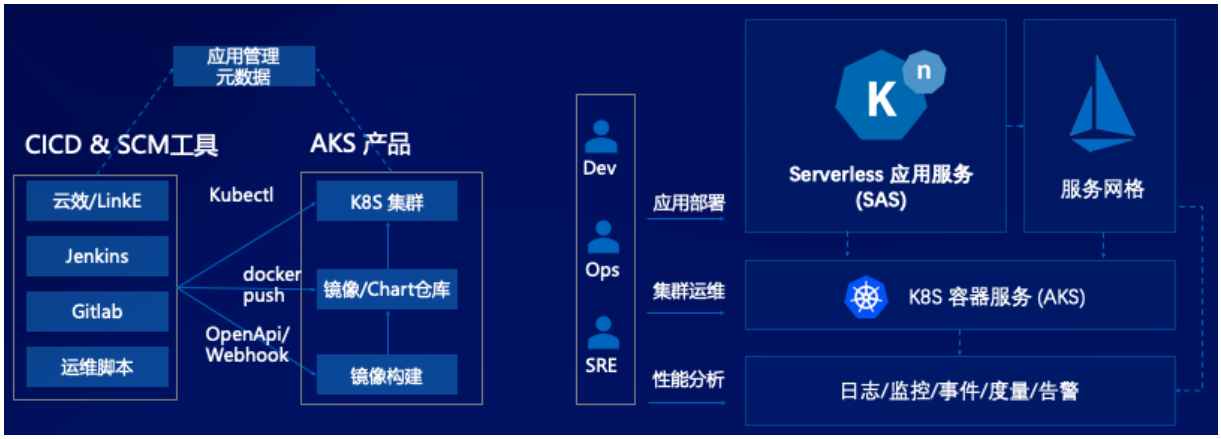
核心部署对象：CafeDeployment

对 K8s 原生发布能力以 CRD 形式作了扩展，以云原生的方式为应用提供管理能力，使标准命令（`kubectl apply -f xxx.yaml`）即可搞定复杂部署。

- 单元化：应用实例需要分割到不同的节点进行部署。
- 精细的发布管理：在 Pod 配置的更新过程中，提供 Beta 发布、分组发布、发布暂停、发布取消等功能。
- 版本控制：在每次 Pod 配置更新后自动保存版本，方便版本回退。
- 原地升级：在保持 Pod IP 不变的前提下实现 Pod 其他配置的变更。

一站落地云原生

容器应用服务已通过了 CNCF 一致性测试，为 K8s 标准原生 Workload 以及相关原生资源提供可视化 UI 界面，将确保提供一致的 Kubernetes 能力，满足容器平台运维的事实标准。
在提供平台能力闭环的同时，确保产品支持开放生态，使开发、运维等角色能够通过所习惯的方式同平台进行交互，使 CICD、代码管理工具、监控系统无缝对接，并能够在容器应用服务之上享受到 ServiceMesh 所带来的优势和前景。



集群管理

提供一键从 0 到 1 搭建集群的能力，并且可以随时按需对集群进行扩容和缩容。另外，提供命名空间管理，可实现业务的逻辑隔离。

元数据管理

在兼容 K8s 原生 Service、Pod、ReplicaSet 等模型的基础上，结合容器应用服务的应用服务概念，提供更加丰富的元数据和资源管理。

网络流量

集成云上负载均衡设备，可将内部流量（负载均衡类型服务）和外部流量（Ingress）转发至应用服务的 Pod。

镜像中心

容器应用服务所包含的镜像中心，提供和用户 Gitlab 代码库关联的镜像构建和版本管理能力，满足从代码到镜像的一键构建需求。同 IAM 对接满足安全保障。

发布部署

针对不同用户场景，同时支持原地升级和原生发布两种应用服务部署形态。

应用商店

支持通过 HelmChart 应用商店一键编排复杂容器应用。

8.1.4. 应用场景

使用 SOFA 技术栈的传统研发运维体系

这类体系的应用使用 SOFABoot 进行开发或直接使用 SOFA Mesh，系统间关系复杂，有依赖关系，并且深度使用了 SOFAStack 产品，需要和现有 PaaS 的发布部署能力做无缝对接：

- 多应用同时发布运维，可做应用分组、应用依赖关系调整等高阶发布运维能力。
- 需要蓝绿发布，单元化发布能力。

这类系统的特点为：

- 应用架构较复杂，且应用间有依赖，需要在发布单中做协调，并做细粒度控制，包括：Pod 分组（包含 beta）、分组重试、单 Pod 重试、Pod 上单任务重试/忽略等现在已有的功能。
- 存在虚拟机和容器混布（迁移）的场景。
- 对蚂蚁中间件强依赖，需要做无损发布（切负载均衡流程流量、切软负载流量、加白名单、发布等流程必须有序操作）。

该场景下，您需要在 AKS 中创建应用服务，使用 SOFABoot 运行时镜像作为基础镜像构建应用镜像，以原地升级方式进行发布单级别的部署，就可使应用服务在 AKS 上运行，与虚拟机提供的服务进行交互，并与现有 SOFAStack 产品做无缝对接。

使用 SOFA 技术栈的轻量级研发运维体系

这类体系的应用使用 SOFABoot 进行开发或者使用 SOFA Mesh，和 SOFAStack 产品有紧密结合。这类应用的特点为：

- 不需要多应用同时发布，即应用可以进行独立发布。
- 需要做无损发布。
- 没有历史包袱，可接受 CloudNative 方式运维。

该场景下，您只需在 AKS 中创建应用服务，使用 SOFABoot 运行时镜像作为基础镜像构建应用镜像，以原地升级方式进行部署管理，即可使应用服务在 AKS 上运行，并和现有蚂蚁科技云产品无缝对接。

使用云原生技术体系

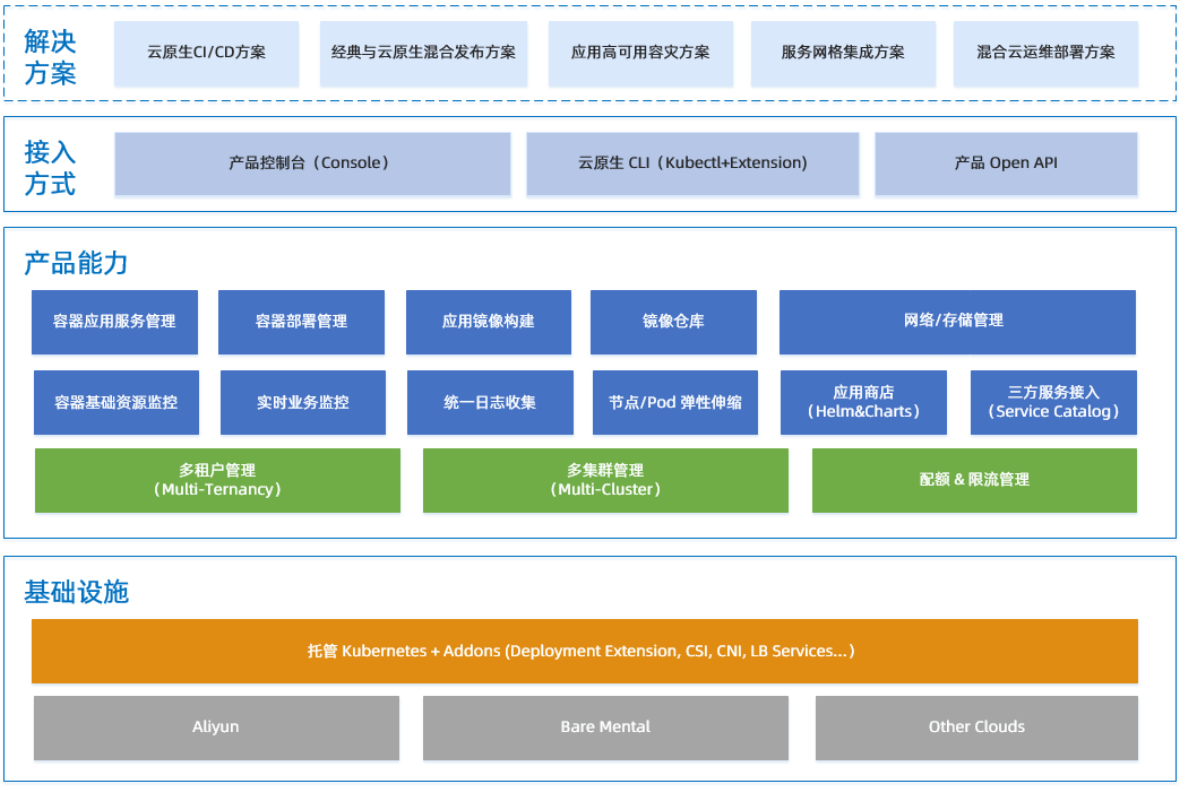
这类体系的应用通常采用传统 Spring 或者 SpringBoot 技术栈，配合 Eureka，ZooKeeper 做服务注册与发现，并配套 CNCF 生态的监控，链路工具，自成一体。这类的应用的特点为：

- 业务对错误容忍性高，或非关键链路系统。
- 进行独立发布。
- 没有历史包袱，可接受 CloudNative 方式运维。
- 不需要和 SOFAStack 产品结合，比如监控，弹性伸缩，容灾，中间件等。

该场景下，直接在 AKS 中创建应用服务并对现有代码或者镜像进行管控，以原生发布方式进行部署管理，其周边使用的配置中心、监控、链路、Service Mesh 等工具可以和原先的方式保持一致，直接通过 Kubernetes 支持的方式，比如 Helm Charts 进行部署和管理。

8.1.5. 架构

SOFAStack 提供基于 Kubernetes 的托管式容器应用服务，服务底层可无缝对接阿里云，裸金属服务器和其他 IaaS 平台（如 OpenStack），并针对不同平台，按照 CSI、CNI 标准提供存储、网络插件，使用 Open Service Broker 标准接入三方服务，如数据库、缓存服务、负载均衡等。
在产品层除了提供业界标准的云原生能力外，还提供具备金融级特性的发布运维能力，比如按序的无损发布策略，针对大规模分布式系统的批量发布运维能力，内置按机房或部署单元的高可用容灾拓扑，并集成 SOFAStack 的实时监控能力。
您可通过产品控制台或 Kubernetes 提供的标准 Kubectl 工具接入容器产品。三方产品亦可通过容器应用服务开放的 OpenAPI 或者 Kubernetes 标准的 API 集成容器应用服务的产品能力，打造各种解决方案，服务用户。



8.1.6. 基础术语

术语	英文	说明
应用服务	Application service	该概念和经典应用服务中的应用服务概念一致。但由于容器有其特殊性，AKS 中的应用服务会包含一些额外的元数信息，比如容器规格配置、镜像、调度策略、日志配置等。
镜像	Image	镜像是应用包，将配置和相关软件等打在一起的二进制包，并且符合 Docker Image 规范。镜像可以来自任何可被 AKS 网络访问到的镜像中心，对于私有镜像中心，需要在 AKS 中配置相应的访问信息。
构建	Build	构建用于描述从应用源代码到制作出镜像过程的配置信息，包括源代码地址、分支信息、源镜像访问信息、目标镜像信息、Dockerfile 位置信息等。
集群	Cluster	AKS 中集群用于描述您所创建的一个工作负载集群，由多个节点组成。
节点	Node	节点表示一台装了 Docker 和 Kubelet，用以运行应用负载的物理机或者虚拟机。

术语	英文	说明
容器组	Pod	Kubernetes 中最小的部署及管理单元。一个 Pod 由多个相关的并且共享磁盘的容器组成。
命名空间	Namespace	命名空间和 Kubernetes 中相应的概念保持一致，用于表示一个逻辑隔离的空间，会将 Pod、Service、ReplicaSet 等元素隔离，但通常来说，网络不隔离。
原地升级	Inplace upgrade	原地升级是指应用服务中 Pod 的更新方式。在该方式下，发布后 Pod 的 IP 和发布前保持一致，所在的节点也尽可能保持不变。
原生部署	Cloud Native deployment	原生部署也是指应用服务中 Pod 的更新方式。在该方式下，发布后 Pod 的 IP 通常和发布前是不一样，所在的节点也可能发生变化。
标签	Label	Kubernetes 的原生概念，用于给相应的资源打上标签，做聚合或者匹配。
污点	Taint	Kubernetes 的原生概念，用于给节点做污点标记，通常用于 Pods 的调度策略。与之相对应的概念为：容忍（tolerance），若 Pods 上有相对应的 tolerance 标记，则可以容忍节点上的污点，并调度到该节点。
保密字典	Secret	Kubernetes 的原生概念，用于存储用户的加密内容。
应用容器	Container	应用程序所运行的隔离工作空间，通常是 Docker 容器或者 Pouch 等兼容 CRI 接口的具有隔离能力的沙箱工作空间。
工作负载	Workload	应用程序运行态的载体及其上层聚合。通常包括：Pod、Deployment、StatefulSet、DeamonSet、Job 等。
配置项	Config Map	Kubernetes 的原生概念，用于存储用户的配置信息。
存储类型	Storage Class	Kubernetes 的原生概念，通常由系统管理员定义，用于指定所支持的存储类别，不同的类别会有不同的存储 SLA、备份策略等差异性。

术语	英文	说明
存储卷	Persistent Volume	Kubernetes 的原生概念，表示一个由系统管理员创建好的存储资源。
存储卷声明	Persistent Volume Claim	Kubernetes 的原生概念，一个存储卷声明绑定一个存储卷。

8.1.7. 限制说明

本文介绍容器应用服务 AKS 的功能限制说明。

应用发布

- 同一个应用服务，不能同时进行发布和运维操作。
- 以下操作都会在发布应用服务时删除并重建 Pod，而非原地升级：
  - 调整应用服务中 Pod 的 CPU 和 memory 的值。
  - 开启实时监控后回滚至关闭监控的版本。
  - 改变 Pod 中容器的数量。
- 在应用服务的版本中更改所关联的 SLB 会导致业务中断。

集群管理

- 不支持集群删除及集群更新。例如，更新集群可用区交换机、安全组、更新集群监控配置等。
- 集群配额：每个工作空间（Workspace）只能创建一个集群。
- 集群中 Pods 网段和 Service 网段创建后不能更改。
- 在删除节点时，包年包月类型的机器不支持删除机器，仅支持清理磁盘操作。
- 通过生成证书去访问集群：
  - 需在 VPC 内网中才能访问，暂不支持公网访问。
  - 不能获取 CafeDeployment、inplaceset 等金融云扩展资源。
- 容器 Web 控制台的超时断连时间为 1 分钟。

镜像

- 镜像构建：限制只能同时进行 3 个构建任务。
- 镜像中心：用户只能往所在租户的 namespace 下 push 镜像，或者从公共的 namespace 拉取镜像。镜像中心只能在 region 内部访问。

8.2. 快速入门

本文将引导您快速熟悉公共云（或阿里云底座的专有云）上 AKS 的操作链路。

步骤一：创建标准工作空间

工作空间（Workspace）本质上是“网络互通、安全策略一致、访问延时极小”的一组资源。您可以通过工作空间方便地将资源进行分组管理，例如：根据不同的研发交付需求，将工作空间划分为开发工作空间、测试工作空间、生产工作空间等。不同工作空间中的资源互相隔离。

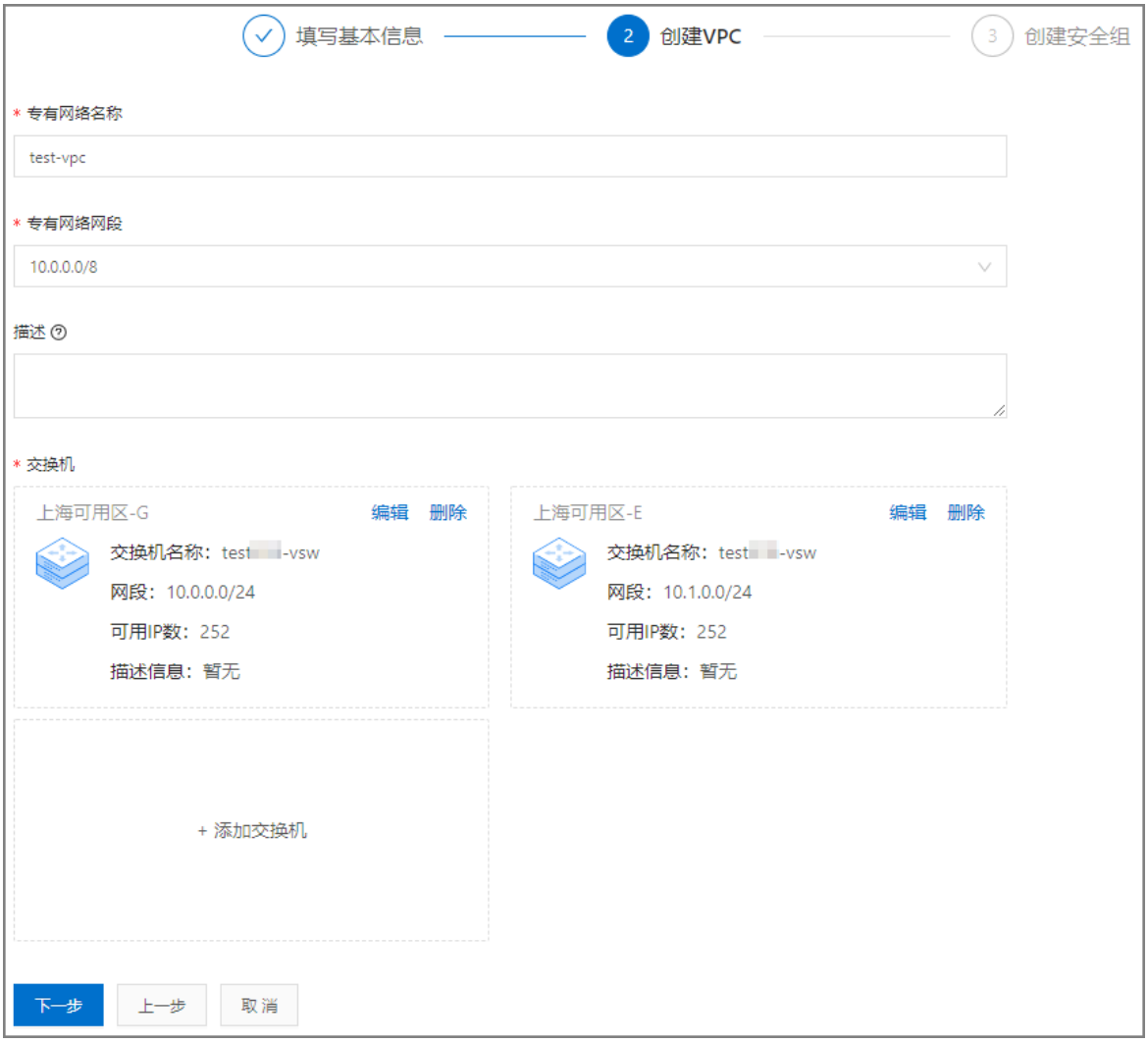
1. 登录 SOFAStack 控制台。
2. 在左侧导航栏，单击下方的 **全局设置**，进入工作空间列表页面。
3. 单击 **创建工作空间**，选择 **标准工作空间** 类型，单击 **创建**。
4. 在 **创建工作空间** 页面，输入以下基本信息。

- **工作空间标识**：2-64 个字符，工作空间的英文标识，全局唯一，一经确定无法修改，例如：dev、test、prod 等。本例中输入：`DemoWS`。
- **工作空间名称**：1-64 个字符，工作空间的显示名称，例如：开发工作空间、测试工作空间、生产工作空间。本例中输入：`DemoWS`。
- **地域**：工作空间所在的地域（Region），一个工作空间一定属于某个地域。

  > 说明
  > 目前开放的地域有 **华东2（上海）** 及 **华东1（杭州）**（金融云）。

- **网络类型**：选择 **VPC 网络**。
- **导入 VPC**：保持关闭。
- **可用区**：工作空间配置的可用区（Available Zone），最多不能超过两个，工作空间只能导入已配置可用区内的资源，且发布部署、容灾等其他功能都是基于工作空间的可用区来做分组发布与容灾切换等操作。推荐配置两个可用区，为支持双机房高可用等架构做准备。

5. 单击 **下一步**，在 **创建 VPC** 页面，输入以下配置信息：



- **专有网络名称**：由 2-128 个英文或中文字符组成，必须以大小写字母或中文开头，可包含数字、下划线（_）或连字符（-），不能以 `http://` 和 `https://` 开头。可以使用系统自动生成的名称 **单元化工作空间标识**-vpc 。
- **专有网络网段**：专有网络的网段，一旦选择便无法更改，专有网络内的所有资源，如 ECS、RDS、SLB 的私网 IP 都在该网段内。可选网段如下：
  - 10.0.0.0/8
  - 172.16.0.0/12
  - 192.168.0.0/16
- **交换机**：单击 **添加交换机**，在弹出的 **创建交换机** 窗口，填写以下信息，单击 **提交**。
  - **名称**：交换机名称。长度为 2-128 个字符，以英文字母或中文开头，可包含数字、下划线（_）和短横线（-）。系统自动生成 **单元化工作空间标识**-vsw 组成的默认名称。
  - **可用区**：交换机的可用区。同一 VPC 内不同可用区的交换机内网互通。您需要为每个可用区创建一台交换机。
  - **自定义网段**：默认关闭。开启后，需填写网段地址。交换机的网段可以和其所属的 VPC 网段相同或者是其 VPC 网段的子网。
  - **子网掩码**：**自定义网段** 关闭时，需分别选择子网掩码和网段地址。默认专有的网段掩码是 24 位，例如 172.31.0.0/24，最多可提供 65536 个私网 IP 地址。范围为 16~29 位之间，可提供 4~65532 个地址。
  - **描述**： 输入交换机的描述信息。可包含 2-256 个中英文字符，不能以 `http://` 和 `https://` 开头。

6. 单击 **下一步**，在 **创建安全组** 页面，单击 **添加安全组**，在弹出的 **添加安全组** 窗口，填写以下信息，单击 **提交**。



- **安全组名称**：长度为 2-128 个英文或中文字符，必须以大小字母或中文开头，不能以 `http://` 和 `https://` 开头。可以包含数字、半角冒号（:）、下划线（_）或者连字符（-）。系统自动生成 `单元化工作空间标识-sg` 组成的默认名称。
- **描述**：可包含 2-256 个中英文字符，不能以 `http://` 和 `https://` 开头。
- **规则**：保持默认设置，接受所有的出入访问。

7. 单击 **确定** 返回 **创建安全组** 页面，单击 **提交**。

步骤二：创建集群

集群是运行工作负载的逻辑分组，包含一组云服务器资源，每台云服务器即集群中的一个节点。首次使用 AKS 时，您需要创建一个初始集群，并添加至少一个节点。

前提条件

- 您拥有 AKS 管理员权限。
- 已创建标准工作空间。
- 首次使用 ACK 集群时，需先授权 ACK 访问云资源的权限，详情参考 首次使用容器服务 Kubernetes 版。

操作步骤

1. 登录 AKS 控制台，在左侧导航栏单击 **集群管理** > **集群详情**。
2. 在集群列表页，单击 **创建集群**。
3. 在 **集群创建** 页面，系统会自动进行预检查，确保相关的产品已经开通并且租户下的账户余额大于 100 元。预检查通过后，单击 **下一步**。

   > **说明**
   > 若检查失败，完成修复失败项目后，可以单击 **重新检查** 操作重新进行预检查。

4. 在 **基本配置** 页面，完成以下集群基础选项配置。
   - **基本信息**：
     - **专有网络**：当前工作空间的专有网络名称。不可修改。
     - **集群名称**：应包含 1~63 个字符，可包含数字、汉字、英文字符或短划线（-）。本例中分别输入：`democluster1` 、 `democluster2` 。
     - **Kubernetes 版本**：选择 Kubernetes 版本。支持 `1.16.9-aliyun.1` 、 `1.18.8-aliyun.1` 。本例中选择 `1.18.8-aliyun.1` 。
     - **容器运行时**：支持 `docker 19.03.5` 、 `containerd 1.4.3` 。本例中选择 `docker 19.03.5` 。
   - **网络配置**：
     - **虚拟交换机**：您可以在已有虚拟交换机列表中，根据可用区选择所需交换机。如果没有您需要的交换机，可以通过单击创建交换机链接进行创建，请参见 创建交换机。

- **网络插件**：设置启用的网络插件，目前支持 Terway 网络插件。

  > 说明
  > Terway 是阿里云容器服务自研的网络插件，将阿里云的弹性网卡分配给容器，支持 Kubernetes 的 Network Policy 来定义容器间的访问策略，支持对单个容器做带宽的限流。具体请参见 Terway。

- **Pod 虚拟交换机**：网络插件使用 Terway 时，需要为 Pod 分配 IP 的虚拟交换机。每个 Pod 虚拟交换机分别对应一个 Worker 实例的虚拟交换机。如果没有您需要的交换机，可以通过单击创建交换机链接进行创建，请参见 创建交换机。
- **Service CIDR**：设置 Service CIDR。网段不能与 VPC 及 Pod 地址段重复，创建成功后不能修改。可选范围：10.0.0.0/16-24，172.16-31.0.0/16-24，192.168.0.0/16-24。

- **高级配置**：保持默认设置。更多信息参考 创建集群。

5. 在 **节点配置** 页面，完成以下 Worker 节点配置。
   - **付费类型**：支持 **按量付费** 和 **包年包月** 两种节点付费类型。选择包年包月时，需设置以下参数。
     - **购买时长**：目前支持选择 1、2、3、6 个月和 1-5 年。
     - **自动续费**：设置是否自动续费。
   - **节点数量**：创建 Worker 实例（ECS 实例）的数量。
   - **实例规格**：支持选择多个实例规格。详情请参见 实例规格族。最多可选择 10 个实例规格。

     > 说明
     > 可以选择多个规格作为备选，每个节点创建时，将从第一个规格开始尝试购买，直到创建成功，最终购买的实例规格可能随库存变化而不同。

   - **系统盘**：支持 **高效云盘**、**ESSD 云盘**。
   - **挂载磁盘**：支持 **高效云盘**、**ESSD 云盘**。
   - **操作系统**：支持 CentOS 和 Alibaba Cloud Linux 操作系统。
   - **登录密码**：设置节点的登录密码。8-30 个字符，且同时包含三项（字母、数字和特殊符号）。
   - **确认密码**：确认设置的节点登录密码。
6. 配置完成后，单击 **下一步**。
7. 在集群的 **配置预览** 页面，确认配置无误后，单击 **提交**。

   > 说明
   > - 一个包含多节点的 Kubernetes 集群的创建时间一般约为十分钟。
   > - 系统会自动跳转到 **创建集群详情** 页面。在集群的创建过程中，若出现任务失败的情况，可单击事件查看具体错误详情，或单击 **重试** 或 **忽略**。

步骤三：创建应用

您将创建 2 个应用：aks-vote-front 及 aks-vote-redis。

操作步骤

1. 登录应用管理控制台，在左侧导航栏单击 **应用列表**。
2. 在 **应用列表** 页面，单击 **创建应用**。
3. 在创建页面输入应用信息，单击 **确定**。

- **应用名称**：输入 aks-vote-front。

  > 说明
  > 应用名称在同一租户内必须是唯一的。

- **技术栈**：选择 Spring Boot。
- **应用分组**：选择应用所属分组。若尚未自定义分组，可以选择系统默认分组。
- **应用标签**：应用分级标签，便于进行应用快速检索，支持添加系统标签和自定义标签。
- **应用描述**：应用描述的长度不能超过 500 个字符。选填，如 `这是一个示例应用` 。

新创建的应用出现在应用列表的最上方。应用状态为 **创建中**，数秒后变为 **创建完成**，此时已完成应用创建。
按上述方式创建另外一个应用，应用名称为：aks-vote-redis。

步骤四：准备镜像

AKS 以镜像的方式部署应用服务，创建应用服务前需先准备好镜像。为方便您测试，AKS 准备了以下可以公开访问的示例镜像：

镜像名称	镜像地址
sofa-samples/aks-vote-front	registry-vpc.cn-shanghai.aliyuncs.com/sofa-samples/aks-vote-front:v1
sofa-samples/aks-vote-redis	registry-vpc.cn-shanghai.aliyuncs.com/sofa-samples/aks-vote-redis:v1

步骤五：创建应用服务

创建 2 个有依赖关系的应用服务：aks-vote-redis-appservice、aks-vote-front-appservice。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用发布 > 应用服务**。
2. 在应用服务列表页，单击 **创建**。
3. 在 **创建应用服务** 页面，填写以下信息，并单击 **提交**。
   - **基本信息**
     - **命名空间**：保持默认 default。
     - **应用服务名称** 及 **所属应用** 见下表。

应用服务名称	所属应用	依赖应用
aks-vote-front-appservice	aks-vote-front	aks-vote-redis
aks-vote-redis-appservice	aks-vote-redis	-

     - **有状态应用**：默认关闭。
   - **Pod 模板配置**
     - **容器名称**：输入名称。
     - **镜像选择**：选择 **镜像仓库**，镜像地址参考 准备镜像。
     - **CPU 配置**：**请求核数** 为 500 millicores，**最大核数** 为 1 core。
     - **内存配置**：**请求内存** 为 128 MiB，**最大内存** 为 1 GiB。

- 在 **高级配置** > **环境变量配置** 中，为 aks-vote-front-appservice 应用服务添加对应的 Redis 应用服务的 DNS name 作为 **环境变量**。如下图所示。



- **伸缩配置**
  - **副本伸缩策略配置**：默认为 **固定副本数**。
  - **副本数**：默认 1。应用运行时保持的 Pod 副本数。
- **访问配置**

  应用服务支持三种访问方式：**集群内访问**、**内网访问**、**公网访问**，您可以根据业务需要做好规划。

  为 Redis 服务设置集群内访问方式

  选择 **集群内访问** 会建一个 Cluster IP 类型的 Service，并将流量转发到容器的相应端口上。您可以在创建应用服务时设置访问方式，也可以应用服务创建完成后添加访问方式。

  a. 在 **访问配置** 页面，单击 **创建服务**。

  b. 在 **创建服务** 窗口，填写以下信息后，单击 **提交**。



  - **服务名称**：输入 `aks-vote-redis-appservice` 。
  - **访问方式**：选择 **集群内访问**。
  - **端口映射**：单击 **+**，填写以下信息。其余保持默认设置。
    - **协议**：选择 **TCP**。目前支持 **TCP**、**UDP**。
    - **访问端口**：容器镜像中工作负载程序实际监听的端口，输入 `6379` 。
    - **容器端口**：容器端口映射到集群虚拟 IP 上的端口，输入 `6379` 。

  为 Vote 服务设置公网访问方式

  配置外部 Service 会创建一个公网的 LoadBalancer，并将流量转发到容器的相应端口上。访问方式由公网负载均衡服务地址以及设置的访问端口组成，例如 `10.117.117.117:80` 。

a. 在 **访问配置** 页面，单击 **创建服务**。

b. 在 **创建服务** 窗口，填写以下信息后，单击 **提交**。

- **服务名称**：输入 `aks-vote-front-appservice` 。
- **访问方式**：选择 **公网访问**。
- **端口映射**：单击 **添加端口映射**，填写以下信息，其余保持默认设置。



- **协议**：选择 **TCP**。目前支持 **TCP**、**HTTP**、**HTTPS**。
- **转发规则**：选择 **按权重**。目前支持 **按权重**、**按权重并且最小连接数**。
- **前端端口**：容器镜像中工作负载程序实际监听的端口，输入 `80` 。
- **后端端口**：容器端口映射到集群虚拟 IP 上的端口，输入 `80` 。

- **部署和调度配置**
  为应用服务 aks-vote-redis-appservice、aks-vote-front-appservice 保持系统默认配置。
- **预览并提交**
  在应用服务 **预览** 页面，确认信息无误。

步骤六：创建发布单

通过发布单同时发布上一步中创建的两个有依赖关系的应用服务。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用发布**。
2. 单击 **发布单** 进入发布部署大盘。
3. 在发布部署大盘，单击 **创建发布单**。
4. 在 **创建发布单** 页面，填写以下发布信息后，单击 **下一步**。
   - **基本信息**
     - **标题**：发布标题。
     - **部署单元**：展示当前工作空间下的所有部署单元。
     - **应用服务发布列表**：在 **待选应用服务列表** 中单击选择需要的应用服务：aks-vote-redis-appservice、aks-vote-front-appservice。单击 **>** 图标将应用服务添加到 **已选应用服务列表** 中。
   - **高级配置**
     - **设置应用服务依赖关系**：aks-vote-front-appservice 依赖于 aks-vote-redis-appservice。

5. 在 **预览** 页面确认信息无误后，单击 **创建**。系统会自动跳转到 **发布单详情** 页面，单击 **整体发布** 即可开始发布。

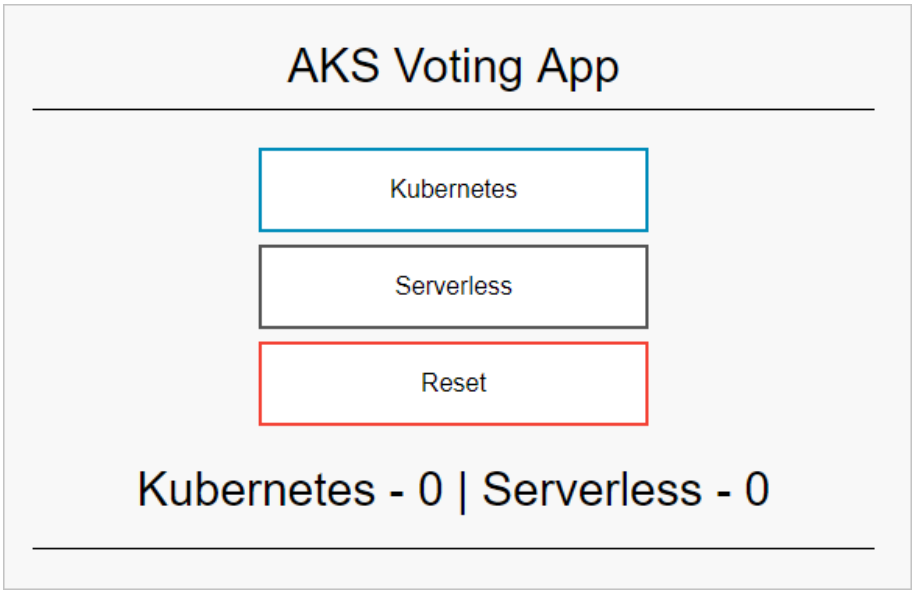
步骤七：验证结果

验证应用服务 aks-vote-front-appservice 及 aks-vote-redis-appservice 是否发布成功。

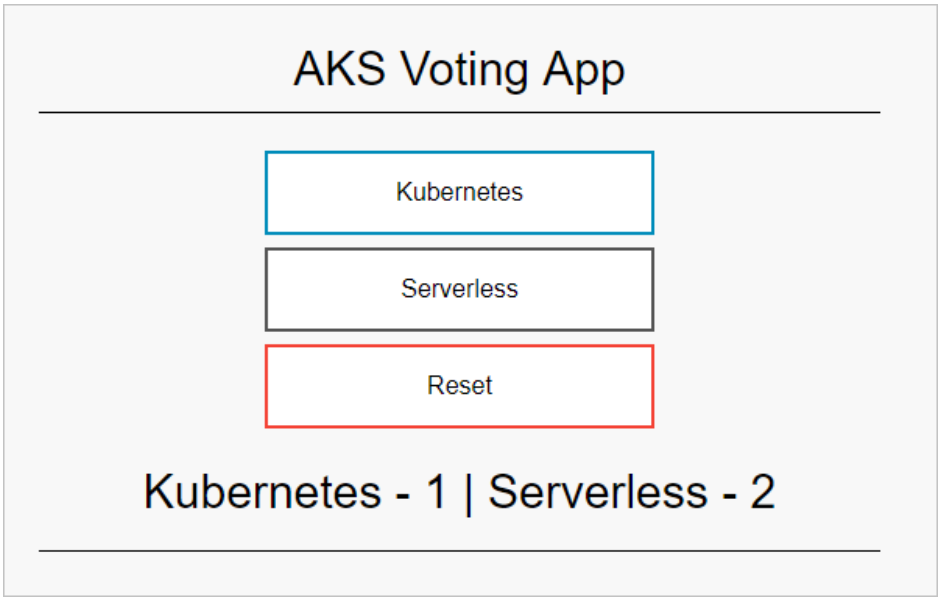
操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用发布**。
2. 在应用服务列表中，单击 **aks-vote-front-appservice** 进入实例详情页。
3. 在 **资源管理** 的 **负载均衡** 页签中找到负载均衡，复制服务地址。
4. 在本地浏览器的地址栏中访问 `http://{负载均衡器服务地址}：80` ，如 `http://10.10.10.1：80` 。

   若页面显示以下内容表示 aks-vote-front-appservice 发布成功。



   单击 kubernetes 或者 Serverless 选项，若页面显示正确投票数据，则表示 aks-vote-redis-appservice 发布成功。如下图所示。



   更多应用服务内容，参考 查看应用服务详情 。

8.3. 集群

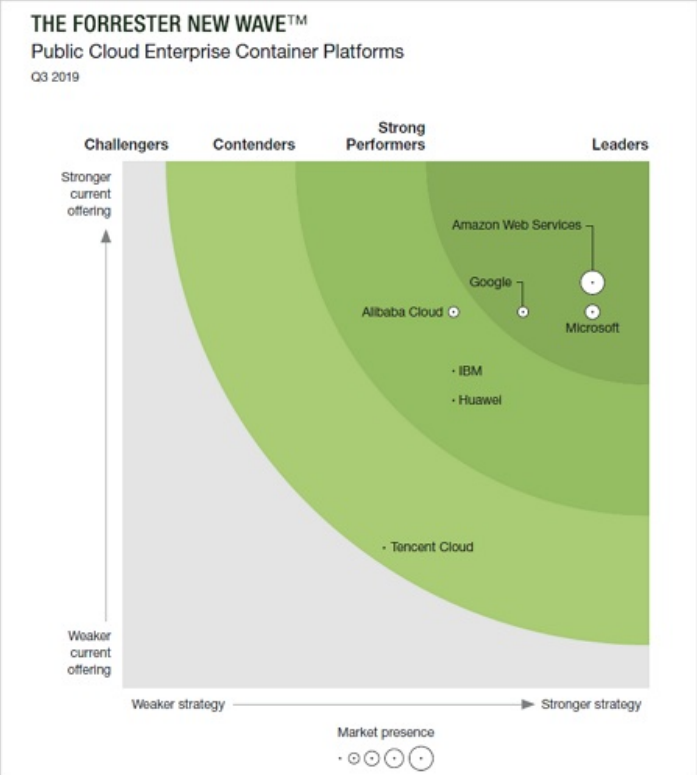
8.3.1. 对接 ACK 产品变更说明

本文介绍 AKS 对接 ACK 托管版集群后，您使用新的 ACK 集群时所需要了解的信息。

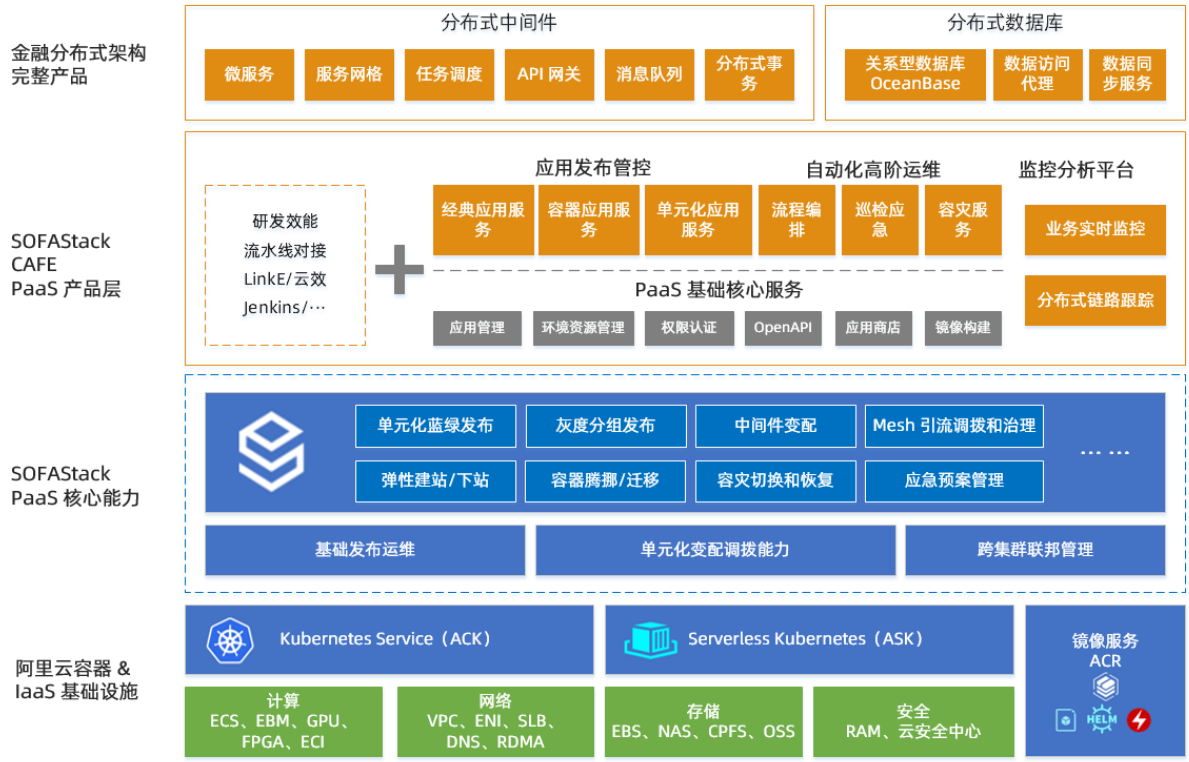
> 说明
>
> - AKS 产品层面兼容新版 ACK 集群和旧版 AKS 集群，两种集群可以共存。已有的旧版 AKS 集群在产品能力上不会受到任何影响。
> - 自功能上线时起（2020 年 8 月 15 号）阿里云底座上新建集群默认为 ACK 托管版集群。

背景信息

容器服务 Kubernetes 版（简称 ACK）提供高性能可伸缩的容器应用管理能力，支持企业级容器化应用的全生命周期管理。ACK 整合阿里云虚拟化、存储、网络和安全能力，打造云端最佳容器化应用运行环境。
2019 年全球知名市场调研机构 Forrester 发布首个企业级公共云容器平台报告。其中，阿里云容器服务的市场表现位居全球前三、中国第一，同时创造中国企业最好成绩，进入强劲表现者象限。



客户价值



AKS 通过深度集成 ACK 托管版集群，希望能够给客户带来如下的业务价值：

- 最大化复用 ACK 容器服务的强大生态和产品能力。
- 方便用户导入现有的 ACK 集群，从而最大化地利用云上资源。
- 集群层面基础运维和能力演进依托 ACK 为主，SOFAStack AKS 将更多专注在上层工作负载建设，为用户提供更加高效易用的企业级分布式应用开发运维平台。

功能演进

AKS 将会分阶段对接 ACK 的产品能力，目前一期实现的 ACK 产品能力如下：

- 集群创建、扩容、导入节点、删除节点等基础能力。
- 容器服务、Service、PV 相关能力。
- ACK 原生工作负载及基础资源监控完成阿里云云监控 CloudMonitor 对接。

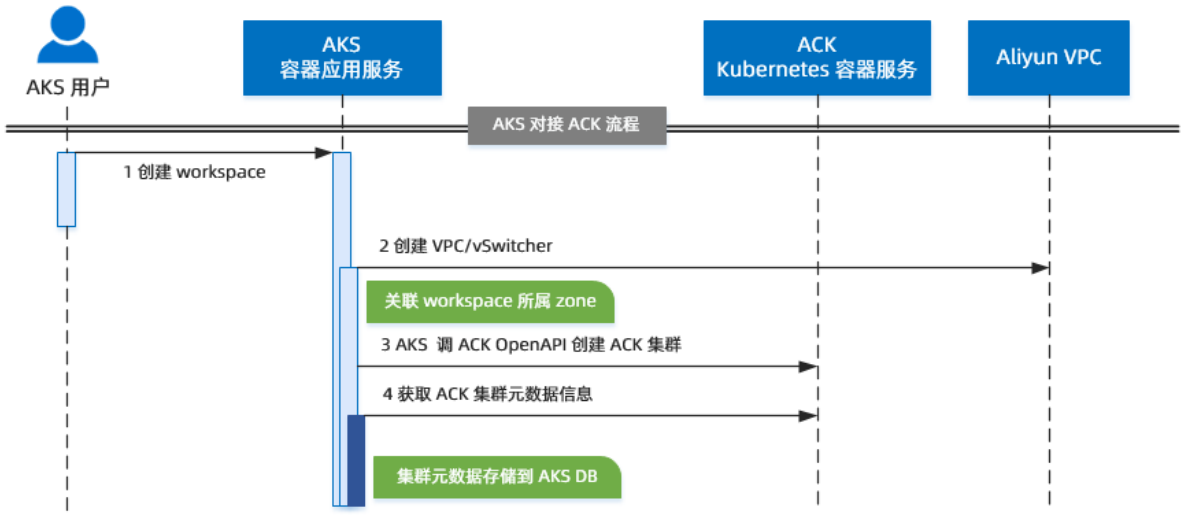
规划二期实现的能力如下：

- 集群导入、释放、删除、更新，实现完整的集群生命周期管理。
- 集群 AddOn 组件管理。
- 对现存的关键 能力限制 的支持。

对接流程

> 说明
> VPC：workspace：ACK cluster 仍保持 1:1:1 的模型关系。一个 workspace 中仅能创建一个 ACK 托管集群。

用户在创建 workspace 过程中创建 VPC, 绑定可用区和交换机。Workspace 创建完成后，在当前 workspace 下创建 ACK 集群。具体的对接流程如下图：



限制与变更

对接 ACK 集群后，原有的 AKS 产品能力的限制和变更如下：

- 目前杭州金区可用区 B 和 D 不支持创建 ACK 集群。
- ACK 事件中心能力降级，切成原生事件中心，事件无持久存储。
- ACK 集群监控与原生负载监控直接对接阿里云 云监控（CloudMonitor），应用服务的 Pod 容器监控暂时空缺，相关页面在 ACK 集群场景下隐藏。
- 节点伸缩（CA）与 Pod 水平伸缩（HPA）功能在 ACK 集群场景下不可用，伸缩功能相关的控制台菜单会自动隐藏。
- KubeConfig 及公网访问功能暂不支持，如需使用需要在 ACK 控制台上获取访问信息。
- 原地升级能力降级，只支持镜像变更场景，其余场景无法保证 IP 不变。
- 集群维度权限仍对接 SOFA 现有的权限体系（FullAccess 和 ReadOnly），详情参考 权限策略。
- 首次使用 ACK 集群时，需先授予 SOFA 操作 ACK 资源及通过 ACK 访问云资源的权限，您需要在 ACK 及 ESS 控制台进行开通并且授权，详情参考 首次使用容器服务 Kubernetes 版 及 开通和授权服务。
- ACK 集群无法并行扩容多个可用区节点。
- PV/Service 只支持 YAML 创建，暂不支持表单创建。

注意事项

对已有原 AKS 集群的用户，如果后面需要新建集群，需要注意以下事项：

- 自功能上线时起（2020 年 8 月 15 号）阿里云底座上默认新建集群为 ACK 集群。

FAQ

Q：同一个租户下的 AKS 集群和 ACK 集群的收费标准是什么？
A：公有云上，ACK 集群保持独立售卖模式，使用过程中产生的集群、带宽、存储、计算相关资源独立计费和出账，并提供独立的 SLA。AKS 当前免费，规划中会根据集群规模进行独立出账。AKS 集群和 ACK 集群的收费标准分别遵照 SOFAStack 产品定价 和 ACK 产品定价 规则。

8.3.2. 创建集群

Kubernetes 托管版集群只需创建 Worker 节点，Master 节点由阿里云容器服务创建并托管，从而让您可以更多关注业务本身。本文介绍如何通过 AKS 控制台创建 Kubernetes 托管版集群。

> 说明
> 从对接 ACK 集群的功能上线时起（2020 年 8 月 15 号），阿里云底座上新建集群默认为 ACK 托管版集群。

前提条件

- 工作空间中没有集群。
- 您拥有 AKS 管理员权限。
- 已有 VPC 网络工作空间。
- 首次使用 ACK 集群时，需先授权 ACK 访问云资源的权限，详情请参考 首次使用容器服务 Kubernetes 版。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **集群管理**。
2. 在 **集群详情** 页，单击 **创建集群**。
3. 在 **集群创建** 页面，AKS 会自动进行预检查，确保已经开通相关产品并且租户下的账户余额多于 100 元。预检查通过后，单击 **下一步**。

   > **说明**
   > - 若检查失败，完成修复失败项目后，可以单击 **重新检查** 操作重新进行预检查。
   > - 由于 SOFAStack 的集群网络插件 CCM 需授权访问云资源以提供负载均衡、统一接入流量管理等服务，预检查步骤会检测当前用户的 RAM Role 列表中是否包含 AliyunSofaCafeCCMRole 角色，若没有需要您按照 CCM 角色授权 进行角色授权。

4. 在 **基本配置** 页面，完成集群配置后，单击 **下一步**。
   i. 完成集群基础选项配置。

配置项	说明
专有网络	当前工作空间的专有网络名称。不可修改。
集群名称	填写集群的名称。 **说明** 集群名称应包含 1~63 个字符，可包含数字、汉字、英文字符或短划线（-）。
Kubernetes 版本	选择 Kubernetes 版本。可选择 `1.16.9-aliyun.1` 或 `1.18.8-aliyun.1`。
容器运行时	支持 `docker 19.03.5`、`containerd 1.4.3`。

ii. 完成集群网络配置。

配置项	说明
虚拟交换机	您可以在已有虚拟交换机列表中，根据可用区选择 1~3 个交换机。 **说明** 如果没有您需要的交换机，可以通过单击创建交换机链接进行创建，请参考 创建交换机。
网络插件	设置启用的网络插件。当前支持 Terway 网络插件。 **说明** Terway 是阿里云容器服务自研的网络插件，将阿里云的弹性网卡分配给容器，支持 Kubernetes 的 Network Policy 来定义容器间的访问策略，支持对单个容器做带宽的限流。
Pod 虚拟交换机	为 Pod 分配 IP 的虚拟交换机。每个 Pod 虚拟交换机分别对应一个 Worker 实例的虚拟交换机。 **说明** 如果没有您需要的交换机，可以通过单击创建交换机链接进行创建，请参考 创建交换机。
Service CIDR	设置 Service CIDR。网段不能与 VPC 及 VPC 内已有 Kubernetes 集群使用的网段重复，创建成功后不能修改。

iii. 完成集群高级配置。

配置项	说明
配置 SNAT	不可修改，默认选中。创建集群时，默认不开通公网。如果您选择的 VPC 不具备公网访问能力，选中为专有网络配置 SNAT 后，ACK 将为您创建 NAT 网关并自动配置 SNAT 规则。
公网访问	设置是否开放使用 EIP 暴露 API Server。API Server 提供了各类资源对象（Pod、Service 等）的增删改查及 watch 等 HTTP Rest 接口。 ■ 若选择开放，会创建一个 EIP，并挂载到内网 SLB 上。此时，Master 节点的 6443 端口（对应 API Server）暴露出来，您可以在外网通过 kubeconfig 连接并操作集群。 ■ 若选择不开放，则不会创建 EIP，您只能在 VPC 内部用 kubeconfig 连接并操作集群。 **说明** 若需获取 kubeconfig 信息，请前往 ACK 控制台。

配置项	说明
kube-proxy 代理模式	支持 iptables 和 IPVS 两种模式。 ■ iptables：成熟稳定的 kube-proxy 代理模式，Kubernetes Service 的服务发现和负载均衡使用 iptables 规则配置，但性能一般，受规模影响较大，适用于集群存在少量的 Service 的场景。 ■ IPVS：高性能的 kube-proxy 代理模式，Kubernetes Service 的服务发现和负载均衡使用 Linux ipvs 模块进行配置，适用于集群存在大量的 Service，对负载均衡有高性能要求的场景。
集群删除保护	防止通过控制台或者 API 误删除集群。默认选中，可修改。
部署单元	默认勾选使用默认配置。若取消勾选，可以为可用区设置不同的部署单元，若需要修改部署单元，可前往 **全局设置 > 工作空间详情** 进行设置。
日志服务	支持 **创建新的 Project**、**使用已有Project**。 ■ **创建新的 Project**：默认处于选中状态，将自动创建名称为 `k8s-log-{ClusterID}` 的 Project。 说明 ClusterID 表示新建集群的唯一标识。 ■ **使用已有Project**：从现有 Project 下拉列表中选择一个 Project 来管理采集的日志。 

5. 在 **节点配置** 页面，完成以下 Worker 节点配置后，单击 **下一步**。

配置项	说明
付费类型	支持 **按量付费** 和 **包年包月** 两种节点付费类型。选择 **包年包月** 时，需设置以下参数。3 ○ **购买时长**：目前支持选择 1、2、3、6 个月和 1~5 年。 ○ **自动续费**：设置是否自动续费。
节点数量	创建 Worker 实例（ECS 实例）的数量。
实例规格	支持选择多个实例规格。详情请参考 实例规格族。
已选实例规格	呈现选中的实例规格。
系统盘	支持 **ESSD 云盘** 和 **高效云盘**。
挂载磁盘	支持 **ESSD 云盘** 和 **高效云盘**。
操作系统	支持 CentOS 和 Aliyun Linux 操作系统。

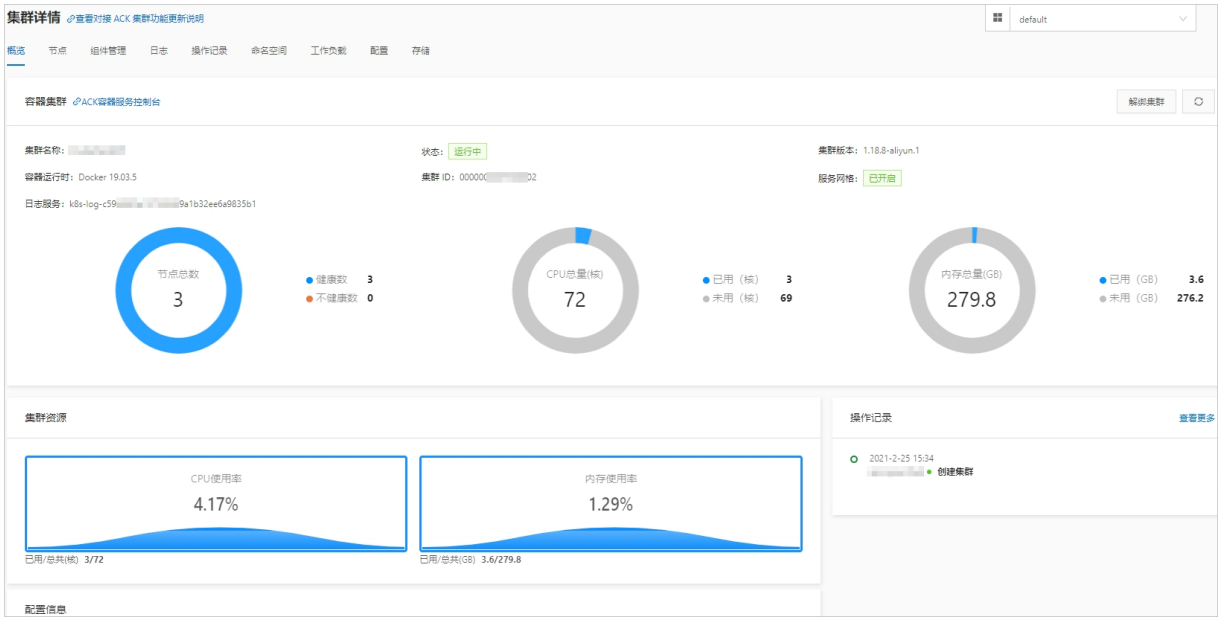
配置项	说明
登录密码	设置节点的登录密码。8～30 个字符，且同时包含三项（字母、数字和特殊符号）。
确认密码	确认设置的节点登录密码。

6. 在集群的 **配置预览** 页面，确认配置无误后，单击 **提交**。

> 说明
> 一个包含多节点的 Kubernetes 集群的创建时间一般约为十分钟。

执行结果

集群创建成功后，您可以在 **集群详情** 页面查看所创建的集群信息。



8.3.3. 集群扩容

集群扩容即扩容集群内的节点资源。通过容器应用服务控制台，您可以根据实际业务需要对集群的节点进行扩容。

若节点资源不足，会导致创建工作负载等操作失败。您可以通过以下两种方式进行扩容：

- 创建新节点
- 添加已有节点

前提条件

已创建集群。

创建新节点

1. 登录容器应用服务控制台，在左侧导航栏单击 **集群管理**。
2. 在 **集群详情** 页，单击 **节点** 页签，进入节点列表页面。
3. 单击 **集群扩容**，完成以下 **集群扩容** 配置。
   - **虚拟交换机**：设置虚拟交换机。您可以在已有虚拟交换机列表中，根据可用区选择 1~3 个交换机 。如果没有您需要的交换机，可以通过单击创建虚拟交换机进行创建。

- **付费类型**：支持 **按量付费** 和 **包年包月** 两种节点付费类型，付费详情请参见计费方式对比。选择 **包年包月** 时，需设置购买时长。目前支持选择 1、2、3、6个月和 1~5 年。需设置是否自动续费。
- **伸缩数量（扩容）**：需要扩容的节点数量。
- **实例规格**：支持选择多个实例规格。详情参见 实例规格族。
- **系统盘**：至少 40 GB。支持 SSD 云盘和高效云盘。
- **挂载磁盘**：至少 40 GB。支持 SSD 云盘和高效云盘。
- **密码**：设置节点的登录密码。如遗忘可登录阿里云 ECS 控制台重置密码，参见 重置实例登录密码。
- **确认密码**：确认设置的节点登录密码。
- **云监控插件**：设置是否启用云监控插件。您可以选中在 ECS 节点上安装云监控插件，从而在云监控控制台查看所创建 ECS 实例的监控信息。
- **标签**：为节点绑定标签。输入键和对应的值，单击 **添加** ，在弹出框中输入 **变量名**、**值**。
- **污点（Taints）**：为 Worker 节点添加污点。单击 **添加**，在弹出框中输入 **变量名**、**值**，选择 **Effect**。Effect 包含以下选项：
  - NoSchedule：不允许无匹配 toleration 的 Pod 调度到该节点。
  - NoExecute：若无匹配 toleration 的 Pod 已经在节点上运行，则将 Pod 驱逐，若无匹配 toleration 的 Pod 尚未在节点上运行，则不会将 Pod 调度到该节点上。
  - PreferNoSchedule：尽量不将无匹配 toleration 的 Pod 调度到该节点。更多详情参见 官方文档。
- **部署单元**：使用默认配置。
- **CPU Policy**：设置 CPU policy。
  - **none**：默认策略，表示启用现有的默认 CPU 亲和方案。
  - **static**：允许为节点上具有某些资源特征的 Pod 赋予增强的 CPU 亲和性和独占性。

4. 完成配置后，单击 **提交**。

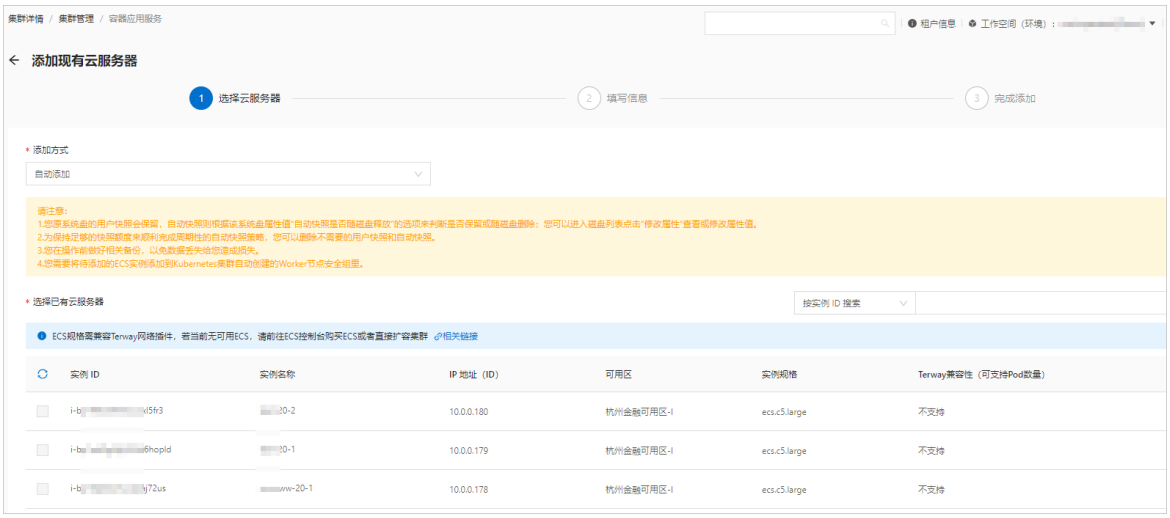


添加已有节点

您可以将工作空间中已有的处于 **运行中** 状态的 ECS 加入到容器服务集群。系统会从当前工作空间中筛选出之前以 **清除机器环境** 方式删除的 ECS，以及从当前工作空间中筛选出由资源管理（RMC）创建的符合条件的 ECS。

> **说明**
> 您需要将待添加的 ECS 实例添加到集群自动创建的节点安全组内。

1. 登录容器应用服务控制台，在左侧导航栏单击 **集群管理**。
2. 在 **集群详情** 页，单击 **节点** 页签，进入节点列表页面。
3. 单击 **添加已有节点**，完成以下 **集群扩容** 配置。



   - **添加方式**：默认为自动添加。自动添加节点方式会列出当前账号下可用的 ECS 云服务器，在 Web 界面进行安装部署，并自动添加到集群。
   - **选择已有云服务器**：勾选要添加的 ECS。
4. 单击 **下一步**，填写以下实例信息。
   - **数据盘挂载**：设置是否将容器和镜像存储在数据盘。

     > **说明**
     > - 如果 ECS 已挂载数据盘，且最后一块数据盘的文件系统未初始化，系统会自动将该数据盘格式化为 ext4，用来存放内容/var/lib/docker、/var/lib/kubelet 。数据盘内原有数据将丢失，请注意备份数据。
     > - 如果 ECS 未挂载数据盘，则不会挂载新的数据盘。

   - **CPU Policy**：设置 CPU Policy。
     - **none**：默认策略，表示启用现有的默认 CPU 亲和方案。
     - **static**：允许为节点上具有某些资源特征的 Pod 赋予增强的 CPU 亲和性和独占性。
   - **密码**：设置节点的登录密码。
   - **确认密码**：确认密码。
   - **部署单元**：使用默认配置。
   - **标签**：为节点绑定标签。输入键和对应的值，单击 **添加**，在弹出框中输入 **变量名**、**值**。
5. 完成配置后，单击 **提交**。

8.3.4. 删除集群

您可以删除不需要的集群，以免产生不必要的费用。本文介绍如何删除集群。

> 说明
>
> - 无法直接通过解绑集群来删除集群，因为解绑集群只是解除集群与当前工作空间的关联关系，集群解绑后的集群不会被删除。
> - 集群解绑后可通过集群导入功能继续使用，详情请参见 导入 ACK 集群。

前提条件

已关闭删除保护设置。

操作步骤

1. 登录 ACK 控制台，在左侧导航栏单击 **集群**。
2. 在 **集群列表** 页，选择所需的集群并单击右侧的 **更多** > **删除**。
3. 在 **删除集群** 弹框中，勾选 **我已知晓以上信息并确认删除集群**，单击 **确定**。



相关操作

关闭删除集群保护设置

1. 登录 ACK 控制台，在左侧导航栏单击 **集群**。
2. 在 **集群列表** 页，单击集群名称 > **基本信息**。
3. 关闭 **删除保护**。

8.3.5. 导入 ACK 集群

AKS 对接 ACK 托管版集群后，您可以将 ACK 集群导入 AKS。本文介绍如何通过 AKS 控制台界面快速导入已经创建好的 ACK 集群。

> 说明
> - 目前，仅支持导入 ACK 标准托管版，且容器运行时必须为 Docker。
> - 待导入的 ACK 集群的 VPC 必须与当前工作空间所属的 VPC 一致。

前提条件

- 您拥有 AKS 管理员权限。
- 您已在 ACK 控制台创建了一个标准托管版集群，并获取了该集群所属的专有网络 ID（VPC ID）。

操作步骤

创建工作空间

1. 进入 SOFAStack 控制台，单击 **全局设置**，进入工作空间管理页面。
2. 单击 **创建工作空间**，配置工作空间：
   - 基本信息：
     - **工作空间标识**：输入工作空间的英文标识，全局唯一，一经确定无法修改。
     - **工作空间名称**：输入工作空间的显示名称。
     - **地域**：选择工作空间所在的地域（Region）。
     - **网络类型**：选择 **VPC 网络**。
     - **导入 VPC**：开启 **导入 VPC** 模式，并在下拉框中选择待导入的 ACK 集群所属的 VPC ID。
     - **可用区**：选择工作空间配置的可用区（Available Zone），最多不能超过两个。
   - 创建 VPC：在 **导入 VPC** 模式下，系统会自动识别 VPC 下的资源，无需操作。
3. 单击 **提交**。

导入集群

1. 进入 **容器应用服务 > 集群管理**，在空集群页面，选择 **导入集群**。
2. 在 **导入集群** 页面的已有集群列表中，选择您需要导入的 ACK 集群。
3. 单击 **提交**，即可导入该集群。

8.3.6. 查看集群详情

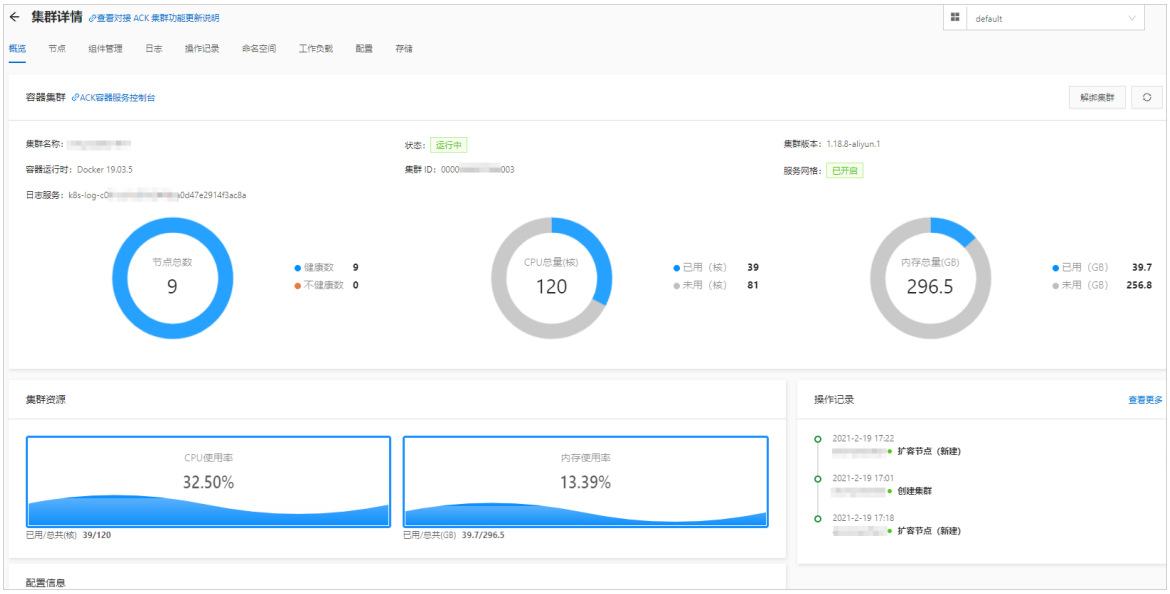
集群创建成功后，您可以在 **集群详情** 页面查看详情，包括集群的基本信息、配置信息、集群资源和集群操作记录，方便您快速了解集群的健康状态信息及资源使用情况。

前提条件

已创建集群。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **集群管理**。
2. 在 **集群详情** 页，查看以下信息：



- **概览**：
  - 集群的节点总数、CPU 使用率、内存使用率、基本配置信息、网络配置信息、操作记录。
  - 单击容器集群旁的 **ACK 容器服务控制台**，可跳转至 ACK 集群详情页，查看该集群的详细信息。
  - 在基本信息区域，可查看服务网格的开关状态。

    > **说明**
    > 目前只能开启服务网格，暂时不支持关闭。

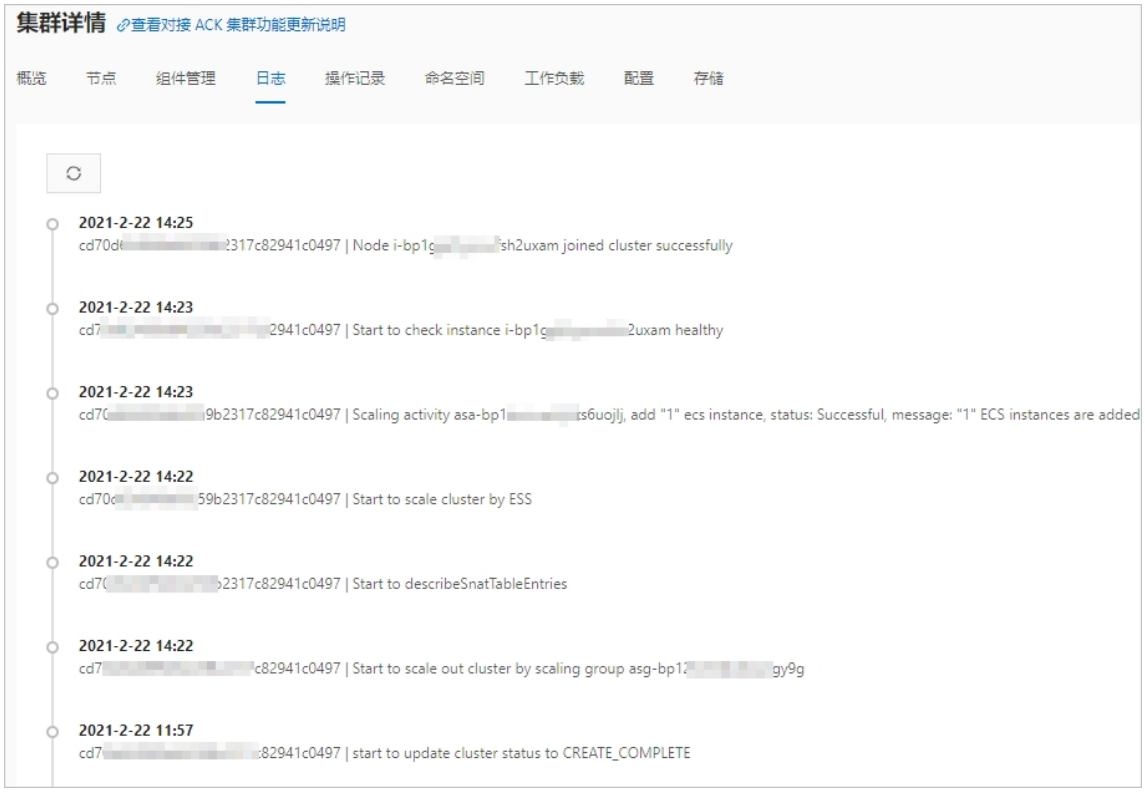
- **节点**：查看节点名称、IP 地址、状态、Pod 数、配置等信息。

  > **说明**
  > 节点相关操作，请参见 节点管理。

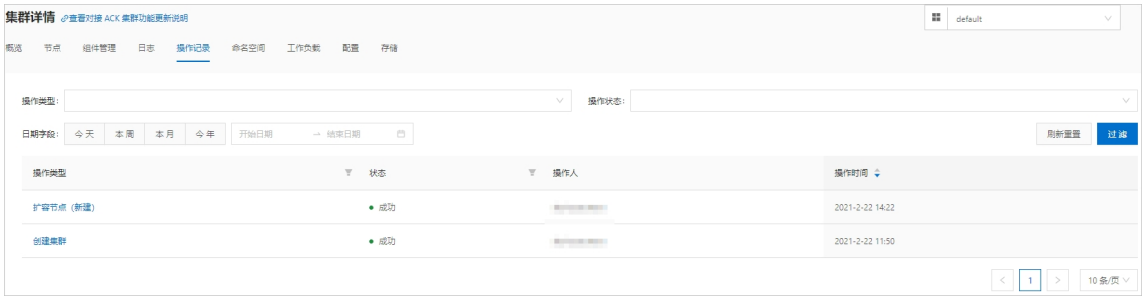
- **组件管理**：提供系统组件和可选组件这两种，用于管理集群的扩展功能。详情参见 组件管理。



- **日志**：查看集群日志。



- **操作记录**：可以查看操作类型、操作状态信息。可根据操作类型、操作状态、日期字段进行筛选查找。



- **命名空间**：可以查看命名空间名称、状态等信息。

> 说明
> 命名空间相关操作，请参见 命名空间。

- **工作负载**：即 Kubernetes 对一组 Pod 的抽象模型，包括 Deployments、Statefulsets、Daemonsets 等多种类型。在该页面可以查看工作负载名称、镜像、Pod 实际数目等信息。

> 说明
> 工作负载相关操作，请分别参见 部署（Deployments）、有状态副本集（Statefulsets）、守护进程集（Daemonsets）、容器组（Pods）。

- **配置**：可以查看名称、命名空间、创建时间等信息。

> 说明
> 配置相关操作，请分别参见 配置项、保密字典。

- **存储**：为工作负载提供存储的组件。目前，AKS 支持 NAS、云盘（cloud_disk）这两种存储类型。

> 说明
> 存储相关操作，请分别参见 存储类型（StorageClass）、存储卷（PV）、存储卷声明（PVC）。

8.3.7. 节点管理

节点是集群组成的基本元素，用以运行应用负载。它既可以是虚拟机，也可以是物理机，具体视业务而定。每个节点都包含运行 Pod 所需的基本组件，包括 Kubelet、Kube-proxy、容器运行时等。您可以通过容器应用服务控制台对节点进行管理。

> 说明
> 本文仅适用于新版 ACK 集群场景，即在 AKS 集成阿里云容器服务（ACK）之后（2020 年 8月 15 号）用户创建的集群。

设置节点调度

您可以通过控制台界面设置节点调度，从而合理分配各节点的负载。

前提条件

已创建集群。

暂停调度

节点暂停调度后，在后续进行应用部署时，Pod 不会再调度到该节点。

1. 登录容器应用服务控制台，在左侧导航栏单击 **集群管理**。
2. 在 **集群详情** 页，单击 **节点** 页签，进入节点列表页面。
3. 选择节点，单击 **更多操作** > **暂停调度**，在弹出框中单击 **确定**。
   回到节点列表页，查看节点状态为 **不可调度**。

恢复调度

1. 登录容器应用服务控制台，在左侧导航栏单击 **集群管理**。
2. 在 **集群详情** 页，单击 **节点** 页签，进入节点列表页面。
3. 选择节点，单击 **更多操作** > **恢复调度**，在弹出框中单击 **确定**。
   回到节点列表页，查看节点状态为 **可调度**。

查看节点详情

前提条件

已创建集群。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **集群管理**。
2. 在 **集群详情** 页，单击 **节点** 页签，进入节点列表页面。



3. 在 **节点列表** 中，单击 **节点名称** 进入节点详情页面。您可以查看以下信息。

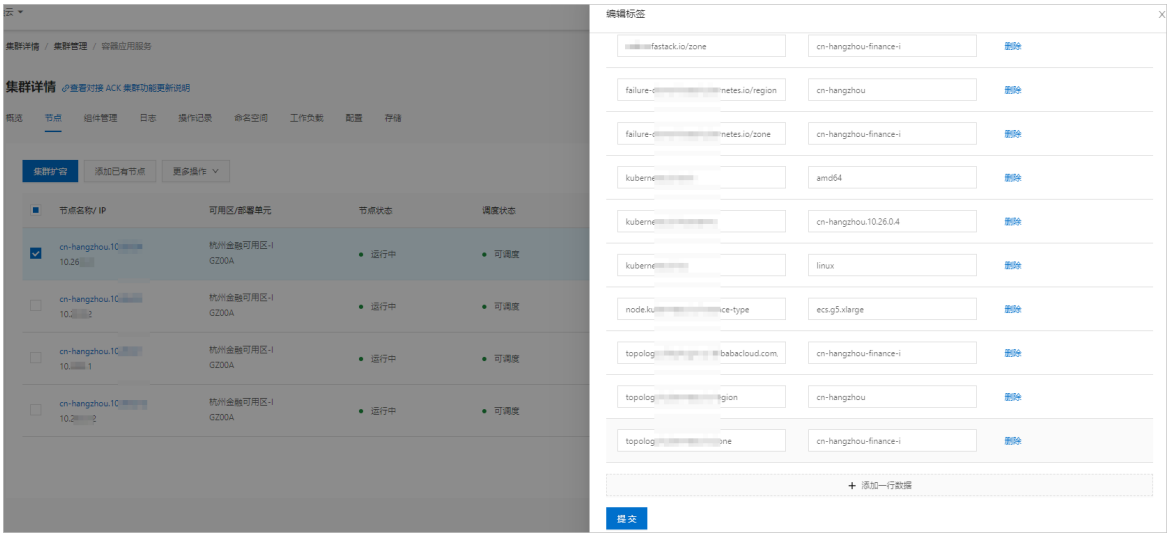
信息模块	具体信息
基本信息	节点服务器 IP
	节点的创建时间
	调度状态： ○ 可调度状态 ○ 不可调度状态
	机器详情：单击 **前往查看** 进入云服务器管理控制台，可查看节点服务器详细信息，例如基本信息及配置、付费、监控等信息。
	注释：系统注释信息
	系统定义的节点标签，可用于节点分组。
Pod 列表	节点所包含的容器列表，包括 Pod 名称、IP、命名空间、 状态、创建时间。单击 Pod 名称跳转到容器组详情页。
节点事件	反映节点生命周期和运行状况的事件日志，包含事件名、次数、事件原因、事件内容等。
YAML	节点的 YAML 文件，支持一键复制。

管理标签

您可以通过容器应用服务对标签进行管理，包括快速添加和删除标签。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **集群管理**。
2. 在 **集群详情** 页，单击 **节点** 页签，进入节点列表页面。
3. 找到目标节点，单击右侧操作栏的更多操作图标 > **编辑标签**。
4. 在 **编辑标签** 弹框中，可以根据需要添加或删除标签：



- **添加标签**：

a. 单击下方 **添加一行数据**。

b. 在弹出框中输入 **key**、**value**。

c. 输入完成后，单击 **提交**。

- **删除标签**：

a. 找到待删除的标签，单击右侧 **操作** 列下的 **删除**。您可以看到该标签消失，标签被删除。

b. 单击 **提交**。

管理 Taint

Taint（污点）用于给节点做污点标记，使 Pod 排斥一类特定的节点，每个节点上都可以应用一个或多个污点。本部分介绍如何对污点进行管理。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **集群管理**。
2. 在 **集群详情** 页，单击 **节点** 页签，进入节点列表页面。
3. 找到目标节点，单击右侧操作栏的更多操作图标 > **编辑 Taint**。
4. 在 **编辑 Taint** 弹框中，可以根据需要添加或删除污点：



- **添加污点**：

a. 单击下方 **添加一行数据**。

b. 在弹出框中输入 **键**（必填）、**值**（选填）。对于 Effect，可在 NoSchedule、NoExecute 或 PreferNoSchedule 中选择。

> 说明
> 当添加多个污点时，各污点的键不能重复，但值可以重复，并且可以为空。

c. 完成配置后，单击 **提交**。

- **删除污点**：

a. 找到待删除的污点，单击右侧 **操作** 列下的 **删除**。您可以看到该污点消失，污点被删除。

b. 单击 **提交**。

监控节点

ACK 集群与云监控服务无缝集成，您可以查看 kubernetes 节点的监控信息，了解 Kubernetes集群下 ECS 实例的节点监控指标。

> 说明
> ACK 集群默认开通了云监控的基础资源监控服务，您还可以在云监控控制台查看原生 k8s 工作负载的监控指标。详情请参见 基础资源监控。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **集群管理**。
2. 在 **集群详情** 页，单击 **节点** 页签，进入节点列表页面。
3. 找到目标节点，单击节点右侧 **操作** 栏的更多图标 > **监控** 进入云监控控制台。
4. 查看对应 ECS 实例的基本监控信息，包括 CPU 使用率、网络流入带宽、网络流出带宽、系统磁盘 BPS、系统盘 IOPS 等指标。

删除节点

当您不再需要该节点继续工作时，可以通过以下操作移除节点。

前提条件

已创建集群。

注意事项

- 删除节点会引起 Pod 迁移，可能会影响业务，建议在业务低峰期操作。
- 操作过程中可能存在非预期风险，请提前做好相关的数据备份。
- 操作过程中，后台会把当前节点设置为不可调度状态。
- 移除节点仅移除 Worker节点，不会移除 Master 节点。
- 移除节点请通过控制台进行操作，如果使用 `kubectl delete node` 方式手动移除节点，则：
  - 移除后的节点无法再添加到其他集群上。
  - 删除集群时，该节点所在的 ECS 实例会被释放。

操作步骤

1. 在左侧导航栏，单击 **集群管理**，进入集群详情页面。
2. 选择 **节点** 页签，找到待删除的节点，单击右侧操作栏的更多操作图标 > **删除**。
3. 在弹出的 **删除警告** 窗口，选择以下删除方式，单击 **确定**。
   - **自动排空节点（drain）**：把待移除节点上的 Pod 转移到其他节点。请确保集群其他节点的资源充足。您还可以通过执行命令 `kubectl drain node-name` 的方式把待移除节点上的 Pod 转移到其他节点。

     > 说明
     > node-name 格式为 `your-region-name.node-id` 。 `your-region-name` 为您集群所在的地域名称。 `node-id` 为待移除节点所在的 ECS 实例 ID。例如 `cn-hanghzou.i-xxx` 。

   - **同时释放 ECS**：释放 ECS 实例仅释放按量付费的 ECS 实例。对于预付费 ECS 实例，计费周期到期后，ECS 实例会自动释放。 您也可以在 ECS 实例到期前：
     - 申请退款，提前释放实例，请参见 退款规则及退款流程。
     - 将计费方式转为按量付费后释放实例，请参见 包年包月转按量付费。

     > 说明
     > 若不选择同时释放 ECS，该节点所在的 ECS 实例会继续计费。

8.3.8. 组件管理

集群 AddOn 组件是容器应用服务提供的扩展功能包，用于管理集群的扩展功能。您可以根据业务需求选择所需的组件。

安装组件

前提条件

已创建集群。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **集群管理 > 集群详情**。
2. 在 **集群详情** 页，选择 **组件管理** 页签。
3. 找到需要安装的组件，单击所需组件卡片右下角的 **安装**。
   - **cafeextcontroller**：系统组件。用于应用发布部署。
   - **oamcontroller**：可选组件。用于 OAM 应用发布部署。
   - **mosn-ingress**：系统组件。用于 Ingress 路由配置。
   - **apiserver-network-proxy**：系统组件。用作集群 kube-apiserver 代理。
   - **cloud-controller-manager**：系统组件。用于为 k8s 应用创建负载均衡及统一接入路由管理。

> 说明
> 系统组件是必装组件，在创建集群时系统默认安装。

执行结果

组件安装完成后，卡片右上角的状态会变为 **已安装**。



重装组件

当组件运行异常时，您可以使用重装功能。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **集群管理 > 集群详情**。
2. 在 **集群详情** 页，选择 **组件管理** 页签。
3. 找到需要重装的组件，单击组件卡片的 **重装 > 确定**。

升级组件

> 注意
> - 请务必通过“组件管理-升级”功能对集群组件进行升级，否则可能会带来不可预知的问题。
> - 升级组件前，请确保集群内所有节点均处于运行中状态，否则可能会导致组件升级失败。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **集群管理 > 集群详情**。
2. 在 **集群详情** 页，选择 **组件管理** 页签。
3. 找到待升级的组件，单击组件卡片的 **升级**。

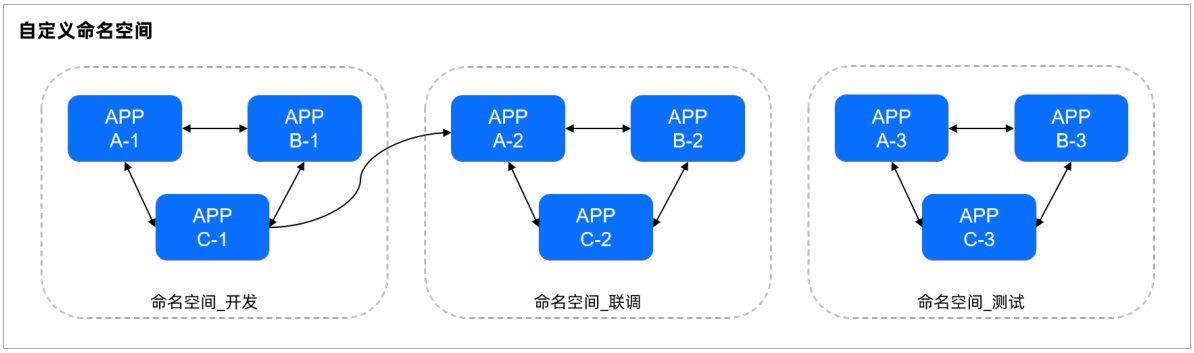
8.3.9. 命名空间

命名空间（namespace）是对一组资源和对象的抽象整合。在同一个集群内可创建不同的命名空间，不同命名空间中的数据彼此隔离，使得它们既可以共享同一个集群的服务，也能够互不干扰。例如，将开发环境、测试环境的业务分别放在不同的命名空间。

命名空间类别

命名空间分为以下两类：

- 系统默认的命名空间：集群在启动时会默认创建以下命名空间。
  - l7-cluster：部署系统组件。
  - default：在不指定命名空间时，默认使用 default。
  - kube-node-lease：系统命名空间。
  - kube-public：用来部署公共插件、容器模板等。
  - kube-system：部署系统组件。
- 用户自定义的命名空间：您可以按照需要创建命名空间，例如开发环境、联调环境和测试环境分别创建对应的命名空间（见下图）。或者按照不同的业务创建对应的命名空间，例如，系统分为登录和游戏，可以分别创建对应命名空间。



创建命名空间

前提条件

已创建集群。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **集群管理**。
2. 在左侧导航栏，单击 **命名空间**，进入命名空间列表页面。
3. 单击 **创建** 进入 **创建命名空间** 页面，输入以下信息，单击 **创建**。
   - **名称**：输入命名空间名称。名称必须唯一。
   - **标签**：为命名空间添加自定义标签。

删除命名空间

删除命名空间将会释放该命名空间下的所有资源。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **集群管理**。
2. 在左侧导航栏，单击 **命名空间**，进入命名空间列表页面。
3. 单击命名空间右侧 **操作** 栏中的 **删除**， 在弹出的对话框中单击 **确认**。
   系统会弹出删除命名空间成功的提示。

查看命名空间详情

您可以通过控制台查看命名空间详情。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **集群管理**。
2. 在左侧导航栏，单击 **命名空间**，进入命名空间列表页面。
3. 单击命名空间名称进入命名空间详情页。您可以查看以下信息：



   - 命名空间名称
   - 状态
   - 创建时间
   - 标签

8.4. 镜像中心

8.4.1. 镜像构建

镜像构建提供将应用的源代码编译成镜像的服务能力。创建构建后，系统会自动在 ACR 镜像中心上创建出以应用名为仓库名称的镜像仓库。镜像成功构建后，会自动推送到该镜像仓库。

创建构建

本部分介绍如何通过容器应用服务控制台完成镜像构建操作。

前提条件

- 已创建应用。
- 拥有 AKS 开发者角色。
- 已在容器镜像服务（ACR）中 创建命名空间。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **镜像中心 > 镜像构建**。
2. 在构建列表页，单击 **创建镜像构建**。
3. 在 **创建镜像构建** 页面，填写以下配置信息，单击 **提交** 完成应用构建。
   - 基础配置
   - 高级配置

构建配置

基础配置

- **所属应用**：选择构建要关联的应用。
- **构建集名称**：选填。
- **仓库命名空间**：选择在 ACR 上创建的命名空间。
- **镜像名**：镜像名称默认与应用名称相同，无法修改。

- **代码库类型**：必须配置代码库才可以进行构建。目前支持 **外部 GitLab**：可配置支持公网访问并且版本在 8.8 及以上的 GitLab （同时支持 GitLab v3 和 v4 版本 API）。 除 **外部 GitLab 地址** 外，还需配置 外部 GitLab Token。

高级配置（选填）



- **源镜像中心访问配置**：Dockerfile 中 FROM 镜像所在的镜像中心访问配置，可配置多个。
- **上传到自定义镜像中心**：默认关闭。用于配置构建成功后额外的 Push 镜像的镜像中心访问信息，如不做配置，默认只会 Push 至 SOFAStack 平台默认的镜像中心。若开启，需完成有以下配置项：



  - **自定义镜像名**：输入一个包含 `registryurl/namespace/repository` 的字符串，比如：`registry.cn-hangzhou.antcloud.com/mytenant/sayhello` 。
  - **用户名**：选填。当所配置的镜像中心需要用户名和密码进行登录时进行配置。
  - **密码**：选填。当所配置的镜像中心需要用户名和密码进行登录时进行配置。

- **构建路径**：选填。执行 `docker build` 所在的工作目录，相对源代码 checkout 后的路径。默认为源代码根目录。
- **Dockerfile 路径**：选填。执行 `docker build` 时的 `--file` 参数。默认为源代码根目录下 Dockerfile 文件。

触发构建

您可以在完成本地代码更新后触发构建镜像。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **镜像中心 > 镜像构建**，进入构建列表页面。
2. 单击 **构建集名称** 进入构建详情页。

3. 在 **构建历史** 页签中，单击 **触发构建**。
4. 在弹出的 **触发构建** 对话框中，填写以下参数，单击 **确定**。
   - **构建分支/标签**：代码的分支，例如：`master` 、 `dev` 。
   - **自定义构建参数**：输入自定义的 key、value 值，作为执行 `docker build` 的 `--build-arg` 参数。

删除构建

您可以在控制台上删除不需要的构建。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **镜像中心 > 镜像构建**，进入构建列表页。
2. 单击构建 **操作** 栏中的 **删除** ，在弹出的 **删除构建** 对话框中单击 **确定**。

查看构建历史

完成镜像构建后，就可以进入该构建详情页面查看构建详情。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **镜像中心 > 镜像构建**，进入构建列表页。
2. 单击 **构建集名称** 进入构建集详情页。
3. 在构建集详情页，查看镜像构建基本信息及构建历史信息。

   您可以单击 **镜像版本** 右侧的复制按钮 复制镜像版本。

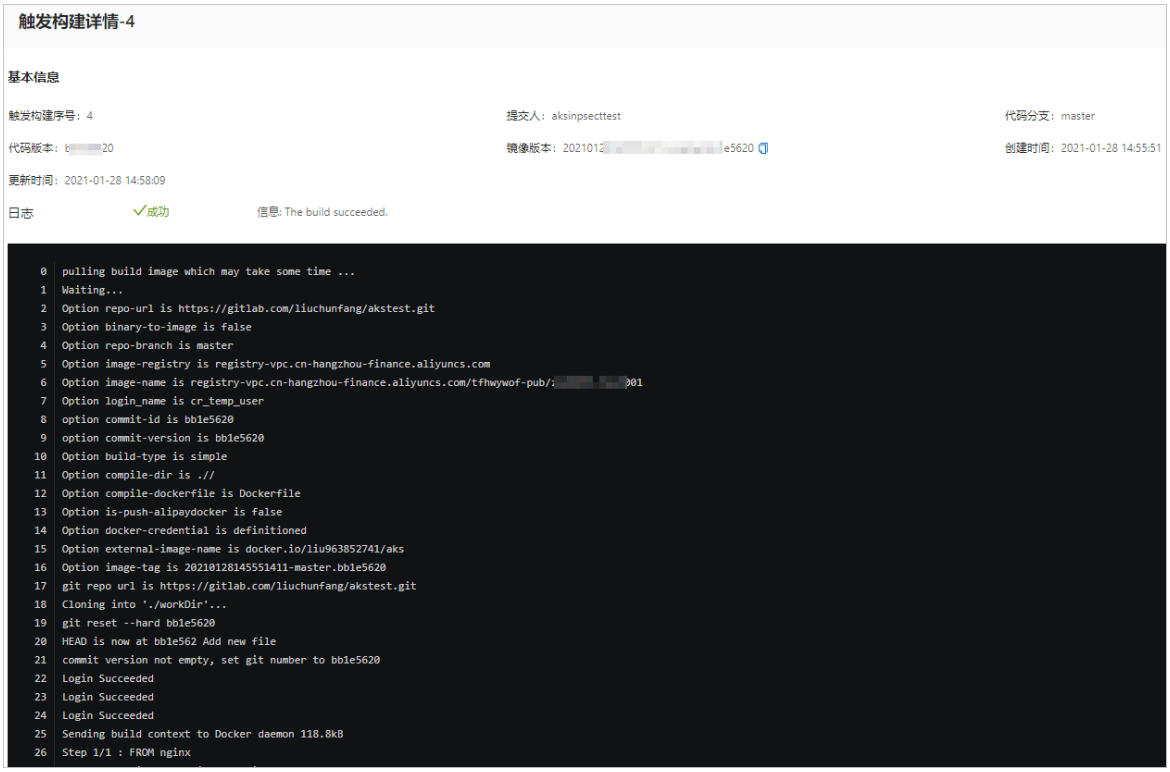


查看触发日志

您可以在控制台查看完整的构建触发日志。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **镜像中心 > 镜像构建**，进入构建列表页。
2. 单击 **构建集名称** 进入构建详情页。
3. 在 **构建历史** 页签中，单击镜像构建的编号，查看详细的触发日志信息。



8.4.2. 镜像仓库

镜像仓库是镜像的集合，用于存储、管理容器镜像。AKS 集成了阿里云容器镜像服务（ACR），您可以在 ACR 控制台上轻松存储、管理、部署容器镜像。本文主要介绍如何上传私有镜像。

> 说明
> 敏捷版环境本不支持阿里云容器镜像服务（ACR），需使用自建的镜像中心。

上传私有镜像

为了给您提供更好的镜像仓库服务，便于您管理私有镜像，AKS 集成了 阿里云容器镜像服务（ACR），您可以开通免费的 ACR 默认实例，用于上传私有镜像，AKS 集群可以正常访问镜像实例的内网地址。

> 说明
> 安全起见，金区暂不支持外网访问，上传镜像时必须通过阿里云的 ECS 上传至镜像仓库，非金区可以通过下面的步骤上传。

前提条件

- 已完成本地 构建 镜像。
- 准备一台安装了 docker 的机器。安装的 Docker 版本必须为 1.6.0 及以上。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **镜像中心** > **镜像仓库**。
2. 单击 **管理镜像中心** 链接进入 ACR 控制台，创建 镜像仓库。

> 说明
> 镜像仓库建议选择与开通 AKS 产品的相同地域。

3. 登录安装了 docker 的机器，执行以下命令将镜像推送到镜像仓库。

```
$ sudo docker login --username=[TestUser] registry.cn-shanghai.aliyuncs.com
$ sudo docker tag [ImageId] registry.cn-shanghai.aliyuncs.com/[命名空间名称]/[仓库名称]:[镜像版本号]
$ sudo docker push registry.cn-shanghai.aliyuncs.com/[命名空间名称]/[仓库名称]:[镜像版本号]
```

> 说明
>
> - 请根据实际镜像信息替换示例中的 `[TestUser]`、`[ImageId]`、`[命名空间名称]`、`[仓库名称]` 和 `[镜像版本号]` 参数。
> - 用于登录的用户名为阿里云账号全名，密码为开通服务时设置的密码。您可以在访问凭证页面修改凭证密码。
> - 上传成功后，您可以在 **默认实例 > 镜像仓库**，单击镜像仓库名称，选择 **镜像版本** 中查看已上传的镜像。您可以在 创建应用服务 时引用镜像的内网地址，AKS 集群可以正常访问镜像实例的内网地址。

更多参考

- 仓库基本使用
- 仓库访问控制
- 官方镜像加速

8.4.3. 向金区镜像仓库上传镜像

为安全起见，无法通过外网访问金区镜像仓库地址。因此，如果需要将镜像上传至金区镜像仓库，可通过同一 VPC 内的 ECS 服务器上传。本文介绍如何通过 ECS 服务器将镜像上传至金区镜像仓库。

前提条件

- 已完成 镜像构建。
- 已创建金区镜像仓库。
- 已安装 Docker。

> 说明
> 由于后续需要通过执行 Docker 命令将镜像打包成压缩包，所以需要先在本地安装 Docker。具体安装步骤可参考 Docker 官网。

操作步骤

1. 将本地镜像打包成压缩包。
   i. 在本地控制台执行如下命令，查看本地镜像信息：

```
$ sudo docker images
```

返回镜像信息示例如下：

```
$ sudo docker images
REPOSITORY                          TAG             IMAGE ID      CREATED     VIRTUAL SIZE
registry.aliyuncs.com/acs/agent     0.7-dfb6816     37bb9c******     7 days ago     37.89 MB
```

ii. 将本地镜像打包成压缩包。

```
$ sudo docker save -o [压缩包名] [REPOSITORY]:[TAG]
```

变量取值示例如下：

变量名	示例值
[压缩包名]	sofaboot_demo.tar.gz
[REPOSITORY]	registry.aliyuncs.com/acs/agent
[TAG]	1.0.0

2. 将压缩包上传至 ECS 服务器。
在本地控制台执行如下命令，将压缩包上传至 ECS 服务器：

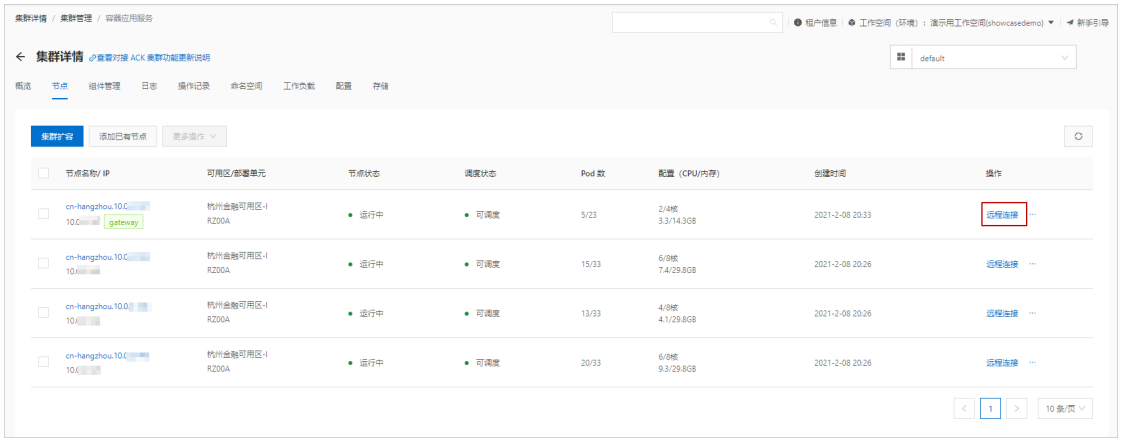
```
$ sudo scp [压缩包名] root@[ ECS 公网 IP ]: [ ECS 服务器上要存放压缩包的文件路径]
```

变量取值示例如下：

变量名	示例值
[ECS 服务器上要存放压缩包的文件路径]	/home/images/
[压缩包名]	sofaboot_demo.tar.gz
[ECS 公网 IP]	121.41.***.***

3. 将镜像上传至镜像仓库。
   i. 登录容器应用服务控制台，在左侧导航栏单击 **集群管理 > 集群详情**。
   ii. 在 **集群详情** 页，单击 **节点** 页签。
   iii. 选择目标节点，单击其右侧操作列中的 **远程连接**。

> **注意**
> ECS 服务器需开通 EIP，即公网 IP。具体步骤可参考 开通新 EIP。



iv. 输入 VNC 密码。

> 说明
> 如果忘记 VNC 密码，单击 **修改 VNC 密码** 进行重置。

v. 输入 ECS 服务器密码。

vi. 将镜像上传至镜像仓库，执行如下命令：

```
$ sudo cd [ ECS 服务器上要存放压缩包的文件路径]
$ sudo docker load -i [压缩包名]
$ sudo docker login --username=[登录账号] [Registry]
$ sudo docker tag [ImageId] [仓库地址]:[镜像版本号]
$ sudo docker push [仓库地址]:[镜像版本号]
```

变量取值示例如下：

变量名	示例值
[ECS 服务器上要存放压缩包的文件路径]	/home/images/
[压缩包名]	sofaboot_demo.tar.gz
[登录账号]	aksinpsecttest@1639881487******
[Registry]	registry-vpc.cn-hangzhou-finance.aliyuncs.com
[ImageId]	798f63******
[仓库地址]	registry-vpc.cn-hangzhou-finance.aliyuncs.com/sofaboot-space/sofaboot_store
[镜像版本号]	1.0.0

8.4.4. 镜像技术栈

镜像技术栈为用户提供基础镜像管理的能力，涵盖官方提供的镜像技术栈以及用户自定义的镜像技术栈。后续在创建应用服务时，您可以上传应用代码包，并选择所需的基础镜像，一键构建出目标镜像 Binary to Image（B2I），无需编写 dockerfile，真正实现无感镜像化。

> 说明
> 目前仅支持浏览和搜索官方镜像技术栈，您可以对自定义的镜像技术栈进行编辑和下架。

创建镜像技术栈

创建镜像技术栈将会定义镜像技术栈的基本元数据。

> 说明
> 目前对可以创建的镜像技术栈个数不做限制。

1. 登录容器应用服务控制台，在左侧导航栏单击 **镜像中心 > 镜像技术栈**，进入镜像技术栈列表页面。
2. 单击 **创建镜像技术栈**，在弹出的窗口中填写以下配置后，单击 **确定**。

- **技术栈名称**：允许包含字母、数字、连字符，且必须以字母开头、以字母或数字结尾。最多允许输入 32 个字符。创建后技术栈的名称不可修改。
- **技术栈描述**：填写该技术栈的描述。
- **技术栈图标**：图标最佳尺寸 80*80 px。

创建后系统会提示创建成功，并以卡片的形式在列表页展示技术栈信息。您可以在技术栈详情页继续编辑技术栈的描述与图标。

创建镜像版本

创建镜像技术栈后，您可以根据需要创建不同的镜像版本。

> 说明
> 目前对可以创建的镜像版本个数不做限制。

1. 单击镜像技术栈的卡片进入技术栈详情页。
2. 单击创建镜像版本，配置以下信息后，单击 **创建**。
   - **镜像版本名称**：允许包含字母、数字、连字符、且必须以字母开头、以字母或数字结尾。最多允许输入 32 个字符。创建后版本的名称不可修改。
   - **基础镜像信息**：选择要使用的镜像仓库以及基础镜像版本。
   - **代码包上传路径地址**：代码包在构建镜像中的目录路径。需以 `/` 开头。系统会将发布包拷贝到基础镜像中用户指定的目录。
   - **备注**：建议填写基础镜像的 Git 地址。

创建后系统会提示创建成功，并以列表形式在技术栈详情页展示。在该页面，您可以对自定义的技术栈版本继续进行浏览、下架的操作。

> 说明
> - 镜像版本号由系统自动生成，目前不支持修改。
> - 镜像版本下架后不可恢复。下架的版本无法在创建应用服务时选择使用。

8.4.5. 示例镜像

容器发布时需要镜像才能进行发布，本文列出各个地域的示例镜像。

容器发布示例镜像

华东1（杭州）

类型	镜像地址
SOFA Service	registry-vpc.cn-hangzhou.aliyuncs.com/sofa-samples/sofahellomeshserver:1.0.1
SOFA Client	registry-vpc.cn-hangzhou.aliyuncs.com/sofa-samples/sofahellomeshclient:1.0.1
Dubbo Service	registry-vpc.cn-hangzhou.aliyuncs.com/sofa-samples/dubboechoprovider:1.0.1
Dubbo Client	registry-vpc.cn-hangzhou.aliyuncs.com/sofa-samples/dubboechoconsumer:1.0.1

类型	镜像地址
Spring Cloud Service	registry-vpc.cn-hangzhou.aliyuncs.com/sofa-samples/springcloudreservationservice:1.1.4
Spring Cloud Client	registry-vpc.cn-hangzhou.aliyuncs.com/sofa-samples/springcloudreservationclient:1.1.4

华东2（上海）

类型	镜像地址
SOFA Service	registry-vpc.cn-shanghai.aliyuncs.com/sofa-samples/sofahellomeshserver:1.0.1
SOFA Client	registry-vpc.cn-shanghai.aliyuncs.com/sofa-samples/sofahellomeshclient:1.0.1
Dubbo Service	registry-vpc.cn-shanghai.aliyuncs.com/sofa-samples/dubboechoprovider:1.0.1
Dubbo Client	registry-vpc.cn-shanghai.aliyuncs.com/sofa-samples/dubboechoconsumer:1.0.1
Spring Cloud Service	registry-vpc.cn-shanghai.aliyuncs.com/sofa-samples/springcloudreservationservice:1.1.4
Spring Cloud Client	registry-vpc.cn-shanghai.aliyuncs.com/sofa-samples/springcloudreservationclient:1.1.4

8.5. 应用服务

8.5.1. 创建应用服务

AKS 应用服务对 Kubernetes 原生的 Deployment 做了能力增强，通过创建应用服务您可以定义容器服务的基本信息、访问策略、发布及调度策略等信息，为后续容器服务的部署做准备。
AKS 产品支持通过以下两种方式创建应用服务：

标准流程 进入应用发布 → 创建应用服务 → 配置应用服务

快捷流程 进入应用商店 → 访问产品模板 → 进入模板详情页 → 产品安装

本文介绍创建应用服务的标准操作流程。通过应用商店的产品模板快捷地创建应用服务，请参见 产品发布。

前提条件

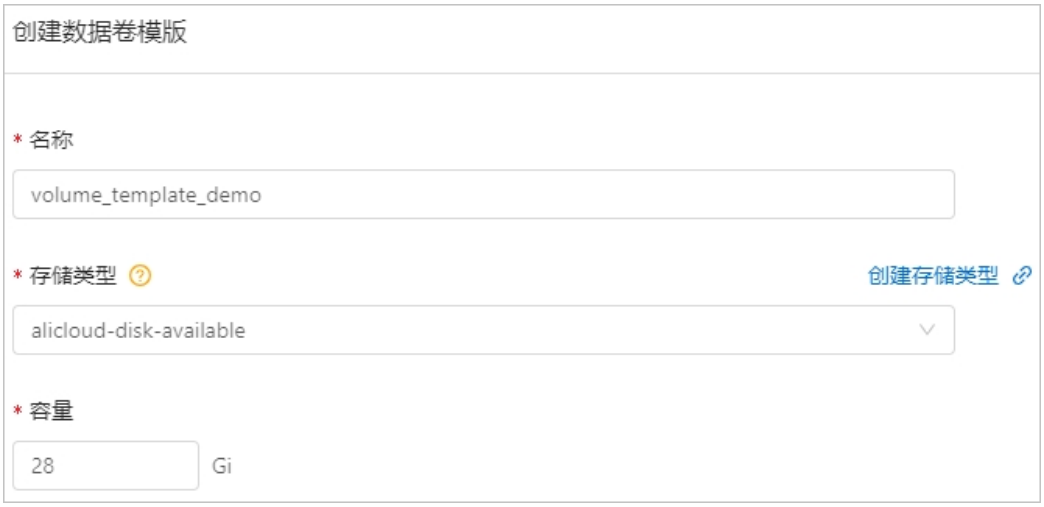
- 已创建应用
- 已创建集群

创建应用服务的标准流程分为以下 6 个步骤：

1. 填写基本信息
2. Pod 模板配置
3. 伸缩配置
4. 访问配置（选填）
5. 部署和调度配置（选填）
6. 预览并提交

一、填写基本信息

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用发布 > 应用服务**。
2. 在应用服务列表页，单击 **创建**。
3. 在 **创建应用服务** 页面，填写以下基础信息，单击 **下一步**。
   - **命名空间**：选择一个集群中的命名空间（namespace），默认为 **default**。
   - **应用服务名称**：容器服务的名称。
   - **所属应用**：选择一个该容器服务所关联的应用。
   - **有状态应用 app**：默认关闭。有状态应用使用 StatefulSet 作为部署工作负载。
     若开启，需配置数据卷模板，具体配置步骤如下：
     a. 单击 **创建新模板**。
     b. 在 **创建数据卷模版** 页面，填写以下信息后，单击 **创建**。



     - **名称**：输入模板名称。允许包含字母、数字、连字符，且必须以字母开头，以字母或数字结尾。
     - **存储类型**：用于指定支持的存储类别。支持 alicloud-disk-available、alicloud-disk-efficiency、alicloud-disk-essd、alicloud-disk-ssd、alicloud-disk-topplogy。
     - **容量**：输入大于 20 的整数。单位 Gi。
   - **描述**：选填。容器服务的描述。

二、Pod 模板配置

对 Pod 里的容器做详细的配置。完成基本选项配置和（可选）高级配置后，单击 **下一步**。

基本选项

- **容器名称**：容器名称。允许包含小写字母、数字、连字符，且必须以小写字母开头，以小写字母或数字结尾。
- **镜像来源**：支持 **镜像仓库**、**构建记录** 及 **从包构建** 三种方式。

- **镜像仓库**：直接输入镜像仓库地址，比如：`registry-cnhz.cloud.alipay.com/aks/nginx:1.8`。
- **构建记录**：选择该类型来源时，会自动关联应用服务所属应用的构建，具体参见 镜像构建。

  > 说明
  > 您可以选择 **使用保密字典拉取镜像**，参见 创建保密字典。

- **从包构建**：支持上传指定技术栈应用（SOFABoot、Java 等）的代码包。系统会根据相关配置信息自动地进行镜像构建，并在后续的容器服务发布中使用系统构建的目标镜像进行发布。
  - **技术栈**：选择镜像技术栈，即构建基于的基础镜像。包括用户自定义的技术栈和官方提供的技术栈。详情参见 镜像技术栈。
  - **技术栈版本**：选择镜像技术栈的版本。
  - **发布包**：上传代码包，无格式限制。
  - **目标镜像中的位置**：代码包在构建镜像中的目录路径。需以 `/` 开头。系统会将发布包拷贝到基础镜像中用户指定的目录。
  - **基础镜像**：根据所选择的技术栈及版本自动生成。
  - **Dockerfile 预览**：构建镜像的 Dockerfile 内容，派生自基础镜像，并只作发布包覆盖。
  - **目标镜像仓库地址（自定义）**：构建出的镜像名称，此处填写目标镜像标签。系统会将构建出来的目标镜像按照此处定义的地址推送到镜像中心，供后续发布所用。

    > 说明
    > - 默认情况下，使用保密字典拉取镜像。若尚未创建保密字典，可单击 **创建保密字典** 链接进行创建。
    > - 默认情况下，会将构建出来的镜像推送到镜像仓库中，因此必须已在容器镜像服务（ACR）中 创建命名空间。

- **CPU 配置**：设置容器使用的 CPU 的数量。**请求核数** 为能保证的最小核数，**最大核数** 为能使用的最大核数。换算方式：1 core = 1000 millicores。
- **内存配置**：设置容器使用的内存的数量。**请求内存** 为能保证的最小内存数量，**最大内存** 为能使用的最大内存数量。换算方式：1024 bytes = 1 KiB；1024 KiB = 1 MiB；1024 MiB = 1 GiB；1024 GiB = 1 TiB。
- **启动命令**：选填。用于指定容器启动时执行的命令。

高级配置

- **环境变量配置**：设置在容器启动时传入应用进程的键值对，例如：`USER=tester`。
- **数据卷配置**：配置容器使用的数据卷，目前仅支持挂在当前容器所在宿主机的目录。
- **健康检查配置**：包含两种检查机制：Readiness 和 Liveness。具体参见 Kubernetes Probe。
- **生命周期事件回调配置**：为容器添加生命周期事件回调，分别在容器启动后和容器停止前执行。
- **日志服务配置**：配置阿里云日志服务（SLS），可选择已有日志库或创建新的日志库。
- **业务实时监控**：添加监控目录，可以在应用服务详情页查看容器的监控数据。

更多信息参见 高级配置说明。

三、伸缩配置

副本伸缩策略配置：目前仅支持 **固定副本数**，默认为 1，即应用服务运行时保持固定数目的 Pod 副本。配置完成后，单击 **下一步**。



四、访问配置（选填）

应用服务支持三种访问方式：集群内访问、内网访问、公网访问，您可以根据业务需要做好规划。

集群内访问

选择 **集群内访问** 会建一个 Cluster IP 类型的 Service，并将流量转发到容器的相应端口上。您可以在创建应用服务时设置访问方式，也可以应用服务创建完成后添加访问方式。

创建应用服务时添加访问方式

1. 在 **访问配置** 页面，单击 **创建服务**。
2. 在 **创建服务** 窗口，填写以下信息后，单击 **提交**。
   - **服务名称**：应用服务名称。
   - **访问方式**：选择 **集群内访问**。
   - **端口映射**：单击 **+**，填写以下信息。



   - **协议**：支持 TCP、UDP 协议。
   - **访问端口**：容器镜像中工作负载程序实际监听的端口，端口范围为 1~65535。
   - **容器端口**：容器端口映射到集群虚拟 IP 上的端口，用虚拟 IP 访问工作负载时使用，端口范围为 1~65535。

内网访问

创建一个内部的 Service，并将流量转发到容器的相应端口上。您可以在创建应用服务时设置访问方式，也可以应用服务创建完成后添加访问方式。

创建应用服务时设置访问方式

1. 在 **访问配置** 页面，单击 **创建服务**。
2. 在 **创建服务** 窗口，填写以下信息后，单击 **提交**。
   - **服务名称**：应用服务名称。

- **访问方式**：选择 **内网访问**。
- **端口映射**：单击 **添加端口映射**，填写以下信息。



  - **协议**：支持 **TCP**、**HTTP**、**HTTPS** 协议。

    > 说明
    > TCP 协议下，默认开启健康检查，不可关闭。

  - **转发规则**：支持 **按权重**、**按权重并且最小连接数**。
  - **前端端口**：容器镜像中工作负载实际监听的端口，端口范围为 1~65535。
  - **后端端口**：容器端口映射到负载均衡实例上的端口，用负载均衡 IP 访问工作负载时使用，端口范围为 1~65535。
  - **健康检查**：若开启，根据选择的协议类型，需填写对应的配置项。

公网访问

配置外部 Service 会创建一个公网的 LoadBalancer，并将流量转发到容器的相应端口上。访问方式由公网负载均衡服务地址以及设置的访问端口组成，例如 `10.117.117.117:80` 。

创建应用服务时设置访问方式

1. 在 **访问配置** 页面，单击 **创建服务**。
2. 在 **创建服务** 窗口，填写以下信息后，单击 **提交**。
   - **服务名称**：应用服务名称。
   - **访问方式**：选择 **公网访问**。



- **端口映射**：单击 **添加端口映射**，填写以下信息。



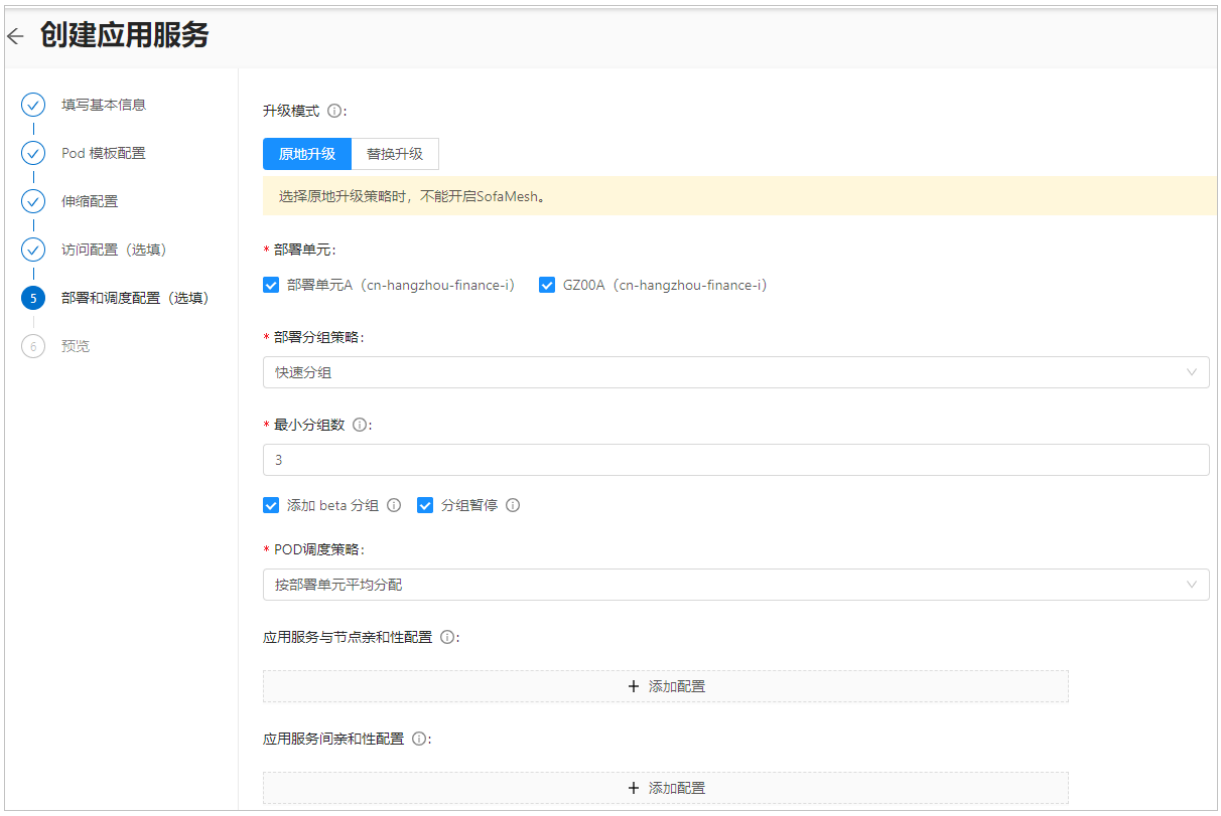
  - **协议**：支持 TCP、HTTP、HTTPS 协议。本例中选择 TCP。

    > 说明
    > TCP 协议下，默认开启健康检查，不可关闭。

  - **转发规则**：支持 **按权重**、**按权重并且最小连接数**。
  - **前端端口**：容器镜像中工作负载实际监听的端口，端口范围为 1~65535。
  - **后端端口**：容器端口映射到负载均衡实例上的端口，用负载均衡 IP 访问工作负载时使用，端口范围为 1~65535。
  - **健康检查**：若开启，根据选择的协议类型，需填写对应的配置项。

五、部署和调度配置（选填）

您可以自定义部署和调度配置，若不修改，发布应用服务时，系统会使用默认配置。
该配置项用于配置容器服务在部署时需要用到的信息，包括以下内容：



- **升级模式**：
  - **原地升级**：默认选项。保持 Pod 不变，IP 不会更改。若选择 **原地升级**，则不能开启 **服务网格**。

    > **说明**
    > 原地升级在敏捷 PaaS 底座的专有云场景和 ACK 集群的公有云场景下会做能力降级：仅升级镜像时，Pod IP 能保持不变，其他变更如环境参数等，会通过滚动升级的方式用新 Pod 替换掉旧 Pod。

  - **替换升级**：会用新的 Pod 替换老的 Pod，类似原生 Deployment 的 RollingUpdate 策略。

    > **说明**
    > 若服务部署后切换升级模式，会造成服务短暂不可用。

- **部署单元**：该配置限定了容器服务的 Pod 只能调度到打了所选中部署单元标签的节点上。默认选中所有部署单元。
- **部署分组策略**：指定发布容器服务时 Pod 的分组策略，有以下几种策略，默认为快速分组。
  - **快速分组**：分组进行发布，并且尽量让每一组的 Pod 都平均分布在每个部署单元。
  - **每个 Pod 一组**：每组一个 Pod，有几个 Pod 就分几组。
  - **共分一组**：所有 Pod 在一组中进行发布。
- **最小分组数**：该选项仅在快速分组时设置。默认为 3。仅当副本数量大于等于该值时生效。每组副本数最多 20 个。

> 说明
> 最小分组数是发布时分组的参考值，实际分组数受发布时 Pod 在部署单元的分布情况影响。

- **添加 Beta 分组**：发布时选取部分 Pod 先行发布，待确认无异常后继续发布。默认选中该选项。
  选择 **添加 Beta 分组** 后，发布时会给应用服务设置一个特殊的 Beta 分组，在该组中，系统会在每个部署单元选择一个 Pod，Beta 分组会在第一组发布。
  - Beta 分组发布完成后系统会自动暂停应用服务发布，等待系统负责人或者运维工程师对应用服务的发布情况进行确认。若容器服务发布正常，则单击 **Beta 分组确认**，使应用服务继续分组发布。
  - Beta 分组可以与所有分组策略共同决定分组。创建新的发布申请时，默认选中 **添加 Beta 分组**。
- **分组暂停**：每组发完后暂停发布，确认发布后继续发布。默认选中该选项。
- **Pod 调度策略**：目前支持 **按部署单元平均分配**。

> 说明
> 该选项仅在容器服务有多个部署单元时有效。当某个部署单元资源不足以调度平均的 pod 数目时，该部署单元上的 pod 会被动态调整到其他单元。

- **应用服务与节点亲和性配置**：添加应用服务在节点级别的亲和性配置。可以通过 node 标签来限定应用服务可以调度的节点范围。

> 说明
> 亲和性需要 Pod 重新创建才会生效。首次发布时，亲和性配置一定会生效，后续发布时，必须选择替换升级方式，才能让亲和性配置生效。

- **应用服务间亲和性配置**：添加应用服务在 Pod 级别的亲和性配置。通过选择与某些应用服务在相同或不同的节点来限定应用服务可以调度的节点范围。

> 说明
> 亲和性需要 Pod 重新创建才会生效。首次发布时，亲和性配置一定会生效，后续发布时，必须选择替换升级方式，才能让亲和性配置生效。

- **注入SOFAMesh**：需集群开启服务网格功能。开启引流配置时，每次发布会按所设置的引流规则转发流量到新版本，也可在发布过程中配置。开启后需填写引流比例，系统按所设置的比例转发流量到新发布的版本，以及引流的目标服务。

六、预览并提交

在应用服务 **预览** 页面，确认信息无误，单击 **提交**。

> 注意
> - 提交前，您可以单击修改图标对应用服务信息进行修改。
> - 应用服务提交后处于 **待部署** 状态，需要单击 **发布** 才会将应用服务部署到集群中。

8.5.2. 高级配置说明

在 创建应用服务 的过程中，您可以在配置 Pod 模板时进行以下高级配置。

环境变量

通过键值对的形式为 Pod 配置环境变量，从而给 Pod 添加环境标志或传递配置等，在容器启动时传入应用进程，例如：`USER=tester`。具体参见 Pod variable。

参数说明如下：



- **来源**：
  - **手动输入**：自行指定环境变量的名称和值。
  - **配置项**：选择已配置的 ConfigMap 来指定环境变量的值。
  - **保密字典**：选择已配置的 Secret 来指定环境变量的值。
- **变量名**：环境变量的名称。
- **变量值/变量引用**：
  - 选择 **手动输入** 时，为环境变量的值。
  - 选择 **配置项** 或 **保密字典** 时，为 ConfigMap 和 Secret 的引用。

数据卷配置

配置容器使用的数据卷，目前仅支持挂载在当前容器所在宿主机的目录。
参数说明如下：



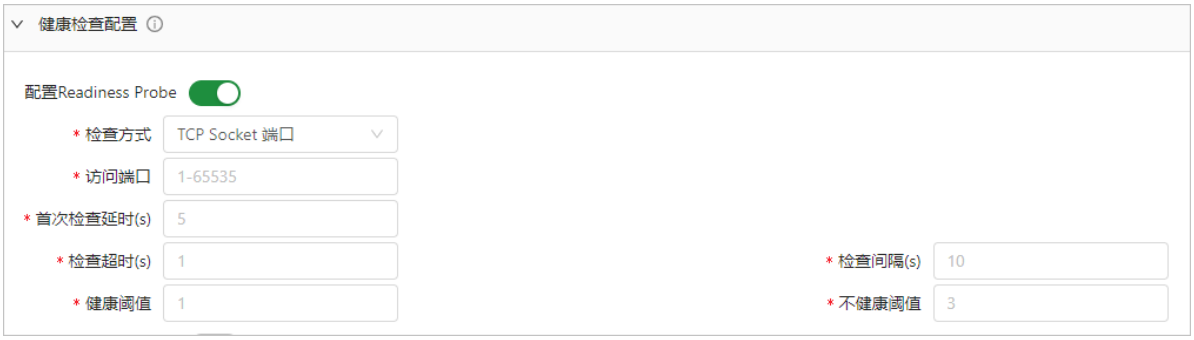
- **类型**：支持 **本地存储**、**配置项**、**保密字典**。更多信息参见 volumes。
- **权限**：支持 **读写** 或 **只读**。
- **容器目录**：容器中的数据卷路径。
- **主机目录/配置项引用**：
  - 选择 **本地存储** 时，为 Pod 所在宿主机的目录。
  - 选择 **配置项** 或 **保密字典** 时，为对应的引用。

健康检查配置

健康检查是指容器运行过程中，基于用户配置的健康检查，定时检查容器健康状况。若不配置健康检查，如果出现业务异常，容器将无法感知，也不会自动重启去恢复业务，进而可能影响业务的正常运作。
目前 AKS 提供两种检查机制：Readiness Probe 和 Liveness Probe。
Readiness Probe 用来判断容器是否准备好接收流量。如果检查失败，集群会屏蔽请求访问该容器；若检查成功，则会开放对该容器的访问。
Liveness Probe 用来检测容器是否正常，如果检查失败，集群会对该容器执行重启操作。

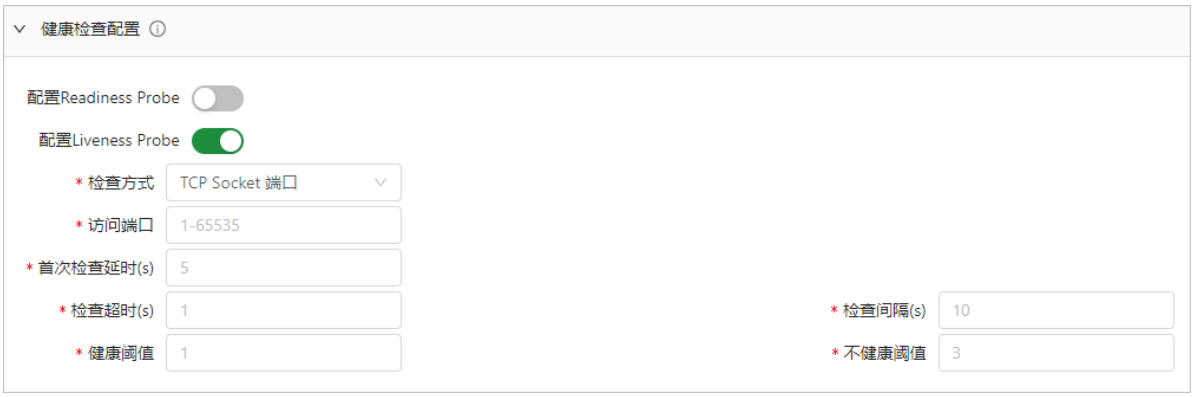
> 说明
> Readiness Probe、Liveness Probe 是 kubernetes 原生概念。更多详情请参见 Kubernetes Probes。

- **配置 Readiness Probe**
  参数说明如下：



- **检查方式**：
  - **TCP Socket 端口**：通过发送 TCP 请求来进行判断，若能成功建立连接，则检查通过。
    - **访问端口**：TCP 访问的容器端口。
  - **HTTP Get 方法**：通过发送 HTTP Get 方法的请求来进行判断，若返回的状态码大于等于 200 且小于 400，则检查通过。
    - **访问路径**：HTTP 请求的访问路径。
    - **访问端口**：HTTP 请求的访问容器端口。范围 1-65535。
  - **在容器中执行命令**：通过在容器中执行命令来进行判断，若命令成功执行，并返回 0，则检查通过。
    - **执行命令**：指定在容器中执行的命令。
- **首次检查延时**：即 initialDelaySeconds，容器启动后第一次执行探测时需要等待多少秒，默认为 5 秒。
- **检查超时**：即 timeoutSeconds，探测超时时间。默认 1 秒，最小 1 秒。
- **检查间隔**：即 periodSeconds，指执行探测的时间间隔，默认为 10 秒，最小为 1 秒。
- **健康阈值**：探测失败后，最少连续探测成功多少次才被认定为成功。默认是 1，最小值是 1。
- **不健康阈值**：探测成功后，最少连续探测失败多少次才被认定为失败。默认是 3，最小值是 1。

- **配置 Liveness Probe**
  参数说明如下：



- **检查方式**：
  - **TCP Socket 端口**：通过发送 TCP 请求来进行判断，若能成功建立连接，则检查通过。
    - **访问端口**：TCP 访问的容器端口。
  - **HTTP Get 方法**：通过发送 HTTP Get 方法的请求来进行判断，若返回的状态码大于等于 200 且小于 400，则检查通过。
    - 访问路径：HTTP 请求的访问路径。
    - 访问端口：HTTP 请求的访问容器端口。范围 1-65535。
  - **在容器中执行命令**：通过在容器中执行命令来进行判断，若命令成功执行，并返回 0，则检查通过。
    - 执行命令：指定在容器中执行的命令。
- **首次检查延时**：即 initialDelaySeconds，容器启动后第一次执行探测时需要等待多少秒，默认为 5 秒。
- **检查超时**：即 timeoutSeconds，探测超时时间。默认 1 秒，最小 1 秒。
- **检查间隔**：即 periodSeconds，指执行探测的时间间隔，默认为 10 秒，最小为 1 秒。
- **健康阈值**：探测失败后，最少连续探测成功多少次才被认定为成功。默认是 1，最小值是 1。
- **不健康阈值**：探测成功后，最少连续探测失败多少次才被认定为失败。默认是 3，最小值是 1。

生命周期事件回调配置

为容器添加生命周期事件回调，分别在容器启动后和容器停止前执行。
目前提供的生命周期回调函数为：Post Start （容器启动后触发）和 Pre Stop（容器停止前触发）。
参数说明如下：



- **配置 Post Start**：
  回调方式：支持 **在容器中执行命令** （需输入执行命令，如 `cat tmp/healthy` ）及 **HTTP GET 方法**（需输入执行路径，如 `/healthy/readiness` ，以及输入访问端口，范围 1-65535）。
- **配置 Pre Stop**：
  回调方式：支持 **在容器中执行命令** （需输入执行命令，如 `cat tmp/healthy` ）及 **HTTP GET 方法**（需输入执行路径，如 `/healthy/liveness` ，以及输入访问端口，范围 1-65535）。

日志服务配置

配置阿里云日志服务（SLS），可选择已有日志库或创建新的日志库。
参数说明如下：



- **日志库**： 从已有的日志库中选择。如果没有可用日志库，可单击 **创建日志库** 来创建新的日志库。
- **日志类型**： 容器文本日志。
- **路径**：收集容器内指定路径的日志。例如： `/home/admin/*.log` 或 `/root/*.log` 。

业务实时监控

添加监控目录，可以在应用服务详情页查看容器的监控数据。目录内容将被监控应用读取。
参数说明如下：
监控目录：指定应用需采集的业务监控日志的输出目录。

8.5.3. 更新应用服务

创建并提交应用服务后会生成对应的版本。每个版本都是无法进行修改的，若更新配置，则会生成新的版本。新的版本需要通过发布才会在线上生效。对于日常更新版本，只需选择镜像而不用其他配置，可在应用服务详情页选择快捷操作。

标准更新

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用发布 > 应用服务**。
2. 在应用服务列表页，单击应用服务的名称进入应用服务详情页。
3. 选择以下方式修改应用服务信息：
   - 单击右上角的 **编辑配置信息**，进入 **编辑版本** 页面。

     > 说明
     > 若当前有未提交的草稿，编辑配置信息会直接编辑草稿。

   - 选择 **版本** 页签，单击 **拷贝到草稿**，单击版本名称，进入 **编辑草稿** 页面。
4. 修改以下配置信息。
   参数说明参考 创建应用服务。
   - **基本信息** 中的 **描述**。
   - **Pod 模板配置** 中除 **容器名称** 以外的所有信息。
   - **访问配置** 中所有信息。
   - **部署调度** 中所有信息。
5. 修改完成后，预览并提交。
   配置信息编辑成功后，您可以在应用服务详情页的版本页签中查看更新的应用服务版本信息。
6. 在应用服务详情页的版本页签，选择新版本，并单击 **发布**。

快捷更新

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用发布 > 应用服务**。
2. 在应用服务列表页，单击应用服务的名称进入应用服务详情页。
3. 单击 **快捷操作**，选择 **更换镜像** 或 **修改副本数**，选择目标版本镜像或修改副本数，单击 **提交**。
4. 提交后会自动生成发布单，单击 **整体发布** 进行发布。

8.5.4. 删除应用服务

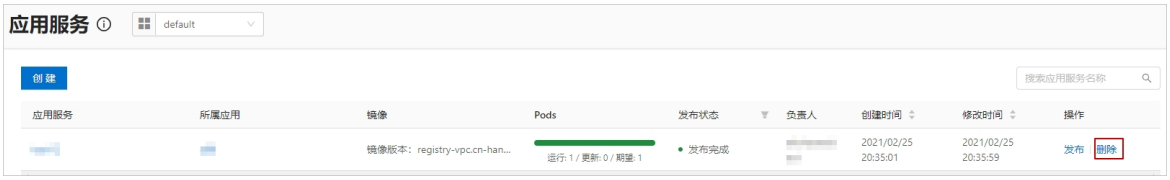
在应用服务列表页，您可以删除不需要的应用服务。删除应用服务后，服务下的资源将全部被释放。本文主要介绍如何删除应用服务。

前提条件

已创建应用服务

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用发布 > 应用服务**。
2. 在应用服务列表页，单击应用服务右侧 **操作** 栏中的 **删除** 选项。



3. 在弹出的对话框中，单击 **确定** 即可删除该应用服务。

> 注意
> 正在发布中的应用服务是无法删除的。

相关操作

您可参考以下文档，了解应用服务更多功能及操作：

- 创建应用服务
- 更新应用服务
- 应用服务详情

8.5.5. 立即发布服务

应用服务创建完成后，您可以立即发布应用服务。

前提条件

发布应用服务前，需确保集群内有充足的节点资源，例如： CPU、Memory，否则 Pod 将无法正常调度到节点。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用发布 > 应用服务**。
2. 在应用服务列表页，单击应用服务名称进入 **应用服务详情** 页，单击右上角的 **发布**。

> 说明
> - 若应用服务存在多个版本，会提示 是否发布该容器服务最近一次变更的版本。
> - 若需发布以前的版本，选择 **版本** 页签，单击版本右侧的 **发布**。

3. 在 **发布预览** 页面，确认信息无误后单击 **发布**。
4. 发布单创建成功后，在 **发布单详情** 页面，单击 **整体发布**。

> 说明
> - 您也可以单击右侧的发布任务，然后单击左侧的 **应用发布**，发布单个应用服务。
> - 待应用服务状态更新为 **发布成功**，即表示该应用服务已经发布成功。

查看发布详情

单击发布后自动跳转到 **发布单详情** 页面。



您可以单击 Pod 名称查看详细的 Pod 事件。

重试回滚

若出现发布失败，您可以单击 **重试**，继续进行发布，或单击 **回滚** 恢复到上一个已发布的版本。

> 说明
> 仅当前应用有其它成功发布的发布包（除当前版本外）时，您才可以回滚应用。若无任何成功发布记录，回滚将释放所有资源。

8.5.6. 发布单

创建发布单

通过发布单可以同时发布多个应用服务。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用发布 > 发布单**。
2. 在发布部署大盘，单击 **创建发布单**。

> 说明
> 目前发布部署大盘页面的搜索框仅支持搜索发布单名称的前缀。

3. 在 **创建发布单** 页面，填写以下发布信息后，单击 **下一步**。
   - **基本信息**
     - **标题**：发布标题。
     - **部署单元**：展示当前工作空间下的所有部署单元。

- **应用服务发布列表**：您可以在 **待选应用服务列表** 中单击选择需要的应用服务，单击 **>** 图标将应用服务添加到 **已选应用服务列表** 中。

  > **说明**
  > 若应用服务存在多个提交版本，需要选择要发布的版本。默认选择最新的版本。

- **高级配置**
  - **设置应用服务依赖关系**：当选择了 2 个及以上应用服务时，可能需要填写应用服务之间的依赖关系。若应用服务之间不存在依赖关系，则直接跳过。
    您可通过逐条添加的方式添加依赖关系。在 **依赖于** 列中选择某个应用服务所依赖的服务名称（可以同时选择多个）。例如，若 appservice1 依赖 appservice2，则在 appservice1 对应的 **依赖于** 列中选择 appservice2 即可，依赖于多个应用服务时同理。

    > **注意**
    > 应用服务之间不可循环依赖，即不可出现应用服务 A 依赖于应用服务 B，同时应用服务 B 又依赖于应用服务 A 这种情况。

4. 在 **预览** 页面确认信息无误后，单击 **创建**。系统会自动跳转到 **发布单详情** 页面，单击 **整体发布** 即可开始发布。

查看发布单详情

发布单创建完成后，系统会自动跳转到 **发布单详情** 页面，您可以查看发布单的执行详情。

> **说明**
> 若发布单初始化失败，则无法进行发布，可以将鼠标悬停到 **失败原因** 上，查看具体原因，或单击 **重新初始化**。

前置任务

- **执行服务配置变更**：若容器服务存在添加或修改 Service 服务配置，执行 Service 服务配置任务，若无则直接跳过。
- **发布日志配置**：若容器服务首次配置日志配置，则执行日志配置前置任务：创建或更新，若无则直接跳过。

应用服务

分组发布应用服务，Pod 状态说明如下：

- **未更新**：Pod 配置未更新为最新版本。
- **更新中**：Pod 配置已更新为最新版本，但状态还未 ready 或保持 ready 状态少于 30 秒。
- **更新成功**：Pod 配置已更新为最新版本，并保持 ready 状态 30 秒以上。
- **更新失败**：上线或更新 Pod 异常。

后置任务

- **清理无效服务配置**：若容器服务存在删除 Service 服务配置，执行 Service 清理任务，若无则直接跳过。
- **发布日志配置**：若容器服务未配置日志配置或取消日志配置，则执行日志配置后置任务：删除，若无则直接跳过。

后置任务（回滚）

回退日志配置：若应用服务需回退至无资源状态，且应用服务的日志配置已生效，则执行日志配置的回退任务，若无则直接跳过。

取消发布

发布单创建完成后，系统会自动跳转到 **发布单详情** 页面，您可以在正式发布应用服务前选择 **取消** 发布，终止发布操作。
在 **发布单详情** 页面，单击 **取消**。
若发布单包含多个应用，单击右侧待发布的应用服务，进入单个应用发布页面，单击 **取消**，会取消当前应用发布，其他未取消的应用可以继续发布。

Beta 确认

配置应用服务的部署和调度策略时，若开通了 Beta 确认功能，在应用服务发布时，系统会给服务器分组设置一个特殊的 Beta 分组，即系统从每个部署单元获取一台机器，加入 Beta 分组，并作为第一组发布。
Beta 分组发布完成后系统会自动暂停应用发布，此时系统负责人或运维人员需要对应用的发布情况进行确认。
若应用发布正常，则单击 **Beta 确认**，使应用继续分组发布。

> 说明
> - 若每个部署单元只有一个 Pod，则 Beta 确认无效。
> - Beta 确认功能目前仅在发布应用服务新版本时生效。
> - 若分组策略中没有勾选 **分组暂停**，Beta 组发布后，剩余的分组会自动分批发布直至发布完成。

重试回滚

若发布失败，您可以单击 **重试** 或 **回滚** 继续进行操作。

重试

若单击 **重试**，系统继续进行发布。

> 说明
> 若应用发布失败，无法创建新的发布单进行发布。

回滚

若单击 **回滚**，根据不同的回滚场景，您需要进行二次确认：

- 取消发布：找不到可回滚版本，继续回滚将终止此次发布。

  > 说明
  > 该场景仅在发布成功过一次，重新发布同一个版本失败时出现。

- 回退至无资源状态：无任何成功发布记录，继续回滚将释放所有资源。
- 回滚至上一版本：即将回滚到上一个成功发布版本。

回滚完成后，单击 **确认回滚**。您可以单击右上角的 **原应用发布详情** 查看原发布单发布详情。
若回滚失败，可以选择 **重试回滚** 或 **取消回滚**。

> 注意
> 取消回滚将终止此次发布，可能导致线上存在多个版本。

查看 Pod 事件

发布过程中或者结束后，您可以查看 Pod 事件，了解 Pod 的调度与运行状态。

- 在 **发布单详情** 页面，单击 Pod 名称查看 Pod 事件。
- 在 **应用服务详情** 页的 Pod 列表中，单击 Pod 名称进入 **Pod 详情** 页，选择 **事件** 页签，查看 Pod 事件。

8.5.7. 应用服务详情

创建完应用服务后，您可以查看应用服务详情，包括应用服务的基本信息、配置信息，方便您快速了解应用服务详细信息，例如 Pod 详情、服务信息、日志配置信息等。同时，应用服务详情页还提供了丰富的功能菜单，能够让您更加快捷高效地进行应用服务的更新操作，包括镜像更换、扩缩容等。

查看应用服务详情

创建完应用服务后，您可以查看应用服务详情。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用发布 > 应用服务**。
2. 在应用服务列表页，单击应用服务实例名称，进入 **应用服务详情** 页面。
3. 查看应用服务配置信息。
   应用服务详情页包括以下内容。

模块	说明
Pod	可以查看容器服务所包含的 Pod，并使用 web shell 登录进入 Pod 内的容器进行操作。
服务	服务名称、服务类型、服务地址、创建时间。
日志	日志服务配置信息。如果没有内容，需要为应用服务配置日志采集，具体参考 配置日志。
版本	应用服务的版本号、发布单、发布状态等信息。

Pod 详情

容器控制台

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用发布 > 应用服务**。
2. 在应用服务列表页，单击应用服务名称进入 **应用服务详情** 页。
3. 在 Pod 列表中，单击 Pod 名称进入 **Pod 详情** 页。
4. 在容器列表中，单击容器右侧的 **前往控制台**。
5. 输入命令管理容器。
   命令样例：`exec [-cl] [-a name] [command [arguments ...]] [redirection ...]`

查看容器详情

在容器列表中，单击容器名称进入容器详情页。在容器列表中，单击容器名称进入容器详情页。您可以查看容器的基本信息。单击下方页签可查看事件、标准日志、环境变量、数据卷配置详情。例如，您可以查看容器在某一时间段的标准日志记录，如下图所示。



查看 Pod 事件

参考 查看 Pod 事件。

服务日志

创建日志库

单击 **创建日志库** 链接，在弹出的页面上输入以下信息，并单击 **提交**。

- **日志库名称**：小写字母、数字、连字符（-）和下划线（_）组成，且以小写字母或者数字开头和结尾，长度为 3-63 字节。
- **保存天数**：采集到日志服务中的日志在日志库中的保存时间，单位为天。可以设置为 1~3000 天。超过该时间后，日志会被删除。
- **shard 数目**：日志库的分区数量，每个 Logstore 可以创建 1~10 个分区。每个 Project 中可以创建最多 200 个分区。

> **说明**
> Logstore 创建完成后，您可以分裂或者合并 shard，从而增加或者减少 shard 的数目。

配置日志

您可以在创建应用服务时的 **Pod 模板配置** 中配置日志服务，选择已有日志库或创建日志库，也可以在应用服务创建后配置日志采集。前种方式可以参考 创建应用服务，本节介绍如何在应用服务创建后配置日志采集。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用发布 > 应用服务**。
2. 在应用服务列表页，单击应用服务名称进入 **应用服务详情** 页。
3. 选择 **日志** 页签，单击 **立即配置**，进入 **编辑应用服务** 页。
4. 在 **Pod 模板配置**，展开 **高级配置** 中的 **日志服务配置**。
5. 单击 **添加日志配置**，填写 **日志库** 与 **路径** 后，单击 **保存**。
   参数说明参考 高级配置说明。 日志路径示例：`/home/admin/release/run/logs/<appName>` 。

查询服务日志

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用发布 > 应用服务**。
2. 在应用服务列表页，单击应用服务名称进入 **应用服务详情** 页。
3. 选择 **日志** 页签，进入 **日志查询** 页。
   您可以输入以下字段搜索日志：
   - **日志库名称**
   - **容器名称**
   - **关键字**
   - **检索时间**
   - **Pod 名称**

服务详情

您可以查看应用服务下的服务详情。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用发布 > 应用服务**。
2. 在应用服务列表页，单击应用服务的名称进入 **应用服务详情** 页。
3. 选择 **服务** 页签进入服务列表页。
4. 单击服务名称进入服务详情页。

模块	说明
基本信息	服务的基本信息，例如服务名称、命名空间、创建时间、注释、标签。
端口	端口映射信息，例如协议、前端端口、后端端口、转发规则、节点端口、峰值带宽、会话保持信息。
Pods	Pod 选择器及后端 Pods 配置信息，例如名称、状态、Pod IP 等信息。
Endpoints	可用地址、不可用地址及端口配置信息。
事件	事件信息，例如类型、次数、事件原因、事件内容等。
负载均衡	显示负载均衡的配置信息。

查看服务版本

您可以查看应用服务的版本信息。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用发布 > 应用服务**。
2. 在应用服务列表页，单击应用服务名称进入 **应用服务详情** 页。
3. 选择 **版本** 页签进入版本列表页。
   默认会将当前版本显示在最上面。您可以查看以下版本信息：
   - **版本**：应用服务的所有提交过的版本 ID，单击版本链接会跳转到查看应用服务的详细配置页面。
   - **发布单**：若服务已经发布，会展示关联的发布单 ID，单击发布单链接会跳转到发布单详情页。
   - **版本状态**：草稿、已提交、已发布。
   - **发布状态**：发布完成、尚未发布。
   - **创建人**：显示应用服务版本的创建人。
   - **创建时间**：显示应用服务版本的创建时间。
   - **操作**：可以单击 **发布** 发布应用服务，或单击 **拷贝到草稿** 生成新的应用服务版本草稿，单击 **提交** 后可进行发布。还可以单击 **版本对比** 查看当前版本与其他版本间的差异。

其他操作

在应用服务详情页，您还可以根据需要进行更多快捷操作，具体可参考下表。

操作	说明
更新应用服务	在应用服务详情页，单击 **编辑配置信息**。 详情参考 更新应用服务。
仅更换镜像	单击 **快捷操作**，选择 **更换镜像**，选择目标版本镜像即可完成镜像更新。 详情参考 更新应用服务。
仅更新伸缩配置	单击 **快捷操作**，选择 **扩缩容**，然后按需修改副本数即可完成伸缩配置更新。 详情参考 更新应用服务。
立即发布应用服务	在应用服务详情页，单击右上角 **发布** 按钮即可立即发布应用服务。 详情参考 立即发布服务。

8.6. 网络

8.6.1. 创建服务

通过创建不同类型的服务（Services），您可以为应用服务设置不同类型的访问策略。AKS 提供通过界面向导和直接编写 YAML 文件两种方式来创建服务。

用控制台向导创建 Service

1. 登录容器应用服务控制台，在左侧导航栏单击 **网络 > 服务（Services）**。
2. 在 **服务（Services）** 页面，单击 **创建服务**。
3. 在 **选择服务类型** 对话框中，选择所需服务类型，单击 **下一步**。
   支持的服务类型如下：
   - **集群 IP**：Cluster IP，支持从集群内访问服务。
   - **节点端口**：Node Port，支持从集群外的静态地址访问服务。
   - **负载均衡**：支持通过负载均衡器访问服务节点端口。
   - **自定义**：提交 YAML 创建自定义服务。

4. 配置服务的 **基本信息**，单击 **下一步**。
   基本信息包含的内容如下：
   - **服务名称**：必填。
   - **自定义 Label**：选填。
   - **自定义 Annotation**：选填。
5. 添加服务 **端口映射**（必填），和 **集群 IP**（选填），单击 **下一步**。
   您可以添加多个端口映射。端口映射包含的内容如下：
   - **名称**：必填。
   - **协议**：必选。
   - **前端端口**：必填。范围 1-65535。
   - **后端端口**：选填。范围 1-65535。
6. 设置 **Pod 选择器** 与 **应用服务选择器**，单击 **提交**。

用 YAML 创建 Service

1. 登录容器应用服务控制台，在左侧导航栏单击 **网络** > **服务（Services）**。
2. 在 **服务（Services）** 页面，单击 **创建服务**。
3. 在 **选择服务类型** 对话框中，选择 **创建自定义服务**。
4. 在输入框中输入 YAML 内容，单击 **提交**。
   样例如下：

```
 apiVersion: v1
 kind:Service
 metadata:
   name: test   # Enter your service-name
namespace:default
 spec:
   externalTrafficPolicy:Cluster
   ports:
- name: http
     nodePort:30184
     port:80
     protocol: TCP
     targetPort:80
- name: https
     nodePort:30487
     port:443
     protocol: TCP
     targetPort:443
   sessionAffinity:None
   type:LoadBalancer
 status:
   loadBalancer:{}
```

> **说明**
> - 若创建失败，在该页面，输入正确的 YAML 内容，单击 **提交**。
> - 若创建成功，系统会自动跳转到服务的详情页面，显示该服务的详细信息，包括事件列表和 YAML 内容。

8.6.2. 创建负载均衡

SOFAStack 提供公网和内网两种类型的负载均衡服务。您可以根据业务场景选择配置对外公开或对内私有的负载均衡，系统会根据您的选择分配公网或内网服务地址。

操作步骤

1. 在创建应用服务的 访问配置 时，单击 **创建负载均衡** 链接进入 **创建负载均衡** 页面。
2. 填写以下信息后，单击 **创建**。
   - **实例数量**：填写实例数量。
   - **实例名称**：长度限制为 1-50 个字符。
   - **实例类型**：选择 **内网** 或 **公网**。
   - **公网带宽计费方式**：选择 **按使用流量计费** 或 **按固定带宽计费**。该选项仅在选择公网类型时出现。
   - **网络类型**：专有网络。仅在选择内网类型时出现。
   - **交换机**：选择已有交换机或者新建交换机。仅当选择内网类型的负载均衡时需设置交换机。
   - **监听器信息**：添加监听器需填写以下信息。
     - **前端协议及端口**：支持 HTTP、HTTPS、TCP 协议。用来接收请求并向后端服务器进行请求转发的负载均衡系统的前端协议和端口。在同一个负载均衡实例内前端端口不可重复。
     - **后端端口**：后端协议端口，ECS 实例上开放的用来接收请求的后端端口，在同一个负载均衡实例内可重复。
     - **转发规则**：轮询模式或最小连接数模式。
     - **峰值带宽**：设定带宽峰值来限定后端 ECS 实例所能对外提供的服务能力。
     - X-Forwarded-For：获取真实 IP，考虑安全原因默认开启。
     - **会话保持**：开启会话保持后，负载均衡监听会把来自同一客户端的访问请求分发到同一台后端服务器上。
     - **健康检查**：开启健康检查功能后，当后端某个 ECS 健康检查出现问题时，负载均衡服务会将请求转发到其它健康检查正常的 ECS 上，而当该 ECS 恢复正常运行时，负载均衡会自动恢复它的请求转发。

8.6.3. 编辑与删除服务

本文介绍如何在 AKS 控制台更新或删除服务。

编辑服务

您可以通过修改服务的配置文件来实时更新服务。

1. 登录容器应用服务控制台，在左侧导航栏单击 **网络 > 服务（Services）**。
2. 在 **服务（Services）** 列表页面，单击服务名称进入详情页。
3. 在详情页，单击右上角的 **编辑**，选择 **编辑 YAML** 或 **编辑**（编辑控制台表单项）。
   编辑控制台表单时，服务的 YAML 文件会在右侧实时展示出来。

   > 说明
   > 专有云上目前只支持编辑 YAML。

4. 完成修改后，单击 **更新**。

删除服务

您可以在控制台上一键删除服务。

1. 登录容器应用服务控制台，在左侧导航栏单击 **网络 > 服务（Services）**。

2. 在 **服务（Services）** 列表页面，单击服务右侧的 **删除**，在弹出框中单击 **确定** 完成删除。

8.6.4. 创建路由

您可以通过创建路由服务，实现灵活可靠的流量路由。本文介绍如何在容器服务控制台通过界面向导创建路由（ingress）。

> **说明**
> 目前，仅支持从 VPC 内部访问内网 SLB，进而路由到 ingress 后端的服务。

前提条件

- 已成功创建一个集群，详见 创建集群。
- 已成功创建路由转发的相应服务，详见 创建服务。

创建路由

1. 登录容器应用服务控制台，在左侧导航栏单击 **网络 > 路由**。
2. 在 **路由** 页面，选择需要创建路由的命名空间后，单击 **创建路由**。
3. 在 **创建路由** 页面，配置路由的基本信息：
   - **路由名称**：必填。支持 1-63 个字符，由小写字母、数字和 - 组成，且必须以字母或数字开头和结尾。
   - **自定义 annotation**：选填。有关 ingress 的注解，详见 Annotations。
4. 配置具体的路由规则：即授权入站请求到达集群服务的路由规则，包括域名、访问路径、后端服务及端口等。
   - **域名**：必填。支持使用以下两种域名：
     - 基于默认域名的路由，即使用集群的默认域名对外提供访问服务。集群的默认域名格式是 `*.[cluster-id].[region-id].alicontainer.com`，可以在集群的基本信息页面中获取。
     - 基于自定义域名的路由，即使用一个虚拟的主机名称作为测试域名对外提供访问服务，如 `www.test.com`。使用自定义测试域名时，您需要在本地 hosts 文件中绑定域名映射规则。示例如下：

       ```
       47.103.95.125 www.test.com
       ```

       > **说明**
       > 该 IP 为 ingress 地址。可前往 **网络** > **服务** > **kube-system** 命名空间下的 **flyover-ingress** 服务对应的 VIP 地址。
   - **HTTP 规则**：必填。配置服务的访问路径、名称、端口以及权重等。
     - 路径：指定服务访问的 URL 路径，每个路径关联一个后端服务。路径必须以 `/` 开头，可直接使用根路径 `/`。

       > **说明**
       > 系统首先会自动判断路径是否为正则表达式，如果是，则采用正则匹配。反之，则采用精确匹配。
     - 服务名：选择路由转发的后端服务，必须在同一个命名空间下。

- 端口：选择服务暴露的前端端口。
- 负载规则：支持轮询、最小连接数、随机。
- 权重：设置该路径下多个服务的权重，服务权重采用相对值计算方式。支持单个路径对应到多个服务的模式，可以灵活配置路由到多个服务的相对权重。

- **Header 规则**：选填。设置 Header 规则后，请求头中满足匹配规则的请求才能被路由到对应路径所指定的服务中，实现基于 Request Header 的流量切分。
  - 关联 HTTP 规则：由 `域名+路径` 组成。
  - 名称、匹配值：自定义的请求字段，名称和匹配值为键值对。
  - 匹配方式：支持正则匹配和精确匹配。

5. 单击 **提交**。

8.6.5. 编辑与删除路由

本文介绍如何在 AKS 控制台更新或删除路由。

前提条件

- 已成功创建了一个集群，详见 创建集群。
- 在该集群下已成功创建了一个路由，详见 创建路由。

编辑路由

1. 进入容器应用服务控制台，在左侧导航栏单击 **网络 > 路由**。
2. 在路由列表页面，单击路由右侧操作列的 **编辑**。
3. 在 **编辑路由** 页面，可修改自定义 annotation 以及具体的路由规则。
4. 修改完成后，单击 **提交**。

删除路由

1. 进入容器应用服务控制台，在左侧导航栏单击 **网络 > 路由**。
2. 在路由列表页面，单击路由右侧操作列的 **删除**，在弹出框中单击 **确定** 完成删除。

8.7. 存储

8.7.1. 创建存储类型（StorageClass）

本文向您介绍如何创建一个存储类型（StorageClass）。
目前，AKS 支持以下两种类型的存储类型：

- NAS 类型
- 云盘（cloud_disk）类型

> 说明
> 目前敏捷 PaaS 底座环境下，AKS 未对接阿里云的 NAS 及云盘存储，若要使用存储功能，需要您安装第三方存储插件。

NAS 类型

> 注意
> NAS 类型的 Storage Class 必须和 AKS 集群节点在相同地域。

AKS 定义的 NAS 类型的 Storage Class 如下：

```
apiVersion: storage.k8s.io/v1

kind:StorageClass
metadata:
    name: csi-nas-capacity
provisioner: antcloud-csi
parameters:
    storage_type:"cloud_nas"# Allowed enum: cloud_disk, cloud_nas
# filesystem parameters
    nasType:"Capacity"# Allowed enum: Capacity, Performance, Extreme *
    protocolType:"NFS"# Allowed enum: NFS, SMB *
    serviceId:"1c7aff75-8b4f-4d2c-8dac-74af0b830d86-aliyun-nas-service-id"
    planId:"44dd37f1-77f9-42c5-91a5-743c007ba484-aliyun-nas-plan-id"
reclaimPolicy:Retain# Allowed enum: Delete, Retain
volumeBindingMode:Immediate# Allowed enum: Immediate, WaitForFirstConsumer
```

参数说明

参数名	参数说明	是否必选	备注
provisioner	antcloud-csi	是	AKS 动态卷控制器名称，不可更改。
storage_type	后端存储类型（块存储，文件存储）	是	目前仅支持： • cloud_disk，表示阿里云盘 • cloud_nas，表示阿里云 NAS。
nasType	NAS 类型	是	可选值： • **Capacity**，详情参见 容量型 NAS。 • **Performance**，详情参见 性能型 NAS。 • **Extreme**，详情参见 极速型 NAS。
protocolType	文件系统类型	否	目前仅支持 NFS
readOnly	是否只读	否	可选值：**true**，**false**
serviceId	Service broker 必选参数	是	必须为：1c7aff75-8b4f-4d2c-8dac-74af0b830d86-aliyun-nas-service-id

参数名	参数说明	是否必选	备注
planId	Service broker必选参数	否	必须为：44dd37f1-77f9-42c5-91a5-743c007ba484-aliyun-nas-plan-id
reclaimPolicy	回收策略	是	• **Retain**：保留后端存储，删除 PV/PVC 不会删除对应的后端存储如云盘。详情参考 Kubernetes 官网文档。 • **Delete**：当删除 PVC 时，自动删除 PV 和后端的存储。详情参考 Kubernetes 官网文档。
volumeBindingMode	存储卷绑定策略	是	控制存储卷绑定和动态供应行为 • **Immediate**：创建 PVC 后立即创建后端存储卷，并且立即绑定新创建的 PV 和 PVC。 • **WaitForFirstConsumer**：当 PVC 被 Pod 使用时，才触发 PV 和后端存储的创建，同时实现 PVC/PV 的绑定，启用该配置后，Storage Class 中的 Zone 和 Region 将不再生效，而是使用 Pod 调度所在节点的 zone 和 region 创建文件系统，保证文件系统能被 Pod 挂载。

云盘（cloud_disk）类型

> **注意**
> 云盘类型的 Storage Class 必须和 AKS 集群节点在相同地域及可用区。

AKS 定义的云盘类型的 Storage Class 如下：

```
apiVersion: storage.k8s.io/v1
kind:StorageClass
metadata:
   name: csi-disk-efficiency
provisioner: antcloud-csi
parameters:
    failure-domain.beta.kubernetes.io/zone:<必须与当前集群中节点可用区匹配>
    failure-domain.beta.kubernetes.io/region:<必须与当前集群所在region匹配>
    fsType:"ext4"
    storage_type:"cloud_disk"# Allowed enum: cloud_disk, cloud_nas
    type:"cloud_ssd"
    readOnly:"false"
    serviceId:"8814fddf-5be7-4f8a-862c-f4a57c25a0b4-aliyun-disk-service-id"
    planId:"02b41c80-915f-43cd-a574-e9cd446f13ea-aliyun-disk-plan-id"
reclaimPolicy:Retain# Allowed enum: Delete, Retain
volumeBindingMode:Immediate# Allowed enum: Immediate, WaitForFirstConsumer
```

参数说明

参数名	说明	是否必填	备注
provisioner	antcloud-csi	是	动态卷控制器名称，不可更改。
failure-domain.beta.kubernetes.io/region	地域	否	该字段当 `volumeBindingMod=Immediate` 时为必选，会被用来创建云盘。例如：CN-SHANGHAI，代表上海地域。
failure-domain.beta.kubernetes.io/zone	可用区	否	该字段当 `volumeBindingMod=Immediate` 时为必选，会被用来创建云盘。例如：CN-SHANGHAI-A，代表上海可用区 A。
fsType	分区格式	否	支持 ext4、ext3，默认 ext4。
storage_type	后端存储类型	是	• cloud_disk，表示阿里云云盘 • cloud_nas, 表示阿里云 NAS。
type	磁盘类型	是	可选值：cloud_ssd、cloud_efficiency、cloud。详情参考 云盘类型。

参数名	说明	是否必填	备注
readOnly	是否只读	否	可选值：true、false。
serviceId	Service broker必选参数	是	必须为： `8814fddf-5be7-4f8a-862c-f4a57c25a0b4-aliyun-disk-service-id`
planId	Service broker必选参数	否	必须为： `02b41c80-915f-43cd-a574-e9cd446f13ea-aliyun-disk-plan-id`
reclaimPolicy	回收策略	是	• **Retain**：保留后端存储，删除 PV/PVC 不会删除对应的后端存储如云盘。详情参考 Kubernetes 官网文档。 • **Delete**：当删除 PVC 时，自动删除 PV 和后端的存储。详情参考 Kubernetes 官网文档。
volumeBindingMode	存储卷绑定策略	是	控制存储卷绑定和动态供应行为。 • **Immediate**：创建 PVC 后立即创建后端存储卷，并且立即绑定新创建的 PV 和 PVC • **WaitForFirstConsumer**：当 PVC 被 pod 使用时，才触发 PV 和后端存储的创建，同时实现 PVC 和 PV 的绑定，启用该配置后，storage class 中的 zone 和 region 将不再被使用，而是使用 pod 调度所在节点的 zone 和 region 创建云盘，保证云盘能被 Pod 挂载。

前提条件

需 AKS 管理员权限。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **存储 > 存储类型**。
2. 在 **存储类型** 页面，单击 **创建存储类型**。在输入框中输入 YAML 内容，单击 **提交**。

8.7.2. 创建存储卷（PV）

您可以使用 YAML 文件或者控制台界面创建存储卷（PV）。

> 说明
> 目前敏捷 PaaS 底座环境下，AKS 未对接阿里云的 NAS 及云盘存储，若要使用存储功能，需要您安装第三方存储插件。

前提条件

已创建存储类型（StorageClass）。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **存储 > 存储卷（PersistentVolume）**。
2. 在 **存储卷** 页面，单击 **创建存储卷**。
3. 在输入框中输入 YAML 格式的内容，单击 **提交**。
   样例如下：

```
apiVersion: v1
kind: PersistentVolume
metadata:
  name: pv-nas
  labels:
    alicloud-pvname: pv-nas
spec:
  capacity:
    storage: 5Gi
  accessModes:
    - ReadWriteMany
  csi:
    driver: nasplugin.csi.alibabacloud.com
    volumeHandle: pv-nas
    volumeAttributes:
      server: "25***49129-ys**7.cn-shenzhen.nas.aliyuncs.com"
      path: "/csi"
  mountOptions:
  - nolock,tcp,noresvport
  - vers=3
```

参数说明

参数名	参数说明
driver	驱动类型。
volumeHandle	配置 PV 的名称。
server	NAS 挂载点。

参数名	参数说明
path	挂载子目录。极速型 NAS 需要以 `/share` 开头。
vers	挂载 NAS 数据卷的 NFS 协议版本号。推荐使用 v3，因为极速 NAS 仅支持 v3。

8.7.3. 创建存储卷声明（PVC）

本文向您介绍如何通过 YAML 文件创建一个存储卷声明（PersistentVolumeClaim）。

> 说明
> 目前敏捷 PaaS 底座环境下，AKS 未对接阿里云的 NAS 及云盘存储，若要使用存储功能，需要您安装第三方存储插件。

前提条件

已 创建存储类型（StorageClass）。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **存储 > 存储卷声明（PersistentVolumeClaim）**。
2. 在 **存储卷声明** 页面，单击 **创建存储卷声明**。
3. 在输入框中输入 YAML 格式的内容，单击 **提交**。
   样例如下：

```
 apiVersion: v1
 kind:PersistentVolumeClaim
 metadata:
   name: test-pvc
namespace:default
 spec:
   accessModes:
-ReadWriteOnce
   resources:
     requests:
       storage:25Gi
   storageClassName: csi-disk-efficiency # 对应 storage class 的名称
```

> 注意
> 对于磁盘类型设置为 cloud_efficiency 和 cloud_ssd 类型的 Storage Class，其对应的最小存储空间是 20 Gi，当 PVC 中指定的大小小于 20 Gi 时，PV 无法创建成功，PVC 会一直处于 pending 状态。

PVC 创建完成后，您可以在 YAML 文件的 status 中查看创建情况。PVC 创建成功，对应的存储卷会展示在 **存储卷** 页面。

8.7.4. 使用 PVC

PVC 创建成功后，您可以在创建 Pod 时使用 PVC。

注意

- 云盘不能跨可用区（zone）挂载，因此在创建 `volumeBindingMode: Immediate` 类型的 PVC 后，使用该 PVC 的pod/workload 需要通过 nodeSelector/nodeAffinity 限制 pod 运行在 PV/PVC 同一个可用区。
- 目前敏捷 PaaS 底座环境下，AKS 未对接阿里云的 NAS 及云盘存储，若要使用存储功能，需要您安装第三方存储插件。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载 > 容器组（Pods）**。
2. 在 Pods 页面，单击 **从 YAML 创建**。
3. 在输入框中输入 YAML 内容，单击 **提交**。

```
 apiVersion: v1
 kind:Pod
 metadata:
   name: pod-with-disk
namespace:default
   labels:
     app: nginx
 spec:
   tolerations:
- key: node.kubernetes.io/unreachable
     effect:NoSchedule
operator:Exists
   automountServiceAccountToken:false
   containers:
- name: busybox
     image: reg-cnsh-nf.cloud.alipay.com/library/busybox
     command:["sleep","3000"]
     ports:
- containerPort:80
     volumeMounts:
- name: data
         mountPath:"/data"
   volumes:
- name: data
       persistentVolumeClaim:
         claimName: test-pvc
```

容器运行后，可以看到 pod 的状态为 running：

```
# kubectl get pod
NAME            READY   STATUS    RESTARTS   AGE
pod-with-disk   1/1Running0103s
```

8.7.5. 使用 PVC 挂载动态存储卷

挂载动态存储卷是基于 Storage Class 实现的，在挂载动态存储卷前，需要 AKS 管理员创建存储类型（StorageClass）。

> 说明
> 目前敏捷 PaaS 底座环境下，AKS 未对接阿里云的 NAS 及云盘存储，若要使用存储功能，需要您安装第三方存储插件。

前提条件

已 创建存储类型。

创建 PVC

您需要在创建 Pod 前创建存储声明，供创建 Pod 时使用。PVC 创建后，会有对应的 PV 自动创建出来。

云盘类型 PVC 样例

```
apiVersion: v1
kind:PersistentVolumeClaim
metadata:
  name: test-pvc
spec:
  accessModes:
-ReadWriteOnce
  resources:
    requests:
      storage:25Gi
  storageClassName: csi-disk-efficiency # csi-disk-ssd
```

参数说明

参数名	参数说明	是否必选	备注
accessModes	只读/读写	否	仅支持：ReadWriteOnce
storage	存储空间大小	是	遵循后端存储限制，参考阿里云云盘限制。对于 csi-disk-ssd 和 csi-disk-efficiency 两种 Storage Class，最小的存储空间都是 20Gi，否则 PV 无法创建成功，PVC 会一直处于 pending 状态。
storageClassName	storage class名字	是	必选指定一个已经存在的名字，否则云盘无法创建。

NAS 类型 PVC 样例

```
apiVersion: v1
kind:PersistentVolumeClaim
metadata:
  name: test-pvc
spec:
  accessModes:
-ReadWriteMany
  resources:
    requests:
      storage:25Gi
  storageClassName: csi-nas-capacity # 或者 csi-nas-performance
```

> **注意**
> 对于 NAS 存储，其对应的存储空间限制能力将不会生效，不会限制容器在运行时数据的写入量，仅仅用做 PV-PVC 在绑定过程的匹配。

参数说明

参数名	参数说明	是否必选	备注
accessModes	只读/读写	否	可选值：-ReadWriteOnce-ReadWriteMany
storage	存储空间大小	是	遵循后端存储限制，如阿里云 NAS 限制。文件系统容量上限都为 1PB（性能型）或者 10PB（容量型）。
storageClassName	storage class 名字	是	必选指定一个已经存在的名字，否则云盘无法创建。

使用 PVC

您可以在创建 Pod 时使用 PVC 挂载动态数据卷。

```
apiVersion: v1
kind: Pod
metadata:
  name: pod-with-nas
  labels:
    app: nginx
spec:
  tolerations:
  - key: node.kubernetes.io/unreachable
    effect: NoSchedule
    operator: Exists
    automountServiceAccountToken: false
  containers:
  - name: busybox
    image: 'reg-cnsh-nf.cloud.alipay.com/library/busybox'
    command: ["sleep","3000"]
    ports:
    - containerPort: 80
    volumeMounts:
    - name: data
      mountPath: "/data"
    volumes:
    - name: data
      persistentVolumeClaim:
        claimName: test-pvc
```

容器运行后，可以看到 Pod 的状态为 running：

```
# kubectl get pod
NAME            READY   STATUS    RESTARTS   AGE
pod-with-nas   1/1Running0103s
```

8.7.6. 使用 PVC 挂载静态存储卷

您可以通过绑定已存在的 PV 为 Pod 挂载静态存储卷。

> 说明
> 目前敏捷 PaaS 底座环境下，AKS 未对接阿里云的 NAS 及云盘存储，若要使用存储功能，需要您安装第三方存储插件。

前提条件

- 已 创建存储类型。
- 同租户下存在可用云盘。

创建静态 PV

创建 PV 并绑定一个已经存在的云盘。您可以使用 YAML 文件或者控制台界面创建 PV。
YAML 样例如下：

```
apiVersion: v1
kind:PersistentVolume
metadata:
  labels:
    cafe.sofastack.io/diskId: d-zm0fd73z6tmf5j7pgbpe
  name: pv-efficient
spec:
  accessModes:
-ReadWriteOnce
  capacity:
    storage:25Gi
  csi:
    driver: antcloud-csi
    fsType: ext4
    volumeAttributes:
      storage_type: cloud_disk
    volumeHandle: instance_test_efficiency
  storageClassName: csi-disk-efficiency
```

参数说明

参数名	参数说明	是否必选	备注
cafe.sofastack.io/diskId	阿里云的云盘 ID，格式为 `d-xxx` 。	是	您可以在资源管理控制台上的 **云服务器 ECS > 存储与快照 > 云盘** 中获取云盘 ID。
volumeAttributes	磁盘属性	是	遵循后端存储限制，如 阿里云盘限制。对于 csi-disk-ssd 和 csi-disk-efficiency 两种 Storage Class，最小的 size 都是 20G。
volumeHandle	表示后端存储的唯一 ID，由用户指定。	是	-

创建 PVC 并指定其绑定到 PV

您可以使用 YAML 文件或者控制台界面创建 PVC。
YAML 样例如下：

```
apiVersion: v1
kind:PersistentVolumeClaim
metadata:
 name: test-pvc
namespace:default
spec:
 accessModes:
-ReadWriteOnce
 resources:
   requests:
     storage:25Gi
 storageClassName: csi-disk-efficiency
 volumeName: pv-efficient
```

> 说明
> 必须要有 volumeName 字段，指定已有的 PV。参数说明参考 使用 PVC 挂载动态存储卷。

使用 PVC

您可以在创建 Pod 时使用 PVC 挂载静态数据卷。

```
apiVersion: v1
kind:Pod
metadata:
  name: pod-with-disk
namespace:default
  labels:
    app: nginx
spec:
  tolerations:
- key: node.kubernetes.io/unreachable
    effect:NoSchedule
operator:Exists
  automountServiceAccountToken:false
  containers:
- name: busybox
    image: reg-cnsh-nf.cloud.alipay.com/library/busybox
    command:["sleep","3000"]
    ports:
- containerPort:80
    volumeMounts:
- name: data
       mountPath:"/data"
  volumes:
- name: data
     persistentVolumeClaim:
       claimName: test-pvc
```

容器运行后，可以看到 Pod 的状态为 running：

```
# kubectl get pod
NAME            READY   STATUS    RESTARTS   AGE
pod-with-disk   1/1Running0103s
```

8.7.7. StatefulSet 挂载动态存储卷

> 说明
> 目前敏捷 PaaS 底座环境下，AKS 未对接阿里云的 NAS 及云盘存储，若要使用存储功能，需要您安装第三方存储插件。

前提条件

- 已 创建存储类型（StorageClass）。
- 已 创建服务。

使用 NAS 创建多实例 StatefulSet

Yaml 样例如下：

```
apiVersion: apps/v1beta2
kind:StatefulSet
metadata:
  name: web
spec:
  selector:
    matchLabels:
      app: nginx
  serviceName:"<指定已有的Service>"
  replicas:2
template:
    metadata:
      labels:
        app: nginx
    spec:
      containers:
- name: nginx
        image: nginx
        ports:
- containerPort:80
          name: web
        volumeMounts:
- name: csi-nas-pvc
          mountPath:/data
  volumeClaimTemplates:
- metadata:
      name: csi-nas-pvc
    spec:
      accessModes:["ReadWriteOnce"]
      storageClassName:"csi-nas-capacity"
      resources:
        requests:
          storage:25G
```

使用云盘创建多实例 StatefulSet

Yaml 样例如下：

```
apiVersion: apps/v1beta2
kind:StatefulSet
metadata:
  name: web
spec:
  selector:
    matchLabels:
      app: nginx
  serviceName:"<指定已有的Service>"
  replicas:2
template:
    metadata:
      labels:
        app: nginx
    spec:
      containers:
- name: nginx
        image: nginx
        ports:
- containerPort:80
          name: web
        volumeMounts:
- name: disk-efficiency
          mountPath:/data
  volumeClaimTemplates:
- metadata:
      name: disk-efficiency
    spec:
      accessModes:["ReadWriteOnce"]
      storageClassName:"csi-disk-efficiency"
      resources:
        requests:
          storage:25G
```

参数说明参考 挂载动态存储卷。

8.8. 配置

8.8.1. 配置项

配置项（ConfigMap）是一种用于存储工作负载所需配置信息的资源类型，需要您自定义配置项的内容。

创建配置项

本部分介绍如何通过容器应用服务控制台完成配置项创建操作。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **配置 > 配置项**。
2. 单击 **创建**，在 **创建配置项** 页面，填写以下信息：
   - **命名空间**：选择命名空间。若不选择，默认为 **default**。
   - **名称**：配置项名称，同一个命名空间里命名必须唯一，最多 32 个字符。
   - **配置源**：
     - 键值对：键代表文件名，值代表文件中的内容。
     - 文件：可上传文件创建配置项。

3. 配置完成后，单击 **创建**。

使用配置项

您可以在 Pod 中使用配置项，主要包括以下使用场景：

- 用配置项定义 Pod 环境变量
- 通过配置项设置命令行参数
- 在数据卷中使用配置项

> **注意**
> 在 Pod 里使用配置项时，需两者处于同一命名空间中。

更多关于配置项的信息，可参见 Configure a Pod to Use a ConfigMap。

用配置项定义 Pod 环境变量

您可以使用配置项定义 Pod 中环境变量，使用 `valueFrom` 引用配置项数据。

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载 > 容器组（Pods）**。
2. 单击 **从 YAML 创建**，在输入框中输入以下内容，单击 **创建**。
   下面是一个编排示例。

```
apiVersion: v1
kind:Pod
metadata:
   name: config-pod-1
spec:
   containers:
- name: test-container
       image: busybox
       command:["/bin/sh","-c","env"]
       env:
- name: SPECIAL_LEVEL_KEY
           valueFrom:##使用valueFrom来指定env引用配置项的value值
             configMapKeyRef:
               name: special-config                ##引用的配置文件名称
               key: SPECIAL_LEVEL                  ##引用的配置项key
   restartPolicy:Never
```

通过配置项设置命令行参数

您可以使用配置项设置容器中的命令或者参数值，使用环境变量替换语法 `$(VAR_NAME)` 来进行。

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载 > 容器组（Pods）**。
2. 单击 **从 YAML 创建**，在输入框中输入以下内容，单击 **创建**。
   下面是一个编排示例。

```
apiVersion: v1
kind:Pod
metadata:
   name: config-pod-3
spec:
   containers:
- name: test-container
       image: busybox
       command:["/bin/sh","-c","echo $(SPECIAL_LEVEL_KEY) $(SPECIAL_TYPE_KEY)"]
       env:
- name: SPECIAL_LEVEL_KEY
           valueFrom:
             configMapKeyRef:
               name: special-config
               key: SPECIAL_LEVEL
- name: SPECIAL_TYPE_KEY
           valueFrom:
             configMapKeyRef:
               name: special-config
               key: SPECIAL_TYPE
   restartPolicy:Never
```

运行该 Pod 后，会输出如下结果。

```
very charm
```

在数据卷中使用配置项

您也可以在数据卷里面使用配置项，在 volumes 下指定配置项名称，会将 key/values 的数据存储到 mountPath 路径下（本例中是 `/etc/config` ）。最终生成以 key 为文件名，values 为文件内容的配置文件。

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载 > 容器组（Pods）**。
2. 单击 **从 YAML 创建**，在输入框中输入以下内容，单击 **创建**。
   下面是一个编排示例。

```
apiVersion: v1
kind:Pod
metadata:
   name: config-pod-4
spec:
   containers:
- name: test-container
       image: busybox
       command:["/bin/sh","-c","ls /etc/config/"]##列出该目录下的文件名
       volumeMounts:
- name: config-volume
         mountPath:/etc/config
   volumes:
- name: config-volume
       configMap:
         name: special-config
   restartPolicy:Never
```

运行 Pod 后，会输出配置项的 key。

```
SPECIAL_TYPE    SPECIAL_LEVEL
```

修改配置项

修改配置文件会影响使用该配置文件的应用。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **配置 > 配置项**。
2. 在配置项列表页，单击 **名称** 进入配置项详情页。
3. 您可以进行以下修改：
   - 单击 **编辑**，修改已有的配置项数据，单击 **提交**。
   - 单击 **添加配置项数据**，增加变量名和值，单击 **提交**。

查看配置项

前提条件

已 创建配置项。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **配置 > 配置项**。
2. 在配置项列表页，单击 **名称** 进入配置项详情页，查看配置项详细数据。

8.8.2. 保密字典

创建保密字典

保密字典（Secret）是一种用于存储工作负载所需要认证信息、密钥的敏感信息等的资源类型。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **配置管理 > 保密字典**。
2. 在保密字典列表页，单击 **创建**。
3. 在 **创建保密字典** 页，填写以下信息：
   - 命名空间：选择命名空间，默认为 default。
   - 名称：保密字典名称。
   - 保密字典类型：
     - Opaque：一般密钥类型。输入键、值。值必须使用 Base64 编码。
     - ImagePullSecret：存放拉取私有仓库镜像所需的认证信息。输入容器注册地址及用户名和密码。
4. 配置完成后，单击 **创建**。

使用保密字典

您可以在工作负载环境变量、命令行参数和数据卷三个场景使用保密字典（Secret），也可以在 创建应用服务 的配置 Pod 模板步骤中使用保密字典拉取镜像。

前提条件

- 已 创建保密字典。
- 在 Pod 里使用保密字典时，需 Pod 和保密字典处于同一命名空间中。
- 本文以下面这个保密字典为例介绍保密字典的用法。

```
apiVersion: v1
kind:Secret
metadata:
  name: mysecret
type:Opaque
data:
  username:my-username #用户名
  password:******#需要用Base64编码
```

用保密字典配置 Pod 的数据卷

保密字典可以在 Pod 中作为文件使用。

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载 > 容器组（Pods）**。
2. 单击 **从 YAML 创建**，在输入框中输入以下内容，单击 **提交**。
   如以下 Pod 示例所示，mysecret 密钥的 username 和 password 以文件方式保存在 `/etc/foo` 目录下。

```
 apiVersion: v1
 kind:Pod
 metadata:
   name: mypod
 spec:
   containers:
- name: mypod
     image: redis
     volumeMounts:
- name: foo
       mountPath:"/etc/foo"
       readOnly:true
   volumes:
- name: foo
     secret:
       secretName: mysecret
```

用保密字典设置 Pod 的环境变量

保密字典可以在 Pod 中设置为环境变量。

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载 > 容器组（Pods）**。
2. 单击 **从 YAML 创建**，在输入框中输入以下内容，单击 **提交**。
   如以下 Pod 示例所示，mysecret 密钥的 username 和 password 配置为 Pod 的环境变量。

```
 apiVersion: v1
 kind:Pod
 metadata:
   name: secret-env-pod
 spec:
   containers:
- name: mycontainer
     image: redis
     env:
- name: SECRET_USERNAME
         valueFrom:
           secretKeyRef:
             name: mysecret
             key: username
- name: SECRET_PASSWORD
         valueFrom:
           secretKeyRef:
             name: mysecret
             key: password
   restartPolicy:Never
```

修改保密字典

对于 Opaque 类型的保密字典，您可以继续修改保密字典数据。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **配置管理 > 保密字典**。
2. 在保密字典列表页，单击 **名称** 进入保密字典详情页。
3. 您可以进行以下修改：
   - 单击 **编辑**，修改已有的配置项数据，单击 **提交**。
   - 单击 **添加保密字典数据**，增加变量名和值，单击 **提交**。

查看保密字典

前提条件

已创建保密字典。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **配置管理 > 保密字典**。
2. 在保密字典列表页，单击 **名称** 进入保密字典详情页。

> **说明**
> 对于 Opaque 类型的保密字典，您可以继续添加保密字典数据。

删除保密字典

您可以删除不再使用的保密字典。

> **注意**
> 删除保密字典可能会影响引用该保密字典的应用容器。建议您在确保没有引用该保密字典的应用容器后，再进行删除操作。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **配置管理** > **保密字典**。
2. 在保密字典列表页，单击 **删除**。
3. 在弹出框中，单击 **确认**。

8.9. 应用商店

8.9.1. 概述

AKS 应用商店集成了 Helm（V3 版本）的能力，并进行了相应扩展，例如提供图形化界面、官方仓库、版本回滚等。

> 说明
> Helm 是管理 Kubernetes 应用程序的打包工具。更多详情请查看 Helm 官网文档。

应用商店支持以下功能：

- **产品模板**：
  - 私有平台：支持上传用户自定义的应用模板、应用模板包、新版本。
  - 公有平台：支持引入并管理第三方应用模板仓库，以及将第三方模板复制到私有平台。
  - 版本控制：针对模板变更自动做版本号的增加，而不是单纯覆盖。
  - 下载模板：支持下载私有平台中的应用模板包。

    > 说明
    > 您可以对私有平台中的产品模板进行增加、删除、编辑内容的操作，公有平台的产品模板仅支持增删操作。若要对公有平台的产品模板进行定制，可以先将第三方模板复制至私有平台。

- **产品发布**：
  - 快速部署：选择私有或公有平台的应用模板，自定义参数并快速地一键部署常用的应用到 Kubernetes 中。
  - 自定义参数：模板安装时用户自定义的参数，会在安装详情页里作持久化保留，而不是覆盖原有商店中的配置。
  - 版本回滚：当前模板有其它成功发布的历史版本（除当前版本外）时，可以进行一键回滚。

8.9.2. 产品模板

产品模板为基于 Kubernetes Helm 标准的应用模板提供了统一高效的管理机制。

私有平台

私有平台相当于用户的私有仓库，存储以下两种类型的模板。

- 平台模板是由 SOFAStack 平台提供和维护的官方示例模板，让您可以即开即用。该类模板仅支持查看和安装操作，用户无法修改模板信息。
- 私有模板满足用户管理和维护自定义模板的需求，私有平台的模板的来源和用途局限于当前集群。您可以对模板内容进行增删改查。

上传私有模板

> 说明
> 目前没有限制模板的上传个数。

前提条件

已准备好符合 Helm 规范的产品模板。 应用开发规范，请参见 官方文档。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用商店 > 产品模板**，进入私有平台页面。
2. 单击 **上传产品模板**，在弹出框中选择准备好的自定义产品模板。

   > 说明
   > 仅允许上传 tgz/zip 格式的文件，推荐使用 `helm package` 打包。

   上传成功后，模板会以卡片的形式展示在列表页。单击卡片进入 **模板详情** 页，您可以继续进行如下操作：
   - 在 **模板版本** 页签中查看、下载、删除历史版本。
   - 在 **模板详情** 页签中查看修改编辑模板内容，修改后单击 **保存**。

公有平台

公有平台作为全局仓库，可以引入第三方模板源。公有平台的模板不允许用户修改内容，您可以安装以及复制模板到私有平台。

添加外部源

> 说明
> 最多可支持添加 5 个外部源。

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用商店 > 产品模板**，单击 **公有平台** 页签进入公有平台页面。
2. 单击 **添加外部源**，在弹出框中填写以下配置：
   - **外部源名称**：必须由英文字母、数字、下划线和中划线组成。
   - **外部源 URL 地址**：必须以 http、https 开头，且需符合 URL 规范。
3. 单击 **确定**。
   添加成功后，外部仓库中的模板会自动导入进来。

导入私有平台

若对公有平台模板有定制化需求，您可以先将公有模板导入到私有平台中。

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用商店 > 产品模板**，单击 **公有平台** 页签进入公有平台页面。
2. 单击模板卡片上的 **导入私有平台**，在弹出框中单击 **确定**。
   导入成功后，您可以导私有平台中搜索到该模板，并可以对模板内容进行编辑和版本管理。

8.9.3. 产品安装

通过安装产品模板，您可以将标准的 Helm 应用模板发布为相应的 Kubernetes workloads。

> 说明
> 如果是通过 OAM 模型定义的应用模板，系统目前仅会生成待发布状态的应用服务版本，需要您继续在应用服务的版本页面进行应用发布，具体步骤请参见 立即发布服务。

前提条件

已 上传产品模板 或 添加外部源。

安装私有模板

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用商店 > 产品模板 > 私有平台 > 私有模板**，进入模板列表页面。
2. 单击模板卡片进入模板详情页，单击右上角的 **产品安装**。
3. 在弹出框中填入以下配置。
   - **发布名称**：支持英文小写字母、数字、中划线，必须以英文小写字母开头，最大长度为 32 个字符。
   - **命名空间**：选择相应的命名空间。
   - **版本**：选择模板的版本。
   - **描述**：选填。最大长度 100 个字符。
4. 单击 **提交**，进入 **产品安装详情** 页。
5. 单击 **开始\重新安装**，在弹出框中单击 **确定**。
   - 若安装失败，可以单击 **发布错误日志** 查看错误日志排查问题。
   - 若要终止安装，可以单击 **取消安装**，在弹出框中单击 **确定**。
   - 安装成功后，可以单击发布详情中的工作负载或者应用服务链接进入对应的详情页面。

发布公有模板

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用商店 > 产品模板**，进入私有模板列表页面。
2. 选择 **公有平台** 页签，进入公有模板列表页面。
3. 单击公有模板卡片进入模板内容详情。
4. 单击右上角的 **产品安装**，在弹出框中填入以下配置。
   - **发布名称**：支持英文小写字母、数字、中划线，必须以英文小写字母开头，最大长度为 32 个字符。
   - **命名空间**：选择相应的命名空间。
   - **版本**：选择模板的版本。
   - **描述**：选填。最大长度 100 个字符。
5. 单击 **提交**，进入 **产品安装详情** 页。
6. 单击 **开始\重新安装**，在弹出框中单击 **确定**。
   - 若安装失败，可以单击 **发布错误日志** 查看错误日志排查问题。
   - 若要终止安装，可以单击 **取消安装**，在弹出框中单击 **确定**。
   - 安装成功后，可以单击发布详情中的工作负载或者应用服务链接进入对应的详情页面。

更新版本

在产品模板安装前后，您可以快速更新要安装的产品模板版本。

1. 在 **产品安装详情** 页，单击右上角的 **更新版本**。
2. 在弹出窗中选择要更新到到版本，单击 **确定**。

删除实例

如果不需要继续使用某个应用，您可以删除实例，删除操作会释放所有后端资源。

1. 在 **产品安装详情** 页，单击右上角的 **删除实例**。
2. 在弹出窗口中单击 **确定**。

回滚历史快照

您可以一键将应用回滚到某个历史版本，默认保留最近 10 个历史版本记录。

> 说明
> 仅成功安装的历史版本可以支持回滚操作。

1. 在 **产品安装详情** 页，选择 **历史版本** 页签。
2. 单击目标版本操作栏中的 **回滚** 选项，在弹出窗口单击 **确定**。

8.9.4. 自定义 OAM 模板

应用商店支持从根据 OAM 模型定义的应用模板一键生成应用服务（CafeDeployment），您可以通过自定义 OAM 模板来快速创建应用服务。本文以示例模板为例阐释如何编写符合 OAM 规范的产品模板。

背景介绍

OAM 是阿里巴巴联合微软在社区推出的一款用于构建和交付云原生应用的标准规范，旨在通过全新的应用定义、运维、分发与交付模型，推动应用管理技术向轻运维的方向迈进。对于 Kubernetes 来说，OAM 既是一个标准的应用定义项目（类比已经不再活跃的 Kubernetes Application CRD 项目），同时也是一个专注于封装、组织和管理 Kubernetes 中各种运维能力、以及连接运维能力与应用的平台层项目。而通过同时提供定义应用和组织管理应用的运维能力这两大核心功能，OAM 项目自然成为了阿里巴巴进行统一应用架构升级和构建下一代 PaaS/Serverless 过程中的关键路径。

在 OAM 中，一个应用程序包含三个核心概念：

- **Component**：组成应用程序的 **组件**，它可能包含微服务集合、数据库和云负载均衡器。
- **Trait**：描述应用程序 **运维特征** 的集合，例如，弹性伸缩和 Ingress 等功能。它们对应用程序的运行至关重要，但在不同环境中其实现方式各不相同。
- **Application Configuration**：最后，为了将这些描述转化为具体的应用程序，运维人员使用 **应用配置** 来组合组件和相应的特征，以构建应部署的应用程序的具体实例。

更多关于 OAM 的介绍，请参考 官方文档。

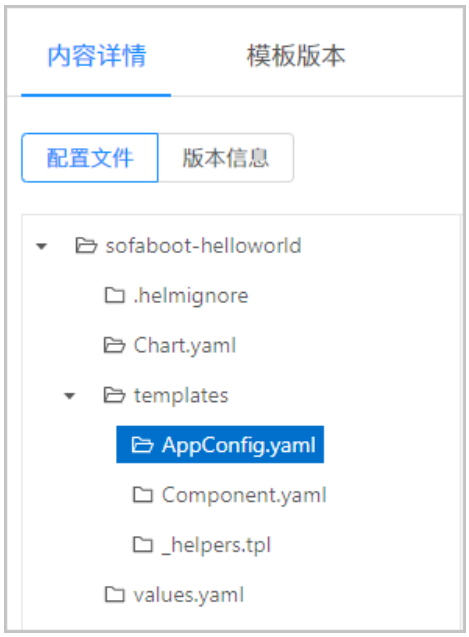
准备工作

下载以下任意模板，并将模板上传至 AKS **应用商店 > 产品模板 > 私有模板** 中。

- sofaboot-helloworld 模板：展示发布策略。
- sofaboot-helloworld-service 模板：展示网络策略。

目录说明

上传模板后，模板中的配置文件目录如下：



您需要关注的配置文件如下：

- templates：
  - AppConfig.yaml：应用配置文件，一般由运维人员编写。
  - Component.yaml：组件配置文件，一般由研发人员编写。
- values.yaml: 应用名称、用户的登录名以及访问中间件服务的四元组信息。

> 说明
> 必须填写 app（应用名称）、access（访问控制键）、secret（访问控制密钥），请参考 创建 AccessKey 获取 access 及 secret。

```
# Default values for mychart.
# This is a YAML-formatted file.
# Declare variables to be passed into your templates.

sofastack:
  access: # 用户AK 必填
  secret: # 用户SK 必填
  app: # 所属应用 必填
```

- Charts.yaml：描述产品模板的元数据信息，包含模板名称、模板版本、应用版本等。

> 说明
> 除 type 的值固定为 application 不可以修改，其他参数值可自定义。

```
apiVersion: v2
appVersion: 0.1.0
description: Demonstration of using Helm with SOFAStack CAFE & OAM.
name: sofaboot-helloworld
type: application
version: 0.1.0
```

模板参数说明

模板的参数说明，请参考示例模板中的注释内容，您可以根据实际情况去修改。

sofaboot-helloworld 模板

Component

```
apiVersion: core.oam.dev/v1alpha2
kind:Component# 组件类型：必须为Component
metadata:
  name:"{{ .Release.Name }}-component"# 组件名称：用于在appconfig中component[].componentName处选择组件
spec:
# 参数列表：包含组件所有可配置的参数，可在appconfig中component[].parameterValues处通过name和value进行配置
  parameters:
- name: imageName     # 参数名称：每个组件内必须唯一
      required:false# 在appconfig中配置该组件时是否必须提供value
      fieldPaths:# workload中的对应字段会被该参数覆盖
-"/spec/containers/0/image"
- name: sofa_instanceid
      required:true
      fieldPaths:
-"/spec/containers/0/env/2/value"
- name: sofa_anvip_endpoint
      required:true
```

```
      required:true
      fieldPaths:
-"/spec/containers/0/env/3/value"
- name: sofa_ak
      required:true
      fieldPaths:
-"/spec/containers/0/env/4/value"
- name: sofa_sk
      required:true
      fieldPaths:
-"/spec/containers/0/env/5/value"
  workload:
    apiVersion: oam.cafe.cloud.alipay.com/v1alpha1
    kind:CafeContainerizedworkload# 现有工作负载类型：CafeContainerizedworkload
    spec:
      containers:
- name: sofaboot-helloworld
          image: registry-vpc.cn-shanghai.aliyuncs.com/sofastackcafe/sofaboot-helloworld:v0.0.
1
          resources:
            limits:
              memory:"700Mi"
              cpu:"700m"
            requests:
              memory:"500Mi"
              cpu:"500m"
          ports:
- containerPort:8080
              protocol: TCP
- containerPort:8341
              protocol: TCP
          env:
- name: run.mode
              value: NORMAL
- name: logging.path
              value:/home/admin/logs
- name: SOFA_INSTANCE_ID
              value:""
- name: SOFA_ANTVIP_ENDPOINT
              value:""
- name: SOFA_ACCESS_KEY
              value:""
- name: SOFA_SECRET_KEY
              value:""
```

Application Configuration

```
apiVersion: core.oam.dev/v1alpha2
kind:ApplicationConfiguration
metadata:
    name:"{{ .Release.Name }}"# 安装模板时输入的发布名称
    annotations:
        app.kubernetes.io/name:"{{ .Values.sofastack.app}}"# 所属应用名（必填）
        app.kubernetes.io/description:"{{ .Release.Name }} from OAM Chart"
        cafe.sofastack.io/loginName:"{{ .Values.sofastack.loginName }}"# 负责人/操作人 登录名（非必填）
        cafe.sofastack.io/userId:"{{ .Values.sofastack.userId }}"# 负责人/操作人 用户ID（非必填）
spec:
    components:
- componentName:"{{ .Release.Name }}-component"# 组件名称：应与组件.metadata.name一致
         parameterValues:# 覆盖组件.spec.parameters相应参数
- name: sofa_instanceid
           value:"{{ .Values.sofastack.instanceid }}"
- name: sofa_anvip_endpoint
           value:"{{ .Values.sofastack.endpoint }}"
- name: sofa_ak
           value:"{{ .Values.sofastack.access }}"
- name: sofa_sk
           value:"{{ .Values.sofastack.secret }}"
         traits:
- trait:
                  apiVersion: core.oam.dev/v1alpha2
                  kind:Strategy# 现有运维特征类型：Strategy、CafeService
                  metadata:
                      annotations:
                          groupStrategy: QUICK        # 分组策略：QUICK（快速分组）、EACH_ONE（每个POD一组）或ALL_ONE（共分一组）
                          groupCount:1# 每组Pod数
                          enableSofaMesh:false# enableSofaMesh：是否开启 Mesh，true 或 false（默认）
                  spec:
                      replicas:1# Pod 副本数
                      strategy:
                          upgradeType:Beta# beta分组
                      needWaitingForConfirm:true# 批量发布时，是否每批都需要等待用户确认
```

sofaboot-helloworld-service 模板 Component

```
apiVersion: core.oam.dev/v1alpha2
kind:Component# 组件类型：必须为Component
metadata:
  name:"{{ .Release.Name }}-component"# 组件名称：用于在appconfig中component[].componentName处选择组件
spec:
# 参数列表：包含组件所有可配置的参数，可在appconfig中component[].parameterValues处通过name和value进行配置
  parameters:
- name: imageName     # 参数名称：每个组件内必须唯一
      required:false# 在appconfig中配置该组件时是否必须提供value
      fieldPaths:# workload中的对应字段会被该参数覆盖
-"/spec/containers/0/image"
```

```
- name: sofa_instanceid
      required:true
      fieldPaths:
-"/spec/containers/0/env/2/value"
- name: sofa_anvip_endpoint
      required:true
      fieldPaths:
-"/spec/containers/0/env/3/value"
- name: sofa_ak
      required:true
      fieldPaths:
-"/spec/containers/0/env/4/value"
- name: sofa_sk
      required:true
      fieldPaths:
-"/spec/containers/0/env/5/value"
  workload:
    apiVersion: oam.cafe.cloud.alipay.com/v1alpha1
    kind:CafeContainerizedworkload# 现有工作负载类型：CafeContainerizedworkload
    spec:
      containers:
- name: sofaboot-helloworld
          image: registry-vpc.cn-shanghai.aliyuncs.com/sofastackcafe/sofaboot-helloworld:v0.0.
1
          resources:
            limits:
              memory:"700Mi"
              cpu:"700m"
            requests:
              memory:"500Mi"
              cpu:"500m"
          ports:
- containerPort:8080
              protocol: TCP
- containerPort:8341
              protocol: TCP
          env:
- name: run.mode
              value: NORMAL
- name: logging.path
              value:/home/admin/logs
- name: SOFA_INSTANCE_ID
              value:""
- name: SOFA_ANTVIP_ENDPOINT
              value:""
- name: SOFA_ACCESS_KEY
              value:""
- name: SOFA_SECRET_KEY
              value:""
```

Application Configuration

```
apiVersion: core.oam.dev/v1alpha2
kind:ApplicationConfiguration
metadata:
```

```
    name:"{{ .Release.Name }}"
    annotations:
        app.kubernetes.io/name:"{{ .Values.sofastack.app}}"# 所属应用名（必填）
        app.kubernetes.io/description:"{{ .Release.Name }} from OAM Chart"
        cafe.sofastack.io/loginName:"{{ .Values.sofastack.loginName }}"# 负责人/操作人 登录名（必
填）
        cafe.sofastack.io/userId:"{{ .Values.sofastack.userId }}"
spec:
    components:
- componentName:"{{ .Release.Name }}-component"# 组件名称：应与组件.metadata.name一致
          parameterValues:# 覆盖组件.spec.parameters相应参数
- name: sofa_instanceid
            value:"{{ .Values.sofastack.instanceid }}"
- name: sofa_anvip_endpoint
            value:"{{ .Values.sofastack.endpoint }}"
- name: sofa_ak
            value:"{{ .Values.sofastack.access }}"
- name: sofa_sk
            value:"{{ .Values.sofastack.secret }}"
          traits:
- trait:
                    apiVersion: core.oam.dev/v1alpha2
                    kind:Strategy# 现有运维特征类型：Strategy、CafeService
                    metadata:
                        annotations:
                            groupStrategy: EACH_ONE     # 分组策略：QUICK（快速分组）、EACH_ONE（
每个POD一组）或ALL_ONE（共分一组）
                            groupCount:2# BatchSize 每组pod数
# enableSofaMesh: true        # enableSofaMesh：true或false（默认）
                    spec:
                        replicas:1# pod总数
                        strategy:
                            batchSize:1# 升级时每批的pod更新数
                            upgradeType:Beta# beta分组
                        needWaitingForConfirm:true# 批量发布时，是否每批都需要等待用户确认
                        topology:
                            unitType:Cell
                            values:# 部署单元
-CellA
# podSetType: InPlaceSet          # 升级策略：InPlaceSet（原地）或ReplicaSet（替换，默认）
- trait:
                apiVersion: core.oam.dev/v1alpha2
                kind:CafeService
                metadata:
                    name: sofaboot-helloworld-service
                    annotations:
                        service.beta.kubernetes.io/antcloud-loadbalancer-address-type: INTER
NET                         # 负载均衡类型（非集群内访问必填）：INTRANET（内网）或INTERNET（外网）
                        service.beta.kubernetes.io/antcloud-loadbalancer-internet-charge-typ
e: PAY_BY_TRAFFIC
                        service.beta.kubernetes.io/antcloud-loadbalancer-listener-healthchec
k:8080:ON;8341:ON
                        service.beta.kubernetes.io/antcloud-loadbalancer-listener-healthchec
k-interval:"8080:2;8341:2"# 健康检查间隔时间（非集群内访问必填）：针对每个端口要单独指定间隔时间，如834
1:3,以英文分号分隔（阿里云默认：45；金融云默认：5）
                        service.beta.kubernetes.io/antcloud-loadbalancer-listener-healthchec
```

```
k-timeout:"8080:5;8341:5"# 健康检查超时时间（非集群内访问必填）：针对每个端口要单独指定超时时间，如8341:3，以英文分号分隔
                          service.beta.kubernetes.io/antcloud-loadbalancer-listener-scheduler:8080:WRR;8341:WRR      # 转发规则（非集群内访问必填）：WRR或WLC
                          service.beta.kubernetes.io/antcloud-loadbalancer-port-protocol:8080:TCP;8341:TCP
                          service.beta.kubernetes.io/antcloud-loadbalancer-lb-name: sofaboot # 负载均衡名称（非集群内访问必填）
# service.beta.kubernetes.io/antcloud-loadbalancer-listener-healthcheck-uri: 8080:/actuator/readiness;8341:/actuator/readiness     # 健康检查路径（非集群内访问且协议类型为http/https必填）：针对每个端口单独设置检查路径，以英文分号分隔
# service.beta.kubernetes.io/antcloud-loadbalancer-listener-healthcheck-domain: 8080:www.alipay.net       # 健康检查域名（可选）
                  spec:
                      type:'LoadBalancer'# 访问方式："ClusterIP"（集群内访问）、"LoadBalancer"（VPC内网访问）、'LoadBalancer'（公网访问）
                      ports:
- protocol: TCP                 # 协议类型（必填）：TCP、UDP、HTTP或HTTPS
                        port:8080# 前端端口
                        targetPort:8080# 后端端口
# name: helloworld-conn         # 服务名称
- protocol: TCP
                        port:8341
                        targetPort:8341
```

8.9.5. OAM 产品模板

AKS 应用商店不仅提供了基于 Kubernetes Helm 标准的应用模板，还提供了符合 OAM 模型规范的应用组件（Components）及运维策略（Traits）模板。本文介绍如何在 AKS 控制台快速添加与发布 OAM 产品模板。

添加 OAM 产品模板

1. 登录容器应用服务控制台，在左侧导航栏单击 **应用商店 > OAM 产品模板**。
2. 单击 **添加产品模板**，在新页面中，配置以下模板信息：
   - **OAM 产品模板名称**：必填。根据业务需求，自定义模板名称。支持小写字母、数字、连字符，且必须以字母开头、以字母或数字结尾。
   - **组件编排**：可选。用于定义应用所包含的各组件模块及组件对应的运维策略。
     a. 单击 **添加组件**，在右侧窗口中，配置以下组件信息后，单击 **提交**。
        - 组件名称：必填。输入组件名称，默认为 helloworld。
        - 工作负载类型：必填。目前仅支持 CafeContainerizedworkload。
        - 容器信息：必填。自定义参数格式为 `{{ .Values.$componentName.imageUrl }}`。其中，`$componentName` 必须与填写的组件名称一致，`imageUrl` 为自定义的参数。
        - 参数：可选。配置环境、身份鉴权等参数，如 accesskey、secretkey、endpoint 以及 instanceid，参数根据用户需求可自行扩展。

     b. 组件添加成功后，单击 **添加运维特征**，选择运维特征（CafeService/Strategy）并配置容器信息，单击 **提交**。
   - **版本号**：必填。为该产品模板定义一个版本号，如 `v1.0.1`，方便后期进行版本维护。
   - **描述**：可选。输入该该产品模板的备注信息。

3. 单击 **提交**，模板创建成功。
4. 进入该产品模板的详情页，单击 **版本生效**。

> 说明
> 版本生效后，无法再修改当前版本的模板信息。您可以再以该版本为基准添加一个新的版本。

发布 OAM 产品模板

1. 在 OAM 产品模板的详情页，选择需要发布的版本，单击 **创建发布**。
2. 在右侧窗口中，根据提示，选择或输入相应的模板发布信息，包括工作空间、命名空间、所属应用以及具体的发布参数。

> 说明
> 请确保发布参数正确，发布时所使用的参数以设置的发布参数为准。

3. 单击 **创建发布**，模板进入创建发布状态。创建成功后会生成发布单，单击 **查看发布单** 可以跳转至发布单页面。

查看 OAM 产品发布

在 **OAM 产品发布** 页面，您可以查看到所有的 OAM 产品模板的发布信息。

1. 在容器应用服务控制台，左侧导航栏单击 **应用商店 > OAM 产品发布**。
2. 选择待查看的 OAM 产品模板发布的命名空间。
3. 在发布的产品模板列表中，找到目标模板，即可获取模板相关信息，或进行以下操作：
   - 可直接查看该模板的版本号、发布状态以及最近一次的操作时间。
   - 单击发布的产品模板名称，进入该模板的详情页。
   - 单击操作列的 **删除**，确认后即可直接删除该发布信息。
   - 对于发布成功的产品模板，单击操作列的 **查看发布单**，可跳转至 **应用发布 > 发布单** 查看详细的发布单信息。

8.10. 工作负载

8.10.1. 部署（Deployments）

部署（Deployments）即运行中始终不保存任何数据或状态的工作负载，例如 nginx。您可以通过 YAML 方式创建部署（Deployments）。完成创建后，还可以对其执行监控、删除等操作。

> 说明
> Deployments 是 kubernetes 原生概念。更多详情请参见 Kubernetes 官网文档。

创建部署（Deployments）

本部分介绍如何通过 YAML 文件创建一个部署（Deployments）。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载 > 部署（Deployments）**。
2. 在 **部署（Deployments）** 页面，单击 **从 YAML 创建**。
3. 在输入框中输入 YAML 内容，单击 **提交**。
   样例如下：

```
 kind:Deployment
 metadata:
   name: deployment-example
   labels:
     k8s-app: deployment-example
namespace:default
 spec:
   replicas:1
   selector:
     matchLabels:
       k8s-app: deployment-example
template:
     metadata:
       labels:
         k8s-app: deployment-example
     spec:
       containers:
- name: nginx
         image: reg-cnsh.cloud.alipay.com/library/nginx:latest # 改成当前 region 的镜像地址
```

创建成功后， Deployments 处于 **部署中** 状态，部署成功后变为 **运行中** 状态。

更新部署（Deployments）

本部分介绍如何通过修改 YAML 文件更新一个部署（Deployments）。

前提条件

已创建 YAML。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载 > 部署（Deployments）**。
2. 在 **部署（Deployments）** 页面，单击 **操作** 栏的 **查看/编辑 YAML**。
3. 在 YAML 编辑框中修改信息，单击 **更新**。
   更新成功后，系统会自动根据新的 YAML 文件信息部署 Deployments。

删除部署（Deployments）

本部分介绍如何删除一个部署（Deployments）。

前提条件

已创建 YAML。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载 > 部署（Deployments）**。
2. 在 **部署（Deployments）** 页面，单击 **操作** 栏的 **删除**，在弹出的对话框中单击 **确定**。

监控部署（Deployments）

本部分介绍如何监控部署（Deployments）。

前提条件

已创建 YAML。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载 > 部署（Deployments）**。
2. 在 **部署（Deployments）** 页面，单击 **操作** 栏的 **监控**，系统将跳转至 云监控控制台。您可以通过该控制台，直观地查看该部署（Deployments）的各项监控指标。

> 说明
> 云监控控制台的相关操作，请参见 查看应用分组。

8.10.2. 有状态副本集（Statefulsets）

有状态副本集（Statefulsets）即运行过程中保存数据或状态的工作负载，例如 MySQL。您可以通过 YAML 方式创建有状态副本集（Statefulsets）。完成创建后，还可以对其执行监控、删除等操作。

> 说明
> Statefulsets 是 kubernetes 原生概念。更多详情请参见 Kubernetes 官网文档。

创建 Statefulsets

本部分介绍如何通过 YAML 文件创建有状态副本集（Statefulsets）。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载 > 有状态副本集（Statefulsets）**。
2. 在 **有状态副本集 (Statefulsets)** 页面，单击 **从 YAML 创建**。
3. 在输入框中输入 YAML 内容，单击 **提交**。
   样例如下：

```
 kind:StatefulSet
 metadata:
   name: stexample
namespace:default
 spec:
   replicas:1
   selector:
     matchLabels:
       k8s-app: stexample
template:
     metadata:
       labels:
         k8s-app: stexample
     spec:
       containers:
- name: nginx
         image: reg-cnsh.cloud.alipay.com/library/nginx:latest
```

创建成功后， Statefulsets 处于 **运行中** 的状态。

更新 Statefulsets

本部介绍如何通过修改 YAML 文件更新一个有状态副本集（Statefulsets）。

前提条件

已创建 YAML。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载 > 有状态副本集（Statefulsets）**。
2. 在 **有状态副本集 (Statefulsets)** 页面，单击 **操作** 栏的 **查看/编辑 YAML**。
3. 在 YAML 编辑框中修改信息，单击 **更新**。
   更新成功后，系统会自动根据新的 YAML 文件信息部署 Statefulsets。

删除 Statefulsets

本部分介绍如何删除有状态副本集（Statefulsets）。

前提条件

已创建 YAML。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载 > 有状态副本集（Statefulsets）**。
2. 在 **有状态副本集 (Statefulsets)** 页面，单击 **操作** 栏的 **删除**，在弹出的对话框中单击 **确定**。

监控 Statefulsets

本部分介绍如何监控有状态副本集（Statefulsets）。

前提条件

已创建 YAML。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载 > 有状态副本集（Statefulsets）**。
2. 在 **有状态副本集（Statefulsets）** 页面，单击 **操作** 栏的 **监控**，系统将跳转至 云监控控制台。您可以通过该控制台，直观地查看该有状态副本集（Statefulsets）的各项监控指标。

> 说明
> 云监控控制台的相关操作，请参见 查看应用分组。

8.10.3. 守护进程集（Daemonsets）

守护进程集（Daemonsets）确保全部（或者部分）节点都运行一个 Pod 实例，支持实例动态添加到新节点。您可以通过 YAML 方式创建守护进程集（Daemonsets）。完成创建后，还可以对其执行监控、删除等操作。

> 说明
> Daemonsets 是 kubernetes 原生概念。更多详情请参见 Kubernetes 官方文档。

创建 Daemonsets

本部分介绍如何通过 YAML 文件创建守护进程集（Daemonsets）。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载 > 守护进程集（Daemonsets）**。
2. 在 **守护进程集 Daemonsets** 页面，单击 **从 YAML 创建**。
3. 在输入框中输入 YAML 内容，单击 **提交**。
   样例如下：

```
 kind:DaemonSet
 metadata:
   name: daemonset-example
   labels:
     k8s-app: daemonset-example
namespace:default
 spec:
   selector:
     matchLabels:
       k8s-app: daemonset-example
template:
     metadata:
       labels:
         k8s-app: daemonset-example
     spec:
       containers:
- name: nginx
         image: reg-cnsh.cloud.alipay.com/library/nginx:latest
```

创建成功后，Daemonsets 处于 **运行中** 的状态。

更新 Daemonsets

本部分介绍如何通过修改 YAML 文件更新守护进程集（Daemonsets）。

前提条件

已创建 YAML。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载** > **守护进程集（Daemonsets）**。
2. 在 **守护进程集 Daemonsets** 页面，单击 **查看/编辑 YAML**。
3. 在 YAML 编辑框中修改信息，单击 **更新**。
   更新成功后，系统会自动根据新的 YAML 文件信息部署 Daemonsets。

删除 Daemonsets

本部分介绍如何删除守护进程集（Daemonsets）。

前提条件

已创建 YAML。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载** > **守护进程集（Daemonsets）**。
2. 在 **守护进程集 Daemonsets** 页面，单击 **操作** 栏的 **删除**，在弹出的对话框中单击 **确定**。

监控 Daemonsets

本部分介绍如何监控守护进程集（Daemonsets）。

前提条件

已创建 YAML。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载** > **守护进程集（Daemonsets）**。
2. 在 **守护进程集（Daemonsets）** 页面，单击 **操作** 栏的 **监控**，系统将跳转至 云监控控制台。您可以通过该控制台，直观地查看该守护进程集（Daemonsets）的各项监控指标。

> 说明
> 云监控控制台的相关操作，请参见 查看应用分组。

8.10.4. 容器组（Pods）

容器组（Pods）是 Kubernetes 中最小可部署单元，代表 Kubernetes 中一个独立的应用程序运行实例，该实例可能由单个容器或者多个紧耦合在一起的容器组成。

> 说明
> Pods 是 kubernetes 原生概念。更多详情请参见 Kubernetes 官方文档。

创建 Pods

本部分介绍如何通过 YAML 文件创建一个容器组（Pods）。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载 > 容器组（Pods）**。
2. 在 **容器组 Pods** 页面，单击 **从 YAML 创建**。
3. 在输入框中输入 YAML 内容，单击 **提交**。
   样例如下：

```
 kind:Pod
 metadata:
   name: nginx
   labels:
     k8s-app: nginx
namespace:default
 spec:
   containers:
- name: nginx
      image: reg-cnsh.cloud.alipay.com/library/nginx:latest
```

创建成功后， Pods 处于 **运行中** 的状态。

编辑 Pods

本部分介绍如何通过修改 YAML 文件更新一个容器组（Pods）。

前提条件

已创建 YAML。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载 > 容器组（Pods）**。
2. 在 **容器组 Pods** 页面，单击 **查看/编辑 YAML**。
3. 在 YAML 编辑框中修改信息，单击 **更新**。
   更新成功后，系统会自动根据新的 YAML 文件信息部署 Pods。

删除 Pods

本部分介绍如何删除一个容器组（Pods）。

前提条件

已创建 YAML。

操作步骤

1. 登录容器应用服务控制台，在左侧导航栏单击 **工作负载 > 容器组（Pods）**。

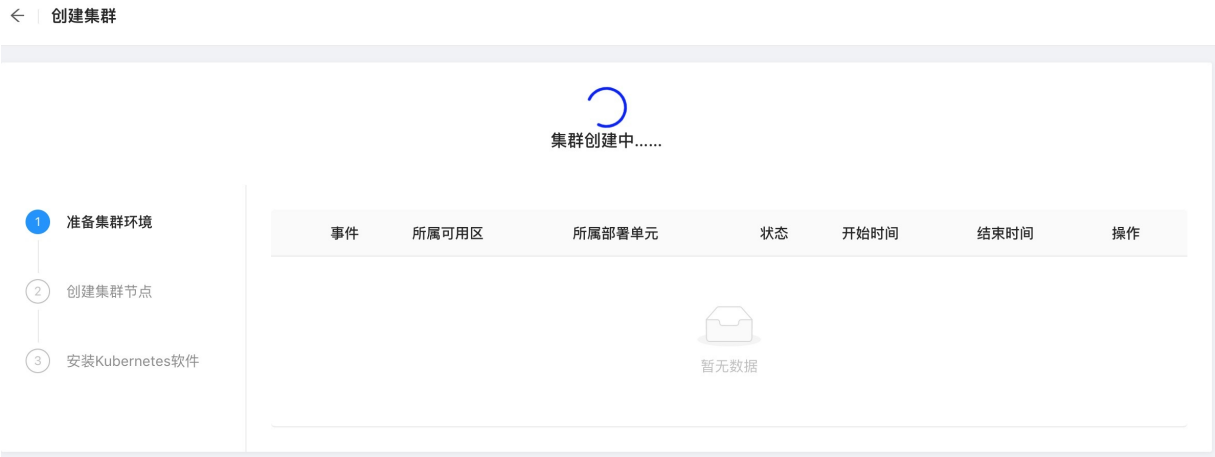
2. 在 **容器组 Pods** 页面，单击 **操作** 栏的 **删除**，在弹出的对话框中单击 **确定**。

8.11. 常见问题

准备集群环境失败，一直在集群创建中

现象：创建集群第一步准备集群环境失败，一直在集群创建中，无任何步骤展示。

解决办法：这种情况不会创建任何资源。等待 2 分钟无响应后，返回重新创建集群即可。



创建 ECS 失败或扩容节点创建 ECS 失败

解决办法：跳转到资源管理平台，单击 **任务记录** 菜单栏，可以查看任务记录列表。找到失败的记录，可以查看失败原因。常见原因有 ECS 规格在阿里云无资源、账号欠费等，可以跳转到 AKS **节点列表** 页面，重新扩容节点。



一直卡在安装 Kubernetes 软件，无响应

解决办法：运维单初始化失败导致。该情况下 ECS 节点已经建出来了，但是安装软件失败。需要到跳转到 AKS **节点列表** 页面，删除安装失败的节点，并重新扩容节点即可。



安装 Kubernetes 软件某个步骤失败

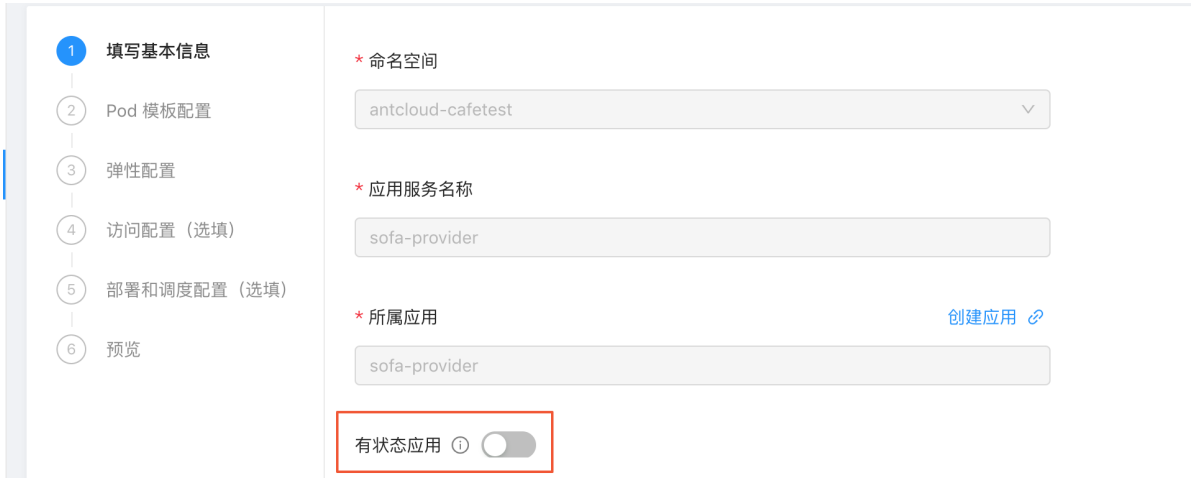
解决办法：单击查看节点安装详情，单击重试节点执行任务即可。



通过镜像发布服务，容器地址会改变

在以下两种情况下，容器地址改变属于正常现象：

- 将无状态应用修改成 **有状态应用**，如下图所示：



- 升级模式选择了 **替换升级**，如下图所示：



解决办法：如果不是以上两种情况，则需要从后台查询一下默认参数配置是否正确。

8.12. 最佳实践

8.12.1. 容器镜像服务使用规范

AKS 集成了 阿里云容器镜像服务（ACR），为您提供更好的镜像仓库服务，便于您管理私有镜像。本文介绍 ACR 使用过程中的一些实践经验与规范建议，帮助您更好地结合使用 AKS 与 ACR。

规格选择

容器镜像服务 ACR 提供了 **默认实例版** 与 **企业版** 两种规格，二者在功能规格、应用场景等均有所不同。详见 规格说明。

- **默认实例版** 提供基础的容器镜像服务，包括安全的应用镜像托管能力、精确的镜像安全扫描功能、稳定的国内外镜像构建服务以及便捷的镜像授权功能，从而方便用户进行镜像全生命周期管理。
- **企业版** 是企业级云原生应用制品管理平台，提供容器镜像、Helm Chart，符合 OCI 规范制品的生命周期管理；支持大规模、多地域、多场景下应用制品的高效分发；与容器服务 ACK 无缝集成，帮助企业降低交付复杂度。

因此，建议根据业务实际需求，如需要的命名空间、仓库数量等，选择合适的服务规格。推荐选择 **容器镜像服务 ACR 企业版**。

资源定义与划分

命名空间

ACR 通过配置命名空间，有效管理各空间下的仓库集合，包括仓库权限和仓库属性。在创建命名空间时，建议按照如下标准：

- 将一个公司或组织的仓库集中在一个命名空间下面。
- 可使用公司名称、团队或组织名称作为命名空间，如 aliyun、alibaba、misaka-team 等。
- 由于命名空间数量是有配额限制的，建议根据团队、组织或者系统的数量，合理划分命名空间。如果需要的命名空间较多，建议购买多个企业版实例。

仓库

镜像命名建议按照如下蚂蚁的统一规范，且一个应用一个镜像库。

- 镜像命名必须以 `reg.docker.alibaba-inc.com/namespace/` 作为前缀。
- 镜像命名建议：

  ```
  reg.docker.alibaba-inc.com/${yournamespace}/alios7u2-{appName}:[{envName}-]{YYYY-mm-dd}-{seq}[.{git commit id}]
  ```

  。其中，线上正式镜像 `{envName}` 留空。

此外，镜像变更后，需修改镜像 ID，然后 push 至镜像中心，不允许相同镜像 ID 重复上传不同镜像。

仓库权限配置

默认情况下，主账号对自己的资源拥有完整的操作权限。主账号可以为不同的子账号授予访问镜像仓库资源的不同权限。权限配置建议如下：

1. 除云管理员之外，其他成员没有命名空间的操作权限。
2. 命名空间管理员具备仓库级别的全部权限，命名空间普通成员具备只读和 push/pull 镜像权限。
3. 非应用/仓库所属团队人员，但是公司的相关开发和运维人员，只具备只读权限，即 `AliyunContainerRegistryReadOnlyAccess` 。
4. 可根据命名空间、仓库名或前缀进行资源筛选，继而给各 RAM 子账号授权各资源的不同权限。

更多权限配置细节，可参考 配置仓库的 RAM 访问控制。

8.12.2. 容器网络与节点规划

本文介绍创建集群时如何合理地选择与规划网络配置以及节点规格等。

网络插件

新版 AKS 在对接 ACK 后，创建集群时，需要在网络配置中选择启动用的网络插件，支持 Flannel 和 Terway 网络插件。

- Flannel：使用的是简单稳定的社区的 Flannel CNI 插件，配合阿里云 VPC 的高速网络，能给集群高性能和稳定的容器网络体验，但功能偏简单，支持的特性少，如不支持基于 Kubernetes 标准的网络策略。
- Terway：ACK 版自研的网络插件，将阿里云的弹性网卡分配给容器，支持基于 Kubernetes 标准的网络策略来定义容器间的访问策略，支持对单个容器做带宽的限流。

因此，建议根据业务实际需求，如不需要使用网络策略，可选择 Flannel 网络插件。其他情况，则建议选择 Terway 网络插件，详见 如何使用 Terway 网络插件。

Terway Pod 地址

Terway Pod 地址空间大小，建议根据 Pod 数量定义。另外，地址不要与 Node 地址以及其他内网地址冲突。

节点规格

从实践经验来看，节点与 Pod 的比例一般 1:4 到 1:6 都有。以通用 Java 应用 4c 为例，节点可能 16c - 64c 都有。但是，如果您选择使用 Terway 网络插件，建议选择较高规格和较新类型的 ECS 机型，需要考虑到 ECS Node 上 ENI 弹性网卡的数量。单节点所支持的最大 Pod 数取决于该节点的弹性网卡（ENI）数。

- 共享 ENI 支持的最大 Pod 数 =（ENI 数 - 1）×单 ENI 的私有 IP 数

- 独占 ENI 支持的最大 Pod 数 = ENI 数 - 1

大多数情况下，节点上可支持的 Pod 数量，通常会远超过 20，所以问题不大。
总体而言，在选择 Worker 节点规格时，主要考虑以下两个因素：

- 确定整个集群的日常使用的总核数以及可用度的容忍度。例如：集群总的核数有 160 核，可以容忍 10% 的错误。那么最小选择 10 台 16 核 ECS，并且高峰运行的负荷不要超过 160*90%=144 核。如果容忍度是 20%，那么最小选择 5 台 32 核 ECS，并且高峰运行的负荷不要超过 160*80%=128 核。这样就算有一台 ECS 出现故障，剩余 ECS 仍可以支持现有业务正常运行。
- 确定 CPU：Memory 比例。对于使用内存比较多的应用例如 Java 类应用，建议考虑使用 1:8 的机型。

使用大规格 ECS 的优势如下：

- 网络带宽大，对于大带宽类的应用，资源利用率高。
- 容器在一台 ECS 内建立通信的比例增大，减少网络传输。
- 拉取镜像的效率更高。因为镜像只需要拉取一次就可以被多个容器使用。而对于小规格的 ECS 拉取镜像的次数就会增多，若需要联动 ECS 伸缩集群，则需要花费更多的时间，反而达不到立即响应的目的。

更多实例规格的选择建议，可参考 ECS 选型 与 实例规格族。

8.12.3. 选择存储卷类型

ACK 集群支持通过 Pod 自动绑定使用阿里云云盘、NAS、OSS 存储服务。在实际业务场景中，您可以根据需求，选择适合的存储卷类型。本文将分别介绍阿里云云盘、NAS 以及 OSS 存储服务及其应用场景。

阿里云云盘

阿里云云盘 是阿里云为云服务器 ECS 提供的数据块级别的块存储产品，具有低时延、高性能、持久性、高可靠等特点。

使用说明

- 云盘为非共享存储，一个云盘只能同时被一个 Pod 挂载。
- 集群中只有与云盘在同一个可用区（Zone）的节点才可以挂载云盘。
- 使用云盘存储卷前，需要先申请一个云盘。详见 创建云盘。

应用场景

- 对磁盘 I/O 要求高的应用，且没有共享数据的需求，如 MySQL、Redis 等数据存储服务。
- 高速写日志。
- 持久化存储数据，不因 Pod 生命周期的结束而消失。

阿里云 NAS

阿里云文件存储 NAS（Apsara File Storage）是面向云服务器 ECS 实例、容器服务等计算节点的文件存储服务，是一种可共享访问、弹性扩展、高可靠、高性能的分布式文件系统。它的读写性能相较于对象存储 OSS 而言较高。

使用说明

- NAS 为共享存储，支持同时被多个 Pod 挂载，即一个 PVC 可以同时被多个 Pod 使用。
- 使用 NAS 数据卷前，您需要在 NAS 管理控制台上创建文件系统，并在文件系统中添加挂载点。

应用场景

- 对磁盘 I/O 要求较高的应用。
- 可实现跨主机文件共享，例如：可作为文件服务器。

阿里云 OSS

阿里云对象存储 OSS（Object Storage Service）是阿里云提供的一种海量、安全、低成本、高持久的云存储服务，适合存放任意类型的文件。

使用说明

- OSS 为共享存储，可以同时为多个 Pod 提供共享存储服务。
- 使用 OSS 存储卷前，您需要先在 OSS 管理控制台上创建 Bucket。详见 创建存储空间。

应用场景

- 对磁盘 I/O 要求低。
- 配置文件、图片、小视频等共享业务。

相关链接

- 有状态服务-使用云盘静态存储卷实现持久化存储
- 有状态服务-动态云盘使用最佳实践
- 有状态服务-NAS使用最佳实践
- 有状态服务-OSS存储使用最佳实践

9.单元化应用服务

9.1. 什么是单元化应用服务

9.1.1. 单元化介绍

在当今的互联网业内，很多大型互联网系统，比如淘宝、支付宝、网商银行等，都已经实现了单元化架构，并从中获益匪浅，更多企业正加入其中。为什么要做单元化，单元化架构能给系统带来什么样的能力。本文结合蚂蚁集团支付宝系统的单元化架构建设实践，阐释单元化的原理与实现。

单点瓶颈

任何一个互联网系统，不论是支付宝、淘宝，还是 Google、Facebook，当发展到一定规模时，都会不可避免地触及到单点瓶颈。这里所说的“单点”，在系统的不同发展阶段表现不同。



服务器和应用单点

在系统发展初期，服务器和应用单点最先成为瓶颈，解决的方法也很简单，加机器、拆应用。

数据库单点

紧接着数据库单点，解决起来就开始不那么容易了，典型的做法是先垂直拆分，再水平拆分，在这个过程中需要解决多数据源、数据分片、透明访问等问题。

机房单点

当应用越来越多、服务器越来越多、数据库越来越多，单个机房的容量开始捉襟见肘，装不下这么多服务器。而且业务量的增长也让系统单机房运行的风险激增，一旦发生机房断电或是其他灾害导致机房故障，就会让整个系统完全瘫痪。机房不能放在一个篮子里，必须让系统在两个或更多 IDC 内运行。

多机房部署通常有以下两种模式：

- **垂直模式**：将全站应用、数据库等划分为几个部分，分别放在不同的机房中，完成一个业务可能需要不同机房中的不同应用提供服务。这就相当于在逻辑层面将一个物理机房放大了，机房容量不足的问题得以解决。
- **水平模式**：每个机房中部署的应用都是相同的，每个机房都有完成全站所有业务的能力，在运行时每个机房都只承担整站的一部分业务流量。

从具体实现角度来看，垂直模式更容易些，能够突破机房容量瓶颈，但不具备容灾能力。而容灾，是一个发展到一定规模的互联网系统的重要诉求，更是金融场景和金融业务必不可少的基础能力，因此在具体实践中，大多数大型系统都采用多机房水平扩展的模式。

单地部署

上面所说的容灾问题，对于一个有着亿级用户的系统，或一个数据系统来说尤为重要，仅仅是机房级容灾还不够，必须要考虑部署地容灾，也就是说不能把所有机房部署在地理上临近的地区，以防发生地震、海啸、核爆等剧烈灾害而导致系统被毁灭性破坏。这类系统的典型代表是银行、第三方支付等金融类系统，比如银行就对机房部署有着经典的“两地三中心”要求。于是，单地部署开始成为制约业务发展的瓶颈。

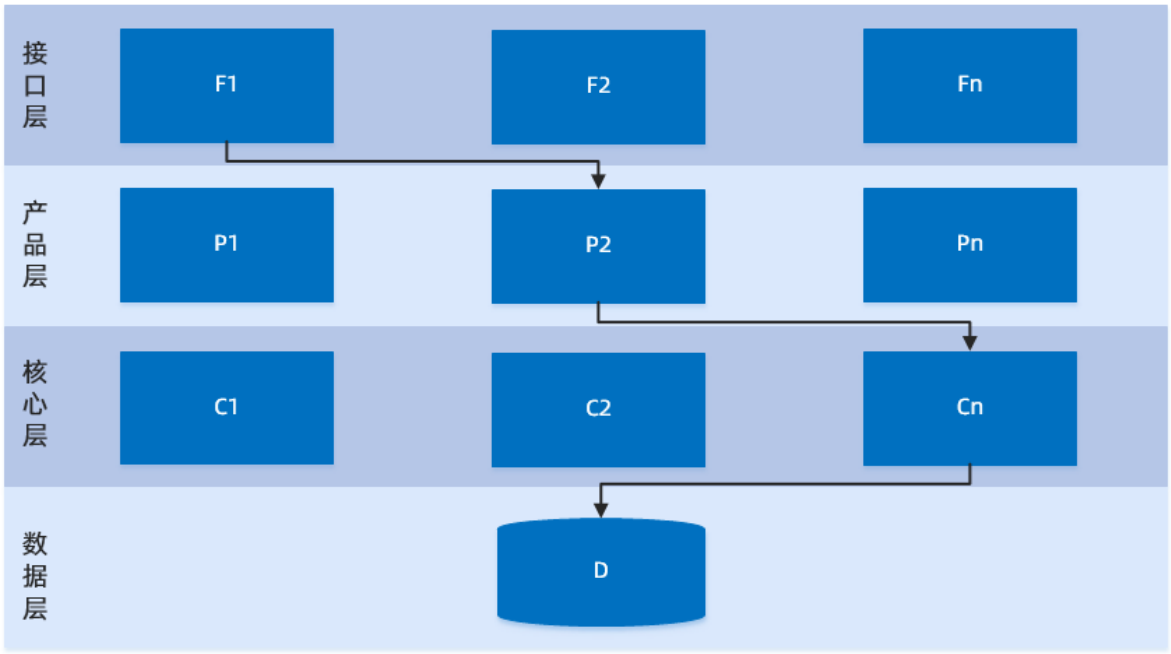
将系统的某个部分部署到距离较远的另外一个地区（城市）的能力，是大型互联网系统走向成熟的标志。多地部署本质上和多机房部署是一样的，面临的问题也几乎一样，除了距离问题。表象上是距离，下面隐藏着的是延时，更远的距离意味着更长的延时，个位数毫秒的延时不会给系统带来什么麻烦，但一旦这个数值变成几十毫秒，量变就引发了质变，很多业务开始不能忍受和忽略延时带来的影响。

单元化

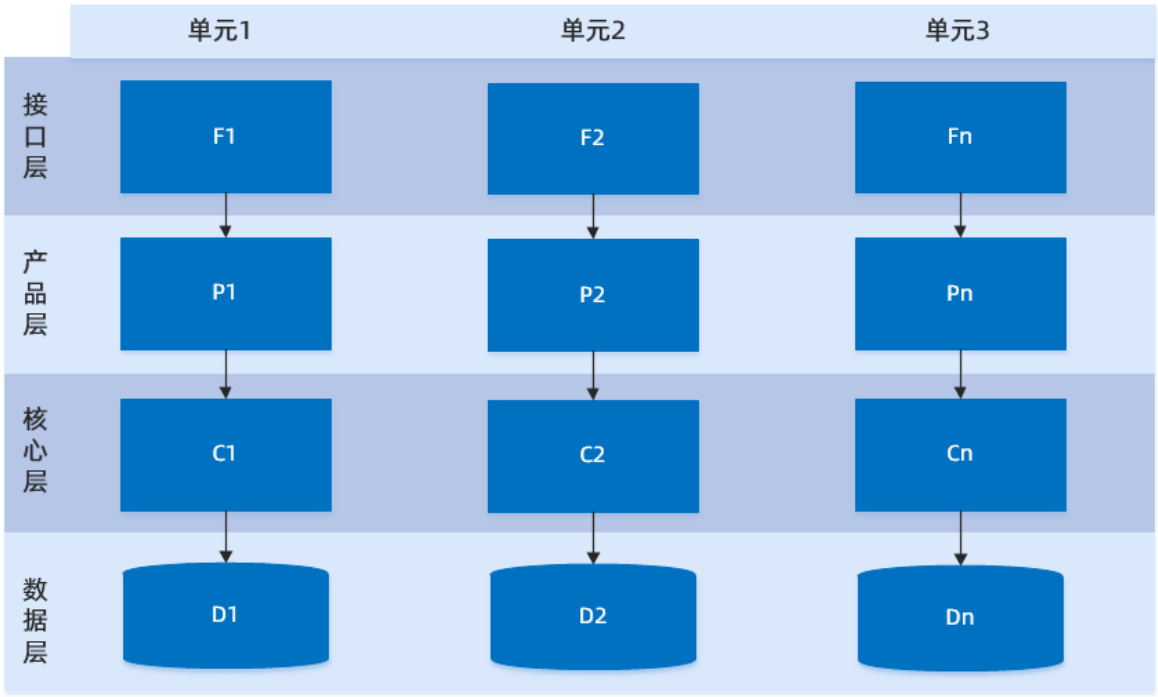
多地多机房部署，是互联网系统的必然发展方向，一个系统要走到这一步，也就必然要解决上面提到的问题：流量调配、数据拆分、延时等。业界有很多技术方案可以用来解决这些问题，而承载这些方案的，是一个部署架构。尽管可采用的部署架构不止一个，但不论是纯理论研究，还是一些先行系统的架构实践，都把“单元化部署”推崇为最佳方案。

单元（即单元化应用服务产品层的部署单元），是指一个能完成所有业务操作的自包含集合，在这个集合中包含了所有业务所需的所有服务，以及分配给这个单元的数据。单元化架构就是将单元作为部署的基本单位，在全站所有机房中部署多个单元，每个机房内单元数目不固定，任一单元均部署系统所需的全部应用，数据则是全量数据按照某种维度划分后的一部分。

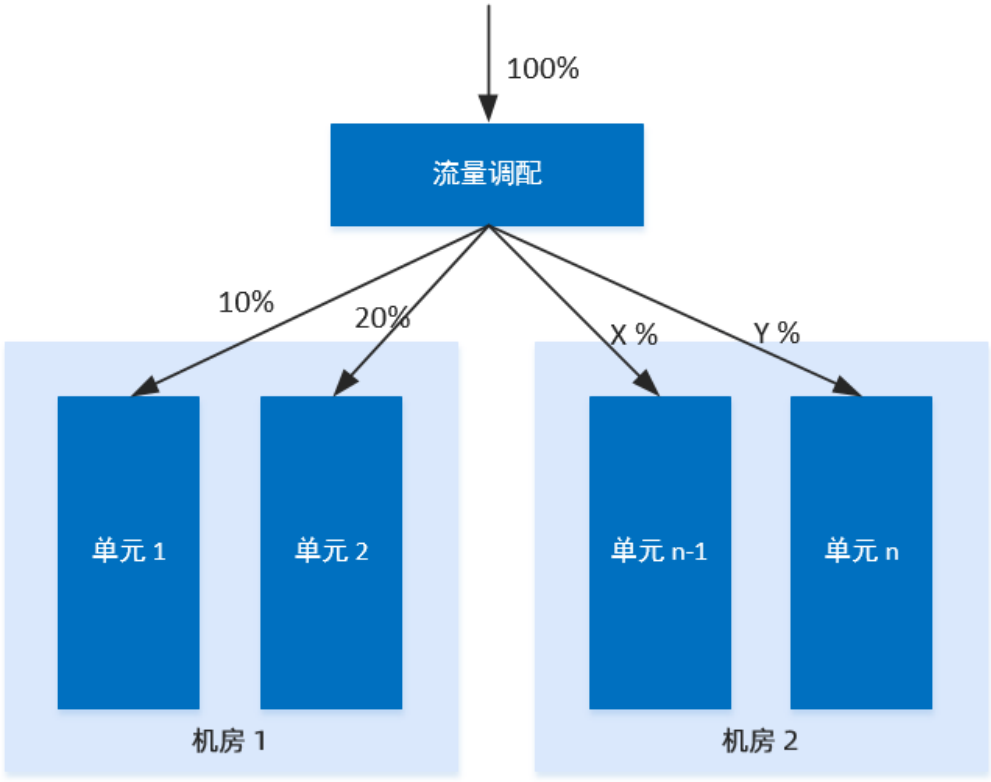
传统意义上的 SOA 化（服务化）架构，服务是分层的，每层的节点数量不尽相同，上层调用下层时，随机选择节点。



单元化架构下，服务仍然是分层的，不同的是每一层中的任意一个节点都属于且仅属于某一个单元，上层调用下层时，仅会选择本单元内的节点。



一个单元，是一个五脏俱全的缩小版整站，它是全能的，因为部署了所有应用；但它不是全量的，因为只能操作一部分数据。能够单元化的系统，很容易在多机房中部署，因为可以轻松将多个单元部署在一个机房内，而将另外几个单元部署在其他机房内。通过在业务入口处设置一个流量调配器，可以调整业务流量在单元之间的比例。



从上述对单元的定义和特性描述中，可以推导出单元化架构要求系统必须具备的一项能力：数据分区，实际上正是数据分区决定了各个单元可承担的业务流量比例。数据分区（shard），即是将全局数据按照某一个维度水平划分开来，每个分区的数据内容互不重叠，这也就是数据库水平拆分所做的事情。

仅把数据分区了还不够，单元化的另外一个必要条件是，全站所有业务数据分区所用的拆分维度和拆分规则都必须一样。若是以用户分区数据，那交易、收单、微贷、支付、账务等全链路业务都应该基于用户维度拆分数据，并且采用一样的规则拆分出同样的分区数。比如，以用户 id 末 2 位作为标识，将每个业务的全量数据都划分为 100 个分区（00-99）。

有了以上两个基础，单元化才可能成为现实。把一个或几个数据分区，部署在某个单元内，这些数据分区占总量数据的比例，就是这个单元能够承担的业务流量比例。 执行数据分区时一个很重要的问题是分区维度的选择，一个好的维度，应该：

- **粒度合适**：粒度过大，会丧失流量调配的灵活性和精细度；粒度过小，会给数据的支撑资源，访问逻辑带来负担。
- **足够平均**：按这个分区维度划分后，每个部署单元的数据量应该是几乎一致的。

以用户为服务主体的系统（很多面向用户的系统，比如支付宝）通常可以按用户维度对数据分区，这是一个最佳实践。

逻辑单元

逻辑单元是单元化架构的基础，一个单元被称为一个 Zone。根据业务特点不同，您可以将系统部署在不同类型的逻辑单元中。更多信息，请参考 创建逻辑单元。

9.1.2. 概述

随着云原生技术在业界持续升温，越来越多的金融客户希望将能将云搭建在拥有云原生能力的平台之上，利用各类云原生的红利迅速实现技术转型，促进业务的敏态发展和持续创新。
单元化应用服务（LDC Hybrid Cloud，简称 LHC）定位于云原生基础设施之上，在多机房、多地域的 Kubernetes 多集群场景，提供应用管理、发布运维、流量调拨、配置同步等能力。
作为开发运维人员日常接触的 PaaS 管控层产品，帮助解决应用和逻辑单元管理、按单元的配置变配、网络流量调拨、监控元数据配置等能力。不仅满足金融场景下多活和容灾的业务需求，同时能够让基础设施享受到容器化基础设施、云原生架构的技术红利。
LHC 旨在提供从单 Kubernetes 集群向多活联邦集群演进的能力，提供具备容灾能力的同城双活、 两地三中心及更多机房级多活容灾场景。可以配合 SOFAStack 各中间件产品、OceanBase 分布式数据库，形成单元化异地多活架构解决方案。

9.1.3. 产品优势

高可用与容灾

金融行业，尤其是银行，出于监管需求，其生产环境通常是要求在自建 IDC 里，即专有云环境。这对云原生底盘 Kubernetes 的高可用和稳定性提出了极高的要求，比如在双机房情况下如何做到双活，在有异地机房时，如何保证通讯的延迟不导致上层业务的大量超时而影响到最终用户的资金、账户安全。在灾难发生时，如何减少影响面，尽可能减少受影响的用户，并缩短影响的时间。

变更管控与保障

针对金融级场景下大规模分布式系统的特点，提供了丰富的发布策略以满足不同的场景，如：分组发布、Beta 发布、灰度发布等，帮助传统架构平滑过渡，适应金融技术风险保障需求，实现大规模金融级运维场景下的容器服务落地。

- 提供发布单追溯功能，记录和升级容器应用变更和版本管理。
- 支持多应用批量发布、设置应用间发布依赖顺序。
- 提供分组分批次、可灰度、可暂停、可回滚的发布策略和控制能力。
- 容器支持按需、按资源水位、按计划弹性伸缩。

无限弹性扩展

随着金融行业思路的转变，将金融业务内嵌到个人生活服务中，必然会产生诸多高频场景。
采用分布式架构，可以让业务在出现热点后，进行拆分、扩容，以应对流量激增。但如果是由于物理资源不够，或者数据层瓶颈，仅采用应用层的分布式就不够了。如何做到从上至下有一套完整的体系，从接入层到应用层最后到数据层进行规整，可以作为一个整体单元进行快弹、快缩。

9.1.4. 产品架构

单元化应用服务，基于 Kubernetes 基础设施提供跨集群资源管理、发布运维及单元化管控能力。其纳管的每个集群，均为完整、标准的 Kubernetes 集群及相关扩展。
在集群之上，通过联邦管控平面，协调各集群资源、应用和配置，以提供应用变更管控、分组发布、镜像管理、流量调拨、元数据管理、集群资源管理等功能。用户可以通过控制台、命令行或 SDK 以标准方式对单集群及联邦管控层进行交互。
作为分布式架构的平台管控层，单元化应用服务还承担与各相关系统的对接能力。

- 与研发效能产品对接，提供持续集成与交付功能，生成 LHC 产品的发布单。
- 与各 PaaS 基础产品对接，提供权限、资源、应用元数据、工作空间、部署单元等领域模型的管理能力。
- 与监控分析平台对接，向用户提供完整集群资源和业务应用的可观测性。



9.1.5. 功能特性

跨集群应用管理

- **应用服务（FedAppService/AppInstanceGroup）**
  - 创建应用服务模板，根据部署单元（Cell）做相应的属性同步或变配。
  - 应用配置版本管理。
- **发布单管理**
  - 有效管控单个或一组应用的发布状态和策略，检查发布变更。
  - 在发布准备时，可以配置应用依赖信息，自动实现串行和并行发布策略。
  - 在发布过程中，确保前置、发布、后置任务事件完整记录，方便追踪发布历程。
- **应用服务级别监控、日志和事件配置和展现**
  - 配置跨集群的监控，并且可以聚合、显示。
  - 配置跨集群的日志收集，并且可以聚合、搜索、显示。

- 应用服务级别事件展示。

跨集群资源管理

- Service：是 Kubernetes 自带的核心资源，用于抽象向多个 Pods 访问的接口，类似于负载均衡的概念。目前支持 Cluster IP、Loadbalancer、Node Port、Headless、ExternalDNS 类型。Service 可通过 YAML 和 UI 方式进行创建和修改。
- ConfigMap：存储应用所需配置信息的资源类型，用于保存配置数据的键值对。ConfigMap 可通过 YAML 和 UI 方式进行创建和修改。
- Secret：一种用于存储工作负载所需要认证信息、密钥的敏感信息等的资源类型。Secret 可通过 YAML和 UI 方式进行创建和修改。
- 工作负载（Kubernetes NativeWorkloads）：工作负载包括 Kubernetes 原生的 Deployment、StatefulSet、DaemonSet、Job、Pod，支持工作负载配置、监控、更新、删除等生命周期管理。

多集群管理

- 集群创建与运维
- 节点扩容与展现
- 命名空间管理
- 提供对工作空间与工作空间组的管理

单元化配置

- 逻辑单元配置
- 部署单元配置

接入层管理

- 接入网关与流量管理
- 应用层流量配置

9.1.6. 应用场景

LHC 服务于云原生模式下，通过一套应用 PaaS 平台，提供统一的应用、资源管理，以及发布运维视图，实现多集群管理、跨集群应用运维发布、资源管理、流量管理。
具体来说，可细分为以下场景：

- **同城双活（active-active）**
  在同一个地域 Region，建立两个或更多可用区下的多个 Kubernetes 集群。
- **两地三中心**
  - 在同城双活的基础上，增加一个异地机房，做数据和应用备份。根据网络延时和带宽情况，可选择异地热备、温备和冷备三种方案。
  - 在异地存在延迟的情况下、业务可接受的范围内，可以选择异地热备。正常情况下访问同城双活侧，在容灾期间切至异地灾备机房，提供业务访问。
- **异地多活（Multi-region active-active）**
  数据层做分片（Sharding），不同的 AZ 可以划分为更多的逻辑单元（Logic Data Center），处理不同的数据分片。 尽量保证数据访问的链路从接入层到应用层再到数据层不会出现跨可用区的调用。这种架构下，可以做到任意数量地域的多活。
- **异构基础设施下的混合云**
  通过 Kubernetes 屏蔽底层 IaaS 的差异性，可充分利用公有云上的资源，将业务同时在专有云和公有云上进行部署，并进行统一运维管控。在该场景下，可以帮助金融客户达到以下目的：
  - 减少开发、测试资源的投入：专有云部署生产应用，公有云按需部署开发测试应用。

- 线下快速容灾需求：应国家监管需求，需要在线下部署一套环境，以应对公有云上的突发情况（例如天弘余额宝）。
- 弹性扩容：结合异地多活架构，使业务能够按需进行机房级的无限水平扩展。

9.1.7. 基础术语

中文	英文	说明
逻辑单元	Zone	一个单元被称为一个 Zone，有 3 种不同类型：RZone、GZone、CZone。单元的特点如下： • 同一个应用在每个单元中拥有独立使用的资源。 • 同一个应用的业务在不同单元中按水平方向拆分。 • 不同单元处理的业务分片不重叠。
单元化架构	-	应用层按照数据层相同的拆片维度，将整个请求链路收敛在一组服务器中，从应用层到数据层就可以组成一个封闭的单元。 数据库只需要承载本单元的应用节点的请求，大大节省了连接数。“单元”可以作为一个相对独立整体来挪动，甚至可以将部分单元部署至异地。
部署单元	Cell	部署单元（Cell），是指一个能完成所有业务操作的自包含集合，在这个集合中包含了所有业务所需的所有服务，以及分配给这个单元的数据。 单元化架构就是将单元作为部署的基本单位，在全站所有机房中部署数个单元，每个机房里的单元数目不定，任意一个单元都部署了系统所需的所有应用，数据则是全量数据按照某种维度划分后的一部分。
应用服务	Application service	该概念和经典应用服务中的应用服务概念一致。 但由于容器有其特殊性，LHC 中的应用服务会包含一些额外的元数信息，比如容器规格配置、镜像、调度策略、日志配置等。
镜像	Image	镜像是应用包，将配置和相关软件等打在一起的二进制包，并且符合 Docker Image 规范。镜像可以来自任何可被 LHC 网络访问到的镜像中心，对于私有镜像中心，需要在 LHC 中配置相应的访问信息。
构建	Build	构建用于描述从应用源代码到制作出镜像过程的配置信息，包括源代码地址、分支信息、源镜像访问信息、目标镜像信息、Dockerfile 位置信息等。
集群	Cluster	LHC 中集群用于描述您所创建的一个工作负载集群，由多个节点组成。
节点	Node	节点表示一台装了 Docker 和 Kubelet，用以运行应用负载的物理机或者虚拟机。
容器组	Pod	Kubernetes 中最小的部署及管理单元。一个 Pod 由多个相关的并且共享磁盘的容器组成。

中文	英文	说明
命名空间	Namespace	命名空间和 Kubernetes 中相应的概念保持一致，用于表示一个逻辑隔离的空间，会将 Pod、Service、ReplicaSet 等元素隔离，但通常来说，网络不隔离。
原地升级	Inplace upgrade	原地升级是指应用服务中 Pod 的更新方式。发布后 Pod 的 IP 通常和发布前无法保持一致，所在的节点也可能发生变化。该更新方式在镜像替换时不会导致 Pod 删除。
标签	Label	Kubernetes 的原生概念，用于给相应的资源打上标签，做聚合或者匹配。
污点	Taint	Kubernetes 的原生概念，用于给节点做污点标记，通常用于 Pods 的调度策略。 与之相对应的概念为：容忍（tolerance），若 Pods 上有相对应的 tolerance 标记，则可以容忍节点上的污点，并调度到该节点。
保密字典	Secret	Kubernetes 的原生概念，用于存储用户的加密内容。
应用容器	Container	应用程序所运行的隔离工作空间，通常是 Docker 容器或者 Pouch 等兼容 CRI 接口的具有隔离能力的沙箱工作空间。
工作负载	Workload	应用程序运行态的载体及其上层聚合。通常包括：Pod、Deployment、StatefulSet、DeamonSet、Job 等。
配置项	Configmap	Kubernetes 的原生概念，用于存储用户的配置信息。
存储类型	Storage Class	Kubernetes 的原生概念，通常由系统管理员定义，用于指定所支持的存储类别，不同的类别会有不同的存储 SLA、备份策略等差异性。
存储卷	Persistent Volume	Kubernetes 的原生概念，表示一个由系统管理员创建好的存储资源。
存储卷声明	Persistent Volume Claim	Kubernetes 的原生概念，一个存储卷声明绑定一个存储卷。

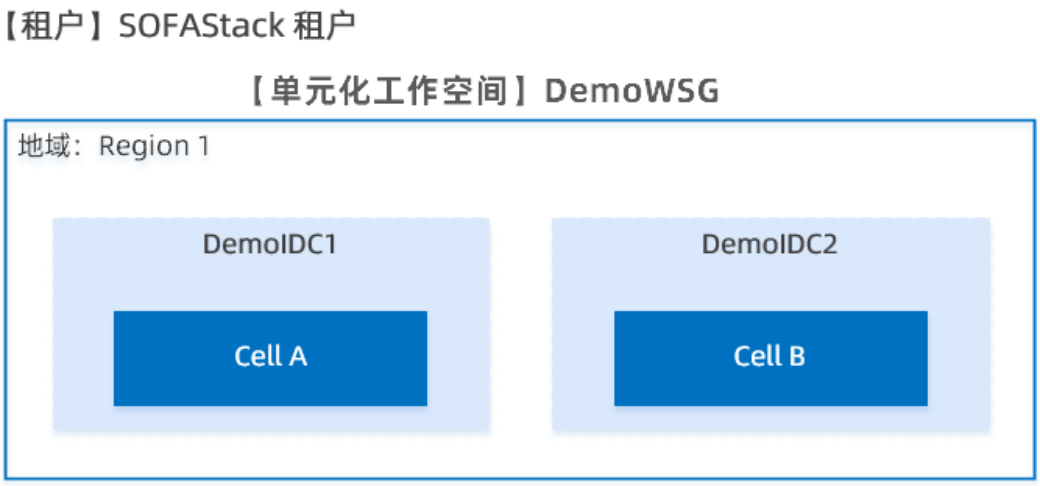
9.2. 快速入门

本文以一个同城双活架构下的双集群发布模式为例，引导您快速熟悉 LHC 的操作链路。
LHC 也支持在单元化架构下发布应用，具体内容参见 单元化入门指南。

> 说明
> 双机房场景下，推荐单元化工作空间与可用区（机房）是 1：2 的关系。

背景信息

部署架构如下图：在 SOFAStack 租户下，在同一地域内，创建一个工作空间组，关联两个工作空间，分别对应两个不同的可用区（机房），每个机房对应一个部署单元（Cell）。



步骤一：创建单元化工作空间

该步骤中您将创建一个包含双可用区的单元化工作空间。

1. 登录 LHC 控制台。
2. 在左侧导航栏，单击下方的 **全局设置**，进入工作空间列表页面。
3. 单击 **创建工作空间**，选择 **单元化工作空间** 类型，单击 **创建**。
4. 在 **创建工作空间** 页面，输入以下基本信息。
   - **工作空间标识**：2~45 个字符，以大小写字母开头，只能包含字母和数字。工作空间标识全局唯一，一经确定无法修改，例如：dev、test、prod 等。本例中输入：`DemoWSG`。
   - **工作空间名称**：1~64 个字符，工作空间的显示名称，例如：开发工作空间、测试工作空间、生产工作空间。本例中输入：`DemoWSG`。
   - **地域**：工作空间所在的地域（Region），一个工作空间一定属于某个地域。
   - **网络类型**：选择 **VPC 网络**。
   - **导入 VPC**：保持默认关闭状态。
5. 单击 **下一步**，在 **创建单元架构** 页面单击 **添加可用区**。 每个工作空间配置的可用区（Available Zone）个数无限制。本例中为工作空间配置两个可用区，为支持双机房高可用等架构做准备。

> 说明
> 系统会根据可用区配置自动为您生成单元化架构拓扑，划分好逻辑单元和部署单元。



6. 单击 **下一步**，在 **创建 VPC** 页面，输入以下配置信息：
   - **专有网络名称**：由 2~128 个英文或中文字符组成，必须以大小字母或中文开头，可包含数字、下划线（_）或连字符（-）。可以使用系统自动生成的名称 `单元化工作空间标识-vpc` 。
   - **专有网络网段**：专有网络的网段，一旦选择便无法更改，专有网络内的所有资源，如 ECS、RDS、SLB 的私网 IP 都在该网段内。可选网段如下：
     - 10.0.0.0/8
     - 172.16.0.0/12
     - 192.168.0.0/16
   - **描述**：非必填，填写专有网络描述内容。不能以 `http://` 和 `https://` 开头。
   - **交换机**：单击 **添加交换机**，在弹出的 **创建交换机** 窗口，填写以下信息，单击 **提交**。
     - **名称**：交换机名称。长度为 2~128 个字符，以英文字母或中文开头，可包含数字、下划线（_）和短横线（-）。系统自动生成 `单元化工作空间标识-vsw` 组成的默认名称。
     - **可用区**：交换机的可用区。同一 VPC 内不同可用区的交换机内网互通。您需要为每个可用区创建一台交换机。
     - **自定义网段**：默认关闭。开启后，需填写网段地址。交换机的网段可以和其所属的 VPC 网段相同或者是其 VPC 网段的子网。
     - **子网掩码**：**自定义网段** 关闭时，需分别选择子网掩码和网段地址。默认专有的网段掩码是 24 位，例如 172.31.0.0/24。范围为 16~29 位之间，可提供 65532 个地址。

- **描述**：输入交换机的描述信息。可包含 2~256 个中英文字符，不能以 `http://` 和 `https://` 开头。

7. 单击 **下一步**，在 **创建安全组** 页面，单击 **添加安全组**，在弹出的 **添加安全组** 窗口，填写以下信息，单击 **确定**。
   - **安全组名称**：长度为 2~128 个英文或中文字符，只能包含字母、数字、短横线以及 `.`，必须以字母开头。系统自动生成 `单元化工作空间标识-sg` 组成的默认名称。
   - **描述**：可包含 2~256 个中英文字符，不能以 `http://` 和 `https://` 开头。
   - **规则**：保持默认设置，接受所有的出入访问。
8. 单击 **确定** 返回 **创建安全组** 页面，单击 **提交**。

步骤二：创建集群

集群是运行工作负载的逻辑分组，包含一组云服务器资源，每台云服务器即集群中的一个节点。首次使用 LHC 时，您需要创建一个初始集群，并添加至少一个节点。

前提条件

- 您拥有 LHC 管理员权限。
- 已创建单元化工作空间。
- 首次使用 ACK 集群时，需先授权 ACK 访问云资源的权限，详情参见 首次使用容器服务 Kubernetes 版。

本例中，您将为两个可用区分别创建一个集群：`democluster1`、`democluster2`，实现同城双活的架构。

> 说明
> 同一个可用区无法创建多个集群。如当前可用区/交换机均已经有集群，请到工作空间管理页面添加新的可用区和交换机到当前工作空间。

操作步骤

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理** > **集群详情**。
2. 在集群列表页，单击 **创建集群**。
3. 在 **集群创建** 页面，系统会自动进行预检查，确保相关的产品已经开通并且租户下的账户余额大于 100 元。预检查通过后，单击 **下一步**。

   > 说明
   > 若检查失败，完成修复失败项目后，可以单击 **重新检查** 操作重新进行预检查。

4. 在 **基本配置** 页面，完成以下集群基础选项配置后，单击 **下一步**。
   - **基本信息**：
     - **专有网络**：Kubernetes 集群仅支持专有网络。您可以在已有 VPC 列表中选择所需的 VPC。
     - **集群名称**：应包含 1~63 个字符，可包含数字、汉字、英文字符或中划线（-）。本例中分别输入：`democluster1`、`democluster2`。
     - **Kubernetes 版本**：选择 Kubernetes 版本。支持 `1.16.9-aliyun.1`、`1.18.8-aliyun.1`。本例中选择 `1.18.8-aliyun.1`。

- **容器运行时**：`docker 19.03.5`。

- **网络配置**：
  - **虚拟交换机**：您可以在已有虚拟交换机列表中，根据可用区选择所需交换机。如果没有您需要的交换机，可以通过单击 **创建虚拟交换机** 进行创建，请参见 创建交换机。
  - **网络插件**：设置启用的网络插件，支持 Terway 网络插件，Terway 是阿里云容器服务自研的网络插件，将阿里云的弹性网卡分配给容器，支持 Kubernetes 的 Network Policy 来定义容器间的访问策略，支持对单个容器做带宽的限流。具体请参见 Flannel 与 Terway。
  - **Pod 虚拟交换机**：网络插件使用 Terway 时，需要为 Pod 分配 IP 的虚拟交换机。每个 Pod 虚拟交换机分别对应一个 Worker 实例的虚拟交换机。如果没有您需要的交换机，可以通过单击 **创建虚拟交换机** 进行创建，请参见 创建交换机。
  - **Service CIDR**：设置 Service CIDR。网段不能与 VPC 及 Pod 地址段重复，创建成功后不能修改。可选范围：10.0.0.0/16-24，172.16-31.0.0/16-24，192.168.0.0/16-24。
- **高级配置**：保持默认设置。更多信息参见 创建集群。

5. 在 **节点配置** 页面，完成以下 Worker 节点配置。
   - **付费类型**：支持 **按量付费** 和 **包年包月** 两种节点付费类型。选择 **包年包月** 时，需设置以下参数。
     - **购买时长**：目前支持选择 1、2、3、6 个月和 1~5 年。
     - **自动续费**：设置是否自动续费。
   - **节点数量**：创建 Worker 实例（ECS 实例）的数量。
   - **实例规格**：支持选择多个实例规格。详情请参见 实例规格族。最多可选择 10 个实例规格。

     > 说明
     > 可以选择多个规格作为备选，每个节点创建时，将从第一个规格开始尝试购买，直到创建成功，最终购买的实例规格可能随库存变化而不同。

   - **系统盘**：支持 SSD 云盘、高效云盘、ESSD 云盘。
   - **挂载磁盘**：支持 SSD 云盘、高效云盘、ESSD 云盘。
   - **操作系统**：支持 CentOS 和 Alibaba Cloud Linux 操作系统。
   - **登录密码**：设置节点的登录密码。8~30 个字符，且同时包含三项（大写字母、小写字母、数字和特殊符号）。
   - **确认密码**：确认设置的节点登录密码。
6. 配置完成后，单击 **下一步**。
7. 在集群的 **配置预览** 页面，确认配置无误后，单击 **提交**。

   > 说明
   > - 一个包含多节点的 Kubernetes 集群的创建时间一般约为十分钟。
   > - 系统会自动跳转到 **创建集群详情** 页面。在集群的创建过程中，若出现任务失败的情况，可单击事件查看具体错误详情，或单击 **重试** 或 **忽略**。

步骤三：创建转账应用

在 **应用管理** 控制台创建名为 transfer-money 的应用。

操作步骤

1. 登录应用管理控制台，在左侧导航栏单击 **应用列表**。

2. 在 **应用列表** 页面，单击右上角的 **创建应用**。
3. 在 **创建应用** 页面输入以下信息，单击 **确定**。
   - **应用名称**：输入 `transfer-money` 。

     > **说明**
     > 应用名称在同一租户内必须是唯一的。

   - **技术栈**：选择 SOFABoot。
   - **应用分组**：选择已创建的应用分组。更多信息参见 管理应用分组。
   - **应用标签**：应用分级标签，便于进行应用快速检索，支持添加系统标签和自定义标签。
   - **应用描述**：选填，应用描述的长度不能超过 500 个字符，如 `这是一个示例应用` 。更多信息参见 管理应用。

步骤四：准备镜像

LHC 以镜像的方式部署应用服务，创建应用服务前需先准备好镜像。为方便您测试，LHC 提供以下可以公开访问的示例镜像：

镜像名称	镜像地址
aks-transfermoney-server	registry-vpc.cn-shanghai.aliyuncs.com/sofa-samples/aks-transfermoney-server:20190726095608708-master.40215e87
aks-transfermoney-client	registry-vpc.cn-shanghai.aliyuncs.com/sofa-samples/aks-transfermoney-client:20190815134744553-master.5ea6278b

步骤五：创建应用服务

创建 2 个有依赖关系的应用服务：aks-transfermoney-server、aks-transfermoney-client。

前提条件

已创建联邦命名空间，例如 `antcloud-demo` 。具体步骤参见 创建联邦命名空间。

创建转账服务端的应用服务

1. 登录控制台，在左侧导航栏单击 **发布运维** > **应用服务**。
2. 在应用服务列表页，单击 **创建应用服务**。
3. 在 **创建应用服务** 页面，填写以下基本信息，单击 **下一步**。
   - **命名空间**：选择一个集群中的命名空间（namespace），默认为列表第一个。
   - **应用服务名称**：容器服务的名称，输入 `aks-transfermoney-server` 。应用服务名称允许包含（小写）字母、数字、连字符，且必须以字母开头，以字母或数字结尾。同一个命名空间下不允许同名。
   - **所属应用**：选择 `transfer-money` 。
   - **描述**：选填。容器服务的描述。
4. 在 **Pod 模板配置** 页面，填写以下信息，单击 **下一步**。
   - **容器名称**：输入名称。容器名称允许包含（小写）字母、数字、连字符、且必须以字母开头、以字母或者数字结尾，最多允许输入 63 个字符。

- **访问类型**：选择 **镜像仓库**，使用示例镜像 aks-transfermoney-server，镜像地址参见 准备镜像。
- **CPU 配置**：**请求核数** 为 200 millicores，**最大核数** 为 500 cores。
- **内存配置**：**请求内存** 为 512 MiB，**最大内存** 为 1 GiB。
- 在 **高级配置** > **环境变量配置** 中，配置 `APPSVC_VERSION = V1` ，`ANTCLOUD_SOFA_PROFILE = prod` 。
- 在 **高级配置** > **健康检查配置** 中，配置 Readiness Probe 检查方法为 `Http GET` 方法 、访问路径为 `/health/readiness` 、访问端口 `8080` 。其余项按照输入框提示配置。
- **配置覆盖**：保持关闭。

5. 在 **弹性配置** 页面，填写以下信息，单击 **下一步**。
   副本伸缩策略配置：目前仅支持 **固定副本数**，默认为 0，勾选部署单元（ `RZ01A` 、 `RZ02A` ）并修改为期望副本数 1，即应用服务运行时保持固定数目的 Pod 副本。
6. 在 **访问配置** 页面，填写以下信息，单击 **下一步**。
   i. 在 **访问配置** 页面，单击 **添加负载均衡**。
   ii. 在 **负载均衡** 窗口，填写以下信息后，单击 **确定**。
   - **负载均衡名称**：填写服务名称。系统默认生成服务名称前缀为 `应用服务名称-` 。

     > 说明
     > 需记录下这个 Service Name，需要以环境变量形式配置到客户端应用。

   - **访问方式**：选择 **内网**。创建一个内部的 Service，并将流量转发到容器的相应端口上。
   - **端口映射**：单击 **添加端口映射**，填写以下信息。其余保持默认设置。
     - **协议**：选择 TCP。
     - **转发规则**：选择 **RR 轮询**。
     - **前端端口**：容器镜像中工作负载程序实际监听的端口，输入 `80` 。
     - **后端端口**：容器端口映射到集群虚拟 IP 上的端口，输入 `8341` 。
7. 在 **部署和调度配置** 页面，为应用服务保持系统默认配置，单击 **下一步**。
8. 在应用服务 **预览** 页面，确认信息无误，单击 **提交**。

> 说明
> 该应用服务配置编辑完成提交后，进行版本备注，方便区分版本。

创建转账客户端的应用服务

1. 登录控制台，在左侧导航栏单击 **发布运维** > **应用服务**。
2. 在应用服务列表页，单击 **创建应用服务**。
3. 在 **创建应用服务** 页面，填写以下基本信息，单击 **下一步**。
   - **命名空间**：选择一个集群中的命名空间（namespace），默认为列表第一个。
   - **应用服务名称**：容器服务的名称，输入 `aks-transfermoney-client` 。服务实例名称允许包含（小写）字母、数字、连字符，且必须以字母开头，以字母或数字结尾。同一个命名空间下不允许同名。

- **所属应用**：选择 `transfer-money` 。
- **负责人**：选择该应用服务的负责人。
- **描述**：选填。容器服务的描述。

4. 在 **Pod 模板配置** 页面，填写以下信息，单击 **下一步**。
   - **容器名称**：输入名称。
   - **访问类型**：选择 **镜像仓库**，使用示例镜像 aks-transfermoney-client，镜像地址参见 准备镜像。
   - **CPU 配置**：**请求核数** 为 200 millicores，**最大核数** 为 500 core。
   - **内存配置**：**请求内存** 为 512 MiB，**最大内存** 为 1 GiB。
   - 在 **高级配置** > **环境变量配置** 中，配置变量名称为 SERVER_TM_ADDRESS，Server 端值为 Service Name。
5. 在 **弹性配置** 页面，填写以下信息，单击 **下一步**。
   副本伸缩策略配置：目前仅支持**固定副本数**，默认为 0，勾选部署单元（ `RZ01A` 、 `RZ02A` ）并修改为期望副本数 1，即应用服务运行时保持固定数目的 Pod 副本。每个 client 能够压测并发上限为 1000。
6. 在 **访问配置** 页面，单击 **添加负载均衡**，填写以下信息，单击 **下一步**。
   i. 在 **负载均衡** 窗口，填写以下信息后，单击 **确定**。
   - **负载均衡名称**：填写服务名称。系统默认生成服务名称前缀为 `应用服务名称-` 。
   - **访问方式**：选择 **外网**。创建一个内部的 Service，并将流量转发到容器的相应端口上。

     > 说明
     > 选取两个应用服务并进行串行发布。配置外部 Service 会创建一个公网的 LoadBalancer，并将流量转发到容器的相应端口上。访问方式由公网负载均衡服务地址以及设置的访问端口组成，例如 `10.117.117.117:80` 。

   - **访问方式**：选择 **外网**。创建一个内部的 Service，并将流量转发到容器的相应端口上。
   - **端口映射**：单击 **添加端口映射**，填写以下信息。其余保持默认设置。
     - **协议**：选择 TCP。
     - **转发规则**：选择 **RR 轮询**。
     - **前端端口**：容器镜像中工作负载程序实际监听的端口，输入 `80` 。
     - **后端端口**：容器端口映射到集群虚拟 IP 上的端口，输入 `8341` 。
7. 在 **部署和调度配置** 页面，为应用服务保持系统默认配置，单击 **下一步**。
8. 在应用服务 **预览** 页面，确认信息无误，单击 **提交**。

   > 说明
   > 该应用服务配置编辑完成提交后，进行版本备注，方便区分版本。

9. 在 **部署和调度配置** 页面，为应用服务保持系统默认配置，单击 **下一步**。
10. 在应用服务 **预览** 页面，确认信息无误，单击 **提交**。

> 说明
> 该应用服务配置编辑完成提交后，进行版本备注，方便区分版本。

步骤六：创建发布单

通过发布单同时发布上一步中创建的两个有依赖关系的应用服务。

> 说明
> 选取两个应用服务并进行串行发布。

操作步骤

1. 登录控制台，在左侧导航栏单击 **发布运维** > **发布单** 进入发布部署大盘。
2. 在发布部署大盘，单击 **创建发布单**。
3. 在 **创建发布单** 页面，填写以下发布信息后，单击 **下一步**。
   - **基本信息**
     - **标题**：发布标题。例如 `Demodeploy` 。
     - **类型**：仅支持分组发布。
     - **命名空间**：选择待发布的应用服务所属的命名空间。
     - **应用服务发布列表**：在 **待选应用服务列表** 中单击选择需要的应用服务：aks-transfermoney-client、aks-transfermoney-server。单击 **>** 图标将应用服务添加到 **已选应用服务列表** 中。

       > 说明
       > 若应用服务存在多个提交版本，需要选择要发布的版本。默认选择最新的版本。

   - **高级配置**
     设置应用服务依赖关系：aks-transfermoney-client 依赖于 aks-transfermoney-server。
4. 在 **预览** 页面确认信息无误后，单击 **创建**。系统会自动跳转到 **发布单详情** 页面，单击 **整体发布** 即可开始发布。

步骤七：验证结果

验证应用服务 aks-transfermoney-client 及 aks-transfermoney-server 是否在双集群中发布成功。

操作步骤

1. 登录控制台，在左侧导航栏单击 **发布运维** > **应用服务**。
2. 在应用服务列表中，分别单击 aks-transfermoney-client 及 aks-transfermoney-server 进入实例详情页。
3. 在 Pod 页签中，查看两个应用服务在两个部署单元中均有一个应用容器运行中。

更多应用服务内容，参见 查看应用服务详情 。

9.3. 准备环境

9.3.1. 创建单元化工作空间

您可以在 **控制台** > **全局设置** 中进行当前租户的所有工作空间的管理，包括进行查看、添加、编辑、删除工作空间等操作。

背景介绍

一个工作空间的本质，是指“网络互通、安全策略一致、访问延时极小”的一组资源。您可以通过工作空间方便地将资源进行分组管理，例如：根据不同的研发交付需求，将工作空间划分为开发工作空间、测试工作空间、生产工作空间等。同时，不同工作空间中的资源互相隔离，可以为每个工作空间分配单独的操作员权限进行管理。为支持高可用（High Availability），工作空间可以跨可用区（Zone），但不能跨地域（Region）。每个工作空间都需要创建相应的基础资源（包括：专有网络、交换机、安全组、ECS 等），并基于这些资源部署集群，从而为应用服务提供服务。

前提条件

了解 单元化架构。

操作步骤

1. 登录控制台，在左侧导航栏单击左下角的 **全局设置**，进入工作空间列表页面。
2. 单击 **创建工作空间**，选择 **单元化工作空间** 类型，单击 **创建**。
   - **标准工作空间**：标准工作空间（Workspace）是 SOFAStack 提供的一种组织机制，用于将服务于不同目标、阶段的资源分组隔离管理。您可以根据研发运维流程，为每个阶段分配工作空间，例如单机房（即单可用区）的开发工作空间、双机房（即包括两个可用区）的生产工作空间。
   - **单元化工作空间**：在标准工作空间的基础上提供了单元化能力，可用于同城双活及异地容灾场景，本质是一组标准工作空间的集合。您可以通过单元化工作空间组对用户资源进行隔离，不同工作空间组下的集群彼此隔离。
3. 在 **创建工作空间** 页面，输入以下基本信息。
   - **工作空间标识**：2~45 个字符，工作空间的英文标识，全局唯一，一经确定无法修改，例如：dev、test、prod 等。
   - **工作空间名称**：1~64 个字符，工作空间的显示名称，例如：开发工作空间、测试工作空间、生产工作空间。
   - **地域**：工作空间所在的地域（Region），一个工作空间一定属于某个地域。
   - **网络类型**：选择 **VPC 网络**。
   - **导入 VPC**：默认关闭。开启后，您可以复用已有的 VPC 资源，将工作空间与已创建的 VPC 绑定。
4. 单击 **下一步**，在 **创建单元架构** 页面单击 **添加可用区**。 每个工作空间配置的可用区（Available Zone）个数不做限制。本例中为工作空间配置两个可用区，为支持双机房高可用等架构做准备。

> **说明**
> 系统会根据可用区配置自动为您生成单元化架构拓扑，划分好逻辑单元和部署单元。更多信息，参考 创建逻辑单元。



5. 单击 **下一步**，在 **创建 VPC** 页面，输入以下配置信息：
   - **专有网络名称**：由 2~128 个英文或中文字符组成，必须以大小字母或中文开头，可包含数字、下划线（_）或连字符（-），不能以 `http://` 和 `https://` 开头。系统自动生成 `单元化工作空间标识-vpc` 组成的默认名称。
   - **专有网络网段**：专有网络的网段，一旦选择便无法更改，专有网络内的所有资源，如 ECS、RDS、SLB 的私网 IP 都在该网段内。可选网段如下：
     - 10.0.0.0/9
     - 172.16.0.0/12
     - 192.168.0.0/16
   - **描述**：选填，可包含 2~256 个中英文字符，不能以 `http://` 和 `https://` 开头。
   - **交换机**：单击 **添加交换机**，在弹出的 **添加交换机** 窗口，填写以下信息，单击 **提交**。
     - **名称**：交换机名称。长度为 2~128 个字符，以英文字母或中文开头，可包含数字、下划线（_）和短横线（-）。系统自动生成 `单元化工作空间标识-vsw` 组成的默认名称。
     - **可用区**：交换机的可用区。同一 VPC 内不同可用区的交换机内网互通。您需要为每个可用区创建一台交换机。
     - **自定义网段**：默认关闭。开启后，需填写网段地址。交换机的网段可以和其所属的 VPC 网段相同或者是其 VPC 网段的子网。

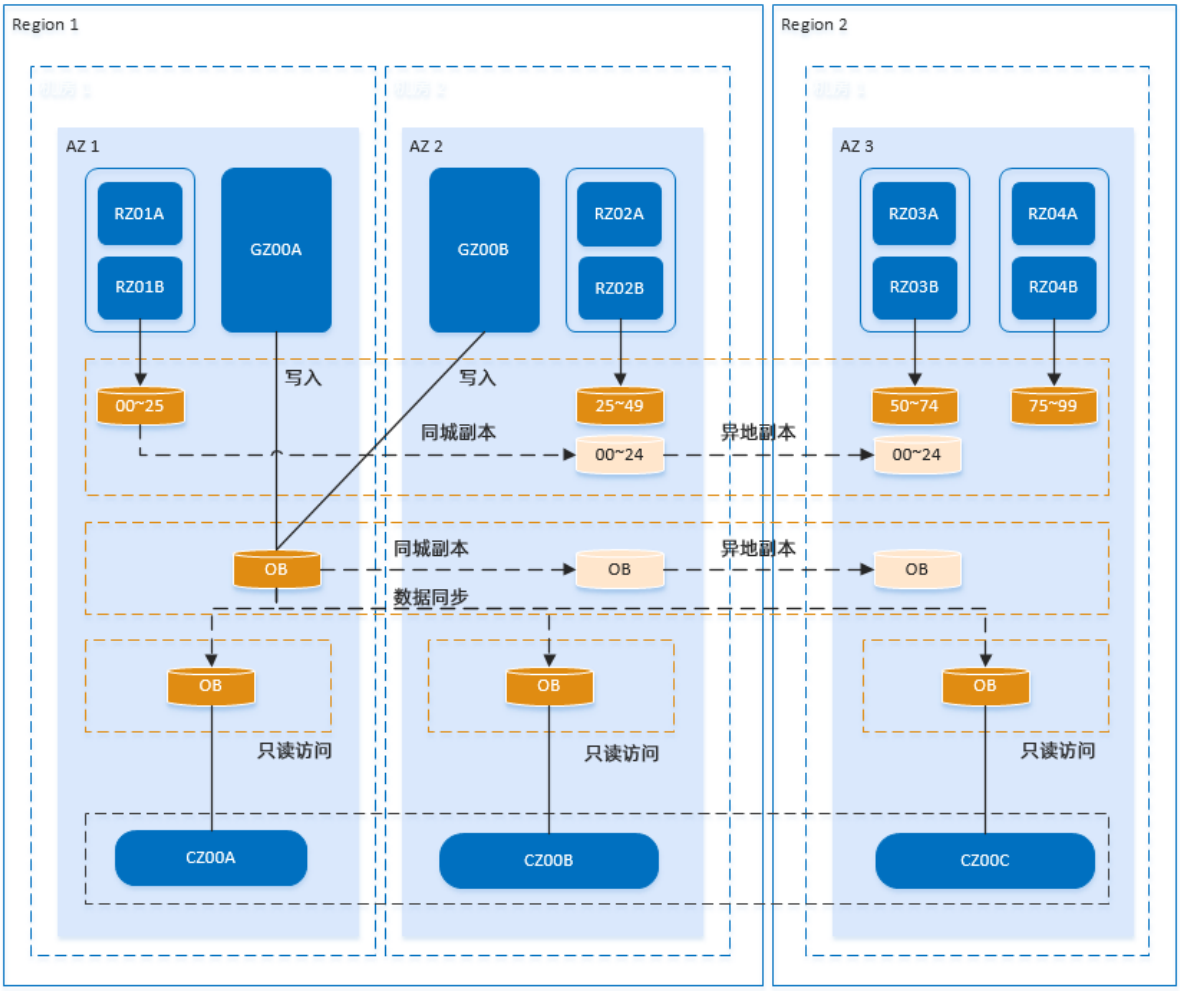
- **子网掩码**：**自定义网段** 关闭时，需分别选择子网掩码和网段地址。默认专有的网段掩码是 24 位，例如 172.31.0.0/24，最多可提供 65536 个私网 IP 地址。范围为 16~29 位之间，可提供 4~65532 个地址。
- **描述**： 输入交换机的描述信息。可包含 2~256 个中英文字符，不能以 `http://` 和 `https://` 开头。

6. 单击 **下一步**，在 **创建安全组** 页面，单击 **添加安全组**，在弹出的 **添加安全组** 窗口，填写以下信息，单击 **确定**。
   - **安全组名称**：长度为 2~128个英文或中文字符，必须以大小字母或中文开头，不能以 `http://` 和 `https://` 开头。可以包含数字、半角冒号（:）、下划线（_）或者连字符（-）。系统自动生成 `单元化工作空间标识-sg` 组成的默认名称。
   - **描述**：可包含 2~256 个中英文字符，不能以 `http://` 和 `https://` 开头。
   - **规则**：单击 **添加规则**，在弹出的 **创建规则** 窗口，填写以下信息，单击 **确定**。
   - **网卡类型**：选择 **内网**。VPC 类型安全组的网卡类型（NicType）只可设置内网（intranet）。
   - **规则方向**：
     - **内网出方向**：是指 ECS 实例访问内网中其他 ECS 实例或者公网上的资源。
     - **内网入方向**：是指内网中的其他 ECS 实例或公网上的资源访问 ECS 实例。
   - **协议类型**：传输层协议。取值大小写敏感。可选项如下：
     - **TCP**：支持 TCP 协议。
     - **UDP**：支持 UDP 协议。
     - **ICMP**：ICMP 协议仅支持 IPv4 地址。
     - **GRE**：支持 GRE 协议。
     - **ALL**：支持所有协议。
   - **端口范围**：目的端安全组开放的传输层协议相关的端口范围，取值范围为 1~65535，设置格式例如“1/200”、“80/80”，其中“-1/-1”不能单独设置，代表不限制端口。取值范围：
     - TCP/UDP 协议：取值范围为 1~65535。使用斜线（/）隔开起始端口和终止端口。正确示范：1/200；错误示范：200/1。
     - ICMP 协议：-1/-1。
     - GRE 协议：-1/-1。
     - 协议类型取值为 ALL：-1/-1。
   - **访问权限**：选择 **接受访问** 或 **拒绝访问**。
   - **优先级**：数字越大代表的优先级越高。选择范围为 1~100。
   - **源端 IP 地址**：填写单一 IP 地址或者 CIDR 网段格式，如：12.1.1.1 或 13.1.1.1/25。如果填写 0.0.0.0/0 表示允许或拒绝所有 IP 地址的访问，设置时请务必谨慎。目前仅支持 IPv4 格式的 IP 地址范围。 默认值：0.0.0.0/0。
   - **目标端 IP 地址**：填写单一 IP 地址或者 CIDR 网段格式，如：12.1.1.1 或 13.1.1.1/25。如果填写 0.0.0.0/0 表示允许或拒绝所有 IP 地址的访问，设置时请务必谨慎。目前仅支持 IPv4 格式的 IP 地址范围。
   - **规则描述信息**：输入安全组规则的描述信息。长度为 1~512 个字符。
7. 单击 **提交**。

9.3.2. 创建逻辑单元

使用 LHC 部署应用服务前，您需要先定义系统的逻辑单元。逻辑单元是单元化架构的基础，一个单元被称为一个 Zone。根据业务特点不同，您可以将系统部署在不同类型的逻辑单元中。

- RZone（Region Zone）：部署按用户维度拆分的关键业务系统。核心业务和数据单元化拆分，拆分后分片均衡，单元内尽量自包含（调用封闭），拥有自己的数据，能完成所有业务。一个可用区可以有多个 RZone。
- GZone（Global Zone）：部署未按用户维度拆分的系统，被 RZone 依赖，提供不可拆分的数据和服务，如配置型的业务。数据库可以和 RZone 共享，多租户隔离，全局只有一组，可以配置流量权重。
- CZone（City Zone）：部署未按用户维度拆分的系统，被 RZone 高频访问 ，解决跨域通信延时问题。为了解决异地延迟问题而特别设计，适合读多写少且不可拆分的业务。 一般每个城市一套应用和数据，是 GZone 的快照。



操作步骤

> 说明
> 在创建单元化工作空间时，系统会自动进行逻辑单元和部署单元的规划，参考 创建单元化工作空间。您可以根据实际需求，在发布应用服务前，对逻辑单元进行修改或增删。

1. 登录控制台，在左侧导航栏单击 **全局设置** 进入工作空间卡片页。
2. 单击需要修改逻辑单元的工作空间卡片进入工作空间详情页。

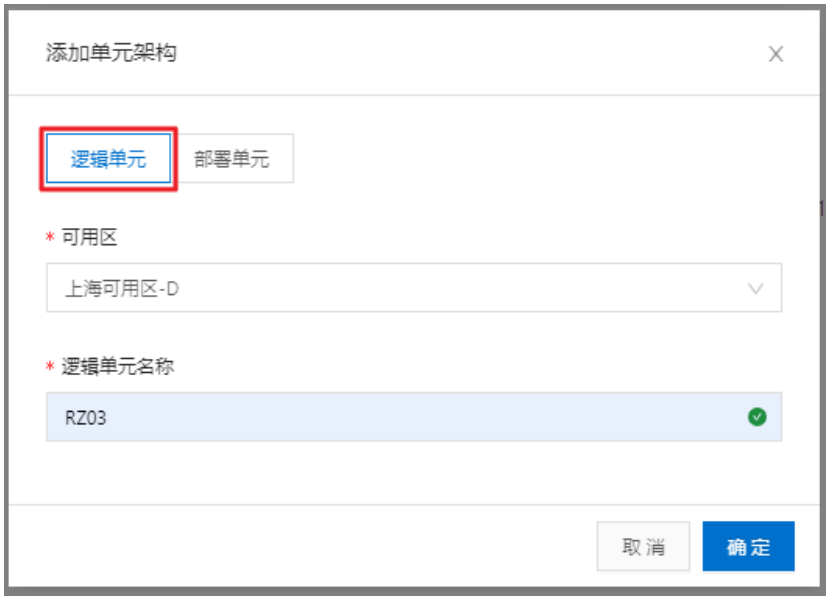
3. 选择 **单元架构** 页签，单击 **编辑**，架构拓扑进入可编辑的状态。



4. 您可以根据需要对逻辑单元进行以下增删操作：
   - **增加**：单击 **添加** 可以选择已有的可用区并增加新的逻辑单元。

     > 说明
     > 仅支持增加 RZone 类型的逻辑单元。逻辑单元名称请以 `RZ` 开头，名称长度限制为 3-8 个字符，建议按照命名规范为 `逻辑单元类型简称+数字` 的形式，如 `RZ03`。

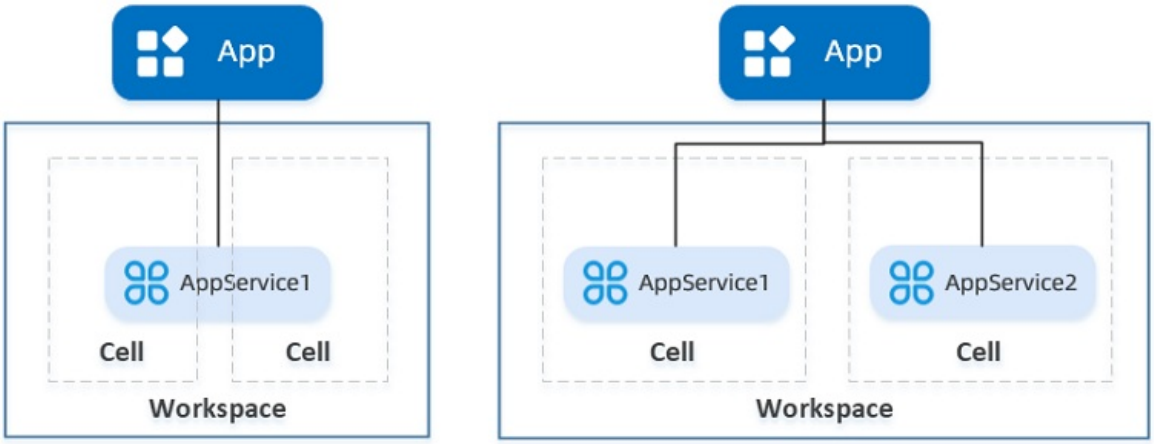
   - **删除**：单击删除图标并在弹出对话框中单击 **确定**，可以将已有的逻辑单元删除。



9.3.3. 创建部署单元

部署单元 Cell 是实现单元化的基础。

- 双机房双活场景下，一个应用的应用服务可以跨部署单元（Cell）部署。
- 单元化场景下，一个应用可以创建多个应用服务在不同部署单元（Cell）部署。



前提条件

已 创建逻辑单元。

操作步骤

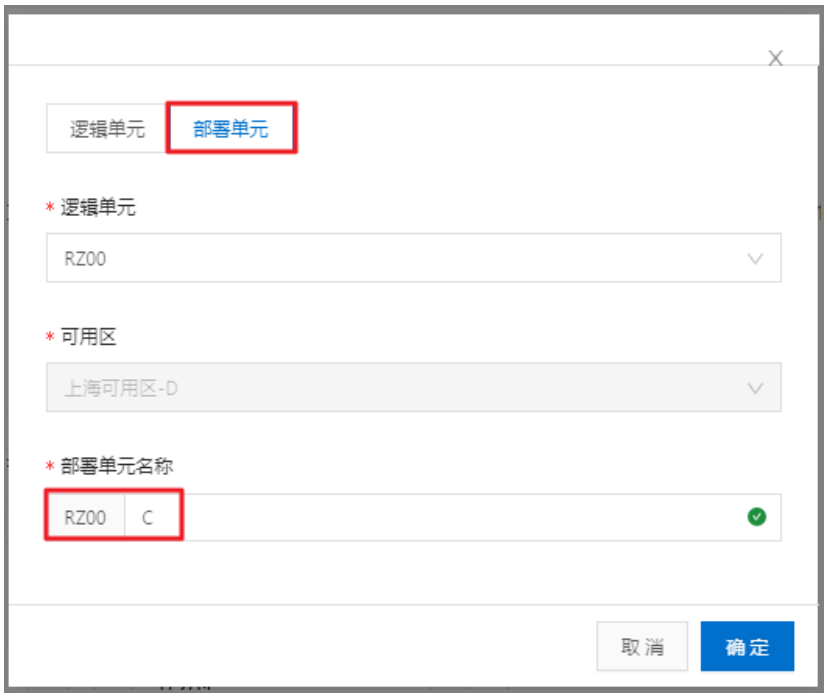
> 说明
> 在创建单元化工作空间时，系统会自动进行逻辑单元和部署单元的规划，参考 创建单元化工作空间。您可以根据实际需求，在发布应用服务前，对部署单元进行增删。

1. 登录控制台，在左侧导航栏单击 **全局设置** 进入工作空间卡片页。
2. 单击需要修改逻辑单元的工作空间卡片进入工作空间详情页。
3. 选择 **单元架构** 页签，单击 **编辑**，架构拓扑进入可编辑的状态。



4. 您可以根据需要对部署单元进行以下增删操作：
   - **增加**：单击 **添加** 填写如下配置信息：
     - **逻辑单元**：选择所属的逻辑单元，支持 RZone 与 GZone 类型。
     - **可用区**：选择所属的可用区。若选择的逻辑单元为 RZone 类型，会自动选中关联的可用区。

- **部署单元名称**：自动生成逻辑单元名称作为前缀，填入字符长度限制为 1-10 个字符，建议使用 `逻辑单元+字母` 方式命名，如 `GZ00A` 。



  - **删除**：单击删除图标并在弹出对话框中单击 **确定**，可以将已有的部署单元删除。

9.3.4. 删除工作空间

您可以在 **控制台 > 全局设置** 中删除当前租户的工作空间。

前提条件

删除工作空间前，需要先清理当前工作空间的所有资源，包括 ECS、VPC、VSwitch、Security Group 资源，当所有资源都被回收之后，才可以删除工作空间，否则会报错。

操作步骤

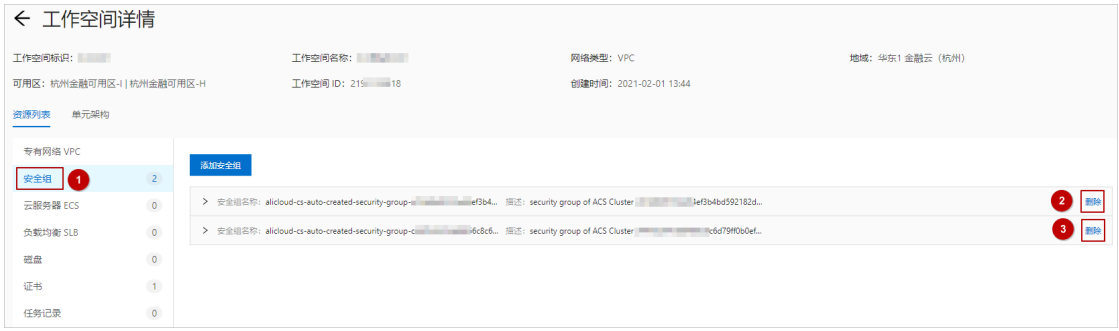
1. 清理 ECS 资源。
   在 **集群管理 >节点列表** 页面选择所有节点，单击 **操作 > 删除**，删除所有节点后，保证当前工作空间的所有 ECS 都被回收。

   > **说明**
   > 清理工作空间前要先删除当前工作空间的集群，这样不会造成闲置集群资源的浪费，因为删除工作空间之后，产品页面上将无法查看到当前工作空间的集群信息，只能在数据库中的进行查找。

2. 删除交换机。
   i. 在 **控制台 > 全局设置** 中单击工作空间卡片，进入 **工作空间详情** 页。
   ii. 在 **资源列表** 页面，单击 **专有网络 VPC**，在 **交换机** 页签中删除所有交换机。



3. 删除安全组。
   i. 在 **控制台 > 全局设置** 中单击 **工作空间**，进入 **工作空间详情** 页。
   ii. 单击 **安全组** 页面右侧的 **删除**。



4. 删除 VPC。
   i. 在 **控制台 > 全局设置** 中单击 **工作空间**，进入 **工作空间详情** 页。
   ii. 单击 **专有网络 VPC**，删除当前工作空间关联的 VPC。

> **说明**
> 删除 VPC 之前要先删除当前工作空间的所有交换机和安全组，因为交换机和安全组资源是创建在当前 VPC 里面的，若直接删除 VPC 会报错。



5. 删除单元架构。
   i. 在 **控制台 > 全局设置** 中单击 **工作空间**，进入 **工作空间详情** 页。

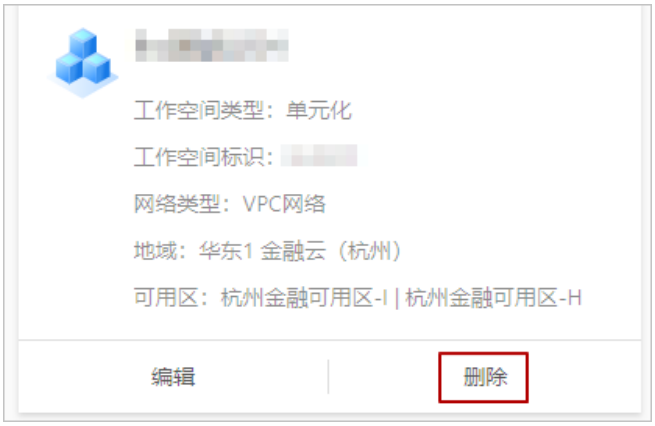
ii. 选择 **单元架构** 页签，单击 **编辑**，架构拓扑进入可编辑的状态。

iii. 单击删除图标，将已有的单元全部删除。



6. 删除工作空间。
   当 ECS、交换机、安全组、VPC 、单元架构都删除成功之后，当前工作空间才可被删除，返回工作空间列表，单击要删除的工作空间卡片上的删除按钮即可。

> 说明
> 删除工作空间会造成当前环境不可用，请谨慎操作。



9.4. 集群

9.4.1. 创建集群

Kubernetes 托管版集群只需创建 Worker 节点，Master 节点由阿里云容器服务创建并托管，从而让您可以更多关注业务本身。本文介绍如何通过 AKS 控制台创建 Kubernetes 托管版集群。

> 说明
> 从对接 ACK 集群的功能上线时起（2020 年 8 月 15 号），阿里云底座上新建集群默认为 ACK 托管版集群。

前提条件

- 您拥有 LHC 管理员权限。
- 已创建单元化工作空间。
- 首次使用 ACK 集群时，需先授权 ACK 访问云资源的权限，详情参考 首次使用容器服务 Kubernetes 版。

操作步骤

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**。
2. 在集群列表页，单击 **创建集群**。
3. 在 **集群创建** 页面，系统会自动进行预检查，确保已经开通相关产品并且租户下的账户余额多于 100 元。预检查通过后，单击 **下一步**。

   > 说明
   > - 若检查失败，完成修复失败项目后，可以单击 **重新检查** 操作重新进行预检查。
   > - 由于 SOFAStack 的集群网络插件 CCM 需授权访问云资源以提供负载均衡、统一接入流量管理等服务，预检查步骤会检测当前用户的 RAM Role 列表中是否包含 AliyunSofaCafeCCMRole 角色，若没有需要您按照 CCM 角色授权 进行角色授权。

4. 在 **基本配置** 页面，完成集群配置后，单击 **下一步**。

i. 完成集群基础选项配置。

配置项	说明
专有网络	Kubernetes 集群仅支持专有网络。您可以在已有 VPC 列表中选择所需的 VPC。如果没有您需要的专有网络，可以在资源管理控制台进行专有网络的创建，参考 创建专有网络。
集群名称	填写集群的名称。 **说明** 集群名称应包含 1~63 个字符，可包含数字、汉字、英文字符或短划线（-）。
Kubernetes 版本	选择 Kubernetes 版本。可选择 `1.20.4-aliyun.1` 或 `1.18.8-aliyun.1` 。
容器运行时	支持 `docker 19.03.5` 、 `containerd 1.4.4` 。

ii. 完成集群网络配置。

配置项	说明
虚拟交换机	您可以在已有虚拟交换机列表中，根据可用区选择 1~3 个交换机。 **说明** 如果没有您需要的交换机，可以通过单击创建交换机链接进行创建，请参考 创建交换机。
网络插件	设置启用的网络插件。当前支持 Terway 网络插件。 **说明** Terway 是阿里云容器服务自研的网络插件，将阿里云的弹性网卡分配给容器，支持 Kubernetes 的 Network Policy 来定义容器间的访问策略，支持对单个容器做带宽的限流。
Pod 虚拟交换机	为 Pod 分配 IP 的虚拟交换机。每个 Pod 虚拟交换机分别对应一个 Worker 实例的虚拟交换机。 **说明** 如果没有您需要的交换机，可以通过单击创建交换机链接进行创建，请参考 创建交换机。
Service CIDR	设置 Service CIDR。网段不能与 VPC 及 VPC 内已有 Kubernetes 集群使用的网段重复，创建成功后不能修改。

iii. 完成集群高级配置。

配置项	说明
配置 SNAT	不可修改，默认选中。创建集群时，默认不开通公网。如果您选择的 VPC 不具备公网访问能力，选中为专有网络配置 SNAT 后，ACK 将为您创建 NAT 网关并自动配置 SNAT 规则。
公网访问	设置是否开放使用 EIP 暴露 API Server。API Server 提供了各类资源对象（Pod、Service 等）的增删改查及 watch 等 HTTP Rest 接口。 ■ 若选择开放，会创建一个 EIP，并挂载到内网 SLB 上。此时，Master 节点的 6443 端口（对应 API Server）暴露出来，您可以在外网通过 kubeconfig 连接并操作集群。 ■ 若选择不开放，则不会创建 EIP，您只能在 VPC 内部用 kubeconfig 连接并操作集群。 **说明** 若需获取 kubeconfig 信息，请前往 ACK 控制台。
kube-proxy 代理模式	支持 **iptables** 和 **IPVS** 两种模式。 ■ **iptables**：成熟稳定的 kube-proxy 代理模式，Kubernetes Service 的服务发现和负载均衡使用 iptables 规则配置，但性能一般，受规模影响较大，适用于集群存在少量的 Service 的场景。 ■ **IPVS**：高性能的 kube-proxy 代理模式，Kubernetes Service 的服务发现和负载均衡使用 Linux ipvs 模块进行配置，适用于集群存在大量的 Service，对负载均衡有高性能要求的场景。
集群删除保护	防止通过控制台或者 API 误删除集群。默认选中，可修改。
部署单元	可勾选 **使用默认配置**。若取消勾选，可以为可用区设置不同的部署单元，若需要修改部署单元，可前往 **全局设置 > 工作空间详情** 进行设置。
日志服务	支持 **创建新的 Project**、**使用已有 Project**。 ■ **创建新的 Project**：默认处于选中状态，将自动创建名称为 `k8s-log-{ClusterID}` 的 Project。 **说明** ClusterID 表示新建集群的唯一标识。 ■ **使用已有 Project**：从现有 Project 下拉列表中选择一个 Project 来管理采集的日志。 

5. 在 **节点配置** 页面，完成以下 Worker 节点配置后，单击 **下一步**。

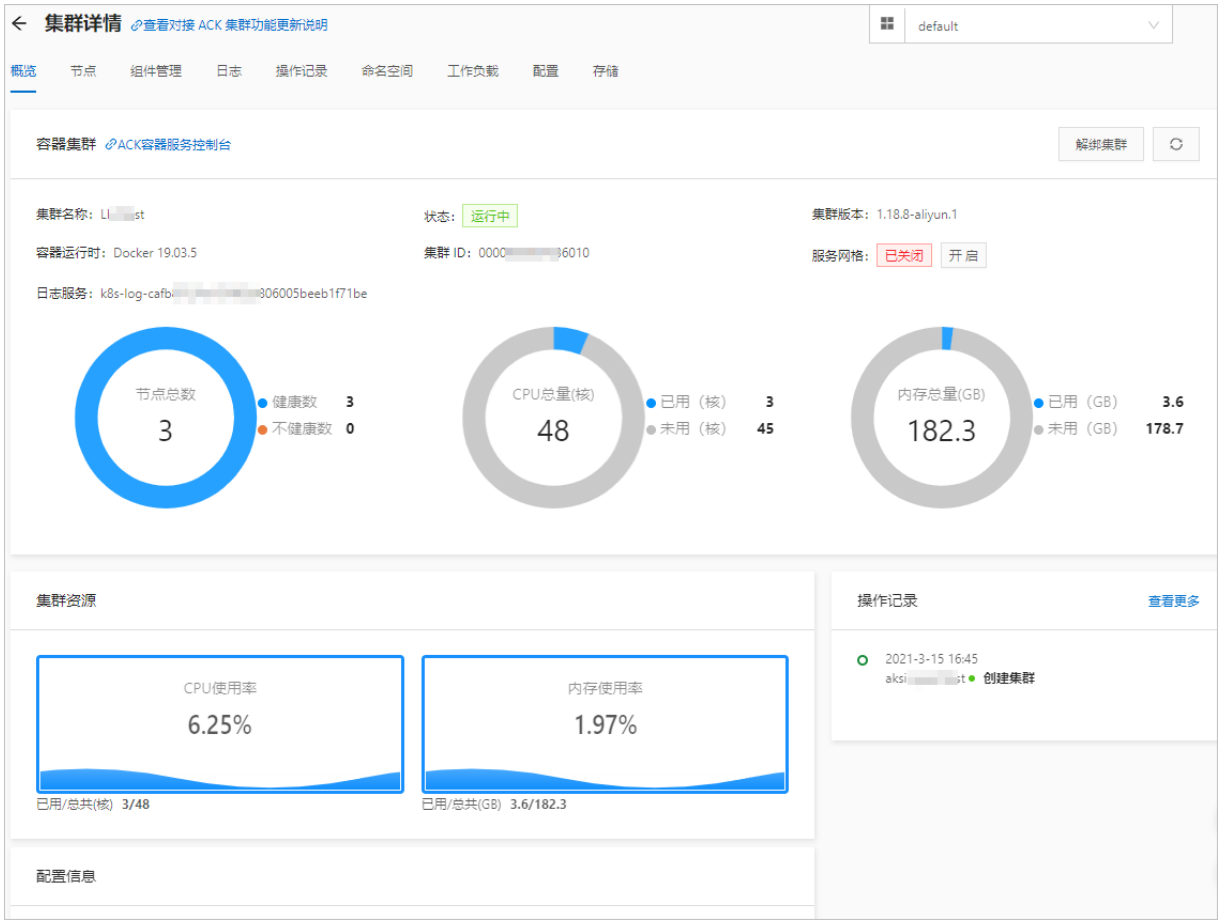
配置项	说明
付费类型	支持 **按量付费** 和 **包年包月** 两种节点付费类型。选择 **包年包月** 时，需设置以下参数。 ○ **购买时长**：目前支持选择 1、2、3、6 个月和 1~5 年。 ○ **自动续费**：设置是否自动续费。
节点数量	创建 Worker 实例（ECS 实例）的数量。
实例规格	支持选择多个实例规格。详情请参考 实例规格族。
已选实例规格	呈现选中的实例规格。
系统盘	支持 **ESSD 云盘** 和 **高效云盘**。
挂载磁盘	支持 **ESSD 云盘** 和 **高效云盘**。
操作系统	支持 CentOS 和 Aliyun Linux 操作系统。
登录密码	设置节点的登录密码。8～30 个字符，且同时包含三项（字母、数字和特殊符号）。
确认密码	确认设置的节点登录密码。

6. 在集群的 **配置预览** 页面，确认配置无误后，单击 **提交**。

> 说明
> 一个包含多节点的 Kubernetes 集群的创建时间一般约为十分钟。

执行结果

集群创建成功后，您可以在集群详情页面查看所创建的集群信息。



9.4.2. CCM 角色授权

由于 SOFAStack 的集群网络插件 CCM 需授权访问云资源以提供负载均衡、统一接入流量管理等服务，在 创建集群 预检查步骤中会检测当前用户的 RAM Role 列表中是否包含 AliyunSofaCafeCCMRole 角色，若没有需要您按照以下操作步骤进行角色授权。

操作步骤

1. 进入 AliyunSofaCafeCCMRole 角色管理 页面。
2. 单击 **修改信任策略**，在 Statement 中 **新增** 以下内容。

> **说明**
>
> `用户账号 ID` 可以从 AliyunSofaCafeCCMRole 角色的 ARN 中获取，例如对于角色 `acs:ram::1688168816881688:role/aliyunsofacafeccmrole`，`用户账号ID` 应该被替换为 `1688168816881688`。

```
{
    "Action": "sts:AssumeRole",
    "Effect": "Allow",
    "Principal": {
        "RAM": [
            "acs:ram::用户账号ID:root"
        ]
    }
},
```

一个已修改好的信任策略如下图所示，其中**红框部分是新增内容**：



9.4.3. 导入集群

LHC 对接 ACK 托管版集群后，您可以将 ACK 集群导入 LHC。本文介绍如何通过 LHC 控制台界面快速导入已经创建好的 ACK 集群。

> 说明
>
> - 目前，仅支持导入 ACK 标准托管版，且容器运行时必须为 Docker。
> - 待导入的 ACK 集群的 VPC 必须与当前工作空间所属的 VPC 一致。
> - 基于 ACK 集群的部署环境下，每个可用区（机房）仅允许导入一个 ACK 集群。

前提条件

- 您拥有 LHC 管理员权限。

- 您已在 ACK 控制台创建了一个标准托管版集群，并且该集群所属的 VPC 与将导入的工作空间所属 VPC 相同。
- 单元化工作空间下的现有集群数量小于可用区数量。

> 说明
> 每个单元化工作空间可纳管的集群数量不能超过工作空间中所包含的可用区数量。

操作步骤

1. 登录 LHC 控制台。
2. 在左侧导航栏中选择 **集群管理 > 集群详情**，在集群列表页面，单击 **导入集群**。
3. 在 **导入集群** 页面，完成以下配置。

配置项	说明
专有网络	从已有 VPC 列表中，选择所需的 VPC。
选择已有集群	从已有集群列表中，选择您需要导入的 ACK 集群。
部署单元	可以为可用区设置不同的部署单元。 如需修改部署单元，可前往 **全局设置 > 工作空间详情** 进行设置。

4. 单击 **提交**，即可导入该集群。

9.4.4. 扩容集群

集群扩容即扩容集群内的节点资源。您可以根据实际业务需要对集群的节点进行扩容。
若节点资源不足，会导致创建工作负载等操作失败。您可以通过以下两种方式进行扩容：

- 创建新节点
- 添加已有节点

前提条件

已创建集群 或 已导入集群

创建新节点

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击目标集群卡片中的 **扩容集群**。



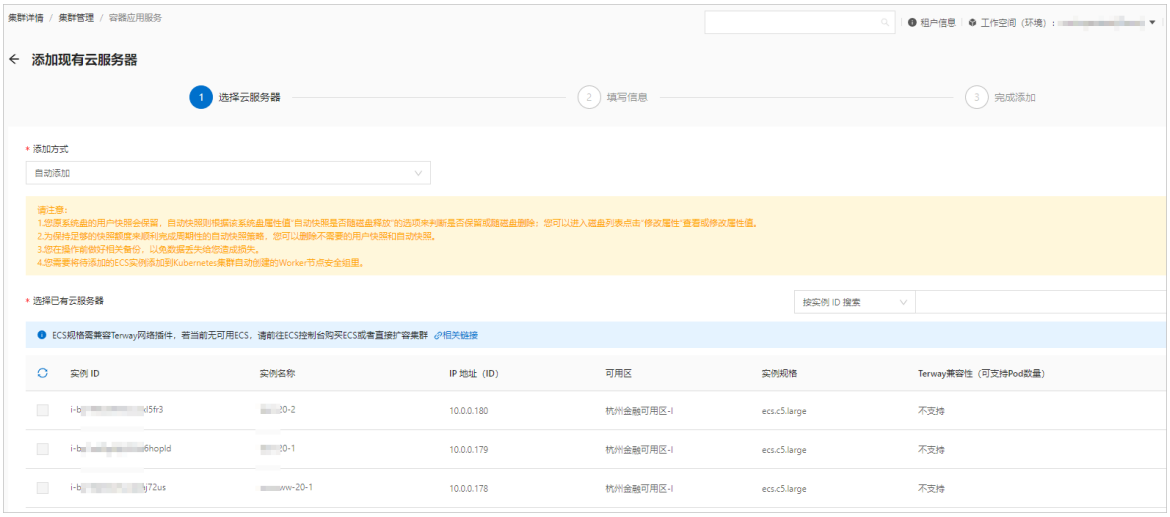
3. 在 **集群扩容** 页面，配置以下节点信息：
   - **虚拟交换机**：选择 ECS 机器所关联的交换机。
   - **付费类型**：支持 **按量付费** 和 **包年包月** 两种节点付费类型。选择 **包年包月** 时，需设置以下参数。
     - **购买时长**：目前支持选择 1、2、3、6 个月和 1~5 年。
     - **自动续费**：设置是否自动续费。
   - **伸缩数量（扩容）**：设置要扩容的机器数量，默认为 1。
   - **实例规格**：支持选择多个实例规格。详情请参见 实例规格族。
   - **系统盘**：支持 SSD 云盘和高效云盘。至少 40 GB。
   - **挂载磁盘**：支持 SSD 云盘和高效云盘。至少 40 GB。
   - **密码**：设置节点的登录密码。8～30 个字符，且同时包含三项（大写字母、小写字母、数字和特殊符号）。若遗忘可登录 ECS 控制台重置密码。
   - **云监控插件**：选择是否在 ECS 节点上安装云监控插件，安装后可以在云监控控制台查看所创建 ECS 实例的监控信息。
   - **标签**：为节点绑定标签，可通过标签来筛选节点。输入键和对应的值，单击 **添加一行数据**，在弹出框中输入 key 和 value。
   - **污点（Taints）**：为 Worker 节点添加污点。单击 **添加一行数据**，在弹出框中输入变量名、值，选择 Effect。更多详情参见 官方文档。
   - **部署单元**：保持默认配置。
   - **统一接入节点**：勾选后，扩容的节点将用于网络统一接入集群节点资源。
4. 完成配置后，单击 **提交**。
   在集群详情页面，如果集群状态显示扩容中，则说明节点正在扩容中。扩容完成后，集群状态显示为运行中。

添加已有节点

您可以将工作空间中已有的处于 **运行中** 状态的 ECS 加入到集群。系统会从当前工作空间中筛选出之前以 **清除机器环境** 方式删除的 ECS，以及从当前工作空间中筛选出符合条件的 ECS。

> 说明
> 您需要将待添加的 ECS 实例添加到集群自动创建的节点安全组内。

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理**。
2. 在 **集群详情** 页，单击 **节点** 页签，进入节点列表页面。
3. 选择目标节点，单击 **添加已有节点**，完成以下 **集群扩容** 配置。



   - **添加方式**：默认为自动添加。自动添加节点方式会列出当前账号下可用的 ECS 云服务器，在 Web 界面进行安装部署，并自动添加到集群。
   - **选择已有云服务器**：勾选要添加的 ECS。
4. 单击 **下一步**，填写以下实例信息。
   - **数据盘挂载**：设置是否将容器和镜像存储在数据盘。

     > 说明
     > - 如果 ECS 已挂载数据盘，且最后一块数据盘的文件系统未初始化，系统会自动将该数据盘格式化为 ext4，用来存放内容/var/lib/docker、/var/lib/kubelet 。数据盘内原有数据将丢失，请注意备份数据。
     > - 如果 ECS 未挂载数据盘，则不会挂载新的数据盘。

   - **CPU Policy**：设置 CPU Policy。
     - **none**：默认策略，表示启用现有的默认 CPU 亲和方案。
     - **static**：允许为节点上具有某些资源特征的 Pod 赋予增强的 CPU 亲和性和独占性。
   - **密码**：设置节点的登录密码。
   - **确认密码**：确认密码。
   - **部署单元**：使用默认配置。
   - **标签**：为节点绑定标签。输入键和对应的值，单击 **添加**，在弹出框中输入 **变量名**、**值**。
5. 完成配置后，单击 **提交**。

9.4.5. 查看集群详情

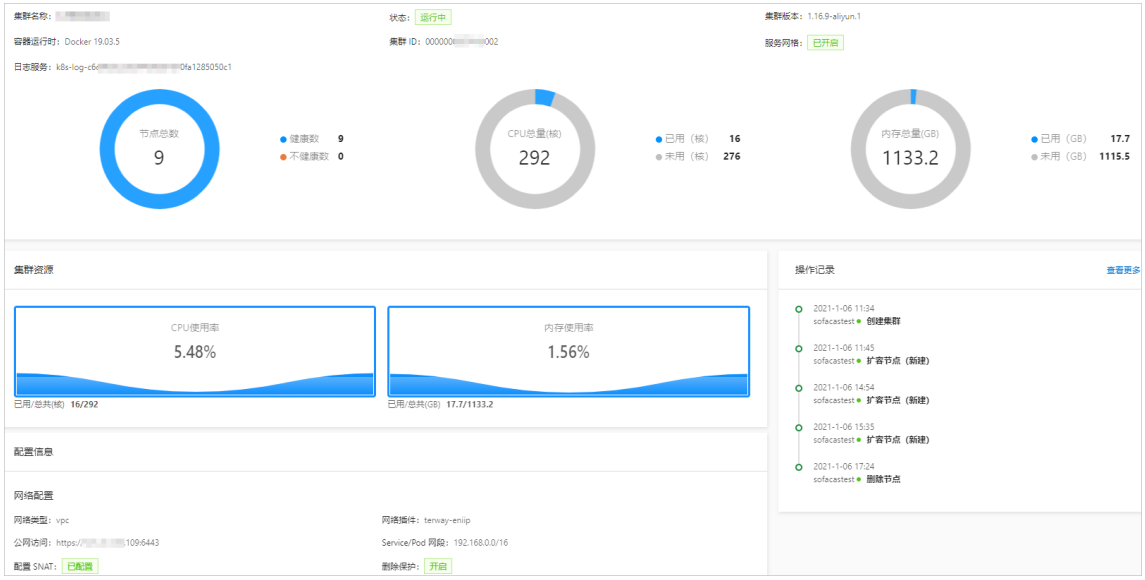
您可以在 LHC 控制台上查看租户下所有的集群信息。

前提条件

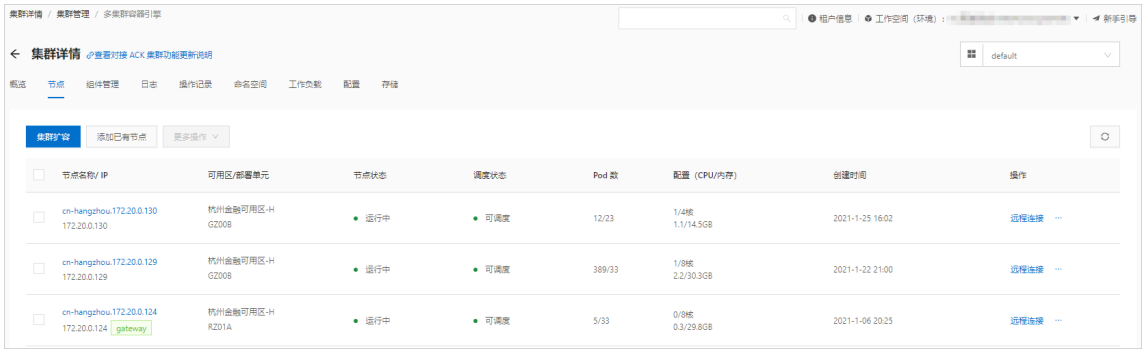
已创建集群 或 已导入集群

操作步骤

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击某个集群名称，进入 **集群详情** 页，查看以下信息：
   - **概览**：
     - **容器集群**：包含集群的基本信息，例如：集群名称、集群 ID、集群版本、网络类型、容器运行时、服务网格开关状态。日志服务、节点总数（健康数、不健康数）、CPU 总核数（已用、未用）、内存总量（已用、未用）。
     - **集群资源**：CPU 使用率、内存使用率。
     - **配置信息**：网络配置包含：网络类型、网络插件、Service/Pod 网段、公网访问地址、配置 SNAT 状态、删除保护状态。
     - **操作记录**：集群操作记录概览。



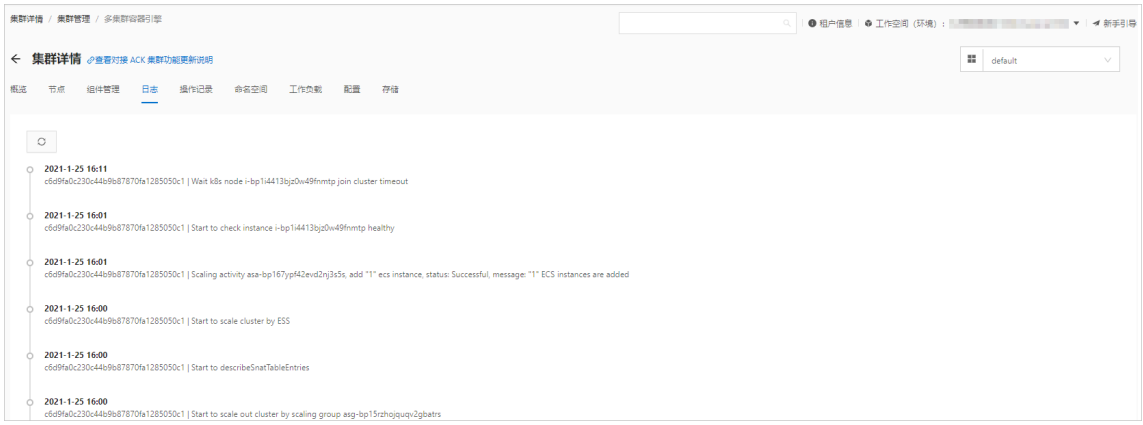
   - **节点**：包含节点名称/ IP、可用区/部署单元、节点状态、调度状态、Pod 数、配置（CPU/内存）、创建时间等信息。



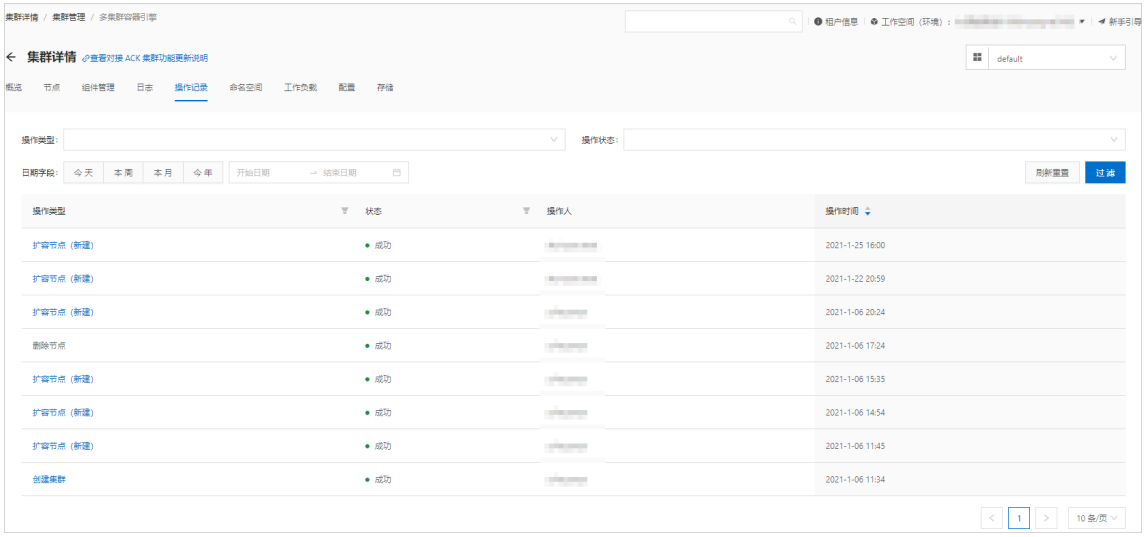
   - **组件管理**：集群的组件管理，主要分为系统组件和可选组件。详情参见 管理组件。



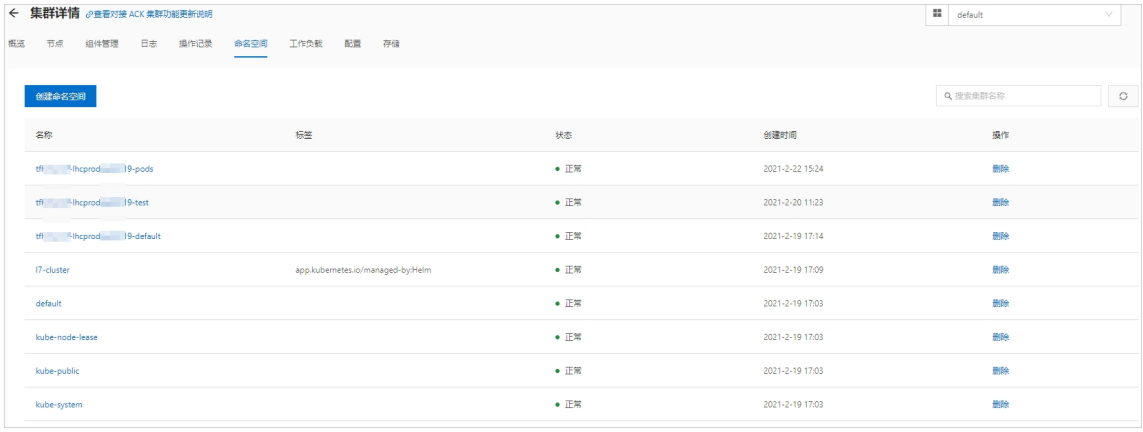
- **日志**：查看集群日志记录。



- **操作记录**：查看集群的历史操作记录。可根据操作类型、操作状态、日期字段进行筛选查找。



- **命名空间**：可以查看命名空间名称、状态等信息。



> **说明**
> 命名空间相关操作，请参见 创建命名空间。

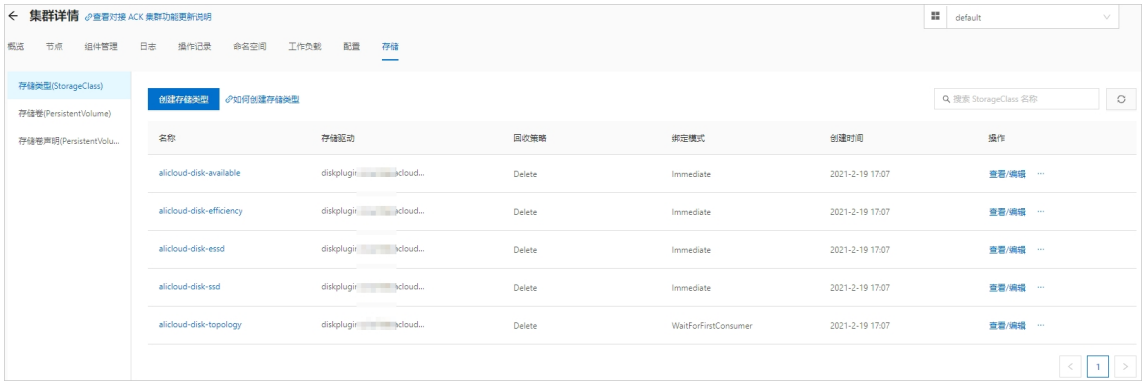
- **工作负载**：即 Kubernetes 对一组 Pod 的抽象模型，包括 Deployments、Statefulsets、Daemonsets 等多种类型。在该页面可以查看工作负载名称、镜像、Pod 实际数目等信息。

  > **说明**
  > 工作负载相关操作，请分别参见 部署（Deployments）、有状态副本集（Statefulsets）、守护进程集（Daemonsets）、容器组（Pods）。

- **配置**：可以查看名称、命名空间、创建时间等信息。

  > **说明**
  > 配置相关操作，请分别参见 管理配置项、管理保密字典。

- **存储**：为工作负载提供存储的组件。可以查看 Pod 信息、事件类型、事件原因、事件内容等信息。



> **说明**
> 存储相关操作，请分别参见 存储类型（StorageClass）、存储卷（PersistentVolume）、存储卷声明（PersistentVolumeClaim）。

9.4.6. 解绑集群

本文介绍如何将集群与 LHC 控制台解绑。解绑集群后，您将无法在 LHC 控制台上管理该集群，但该集群不会被删除。

> 说明
> - 解绑集群是高危操作，会导致当前环境不可用，请谨慎操作。
> - 集群解绑后如想继续使用，可通过集群导入操作将集群加入当前工作空间，详情请参见 导入集群。

前提条件

- 已导入集群。
- 您拥有 LHC 管理员权限。

操作步骤

1. 登录 LHC 控制台。
2. 在左侧导航栏中选择 **集群管理 > 集群详情**，进入集群列表页面。
3. 单击某个集群卡片中的 **解绑集群**。



4. 在 **解绑集群** 弹框中，输入 **解绑集群** 完成校验。
5. 单击 **确认**。

9.4.7. 节点管理

9.4.7.1. 节点调度

您可以通过控制台设置节点调度状态，从而合理分配各节点的负载。

前提条件

已创建集群 或 已导入集群

暂停调度

节点暂停调度后，在后续进行应用部署时，Pod 不会再调度到该节点。

1. 登录控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击目标集群名称，进入该集群的 **集群详情** 页。
3. 单击 **节点** 页签，进入节点列表页面。
4. 勾选目标节点，单击 **更多操作 > 暂停调度**，在弹出框中单击 **确定**。

返回到节点详情页，可以看到节点状态将变为 **暂停调度**。



恢复调度

1. 登录控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击目标集群名称，进入该集群的 **集群详情** 页。
3. 单击 **节点** 页签，进入节点列表页面。
4. 勾选目标节点，单击 **更多操作 > 恢复调度**，在弹出框中单击 **确定**。
   返回到节点详情页，可以看到节点状态将变为 **可调度**。



9.4.7.2. 管理节点标签

节点标签可以给节点定义不同的属性，通过这些标签可以快速的了解各个节点的特点，便于节点的管理和工作负载的调度。

前提条件

已创建集群 或 已导入集群

添加标签

1. 登录控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击某个集群卡片，进入 **集群详情** 页。
3. 单击 **节点** 页签，进入节点列表页面。
4. 选择节点，单击 **更多操作 > 编辑标签**。
5. 在 **添加标签** 页面，单击 **添加一行数据**，输入变量名和值，单击 **提交**。

删除标签

删除标签只能够删除用户自定义的标签，节点固有的标签不能删除。

1. 登录控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击某个集群卡片，进入 **集群详情** 页。
3. 单击 **节点** 页签，进入节点列表页面。
4. 选择节点，单击 **更多操作 > 编辑标签**。
5. 在 **添加标签** 页面，单击标签右侧的 **删除**。

9.4.7.3. 管理 Taint

Taint（污点）用于给节点做污点标记，使 Pod 排斥一类特定的节点，每个节点上都可以应用一个或多个污点。本部分介绍如何添加和删除 Taint。

前提条件

已创建集群 或 已导入集群

添加 Taint

1. 登录控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击某个集群卡片，进入 **集群详情** 页。
3. 单击 **节点** 页签，进入节点列表页面。
4. 选择节点，单击 **更多操作 > 编辑 Taint**。
5. 在 Taint 列表中，单击 **添加一行数据**，在弹出框中输入 **键**、**值**，选择 **Effect**，单击 **确定**。Effect 包含以下选项：
   - **NoSchedule**：不允许无匹配 toleration 的 Pod 调度到该节点。
   - **NoExecute**：若无匹配 toleration 的 Pod 已经在节点上运行，则将 Pod 驱逐，若无匹配 toleration 的 Pod 尚未在节点上运行，则不会将 Pod 调度到该节点上。
   - **PreferNoSchedule**：尽量不将无匹配 toleration 的 Pod 调度到该节点。

更多详情参见 官方文档。

删除 Taint

1. 登录控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击某个集群卡片，进入 **集群详情** 页。
3. 单击 **节点** 页签，进入节点列表页面。
4. 选择节点，单击 **更多操作 > 编辑 Taint**。
5. 在 Taint 列表中，单击 Taint 右侧的 **删除**。

9.4.7.4. 查看节点详情

集群创建或导入完成后，如需查看集群中节点的详细信息，您可以通过 LHC 控制台查看节点信息，包括节点服务器 IP、节点创建时间、Pod 列表以及节点事件等信息。

> 说明
> 从对接 ACK 集群的功能上线时起（2020 年 8 月 15 号），阿里云底座上新建集群默认为 ACK 托管版集群。

前提条件

已创建集群 或 已导入集群

操作步骤

1. 单击某个集群名称，进入 **集群详情** 页。
2. 单击 **节点** 页签，进入 **节点列表** 页面。
3. 在 **节点列表** 中，单击 **节点名称** 进入节点详情页面。您可以查看以下信息。

信息模块	具体信息
基本信息	节点服务器 IP
	节点的创建时间
	调度状态： ○ 可调度状态 ○ 不可调度状态
	机器详情：单击详情链接跳转至工作空间资源管理页面，可查看 ECS 详情。
	注释：系统注释信息
	系统定义的节点标签，可用于节点分组。
Pod 列表	节点所包含的容器列表，包括 Pod 名称、IP、命名空间、 状态、创建时间。 单击 Pod 名称跳转到 Pod 详情页。
节点事件	反映节点生命周期和运行状况的事件日志，包含事件类型、事件原因、事件内容。 事件类型分为 Normal、Warning、Error。您可以查看 Warning 和 Error 类型事件进行故障排查。
YAML	节点的 YAML 文件，可以一键复制。

9.4.7.5. 删除节点

当您不再需要某个节点继续工作时，可以将其移除。本文介绍如何通过 LHC 控制台删除节点。

> 说明
> 从对接 ACK 集群的功能上线时起（2020 年 8 月 15 号），阿里云底座上新建集群默认为 ACK 托管版集群。

前提条件

已创建集群 或 已导入集群

> 注意
> - 删除节点会自动驱逐 Pod 引起 Pod 迁移，可能会影响业务，建议在业务低峰期操作。
> - 操作过程中可能存在非预期风险，请提前做好相关的数据备份。

操作步骤

1. 在左侧导航栏，单击 **集群管理** > **集群详情**，进入集群列表页面。
2. 单击集群名称进入该集群的详情页，单击 **节点** 页签，进入节点列表页面。

3. 勾选待删除节点，单击 **更多操作 > 删除**。
4. 在弹出窗口中，选择以下删除方式，单击 **确定**。
   - **自动排空节点（drain）**：把待移除节点上的 Pod 转移到其他节点。请确保集群其他节点的资源充足。您还可以通过执行命令 `kubectl drain node-name` 的方式把待移除节点上的 Pod 转移到其他节点。

     > 说明
     > node-name 格式为 `your-region-name.node-id` 。例如 `cn-hangzhou.i-xxx` 。
     > `your-region-name` 为您集群所在的地域名称。node-id 为待移除节点所在的 ECS 实例 ID。

   - **同时释放 ECS**：彻底删除服务器。
     - 释放 ECS 实例仅释放按量付费的 ECS 实例。

       > 说明
       > - 如果节点上存在 Pod，请选择 **自动排空节点**，并确保集群资源充足，Pod 将自动迁移到其他节点。
       > - 释放 ECS 仅释放按量付费类型的 ECS 实例，不被释放的 ECS 会继续收费。
       > - 移除节点是异步操作，请在节点列表中查看移除进度。

     - 对于预付费 ECS 实例，计费周期到期后，ECS 实例会自动释放。
     - 您也可以在 ECS 实例到期前：
       - 申请退款，提前释放实例，请参见 退款规则及退款流程。
       - 将计费方式转为按量付费后释放实例，请参见 包年包月转按量付费。
       - 若不选择同时释放 ECS，该节点所在的 ECS 实例会继续计费。

         > 说明
         > - 如果节点上存在 Pod，请选择 **自动排空节点**，并确保集群资源充足，Pod 将自动迁移到其他节点。
         > - 释放 ECS 仅释放按量付费类型的 ECS 实例，不被释放的 ECS 会继续收费。
         > - 移除节点是异步操作，请在节点列表中查看移除进度。

9.4.8. 命名空间

9.4.8.1. 创建命名空间

命名空间（Namespace）是对一组资源和对象的抽象整合。您可以跨集群或者在同一个集群内创建不同的命名空间，不同命名空间中的数据彼此隔离，使得它们既可以共享服务，也能够互不干扰。例如，将开发环境、测试环境的业务分别放在不同的命名空间。

命名空间类别

LHC 的命名空间分为以下两类：

- 联邦命名空间（Federated Namespace）：跨集群的命名空间，需要自定义。
- 单集群命名空间（Namespace）：在单个集群中，您可以按照需要创建命名空间，例如开发环境、联调环境和测试环境分别创建对应的命名空间。或者按照不同的业务创建对应的命名空间，例如，系统分为登录和游戏，可以分别创建对应命名空间。LHC 默认会创建以下集群内的系统命名空间。

- **default**：在不指定命名空间时，默认使用 default。
- **kube-public**：用来部署公共插件、容器模板等。
- **kube-system**：部署系统组件。
- **sigma-operator-machine-conditions**：节点运维所需的系统命名空间。
- **operator-extensions**：节点运维所需的系统名称空间。
- **antstack**：系统应用所在的命名空间。仅存在于 AntStack 底座的专有云上。

前提条件

已创建集群 或 已导入集群

创建联邦命名空间（Federated Namespace）

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 命名空间**，进入命名空间列表页。
2. 单击 **创建** 进入 **创建命名空间** 页面，输入以下信息，单击 **创建**。
   - **名称**：输入命名空间名称。默认名称的前缀为 `租户名-单元化工作空间标识-` 。
   - **添加标签**：为命名空间添加自定义标签。

创建单集群命名空间（Namespace）

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击某个集群名称，进入 **集群详情** 页，单击 **命名空间** 页签。
3. 单击 **创建命名空间**，在 **创建命名空间** 页面，输入以下配置信息：
   - **名称**：命名空间名称。
   - **标签**：为命名空间添加自定义标签。
4. 完成配置后，单击 **提交**。

9.4.8.2. 查看命名空间详情

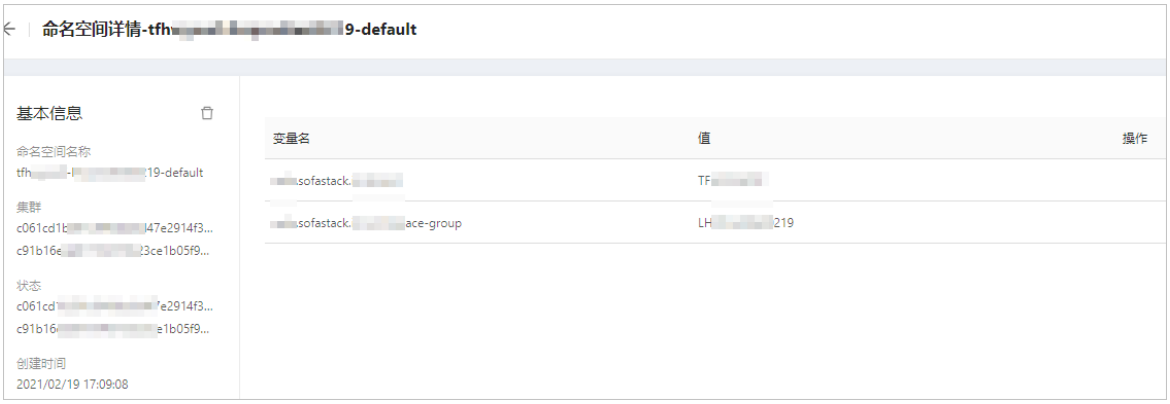
您可以通过控制台查看命名空间详情。

查看联邦命名空间

1. 登录控制台，在左侧导航栏单击 **集群管理 > 命名空间**，进入命名空间列表页。



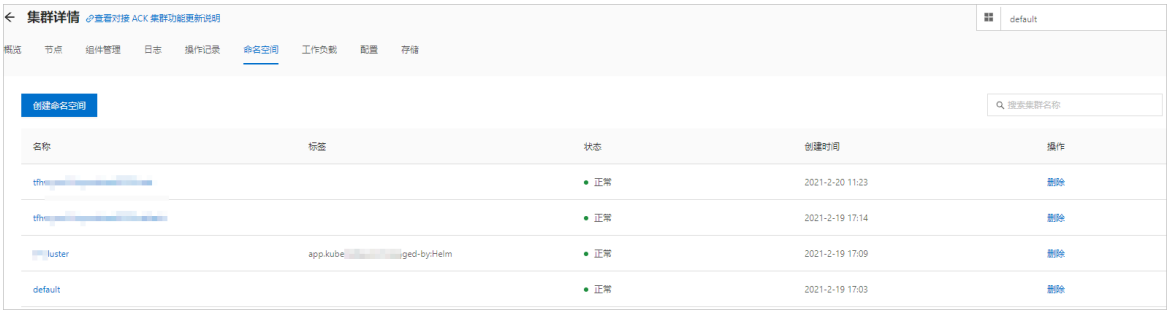
2. 单击命名空间名称，进入 **命名空间详情** 页。您可以查看如下信息：



- 基本信息
  - 命名空间名称
  - 集群
  - 状态
  - 创建时间
- 标签信息
  - 变量名
  - 值

查看单集群命名空间

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击某个集群名称，进入 **集群详情** 页，单击 **命名空间** 页签。



3. 单击命名空间名称，进入命名空间详情页。您可以查看如下信息：



- 基本信息
  - 命名空间名称
  - 状态

    - 创建时间
  - 标签信息
    - 变量名
    - 值

9.4.8.3. 删除命名空间

> 注意
> - 删除命名空间将会释放该命名空间下的所有资源。
> - 请勿删除系统命名空间。

删除联邦命名空间

1. 登录控制台，在左侧导航栏单击 **集群管理 > 命名空间**，进入命名空间列表页。
2. 单击命名空间右侧 **操作** 栏中的 **删除**， 在弹出的对话框中单击 **确认**。系统会弹出删除命名空间成功的提示。

删除单集群命名空间

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击某个集群名称，进入 **集群详情** 页，单击 **命名空间** 页签。
3. 单击命名空间右侧 **操作** 栏中的 **删除**， 在弹出的对话框中单击 **确认**。
   系统会弹出删除命名空间成功的提示。

9.4.8.4. 设置资源配额

通过设置命名空间级资源配额，实现多团队或多用户在共享集群资源的情况下限制团队或用户可以使用的资源总量，包括限制命名空间下创建某一类型对象的数量以及对象消耗计算资源（CPU、内存）的总量。本文介绍如何通过控制台设置资源配额。

> 说明
> 目前，仅单集群命名空间支持设置资源配额。

背景信息

默认情况下，运行中的 Pod 可无限制使用节点（Node）上的 CPU 和内存，即任何一个 Pod 都可以无节制地使用集群的计算资源，导致某个命名空间的 Pod 可能会耗尽集群的资源。
命名空间的一个重要的作用是充当一个虚拟的集群，用于多种工作用途，满足多用户的使用需求。因此，为命名空间配置资源额度是一种最佳实践。您可以为命名空间配置包括 CPU、内存、Pod 数量等资源的额度。
更多信息请参见 Resource Quotas。

工作原理

资源配额（Resource Quotas）是对命名空间进行资源配额，限制资源使用的一种策略。

- 资源配额应用在命名空间上，可以为每个命名空间创建一个或多个资源配额对象。
- 开启计算机资源限制（比如 CPU 或内存限制）后，创建容器组时必须配置计算资源请求或限制。
- 若使用的资源用量超过设定的资源配额，将禁止创建新的资源。

前提条件

- 已导入集群

- 已创建单集群命名空间

操作步骤

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击目标集群名称，进入 **集群详情** 页，单击 **命名空间** 页签。
3. 单击目标命名空间右侧 **操作** 栏中的 **资源配额**。
4. 在 **资源配额** 对话框中，根据需要完成资源配额配置。如果不勾选，则代表不限制该资源的使用。
   i. 完成计算机资源限制配置。

   > 说明
   > 对命名空间设置 CPU 或内存配额后，创建容器组时，必须指定 CPU 或内存资源限制。详情请参见 Resource Quotas。

配置项	说明	单位
CPU 限制	用于限制命名空间下 Pod 能申请到的 CPU 资源的最大值。	核
内存限制	用于限制命名空间下 Pod 能申请到的内存资源的最大值。	Mi、Gi、Ti、M、G、T

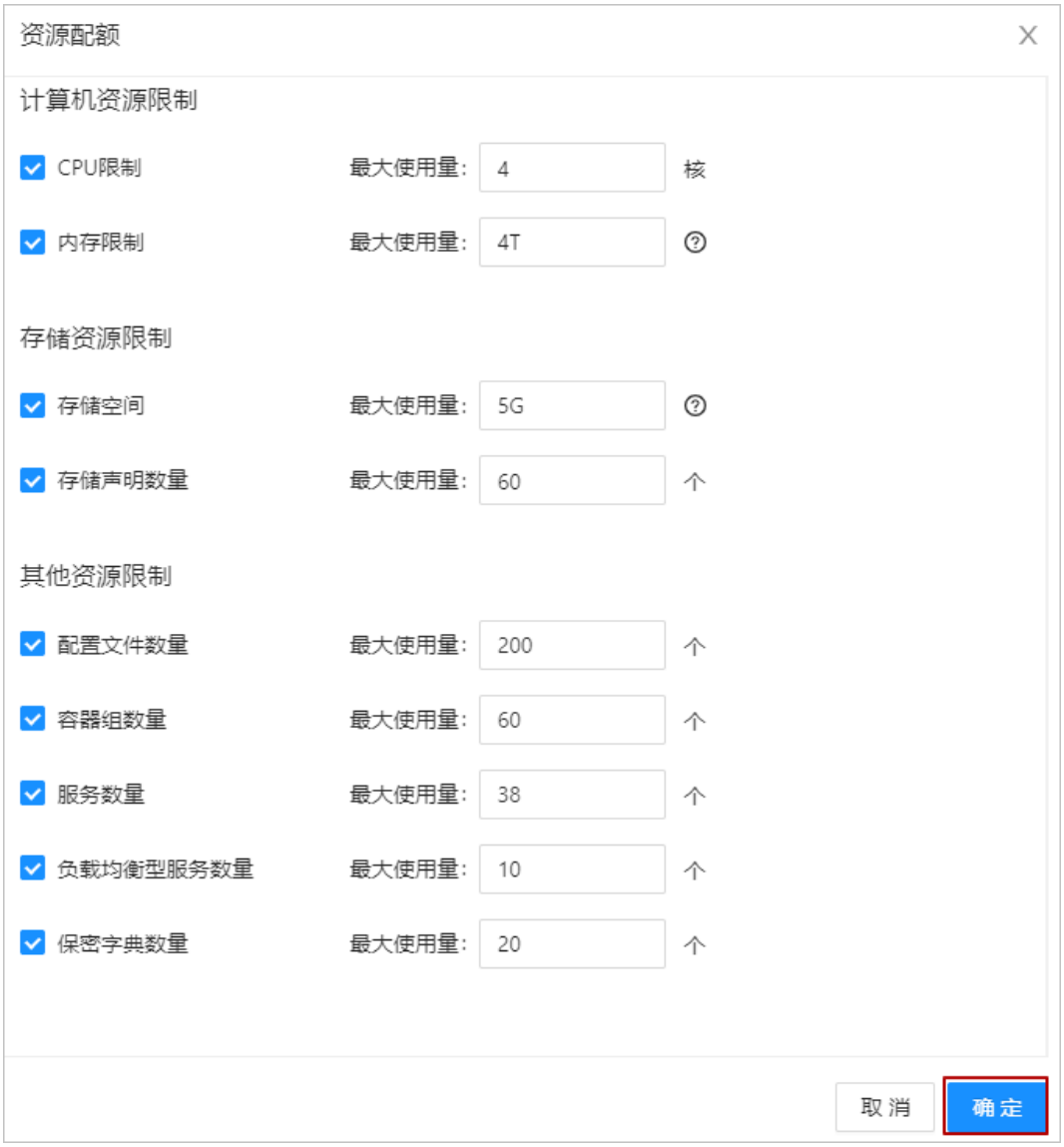
   ii. 完成存储资源限制配置。

配置项	说明	单位
存储空间	用于限制命名空间下允许的存储空间上限。	Mi、Gi、Ti、M、G、T
存储声明数量	用于限制命名空间下允许的存储声明的最大数量。	个

   iii. 完成其他资源限制配置。

配置项	说明	单位
配置文件数量	该命名空间下允许存在的配置文件数量上限。	个
容器组数量	该命名空间下允许存在的容器组数量上限。	个
服务数量	该命名空间下允许存在的服务数量上限。	个
负载均衡型服务数量	该命名空间下允许存在的负载均衡型服务数量上限。	个
保密字典数量	该命名空间下允许存在的保密字典数量上限。	个

5. 完成资源配额配置后，单击 **确定**。



9.4.9. 管理组件

集群 AddOn 组件是单元化应用服务提供的扩展功能包，用于管理集群的扩展功能。您可以根据业务需求选择所需的组件。

安装组件

前提条件

已创建集群 或 已导入集群

操作步骤

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击目标集群卡片，进入 **集群详情** 页。
3. 单击 **组件管理** 页签，进入组件列表页面。
4. 找到需要安装的组件，单击所需组件卡片右下角的 **安装**。关于组件说明可参考下表。

组件名称	功能	是否必选	是否推荐安装
apiserver-network-proxy	用作集群 kube-apiserver 代理。	是	创建集群过程中会自动安装。
cafeextcontroller	CAFE 发布部署核心组件，用于应用发布部署。	是	创建或导入集群过程中会自动安装。
cloud-controller-manager	使用 Cloud Controller Manager 为 K8s 应用创建负载均衡及统一接入路由管理。	是	创建或导入集群过程中会默认安装。
pontus-pontusagent	监控 Agent 组件，Daemonset 方式部署。	否	使用 监控产品 RMS 时必须安装，否则无需安装。
chaosblade-operator	面向云原生领域的混沌实验注入工具，配合金融云高可用管理平台使用，可以实现针对基础资源、用户应用、蚂蚁产品的故障演练。	否	使用 高可用管理产品 HAS 时必须安装，否则无需安装。
mosn-ingress	用于 Ingress 路由配置。	是	创建集群过程中会自动安装。

执行结果

组件安装完成后，卡片右上角的状态会变为 **已安装**。



重装组件

当组件运行异常时，您可以使用重装功能。

操作步骤

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击目标集群卡片，进入 **集群详情** 页。
3. 单击 **组件管理** 页签，进入组件列表页面。
4. 找到需要重装的组件，单击组件卡片的 **重装 > 确定**。

升级组件

> 注意
>
> - 请务必通过 **组件管理 > 升级** 对集群组件进行升级，否则可能会带来不可预知的问题。
> - 升级组件前，请确保集群内所有节点均处于运行中状态，否则可能会导致组件升级失败。
> - 组件升级过程中，请勿发布应用，可能出现非预期问题。

操作步骤

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。

2. 单击目标集群卡片，进入 **集群详情** 页。
3. 单击 **组件管理** 页签，进入组件列表页面。
4. 找到待升级的组件，单击组件卡片的 **升级**。

9.5. 应用服务

9.5.1. 创建应用服务

LHC 应用服务对 Kubernetes 原生的 Deployment 做了能力增强，您可以通过创建应用服务定义容器服务的基本信息、访问策略、发布及调度策略等信息，为后续容器服务的部署做准备。

前提条件

- 已创建应用
- 已创建集群 或 已导入集群

创建应用服务的过程分为以下 6 个步骤：

1. 填写基本信息
2. Pod 模板配置
3. 弹性配置
4. 访问配置
5. 部署和调度配置
6. 预览并提交

一、填写基本信息

1. 登录控制台，在左侧导航栏单击 **发布运维 > 应用服务**。
2. 在应用服务列表页，单击 **创建应用服务**。
3. 在 **创建应用服务** 页面，填写以下基本信息，单击 **下一步**。
   - **命名空间**：选择命名空间，默认为列表第一个。
   - **应用服务名称**：容器服务的名称。服务实例名称允许包含（小写）字母、数字、连字符，且必须以字母开头，以字母或数字结尾。同一个工作空间组下不允许同名。
   - **所属应用**：选择一个该容器服务所关联的应用。
   - **有状态模式**：默认关闭。开启后，Pod 名称会加上 `-0` 、 `-1` ... 数字后缀，重建过程中不会生成新 Pod 名称。相关操作，请参见 LHC Pod 域名配置。
   - **数据卷模版配置**：可选，用于挂载存储。单击 **添加配置** 可设置多个。具体配置项如下：
     - **名称**：输入模板名称。允许包含字母、数字、连字符，且必须以字母开头，以字母或数字结尾。
     - **存储类型**：用于指定支持的存储类别。单击 **创建存储类型** 将跳转至存储页。

       > 说明
       > - 对于存储类型，通常由系统管理员定义，具体类别因存储 SLA、备份策略等而异。
       > - 对于多集群下的存储类型，需要多集群下都存在此存储类型才可以被调用。

- **容量**：设置数据卷模板的容量。单位为 GiB。

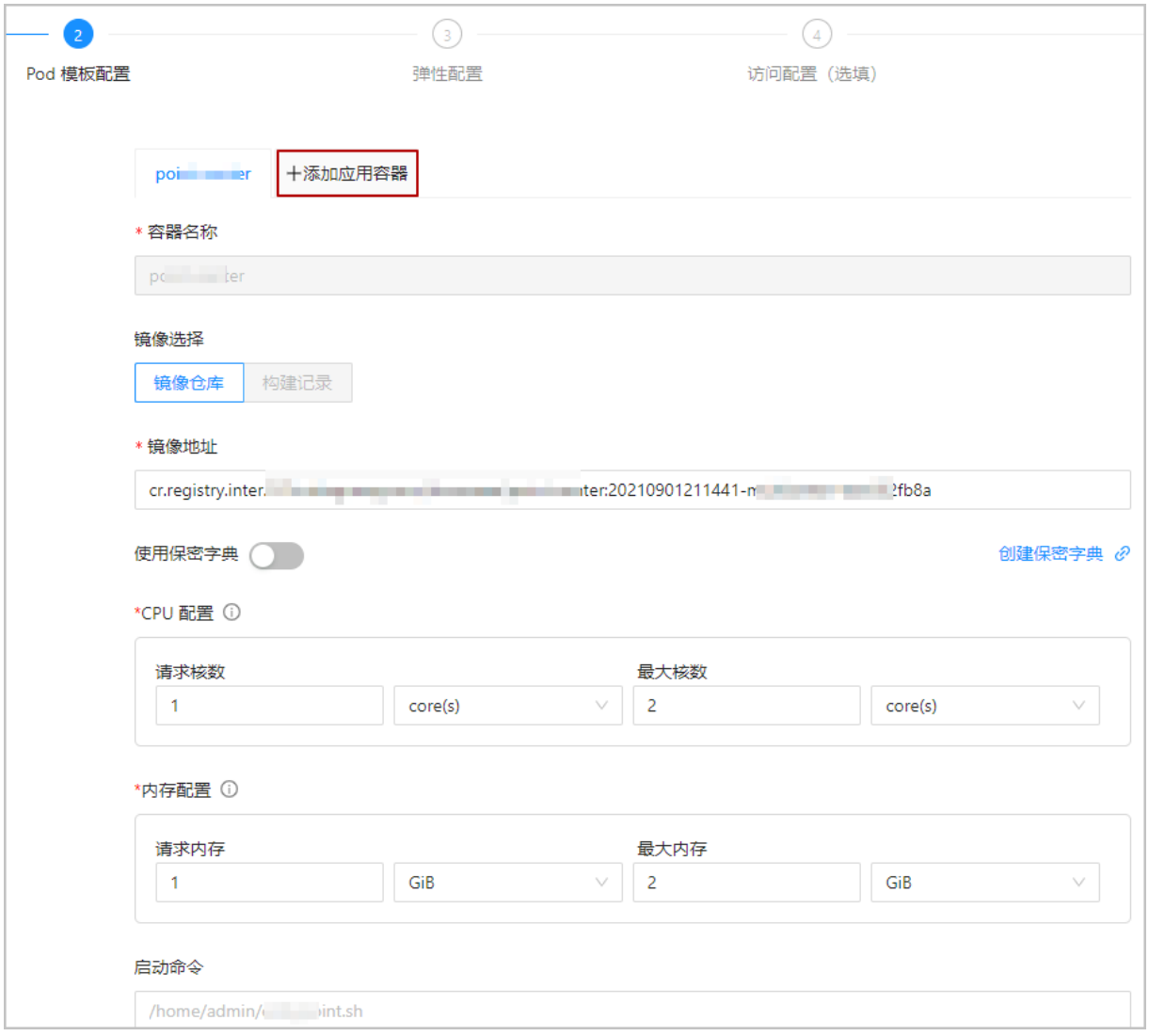
> **说明**
> 数据卷容量大小受设置的 PV 容量的限制。关于 PV 容量设置，请参见 创建存储卷。

- **描述**：选填。容器服务的描述。

二、Pod 模板配置

对 Pod 里的容器做详细的配置。Pod 与容器之间的比例并非 1:1，可通过单击 **添加应用容器** 为单 Pod 配置多个容器。

> **说明**
> 对于可添加的容器数量，暂无限制。



基本选项

- **容器名称**：容器名称。
- **镜像来源**：支持 **镜像仓库**、**构建记录** 两种方式。

- **镜像仓库**：直接输入镜像仓库地址，比如：`registry-cnhz.cloud.alipay.com/aks/nginx:1.8`。

  > 说明
  > 如果配置多个容器，输入的镜像地址不得相同。

- **构建记录**：选择该类型来源时，会自动关联应用服务所属应用的构建，具体参见 镜像构建。

  > 说明
  > 目前暂不支持 **构建记录**。如有需要，请提交工单进行处理。

- **使用保密字典**：若该镜像属于私有镜像，需要访问认证，您可以开启 **使用保密字典** 拉取镜像。若尚未创建保密字典，可单击创建保密字典，详情请参见 创建保密字典。

- **CPU 配置**：设置容器使用的 CPU 的数量。**请求核数** 为能保证的最小核数，**最大核数** 为能使用的最大核数。换算方式：1 core = 1000 millicores。
- **内存配置**：设置容器使用的内存的数量。**请求内存** 为能保证的最小内存数量，**最大内存** 为能使用的最大内存数量。换算方式：1024 Bytes = 1 KiB；1024 KiB = 1 MiB；1024 MiB = 1 GiB；1024 GiB = 1 TiB。
- **启动命令**：选填。用于指定容器启动时执行的命令。

高级配置

默认关闭 。如果开启 ，需要完成以下选项配置。

- **hostNetwork 配置**：默认关闭。开启后，Pod IP 地址与节点 IP 地址保持一致。在 host 模式下，一个节点最多只能挂载 1 个 Pod。
- **环境变量配置**：设置在容器启动时传入应用进程的键值对，例如：`USER=tester`。
- **数据卷配置**：配置容器使用的数据卷，目前仅支持挂在当前容器所在宿主机的目录。
- **健康检查配置**：包含两种检查机制：Readiness 和 Liveness。具体参见 Kubernetes Probe。
- **生命周期事件回调配置**：为容器添加生命周期事件回调，分别在容器启动后和容器停止前执行。
- **日志服务配置**：配置日志服务（SLS），可选择已有日志库或创建新的日志库。

  > 说明
  > 首次使用日志服务配置功能，需前往 日志服务控制台 开启配置索引。详情请参见 配置索引。

更多信息参见 高级配置。

配置覆盖

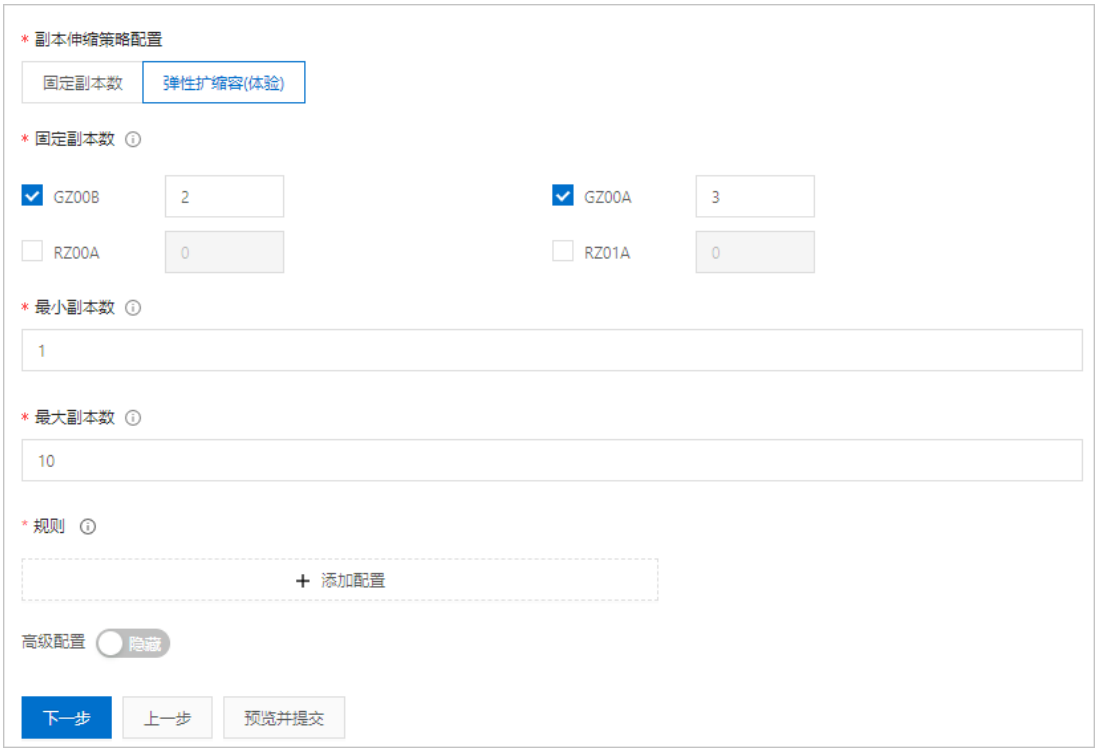
默认关闭。如果开启，需要完成以下选项配置。

- **环境变量配置**：您可以为每个部署单元单独设置环境变量，覆盖应用服务中的默认配置，但仅针对于手动输入类型的环境变量。
- **镜像地址**：您可以为每个部署单元单独配置不同的镜像地址，覆盖应用服务中的默认配置。

三、弹性配置

副本伸缩策略配置：支持以下两种弹性策略。

- **固定副本数**：默认为 0，可勾选部署单元并修改为期望副本数，即应用服务运行时保持固定数目的 Pod 副本。
- **弹性扩缩容**：可对应用服务自动伸缩进行如下弹性规则配置。



- **固定副本数**：默认为 0，可勾选部署单元并修改为期望副本数。
- **最小副本数**：弹性扩缩容时 Pod 的最小副本数。
- **最大副本数**：弹性扩缩容时 Pod 的最大副本数。
- **规则**：当前伸缩规则基于的指标类型。单击 **+ 添加配置**，配置以下选项。



  - **规则类型**：当前伸缩规则基于的规则类型，目前支持 CPU、Memory、QPS、响应时间。
  - **目标类型**：当前伸缩规则基于的指标类型所期望达到的值的类型，目前支持 **使用率** 与 **平均资源使用**。
  - **使用率/平均值**：输入使用率/平均值。
  - **单位**：无需输入，根据所选的规则类型自动匹配。

- **高级配置**：默认关闭 。如果开启，需要完成以下选项配置。



  - **幅度**：可以通过设置按百分比或指定个数来调整扩缩容的速率，单位为百分比或个数。
  - **间隔时间**：上次扩缩容操作与本次扩缩容操作之间的时间间隔，单位为秒。
  - **冷却时间**：冷却扩缩容操作的时间，即完成一次扩缩容操作之后不再次触发扩缩容操作的时间窗口，单位为秒。

> **说明**
> 一个扩缩容活动执行完成后，在设定的冷却时间内不执行其他扩缩容活动。

四、访问配置（选填）

应用服务支持统一接入及负载均衡两种访问方式。负载均衡是基于端口的请求负载均衡，统一接入是基于规则的请求负载均衡。您可以根据业务需要做好规划。

负载均衡

您可以在创建应用服务时设置访问方式，也可以应用服务创建完成后添加访问方式。

创建应用服务时设置访问方式

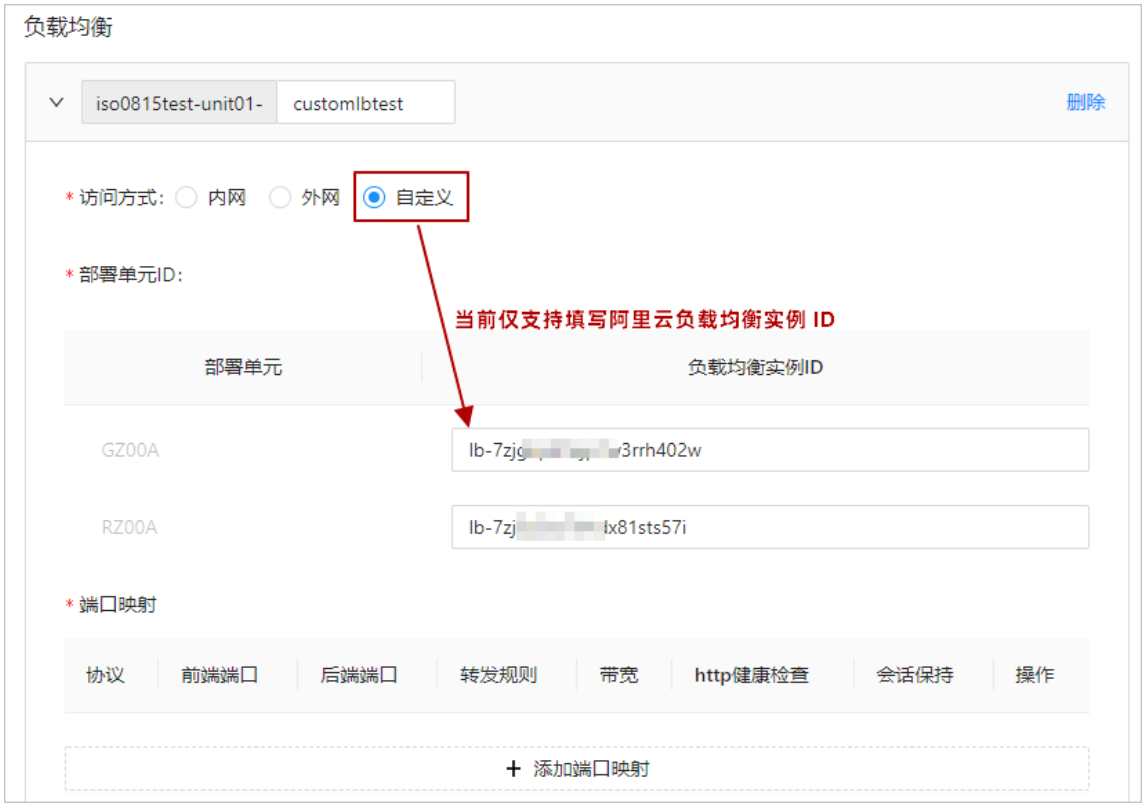
1. 在 **访问配置** 页面，单击 **添加负载均衡**。
2. 在 **负载均衡** 配置中，填写以下信息后，单击 **下一步**。
   - **负载均衡名称**：填写服务名称。系统默认生成服务名称前缀为 `应用服务名称-` 。服务名称允许包含（小写）字母、数字、连字符且必须以字母开头、以字母或者数字结尾。
   - **访问方式**：选择负载均衡服务的访问方式。支持以下选项：
     - **内网**：创建一个内网的 LoadBalancer，并将流量转发到容器的相应端口上。
     - **外网**：创建一个公网的 LoadBalancer，并将流量转发到容器的相应端口上。访问方式由公网负载均衡服务地址以及设置的访问端口组成，例如 `10.117.117.117:80` 。

- **自定义**：在自定义模式下，会在选定的部署单元内分别创建一个 LoadBalancer 类型的 Service，通过用户填入的 ID 复用已有的负载均衡实例。

  > **注意**
  > - 当前仅支持填入阿里云负载均衡实例 ID。更多合作伙伴型号支持，敬请期待。
  > - 阿里云场景下，被复用的 SLB 必须要满足如下条件：
  >   - 确保被复用的 SLB 实例是在 SLB 控制台手动创建的。
  >   - 确保期望使用的监听器端口尚未被占用。
  >   - 如果是内网地址类型的 SLB 实例，须确保与部署单元对应的集群同属于同一个 VPC。

- **部署单元 ID**：仅当访问方式选择 **自定义** 时显示该此选项。为各部署单元填写对应的负载均衡实例 ID。

  > **说明**
  > - 对于同一应用服务下的不同部署单元，支持设置相同的负载均衡实例 ID。
  > - 确保应用服务端口映射中配置的前端端口在已有负载均衡实例上尚未被其他应用服务占用，否则会因端口冲突而导致应用服务发布失败。



- **联邦负载均衡实例**：仅当访问方式选择 **内网** 或 **外网** 且开启 **使用联邦负载均衡** 时显示该此选项。从下拉列表中选择对应联邦负载均衡实例，可选择到的联邦负载均衡的部署单元必须能够覆盖应用服务 弹性配置 步骤中选择的所有部署单元。

> 说明
> - 如需创建新的联邦负载均衡实例，请参见 创建联邦负载均衡。
> - 如需修改联邦负载均衡配置的部署单元，请参见 配置联邦负载均衡。



- **端口映射**：单击 **添加端口映射**，填写以下信息。
  - **协议**：支持 TCP、 HTTP、HTTPS 协议。
  - **转发规则**：支持 RR 轮询、WLC 最小连接数。
  - **前端端口**：容器镜像中工作负载实际监听的端口，端口范围为 1~65535。
  - **后端端口**：容器端口映射到负载均衡实例上的端口，用负载均衡 IP 访问工作负载时使用，端口范围为 1~65535。
  - **带宽**：单位为 Mbps，默认值为 -1，表示无限制。支持 -1 或 1~5120 范围内的整数。
  - **HTTP 健康检查**：HTTP 模式的健康检查是通过发送 HEAD 或 GET 请求模拟浏览器的访问行为来检查后端 Pod 是否健康。若开启，需填写以下的配置项。

    > 说明
    > TCP 端口默认开启 TCP 健康检查，此处可按需开启额外 HTTP 健康检查。

    - **URL**：指定具体的检查路径。
    - **域名**：非必填。如果您的应用 Pod 需要校验请求的 host 字段，则需要配置相关域名，确保健康检查正常工作。若在健康检查中配置了域名，SLB 会将域名配置到 host 字段中去；反之，如果没有配置域名，SLB 则不会在请求中附带 host 字段，因此健康检查请求就会被服务器拒绝，可能导致健康检查失败。
    - **端口**：健康检查服务访问后端 Pod 时的探测端口。端口范围 1~65535。
    - **超时时间（s）**：接收响应需要等待的时间。如果后端 Pod 在指定的时间内没有正确响应，则判定为健康检查失败。默认为 3 秒。
    - **时间间隔（s）**：负载均衡的健康检查请求进行健康检查的时间间隔。默认值为 5 秒。
    - **http 健康状态码**：选择健康检查正常的 HTTP 状态码。默认值为 http_2xx。

- **会话保持**：开启会话保持后，负载均衡监听会把来自同一客户端的访问请求分发到同一台后端 Pod 上。
  HTTP 协议会话保持基于 Cookie。负载均衡提供了两种 Cookie 处理方式：
  - **植入 Cookie**：您只需要指定 Cookie 的过期时间。客户端第一次访问时，负载均衡会在返回请求中植入Cookie（即在 HTTP 或 HTTPS 响应报文中插入 SERVERID），下次客户端携带此 Cookie 访问，负载均衡服务会将请求定向转发给之前记录到的后端服务器上。
  - **重写 Cookie**：可以根据需要指定 HTTPS 或 HTTP 响应中插入的 Cookie。您需要在后端 Pod 上维护该 Cookie 的过期时间和生存时间。
    负载均衡服务发现用户自定义了 Cookie，将会对原来的 Cookie 进行重写，下次客户端携带新的 Cookie 访问，负载均衡服务会将请求定向转发给之前记录到的后端 Pod。详情请参见 会话保持规则配置。

统一接入

统一接入 支持配置转发规则，将流量转发到容器的相应端口上。您可以在创建应用服务时设置访问方式，也可以应用服务创建完成后添加访问方式。

创建应用服务时添加访问方式

1. 在 **访问配置** 页面，单击 **添加统一接入**。
2. 在 **统一接入** 配置中，填写以下信息后，单击 **下一步**。
   - **服务名称**：填写服务名称。系统默认生成服务名称前缀为 `应用服务名称-` 。服务名称允许包含（小写）字母、数字、连字符、且必须以字母开头、以字母或者数字结尾。
   - **所属统一接入实例**：选择已创建的统一接入实例。详情参见 创建统一接入实例。
   - **接入协议**：选择统一接入实例上支持的协议，一般有 **HTTP** 或 **HTTPS**。
   - **路由模式**：选择 **单元化** 或 **非单元化**。
     - **单元化** 代表为这个统一接入开启单元化转发的能力，即会根据请求 Cookie 内的特定 UID 进行单元化切片，每个 Zone 处理固定区间的用户请求。
     - **非单元化** 代表随机分发。
   - **统一接入规则**：单击 **添加统一接入规则** 转发规则，填写以下信息。
     - **域名**：填写需要转发的请求域名。
     - **后端端口**：选择容器镜像中工作负载程序实际监听的端口，端口范围为 1~65535。
     - **转发路径**：请求转发到的具体容器路径地址。默认为根目录。

> **注意**
> 添加或删除已有转发规则都会导致发布时线上服务不可用。

Headless Service

Headless Service 对应的每一个 Endpoint，即每一个 Pod，都会有对应的 DNS 域名。配置 Headless Service 名称，这样各 Pod 之间便可以通过域名互相访问。Headless Service 不会分配 Cluster IP，kube-proxy 不会处理此类 Service，但可以通过域名解析访问 Pod。相关操作，请参见 LHC Pod 域名配置。
您可以在创建应用服务时添加 Headless Service 名称，也可以应用服务创建完成后重新编辑应用服务进行添加。添加之后需要发布应用服务才会生效。

创建应用服务时添加 Headless Service

> **说明**
> 一个应用服务仅支持添加一个 Headless Service 。

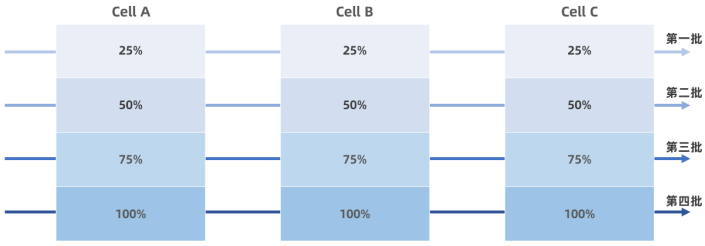
1. 在 **访问配置** 页面，单击 **添加 Headless Service**。
2. 在 **Headless Service** 配置中，填写 Headless Service 名称。名称由服务名称前缀、Headless Service 名称组成。
   - 前缀：系统默认生成服务名称前缀 `应用服务名称-` 。
   - Headless Service 名称：输入名称，允许包含小写字母、数字、连字符（-）。最大长度不超过 20 个字符。
3. 单击 **下一步**。

五、部署和调度配置（选填）

您可以自定义部署和调度配置，若不修改，发布应用服务时，系统会使用默认配置。
该配置项用于配置容器服务在部署时需要用到的信息，包括以下内容：

- **是否使用 Beta 验证**：发布时选取部分 Pod 先行发布，待确认无异常后继续发布。默认开启。
  开启 Beta 分组后，发布时会给应用服务设置一个特殊的 Beta 分组，在该组中，系统会在每个部署单元选择一个 Pod，Beta 分组会在第一组发布。
  - Beta 分组发布完成后系统会自动暂停应用服务发布，等待系统负责人或者运维工程师对应用服务的发布情况进行确认。若容器服务发布正常，则单击 **Beta 分组确认**，使应用服务继续分组发布。
  - Beta 分组可以与所有分组策略共同决定分组。创建新的发布申请时，默认开启 **添加 Beta 分组**，此时同一个发布单上的所有应用服务都设置 Beta 分组。
- **部署分组策略**：指定发布容器服务时 Pod 的分组策略，支持以下几种策略，具体可参考下表。

策略类型	说明
按部署单元分组	默认为 **按部署单元分组**。 按部署单元维度发布，尽可能让 Pod 均匀分布在各部署单元中。
每个 Pod 一组	每组一个 Pod，有几个 Pod 便分几组。
共分一组	所有 Pod 在一组中进行发布。
快速分组	在发布时按照组的维度进行发布。

策略类型	说明
共分一组（Beta 单元模式）	在发布时按设定的每次最大同时变更 Pod 数量进行发布。选定后，可以根据需要配置以下选项。 ○ **最大步长**：每个普通分组最大变更 Pod 数（不包含 Beta 分组）。如果将最大步长设为 2，则普通分组的 Pod 数不超过 2。 ○ **部署单元发布顺序**： ■ 如果不设置部署单元发布顺序，按默认的部署单元顺序对各部署单元的 Pod 进行发布或变更。 ■ 如果设置了部署单元发布顺序，按指定顺序对各部署单元的 Pod 进行发布或变更。 ○ **是否开启 Beta 验证**： ■ 如果未开启，所有 Pod 会依次发布，无 Beta 分组，无分组暂停，即用户不需要确认任何分组，执行完一个普通分组后自动执行下一个分组。 ■ 如果开启，部署顺序中的第一个单元会整体作为 Beta 分组进行发布，需要用户确认。其余单元的所有 Pod 会依次发布，这些分组无分组暂停。 **说明** 选择 **共分一组（Beta 单元模式）** 后，在某一时刻可能出现某一部署单元内的 Pod 同时在发布，以致 Pod 不可用的情形。如需规避此现象，即在加快发布进度的同时保证一定 Pod 可用，建议选择 **按部署单元百分比并发模式**。
按部署单元百分比并发模式	如需加快按部署单元发布进度，可选择该分组策略。选定后，需指定一个百分比值。发布变更时，按每个部署单元该百分比数量的 Pod 为一批。 如果将百分比设为 25%，则第一批发布所有部署单元的前 25% 的 Pod，第二批发布所有部署单元的 25%-50% 的 Pod，以此类推直至 100%，可参考以下示例图。 

- **部署单元发布顺序**：仅当分组策略选择 **按部署单元分组** 或 **共分一组（Beta 单元模式）** 时显示该此选项。默认关闭。用于重新编排原有部署单元发布顺序。开启后，支持通过拖拽操作对部署单元进行排序，便于根据实际业务需要灵活调整发布顺序。



- **最小分组数**：分组策略选择 **按部署单元分组** 或 **快速分组** 时需设置。

  > **说明**
  > 最小分组数是发布时分组的参考值，实际分组数受发布时 Pod 在部署单元的分布情况影响。

- **分组暂停**：支持仅第一个分组或每组暂停发布，待确认发布后继续发布。
  - **全部暂停**：默认选中。发布单会在每个分组暂停发布。
  - **首批暂停**：选择 **首批暂停** 后，发布单只会在第一个分组暂停发布。
- **自定义标签**：发布应用服务时，将会以 label 的形式发布在 Pod 上。以 `key=value` 的格式输入，多个 label 间用英文逗号分隔。
- **应用服务与节点亲和性配置**：添加应用服务在节点级别的亲和性配置。可以通过 node 标签来限定应用服务可以调度的节点范围。

  > **说明**
  > 亲和性需要 Pod 重新创建才会生效。首次发布时，亲和性配置一定会生效，后续发布时，必须选择替换升级方式，才能让亲和性配置生效。

- **应用服务间亲和性配置**：添加应用服务在 Pod 级别的亲和性配置。通过选择与某些应用服务在相同或不同的节点来限定应用服务可以调度的节点范围。

  > **说明**
  > 亲和性需要 Pod 重新创建才会生效。首次发布时，亲和性配置一定会生效，后续发布时，必须选择替换升级方式，才能让亲和性配置生效。

- **应用服务污点容忍配置**：配置完成后应用服务将部署到对应污点的节点上。在弹出框中输入 **键**、**值**，选择 **Effect**。Effect 包含以下选项：
  - **NoSchedule**：不允许无匹配 toleration 的 Pod 调度到该节点。
  - **NoExecute**：若无匹配 toleration 的 Pod 已经在节点上运行，则将 Pod 驱逐，若无匹配 toleration 的 Pod 尚未在节点上运行，则不会将 Pod 调度到该节点上。
  - **PreferNoSchedule**：尽量不将无匹配 toleration 的 Pod 调度到该节点。
- **注入 SOFAMesh**：需集群开启 Mesh 功能。开启引流配置时，每次发布会按所设置的引流规则转发流量到新版本，也可在发布过程中配置。开启后，会为应用服务开启服务网格功能，可以在服务网格控制台对应用服务进行管控和治理。

六、预览并提交

在应用服务 **预览** 页面，确认信息无误，单击 **提交**。

> 说明
> - 提交前您可以单击修改图标对应用服务信息进行修改。
> - 应用服务提交后处于 **待部署** 状态，需要单击 **发布** 才会将应用服务部署到集群中。

9.5.2. 高级配置

在创建应用服务过程中，您可以在配置 Pod 模板时选择是否启用高级配置，以实现对应用服务性能优化和灵活性提升。本文主要对配置 Pod 模板时的高级配置项进行说明。

hostNetwork 配置

> 说明
> 使用 hostNetwork 时，为避免端口监听冲突，1 个节点最多只能挂载 1 个 Pod。

在某些场景下，需要将 Pod 的网络从容器网络修改为宿主机网络。比如，需要解决网络的连通性问题等。
如果使用 hostNetwork 配置网络，在应用服务发布成功后，Pod IP 将与节点 IP 保持一致。

环境变量

通过键值对的形式为 Pod 配置环境变量，从而给 Pod 添加环境标志或传递配置等，在容器启动时传入应用进程，例如：`USER=tester`。更多信息，请参见 Pod variable。

参数说明如下表所示。

参数名	说明
来源	环境变量配置来源。取值如下： • **手动输入**：自行指定环境变量的名称和值。 • **配置项**：选择已配置的 ConfigMap 来指定环境变量的值。 • **保密字典**：选择已配置的 Secret 来指定环境变量的值。
变量名	环境变量的名称。 > 说明 > 变量名必须由字母、数字、下划线（_）、短横线（-）或英文句号（.）组成，且不能以数字开头。
变量值/变量引用	根据选择的环境变量来源，配置不同的变量值或引用。 • 当选择 **手动输入** 时，输入环境变量的值。 • 当选择 **配置项** 或 **保密字典** 时，输入 ConfigMap 和 Secret 的引用。

数据卷配置

配置容器使用的数据卷，目前仅支持挂载在当前容器所在宿主机的目录。
参数说明如下表所示。

参数名	说明
类型	支持的类型如下： • **本地存储** • **配置项** • **保密字典** • **PVC 模板** • **PVC** 更多信息，请参见 Volumes。
权限	支持 **读写**、**只读**。
容器目录	容器中的数据卷路径。
主机目录/配置项引用/PVC 名称	根据选择的数据卷配置类型，设置对应的主机目录或配置项引用。 • 当选择 **本地存储** 时，输入 Pod 所在宿主机的目录。 • 当选择 **配置项** 时，从下拉列表中选择对应的配置项。 • 当选择 **保密字典** 时，从下拉列表中选择对应的保密字典。 • 当选择 **PVC 模板** 时，从下拉列表中选择对应的 PVC 模板。 **说明** 如果尚未配置 PVC 模板，可在创建或编辑应用服务时进行配置。更多信息，请参见 数据卷模板配置。 • 当选择 **PVC** 时，从下拉列表中选择对应的 PVC。
subPath/subPathExpr	输入 subPath 或 subPathExpr 名称，用于指定所引用卷内的子路径，而不是其根路径。配置完成后，容器目录下已有的文件不会被覆盖。

健康检查配置

健康检查是指容器运行过程中，基于用户配置的健康检查，定时检查容器健康状况。若不配置健康检查，如果出现业务异常，容器将无法感知，也不会自动重启去恢复业务，进而可能影响业务的正常运作。
目前提供两种检查机制：Readiness Probe 和 Liveness Probe。更多信息，请参见 Kubernetes Probe。

说明
- Readiness Probe：用于检查容器是否准备好接收流量。
  - 如果检查失败，集群会屏蔽请求访问该容器。
  - 如果检查成功，会开放对该容器的访问。
- Liveness Probe：检测容器是否正常，如果检查失败，集群会对该容器执行重启操作。

配置 Readiness Probe、Liveness Probe 参数说明如下表所示。

参数名	说明

参数名	说明
检查方式	• **TCP Socket 端口**：通过发送 TCP 请求来进行判断，若能成功建立连接，则检查通过。 • **HTTP Get 方法**：通过发送 HTTP Get 方法的请求来进行判断，若返回的状态码大于等于 200 且小于 400，则检查通过。 • **在容器中执行命令**：通过在容器中执行命令来进行判断，若命令成功执行，并返回 0，则检查通过。
执行命令	当 **检查方式** 选择 **HTTP Get 方法** 或 **在容器中执行命令** 时需设置。 • 选择 **HTTP Get 方法** 时，指定 HTTP 请求的执行命令 。 • 选择 **在容器中执行命令** 时，指定在容器中执行的命令。
访问端口	当 **检查方式** 选择 **TCP Socket 端口** 或 **HTTP Get 方法** 时需设置。 • 选择 **TCP Socket 端口** 时，填写 TCP 访问的容器端口。 • 选择 **HTTP Get 方法** 时，填写 HTTP 请求的访问容器端口。
首次检查延时	initialDelaySeconds，容器启动后第一次执行探测时需要等待多少秒。默认为 3 秒。
检查超时	timeoutSeconds，探测超时时间。默认 1 秒，最小 1 秒。
检查间隔	periodSeconds，指执行探测的时间间隔。默认为 10 秒，最小为 1 秒。
健康阈值	探测失败后，最少连续探测成功多少次才被认定为成功。默认是 1，最小值是 1。
不健康阈值	探测成功后，最少连续探测失败多少次才被认定为失败。默认是 3。最小值是 1。

生命周期事件回调配置

为容器添加生命周期事件回调，分别在容器启动后和容器停止前执行。
配置 Post Start、Pre Stop 参数说明如下表所示。

参数名	说明
回调方式	支持 **在容器中执行命令**、**HTTP GET 方法**。
访问路径	当 **回调方式** 选择 **HTTP Get 方法** 时需设置。 输入访问路径，如 `/healthy/readiness` 。
访问端口	当 **回调方式** 选择 **HTTP Get 方法** 时需设置。 输入访问端口，范围为 1~65535。
执行命令	当 **回调方式** 选择 **在容器中执行命令** 时需设置。 输入执行命令，如 `cat tmp/healthy` 。

日志服务配置

配置日志服务，可选择已有日志库或创建新的日志库。

> 说明
> 首次使用日志服务配置功能，需前往 日志服务控制台 开启配置索引。详情请参见 配置索引。

参数说明如下表所示。

参数名	说明
日志库	在日志服务中生成一个对应的 logstore，用于存储采集到的日志。 从已有的日志库中选择。如果没有可用日志库，可单击 **创建日志库** 来创建新的日志库。
日志类型	容器文本日志。
路径	收集容器内指定路径的日志。例如：`/home/admin/*.log` 或 `/root/*.log` 。

9.5.3. 发布应用服务

9.5.3.1. 立即发布

应用服务创建完成后，您可以通过 LHC 控制台发布该应用服务。
可以通过如下两种方式发布应用服务：

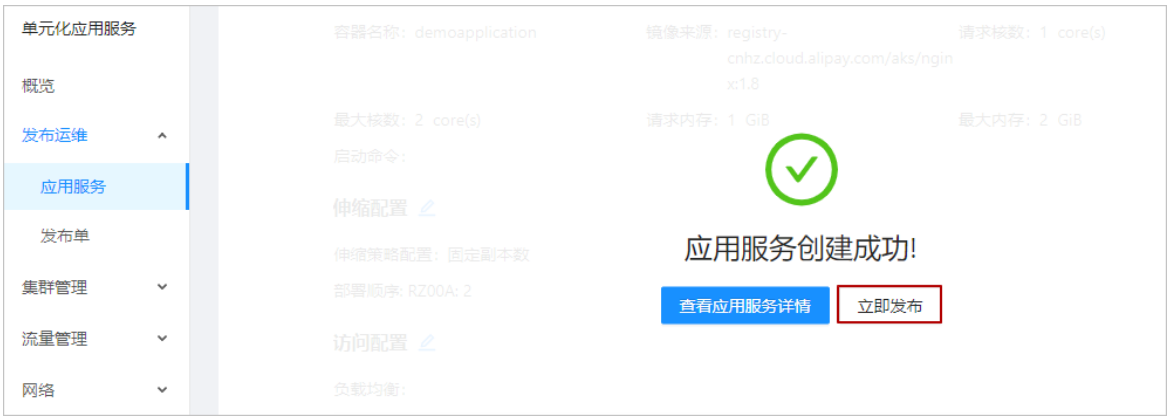
- 方式一：应用服务创建成功后立即发布
- 方式二：通过应用服务详情页发布

前提条件

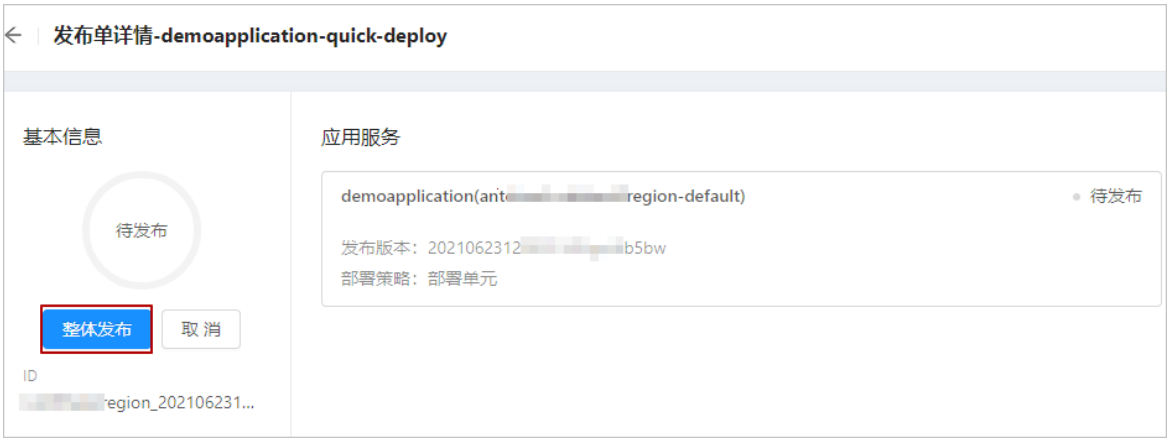
发布应用服务前，需确保集群内有充足的节点资源（如 CPU、Memory），否则 Pod 将无法正常调度到节点。

方式一：应用服务创建成功后立即发布

1. 完成应用服务创建后，单击 **立即发布**。具体创建步骤可参考 创建应用服务。



2. 在 **发布单详情** 页，单击 **整体发布**。



> **说明**
> 待应用服务状态更新为 **发布成功**，即表示该应用服务已经发布成功。

方式二：通过应用服务详情页发布

1. 登录 LHC 控制台。
2. 在左侧导航栏上，单击 **发布运维** > **应用服务**。
3. 在应用服务列表页，单击应用服务名称进入 **应用服务详情** 页。
4. 单击右上角的 **发布**。



> **说明**
>
> - 若应用服务存在多个版本，会提示 是否发布该容器服务最近一次变更的版本。
> - 如需发布以前的版本，选择 **版本** 页签，单击版本右侧的 **发布**。

5. 在 **发布预览** 页面，确认信息无误后单击 **开始发布**。
6. 发布单创建成功后，在 **发布单详情** 页面，单击 **整体发布**。

> **说明**
> 待应用服务状态更新为 **发布成功**，即表示该应用服务已经发布成功。

后续操作

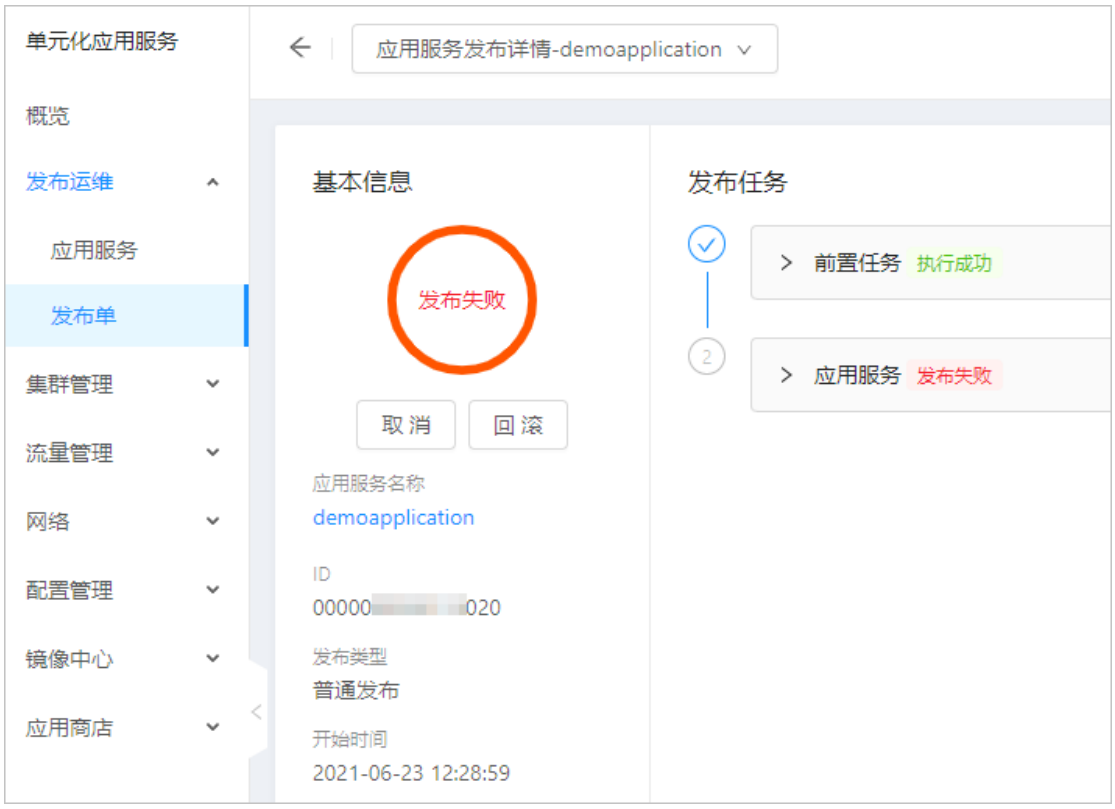
查看发布详情

单击 **发布** 后自动跳转到 **发布单详情** 页面。您可以展开右侧的 **发布任务** 查看详细的发布状态。



重试回滚

若出现发布失败，您可以单击 **重试**，继续进行发布，或单击 **回滚** 恢复到上一个已发布的版本。



> **说明**
> 仅当前应用有其它成功发布的发布包（除当前版本外）时，您才可以回滚应用。若无任何成功发布记录，回滚将释放所有资源。

9.5.3.2. 发布单

通过发布单可以同时发布多个应用服务。本文介绍如何通过创建发布单发布应用服务，以及查看发布单详情等。

创建发布单

1. 登录 LHC 控制台。
2. 在左侧导航栏单击 **发布运维** > **发布单**。
3. 在发布部署大盘，单击 **创建发布单**。

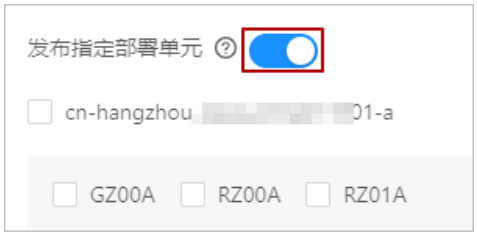
   > **说明**
   > 目前发布部署大盘页面的搜索框仅支持搜索发布单名称的前缀。

4. 在 **创建发布单** 页面，填写以下发布信息后，单击 **下一步**。
   - **基本信息**
     - **标题**：输入发布标题。
     - **类型**：发布类型。支持 **分组发布**、**蓝绿发布**。本例中选择 **分组发布**。蓝绿发布相关操作，请参见 蓝绿发布。
     - **命名空间**：选择待发布的应用服务所属的命名空间。
     - **应用服务发布列表**：您可以在 **待选应用服务列表** 中单击选择需要的应用服务，单击 将应用服务添加到 **已选应用服务列表** 中。

       > **说明**
       > 若应用服务存在多个提交版本，需要选择要发布的版本。默认选择最新的版本。

     - **发布指定部署单元**：默认关闭。如果开启，需根据实际业务诉求勾选部署单元。应用服务将仅发布至已选择且当前版本配置中的部署单元，不会对未选中部署单元的已有 Pod 做任何变更。

       > **警告**
       > 此功能仅适用于机房容灾场景，即通过勾选部署单元，实现容灾场景下应用服务的正常发布。关于容灾更多信息，请参见 多集群联邦发布容灾。

   - **高级配置**
     设置应用服务依赖关系：当选择了 2 个及以上应用服务时，可按需填写应用服务之间的依赖关系。若应用服务之间不存在依赖关系，则直接跳过。您可通过逐条添加的方式添加依赖关系。在 **依赖于** 列中输入某个应用服务所依赖的服务名称（可以同时选择多个）。例如，若 appservice1 依赖 appservice2，则在 appservice1 对应的 **依赖于** 列中输入 appservice2 即可，依赖于多个应用服务时同理。

     > **注意**
     > 应用服务之间不可循环依赖，即不可出现应用服务 A 依赖于应用服务 B，同时应用服务 B 又依赖于应用服务 A 这种情况。

5. 在 **预览** 页面确认信息无误后，单击 **创建**。系统会自动跳转到 **发布单详情** 页面，单击 **整体发布** 即可开始发布。

查看发布单详情

发布单创建完成后，系统会自动跳转到 **发布单详情** 页面，您可以查看发布单的执行详情。

> 说明
> 若发布单初始化失败，则无法进行发布，可以将鼠标悬停到 **失败原因** 上，查看具体原因，或单击 **重新初始化**。

前置任务

- **执行服务配置变更**：若容器服务存在添加或修改 Service 服务配置，执行 Service 服务配置任务，若无则直接跳过。
- **发布日志配置**：若容器服务首次配置日志配置，则执行日志配置前置任务：创建或更新，若无则直接跳过。

应用服务

分组发布应用服务，Pod 状态说明如下：

- **未更新**：Pod 配置未更新为最新版本。
- **更新中**：Pod 配置已更新为最新版本，但状态还未 ready 或保持 ready 状态少于 30 秒。
- **更新成功**：Pod 配置已更新为最新版本，并保持 ready 状态 30 秒以上。
- **更新失败**：上线或更新 Pod 异常。

后置任务

- **清理无效服务配置**：若容器服务存在删除 Service 服务配置，执行 Service 清理任务，若无则直接跳过。
- **发布日志配置**：若容器服务未配置日志配置或取消日志配置，则执行日志配置后置任务—删除，若无则直接跳过。

后置任务（回滚）

回退日志配置：若应用服务需回退至无资源状态，且应用服务的日志配置已生效，则执行日志配置的回退任务，若无则直接跳过。

取消发布

发布单创建完成后，系统会自动跳转到 **发布单详情** 页面，您可以在正式发布应用服务前选择 **取消** 发布，终止发布操作。

在 **发布单详情** 页面，单击 **取消**。

若发布单包含多个应用，单击右侧待发布的应用服务，进入单个应用发布页面，单击 **取消**，会取消当前应用发布，其他未取消的应用可以继续发布。

Beta 确认

配置应用服务的部署和调度策略时，若开通了 Beta 确认功能，在应用服务发布时，系统会给服务器分组设置一个特殊的 Beta 分组，即系统从每个部署单元获取一台机器，加入 Beta 分组，并作为第一组发布。

Beta 分组发布完成后系统会自动暂停应用发布，此时系统负责人或运维人员需要对应用的发布情况进行确认。

若应用发布正常，则单击 **Beta 确认**，使应用继续分组发布。

> 说明
> - 若每个部署单元只有一个 Pod，则 Beta 确认无效。
> - Beta 确认功能目前仅在发布应用服务新版本时生效。
> - 若分组策略中没有勾选 **分组暂停**，Beta 组发布后，剩余的分组会自动分批发布直至发布完成。

重试回滚

若发布失败，您可以单击 **重试** 或 **回滚** 继续进行操作。

重试

若单击 **重试**，系统继续进行发布。

> 说明
> 若应用发布失败，无法创建新的发布单进行发布。

回滚

若单击 **回滚**，根据不同的回滚场景，您需要进行二次确认：

- **取消发布**：找不到可回滚版本，继续回滚将终止此次发布。

  > 说明
  > 该场景仅在发布成功过一次，重新发布同一个版本失败时出现。

- **回退至无资源状态**：无任何成功发布记录，继续回滚将释放所有资源。
- **回滚至上一版本**：即将回滚到上一个成功发布版本。

回滚完成后，单击 **确认回滚**。您可以单击右上角的 **原应用发布详情** 查看原发布单发布详情。
若回滚失败，可以选择 **重试回滚** 或 **取消回滚**。

> 注意
> 取消回滚将终止此次发布，可能导致线上存在多个版本。

查看 Pod 事件

发布过程中或者结束后，您可以查看 Pod 事件，了解 Pod 的调度与运行状态。

- 在 **发布单详情** 页面，单击 Pod 名称查看 Pod 事件。
- 在 **应用服务详情** 页的 Pod 列表中，单击 Pod 名称进入 **Pod 详情** 页，选择 **事件** 页签，查看 Pod 事件。

9.5.3.3. 蓝绿发布

蓝绿发布是迭代产品在生产环境安全上线的一种重要手段，提供零宕机部署方式。本文介绍如何通过创建蓝绿发布单发布应用服务。

> 说明
> 蓝绿发布单在应用服务维度互斥，即不能通过两个发布单同时发布一个应用服务，且与普通发布单互斥。

前提条件

已创建应用服务，且配置应用服务的 部署和调度策略 时，将部署分组策略设为 **共分一组** 或 **快速分组**。
关于分组策略限制说明可参考下表。

部署分组策略	蓝绿发布是否支持	说明
每个 Pod 一组	×	由于蓝绿发布的灰度验证是通过 spanner 来控制，而不是通过应用服务内百分比升级的 Pod 来控制的。
按部署单元分组	×	假设流量配比为 100：0，对于流量为 0 的区域，若按部署单元分组，流量无法调配至该区域。
共分一组	√	/

部署分组策略	蓝绿发布是否支持	说明
快速分组	√	/
共分一组（Beta 单元模式）	×	/
按部署单元百分比并发模式	×	/

创建蓝绿发布单

1. 登录 LHC 控制台。
2. 在左侧导航栏单击 **发布运维** > **发布单**。
3. 在发布部署大盘，单击 **创建发布单**。

> 说明
> 目前发布部署大盘页面的搜索框仅支持搜索发布单名称的前缀。

4. 在 **创建发布单** 页面，填写以下发布信息后，单击 **下一步**。
   - **基本信息**
     - **标题**：发布标题。
     - **类型**：选择 **蓝绿发布**。
     - **命名空间**：选择待发布的应用服务所属的命名空间。
     - **应用服务发布列表**：您可以在 **待选应用服务列表** 中单击选择需要的应用服务，单击 **添加** 将应用服务添加到 **已选应用服务列表** 中。

       > 说明
       > - 若应用服务存在多个提交版本，需要选择要发布的版本。默认选择最新的版本。
       > - 蓝绿发布单选择的应用服务上一次部署状态必须是成功的，保证有一个完整的蓝版本在线。
       > - 蓝绿发布单选择的版本与当前运行的版本对比，不能有统一接入的配置变更。
       > - 蓝绿发布单选择的两个版本不能有拓扑变更，比如从 Cell A 变为 Cell B。对于副本数变更，不属于拓扑变更。

   - **高级配置**
     设置应用服务依赖关系：当选择了 2 个及以上应用服务时，可能需要填写应用服务之间的依赖关系。如果应用服务之间不存在依赖关系，则直接跳过。
     在 **依赖于** 列中输入某个应用服务所依赖的服务名称（可以同时选择多个）。您可通过逐条添加的方式添加依赖关系。
     例如，如果 appservice1 依赖 appservice2，则在 appservice1 对应的 **依赖于** 列中输入 appservice2 即可，依赖于多个应用服务时同理。

     > 注意
     > 应用服务之间不可循环依赖，即不可出现应用服务 A 依赖于应用服务 B，同时应用服务 B 又依赖于应用服务 A 这种情况。

5. 在 **预览** 页面确认信息无误后，单击 **创建**。系统会自动跳转到 **发布单详情** 页面，单击 **整体发布** 即可开始发布。

查看蓝绿发布单详情

发布单创建完成后，系统会自动跳转到 **发布单详情** 页面，您可以查看发布单的执行详情。

> 说明
>
> - 若发布单初始化失败，则无法进行发布，可以将鼠标悬停到 **失败原因** 上，查看具体原因，或单击 **重新初始化**。
> - 蓝绿发布单的蓝版本为当前在线版本。

前置任务

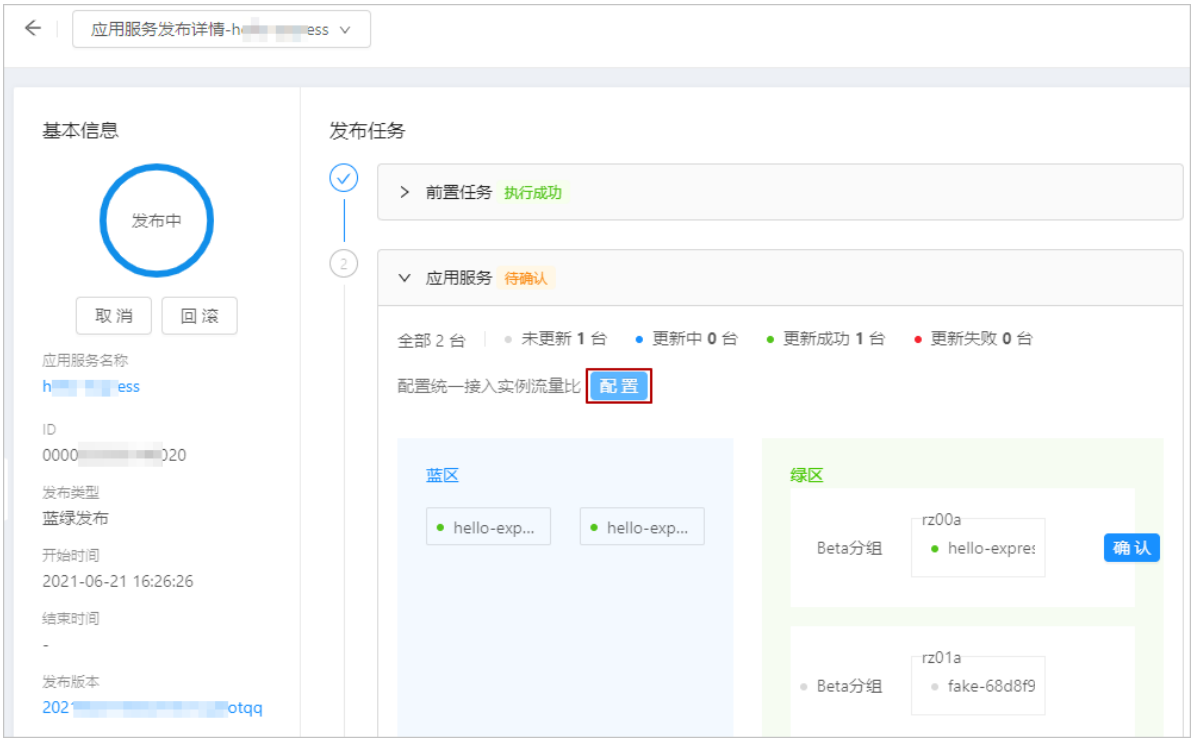
- **执行服务配置变更**：若容器服务存在添加或修改 Service 服务配置，执行 Service 服务配置任务，若无则直接跳过。
- **发布日志配置**：若容器服务首次配置日志配置，则执行日志配置前置任务：创建或更新，若无则直接跳过。

应用服务

若应用服务带有统一接入，可配置蓝绿区流量配比。

> 说明
>
> 仅应用服务配置了统一接入，才可以为蓝绿区设置流量配比，即在绿版本发布的过程中进行流量调拨。关于如何为应用服务配置统一接入的信息，请参见 配置统一接入。

- 如果需要自定义蓝绿区流量配比，在应用服务区域单击 **配置**，在弹框中输入蓝区、绿区流量配比即可。



- 如果未自定义蓝绿区流量配比，流量配比将按 `蓝区：绿区 = 100：0（%）` 来确定流量分发的比重。随后，将在执行后置任务时，全部流量将流向绿区。

后置任务

- **流量切换到绿区**：若蓝绿发布应用服务，执行流量切换到绿区后置任务。
- **清理无效服务配置**：若容器服务存在删除 Service 服务配置，执行 Service 清理任务，若无则直接跳过。
- **发布日志配置**：若容器服务未配置日志配置或取消日志配置，则执行日志配置后置任务—删除，若无则直接跳过。

后置任务（回滚）

回退日志配置：若应用服务需回退至无资源状态，且应用服务的日志配置已生效，则执行日志配置的回退任务，若无则直接跳过。

其他操作

- 蓝绿发布单创建完成后，如需终止蓝绿发布，可单击 **取消**。具体操作步骤与普通发布的相同，请参见 取消发布。

  > **说明**
  > 蓝绿发布单取消之后，可能会同时出现多个运行版本，需等到下一次用户主动发布才会自动清除多余版本。

- 配置应用服务的部署和调度策略时，若开通了 Beta 确认功能，在应用服务发布时，单击 **Beta 确认**，使应用继续分组发布。具体操作步骤与普通发布的相同，请参见 Beta 确认。
- 若蓝绿发布执行失败，可单击 **重试** 或 **回滚** 继续进行操作。具体操作步骤与普通发布的相同，请参见 重试回滚。
- 发布过程中或完成后，您可以查看 Pod 事件，了解 Pod 的调度与运行状态。具体操作步骤与普通发布的相同，请参见 查看 Pod 事件。

9.5.3.4. 多集群联邦发布容灾

本文介绍 LHC 多集群联邦发布容灾相关信息，包括基本概念、灾难感知等。

背景信息

当站点因为不可抗力因素或者设备故障导致应用在短时间内无法恢复时，LHC 可提供多集群联邦发布容灾保护。当站点故障时，通过简单的配置，即可在容灾站点迅速恢复业务。

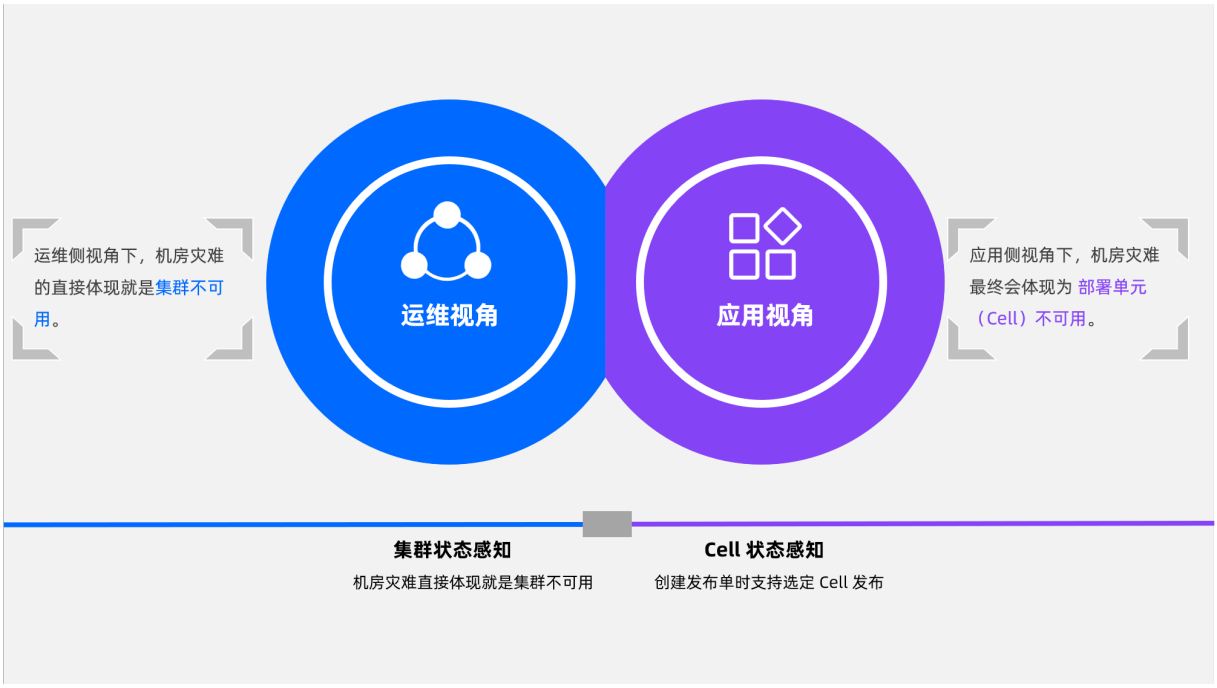
容灾是一个范畴比较广泛的概念。广义上，容灾是一个系统工程，包括所有与业务连续性相关的内容。狭义的容灾是指建立两套或多套功能相同的 IT 系统，互相之间可以进行状态监视和功能切换，当主要站点因意外停止工作时，整个应用系统可以利用辅助站点快速恢复，并继续工作。

容灾的主要目的是当自然或人为的原因导致生产系统发生灾难时，能够尽可能地保证业务的连续性。

机房灾难感知

为实现容灾场景下 LHC 多集群的正常发布，可通过以下方式来判断机房灾难。

- **集群状态**：运维侧视角下，机房灾难的直接体现就是集群不可用。
- **部署单元状态**：应用侧视角下，机房灾难最终会体现为部署单元（Cell）不可用，导致用户通过发布单发布应用服务时出现不符合预期的发布。



机房灾难下的多集群联邦发布

对于应用发布而言，容灾场景下存在多种可选项。在发布单执行前，用户可通过仅指定可用的 Cell 执行发布。相关操作请参见：创建发布单。

9.5.4. 更新应用服务

创建并提交应用服务后会生成对应的版本。每个版本都是无法进行修改的，若更新配置，则会生成新的版本。新的版本需要通过发布才会在线上生效。对于日常更新版本，只需选择镜像而不用其他配置，可在应用服务详情页选择快捷操作。

标准更新流程

1. 登录控制台，在左侧导航栏单击 **发布运维 > 应用服务**。
2. 在应用服务列表页，单击应用服务的名称进入 **应用服务详情** 页。
3. 单击右上角的 **编辑配置信息**，进入 **编辑版本** 页面。

   > 说明
   > 若当前有未提交的草稿，在弹出的对话框中选择是否恢复打开草稿，若选择 否，会删除当前草稿。

4. 修改以下配置信息。参数说明参考 创建应用服务。
   - **基本信息** 中的 **描述**。
   - **Pod 模板配置** 中除 **容器名称** 以外的所有信息。
   - **弹性配置** 中的副本数信息。
   - **访问配置** 中除服务名称以外的所有信息。

     > 说明
     > 在编辑应用服务页对访问配置 已创建的自定义 LB 执行删除时，只会删除配置的元数据，而自定义的负载均衡实例本身不会被删除。

   - **部署调度** 中所有信息。
5. 修改完成后，预览并提交。

配置信息编辑成功后，您可以在应用服务详情页的版本页签中查看更新的应用服务版本信息。

6. 在应用服务详情页的版本页签，选择新版本，并单击 **发布**。

快捷更新流程

1. 登录控制台，在左侧导航栏单击 **发布运维 > 应用服务**。
2. 在应用服务列表页，单击应用服务的名称进入 **应用服务详情** 页。
3. 单击 **快捷操作**，选择 **更换镜像** 或 **扩缩容**，选择目标版本镜像或修改副本数，单击 **提交**。
4. 提交后会自动生成发布单，单击 **整体发布** 进行发布。

9.5.5. 导入与导出应用服务

应用服务导入功能支持将本地应用服务数据导入到 LHC 控制台的应用服务创建页。本文介绍如何通过 LHC 控制台导入与导出应用服务。

背景介绍

企业级应用研发一般会有多个环境，比如开发测试环境、生产环境等，在不同的环境中需要配置相同或相近的应用服务。如果在开发测试环境创建和发布了应用服务 A，进行测试验证后，需要在生产环境重新发布和发布应用服务 A，可能会带来如下问题：



- **重复配置**：每个应用服务在新的环境都需要手动配置，导致很大的重复工作量。
- **测试生产不一致**：手动填写应用服务创建页会出现遗漏现象，可能导致生产和测试环境配置不一样，无法达到开发测试验证的预期效果。

基于以上两个问题，LHC 提供应用服务导入与导出功能，支持将应用服务以 JSON 格式导出，然后导入其他环境，极大简化了应用服务的创建和发布操作，进而有效提升运维效率，保障用户体验。

导入应用服务

> 注意
> 应用服务导入功能，并非实际创建应用服务，而是基于导入的应用服务数据预先填充应用服务创建页。

您可以通过 LHC 控制台将 JSON 格式的应用服务配置信息导入当前环境。具体操作如下：

1. 登录控制台，在左侧导航栏单击 **发布运维 > 应用服务**。
2. 在应用服务列表页，单击 **导入**。
3. 在弹窗中选择本地保存的 JSON 格式的应用服务数据，单击 **打开**。

导入后，可根据实际业务需要在自动填充的应用服务创建页修改具体的配置项信息。

导出应用服务

将当前环境下的目标应用服务导出为 JSON 格式文件。具体操作如下：

1. 登录控制台，在左侧导航栏单击 **发布运维 >应用服务**。
2. 在应用服务列表页，单击应用服务名称进入 **应用服务详情** 页。
3. 选择 **版本** 页签进入版本列表页。
4. 单击目标应用服务版本右侧 **操作** 栏中的 **导出**，以 JSON 格式将应用服务数据保存至本地。

9.5.6. 查看应用服务详情

创建完应用服务后，您可以查看应用服务详情，包括应用服务的基本信息、配置信息，方便您快速了解应用服务详细信息，例如 Pod 详情、负载均衡信息、统一接入信息等。同时，应用服务详情页还提供了丰富的功能菜单，能够让您更加快捷高效地进行应用服务的更新操作，包括镜像更换、扩缩容等。

操作步骤

1. 登录控制台，在左侧导航栏单击 **发布运维 > 应用服务**。
2. 在应用服务列表页，单击应用服务实例名称，进入 **应用服务详情** 页面。
3. 查看应用服务配置信息。
   应用服务详情 页包括以下内容。

模块	说明
基本信息	应用服务的基本信息，例如应用服务名称、所属应用、创建人信息、技术栈、镜像来源、请求 CPU、请求内存、最大 CPU、最大内存、命名空间、副本个数等。
Pod	显示部署单元维度容器服务所包含的 Pod。可使用 web shell 登录进入 Pod 内的容器进行操作。详情参考 Pod 详情。 在 Pod 列表页，单击 **操作** 列中的 **监控**，可监控 Pod 详情。
负载均衡	显示服务名称、网络类型、负载均衡地址、NodePort、部署单元、创建时间。
统一接入	显示统一接入名称、统一接入实例名称、部署单元等信息。 单击统一接入实例名称可跳转至 **实例详情** 页。
版本	显示应用服务的版本号、发布单、发布状态等信息。详情参考 查看应用服务版本。
事件	显示事件数量、事件内容、创建时间。 支持按自定义的时间段筛选，只需在日历中选择开始和结束时间，或者在文本框内手动输入即可。

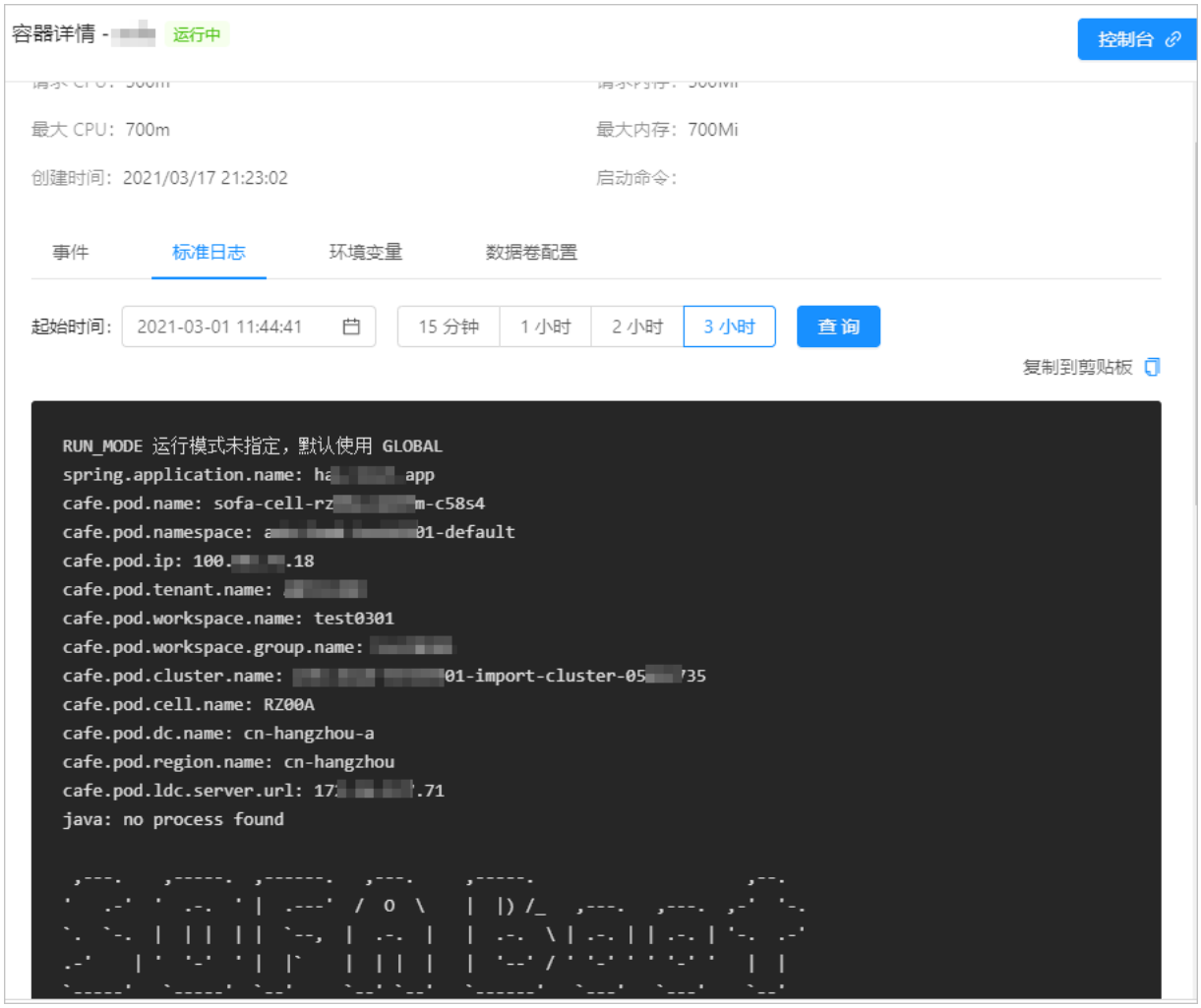
Pod 详情

容器控制台

1. 登录控制台，在左侧导航栏单击 **发布运维 > 应用服务**。
2. 在应用服务列表页，单击应用服务名称进入 **应用服务详情** 页。
3. 在 Pod 列表中，单击 Pod 名称进入 **Pod 详情** 页。
4. 在容器列表中，单击容器右侧的 **前往控制台**。
5. 输入命令管理容器。
   命令样例：`exec [-cl] [-a name] [command [arguments ...]] [redirection ...]`

查看容器详情

在容器列表中，单击容器名称进入容器详情页。您可以查看容器的基本信息。单击下方页签可查看事件、标准日志、环境变量、数据卷配置详情。例如，您可以查看容器在某一时间段的标准日志记录，如下图所示。



查看 Pod 事件

参考 查看 Pod 事件。

查看应用服务版本

1. 登录控制台，在左侧导航栏单击 **发布运维 >应用服务**。
2. 在应用服务列表页，单击应用服务名称进入 **应用服务详情** 页。
3. 选择 **版本** 页签进入版本列表页。默认会将当前版本显示在最上面。您可以查看以下版本信息：
   - **应用服务版本**：应用服务的所有提交过的版本 ID，单击版本链接会跳转到查看应用服务的详细配置页面。
   - **发布单**：若服务已经发布，会展示关联的发布单 ID，单击发布单链接会跳转到发布单详情页。
   - **发布状态**：发布完成、发布失败、待部署、已取消、已回滚等。
   - **版本状态**：版本的提交状态，如已提交。
   - **创建人**：显示应用服务版本的创建人。
   - **创建时间**：显示应用服务版本的创建时间。

- **操作**：支持版本对比、导出操作。
  - **版本对比**：显示当前版本与其他版本间的差异。
  - **导出**：将应用服务数据导出至本地。详情参考 导出应用服务。

其他操作

在应用服务详情页，您还可以根据需要进行更多快捷操作，具体可参考下表。

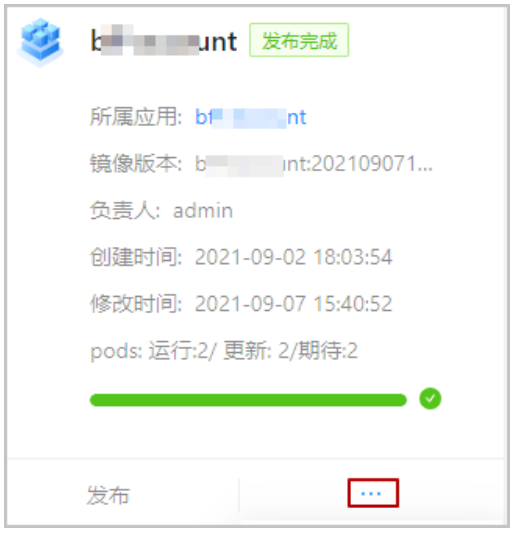
操作	说明
更新应用服务	在应用服务详情页，单击 **编辑配置信息**。 详情参考 更新应用服务。
仅更换镜像	单击 **快捷操作**，选择 **更换镜像**，选择目标版本镜像即可完成镜像更新。 详情参考 更新应用服务。
仅更新伸缩配置	单击 **快捷操作**，选择 **扩缩容**，然后按需修改副本数即可完成伸缩配置更新。 详情参考 更新应用服务。
立即发布应用服务	在应用服务详情页，单击右上角 **发布** 按钮即可立即发布应用服务。 详情参考 立即发布应用服务。

9.5.7. 删除应用服务

在应用服务列表页，您可以删除不再需要的应用服务。删除应用服务后，服务下的资源将全部被释放。本文主要介绍如何删除应用服务。

操作步骤

1. 登录控制台，在左侧导航栏单击 **发布运维** > **应用服务**。
2. 可以通过如下方式之一删除应用服务：
   - 找到需要删除的应用服务，单击其卡片上的更多图标 ··· > **删除**。



   - 单击右上角列表图标进入应用服务列表页，单击目标应用服务右侧 **操作** 栏中的 **删除** 。



3. 在弹出的删除验证对话框中，输入应用服务名称后，单击 **确定**。
   删除时，该应用服务下所有 Pod 也会随之一并删除。

> **注意**
> 正在发布中的应用服务是无法删除的。

相关文档

您可以参考以下文档，了解应用服务更多功能及操作：

- 创建应用服务
- 发布应用服务
- 更新应用服务
- 导入与导出应用服务
- 查看应用服务详情

9.5.8. 系统环境变量

作为基于 Kubernetes 的容器服务平台，LHC 默认会在容器里注入一系列平台相关的环境变量，这样通过单元化应用服务部署的应用可以很方便地通过环境变量获取许多基础信息。平台默认注入的环境变量列表如下：

> **说明**
> 系统环境变量为只读属性，用户不可以自行修改。

环境变量名称	环境变量解释	示例
CAFE_ENV_TENANT_NAME	租户名称	ANTCLOUD
CAFE_ENV_APPNAME	应用名称	transform
CAFE_ENV_DC_NAME	机房名称	cn-hangzhou
CAFE_ENV_REGION_NAME	Region 名称	杭州金区
CAFE_ENV_CELL_NAME	部署单元名称	CellA/RZ01A（单元化架构）
CAFE_ENV_WSG_NAME	工作空间组名称	dev
CAFE_ENV_WS_NAME	工作空间（单元化工作空间）名称	dev
CAFE_ENV_POD_NAME	Pod 名称	transform-01
CAFE_ENV_POD_NAMESPACE	Pod 所属 namespace	default

环境变量名称	环境变量解释	示例
CAFE_ENV_POD_IP	Pod IP	127.0.0.1
CAFE_ENV_HOST_NAME	Pod 所属机器的名称	node01
CAFE_ENV_HOST_IP	Pod 所属机器的 IP	127.0.0.1
app.env.zmode	是否是单元化模式（中间件依赖）	true、false
app.env.dbmode	数据库 mode（中间件依赖）	test、pre、prod

9.5.9. 容器水平伸缩 HPA

容器水平伸缩（Horizontal Pod Autoscaler，简称 HPA）是 Kubernetes 中实现 Pod 水平自动伸缩的功能。对于 Kubernetes 中的 Pod 而言，HPA 可以实现很多自动化功能，比如当 Pod 中工作负载上升时，会基于设定的扩容规则，动态调整副本数，从而保证业务系统稳定运行；当 Pod 中工作负载下降时，会基于设定的缩容规则，动态调整副本数，从而提高资源利用率。

HPA 基本原理

HPA 是用来控制 Pod 水平伸缩的控制器，基于设定的扩缩容规则，实时采集监控指标数据，根据用户设定的指标阈值计算副本数，进而调整目标资源的副本数量，完成扩缩容操作。

HPA 工作流程如下：

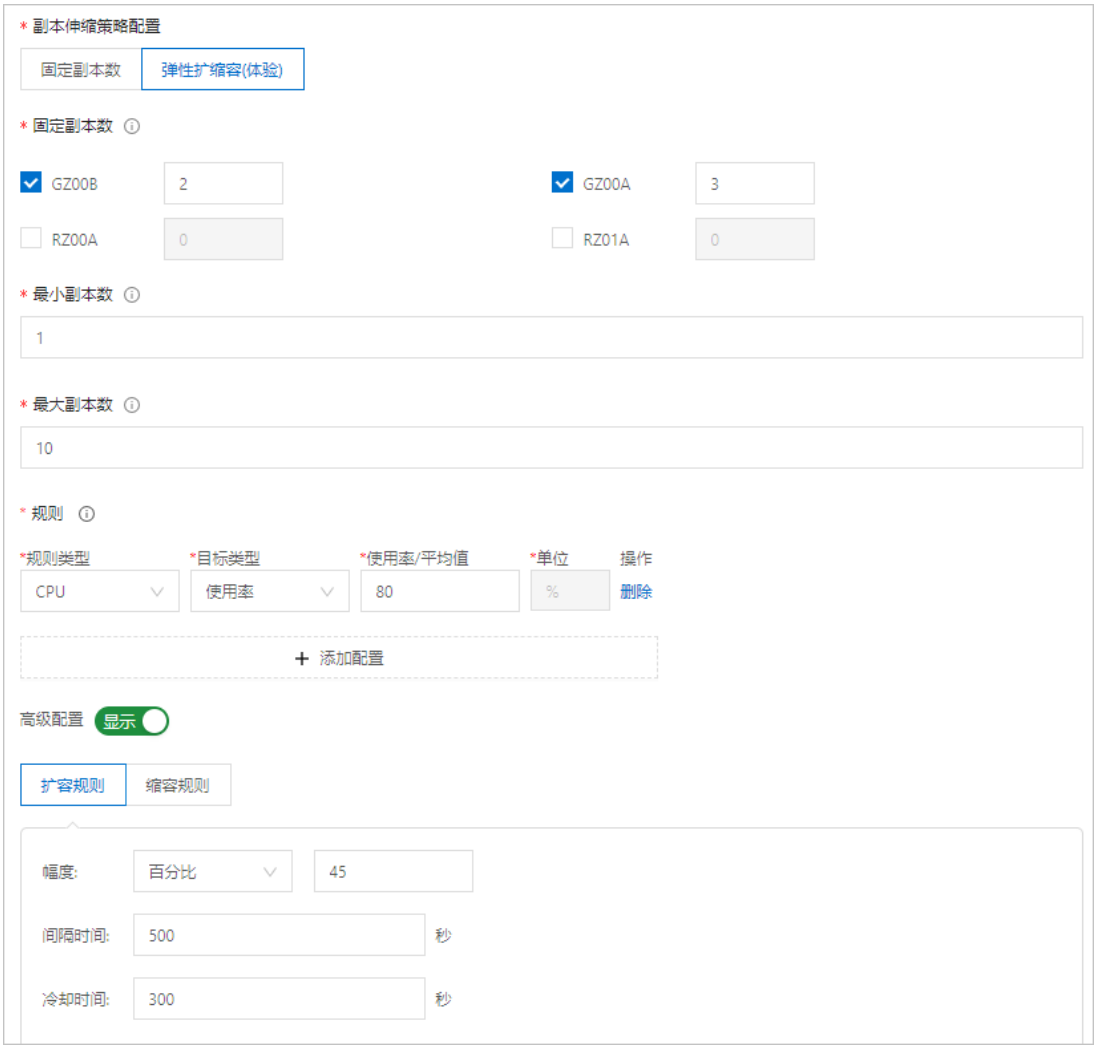


通过控制台创建 HPA

LHC 支持在控制台界面上快速创建支持 HPA 的应用服务，您可以轻松通过单元化应用服务控制台在创建应用服务时进行创建。

操作步骤

1. 登录控制台，在左侧导航栏单击 **发布运维 > 应用服务**。
2. 在应用服务列表页，单击 **创建应用服务**。
3. 在 **创建应用服务** 页面，完成基本信息、Pod 模板配置后，单击 **下一步**。
4. 在 **弹性配置** 页面，完成以下配置后，单击 **下一步**。



- **副本伸缩策略配置**：选择 **弹性扩缩容**。
- **固定副本数**：勾选部署单元并修改为期望副本数。
- **最小副本数**：弹性扩缩容时 Pod 的最小副本数。默认值为 1。
- **最大副本数**：弹性扩缩容时 Pod 的最大副本数。默认值为 10。
- **规则**：当前伸缩规则基于的指标类型。单击 **+ 添加配置**，配置以下选项：
  - **规则类型**：当前伸缩规则基于的规则类型。目前支持 **CPU**、**Memory**、**QPS**、**响应时间**。
  - **目标类型**：当前伸缩规则基于的指标类型所期望达到的值的类型，目前支持 **使用率** 与 **平均资源使用**。
  - **使用率/平均值**：输入使用率/平均值。
  - **单位**：无需输入，根据所选的规则类型自动匹配。
- **高级配置**：默认关闭。如果开启 显示，需要为扩缩容规则配置幅度、间隔时间、冷却时间。
  - **幅度**：可以通过设置按百分比或指定个数来调整扩缩容的速率，单位为百分比或个数。
  - **间隔时间**：上次扩缩容操作与本次扩缩容操作之间的时间间隔，单位为秒。

- **冷却时间**：冷却扩缩容操作的时间，即完成一次扩缩容操作之后不再次触发扩缩容操作的时间窗口，单位为秒。

  > **说明**
  > 一个扩缩容活动执行完成后，在设定的冷却时间内不执行其他扩缩容活动。

5. （可选）完成访问配置、部署和调度配置后，单击 **提交**。

通过 kubectl 命令创建

您还可以通过编排模板来手动创建 HPA，并将其绑定到要伸缩的 Deployment 对象上，通过 kubectl 命令实现容器自动伸缩配置。
下面针对一个 Nginx 应用进行举例。

1. 创建并复制以下内容到 nginx.yml 中。Deployment 的编排模板如下。

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: nginx
  labels:
    app: nginx
spec:
  replicas: 2
  selector:
    matchLabels:
      app: nginx
  template:
    metadata:
      labels:
        app: nginx
    spec:
      containers:
      - name: nginx
        image: nginx:1.7.9 # replace it with your exactly <image_name:tags>
        ports:
        - containerPort: 80
        resources:
          requests:                          ##必须设置，否则 HPA 无法运行。
            cpu: 500m
```

> **说明**
> 需要为 Pod 设置 request 资源，如果不设置，HPA 不会运行。

2. 执行以下命令，创建 Nginx 应用。

```
kubectl create -f nginx.yml
```

3. 创建 HPA。通过 scaleTargetRef 设置当前 HPA 绑定的对象。在本例中，绑定对象是名为 nginx 的 Deployment。

```
apiVersion: autoscaling/v2beta2
kind: HorizontalPodAutoscaler
metadata:
  name: nginx-hpa
  namespace: default
spec:
  scaleTargetRef:                                ##绑定名为 nginx 的 Deployment。
    apiVersion: apps/v1
    kind: Deployment
    name: nginx
  minReplicas: 1
  maxReplicas: 10
  metrics:
  - type: Resource
    resource:
      name: cpu
      targetAverageUtilization: 50
```

4. 执行 `kubectl describe hpaname` 会发现有类似以下信息的警告信息。

```
Warning  FailedGetResourceMetric       2m (x6 over 4m)  horizontal-pod-autoscaler  missing
request for cpu on container nginx in pod default/nginx-deployment-basic-75675f5897-mqzs7
Warning  FailedComputeMetricsReplicas  2m (x6 over 4m)  horizontal-pod-autoscaler  failed
to get cpu utilization: missing request for cpu on container nginx in pod default/nginx-de
ployment-basic-75675f5创建好HPA后，再次执行kubectl describe hpa
```

5. 创建好 HPA 后，再次执行 `kubectl describe hpaname` 命令。如果看到以下信息，则表示 HPA 已正常运行。

```
Normal SuccessfulRescale 39s horizontal-pod-autoscaler New size: 1; reason: All metrics be
low target
```

此时，当 Nginx 的 Pod 的利用率超过本例中设置的 50% 利用率时，则会进行水平扩容，低于 50% 时会进行缩容。

9.5.10. 运维单

借助运维单功能，您可以重启或替换单个或多个应用服务的 Pod。本文介绍如何通过控制台创建运维单，以及查看运维单详情等。

背景信息

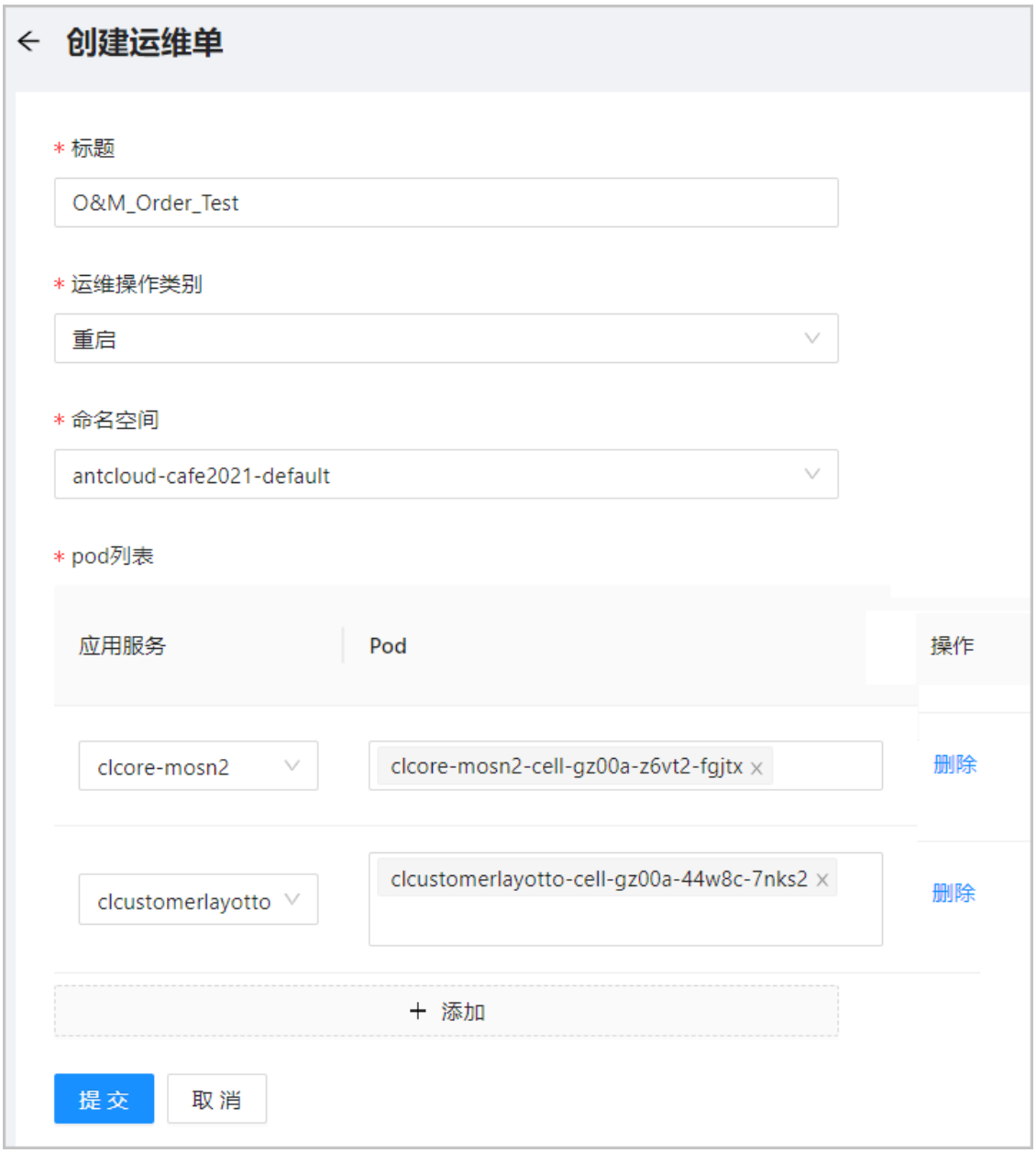
运维单发布提供如下运维功能：

- **重启**：通过指定 Pod 进行重启操作。
- **替换**：通过直接删除 Pod，CafeDeployment 会自动生成新的 Pod，从而实现 Pod 的替换。

创建运维单

1. 登录控制台。
2. 在左侧导航栏单击 **发布运维** > **运维单**。
3. 在运维单列表页，单击 **创建运维单**。
4. 在 **创建运维单** 页，填写以下信息后，单击 **提交**。
   - **标题**：输入运维单标题。
   - **运维操作类别**：支持 **重启**、**替换**。本例中选择 **重启**。

- **命名空间**：从下拉列表选择对应命名空间。
- **Pod 列表**：单击 **添加**，选择应用服务、Pod。



5. 单击 **提交**。
   在运维单列表页，单击创建的运维单会跳转到 **运维单详情** 页面，单击 **整体发布** 即可开始发布。

查看运维单详情

在 **运维单详情** 页面，您可以查看运维单的基本信息和执行详情。

- **基本信息**：运维单 ID、执行人、发布类型等。
- **应用服务**：分组发布应用服务，Pod 状态说明如下：
  - **未更新**：Pod 配置未更新为最新版本。
  - **更新中**：Pod 配置已更新为最新版本，但状态还未 ready 或保持 ready 状态少于 30 秒。
  - **更新成功**：Pod 配置已更新为最新版本，并保持 ready 状态 30 秒以上。

- **更新失败：上线或更新 Pod 异常。**



其他操作

- 如需在发布运维单过程中终止发布操作，在 **运维单详情** 页，单击 **取消** 即可取消当前运维单发布。
- 若运维单初始化失败，则无法进行发布，可以将鼠标悬停到 **失败原因** 上，查看具体原因，或单击 **重新初始化**。

9.6. 流量管理

9.6.1. 应用层流量管理

应用层流控

单元化架构下应用层流控规则主要涉及两块内容：UID 分配和部署单元的流量比。UID 分配指各个本地域逻辑单元组（RZoneGroup）所对应的 UID 分片（00-99）；部署单元流量比指各个部署单元所占对应的逻辑单元组（GZoneGroup、CZoneGroup、RZoneGroup）的流量权重。



上图中流量分配情况表示两个 RZone（RZ01 和 RZ02） 都承载流量，UID 分片在 0-19 的请求会访问到 RZ01，20-99 UID 分片的请求会访问到 RZ02。在 RZ01 和 RZ02 这两个逻辑单元内，只存在一个部署单元，分别是 RZ01A 和 RZ02A，所以每个部署单元就承载了 100% 的逻辑单元流量。逻辑单元 GZ01 包含两个部署单元 GZ01A 和 GZ01B，各占 50% 的逻辑单元流量。

正式流量

> 说明
> - 压测流量与正式流量的功能使用一致，唯一区别在于压测流量规则仅适用于压测场景。
> - 压测流量仅在专有云环境中支持。

在左侧导航栏单击 **流量管理 > 应用层**，在 **正式流量** 页签中，系统会先展示最近一次推送成功的全局流量规则，即当前生效的全局流量快照。当环境初始化时，在没有任何流量推送记录的情况下，或者推送的不是全局流量规则的情况下，不会展示任何内容。



规则配置

1. 单击 **规则配置** 进入 **正式流量规则配置** 页面，该页面展示的是当前已配置的规则（只是配置，不代表已生效）。
2. 单击逻辑单元卡片或分片数据即可进入编辑页面，填写以下配置信息：
   - 对于 RZone 类型的部署单元，配置对应部署单元的 UID 分片配置和流量权重配置。
   - 对于 GZone 类型的部署单元，配置对应部署单元的流量权重配置。

> **说明**
> 请确保调整后的各个部署单元的流量比总和为 100%。UID 分片的取值范围为 00~ 99。

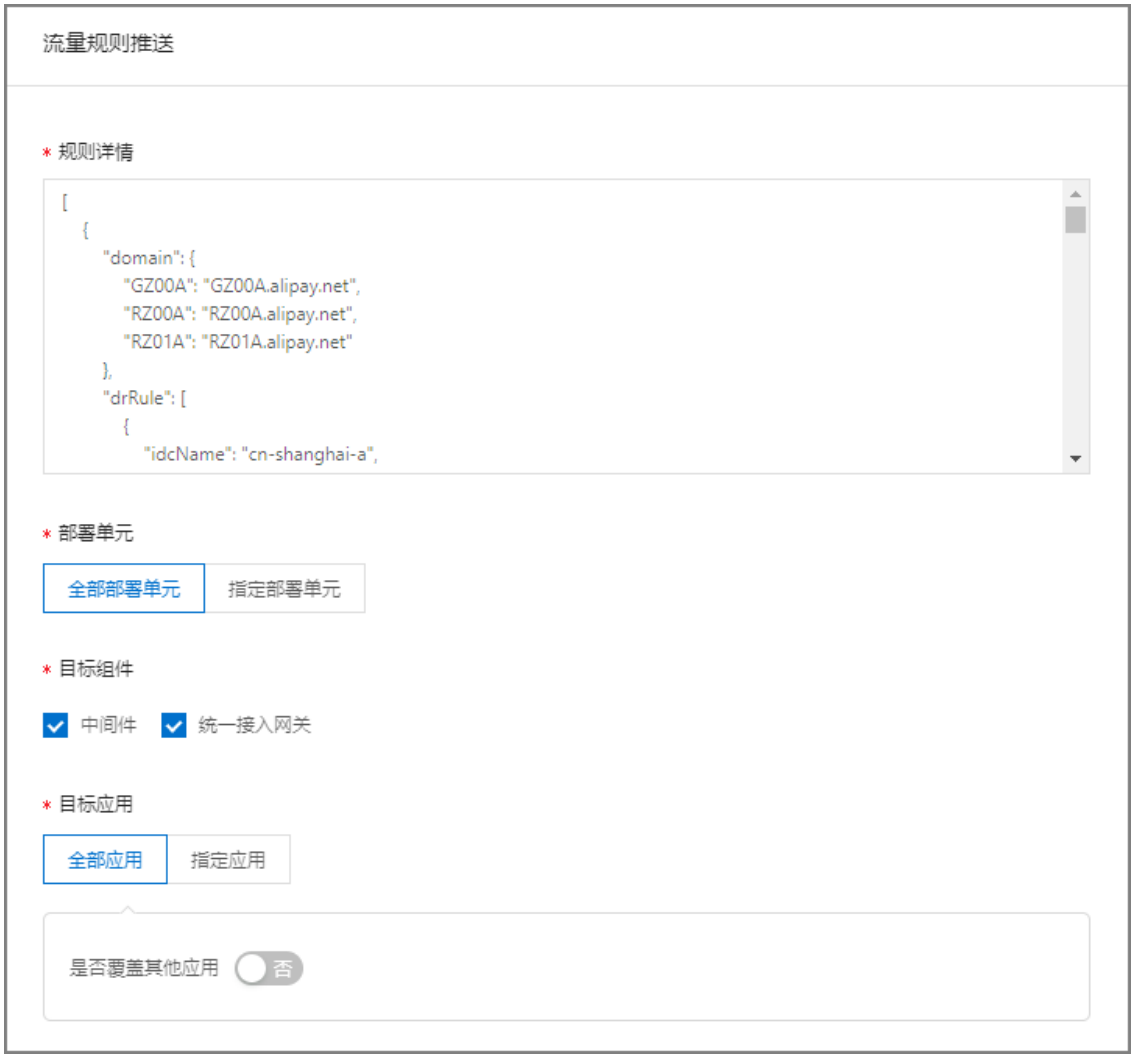




3. 调整权重或 UID 分片后，单击 **保存** 以确保配置生效。

流量规则推送

1. 在左侧导航栏单击 **流量管理 > 应用层。**
2. 单击 **规则配置** 进入 **正式流量规则配置** 页面。
3. 单击 **流量规则推送**，进入 **流量规则推送** 页面。
4. 设置以下推送配置：
   - **规则详情**：系统会根据当前配置的流量规则自动生成 JSON 格式的规则文本，该文本会作为实际推送到应用或中枢服务的参数，让全局应用或指定应用感知到流量变化。
   - **部署单元**：选择 **全部部署单元** 或 **指定部署单元**。可指定部署单元进行规则推送，针对局部部署单元内的应用生效。
   - **目标组件**：选择 **中间件** 或 **统一接入网关**。一般情况下流量规则的变更都需要让这两个组件感知，除非一些灰度测试场景。
   - **目标应用**：选择需要感知规则变化的目标应用，一般来说会选择全局应用推送，但在灰度或测试场景也可以只让部分应用感知新的流量规则做一些流量验证。
   - **是否覆盖其他应用**：选择推送到全局应用时，若开启覆盖其他应用，会用新配置的流量规则覆盖掉应用当前生效的规则。不覆盖，则会保留之前应用单独推送的规则。



5. 单击 **推送**，开始推送当前配置完的流量规则。

推送历史

在 **推送历史** 页签中，您可以查看最近推送的记录及状态。

- **关联应用**：routeClient 表示推送全局应用规则，否则会展示实际推送勾选的目标应用。
- **目标部署单元/机房**：展示推送的目标机房或部署单元。
- **规则类型**：目前仅支持路由规则。
- **推送结果**：展示成功或失败。错误信息会在推送失败后展示一些可供排查的失败原因。

	关联应用	目标部署单元/机房	规则类型	推送结果	错误信息	操作人	推送时间
+	routeClient	全部	路由规则	成功		zonemng	2020/03/17 23:19:37
+	routeClient	全部	路由规则	成功		zonemng	2020/03/17 22:41:00
+	routeClient	全部	路由规则	成功		zonemng	2020/03/17 15:50:11
+	routeClient	全部	路由规则	成功		zonemng	2020/03/17 14:20:34

9.7. 网络

9.7.1. 打通多集群容器网络

目前在单元化架构场景中，多个 ACK 集群的容器之间网络互通需要通过为集群安全组添加访问授权的方式实现。

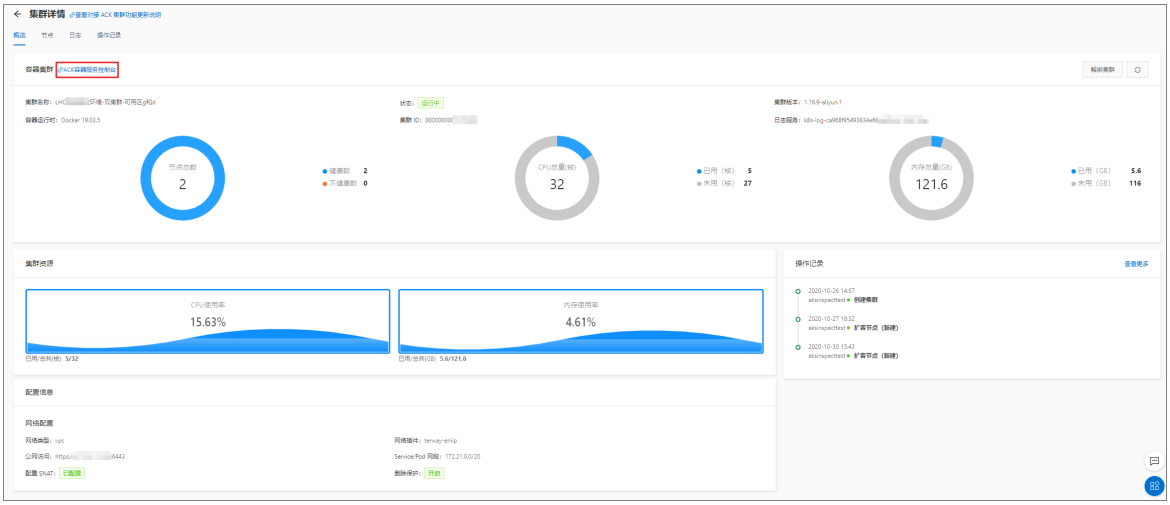
> **说明**
> 某个集群的安全组必须和其他所有集群的安全组打通，若总集群个数为 N，则每个集群的打通条目为 N-1，整体需要做的打通次数是 N*(N-1)。

本文以双集群环境为例，打通集群 1 和集群 2 的安全组。

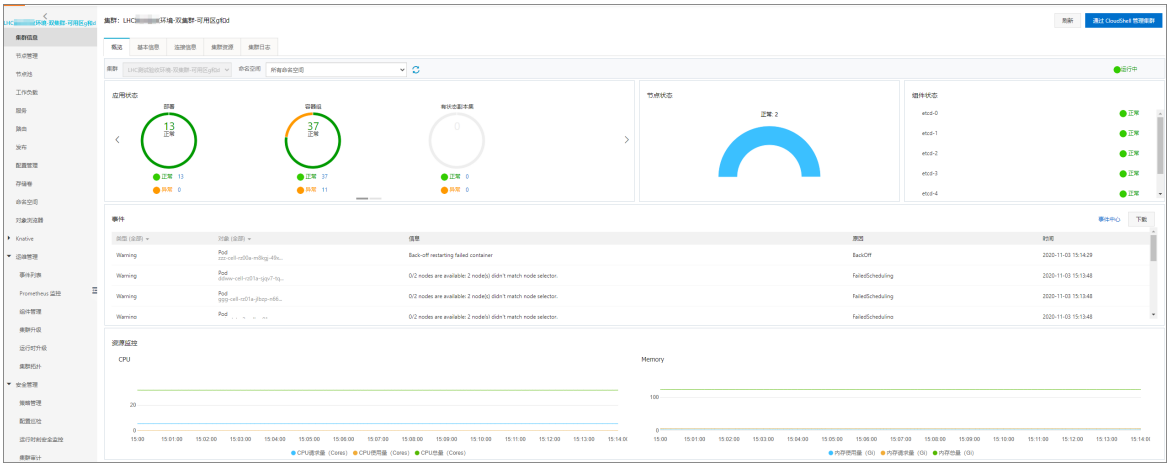


操作步骤

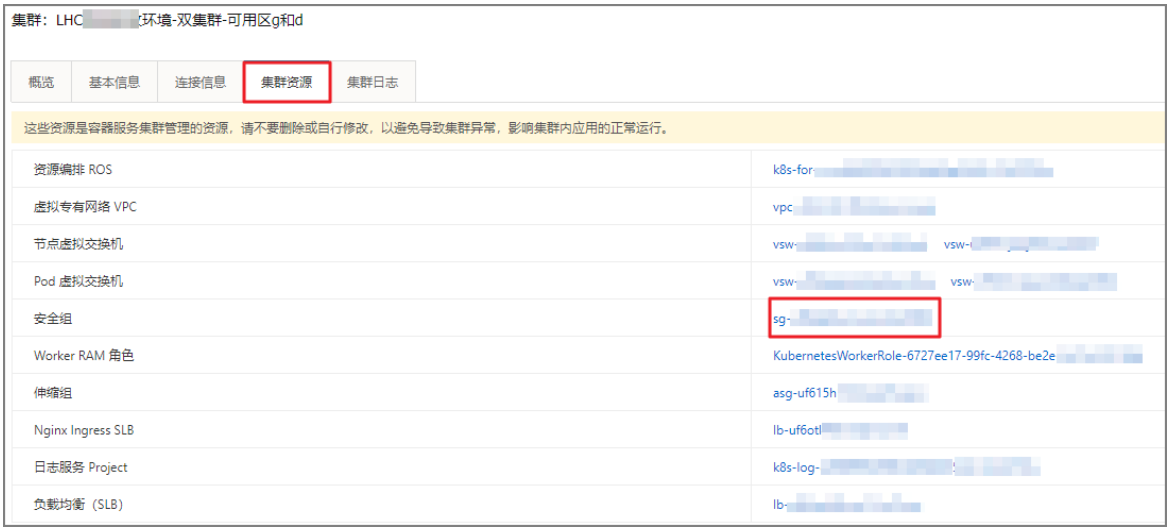
1. 登录控制台。在左侧导航栏选择 **集群管理 > 集群详情**。
2. 单击目标集群卡片进入 **集群详情** 页。



3. 单击左上角的 **ACK 容器服务控制台** 跳转到 ACK 的集群信息概览页。



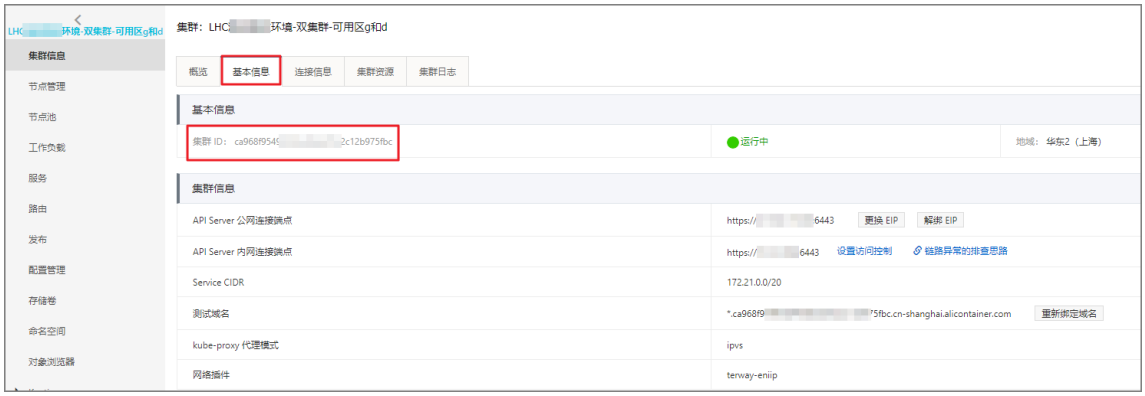
4. 单击 **集群资源** 页签，在集群管理的资源列表中，单击安全组链接，进入 **安全组规则** 页面。



5. 在 **入方向** 页签中，单击 **手动添加**，填入以下信息后，单击 **保存**。
   - **授权策略**：保持默认 **允许**。
   - **优先级**：优先级可选范围为 1~100，默认值为 1，即最高优先级。保持默认。
   - **协议类型**：选择 **全部**。
   - **端口范围**：保持默认 `-1/-1`，代表不限制端口。
   - **授权对象**：选择待被授权集群所在的安全组 ID。



> **说明**
> 安全组 ID 后半部分是集群 ID，可以在 **ACK 控制台 > 集群信息 > 基本信息** 页签获取。



9.7.2. 创建统一接入集群

统一接入集群承担统一接入和单元化网关的角色，在每个 K8s 集群内统一接入集群平行发布运维，多个平行的统一接入集群联合成一个联邦接入集群，是一个逻辑的集群实体。统一接入网关以集群的方式来承载业务负载均衡的流量，您需要根据流量的类型来规划集群，比如公网、内网等。

> 说明
> 集群上线后，您需要创建统一接入实例，并为应用服务创建负载均衡/统一接入类型的访问配置，才能进行各类业务流量的接入及路由，具体的转发及监听规则需在 应用服务访问配置 中设定。

组件作用

统一接入网关的作用主要是请求转发，即根据用户请求中的路由类型、域名和路径信息，识别出用户请求的目标应用并转发。

路由类型

路由类型主要分为 Gzone 路由（非单元化路由）和 Rzone 路由（单元化路由），详情如下

- Gzone 路由：根据目标应用和目标单元类型，随机找出其路由的目标服务器组，如果存在多个目标服务器组，则按照目标服务器组的权重来进行流量分发。
- Rzone 路由：当识别到用户请求最终目标为 Rzone 服务器时，通过用户请求的 Cookie 对锁进行解析，并计算出确定的 Rzone 路由目标。解析步骤如下：
  i. 锁的 Key 为 zone_op，通过 Key 计算出其 Value 值。
  ii. 根据该 Value 值进行目标单元计算和转发。

路由示例

根据 zone_op 对应 Value 值的范围，假设我们实现设置了如下路由规则：

- [00,49]， 路由到 rz00 单元。
- [50,99]， 路由到 rz01 单元。

则路由结果如下：

- 当 zone_op 的 Value 值不在 [00,99] 范围内：会将流量路由到当前 spanner 所在机房的目标应用对应的服务器组中，如果存在多个目标服务器组，则按照目标服务器的权重进行流量分发。
- 当 zone_op 的 Value 值位于 [00,49] 范围内：会将流量路由到目标应用的 rz00 单元所在目标服务器中，如果存在多个目标服务器组，则按照目标服务器的权重进行流量分发。
- 当 zone_op 的 Value 值位于 [50,99] 范围内：会将流量路由到目标应用的 rz01 单元所在目标服务器中，如果存在多个目标服务器组，则按照目标服务器的权重进行流量分发。

使用场景

主要使用场景如下：

- 在异地多活、弹性流量和容灾切换场景下，需要对机房流量进行分配和再分配（划拨、切入和切出）。

- 动态修改路由规则之后，接入路由组件会根据新的路由规则，将业务流量路由至新的目标机房，从而实现流量分配、弹性容量和容灾切换。
- 部署方式：按机房部署，即各个机房各部署一套接入路由集群。

注意事项

- 统一接入集群的容器只能调度到专有的统一接入类型的节点上，因此，创建统一接入集群时需要保证有足够的资源供调度。
- 在创建集群后报资源不足调度失败错误时，您需要在集群管理页面进行节点扩容，并选择 **统一接入节点** 类型。详情请参考 扩容集群。
- 统一接入集群横跨各可用区，所以需要同时保证各可用区资源充足。

操作步骤

1. 登录控制台。
2. 在左侧导航栏中选择 **网络 > 统一接入集群**，进入网管集群列表页面。
3. 单击 **创建网关集群**，完成以下配置后，单击 **提交**。

i. 完成基本信息配置。

配置项	说明
集群名称	只允许包含小写字母、数字、中划线，且必须以字母开头、以字母或数字结尾。长度为 1-29 个字符。
网络类型	可指定的网络类型包括内网、公网。指定集群的网络类型后，该集群上仅能创建与之相同网络类型的统一接入实例。
初始集群状态	可指定 **在线** 和 **维护**。 ■ **在线**：网关集群对外提供流量转发服务，同时可以接受转发规则的更新。 ■ **维护**：网关集群仅对外提供流量转发服务，但不接受转发规则的更新请求。
容器规格	统一接入集群内单个节点的总资源等于控制平面固定资源和数据平面自定义资源之和。容器的规格和节点的数量决定了该集群的请求处理能力。具体数据请参考 压测结果。 ■ **控制平面固定资源**：固定规格为 1C/1G/10G（CPU/内存/磁盘）。 ■ **数据平面自定义资源**：可自定义，推荐使用的最小规格为 4C/8G/40G，CPU：内存：磁盘容量的配比为 1：2：10。通常情况下，每 1C/2G/10G 的数据平面规格可以支持 3200 QPS 的 HTTP 短连接。

配置项	说明
host 网络模式	网关集群所采用的网络模型。 ■ 若选择 是，Pod 会复用容器所在宿主机的 IP。 ■ 若选择 否，每个 Pod 会拥有独立 IP。 说明 高密度部署模式下，仅可以选择 否，即选择非 host 模式。

ii. 完成容器版本配置。

配置项	说明
名称	网关容器的版本名称。长度为 1-5 个字符。只包含小写字母、数字、中划线，必须以字母开头，以字母或数字结尾。
容器镜像	网关容器的镜像地址，输入 `registry.cn-hangzhou.aliyuncs.com/sofastack/spannerplus:1.1.0` 。
副本数	网关节点的数量。出于容灾需要，建议一个集群至少包含 3 个以上节点。可以按照对容量的需求，合理规划容器规格与节点数量。 说明 创建副本数为 0 的网关集群会导致集群不可用，需要重新扩容。
标签	输入 key、value，为该版本的网关容器添加对应的 label。

9.7.3. 统一接入节点性能压测

本文介绍 LHC 统一接入节点性能压测相关内容，包括压测环境、压测方案和压测结果等信息。

压测概述

产品名称	部署拓扑/规格	性能指标概述
单元化应用服务 LHC	4C8G，4 个统一网关集群节点可以支持的最大 QPS 数量。	• HTTP 短连接：9000 QPS • HTTP 长连接：11000 QPS • HTTPS 短连接（不复用 session）：1500 QPS • HTTPS 短连接（复用 session）：4000 QPS • HTTPS 长连接：9000 QPS

压测环境

本次压测所使用的机器配置信息如下：

- ECS 类型：ecs.sn2ne.2xlarge 系列，规格为 8 vCPU、32 GiB 内存。

- **操作系统**：CentOS Linux 7 (Core)
- **内核版本**：3.10.0-1062.18.1.el7.x86_64

压测方案

> 说明
> 本次 Spanner 跟应用混部，节点上没有特殊的网络配置。

- **配置**：
  - upstream 关闭长连接。
  - 采用 rule-tt-vsg 精细化引流配置。
- **压测 endpoint**：直接压测 Spanner IP。
- **压测工具**：wrk
- **后端服务**：nginx
- **返回 body 大小**：本次压测没有考虑不同 body 大小的场景。
- **Spanner 配置**：单 worker 节点配置为 4C8G，4 个 worker 节点。
- HTTPS 请求协议为 **TLSv1.2**、**加密套件**：`ECDHE-RSA-AES256-GCM-SHA384`
- **统计方式**：`tail -f access.log | grep " 200 " | awk '{print $3}' | uniq -c`

测试详情

HTTP 短连接

```
wrk -t4 -c 60 -d 200   -H 'Connection: Close' -H 'Host: XXX' ``http://IP:PORT
```

9000 QPS 左右

```
  8820 [20/Nov/2020:17:10:16
  8601 [20/Nov/2020:17:10:17
  9232 [20/Nov/2020:17:10:18
  9243 [20/Nov/2020:17:10:19
  9291 [20/Nov/2020:17:10:20
  8903 [20/Nov/2020:17:10:21
  9446 [20/Nov/2020:17:10:22
  9337 [20/Nov/2020:17:10:23
  9346 [20/Nov/2020:17:10:24
  9244 [20/Nov/2020:17:10:25
  9454 [20/Nov/2020:17:10:26
  9086 [20/Nov/2020:17:10:27
  9474 [20/Nov/2020:17:10:28
  9016 [20/Nov/2020:17:10:29
  9315 [20/Nov/2020:17:10:30
  9315 [20/Nov/2020:17:10:31
  9317 [20/Nov/2020:17:10:32
  9062 [20/Nov/2020:17:10:33
  9088 [20/Nov/2020:17:10:34
  9252 [20/Nov/2020:17:10:35
  9073 [20/Nov/2020:17:10:36
  8057 [20/Nov/2020:17:10:37
  9244 [20/Nov/2020:17:10:38
  9194 [20/Nov/2020:17:10:39
  9154 [20/Nov/2020:17:10:40
  9451 [20/Nov/2020:17:10:41
  9316 [20/Nov/2020:17:10:42
  9252 [20/Nov/2020:17:10:43
  9271 [20/Nov/2020:17:10:44
  9246 [20/Nov/2020:17:10:45
  8004 [20/Nov/2020:17:10:46
```

HTTP 长连接

```
wrk -t4 -c 60 -d 200 -H 'Host: XXX' http://IP:PORT
```

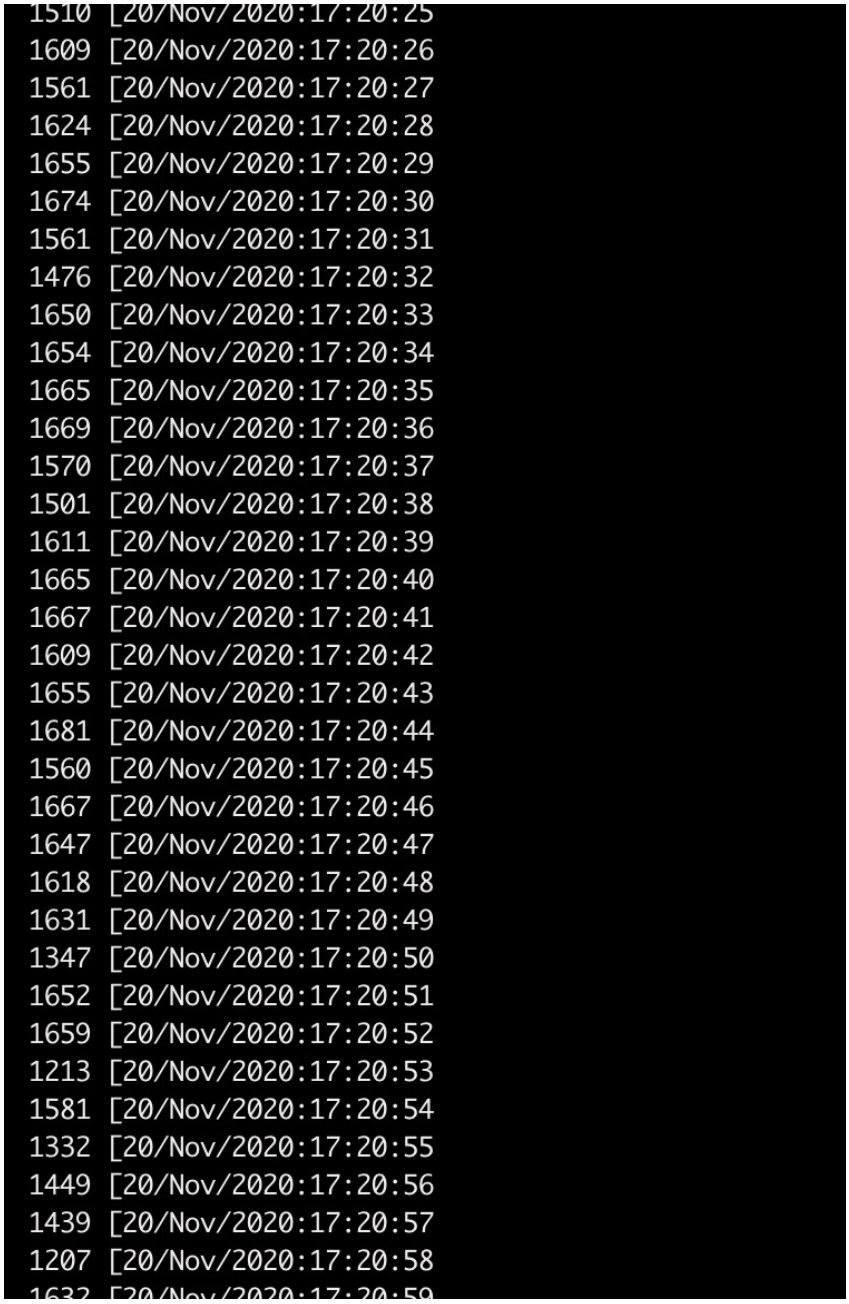
11000 QPS 左右

```
 6121 [20/Nov/2020:17:18:58
11572 [20/Nov/2020:17:18:59
11854 [20/Nov/2020:17:19:00
11576 [20/Nov/2020:17:19:01
11645 [20/Nov/2020:17:19:02
11885 [20/Nov/2020:17:19:03
11679 [20/Nov/2020:17:19:04
11613 [20/Nov/2020:17:19:05
11805 [20/Nov/2020:17:19:06
11655 [20/Nov/2020:17:19:07
11733 [20/Nov/2020:17:19:08
11543 [20/Nov/2020:17:19:09
11673 [20/Nov/2020:17:19:10
11638 [20/Nov/2020:17:19:11
11366 [20/Nov/2020:17:19:12
11496 [20/Nov/2020:17:19:13
11686 [20/Nov/2020:17:19:14
11576 [20/Nov/2020:17:19:15
11578 [20/Nov/2020:17:19:16
11508 [20/Nov/2020:17:19:17
11477 [20/Nov/2020:17:19:18
11641 [20/Nov/2020:17:19:19
11464 [20/Nov/2020:17:19:20
11484 [20/Nov/2020:17:19:21
11503 [20/Nov/2020:17:19:22
11481 [20/Nov/2020:17:19:23
 3748 [20/Nov/2020:17:19:24
```

HTTPS 短连接，不复用 session

```
wrk -t4 -c 60 -d 200   -H 'Host: XXX' -H 'Connection: Close' -m 0  https://IP:PORT
```

1500 QPS



HTTPS 短连接，复用 session

```
wrk -t4 -c 60 -d 200   -H 'Host: XXX' -H 'Connection: Close'  https://IP:PORT
```

4000 QPS

```
4923 [20/Nov/2020:17:22:28
5098 [20/Nov/2020:17:22:29
4899 [20/Nov/2020:17:22:30
4989 [20/Nov/2020:17:22:31
4355 [20/Nov/2020:17:22:32
4272 [20/Nov/2020:17:22:33
4372 [20/Nov/2020:17:22:34
4518 [20/Nov/2020:17:22:35
3844 [20/Nov/2020:17:22:36
4650 [20/Nov/2020:17:22:37
4068 [20/Nov/2020:17:22:38
3822 [20/Nov/2020:17:22:39
5085 [20/Nov/2020:17:22:40
4390 [20/Nov/2020:17:22:41
4485 [20/Nov/2020:17:22:42
3971 [20/Nov/2020:17:22:43
4137 [20/Nov/2020:17:22:44
4193 [20/Nov/2020:17:22:45
4411 [20/Nov/2020:17:22:46
3640 [20/Nov/2020:17:22:47
3927 [20/Nov/2020:17:22:48
4002 [20/Nov/2020:17:22:49
```

HTTPS 长连接

9000 QPS 左右

```
 180 [20/Nov/2020:17:23:27
9783 [20/Nov/2020:17:23:28
8667 [20/Nov/2020:17:23:29
9043 [20/Nov/2020:17:23:30
8692 [20/Nov/2020:17:23:31
9046 [20/Nov/2020:17:23:32
9226 [20/Nov/2020:17:23:33
8980 [20/Nov/2020:17:23:34
8691 [20/Nov/2020:17:23:35
9837 [20/Nov/2020:17:23:36
9471 [20/Nov/2020:17:23:37
9286 [20/Nov/2020:17:23:38
9608 [20/Nov/2020:17:23:39
9106 [20/Nov/2020:17:23:40
9538 [20/Nov/2020:17:23:41
8531 [20/Nov/2020:17:23:42
9006 [20/Nov/2020:17:23:43
8830 [20/Nov/2020:17:23:44
9395 [20/Nov/2020:17:23:45
8491 [20/Nov/2020:17:23:46
9671 [20/Nov/2020:17:23:47
8093 [20/Nov/2020:17:23:48
8829 [20/Nov/2020:17:23:49
```

压测结果

在上述压测环境下，单个网关节点支持的 QPS 如下：

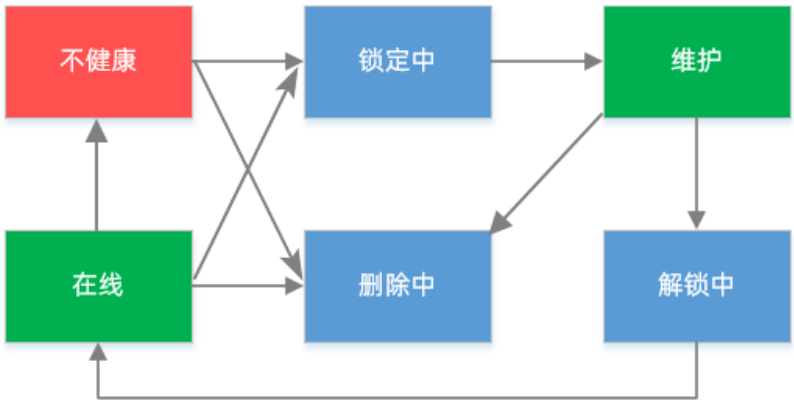
> 说明
> 不同 ECS 系列、CPU 规格、网卡型号、内核参数、及不同的请求都会影响网关节点的转发性能，具体以业务压测结果为准。

容器规格	HTTP 短连接	HTTP 长连接	HTTPS 短连接（不复用 session）	HTTPS 短连接（复用 session）	HTTPS 长连接
4C8G	9000 QPS	11000 QPS	1500 QPS	4000 QPS	9000 QPS

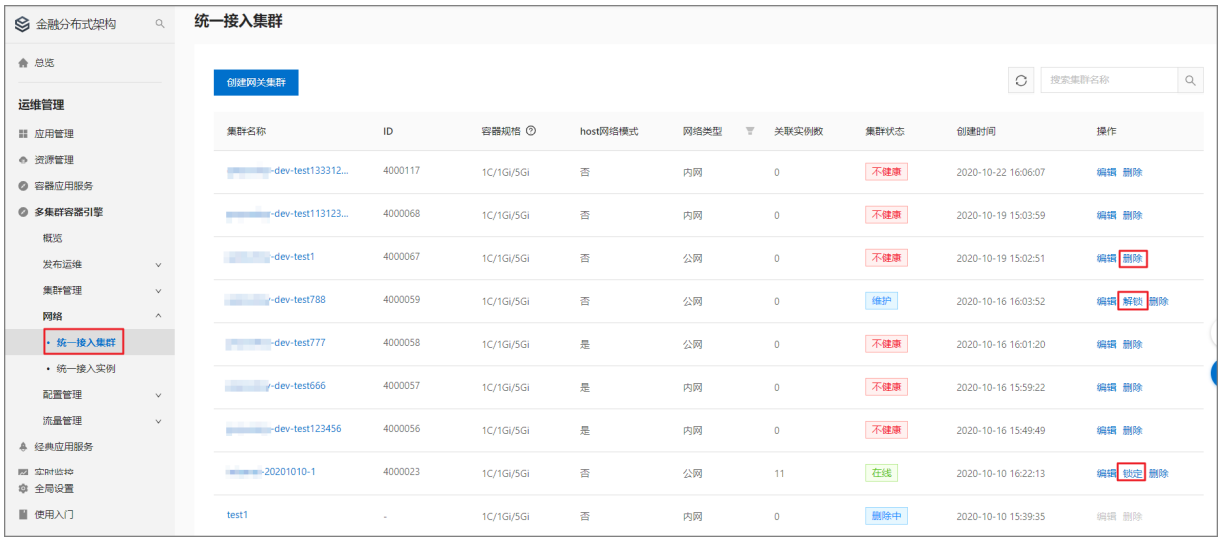
9.7.4. 网关集群状态说明

本文介绍统一接入集群在不同状态下的含义，便于您更深入地理解统一接入集群处于不同阶段的运行情况。
统一接入集群的状态可参考下表。

状态	状态属性	说明
不健康	稳定状态	统一接入集群处于不健康状态。 如果底层副本数不对或者集群创建出了问题，会显示 **不健康**。
健康—在线	稳定状态	正常，网关集群对外提供流量转发服务，同时可以接受转发规则的更新。
健康—维护	稳定状态	锁定，网关集群仅对外提供流量转发服务，但不接受转发规则的更新请求。
锁定中	中间状态	统一接入集群处于正在被锁定的状态。
删除中	中间状态	统一接入集群处于正在被删除的状态。



在创建统一接入集群时，您可以设置初始集群状态为在线或维护，集群创建成功后，可以通过操作栏中的 **锁定**、**解锁**、**删除** 选项来变更集群状态。

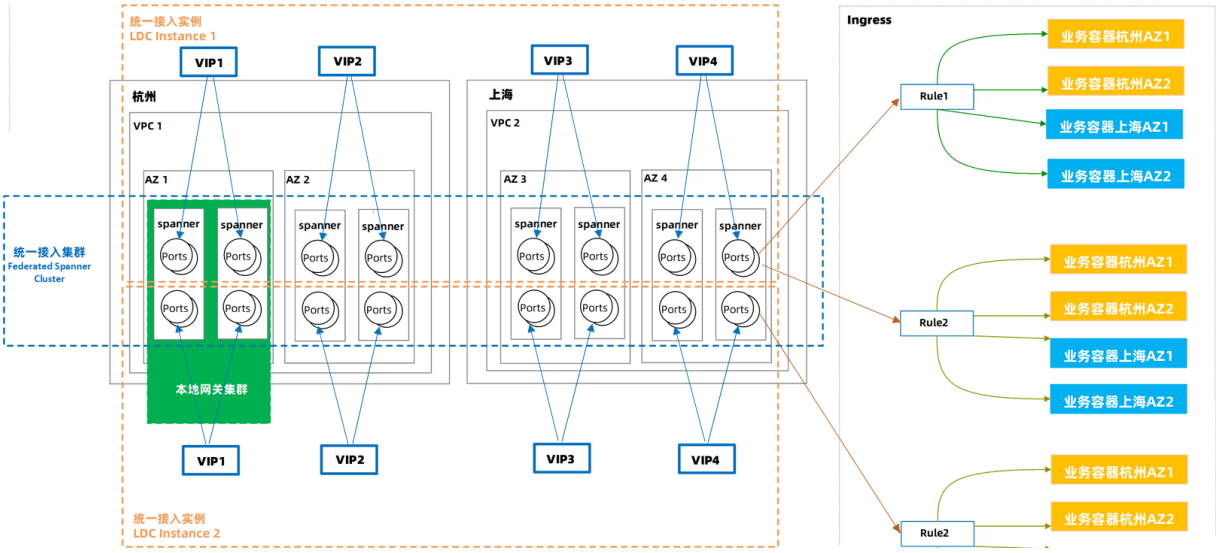


9.7.5. 创建统一接入实例

创建完统一接入网关集群后，您需要继续为网关集群节点绑定的统一接入实例，才能为业务应用提供统一的访问入口。

统一接入实例为一个 VIP 独立、统一接入集群共享的逻辑实例，在多个集群的情况下，该实例应该包含多个 VIP，需要前置域名作为统一入口。

统一接入实例本质上是共享某个统一接入集群的协议入口集合，目前统一接入实例支持 HTTP 和 HTTPS 两种协议，同时统一接入集群也会为每个实例中的每种协议生成一个单独的端口。统一接入实例落在联邦网关集群上，一个实例横跨多个用户集群，作为统一入口。



前提条件

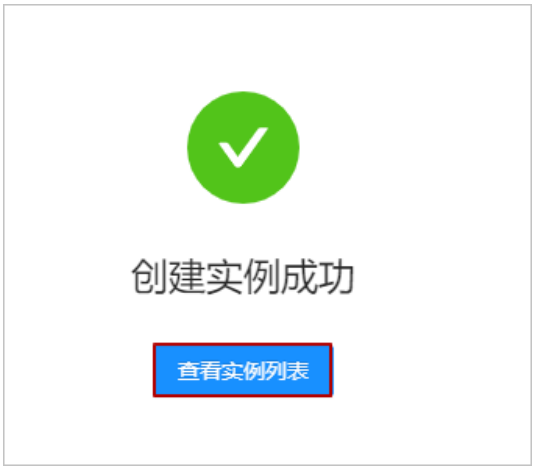
- 已创建统一接入集群。
- 网关集群的状态是 **在线** 或 **维护中**，且网络类型与接入实例的网络类型一致。

操作步骤

1. 登录 LHC 控制台。
2. 在左侧导航栏中选择 **网络 > 统一接入实例**，进入统一接入实例列表页面。
3. 单击 **创建实例**，填写以下配置信息。
   - **基本信息**：
     - **实例名称**：必填，允许包含小写字母、数字、中划线、且必须以字母开头、以字母或者数字结尾。允许长度为 1~34 个字符。不可与已有实例名称重复。
     - **网络类型**：支持 **内网** 或 **公网**。
     - **统一接入集群**：选择可用的接入集群。选择网络类型后，可选集群仅显示网络类型相符合的统一接入集群列表。
   - **配置信息**：
     - **HTTP 协议**：必填。输入前端端口号，范围为 1~65535。
     - **HTTPS 协议**：选填。单击 **添加 HTTPS 配置**，并输入前端端口号和证书 ID。

     > **说明**
     > 目前 HTTPS 协议的支持仍在内测中，不建议在生产环境中使用。若需测试，可以提交工单获取证书 ID。

4. 单击 **提交**。
   系统将弹出“创建实例成功”对话框，提示已成功创建统一接入实例。单击 **查看实例列表**，即可看到创建的实例出现在统一接入实例列表中。



9.7.6. 联邦负载均衡

9.7.6.1. 创建联邦负载均衡

联邦负载均衡，即不同部署单元负载均衡所引入的抽象层，每次在应用服务内应用时，只需引用了一个联邦的负载均衡 ID 即可，剩下关联不同部署单元的实际负载均衡，便交由联邦负载均衡管控。您可以根据业务场景选择配置对外公开或对内私有的联邦负载均衡，系统会根据您的选择分配公网或内网地址。本文介绍如何通过控制台创建联邦负载均衡。

操作步骤

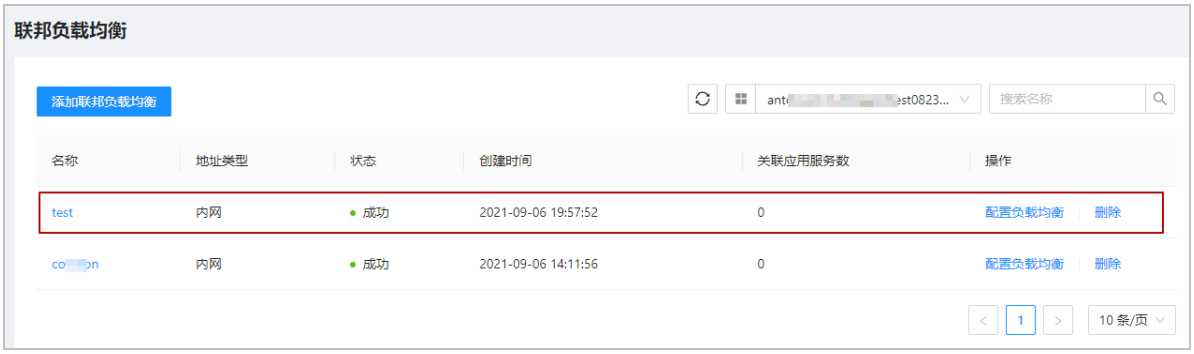
1. 登录控制台。
2. 在左侧导航栏单击 **网络 > 联邦负载均衡**，进入联邦负载均衡列表页面。
3. 单击 **添加联邦负载均衡**，在对话框中完成以下配置。

配置项	说明
名称	自定义新建联邦负载均衡名称。 只允许包含小写字母、数字、中划线，且必须以字母开头、以字母或数字结尾。长度为 1~29 个字符。
地址类型	可指定的地址类型包括内网、外网。 根据业务场景选择配置对外公开或对内私有的联邦负载均衡实例，系统会根据您的选择分配公网或私网服务地址。 ○ **外网**：外网实例仅提供公网 IP，可以通过 Internet 访问负载均衡服务。 ○ **内网**：内网实例仅提供私网 IP，只能通过内部网络访问该负载均衡服务，无法从 Internet 访问。
规格	选择实例规格类型。 说明 如需更高性能负载均衡规格，请前往云平台 SLB 控制台进行升配操作。
部署单元列表	根据实际业务诉求勾选部署单元。

4. 单击 **提交**。

结果验证

- 创建完成后，该联邦负载均衡将显示在联邦负载均衡列表中。



- 如果在创建应用服务的 访问配置 过程中选择使用该联邦负载均衡，**关联应用服务数** 列中的数量将更新。



9.7.6.2. 配置联邦负载均衡

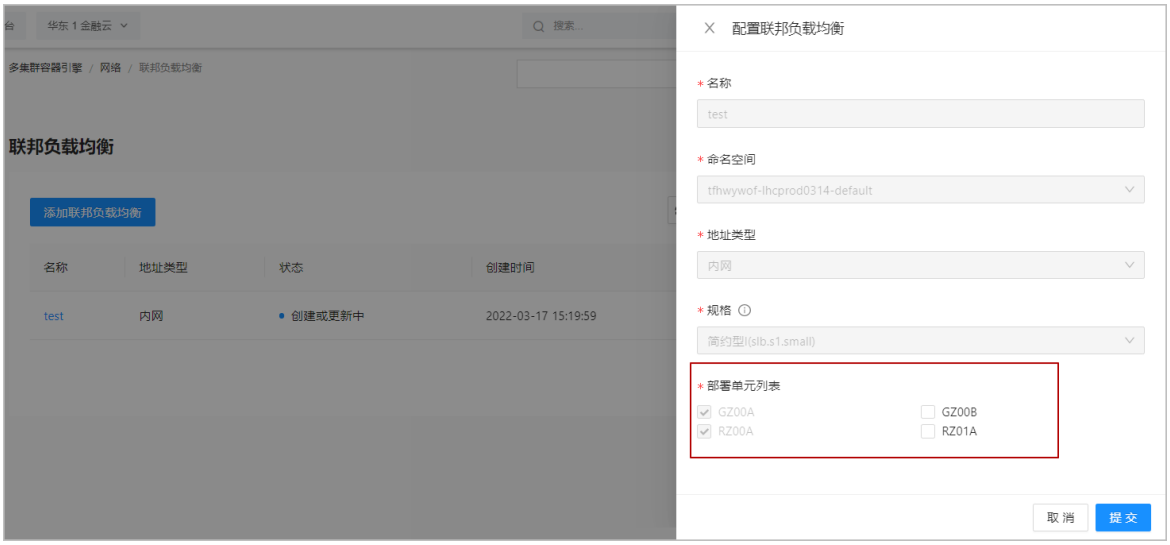
本文介绍如何通过控制台配置联邦负载均衡实例。

前提条件

已创建联邦负载均衡

操作步骤

1. 登录控制台。
2. 在左侧导航栏单击 **网络 > 联邦负载均衡**，进入联邦负载均衡列表页面。
3. 单击目标联邦负载均衡右侧操作列中的 **配置负载均衡。**
4. 在 **配置联邦负载均衡** 页，根据需求选择部署单元。



> 说明
> - 仅支持选择部署单元，其他配置项均不支持修改。
> - 对于部署单元列表内选中的部署单元，不支持取消选择。

5. 单击 **提交**。

9.7.6.3. 删除联邦负载均衡

在联邦负载均衡列表页，您可以删除不再需要的联邦负载均衡实例。本文主要介绍如何通过控制台删除联邦负载均衡。

前提条件

如果已使用联邦负载均衡（即已关联应用服务），需要先删除关联的应用服务（详见 删除应用服务），否则无法删除联邦负载均衡。

> 说明
> 后续应用服务版本管理将提供 **废弃历史版本** 能力，届时只需废弃关联了联邦负载均衡的应用服务版本后，即可删除关联的联邦负载均衡实例。

操作步骤

1. 登录控制台。
2. 在左侧导航栏单击 **网络 > 联邦负载均衡**，进入联邦负载均衡列表页面。
3. 单击目标联邦负载均衡右侧操作列中的 **删除 > 确定**。

相关文档

您可以参考以下文档，了解联邦负载均衡更多功能及操作：

- 创建联邦负载均衡
- 配置联邦负载均衡

9.8. 配置管理

9.8.1. 管理配置项

配置项（ConfigMap）是一种用于存储工作负载所需配置信息的资源类型，需要您自定义配置项的内容。

创建配置项

您可以添加联邦配置项，也可以为单个的集群添加配置项。

创建联邦配置项

1. 登录 LHC 控制台，在左侧导航栏单击 **配置管理 > 配置项**。
2. 单击 **创建**，在 **创建配置项** 页面，填写以下信息：
   - **命名空间**：选择所属命名空间。
   - **名称前缀**：配置项名称，同一个命名空间里命名必须唯一，最多 32 个字符。允许包含（小写）字母、数字、连字符，且必须以字母开头、以字母或者数字结尾。
   - **配置源**：添加键值对。变量名代表配置项数据的变量名，值代表配置项数据的值。
   - **自定义配置**：默认关闭。开启后，可以自定义部署单元级别的配置项。
3. 配置完成后，单击 **创建**。

创建单集群配置项

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击某个集群名称，进入 **集群详情** 页，单击 **配置** 页签。
3. 在左侧导航栏，选择 **配置项 (Configmaps)**。
4. （可选）选择所属命名空间，如果不选择，默认使用 **default**。
5. 单击 **创建**，在 **创建配置项** 页面，填写以下信息：
   - **命名空间**：所属的命名空间（已在上一步中选定，无需填写，默认为 **default**）。
   - **名称**：配置项名称，同一个命名空间里命名必须唯一，最多 32 个字符。允许包含（小写）字母、数字、连字符，且必须以字母开头、以字母或者数字结尾。
   - **配置源**：
     - **键值对**：添加键值对。变量名代表配置项数据的变量名，值代表配置项数据的值。
     - **文件**：可上传文件创建配置项。
6. 配置完成后，单击 **提交**。

修改配置项

修改配置文件会影响使用该配置文件的应用。

1. 在配置项列表页，单击 **名称** 进入配置项详情页。
2. 您可以进行以下修改：
   - 单击 **编辑**，修改已有的配置项数据，单击 **提交**。
   - 单击 **添加**，增加变量名和值，单击 **提交**。

9.8.2. 使用配置项

您可以在 Pod 中使用配置项，主要包括以下使用场景：

- 用配置项定义 Pod 环境变量
- 通过配置项设置命令行参数
- 在数据卷中使用配置项

> **说明**
> 在 Pod 里使用配置项时，需两者处于同一命名空间中。

更多关于配置项的信息，参见 Configure a Pod to Use a ConfigMap。

用配置项定义 Pod 环境变量

您可以使用配置项定义 Pod 中环境变量，使用 `valueFrom` 引用配置项数据。

1. 登录控制台，在左侧导航栏单击 **集群管理 > 集群详情。**
2. 在集群列表页，单击集群名称进入集群详情页。
3. 在左侧导航栏单击 **工作负载 > 容器组（Pods）**。
4. 单击 **从 YAML 创建**，在输入框中输入以下内容，单击 **创建**。下面是一个编排示例。

```
apiVersion: v1
kind:Pod
metadata:
name: config-pod-1
namespace: yournamespace  #改成当前的命名空间名称
spec:
containers:
- name: test-container
    image: busybox
    command:["/bin/sh","-c","env"]
    env:
- name: SPECIAL_LEVEL_KEY
        valueFrom:#使用valueFrom来指定env引用配置项的value值
          configMapKeyRef:
            name: special-config                #引用的配置文件名称
            key: SPECIAL_LEVEL                  #引用的配置项key
restartPolicy:Never
```

通过配置项设置命令行参数

您可以使用配置项设置容器中的命令或者参数值，使用环境变量替换语法 `$(VAR_NAME)` 来进行。

1. 登录控制台，在左侧导航栏单击 **集群管理 > 集群详情。**
2. 在集群列表页，单击集群名称进入集群详情页。
3. 在左侧导航栏单击 **工作负载 > 容器组（Pods）**。
4. 单击 **从 YAML 创建**，在输入框中输入以下内容，单击 **创建。**下面是一个编排示例。

```
apiVersion: v1
kind:Pod
metadata:
name: config-pod-3
namespace: yournamespace  # 改成当前的命名空间名称
spec:
containers:
- name: test-container
    image: busybox
    command:["/bin/sh","-c","echo $(SPECIAL_LEVEL_KEY) $(SPECIAL_TYPE_KEY)"]
    env:
- name: SPECIAL_LEVEL_KEY
        valueFrom:
          configMapKeyRef:
            name: special-config
            key: SPECIAL_LEVEL
- name: SPECIAL_TYPE_KEY
        valueFrom:
          configMapKeyRef:
            name: special-config
            key: SPECIAL_TYPE
restartPolicy:Never
```

运行该 pod 后，会输出如下结果。

```
very charm
```

在数据卷中使用配置项

您也可以在数据卷里面使用配置项，在 volumes 下指定配置项名称，会将 key/values 的数据存储到 mountPath 路径下（本例中是 `/etc/config` ）。最终生成以 key 为文件名，values 为文件内容的配置文件。

1. 登录控制台，在左侧导航栏单击 **集群管理 > 集群详情。**
2. 在集群列表页，单击集群名称进入集群详情页。
3. 在左侧导航栏单击 **工作负载 > 容器组（Pods）** 。
4. 单击 **从 YAML 创建**，在输入框中输入以下内容，单击 **创建。**下面是一个编排示例。

```
apiVersion: v1
kind:Pod
metadata:
name: config-pod-4
namespace: yournamespace  #改成当前的命名空间名称
spec:
containers:
- name: test-container
    image: busybox
    command:["/bin/sh","-c","ls /etc/config/"]#列出该目录下的文件名
    volumeMounts:
- name: config-volume
      mountPath:/etc/config
volumes:
- name: config-volume
    configMap:
      name: special-config
restartPolicy:Never
```

运行 pod 后，会输出配置项的 key。

```
SPECIAL_TYPE  SPECIAL_LEVEL
```

9.8.3. 管理保密字典

保密字典（Secret）是一种用于存储工作负载所需要认证信息、密钥的敏感信息等的资源类型。

创建保密字典

您可以创建全局的保密字典，也可以为单个集群创建保密字典。

创建联邦保密字典

1. 登录 LHC 控制台，在左侧导航栏单击 **配置管理 > 保密字典。**
2. 在保密字典列表页，单击 **创建。**
3. 在 **创建保密字典** 页，填写以下信息：
   - **命名空间**：选择所属命名空间。
   - **名称**：保密字典名称。
   - **保密字典类型**：
     - Opaque：一般密钥类型。输入变量名、值。值必须使用 Base64 编码。
     - ImagePullSecret：存放拉取私有仓库镜像所需的认证信息。输入容器注册地址、用户名和登录密码及（可选）邮箱。
   - **自定义配置**：开启后可以自定义部署单元级别的配置项，默认关闭。

4. 配置完成后，单击 **创建**。

创建单集群保密字典

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击目标集群名称，进入 **集群详情** 页，单击 **配置** 页签。
3. 在左侧导航栏，选择 **保密字典 (Secrets)**。
4. （可选）选择所属命名空间，如果不选择，默认使用 **default**。
5. 单击 **创建**，在 **创建保密字典** 页面，填写以下信息：
   - **命名空间**：所属的命名空间（已在上一步中选定，无需填写，默认为 **default**）。
   - **名称**：保密字典名称。
   - **保密字典类型**：
     - **Opaque**：一般密钥类型。输入变量名、值。值必须使用 Base64 编码。
     - **ImagePullSecret**：存放拉取私有仓库镜像所需的认证信息。输入容器注册地址、用户名及登录密码。
6. 配置完成后，单击 **提交**。

修改保密字典

对于 Opaque 类型的保密字典，您可以继续修改保密字典数据。

1. 登录 LHC 控制台，在左侧导航栏单击 **配置管理 > 保密字典**。
2. 在保密字典列表页，单击 **名称** 进入保密字典详情页。
3. 您可以进行以下修改：
   - 单击 **编辑**，修改已有的配置项数据，单击 **提交**。
   - 单击 **添加保密字典数据**，增加变量名和值，单击 **提交**。

查看保密字典

1. 登录 LHC 控制台，在左侧导航栏单击 **配置管理 > 保密字典**。
2. 在保密字典列表页，单击 **名称** 进入保密字典详情页。

> 说明
> 对于 Opaque 类型的保密字典，您可以继续添加保密字典数据。

删除保密字典

> 注意
> 删除保密字典可能会影响引用该保密字典的应用容器。建议您在确保没有引用该保密字典的应用容器后，再进行删除操作。

1. 在保密字典列表页，单击 **删除**。
2. 在弹出框中，单击 **确认**。

9.8.4. 使用保密字典

您可以在工作负载环境变量、命令行参数和数据卷三个场景使用保密字典（Secrets），也可以在 创建应用服务 的 配置 Pod 模板 步骤中使用保密字典拉取镜像。

前提条件

已创建保密字典。

> 说明
> 在 Pod 中使用保密字典时，Pod 和保密字典需处于同一命名空间中。

本文以如下保密字典为例，介绍保密字典的用法。

```
  apiVersion: v1
  kind:Secret
  metadata:
    name: mysecret
namespace: yournamespace  #改成当前的命名空间名称
  type:Opaque
  data:
    username:my-username #用户名
    password:******#需要用Base64编码
```

用保密字典配置 Pod 的数据卷

保密字典可以在 Pod 中作为文件使用。

1. 登录控制台，在左侧导航栏单击 **集群管理 > 集群详情**。
2. 在集群列表页，单击目标集群名称进入 **集群详情** 页。
3. 在左侧导航栏，单击 **工作负载 > 容器组（Pods）**。
4. 单击 **从 YAML 创建**，在输入框中输入以下内容，单击 **提交**。
   如以下 Pod 示例所示，mysecret 密钥的 username 和 password 以文件方式保存在 `/etc/foo` 目录下。

```
 apiVersion: v1
 kind:Pod
 metadata:
   name: mypod
namespace: yournamespace  #改成当前的命名空间名称
 spec:
   containers:
- name: mypod
     image: redis
     volumeMounts:
- name: foo
       mountPath:"/etc/foo"
       readOnly:true
   volumes:
- name: foo
     secret:
       secretName: mysecret
```

用保密字典设置 Pod 的环境变量

保密字典可以在 Pod 中设置为环境变量。

1. 登录控制台，在左侧导航栏单击 **集群管理 > 集群详情**。
2. 在集群列表页，单击目标集群名称进入集群详情页。
3. 在左侧导航栏，单击 **工作负载 > 容器组（Pods）**。

4. 单击 **从 YAML 创建**，在输入框中输入以下内容，单击 **提交**。
   如以下 Pod 示例所示，mysecret 密钥的 username 和 password 配置为 Pod 的环境变量。

```
 apiVersion: v1
 kind:Pod
 metadata:
   name: secret-env-pod
namespace: yournamespace  #改成当前的命名空间名称
 spec:
   containers:
- name: mycontainer
     image: redis
     env:
- name: SECRET_USERNAME
         valueFrom:
           secretKeyRef:
             name: mysecret
             key: username
- name: SECRET_PASSWORD
         valueFrom:
           secretKeyRef:
             name: mysecret
             key: password
   restartPolicy:Never
```

9.9. 工作负载

9.9.1. 部署（Deployments）

部署（Deployments）即运行中始终不保存任何数据或状态的工作负载，例如 Nginx。您可以在 LHC 控制台上轻松管理部署（Deployments），如执行创建、更新、删除等操作。

> 说明
> Deployments 是 kubernetes 原生概念。更多详情请参见 Kubernetes 官网文档。

创建 Deployments

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击目标集群名称，进入 **集群详情** 页，单击 **工作负载** 页签。
3. 在左侧菜单选项，选择 **部署（Deployments）**，然后单击 **创建 Deployments**。
4. 在弹出的输入框中输入 YAML 内容，单击 **提交**。
   YAML 样例如下：

```
 kind:Deployment
 metadata:
   name: deployment-example
   labels:
     k8s-app: deployment-example
namespace:default
 spec:
   replicas:1
   selector:
     matchLabels:
       k8s-app: deployment-example
template:
     metadata:
       labels:
         k8s-app: deployment-example
     spec:
       containers:
- name: nginx
         image: reg-cnsh.cloud.alipay.com/library/nginx:latest # 改成当前 region 的镜像地址
```

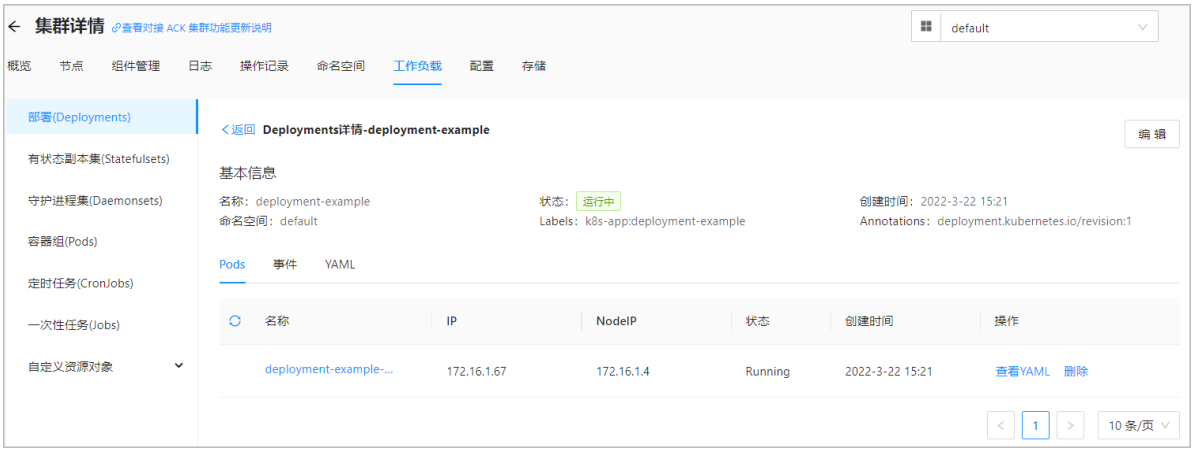
执行结果

创建成功后，名为 **deployment-example** 的 Deployments 将出现在列表页面中。



相关操作

- 如需查看 Deployments 详情，请在 Deployments 列表页单击目标 Deployments 名称，进入 Deployments 详情页。在该页面上，可以查看 Deployments 基本信息、Pod 列表信息、事件信息及 YAML 配置信息。



- 如需搜索 Deployments，请在 Deployments 列表页右上角的输入栏，输入 Deployments 名称关键字即可进行搜索。
- 如需更新 Deployments，请在 Deployments 列表页单击操作栏中的 **查看/编辑**，并在编辑 Deployments 对话框中编辑 YAML 信息，然后单击 **提交**。

- 如需删除 Deployments，请在 Deployments 列表页单击操作栏中的更多图标 > **删除**，然后在弹出的提示对话框中单击 **确定**。

9.9.2. 有状态副本集（Statefulsets）

有状态副本集（Statefulsets）即运行过程中保存数据或状态的工作负载，例如 MySQL。您可以在 LHC 控制台上轻松管理有状态副本集（Statefulsets），如执行创建、更新、删除等操作。

> **说明**
> Statefulsets 是 kubernetes 原生概念。更多详情请参见 Kubernetes 官网文档。

创建 Statefulsets

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击目标集群名称，进入 **集群详情** 页，单击 **工作负载** 页签。
3. 在左侧菜单选项，选择 **有状态副本集（Statefulsets）**，然后单击 **创建 Statefulsets**。
4. 在弹出的输入框中输入 YAML 内容，单击 **提交**。
   YAML 样例如下：

```
kind: StatefulSet
metadata:
  name: statefulset-example
  labels:
    k8s-app: statefulset-example
  namespace: default
spec:
  replicas: 1
  selector:
    matchLabels:
      k8s-app: statefulset-example
  template:
    metadata:
      labels:
        k8s-app: statefulset-example
    spec:
      containers:
        - name: nginx
          image: 'registry-vpc.cn-hangzhou.aliyuncs.com/sofa-samples/nginx:latest'
```

执行结果

创建成功后，名为 **statefulset-examplee** 的 Statefulsets 将出现在列表页面中。



相关操作

- 如需查看 Statefulsets 详情，请在 Statefulsets 列表页单击目标 Statefulsets 名称，进入 Statefulsets 详情页。在该页面上，可以查看 Statefulsets 基本信息、Pod 列表信息、事件信息及 YAML 配置信息。
- 如需搜索 Statefulsets，请在 Statefulsets 列表页右上角的输入栏，输入 Statefulsets 名称关键字即可进行搜索。
- 如需更新 Statefulsets，请在 Statefulsets 列表页单击操作栏中的 **查看/编辑**，并在编辑 Statefulsets 对话框中编辑 YAML 信息，然后单击 **提交**。
- 如需删除 Statefulsets，请在 Statefulsets 列表页单击操作栏中的更多图标 > **删除**，然后在弹出的提示对话框中单击 **确定**。

9.9.3. 守护进程集（Daemonsets）

守护进程集（Daemonsets）确保全部（或者部分）节点都运行一个 Pod 实例，支持实例动态添加到新节点。您可以在 LHC 控制台上轻松管理守护进程集（Daemonsets），如执行创建、更新、删除等操作。

> **说明**
> Daemonsets 是 kubernetes 原生概念。更多详情请参见 Kubernetes 官方文档。

创建 Daemonsets

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击目标集群名称，进入 **集群详情** 页，单击 **工作负载** 页签。
3. 在左侧菜单选项，选择 **守护进程集（Daemonsets）**，然后单击 **创建 Daemonsets**。
4. 在弹出的输入框中输入 YAML 内容，单击 **提交**。
   YAML 样例如下：

```
 kind:DaemonSet
 metadata:
   name: daemonset-example
   labels:
     k8s-app: daemonset-example
namespace:default
 spec:
   selector:
     matchLabels:
       k8s-app: daemonset-example
template:
     metadata:
       labels:
         k8s-app: daemonset-example
     spec:
       containers:
- name: nginx
         image: reg-cnsh.cloud.alipay.com/library/nginx:latest
```

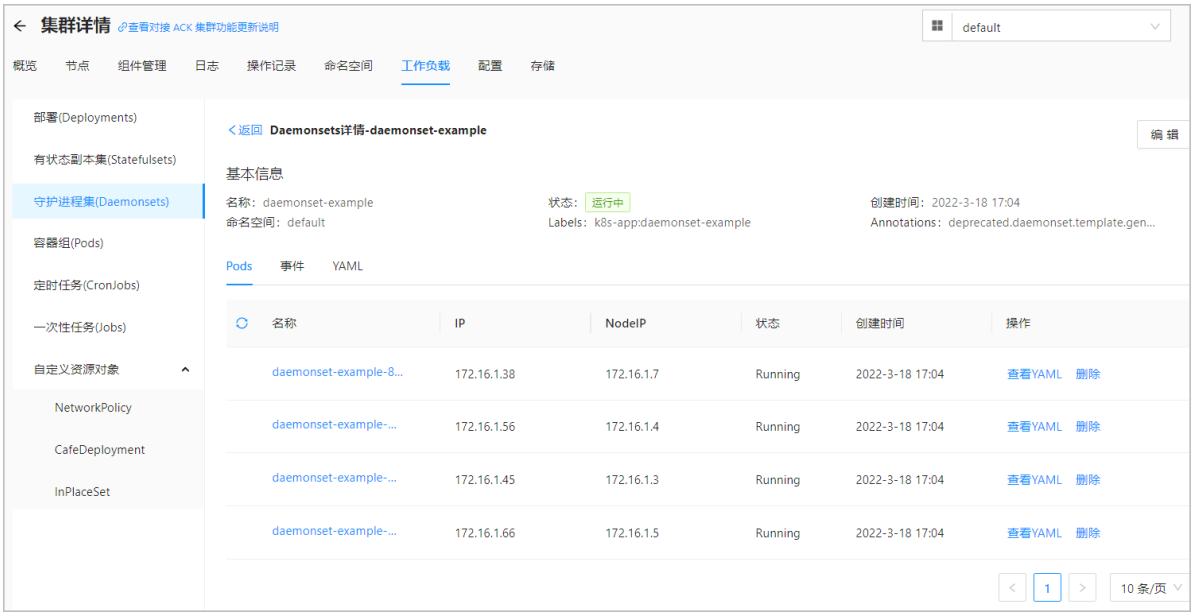
执行结果

创建成功后，名为 **daemonset-example** 的 Daemonsets 将出现在列表页面中。



相关操作

- 如需查看 Daemonsets 详情，请在 Daemonsets 列表页单击目标 Daemonsets 名称，进入 Daemonsets 详情页。在该页面上，可以查看 Daemonsets 基本信息、Pod 列表信息、事件信息及 YAML 配置信息。



- 如需搜索 Daemonsets，请在 Daemonsets 列表页右上角的输入栏，输入 Daemonsets 名称关键字即可进行搜索。
- 如需更新 Daemonsets，请在 Daemonsets 列表页单击操作栏中的 **查看/编辑**，并在编辑 Daemonsets 对话框中编辑 YAML 信息，然后单击 **提交**。
- 如需删除 Daemonsets，请在 Daemonsets 列表页单击操作栏中的更多图标 > **删除**，然后在弹出的提示对话框中单击 **确定**。

9.9.4. 容器组（Pods）

容器组（Pods）是 Kubernetes 中最小可部署单元，代表 Kubernetes 中一个独立的应用程序运行实例，该实例可能由单个容器或者多个紧耦合在一起的容器组成。

> 说明
> Pods 是 kubernetes 原生概念。更多详情请参见 Kubernetes 官方文档。

创建 Pods

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击目标集群名称，进入 **集群详情** 页，单击 **工作负载** 页签。
3. 在左侧菜单选项，选择 **容器组（Pods）**，然后单击 **创建 Pods**。
4. 在弹出的输入框中输入 YAML 内容，单击 **提交**。

YAML 样例如下：

```
kind: Pod
metadata:
  name: pod-example
  labels:
    k8s-app: pod-example
  namespace: default
spec:
  containers:
    - name: nginx
      image: 'registry-vpc.cn-hangzhou.aliyuncs.com/sofa-samples/nginx:latest'
```

执行结果

创建成功后，名为 **pod-example** 的 Pods 将出现在列表页面中。



相关操作

- 如需查看 Pods 详情，请在 Pods 列表页单击目标 Pods 名称，进入 Pods 详情页。在该页面上，可以查看 Pods 基本信息、容器及事件信息。



在 Pods 详情页，您还可以根据需要执行如下操作：

- 查看容器详情：单击想要查看详情的 Pod 名称，即可进入容器详情页，如下图所示。



- 前往容器服务控制台：在 Pods 详情页单击目标 Pod **操作** 列的 **前往控制台**，或者单击容器详情页上的 **控制台**，将跳转至容器服务控制台。可按需输入命令来管理容器。

- 如需搜索 Pods，请在 Pods 列表页右上角的输入栏，输入 Pods 名称关键字即可进行搜索。
- 如需更新 Pods，请在 Pods 列表页单击操作栏中的 **查看/编辑**，并在编辑 Pods 对话框中编辑 YAML 信息，然后单击 **提交**。
- 如需删除 Pods，请在 Pods 列表页单击操作栏中的更多图标 > **删除**，然后在弹出的提示对话框中单击 **确定**。

9.9.5. 定时任务（CronJobs）

定时任务（CronJobs ）用于创建周期性以及重复性的任务，例如数据库备份或者发送邮件。Job 负责批量处理短暂的一次性任务，即仅执行一次的任务，而 CronJob 则是基于 Job 增加了时间调度，即在指定时间周期内运行指定 Job。 本文介绍如何通过控制台创建 CronJobs 及其他常用操作。

> 说明
> CronJob 是 kubernetes 原生概念。更多详情请参见 Kubernetes 官方文档。

创建 CronJobs

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击目标集群名称，进入 **集群详情** 页，单击 **工作负载** 页签。
3. 在左侧菜单选项，选择 **定时任务（CronJobs）**，然后单击 **创建 CronJobs**。
4. 在弹出的输入框中输入 YAML 内容，单击 **提交**。
   YAML 样例如下：

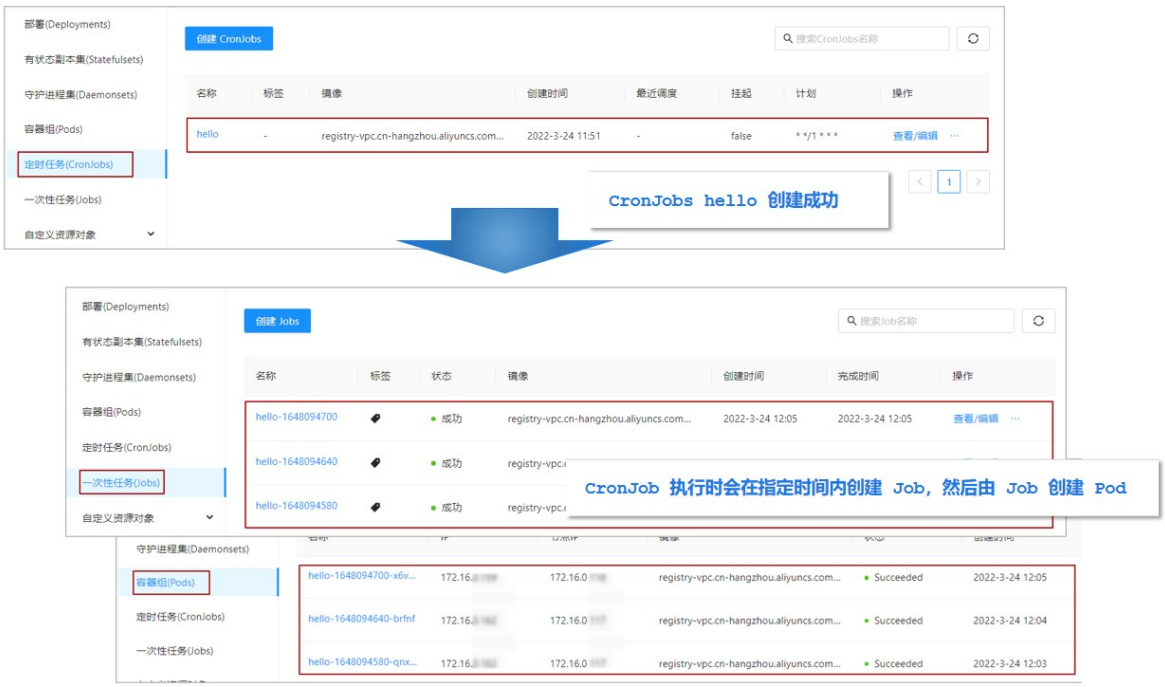
```
apiVersion: batch/v1beta1
kind: CronJob
metadata:
  name: hello
spec:
  jobTemplate:
    spec:
      template:
        spec:
          containers:
            - command:
                - /bin/sh
                - '-c'
                - date; echo Hello from the Kubernetes cluster
              image: 'registry-vpc.cn-hangzhou.aliyuncs.com/sofa-samples/nginx:latest'
              imagePullPolicy: IfNotPresent
              name: hello
          restartPolicy: Never
  schedule: '* */1 * * *'
```

参数说明可参考下表。

参数	说明
kind	标识 CronJob 资源类型。
metadata	CronJob 的基本信息，比如 CronJob 的名称。
spec.schedule	CronJob 执行的 Cron 的策略。即指定任务运行周期，格式同 Cron schedule。 `# 格式说明` `#  ┌──分钟（0 - 59）` `#	┌──小时（0 - 23）` `#		┌──日（1 - 31）` `#			┌──月（1 - 12）` `#				┌──星期（0 - 6）` `#				|` `# * * * * *`
spec.jobTemplate	Cron 执行的 Job 模板。即指定需要运行的任务，格式同 Job patterns。

执行结果

创建成功后，名为 **hello** 的 CronJobs 将出现在列表页面中。CronJob 执行时会在指定的时间创建出 Job，然后由 Job 创建出 Pod。



相关操作

- 如需查看 CronJobs 详情，请在 CronJobs 列表页单击目标 CronJobs 名称，进入 CronJobs 详情页。在该页面上，可以查看 CronJobs 基本信息、任务列表信息及事件信息。
- 如需搜索 CronJobs，请在 CronJobs 列表页右上角的输入栏，输入 CronJobs 名称关键字即可进行搜索。
- 如需更新 CronJobs，请在 CronJobs 列表页单击操作栏中的 **查看/编辑**，并在编辑 CronJobs 对话框中编辑 YAML 信息，然后单击 **提交**。
- 如需删除 CronJobs，请在 CronJobs 列表页单击操作栏中的更多图标 > **删除**，然后在弹出的提示对话框中单击 **确定**。

> 说明
> 对 CronJobs 执行删除时，不会删除 CronJobs 创建的 Jobs，如需删除对应的 Jobs，请前往 Jobs 列表页手动删除。

9.9.6. 一次性任务（Jobs）

一次性任务（Jobs）负责批量处理短暂的一次性任务，即仅执行一次的任务，保证批处理任务的一个或多个 Pod 成功结束。本文介绍如何通过控制台创建 Jobs 及其他常用操作。

> 说明
> Jobs 是 kubernetes 原生概念。更多详情请参见 Kubernetes 官方文档。

创建 Jobs

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击目标集群名称，进入 **集群详情** 页，单击 **工作负载** 页签。
3. 在左侧菜单选项，选择 **部署（Jobs）**，然后单击 **创建 Jobs**。
4. 在弹出的输入框中输入 YAML 内容，单击 **提交**。
   YAML 样例如下：

```
apiVersion: batch/v1
kind: Job
metadata:
  name: hello
spec:
  template:
    spec:
      containers:
        - name: hello
          image: 'registry-vpc.cn-hangzhou.aliyuncs.com/sofa-samples/nginx:latest'
          command:
            - /bin/sh
            - '-c'
            - date; echo Hello from the Kubernetes cluster
      restartPolicy: Never
  backoffLimit: 1
```

参数说明可参考下表。

参数	说明
kind	标识 Job 资源类型。
metadata	Job 的基本信息，比如 Job 的名称。
spec.Template	Job 管理的 Pod 的详细模板配置。
restartPolicy	任务的重启策略。支持 Never、OnFailure。 ○ Never：表示在任务需要重新运行时执行。 ○ OnFailure：表示在任务失败时再执行。
backoffLimit	任务的重试次数。用于控制任务重试次数。

执行结果

创建成功后，名为 **hello** 的 Jobs 将出现在列表页面中。



相关操作

- 如需查看 Jobs 详情，请在 Jobs 列表页单击目标 Jobs 名称，进入 Jobs 详情页。在该页面上，可以查看 Jobs

基本信息、Pod 列表信息、事件信息及 YAML 配置信息。



- 如需搜索 Jobs，请在 Jobs 列表页右上角的输入栏，输入 Jobs 名称关键字即可进行搜索。
- 如需更新 Jobs，请在 Jobs 列表页单击操作栏中的 **查看/编辑**，并在编辑 Jobs 对话框中编辑 YAML 信息，然后单击 **提交**。
- 如需删除 Jobs，请在 Jobs 列表页单击操作栏中的更多图标 > **删除**，然后在弹出的提示对话框中单击 **确定**。

9.9.7. 自定义资源对象

9.9.7.1. NetworkPolicy 概述

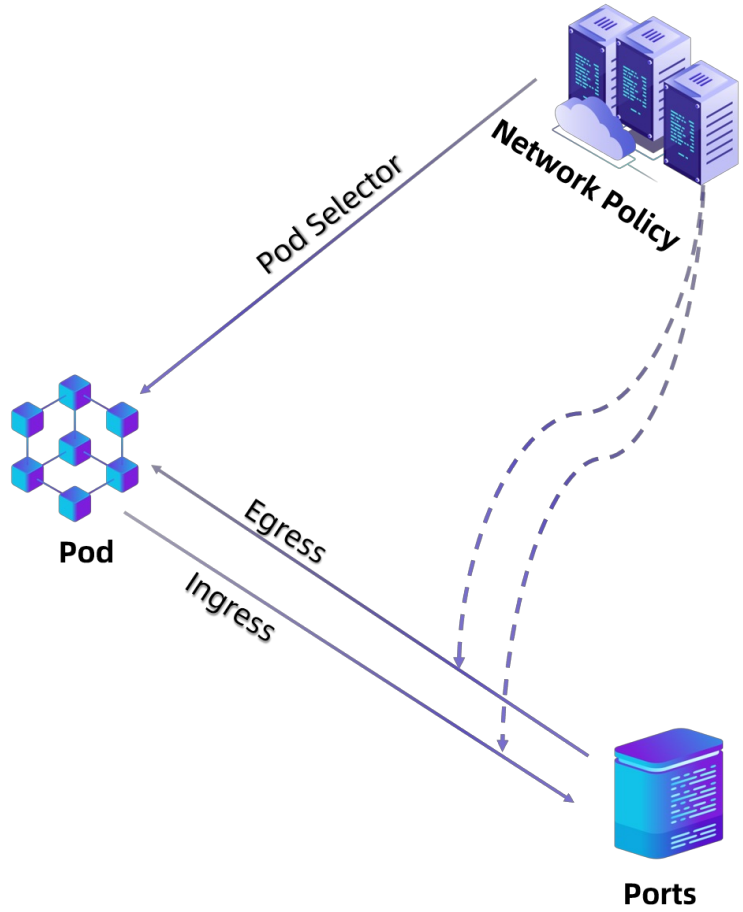
网络策略（NetworkPolicy）是 Kubernetes 提供的一种资源，提供基于策略的网络控制。如果您希望在 IP 地址或端口层面（OSI 三层或四层）控制网络流量，您可以针对集群内特定应用采用网络策略。

背景介绍

随着业务逻辑的复杂化，传统的单一外部防火墙或依照应用分层的防火墙已逐渐无法满足需求。
LHC 基于 Kubernetes 网络策略功能作了增强，通过配置网络策略，允许在集群内实现网络隔离，也就是可以在多组 Pod 之间架起防火墙。

基本原理

NetworkPolicy 会应用在与其相匹配的 PodSelector 的 Pod 上，基于入向（Ingress）或出向（Egress）规则指定允许访问这些 Pod 的源地址或这些 Pod 可以访问的目标地址。
Ingress 和 Egress 这两种规则可匹配由 PodSelector 选出的 Pod 或命名空间内的所有 Pod，或者通过 CIDR 指定的 IP 地址段。



说明

- Ingress：入方向流量规则，用于配置访问服务的源地址或端口白名单。
- Egress：出方向流量规则，用于配制 Pod 访问外部服务的目的地址或端口白名单。

相关文档

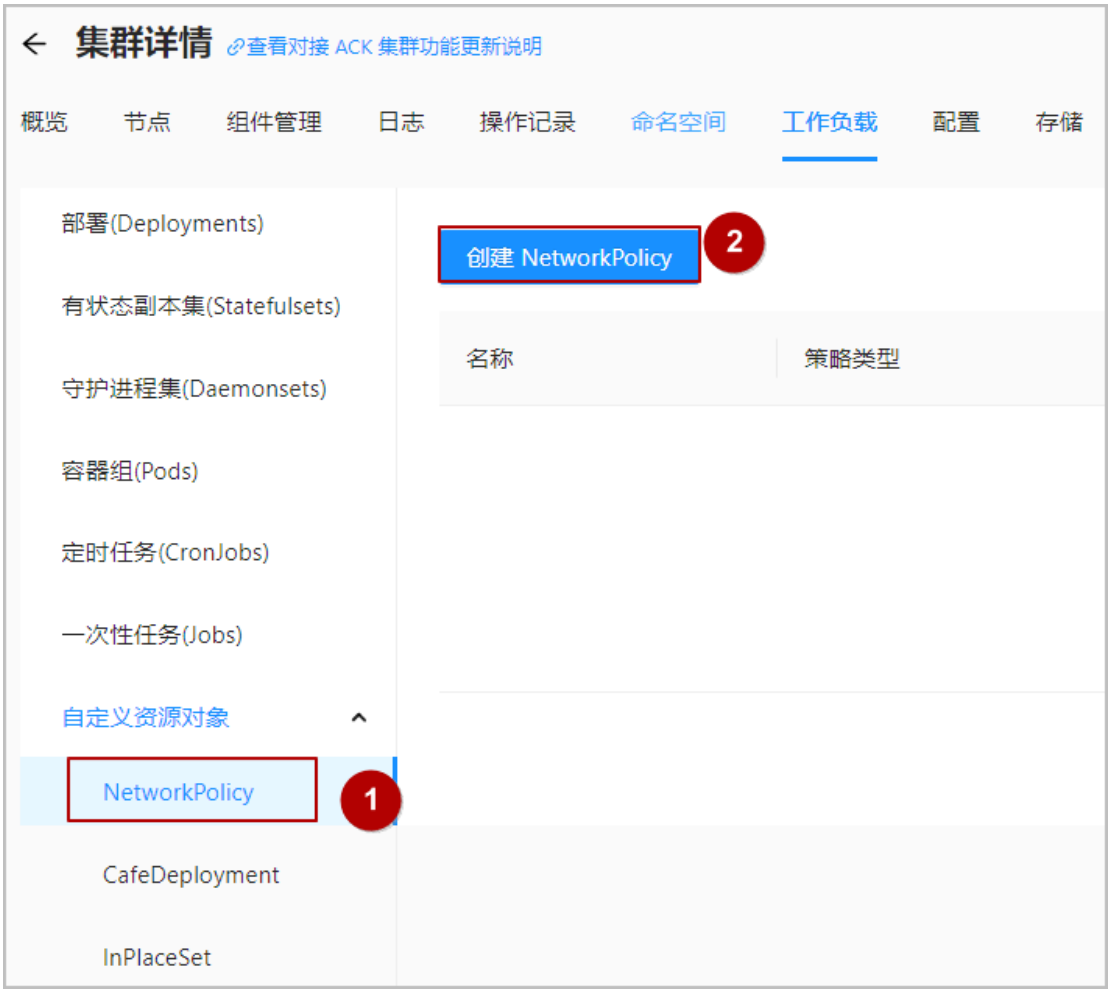
- 关于 Kubernetes NetworkPolicy 的更多信息，请参见 Kubernetes 官方文档。
- NetworkPolicy 具体操作可参考 管理 NetworkPolicy。

9.9.7.2. 管理 NetworkPolicy

您可以在 LHC 控制台上轻松管理 NetworkPolicy，如执行创建、更新、删除等操作。

创建 NetworkPolicy

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**。
2. 单击目标集群名称，进入 **集群详情** 页，单击 **工作负载** 页签。
3. 在左侧菜单选项，选择 **自定义资源对象 > NetworkPolicy**，单击 **创建 NetworkPolicy**。



4. 在输入框中输入 YAML 内容，单击 **提交**。
样例如下：

```
kind: NetworkPolicy
apiVersion: networking.k8s.io/v1
metadata:
  name: allow-all-ingress-example
spec:
  podSelector:
    matchLabels:
      run: nginx
  ingress:
  - from:
    - podSelector:
        matchLabels:
          access: "true"
```

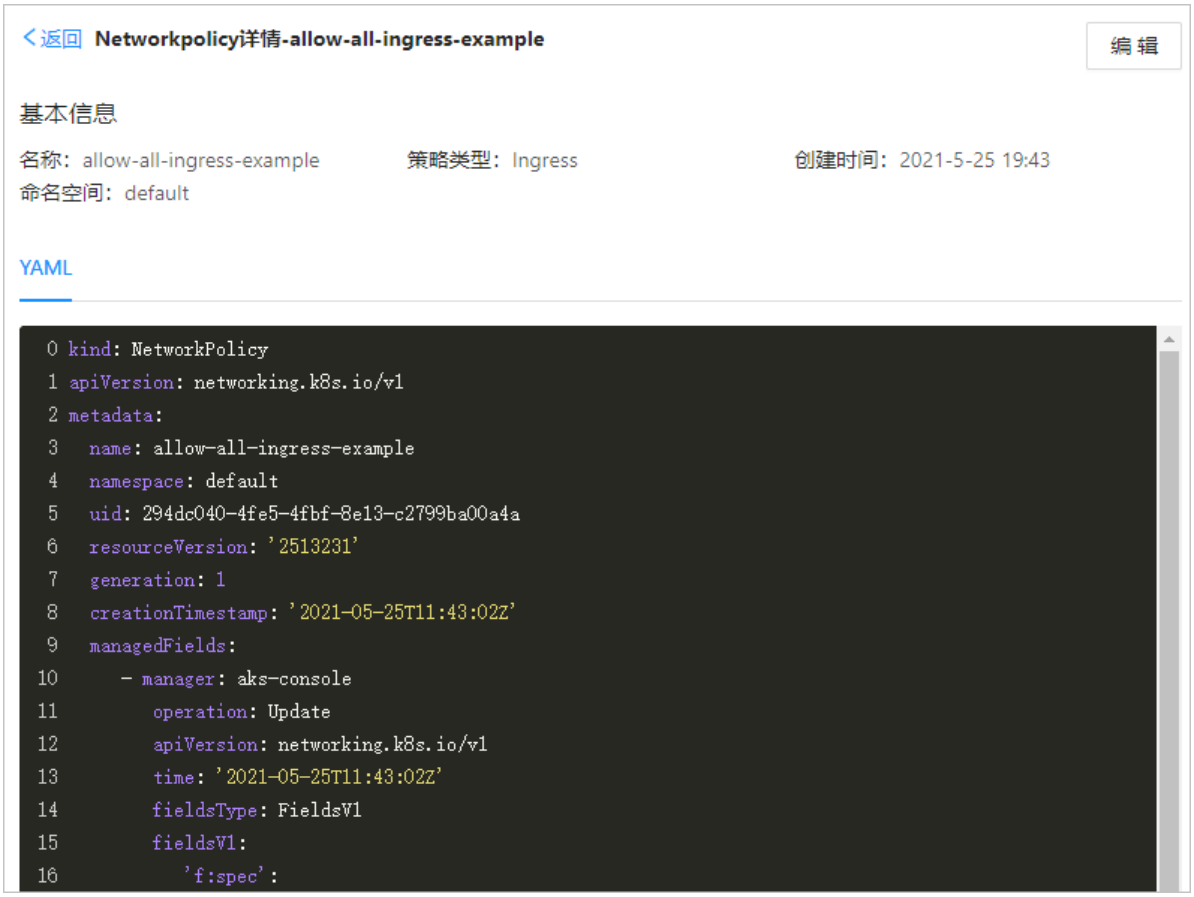
执行结果

创建成功后， 名为 **allow-all-ingress-example** 的 NetworkPolicy 将出现在列表页面中。

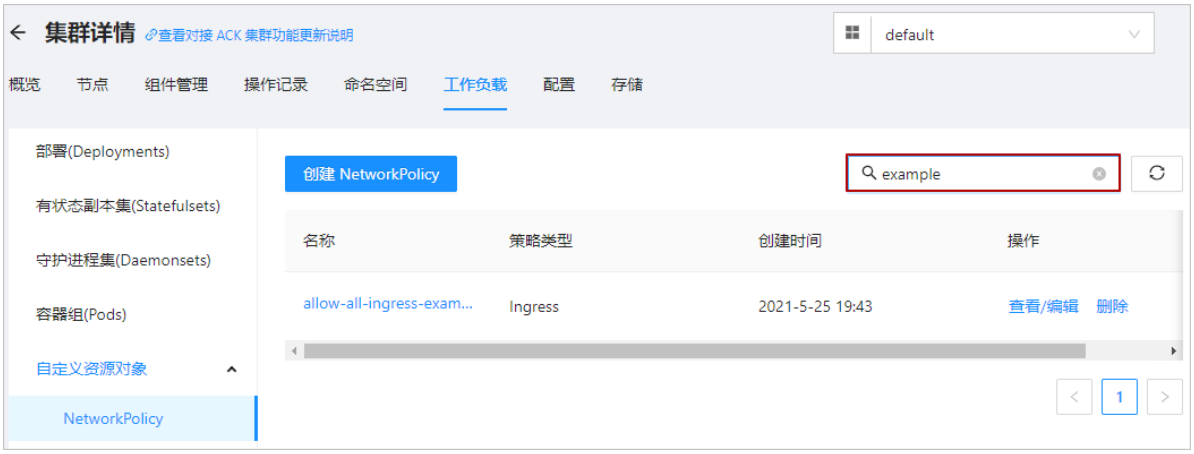


相关操作

- 如需查看 NetworkPolicy 详情，请在 NetworkPolicy 列表页单击目标 NetworkPolicy 名称，进入 NetworkPolicy 详情页。在该页面上，可以查看 NetworkPolicy 基本信息及 YAML 配置信息。



- 如需搜索 NetworkPolicy，请在 NetworkPolicy 列表页右上角的输入栏，输入 NetworkPolicy 名称关键字即可进行搜索。本例中输入 **example**，查询结果如下图所示。



- 如需更新 NetworkPolicy，请在 NetworkPolicy 列表页单击操作栏中的 **查看/编辑**，并在编辑 NetworkPolicy 对话框中编辑 YAML 信息，然后单击 **提交**。
- 如需删除 NetworkPolicy，请在 NetworkPolicy 列表页单击操作栏中的 **删除**，然后在弹出的提示对话框中单击 **确定**。

9.9.7.3. CafeDeployment

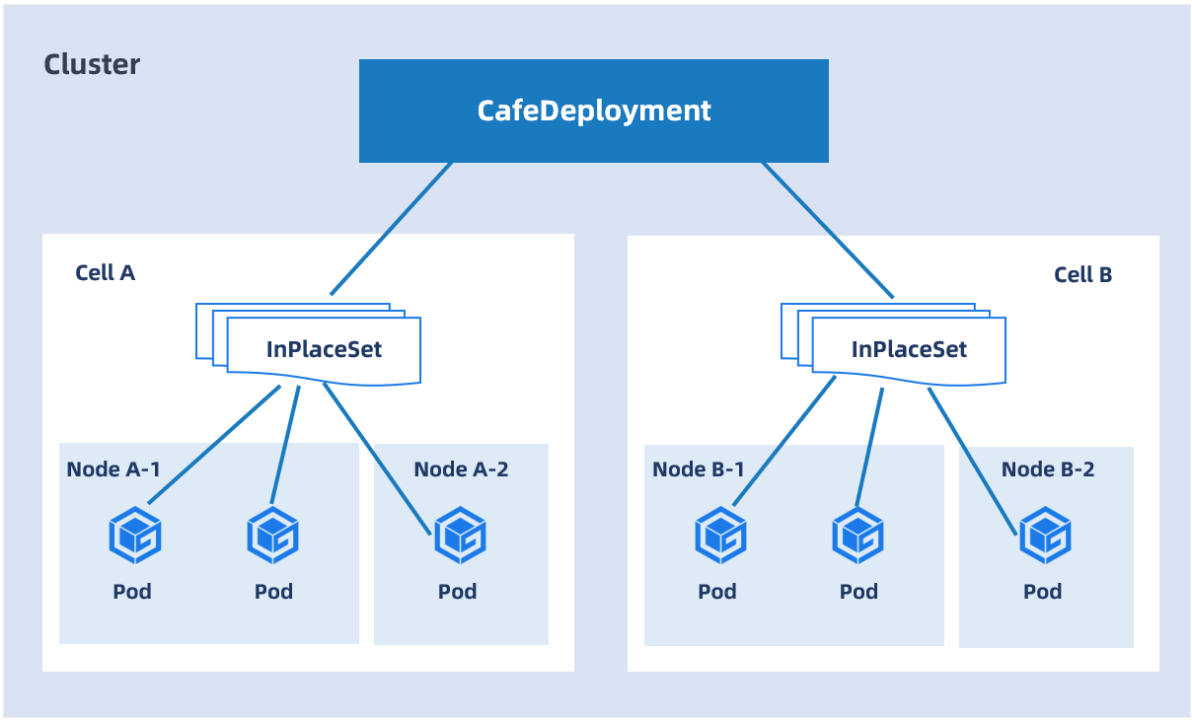
CafeDeployment 是 CAFE 产品设计的高阶工作负载（workload），为 InPlaceSet（原地升级）工作负载提供高阶发布能力，提供 Pod 单元化管理、Beta 和分组发布功能以及版本控制能力，实现 Pod 原地升级。本文介绍如何查询 CafeDeployment。

使用限制

当前该资源类型仅支持查询，无法修改、增加、删除。

背景信息

- **CafeDeployment 与应用服务**
  在 LHC 中，应用服务是一个中枢的 fedcafedeployement，它会在每一个要下发的部署单元内创建一个 Cafedeployment，而 Cafedeployement 下一层挂载 Inplaceset，再下层是实体 Pod。应用服务创建并发布后，会创建 CafeDeployment。
- **CafeDeployment 与 InPlaceSet**
  CafeDeployment 主要提供跨部署单元的管理功能，其下管理多个 InPlaceSet。每个 InPlaceSet 对应一个部署单元。部署单元是逻辑概念，它通过 Node 上的 label 来划分集群中的节点，而 InPlaceSet 则通过 NodeAffinity 能力，将其下的 Pod 部署到同一个部署单元的机器上。由此实现 CafeDeployment 跨部署单元的管理。



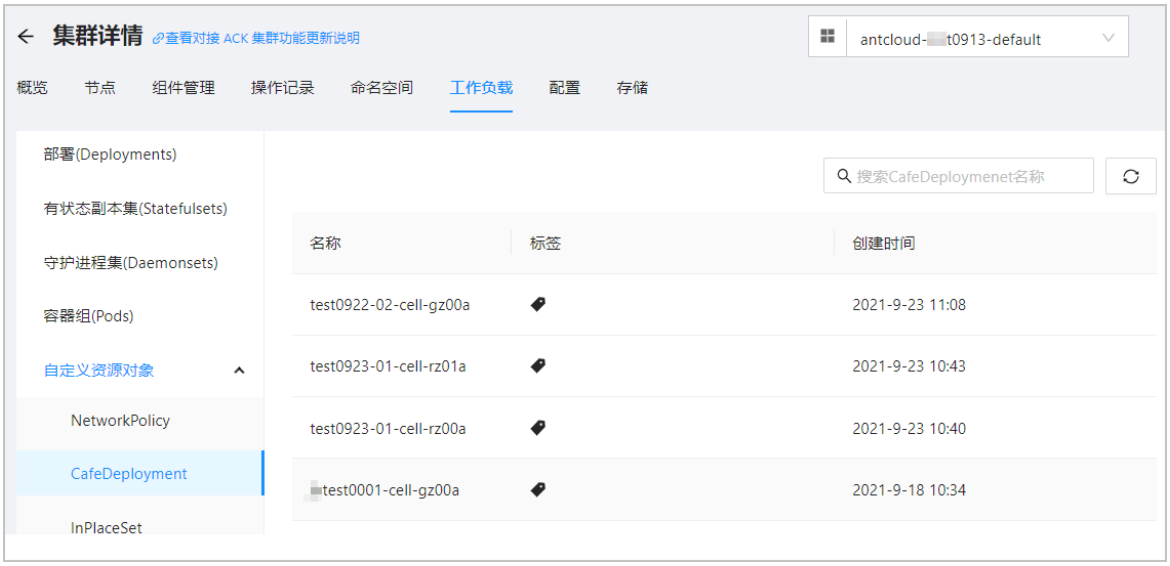
应用场景

CafeDeployment 的一些典型使用场景如下：

- 单元化：应用实例需要分割到不同的节点进行部署。
- 精细的发布管理：在 Pod 配置的更新过程中，提供了 Beta 发布、分组发布、发布暂停、发布取消等功能。
- 版本控制：在每次 Pod 配置更新后自动保存版本，方便版本回退。
- 原地升级：在保持 Pod IP 不变的前提下实现 Pod 其他配置的变更。

操作步骤

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**。
2. 单击目标集群名称，进入 **集群详情** 页，单击 **工作负载** 页签。
3. 在左侧菜单选项，选择 **自定义资源对象 > CafeDeployment**。
4. 在 **CafeDeployment** 列表页，查询 CafeDeployment 数据。
   该页面列出了支持的全部 CafeDeployment、标签以及创建时间。您还可以通过右侧搜索框，使用关键词搜索 CafeDeployment。



9.9.7.4. InPlaceSet

InPlaceSet 是 CAFE 产品设计的基础工作负载（workload），负责管理 Pod 的副本和版本，并管理 Pod 生命周期，进而提供高阶的能力。本文介绍如何查询 InPlaceSet。

背景信息

InPlaceSet 是一种新的工作负载抽象自定义资源，主要用于 Pod 原地升级，提升整体发布效率。

LHC 作为蚂蚁集团全新打造的基于 Kubernetes 的容器产品，发布部署能力的建设是其非常重要的一环，通过操作单一 Pod 对象，目前已具备基础的发布部署能力，所采用的模式也是原地升级。

> 说明
> InPlaceSet、应用服务、CafeDeployment 之间的关系可参考 CafeDeployment 与应用服务、CafeDeployment 与 InPlaceSet。

使用限制

当前该资源类型仅支持查询，无法修改、增加、删除。

操作步骤

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**。
2. 单击目标集群名称，进入 **集群详情** 页，单击 **工作负载** 页签。
3. 在左侧菜单选项，选择 **自定义资源对象 > InPlaceSet**。
4. 在 **InPlaceSet** 列表页，查询 InPlaceSet 数据。
   该页面列出了支持的全部 InPlaceSet、标签以及创建时间。您还可以通过右侧搜索框，使用关键词搜索 InPlaceSet。



9.10. 存储

9.10.1. 管理存储类型（StorageClass）

您可以在 LHC 控制台上轻松管理存储类型（StorageClass），如创建、删除等操作。
目前，LHC 支持以下两种类型的存储类型：

- NAS 类型
- 云盘（cloud_disk）类型

> **说明**
> 目前 LHC 未对接阿里云的 NAS 及云盘存储，若要使用存储功能，需要安装第三方存储插件。

NAS 类型

> **注意**
> NAS 类型的 Storage Class 必须和 LHC 集群节点在相同地域。

LHC 定义的 NAS 类型的 Storage Class 如下：

```
apiVersion: storage.k8s.io/v1

kind:StorageClass
metadata:
    name: csi-nas-capacity
provisioner: antcloud-csi
parameters:
    storage_type:"cloud_nas"# Allowed enum: cloud_disk, cloud_nas
# filesystem parameters
    nasType:"Capacity"# Allowed enum: Capacity, Performance, Extreme *
    protocolType:"NFS"# Allowed enum: NFS, SMB *
    serviceId:"1c7aff75-8b4f-4d2c-8dac-74af0b830d86-aliyun-nas-service-id"
    planId:"44dd37f1-77f9-42c5-91a5-743c007ba484-aliyun-nas-plan-id"
reclaimPolicy:Retain# Allowed enum: Delete, Retain
volumeBindingMode:Immediate# Allowed enum: Immediate, WaitForFirstConsumer
```

参数说明

参数名	参数说明	是否必选	备注
provisioner	antcloud-csi	是	LHC动态卷控制器名称，不可更改。
storage_type	后端存储类型（块存储，文件存储）	是	目前仅支持： • cloud_disk，表示阿里云盘。 • cloud_nas，表示阿里云NAS。
nasType	NAS 类型	是	可选值： • **Capacity**，详情参考 容量型 NAS。 • **Performance**，详情参考 性能型 NAS。 • **Extreme**，详情参考 极速型 NAS。
protocolType	文件系统类型	否	目前仅支持 NFS
readOnly	是否只读	否	可选值：**true**，**false**
serviceId	Service broker 必选参数	是	必须为：1c7aff75-8b4f-4d2c-8dac-74af0b830d86-aliyun-nas-service-id
planId	Service broker必选参数	否	必须为：44dd37f1-77f9-42c5-91a5-743c007ba484-aliyun-nas-plan-id
reclaimPolicy	回收策略	是	• **Retain**：保留后端存储，删除 PV/PVC 不会删除对应的后端存储如云盘。详情参考 Kubernetes 官网文档。 • **Delete**：当删除 PVC 时，自动删除 PV 和后端的存储。详情参考 Kubernetes 官网文档。

参数名	参数说明	是否必选	备注
volumeBindingMode	存储卷绑定策略	是	控制存储卷绑定和动态供应行为 • **Immediate**：创建 PVC 后立即创建后端存储卷，并且立即绑定新创建的 PV 和 PVC。 • **WaitForFirstConsumer**：当 PVC 被 Pod 使用时，才触发 PV 和后端存储的创建，同时实现 PVC/PV 的绑定，启用该配置后，Storage Class 中的 Zone 和 Region 将不再生效，而是使用 Pod 调度所在节点的 zone 和 region 创建文件系统，保证文件系统能被 Pod 挂载。

云盘（cloud_disk）类型

> **注意**
> 云盘类型的 Storage Class 必须和 LHC 集群节点在相同地域及可用区。

LHC 定义的云盘类型的 Storage Class 如下：

```
apiVersion: storage.k8s.io/v1
kind:StorageClass
metadata:
   name: csi-disk-efficiency
provisioner: antcloud-csi
parameters:
    failure-domain.beta.kubernetes.io/zone:<必须与当前集群中节点可用区匹配>
    failure-domain.beta.kubernetes.io/region:<必须与当前集群所在region匹配>
    fsType:"ext4"
    storage_type:"cloud_disk"# Allowed enum: cloud_disk, cloud_nas
    type:"cloud_ssd"
    readOnly:"false"
    serviceId:"8814fddf-5be7-4f8a-862c-f4a57c25a0b4-aliyun-disk-service-id"
    planId:"02b41c80-915f-43cd-a574-e9cd446f13ea-aliyun-disk-plan-id"
reclaimPolicy:Retain# Allowed enum: Delete, Retain
volumeBindingMode:Immediate# Allowed enum: Immediate, WaitForFirstConsumer
```

参数说明

参数名	说明	是否必填	备注
provisioner	antcloud-csi	是	动态卷控制器名称，不可更改。

参数名	说明	是否必填	备注
failure-domain.beta.kubernetes.io/region	地域	否	该字段当 `volumeBindingMod=Immediate` 时为必选，会被用来创建云盘。例如：CN-SHANGHAI，代表上海地域。
failure-domain.beta.kubernetes.io/zone	可用区	否	该字段当 `volumeBindingMod=Immediate` 时为必选，会被用来创建云盘。例如：CN-SHANGHAI-A，代表上海可用区 A。
fsType	分区格式	否	支持 ext4、ext3，默认 ext4。
storage_type	后端存储类型	是	• cloud_disk，表示阿里云云盘 • cloud_nas, 表示阿里云 NAS。
type	磁盘类型	是	可选值：cloud_ssd、cloud_efficiency、cloud。详情参考 云盘类型。
readOnly	是否只读	否	可选值：true、false。
serviceId	Service broker 必选参数	是	必须为：`8814fddf-5be7-4f8a-862c-f4a57c25a0b4-aliyun-disk-service-id`
planId	Service broker 必选参数	否	必须为：`02b41c80-915f-43cd-a574-e9cd446f13ea-aliyun-disk-plan-id`
reclaimPolicy	回收策略	是	• **Retain**：保留后端存储，删除 PV/PVC 不会删除对应的后端存储如云盘。详情参考 Kubernetes 官网文档。 • **Delete**：当删除 PVC 时，自动删除 PV 和后端的存储。详情参考 Kubernetes 官网文档。

参数名	说明	是否必填	备注
volumeBindingMode	存储卷绑定策略	是	控制存储卷绑定和动态供应行为。 • **Immediate**：创建 PVC 后立即创建后端存储卷，并且立即绑定新创建的 PV 和 PVC • **WaitForFirstConsumer**：当 PVC 被 pod 使用时，才触发 PV 和后端存储的创建，同时实现 PVC 和 PV 的绑定，启用该配置后，storage class 中的 zone 和 region 将不再被使用，而是使用 pod 调度所在节点的 zone 和 region 创建云盘，保证云盘能被 Pod 挂载。

创建存储类型

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击某个集群名称，进入 **集群详情** 页，单击 **存储** 页签。
3. 在左侧菜单选项，选择 **存储类型 (StorageClass)**，然后单击 **创建存储类型**。
4. 在输入框中输入 YAML 内容，单击 **提交**。
   - NAS 类型样例：见 NAS 类型。

更新存储类型

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击某个集群名称，进入 **集群详情** 页，单击 **存储** 页签。
3. 在左侧菜单选项，选择 **存储类型 (StorageClass)**，然后单击 **操作** 栏的 **查看/编辑**。
4. 在 YAML 编辑框中修改信息，单击 **提交**。
   更新成功后，系统会自动根据新的 YAML 文件信息更新存储类型。

删除存储类型

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击某个集群名称，进入 **集群详情** 页，单击 **存储** 页签。
3. 在左侧菜单选项，选择 **存储类型 (StorageClass)**。
4. 单击 **操作** 栏的更多图标 > **删除**。
5. 在弹出的对话框中单击 **确定**。

9.10.2. 管理存储卷（PersistentVolume）

您可以在 LHC 控制台上轻松管理存储卷（PersistentVolume），如创建、删除等操作。

创建存储卷

前提条件

已创建存储类型（StorageClass）。

操作步骤

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击某个集群名称，进入 **集群详情** 页，单击 **存储** 页签。
3. 在左侧菜单选项，选择 **存储卷（PersistentVolume）**，然后单击 **存储卷**。
4. 在输入框中输入 YAML 内容，单击 **提交**。
   样例如下：

```
apiVersion: v1
kind: PersistentVolume
metadata:
  name: pv-nas
spec:
  capacity:
    storage: 5Gi
  storageClassName: nas
  accessModes:
    - ReadWriteMany
```

更新存储卷

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击某个集群名称，进入 **集群详情** 页，单击 **存储** 页签。
3. 在左侧菜单选项，选择 **存储卷（PersistentVolume）**，然后单击 **操作** 栏的 **查看/编辑**。
4. 在 YAML 编辑框中修改信息，单击 **提交**。
   更新成功后，系统会自动根据新的 YAML 文件信息更新存储卷。

删除存储卷

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击某个集群名称，进入 **集群详情** 页，单击 **存储** 页签。
3. 在左侧菜单选项，选择 **存储卷（PersistentVolume）**。
4. 单击 **操作** 栏的更多图标 > **删除**。
5. 在弹出的对话框中单击 **确定**。

9.10.3. 管理存储卷声明（PVC）

您可以在 LHC 控制台上轻松管理存储卷声明（PersistentVolumeClaim），如创建、删除等操作。

创建存储卷声明

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击某个集群名称，进入 **集群详情** 页，单击 **存储** 页签。
3. 在左侧菜单选项，选择 **存储卷声明（PersistentVolumeClaim）**，然后单击 **创建存储卷声明**。
4. 在输入框中输入 YAML 内容，单击 **提交**。
   样例如下：

```
apiVersion: v1
kind: PersistentVolumeClaim
metadata:
  name: test-pvc
  namespace: default
spec:
  accessModes:
  - ReadWriteOnce
  resources:
    requests:
      storage: 25Gi
  storageClassName: csi-disk-efficiency # 对应 storage class 的名称
```

> 注意
> 对于磁盘类型设置为 cloud_efficiency 和 cloud_ssd 类型的 Storage Class，其对应的最小存储空间是 20 Gi，当 PVC 中指定的大小小于 20 Gi 时，PV 无法创建成功，PVC 会一直处于 pending 状态。

PVC 创建完成后，您可以在 YAML 文件的 status 中查看创建情况。PVC 创建成功，对应的存储卷会展示在**存储卷** 页面。

更新存储卷声明

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击某个集群名称，进入 **集群详情** 页，单击 **存储** 页签。
3. 在左侧菜单选项，选择 **存储卷声明**，然后单击 **操作** 栏的 **查看/编辑**。
4. 在 YAML 编辑框中修改信息，单击 **提交**。
   更新成功后，系统会自动根据新的 YAML 文件信息更新存储卷声明。

删除存储卷声明

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击某个集群名称，进入 **集群详情** 页，单击 **存储** 页签。
3. 在左侧菜单选项，选择 **存储卷声明**。
4. 单击 **操作** 栏的更多图标 > **删除**。
5. 在弹出的对话框中单击 **确定**。

9.10.4. 使用存储卷声明（PVC）

PVC 创建成功后，您可以在创建 Pod 时使用 PVC。

> 注意
> - 云盘不能跨可用区（zone）挂载，因此在创建 `volumeBindingMode：Immediate` 类型的 PVC 后，使用该 PVC 的 pod/workload 需要通过 nodeSelector/nodeAffinity 限制 pod 运行在 PV/PVC 同一个可用区。
> - 目前 LHC 未对接阿里云的 NAS 及云盘存储，若要使用存储功能，需要安装第三方存储插件。

前提条件

已创建 PVC。

操作步骤

1. 登录 LHC 控制台，在左侧导航栏单击 **集群管理 > 集群详情**，进入集群列表页。
2. 单击某个集群名称，进入 **集群详情** 页，单击 **工作负载** 页签。
3. 在左侧菜单选项，选择 **容器组（Pods）** ，单击 **从 YAML 创建**。
4. 在输入框中输入 YAML 内容，单击 **提交**。
   样例如下：

```
 apiVersion: v1
 kind:Pod
 metadata:
   name: pod-with-disk
namespace:default
   labels:
     app: nginx
 spec:
   tolerations:
- key: node.kubernetes.io/unreachable
     effect:NoSchedule
operator:Exists
   automountServiceAccountToken:false
   containers:
- name: busybox
     image: reg-cnsh-nf.cloud.alipay.com/library/busybox
     command:["sleep","3000"]
     ports:
- containerPort:80
     volumeMounts:
- name: data
         mountPath:"/data"
   volumes:
- name: data
       persistentVolumeClaim:
         claimName: test-pvc
```

容器运行后，可以看到 pod 的状态为 running：

```
# kubectl get pod
NAME            READY    STATUS     RESTARTS    AGE
pod-with-disk   1/1Running0103s
```

9.10.5. 使用 PVC 挂载动态存储卷

挂载动态存储卷是基于 Storage Class 实现的，在挂载动态存储卷前，需要先创建存储类型（StorageClass）。

> 说明
> 目前 LHC 未对接阿里云的 NAS 及云盘存储，若要使用存储功能，需要安装第三方存储插件。

前提条件

已创建存储类型。

步骤一：创建 PVC

您需要在创建 Pod 前创建存储声明，供创建 Pod 时使用。PVC 创建后，会有对应的 PV 自动创建出来。

云盘类型 PVC 样例

```
apiVersion: v1
kind:PersistentVolumeClaim
metadata:
  name: test-pvc
spec:
  accessModes:
-ReadWriteOnce
  resources:
    requests:
      storage:25Gi
  storageClassName: csi-disk-efficiency # csi-disk-ssd
```

参数说明

参数名	参数说明	是否必选	备注
accessModes	只读或读写	否	仅支持：ReadWriteOnce
storage	存储空间大小	是	遵循后端存储限制，参考 阿里云云盘限制。对于 csi-disk-ssd 和 csi-disk-efficiency 两种 Storage Class，最小的存储空间都是 20 GiB，否则 PV 无法创建成功，PVC 会一直处于 pending 状态。
storageClassName	Storage Class 名称	是	必须指定一个已经存在的名字，否则云盘无法创建。

NAS 类型 PVC 样例

```
apiVersion: v1
kind:PersistentVolumeClaim
metadata:
  name: test-pvc
spec:
  accessModes:
-ReadWriteMany
  resources:
    requests:
      storage:25Gi
  storageClassName: csi-nas-capacity # 或者 csi-nas-performance
```

> 注意
> 对于 NAS 存储，其对应的存储空间限制能力将不会生效，不会限制容器在运行时数据的写入量，仅用做 PV-PVC 在绑定过程的匹配。

参数说明

参数名	参数说明	是否必选	备注

参数名	参数说明	是否必选	备注
accessModes	只读或读写	否	可选值：- ReadWriteOnce- ReadWriteMany
storage	存储空间大小	是	遵循后端存储限制，如阿里云 NAS 限制。文件系统容量上限都为 1 PB（性能型）或者 10 PB（容量型）。
storageClassName	Storage Class 名称	是	必须指定一个已经存在的名字，否则云盘无法创建。

步骤二：使用 PVC

您可以在创建 Pod 时使用 PVC 挂载动态数据卷。

```
apiVersion: v1
kind: Pod
metadata:
  name: pod-with-nas
  labels:
    app: nginx
spec:
  tolerations:
  - key: node.kubernetes.io/unreachable
    effect: NoSchedule
    operator: Exists
    automountServiceAccountToken: false
  containers:
  - name: busybox
    image: 'reg-cnsh-nf.cloud.alipay.com/library/busybox'
    command: ["sleep","3000"]
    ports:
    - containerPort: 80
    volumeMounts:
    - name: data
      mountPath: "/data"
    volumes:
    - name: data
      persistentVolumeClaim:
        claimName: test-pvc
```

容器运行后，可以看到 Pod 的状态为 Running：

```
# kubectl get pod
NAME            READY   STATUS    RESTARTS   AGE
pod-with-nas   1/1Running0103s
```

9.10.6. 使用 PVC 挂载静态存储卷

您可以通过绑定已存在的 PV 为 Pod 挂载静态存储卷。

> 说明
> 目前 LHC 未对接阿里云的 NAS 及云盘存储，若要使用存储功能，需要安装第三方存储插件。

前提条件

- 已创建存储类型。
- 同租户下存在可用云盘。

步骤一：创建静态 PV

创建 PV 并绑定一个已经存在的云盘。您可以使用 YAML 文件或者控制台界面创建 PV。
YAML 样例如下：

```
apiVersion: v1
kind:PersistentVolume
metadata:
  labels:
    cafe.sofastack.io/diskId: d-zm0fd73z6tmf5j7pgbpe
  name: pv-efficient
spec:
  accessModes:
-ReadWriteOnce
  capacity:
    storage:25Gi
  csi:
    driver: antcloud-csi
    fsType: ext4
    volumeAttributes:
      storage_type: cloud_disk
    volumeHandle: instance_test_efficiency
  storageClassName: csi-disk-efficiency
```

参数说明

参数名	参数说明	是否必选	备注
cafe.sofastack.io/diskId	阿里云的云盘 ID，格式为 `d-xxx` 。	是	您可以在资源管理控制台上的 **云服务器 ECS** > **存储与快照** > **云盘** 中获取云盘 ID。
volumeAttributes	磁盘属性	是	遵循后端存储限制，如 阿里云盘限制。对于 csi-disk-ssd 和 csi-disk-efficiency 两种 Storage Class，最小的 size 都是 20G。
volumeHandle	表示后端存储的唯一 ID，由用户指定。	是	-

步骤二：创建 PVC 并指定其绑定到 PV

您可以使用 YAML 文件或者控制台界面创建 PVC。
YAML 样例如下：

```
apiVersion: v1
kind:PersistentVolumeClaim
metadata:
 name: test-pvc
namespace:default
spec:
 accessModes:
-ReadWriteOnce
 resources:
   requests:
     storage:25Gi
 storageClassName: csi-disk-efficiency
 volumeName: pv-efficient
```

> **说明**
> 必须要有 volumeName 字段，指定已有的 PV。参数说明参考 使用 PVC 挂载动态存储卷。

步骤三：使用 PVC

您可以在创建 Pod 时使用 PVC 挂载静态数据卷。

```
apiVersion: v1
kind:Pod
metadata:
  name: pod-with-disk
namespace:default
  labels:
    app: nginx
spec:
  tolerations:
- key: node.kubernetes.io/unreachable
    effect:NoSchedule
operator:Exists
  automountServiceAccountToken:false
  containers:
- name: busybox
    image: reg-cnsh-nf.cloud.alipay.com/library/busybox
    command:["sleep","3000"]
    ports:
- containerPort:80
    volumeMounts:
- name: data
       mountPath:"/data"
  volumes:
- name: data
     persistentVolumeClaim:
       claimName: test-pvc
```

容器运行后，可以看到 Pod 的状态为 running：

```
# kubectl get pod
NAME            READY   STATUS    RESTARTS   AGE
pod-with-disk   1/1Running0103s
```

9.10.7. StatefulSet 挂载动态存储卷

> 说明
> 目前 LHC 未对接阿里云的 NAS 及云盘存储，若要使用存储功能，需要安装第三方存储插件。

前提条件

- 已创建存储类型。
- 已创建应用服务。

使用 NAS 创建多实例 StatefulSet

YAML 样例如下：

```
apiVersion: apps/v1beta2
kind:StatefulSet
metadata:
  name: web
spec:
  selector:
    matchLabels:
      app: nginx
  serviceName:"<指定已有的Service>"
  replicas:2
template:
    metadata:
      labels:
        app: nginx
    spec:
      containers:
- name: nginx
        image: nginx
        ports:
- containerPort:80
          name: web
        volumeMounts:
- name: csi-nas-pvc
          mountPath:/data
  volumeClaimTemplates:
- metadata:
      name: csi-nas-pvc
    spec:
      accessModes:["ReadWriteOnce"]
      storageClassName:"csi-nas-capacity"
      resources:
        requests:
          storage:25G
```

使用云盘创建多实例 StatefulSet

YAML 样例如下：

```
apiVersion: apps/v1beta2
kind:StatefulSet
metadata:
  name: web
spec:
  selector:
    matchLabels:
      app: nginx
  serviceName:"<指定已有的Service>"
  replicas:2
template:
    metadata:
      labels:
        app: nginx
    spec:
      containers:
- name: nginx
        image: nginx
        ports:
- containerPort:80
          name: web
        volumeMounts:
- name: disk-efficiency
          mountPath:/data
  volumeClaimTemplates:
- metadata:
      name: disk-efficiency
    spec:
      accessModes:["ReadWriteOnce"]
      storageClassName:"csi-disk-efficiency"
      resources:
        requests:
          storage:25G
```

参数说明参考 使用 PVC 挂载动态存储卷。

9.11. 镜像中心

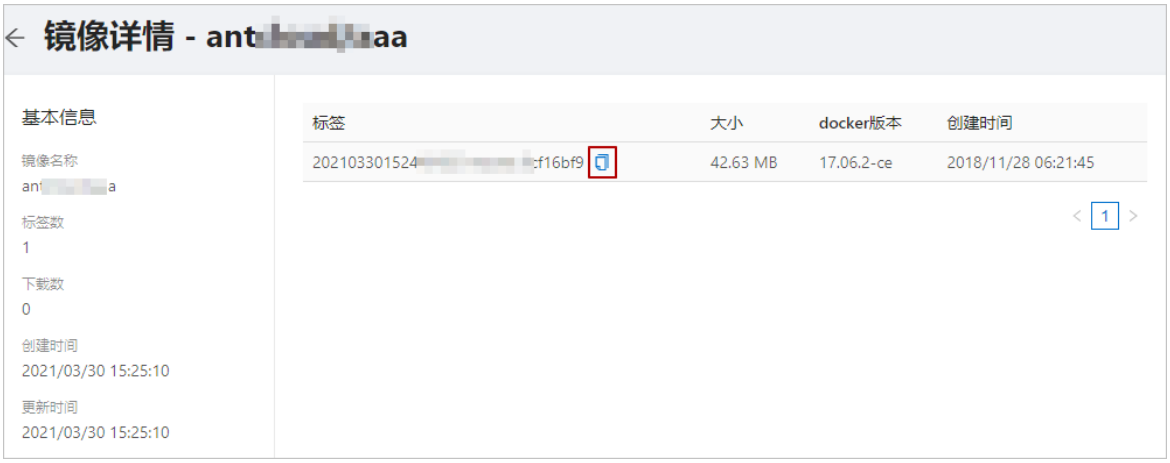
9.11.1. 镜像仓库

镜像仓库是镜像的集合，用于存储、管理容器镜像。您可以在控制台上轻松存储、管理、部署容器镜像。本文主要介绍如何下载平台镜像、上传私有镜像。

下载平台镜像

平台镜像中提供常用的样例镜像。

1. 登录控制台，在左侧导航栏单击 **镜像中心** > **镜像仓库**。
2. 选择 **平台镜像** 页签。平台提供的镜像以卡片形式展示。
3. 单击卡片进入镜像信息页面，单击标签右侧的复制图标，复制镜像地址，用 `docker pull` 命令下载镜像。



上传私有镜像

前提条件

- 准备一台装有 Docker 的机器。

> **说明**
> Docker 版本必须为 1.6.0 及以上。

- 租户初次使用服务账号登录，需先在租户内新建一个用户。

操作步骤

1. 登录装有 Docker 的机器。
2. 获取用户的 accessKey 和 accessSecret，分别用作用户名和密码登录镜像中心。
   格式：`docker login` **镜像中心域名**，示例：`docker login hub.alipay.net`。

   > **说明**
   > accessKey 和 accessSecret，可以在 **控制台 > 账户信息 > AccessKey 信息** 中获取。

3. 将镜像命名为 **镜像中心域名/租户名/应用名称:标签** 的格式。
   示例：`hub.alipay.net/tenantname/appname:imagetag`。
4. 执行以下命令上传镜像：
   ```
   docker push hub.alipay.net/tenantname/appname:imagetag
   ```
   上传后的镜像会存储在 **镜像中心> 镜像仓库 > 私有镜像** 中。单击卡片进入镜像信息页面，单击标签右侧的复制图标复制镜像地址，您可以用 `docker pull` 命令下载镜像。

9.11.2. 构建镜像

镜像构建提供将应用的源代码编译成镜像的服务能力。镜像成功构建后，默认会推送至 LHC 默认的镜像中心。

创建构建

本部分介绍如何通过单元化应用服务控制台完成镜像构建操作。

前提条件

已创建应用。

操作步骤

1. 登录控制台，在左侧导航栏单击 **镜像中心 > 镜像构建**。
2. 在构建列表页，单击 **创建**。
3. 在 **创建镜像构建** 页面，填写以下配置信息，在 **确认配置** 页面单击 **提交** 完成应用构建。
   - **基础配置**
     - **所属应用**：选择构建要关联的应用。
     - **构建集名称**：选填。
     - **仓库命名空间**：选择在 ACR 上创建的命名空间。
     - **镜像名**：镜像名称默认与应用名称相同，无法修改。
     - **代码库类型**：必须配置代码库才可以进行构建。目前支持 **外部 GitLab**：可配置支持公网访问并且版本在 8.8 及以上的 GitLab （同时支持 GitLab v3 和 v4 版本 API）。 除 **外部 GitLab 地址** 外，还需配置 外部 GitLab Token。
   - **高级配置（选填）**
     - **源镜像中心访问配置**：Dockerfile 中 FROM 镜像所在的镜像中心访问配置，可配置多个。
     - **上传到自定义镜像中心**：该项用于配置构建成功后额外的 Push 镜像的镜像中心访问信息，如不做配置，默认只会 Push 至 SOFAStack 平台默认的镜像中心。有以下配置项：
       - **自定义镜像名**：输入一个包含 `registryurl/namespace/repository` 的字符串，比如：`registry.cn-hangzhou.antcloud.com/mytenant/sayhello` 。
       - **用户名**：选填。当所配置的镜像中心需要用户名和密码进行登录时进行配置。
       - **密码**：选填。当所配置的镜像中心需要用户名和密码进行登录时进行配置。
     - **构建路径**：选填。执行 `docker build` 所在的工作目录，相对源代码 checkout 后的路径。默认为源代码根目录。
     - **Dockerfile 路径**：选填。执行 `docker build` 时的 `--file` 参数。默认为源代码根目录下 Dockerfile 文件。

触发构建

您可以在完成本地代码更新后触发构建镜像。

操作步骤

1. 登录控制台，在左侧导航栏单击 **镜像中心 > 镜像构建**，进入构建列表页面。
2. 单击 **构建集名称** 进入构建详情页。
3. 在 **构建历史** 页签中，单击 **触发构建**。
4. 在弹出的 **构建** 对话框中，填写以下参数，单击 **确定**。
   - **分支**：代码的分支，例如：`master` 、 `dev` 。
   - **构建参数**：作为执行 `docker build` 的 `--build-arg` 参数。

删除构建

您可以在控制台上删除不需要的构建。

操作步骤

1. 登录控制台，在左侧导航栏单击 **镜像中心 > 镜像构建**，进入构建列表页。
2. 单击构建 **操作** 栏中的 **删除** ，在弹出的 **删除构建** 对话框中单击 **确定**。

您可以勾选 **同时删除镜像**，将构建关联的镜像一起删除。

查看构建历史

完成镜像构建后，就可以进入该构建详情页面查看构建详情。

操作步骤

1. 登录控制台，在左侧导航栏单击 **镜像中心 > 镜像构建**，进入构建列表页。
2. 单击 **构建集名称** 进入构建集详情页。
3. 在 **构建历史** 页签中，查看构建历史信息。您可以单击 **镜像版本** 右侧的复制按钮复制镜像地址，或者单击镜像构建记录序号查看详细构建日志，可参考 查看触发日志。



查看触发日志

您可以在控制台查看完整的构建触发日志。

操作步骤

1. 登录控制台，在左侧导航栏单击 **镜像中心 > 镜像构建**，进入构建列表页。
2. 单击 **构建集名称** 进入构建详情页。
3. 在 **构建历史** 页签中，单击镜像构建记录序号，查看详细的触发日志信息。



9.11.3. 镜像技术栈

镜像技术栈为用户提供基础镜像管理的能力，可自定义镜像技术栈。

创建镜像技术栈

创建镜像技术栈将会定义镜像技术栈的基本元数据。

> 说明
> 目前对可以创建的镜像技术栈个数不作限制。

1. 登录控制台，在左侧导航栏单击 **镜像中心 > 镜像技术栈**，进入镜像技术栈列表页面。
2. 单击 **创建镜像技术栈**，在弹出的窗口中完成以下配置。

配置项	说明
技术栈名称	必填。允许包含字母、数字、连字符，且必须以字母开头、以字母或数字结尾。最多允许输入 32 个字符。 创建后技术栈的名称不可修改。
技术栈描述	必填。填写技术栈的描述。

3. 单击 **提交**。

创建镜像版本

创建镜像技术栈后，您可以根据需要创建不同的镜像版本。

> 说明
> 目前对可以创建的镜像版本个数不作限制。

1. 在镜像技术栈列表页，单击需要创建镜像版本的技术栈名称，进入技术栈详情页。
2. 单击 **创建镜像版本**，完成以下配置。

配置项	说明
镜像版本名称	允许包含字母、数字、连字符、且必须以字母开头、以字母或数字结尾。最多允许输入 32 个字符。创建后版本的名称不可修改。
基础镜像信息	选择要使用的镜像仓库和基础镜像版本。 > 说明 如果基础镜像来自外部镜像中心，需提供用户名和密码。
代码包上传路径地址	代码包在构建镜像中的目录路径。需以 `/` 开头。系统会将发布包拷贝到基础镜像中用户指定的目录。
备注	输入基础镜像的 Git 地址。

3. 单击 **提交**。
   创建后系统会提示创建成功，并以列表形式在技术栈详情页展示。在该页面，您可以对自定义的技术栈版本继续进行浏览、下架的操作。

> 说明
> - 镜像版本号由系统自动生成，目前不支持修改。
> - 镜像版本下架后不可恢复。下架的版本无法在创建应用服务时选择使用。

9.12. 应用商店

9.12.1. 应用商店概述

LHC 应用商店功能集成了 Helm 能力，提供 Helm 相关功能，并进行了相应扩展，例如提供图形化界面、官方仓库、版本回滚等。

背景信息

Helm 是管理 Kubernetes 应用程序的打包工具，主要用来管理 Charts。Helm Chart 是用来封装 Kubernetes 原生应用的一系列 YAML 文件。可在您部署应用时自定义应用的一些 Metadata，以便于应用的分发。
对于应用发布者而言，可通过 Helm 打包应用、管理应用依赖关系、管理应用版本等。对于使用者而言，使用 Helm 后无需编写复杂的应用部署文件，即可轻松查找、安装、升级、回滚及卸载应用。
LHC 应用商店基于 Helm 能力作了增强，通过产品模版简化部署复杂应用的过程，并根据行业特性构建行业交付标准、交付流程和交付路径的基础，旨在打造行业通用的应用商店，可根据不同需求应对不同的业务场景。

功能说明

LHC 应用商店支持如下功能：

- **产品模板**
  - 私有平台：支持上传用户自定义的应用模板、应用模板包、新版本。
  - 公有平台：支持引入并管理第三方应用模板仓库，以及将第三方模板复制到私有平台。
  - 版本控制：针对模板变更自动做版本号的增加，而不是单纯覆盖。
  - 下载模板：支持下载私有平台中的应用模板包。

> 说明
> 您可以对私有平台中的产品模板进行增加、删除、编辑内容的操作，公有平台的产品模板仅支持增删操作。若要对公有平台的产品模板进行定制，可以先将第三方模板复制至私有平台。

- **产品发布**
  - 快速部署：选择私有或公有平台的应用模板，自定义参数并快速地一键部署常用的应用到 Kubernetes 中。
  - 自定义参数：模板安装时用户自定义的参数，会在安装详情页里作持久化保留，而不是覆盖原有商店中的配置。
  - 版本回滚：当前模板有其它成功发布的历史版本（除当前版本外）时，可以进行一键回滚。

相关文档

- 关于 Helm 的更多信息，可参考 Helm 官网文档。
- 关于产品模板的更多信息，可参考 产品模板。
- 关于产品发布的更多信息，可参考 产品安装。

9.12.2. 产品模板

产品模板为基于 Kubernetes Helm 标准的应用模板提供了统一高效的管理机制。本文介绍如何在私有平台上传私有模板、如何在公有平台添加外部源，以及如何将公有平台模板导入私有平台等。

私有平台

私有平台相当于用户的私有仓库，存储以下两种类型的模板。

- **平台模板**
  由 SOFAStack 平台提供和维护的官方示例模板，当前支持 redis、etcd、mysql、mongodb、elasticsearch-exporter、kibana 等示例模板，让您可以即开即用。

> 说明
> 此类模板仅支持查看和安装操作，用户无法修改模板信息。

- **私有模板**
  满足用户管理和维护自定义模板的需求，私有平台的模板的来源和用途局限于当前集群。您可以对模板内容进行增删改查。

上传私有模板

> 说明
> 目前没有限制模板的上传个数。

前提条件

已准备好符合 Helm 规范的产品模板。

操作步骤

1. 登录 LHC 控制台，在左侧导航栏单击 **应用商店 > 产品模板**，进入私有平台页面。
2. 单击 **上传产品模板**，在弹出框中选择准备好的自定义产品模板。

> 说明
> 仅允许上传 tgz/zip 格式的文件，推荐使用 `helm package` 打包。

上传成功后，模板会以卡片的形式展示在列表页。单击卡片进入 **模板详情** 页，您可以继续进行如下操作：

- 在 **模板版本** 页签中查看、下载、删除历史版本。
- 在 **内容详情** 页签中查看、修改、编辑模板内容，修改后单击 **保存**。

公有平台

公有平台作为全局仓库，可以引入第三方模板源。公有平台的模板不支持直接修改内容。如需修改，可安装和复制模板到私有平台。

添加外部源

您可以通过 LHC 控制台将第三方库加入到源列表，此操作等效于 `helm repo add [NAME] [URL]` 命令。

> 说明
> 最多可添加 5 个外部源。

1. 登录 LHC 控制台，在左侧导航栏单击 **应用商店 > 产品模板**，单击 **公有平台** 页签进入公有平台页面。
2. 单击 **添加外部源**，在弹出框中完成以下配置后，单击 **确定**。



- **外部源名称**：Helm Repo 的标记名称，可包含英文字母、数字、下划线或中划线。本例中输入 `stable`。
- **外部源 URL 地址**：Helm Repo 的路径地址，必须以 `http` 或 `https` 开头，且需符合 URL 规范。本例中输入 `https://apphub.aliyuncs.com/stable`。

添加成功后，外部仓库中的模板会自动导入。

导入私有平台

若对公有平台模板有定制化需求，您可以先将公有模板导入私有平台。

1. 登录 LHC 控制台，在左侧导航栏单击 **应用商店 > 产品模板**，单击 **公有平台** 页签进入公有平台页面。
2. 单击目标模板卡片上的 **导入私有平台**，在弹出框中单击 **确定**。



导入成功后，您可以在私有平台搜索到该模板，并对模板内容进行编辑和版本管理。

9.12.3. 产品安装

通过安装产品模板，您可以将标准的 Helm 应用模板发布为相应的 Kubernetes workloads。

前提条件

已 上传产品模板 或 添加外部源。

安装私有模板

1. 登录 LHC 控制台。
2. 在左侧导航栏上，单击 **应用商店 > 产品模板。**
3. 单击 **私有平台 > 私有模板**，进入模板列表页面。
4. 单击模板卡片进入模板详情页，单击右上角的 **产品安装**。
5. 在弹出框中填入以下配置。
   - **发布名称**：输入自定义的发布产品名称。支持英文小写字母、数字、中划线，必须以英文小写字母开头，最大长度为 32 个字符。
   - **集群**：选择对应集群。
   - **命名空间**：选择相应的命名空间。
   - **版本**：选择模板的版本。
   - **描述**：选填。最大长度 100 个字符。
6. 单击 **提交**，进入 **产品安装详情** 页。
7. 单击 **开始/重新安装**，在弹出框中单击 **确定**。
   安装成功后，可以单击发布详情中的工作负载或者应用服务链接进入对应的详情页面。

> 说明
> - 如果安装失败，可以单击 **发布错误日志** 查看错误日志排查问题。
> - 如果要中止安装，可以单击 **取消安装**，在弹出框中单击 **确定**。

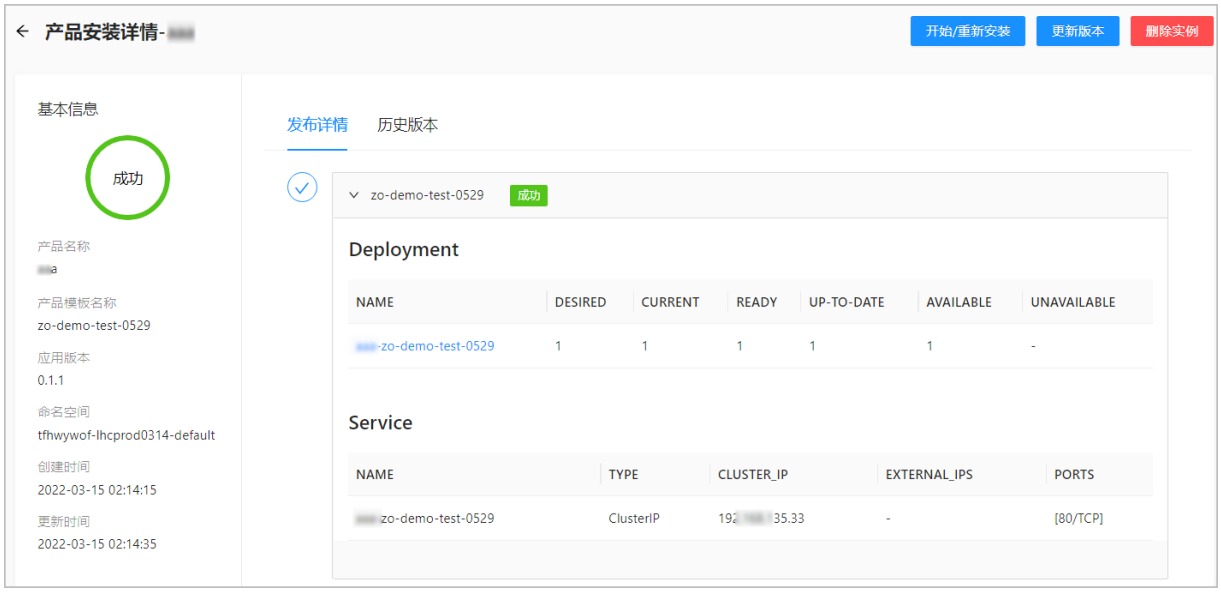
发布公有模板

1. 登录 LHC 控制台。
2. 在左侧导航栏上，单击 **应用商店** > **产品模板**，进入私有模板列表页面。
3. 选择 **公有平台** 页签，进入公有模板列表页面。
4. 单击公有模板卡片进入模板内容详情。
5. 单击右上角的 **产品安装**，在弹出框中完成以下配置。
   - **发布名称**：输入自定义的发布产品名称。支持英文小写字母、数字、中划线，必须以英文小写字母开头，最大长度为 32 个字符。
   - **命名空间**：选择相应的命名空间。
   - **版本**：选择模板的版本。
   - **描述**：选填。最大长度 100 个字符。
6. 单击 **提交**，进入 **产品安装详情** 页。
7. 单击 **开始/重新安装**，在弹出框中单击 **确定**。
   安装成功后，可以单击发布详情中的工作负载或者应用服务链接进入对应的详情页面。

> 说明
> - 如果安装失败，可以单击 **发布错误日志** 查看错误日志排查问题。
> - 如果要中止安装，可以单击 **取消安装**，在弹出框中单击 **确定**。

更多操作

模板安装完成后，在 **产品安装详情** 页您还可以根据实际业务需要执行更多操作。



具体操作可参考下表。

操作项	操作场景	操作步骤
更新模板版本	在产品模板安装前后，您可以快速更新要安装的产品模板版本。	1. 单击 **更新版本**。 2. 在弹出窗中选择要更新到的版本，单击 **确定**。
删除实例	如果不需要继续使用某个应用，您可以删除实例，删除后会释放所有后端资源。	单击 **删除实例** > **确定**。 **注意** 删除后无法恢复，请谨慎操作。
回滚历史快照	如果需要将应用回滚到之前的版本，您可以执行回滚操作。	1. 选择 **历史版本** 页签。 2. 单击目标版本操作栏中的 **回滚** ，然后单击 **确定**。 **说明** • 仅成功安装的历史版本支持回滚操作。 • 支持一键将应用回滚到某个历史版本，默认保留最近 10 个历史版本记录。

9.13. 基础元数据

9.13.1. 逻辑单元

在创建单元化工作空间或者工作空间组时，系统会自动进行逻辑单元和部署单元的规划。您可以根据实际需求，对逻辑单元进行修改或删除。本文介绍如何通过控制台为逻辑单元配置灾备及相关操作。

背景信息

逻辑单元是单元化架构的基础，一个单元被称为一个 Zone。根据业务特点不同，您可以将系统部署在不同类型的逻辑单元中。

- RZone（Region Zone）：部署按用户维度拆分的关键业务系统。核心业务和数据单元化拆分，拆分后分片均衡，单元内尽量自包含（调用封闭），拥有自己的数据，能完成所有业务。一个可用区可以有多个 RZone。
- GZone（Global Zone）：部署未按用户维度拆分的系统，被 RZone 依赖，提供不可拆分的数据和服务，如配置型的业务。数据库可以和 RZone 共享，多租户隔离，全局只有一组，可以配置流量权重。
- CZone（City Zone）：部署未按用户维度拆分的系统，被 RZone 高频访问 ，解决跨域通信延时问题。为了解决异地延迟问题而特别设计，适合读多写少且不可拆分的业务。 一般每个城市一套应用和数据，是 GZone 的快照。

配置灾备

灾备即容灾与备份结合的保障企业数据、业务高可用的方案之一。具体配置操作如下：

1. 登录控制台，在左侧导航栏单击 **基础元数据** > **逻辑单元**。
2. 在逻辑单元列表页，选择需要配置灾备信息的逻辑单元，单击 **灾备信息** 列中的 **配置**。
3. 在对话框中完成以下配置后，单击 **提交**。

配置项	说明
同城灾备组	可选。同城容灾场景下的灾备分组。 说明 不支持和 异地灾备组 选择相同的逻辑单元。
异地灾备组	可选。异地容灾场景下的灾备分组。 说明 不支持和 同城灾备组 选择相同的逻辑单元。
正式容灾状态	正式容灾或容灾演练场景下的状态。根据实际容灾场景进行选择。支持 正常、同城、异地。

查看灾备信息

您可以在逻辑单元列表页查看灾备信息。

1. 登录控制台，在左侧导航栏单击 **基础元数据 > 逻辑单元**。
2. 在逻辑单元列表页，选择需要查看灾备信息的逻辑单元，将鼠标光标悬浮于 **灾备信息** 列中的 **查看** 上方，即可查看当前逻辑单元配置的灾备信息。

删除逻辑单元

如果不再需要某个逻辑单元，您可以通过 **基础元数据 > 逻辑单元**，也可以通过 **全局设置** 进行删除（更多信息，请参见 创建逻辑单元）。

> 注意
> 删除逻辑单元前要先删除部署单元，若直接删除会报错。

1. 登录控制台，在左侧导航栏单击 **基础元数据 > 逻辑单元**。
2. 在逻辑单元列表页，选择需要删除的逻辑单元，单击 **操作** 列中的 **删除**。
3. 在弹出框中，单击 **确定**。

> 说明
> 删除逻辑单元会导致应用服务不可用，请谨慎操作。

9.13.2. 部署单元

在创建单元化工作空间或者工作空间组时，系统会自动进行逻辑单元和部署单元的规划。您可以根据实际需求，对部署单元进行修改或删除。

背景信息

部署单元（Cell）定义了一个服务通信的边界，用以规整通信的链路，为发布部署时切流做无损发布、高可用应用架构，以及容灾切换和恢复打下基础。部署单元（Cell）也是实现单元化的基础。

- 双机房双活场景下，一个应用的应用服务可以跨部署单元（Cell）部署。
- 单元化场景下，一个应用可以创建多个应用服务在不同部署单元（Cell）部署。

修改部署单元

1. 登录控制台，在左侧导航栏单击 **基础元数据 > 部署单元**。

2. 在部署单元列表页，选择需要修改的部署单元，单击 **操作** 列中的 **编辑**。
3. 在 **编辑部署单元** 页，根据需要修改相应信息。仅支持修改默认权重，即部署单元所承载的业务流量权重。
4. 完成后单击 **提交** 保存修改。

删除部署单元

如果不再需要某个部署单元，可以在部署单元列表页将其删除。

1. 登录控制台，在左侧导航栏单击 **基础元数据** > **部署单元**。
2. 在部署单元列表页，选择需要删除的部署单元，单击 **操作** 列中的 **删除**。
3. 在弹出框中，单击 **确定**。

> 说明
> 删除部署单元会导致应用服务不可用，请谨慎操作。

9.14. 常见问题

本文介绍通过产品使用过程中的一些常见问题，便于帮助您快速定位问题。

SOFAStack 常见基础问题

SOFAStack 常见基础问题如下：

- SOFAStack 是否可以部署在虚拟机、物理机上？
- SOFAStack 当前是否只支持阿里云飞天 ACK？
- SOFAStack 是否可以被客户侧现有系统集成？

SOFAStack 是否可以部署在虚拟机、物理机上？

可以，SOFA 部署支持物理机、虚机 VM。

SOFAStack 当前是否只支持阿里云飞天 ACK？

SOFA 部署支持多云异构、目前支持华为云、开源 OpenStack 等。

SOFAStack 是否可以被客户侧现有系统集成？

可以，SOFAStack 提供开放标准 API 接口，供客户侧系统调用集成。

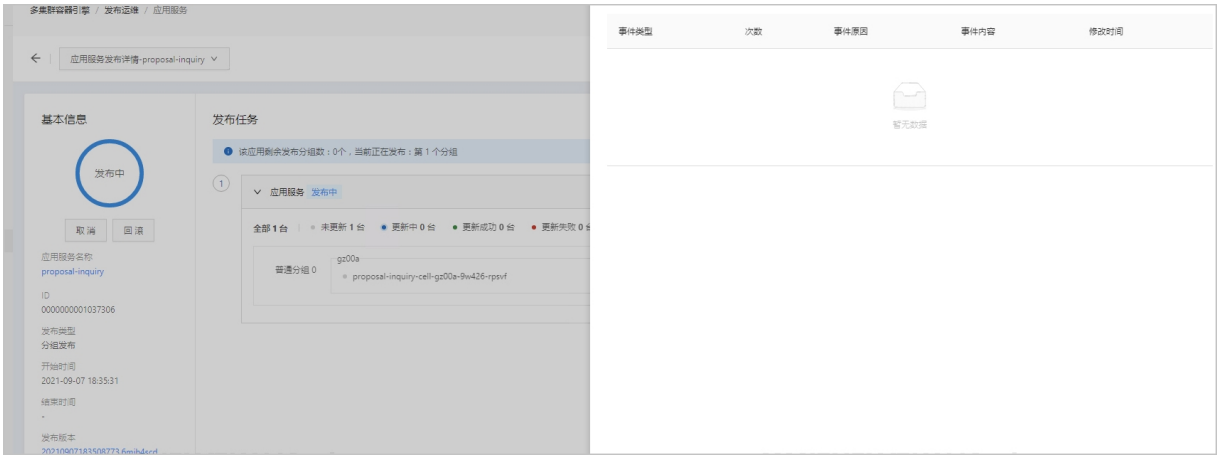
应用发布常见问题

发布卡单且 Pod 无事件

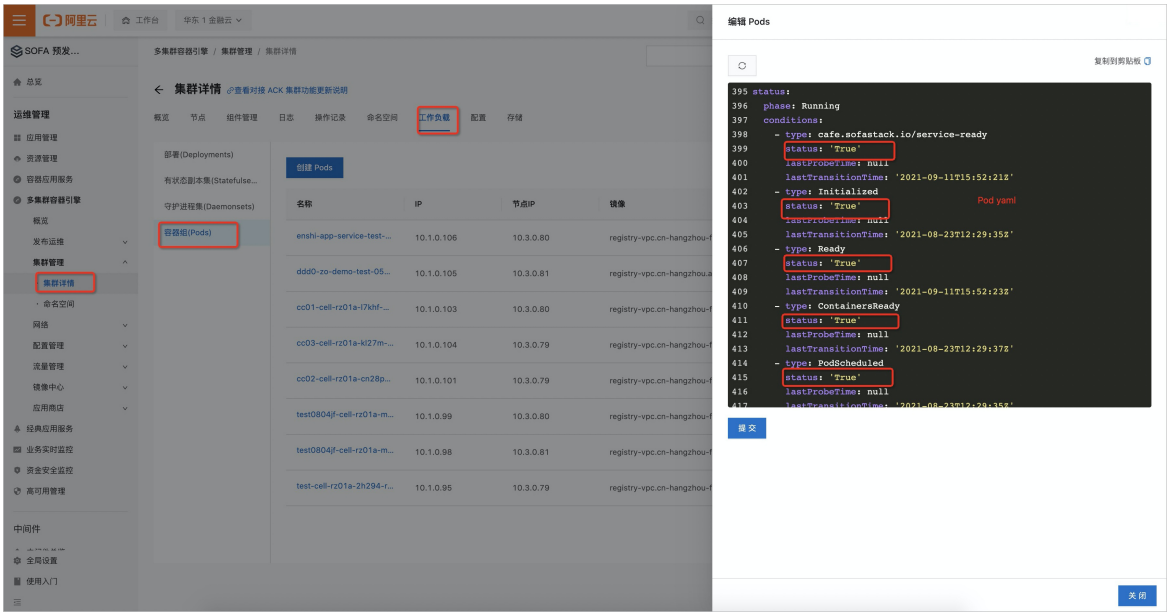
> 注意
> 只有出现的问题与 **问题描述** 中描述的完全匹配时，才可按照本节所列的办法处理。否则，仍须查看 Pod 的事件查找错误原因。

问题描述：

发布单一直处于发布中，Pod 状态没有更新，且事件列表为空，见下图。



1. 单击 **集群管理** > **集群详情**。
2. 在 **集群详情** 页，单击 **工作负载** > **容器组（Pods）**。
3. 查看 Pod YAML，发现 `status` 内所有状态全部是 true，见下图。



4. 等待 10 分钟左右，发布失败，Pod 状态依然是灰色的，见下图。

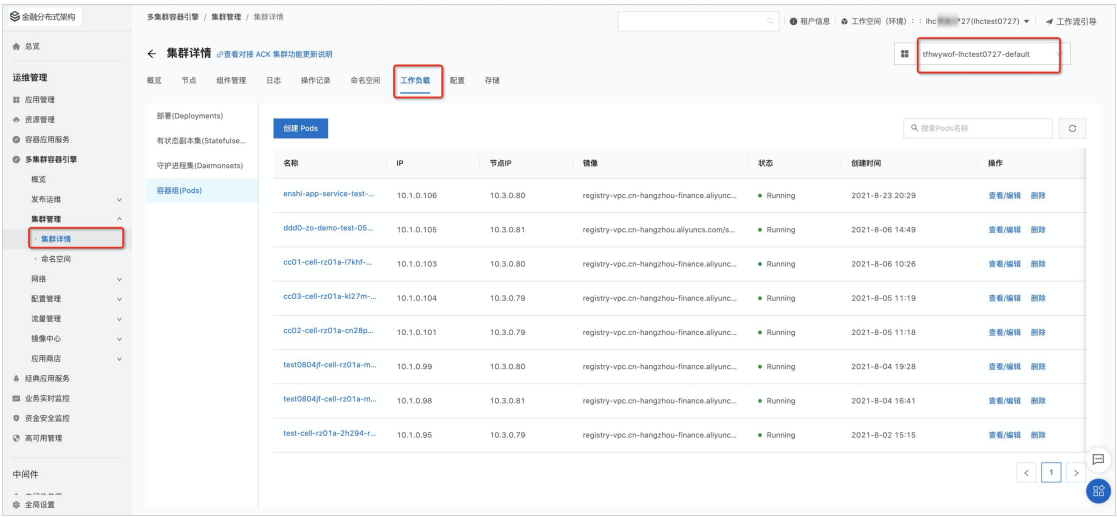


解决办法：

方法一：直接删除 Pod

> **注意**
> 方法一直接删除 Pod 会导致 Pod 无法原地升级，并且新创建出来的 Pod name 和 IP 都会发生变化。

1. 找到卡单的 Pod。
   i. 单击 **集群管理** > **集群详情**。
   ii. 在 **集群详情** 页，单击 **工作负载** > **容器组（Pods）**。
   iii. 切换命名空间，找到卡单的 Pod。



2. 直接删除 Pod，会自动创建新的 Pod。
3. 前往 **发布单** 页面，单击 **重试**，观察 Pod 的状态变化。

方法二：为 Pod 添加 Label

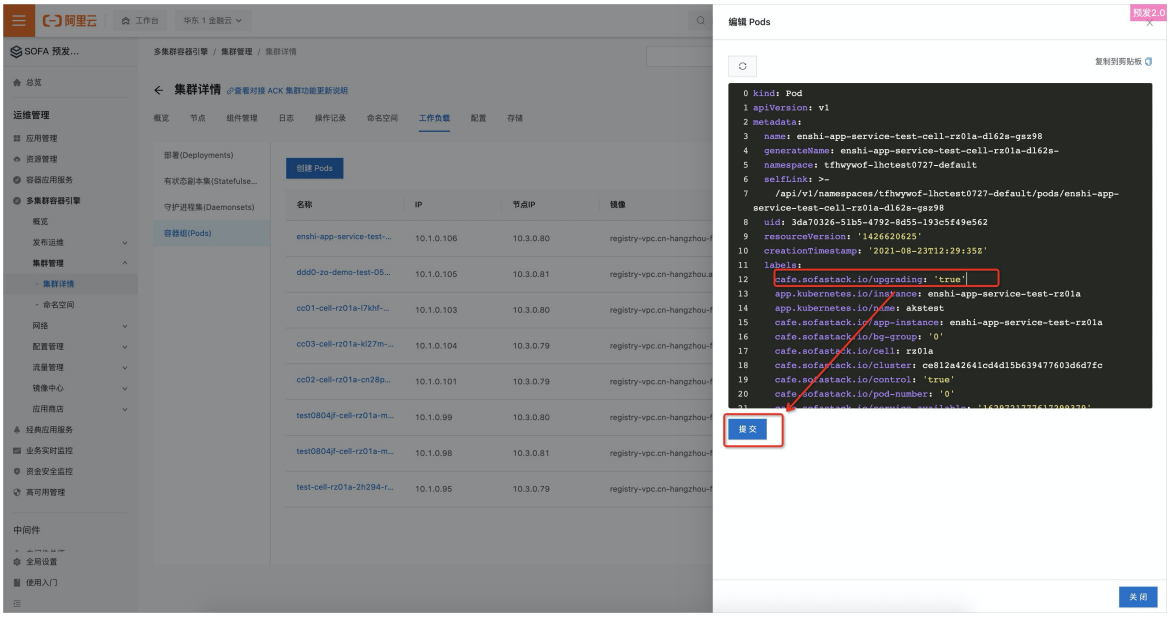
如果本次发布只是更新了镜像，并且希望 Pod name 和 IP 保持不变，请使用方法二为 Pod 添加 Label 进行解决。

1. 找到卡单的 Pod。
   i. 单击 **集群管理** > **集群详情**。

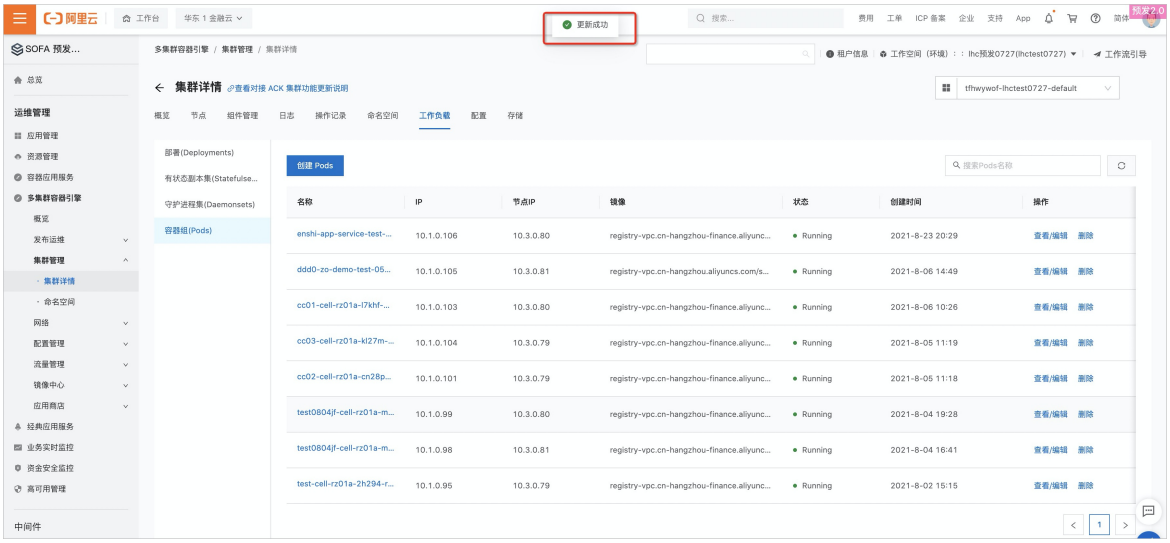
ii. 在 **集群详情** 页，单击 **工作负载 > 容器组（Pods）**。

iii. 切换命名空间，找到卡单的 Pod。

2. 单击 **查看/编辑**，手动为 Pod 增加 Label `cafe.sofastack.io/upgrading: 'true'` 后，单击 **提交**。



3. 提交成功之后，页面会提示“更新成功”。



4. 更新完成之后，回到发布单页面单击 **重试**，观察 Pod 的状态变化。

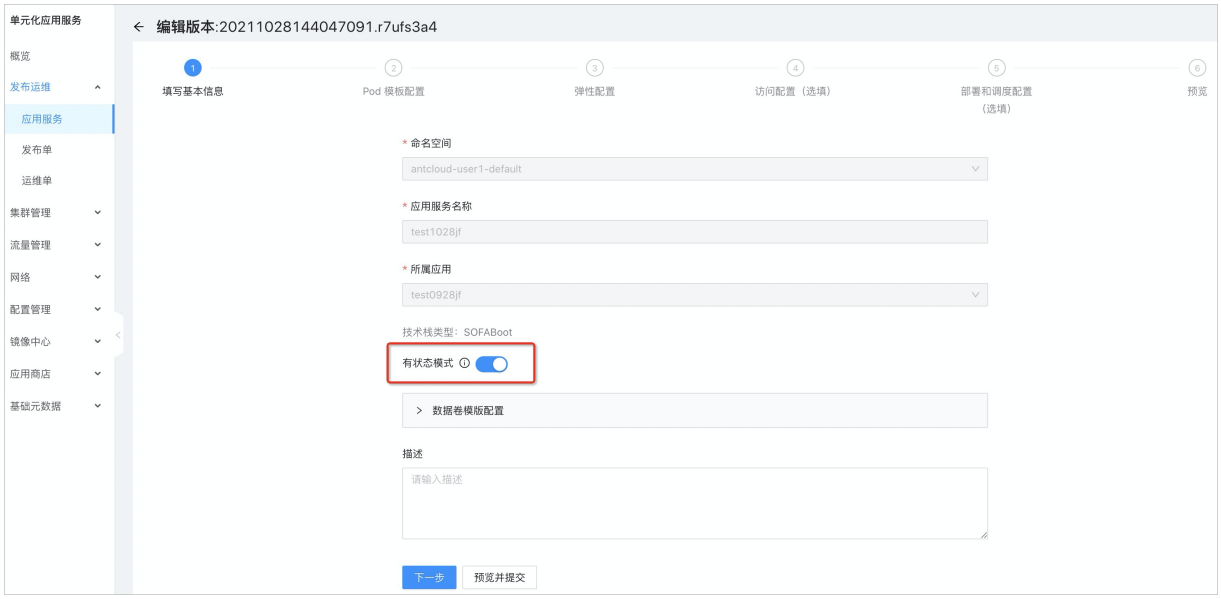
9.15. 最佳实践

9.15.1. LHC Pod 域名配置

本文介绍如何通过域名访问 Pod。

步骤一：开启有状态模式

在创建应用服务或者编辑应用服务的 填写基本信息 阶段，开启 **有状态模式**。



开启后，Pod 名称会加上 `-0`、`-1` ... 数字后缀。待应用服务发布成功后，可前往应用服务详情页查看 Pod 名称是否已加上 `-0`、`-1` ... 数字后缀。



步骤二：配置 Headless Service

在创建应用服务或者编辑应用服务的 访问配置 阶段，添加 Headless Service 名称。本例中输入 **svc**。Headless Service 名称即为 **test0928jf-svc**。



步骤三：基于域名规则计算 Pod 域名

域名规则如下：

- Service 命名规则：<Headless Service 名称>-cell-<部署单元名>

示例：test0928jf-svc-cell-rz00a

- Hostname 命名规则：<应用服务名>-<部署单元名>-<Pod number>

> 说明
> - Pod Number：Cell 维度从 0 开始递增的数字。
> - 应用服务发布完成后，可前往应用服务详情页查看 Pod 名称，Pod 名称最后一位即为 Pod Number。

示例：test0928jf-rz00a-0

- Pod 域名规则：<hostname>.<service name>
  示例：test0928jf-rz00a-0.test0928jf-svc-cell-rz00a

Pod 域名规则		
< hostname >	.	< service name >
示例：test0928jf-rz00a-0.test0928jf-svc-cell-rz00a		

Hostname 命名规则	Service 命名规则
< 应用服务名 >-< 部署单元名 >-< Pod number >	< Headless Service 名称 >-cell-< 部署单元名 >
示例：test0928jf-rz00a-0	示例：test0928jf-svc-cell-rz00a

假设部署单元 rz00a 有 2 个副本，基于上述规则，计算结果如下：

计算项	计算结果
Service	test0928jf-svc-cell-rz00a
Hostname	test0928jf-rz00a-0 test0928jf-rz00a-1
Pod 域名	test0928jf-rz00a-0.test0928jf-svc-cell-rz00a test0928jf-rz00a-1.test0928jf-svc-cell-rz00a

通过 Ping 命令验证能否访问 Service：test0928jf-svc-cell-rz00a。



通过 Ping 命令验证能否访问 Pod 域名：test0928jf-rz00a-0.test0928jf-svc-cell-rz00a、test0928jf-rz00a-1.test0928jf-svc-cell-rz00a。

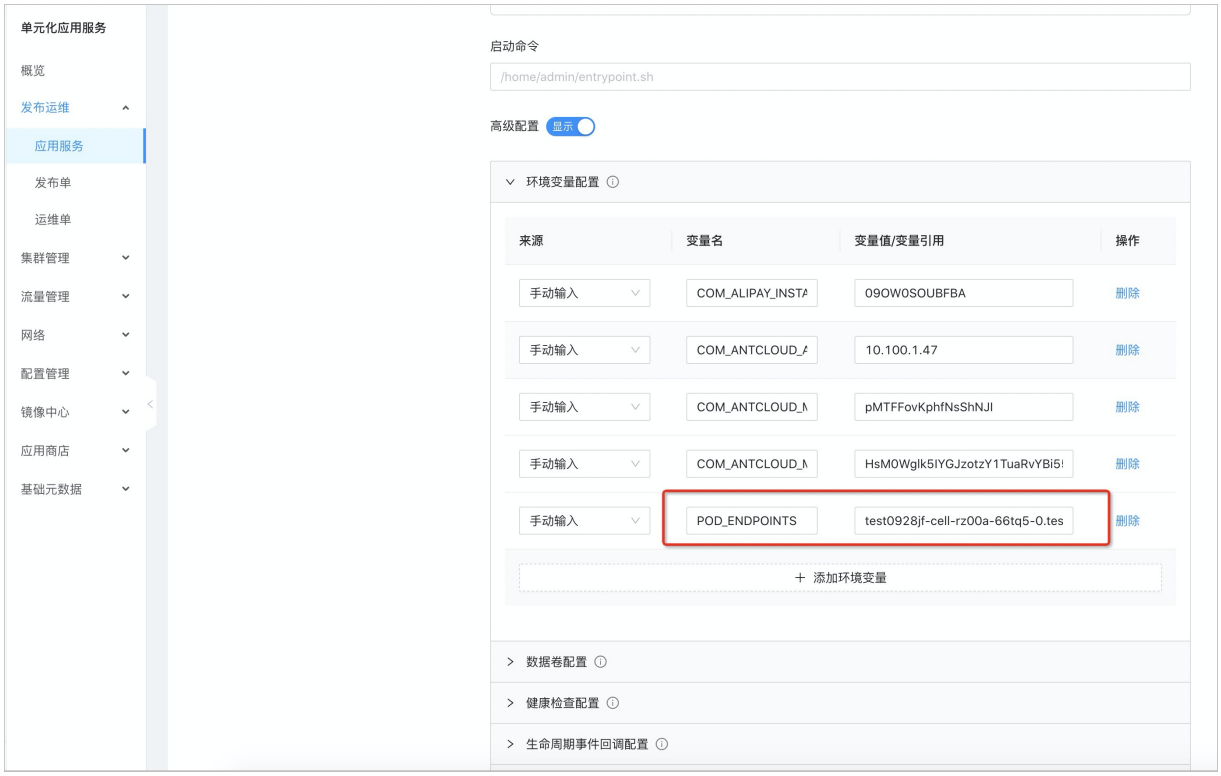
```
sh-4.2#
sh-4.2# ping test0928jf-rz00a-0.test0928jf-cell-rz00a
PING test0928jf-rz00a-0.test0928jf-cell-rz00a.antcloud-user1-default.svc.cluster.local (10.110.3.184) 56(84) bytes of data.
64 bytes from test0928jf-rz00a-0.test0928jf-cell-rz00a.antcloud-user1-default.svc.cluster.local (10.110.3.184): icmp_seq=1 ttl=62 time=0.219 ms
64 bytes from test0928jf-rz00a-0.test0928jf-cell-rz00a.antcloud-user1-default.svc.cluster.local (10.110.3.184): icmp_seq=2 ttl=62 time=0.201 ms
^C
--- test0928jf-rz00a-0.test0928jf-cell-rz00a.antcloud-user1-default.svc.cluster.local ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 1000ms
rtt min/avg/max/mdev = 0.201/0.210/0.219/0.009 ms
sh-4.2#
sh-4.2#
sh-4.2# ping test0928jf-rz00a-1.test0928jf-cell-rz00a
PING test0928jf-rz00a-1.test0928jf-cell-rz00a.antcloud-user1-default.svc.cluster.local (10.110.3.183) 56(84) bytes of data.
64 bytes from test0928jf-rz00a-1.test0928jf-cell-rz00a.antcloud-user1-default.svc.cluster.local (10.110.3.183): icmp_seq=1 ttl=62 time=0.253 ms
64 bytes from test0928jf-rz00a-1.test0928jf-cell-rz00a.antcloud-user1-default.svc.cluster.local (10.110.3.183): icmp_seq=2 ttl=62 time=0.216 ms
64 bytes from test0928jf-rz00a-1.test0928jf-cell-rz00a.antcloud-user1-default.svc.cluster.local (10.110.3.183): icmp_seq=3 ttl=62 time=0.214 ms
^C
--- test0928jf-rz00a-1.test0928jf-cell-rz00a.antcloud-user1-default.svc.cluster.local ping statistics ---
3 packets transmitted, 3 received, 0% packet loss, time 2000ms
rtt min/avg/max/mdev = 0.214/0.227/0.253/0.025 ms
```

> **注意**
> Pod 域名只能在集群内部访问，如果应用服务发布的部署单元属于不同的集群，则不同的部署单元之间通过 Pod 域名无法访问。

步骤四：添加环境变量 POD_ENDPOINTS = <Pod 域名>

在编辑应用服务的 Pod 模板配置 阶段，开启 **覆盖配置** 后，新建环境变量，将上面的域名列表写入环境变量，变量名支持自定义：

POD_ENDPOINTS= test0928jf-rz00a-0.test0928jf-svc-cell-rz00a,test0928jf-rz00a-1.test0928jf-svc-cell-rz00a



待应用服务发布后，即可从环境变量中获取所有 Pod 的域名列表。环境变量名可以按需定义。

10.实时监控（RMS 3.0）

10.1. 产品简介

10.1.1. 什么是实时监控

实时监控服务（Real-time Monitoring Service，简称 RMS）是一款具有可视化监测能力的金融级监控产品。基于日志、指标、链路等海量数据进行多维聚合，向用户提供业务监控、应用监控、云原生监控、基础资源监控、日志查询分析、分布式链路等多角度的可视化监测功能，有丰富的可视化大盘，并提供了告警订阅功能。

该服务可以帮助运维、研发、SRE（Site Reliability Engineer）等快速地发现问题、定位问题、分析问题、解决问题，为线上系统可用率提供有效保障。

全方位实时监控

提供业务、应用、基础资源、云原生等各种视角的监控能力，可实现关键指标秒级、普通指标分钟级的监控，具有高可靠、高时效、低延迟的特点。

灵活的报警规则

可根据业务特征、时间段、重要程度等维度设置报警规则，实现不误报、不漏报。

便捷的自定义配置

具有丰富的自定义产品配置功能，可便捷、高效地进行产品、报警配置。

开放的技术栈配置

实现 Kubernetes、SOFA 技术栈应用部署即监控，通过简单技术栈配置，即可接入并监控非标业务应用。

可视化大盘

丰富的可视化大盘，帮助您定制个性化的监控大盘。

低资源占用

在可靠传输大量监控数据时，保证对宿主机的 CPU、内存等资源的极低占用率。

高可用

提供万台设备的分钟级监控部署能力，故障自动恢复，集群可伸缩。

稳定高效的时序和数据存储

可在线持续聚合数据，保证数据容量可控，提供智能分级存储、存放策略。

分布式链路

提供应用拓扑和链路查询功能，观测应用及服务之间的复杂调用关系与性能指标，从而实现故障根因分析、服务治理、应用开发调试、性能管理、性能调优、架构管控、故障定责等运维开发工作。

10.1.2. 功能架构

RMS 通过采集各种指标、日志和链路等数据，并进行海量数据的清洗、计算。以此来支撑产品层的应用监控、业务监控、平台监控、基础设施监控、告警管理和分布式链路等能力。作为一个企业级产品，RMS 提供资源租户隔离、访问鉴权控制和监控配置模板等企业级特性。并且提供数据高可用、服务高可靠、双机房容灾部署等平台特性。
基于 RMS 丰富数据和强大功能，可以支撑容灾巡检、故障重放、弹性扩缩、微服务治理和全链路压测等场景。



- **应用监控**
  RMS 能同时监控容器应用和经典应用，并通过 LDC、IDC 和单机实例等多视角、多维度逐层下钻分析，实时展现服务实例、依赖的中间件和基础资源运行状态、使用趋势和告警信息，发掘应用故障所在的层级和对象，保证应用的流畅运行。
- **业务监控**
  RMS 提供了灵活的、基于业务场景的自定义业务监控，通过业务监控可将不同监控图表展示到同一个屏幕上，通过不同的大盘模板、统计模板等形式来展示可观测数据，例如，分钟级多 Key、TopN 等，让用户可以全面、深入地掌握业务数据。
- **中间件监控**
  RMS 默认集成了对消息、Mesh、数据库等中间件的监控，并支持在应用监控中查看应用所调用各中间件的情况。
- **基础资源监控**
  支持对物理机、虚拟机、Kubernetes 集群和原生容器等资源的监控。
- **分布式链路**

分布式链路帮助运维人员、开发人员和架构师看清楚复杂的大规模微服务架构下的应用及服务之间的复杂调用关系、性能指标、出错信息与关联日志，从而实现故障根因分析、服务治理、应用开发调试、性能管理、性能调优、架构管控、故障定责等运维开发工作。

- **告警管理**
  针对各资源对象，RMS 允许用户灵活地配置自定义告警规则，并支持多种订阅方式，如邮件、短信、钉钉等。当监控数据满足阈值条件时，第一时间通知对应的运维人员，帮助其发现异常及原因。
- **企业级特性**
  在企业级特性层面，RMS 包含的能力如下：
  - 提供多层级业务和资源隔离机制，比如租户、workspace 等，保证监控数据的安全性。
  - 支持监控具有多 VPC 服务实例的应用，并以应用中心视角展示监控数据。
  - 支持监控配置模板导入与导出，使测试环境的应用监控配置可在生产环境一键激活，避免了重复配置，也保证了发布流程的标准化和可控性。

10.1.3. 功能概览

本文介绍实时监控 RMS 常用功能。在使用本产品之前，建议您先了解功能架构、基本概念等，以便更好地理解 RMS 提供的功能。

风险发现

RMS 提供风险发现相关功能项如下：

功能分类	功能名称	功能说明
监控	业务监控	• 支持对应用业务日志进行灵活切割。 • 按业务维度聚合数据，提供一套完整的业务实时分析能力。 • 主要面向个性化的非标业务监控场景，比如交易量、放贷量等。 • 实现关键指标秒级、普通指标分钟级等高可靠、低延迟能力。
	应用监控	• 支持对容器应用和经典应用双重监控。 • 提供对应用从内到外的多维监控，比如服务指标、系统指标、JVM 指标、端口探测等。 • 按照单实例、机房、单元化等视角对监控指标进行多视角聚合，提供应用可用性观测保证。
	主机监控	支持对物理机和虚拟机的监控，比如系统指标、端口探测等。
	云原生监控	• 提供对云原生丰富的监控能力，从 Kubernetes 集群、Kubernetes 组件以及容器运行时等多维度进行监控。默认提供云原生最佳监控的可视化面板。 • 提供多集群、集群、应用等多视角进行对比分析，让运维人员切实感知 PaaS 层稳定性。
	Prometheus 监控	支持采集符合 Open-Metrics 的 Prometheus 指标，作为统一标准格式存储为时序数据，并提供可视化大盘配置。
	蚂蚁产品监控	原生采集 SOFA 框架日志的监控指标，默认提供对蚂蚁各产品的监控和告警能力。

功能分类	功能名称	功能说明
	自监控	提供对 RMS 自身组件可用性的监控告警和巡检能力，以及支持部署 RMS- SM 实现对 RMS 整体监控。
告警	告警规则	支持灵活的为监控对象配置告警规则。 当规则条件被触发时，系统将产生告警事件，例如为告警规则订阅通知对象，系统将为通知对象发送告警消息，以提醒用户采取必要的问题解决措施。

风险定位

RMS 提供的风险定位相关功能项如下：

功能分类	功能名称	功能说明
链路追踪	全局应用拓扑	持续自动发现整个应用间调用，以及对中间件的依赖调用架构关系，绘制完整的全局拓扑，并在拓扑中展示节点和连线异常状态，帮助用户更快感知到故障问题。
	应用下钻和性能分析	支持基于应用 >上下游应用 > 接口等逐层下钻分析，并展现其黄金性能指标（比如吞吐量、平均响应时、错误率、满意度等），建立由底层到上层间的数据关联信息，从而深度分析分布式场景下的影响应用性能的问题根因。
	多维链路查询	支持按照多维业务条件检索调用链，包括 RPC、MVC、REST 、DB 等不同类型调用方式，可实现按照各场景自定义 Tag（比如交易号）查询对应的调用链路集合。
	链路详情	提供系统间调用的链路图、链路详情、时序图等，清晰全方位的展现系统间调用的交互、性能和异常信息。
	兼容 Open-telementry	原生与 SOFAStack 上的应用进行无缝对接，用户的业务代码无需任何修改即可轻松接入，也支持接收 zipkin、skywalking 等符合 Open-Tracing 的链路数据，统一生成应用拓扑和链路追踪。

平台特性

RMS 平台特性如下：

- **低资源占用**
  在完成大量数据可靠采集和传输的同时，保证对宿主机的 CPU、内存等资源的使用限制和极低占用率。
- **稳定高效的数据计算和存储**
  支持分钟级对万台设备的实时监控和在线持续聚合数据，智能分级存储存放策略 ，时序数据库保证数据容量可控和弹性扩缩平台。
- **高可用**
  支持双机房容灾架构，组件多实例运行，数据库多副本同步，并提供自监控、故障自动恢复等能力。
- **开放自主**
  所有组件皆为自研可控，并经过大规模生产验证；提供完善的产品功能和数据的 OpenAPI，更好的和第三方系统集成对接。

10.1.4. 实现原理

监控的通用逻辑



- **定义数据源**：首先，需要建立元数据基本模型，元数据是监控系统的根基。例如：定义应用、部署实例、机房信息，单元化信息等。在 SOFAStack 场景下，这部分信息需要从 PaaS 侧自动获取。
- **数据采集**：通过安装在机器上的 Agent 采集监控数据，这部分可以是日志，也可以是通过各种方式获取到的指标。
- **数据清洗**：将采集得到的非结构化的数据，解析为监控系统可以处理的结构化信息。这部分解析能力需要是事先通过监控系统预设好的某种规则。例如，日志解析规则、Prometheus 协议规则等。
  举个例子，有业务日志样例如下，希望通过监控获取每一分钟各商品各自创建了多少笔交易这一数据。

```
2012-11-11 11:11:11,2950211004,衣服,交易,创建,7,Y,
2012-11-11 11:11:12,2950211005,衣服,交易,创建,8,Y,
2012-11-11 11:11:13,2950211006,食品,交易,创建,9,Y,
2012-11-11 11:11:14,2950211007,家居,交易,创建,11,Y,
2012-11-11 11:11:15,2950211008,食品,交易,创建,7,N,
……
```

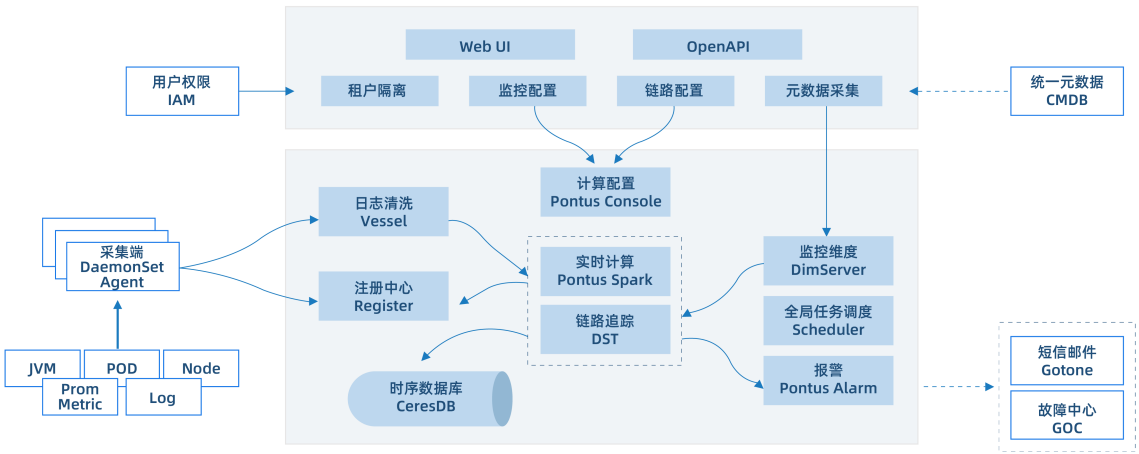
> 说明
> 日志格式为：`日志打印时间,订单id,商品,服务名,方法名,耗时,结果` 。

- **数据统计**：由于监控系统面对的是海量集群，在获取结构化数据之后，需要对数据进行数据统计。统计方式可以有很多种，比如求行数、求和、求平均、最大/最小值等统计方法。
- **预警配置**：可在统计数据上定义类型丰富的预警规则，包括绝对值、同比、环比、最近 N 分钟求和等，支持定义多项规则和逻辑运算。

实现原理

RMS 的基本实现原理是在所有的应用主机上部署 Agent（代理），从而采集主机和应用的各种指标、日志和链路等数据，并将其存储在 CeresDB 中，通过在 PontusSpark 中进行分析计算，实现在不同产品场景中的可视化展现和告警通知。

RMS 实现原理示意图



采集层

DaemonSet Agent（又称 Pontus-Agent）日志采集代理

- 负责所有的数据的最终采集，包括原始日志拉取和指标型数据采集。
- 具有插件化的能力，统一本地调度能力，负责拉起插件和解析数据，解析完了之后做一次聚合计算。然后放入本地的缓存中，等待 Vessel 来拉取数据。
- 提供一次性的任务执行能力，主要用于性能分析。
- 启动时会和 Register 进行通信。通信之后会获取 Agent 运行所需要的所有配置，包含运行时采集配置，Vessel 建连的配置等。

Register 注册中心

- 负责与 Agent 之间保持网络心跳，收集所有 Agent 的状态，版本透视等信息。
- 负责 Agent 与 Vessel 的配置生成与下发，即负责管控所有的采集配置。
- 负责管控 Agent 与 Vessel 之间的连接与会话（session），Register 感知 Vessel 集群负载，负责通知 Agent，应该和哪个 Vessel 地址进行建连。

Vessel 流量清洗组件

- 抽象来看，Vessel 其实就是一个独立部署的 Agent，这个 Agent 通过 remote input 的形式，拿到了日志或者其他的原始数据。
- 一个 Vessel 服务可以托管 N 个 Agent，即从 N 个 Agent 中收集数据。
- Vessel 可以通过一系列的解析规则，最终将非结构化的数据变成结构化的数据并返回，也就是所谓的数据清洗。
- 所有上层组件（包括计算层与链路层组件）都是通过 Vessel 来获取数据。

计算层

DimServer维度数据服务（又称元数据服务）

- 包括维度数据的读和写的服务。在这一层上无业务语义，业务语义为上层产品层注入。
- 计算和采集层都会用到维度数据（元数据），通过 DimServer 提供的 Java Client 来实现数据通讯。
- 为适配任意维度数据结构，DimServer 实现了维度数据 Table 化，即使用若干张宽表来处理所有可能的元数据表结构，这类宽表被称为维度表。

PontusConsole 监控数据平台的管控端

- 主要负责内部运维性管控、系统配置，以及提供给产品使用的配置接口。
- 提供操作维度表、指标表的标准数据服务接口。

Scheduler 全局任务调度器

- 不依赖任何中间件，实现全局任务调度能力，仅用于监控系统内部，实现功能自闭环。
- 负责根据配置信息定时生成并分发监控的计算任务。
- Scheduler 可以实现多机房部署，通过数据库维持高可用。

PontusSpark 分布式计算引擎

- 功能强大的 Spark 计算集群，负责接收产品层的计算配置，并对数据进行离线计算和统计。
- 从 Register 中获取监控数据采集的 Vessel 地址，并通过 Vessel 获取必要的监控数据。

PontusAlarm 监控告警组件（又称 RMSAlarm）

- 主要负责将监控计算完成的告警信息以短信、邮件、钉钉等方式向外推送。
- 在 SOFAStack 输出的模式下，告警渠道主要通过 Gotone 服务提供。

MonitorGateway 分布式链路追踪

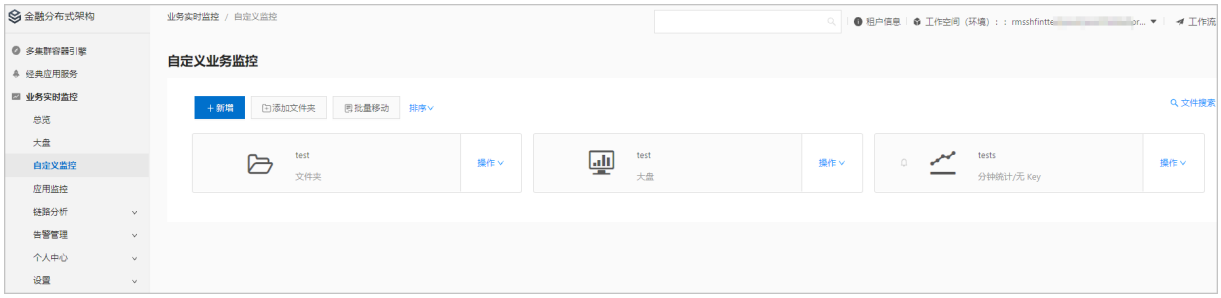
- 主要负责分布式链路处理的相关功能。
- 对于链路信息的收集，提供三种方式，分别为主动日志采集、SLS 日志采集以及链路日志上报。

数据层

主要是由时序数据库 CeresDB 来承载。CeresDB 是蚂蚁自研的时序数据库，时序引擎是一种存储和管理时间序列数据的分布式数据库，为时间序列提供高性能读写、预处理计算、可视化查询等功能。

产品层

产品层（应用名为 monitorprod）承载了 RMS 的所有页面以及用户交互、配置逻辑，是监控对客的主要门户。如下图所示：



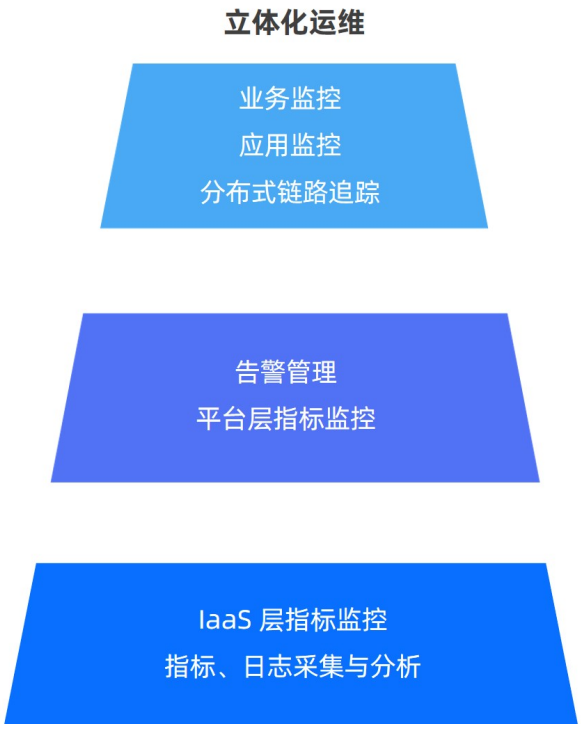
产品层负责与用户权限系统打通，读取用户身份信息。同时，产品层也负责打通 PaaS 元数据，将元数据同步至 DimServer 。

> 说明
> PaaS 元数据，是指通过 PaaS 系统发布的描述应用的基本信息，例如应用 IP、机房、单元化等信息。

10.1.5. 应用场景

立体化运维

深度对接 Kubernetes、SOFA 等技术栈应用服务，一站式收集基础设施、中间件、应用运行时数据和业务数据，通过指标监控、日志分析、链路追踪、告警订阅等功能，对应用性能、运行状态和资源使用等提供立体化运维分析，及时发现并定位应用、资源及平台的问题。



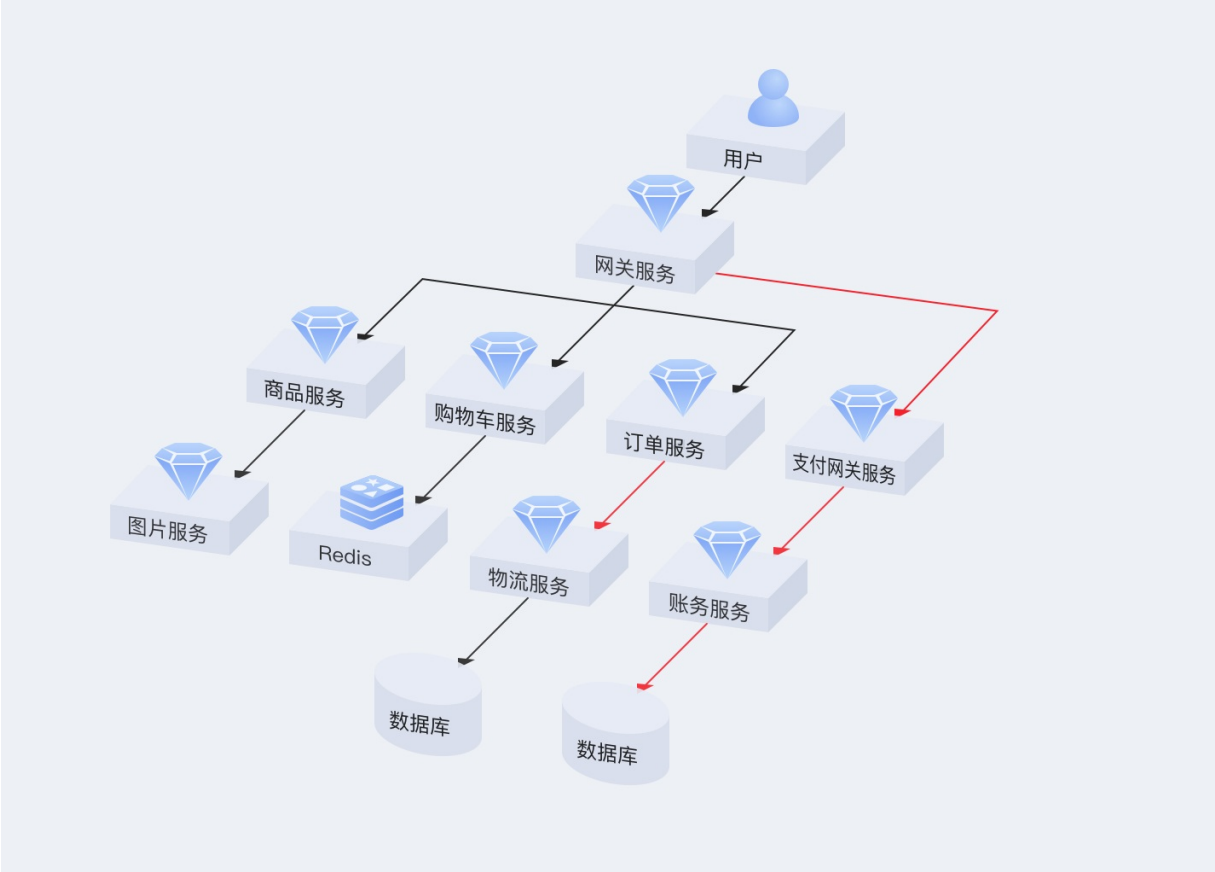
- **一站式分析**：在应用总览中可查看应用错误指标趋势、应用链路、应用指标和系统指标等，提供一站式的应用分析能力。
- **全方位监控**：覆盖基础设施、云数据库、云中间件、应用等多个维度监控，提供一站式运维能力。
- **故障关联分析**：以应用为中心，覆盖组件、实例、主机、云资源等多维度关联分析，迅速找到异常故障点。

问题分析与快速定位

在分布式场景下，服务调用错综复杂，问题分析与定位非常困难，分布式链路跟踪系统能迅速定位到有问题的服务，协助快速解决问题节点。

- **完整的应用调用拓扑关系**：自动发现该服务的历史调用，以及对所有中间件的调用，绘制整个系统调用关系的拓扑图。
- **快速定位不健康应用**：在调用关系拓扑中，对不健康应用进行显式标识，便于快速发现有问题应用并进行分析。
- **服务性能详情**：调用拓扑中的应用都可以单独进行下钻分析，可以从吞吐量、错误率、响应时间等指标出发，对应用性能进行详细分析。

问题分析与快速定位示意图

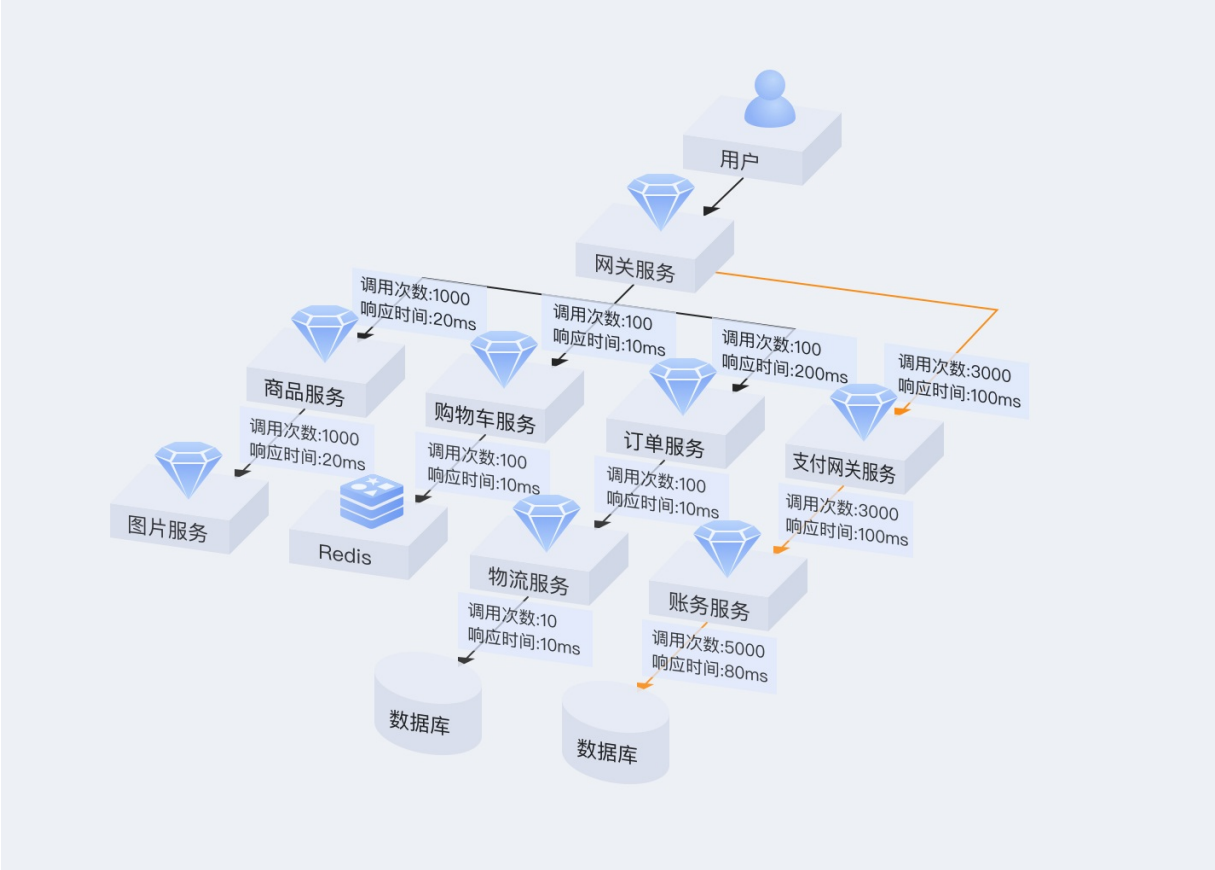


应用性能优化

在调用关系拓扑中，可以对各个应用的调用次数以及耗时情况进行分析，找到负载较高以及负载较少的应用，从而对资源进行合理利用。

- **调用链路聚合汇总**：对所有的调用信息进行聚合汇总，对各个应用的调用情况以及响应情况进行分析。
- **关键路径**：快速发现整个系统调用拓扑中，关键应用的路径。
- **优化不合理调用**：及时发现某些不合理的调用并进行处理，如频繁进行数据库操作等。

性能优化示意图



10.1.6. 基本概念

基本概念表中的基本概念按中文拼音进行排序。

报表（Report）

指包含单个或多个数据源的视图，集中展示各数据源的监控结果。

大盘（Dashboard）

指包含多个报表的一个页面。

单机视角

从应用实例（单机）维度进行指标数据的聚合，从单机的空间分布和时间分布上对应用监控指标数据进行对比分析。

订阅

订阅后，通知组内的成员会收到监控告警通知。

服务指标

服务指标将应用相关的 Error、Service、SAL、CAL、DAL 等服务指标进行聚合透出，从 IDC（机房）、LDC（单元化）、单机的空间分布和时间分布上进行对比分析，一个入口总览分析应用相关的所有监控数据。

IDC 视角

从 IDC（机房）维度进行指标数据的聚合，从单机的空间分布和时间分布上对应用监控指标数据进行对比分析。

监控产品（Monitoring Products）

指为实现自定义监控而提供的一些工具，可用于配置监控指标、生成数据源。

监控对象

由一个或者多个维度描述的被监控目标，比如说一个容器就可以用单个容器的维度描述。一个应用的一个逻辑 Zone 就需要用两个维度来描述（App + Zone）。监控对象实际上是监控数据的归属。

框架配置

基于 SOFA/SOFABOOT 框架开发的应用，PV、Service、SAL、SQL、Cal、CE Thread 等监控项会自动基于默认框架下的日志路径采集指标，仅需开启监控即可。

LDC 视角

从逻辑机房（LDC）维度进行指标数据的聚合，从单机的空间分布和时间分布上对应用监控指标数据进行对比分析。

配置模板

RMS 提供监控配置模板功能，支持将应用和自定义监控告警以配置模板（JSON 文件）导出，然后一键导入其他环境，如该环境已部署同名应用，则监控告警配置即生效，帮助投产运维更加高效便捷和一致。

数据源（Data Source）

指通过监控产品配置的监控指标。

通知人（Notificant）

指报警消息的接收人，接收渠道为手机短信。

通知组（Notification Group）

每个通知组可包含一个或多个通知人。在管理通知时，可通过通知组订阅，将报警内容发送给通知组内所有的通知人。

预警（Alert）

指通过配置的报警规则，触发报警通知。

指标

一个被监控对象数量特征的概念和数值。通常可以用若干个指标来描述一个监控对象随时间变化的情况，达到监控谁发生了什么事的最终效果。

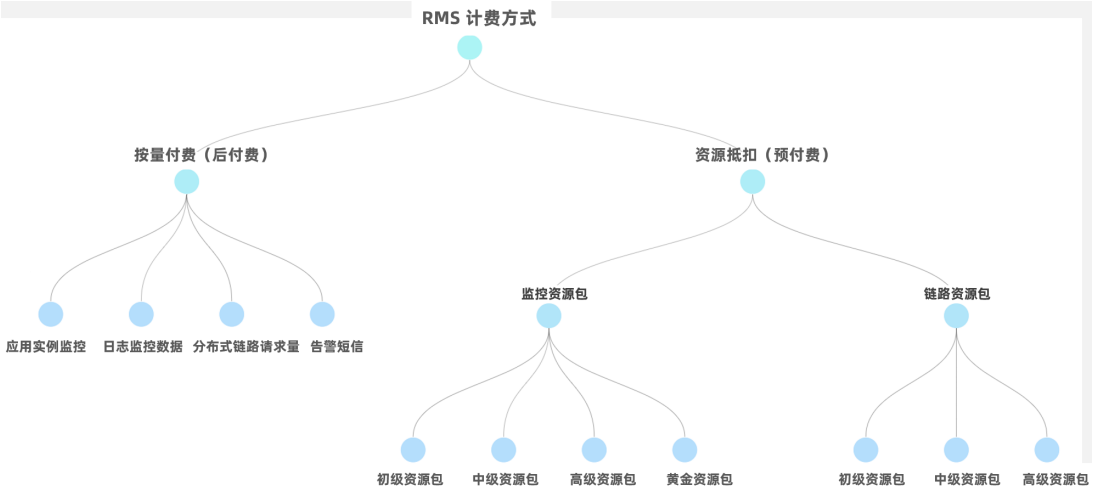
自定义监控配置

非 SOFA/SOFABoot 框架开发的应用，需为每个监控项配置采集日志路径和列值，比如 Error、Dal。如果是 SOFA/SOFABoot 框架开发的应用，但不希望使用默认日志路径监控的，也可以切换成 **自定义监控配置**。

10.2. 产品计费

10.2.1. 计费项概述

RMS 提供应用监控、自定义监控、基础资源监控、分布式链路和告警通知等能力。本文介绍 RMS 的计费资源项以及各资源的计费方式。



计费项目

RMS 计费资源项如下表所示：

计费项	说明	计费方式
应用实例	RMS 自动对应用实例部署所在的 ECS 进行安装 Agent，对应用的监控、链路和日志等数据进行采集，提供相关的可视化能力。并且默认提供应用的系统指标、JVM 指标、指定服务指标的监控。	• 按量付费 • 资源包抵扣

计费项	说明	计费方式
日志监控数据	Agent 直接采集应用业务日志数据，之后对日志进行处理、计算为指标。日志监控数据计量仅针对通过自定义监控配置采集日志。 按照每天总的日志摄入量进行计算。指标数据默认存储 21 天。	• 按量付费 • 资源包抵扣
分布式链路请求量	根据实际链路上报的 span 数决定。 按照每天总 span 量进行计量。链路数据默认存储 21 天。	• 按量付费 • 资源包抵扣

计费方式

RMS 各资源的计费方式如下表所示：

计费资源	计费方式	计费规则	相关文档
应用实例	按量付费	费用 = 实例单价 * 计费时长。 单位：元/个	应用实例监控计量
	资源包抵扣	在每个计费周期内按应用实例使用量扣除资源包额度。	监控资源包预付费
日志监控数据	按量付费	根据日志摄入量按阶梯计费。 单位：元/GB	日志监控数据计量
	资源包抵扣	在每个计费周期内按日志摄入量扣除资源包额度。	监控资源包预付费
分布式链路请求量	按量付费	费用 = 每日请求量 * 计费时长。 单位：元/百万	分布式链路请求量计量
	资源包抵扣	在每个计费周期内按每日请求量扣除资源包额度。	链路资源包预付费

10.2.2. 按量付费（后付费）

本文介绍按量付费（后付费）的购买说明及计费规则。

购买和使用须知

- 对于按量付费计费方式，开通服务后即可使用，使用即开始收费。
- 对于按量付费的用户，将在次日 00:00 扣除前一天的费用。

应用实例监控计量

应用实例分为 Pod 和经典应用进程两种实例，实例按照每小时进行计费，比如监控一个实例一天的计量为 1 实例 * 24 小时。应用实例监控计量计费规则可参考下表。

计费项	单位	阶梯范围（个）	公共云单价（元）	金融云单价（元）
实例数	元/个	不限	0.04	0.08

计费示例

以华东1（杭州）地域为例，假设平均每天需要监控的 Pod 有 100 个，则每个月的消费金额计算方法如下所示：

- 使用按量后付费，非金区后付费金额（元）= 100 * 24 * 30 * 0.04（元/Agent*Hour）= 2880
- 使用按量后付费，金区后付费金额（元）= 100 * 24 * 30 * 0.08（元/Agent*Hour）= 5760

日志监控数据计量

日志监控数据计量计费规则可参考下表。

计费项	单位	阶梯范围（GB）	公共云单价（元）	金融云单价（元）
日志摄入量	元/GB	(0，50]	3	6.0
		(50，100]	2.7	5.4
		(100，1000]	2.4	4.8
		(1000，10000]	2.1	4.2
		10000 以上	1.8	3.6

计费示例

以华东1（杭州）地域为例，假设平均每天需要监控的 Pod 有 100 个，预计一天的日志摄入量大概为 100 GB，则每个月的消费金额计算方法如下所示：

- 使用按量后付费，非金区后付费金额（元）= （50 * 3 + 50 * 2.7）* 30（元/GB）= 8550
- 使用按量后付费，金区后付费金额（元）= （50 * 6 + 50 * 5.4）* 30（元/GB）= 17110

分布式链路请求量计量

分布式链路请求量计费规则可参考下表。

计费项	单位	阶梯范围	公共云单价（元）	金融云单价（元）
span 数	元/百万	不限	0.5	1

计费示例

以华东 1（杭州）地域为例，假设平均每天需要监控的 Pod 有 100 个，预计一天产生 10 百万的请求量，则每个月的消费金额计算方法如下所示：

- 使用按量后付费，非金区后付费金额（元）= 10 * 0.5 * 30（元/GB）= 150
- 使用按量后付费，金区后付费金额（元）= 10 * 1 * 30（元/GB）= 300

告警短信计量

单日告警短信数量不超过 10 条不收取任何费用。若单日告警短信数超过 10 条，按以下计费逻辑收取费用：

- 公共云：短信走阿里云渠道，按每条 0.1 元收费，RMS 不做额外收费。
- 金区：RMS 收取额外的短信费用。

具体告警短信计量计费规则可参考下表。

计费项（防止滥用）	单位	阶梯范围（个）	公共云单价（元）	金融云单价（元）
告警数量（单日）	元/每个	10 个	免费	免费
		>10 个	0.1	0.2

10.2.3. 资源包抵扣（预付费）

本文介绍资源包抵扣（预付费）的购买说明和计费规则。

购买和使用须知

对于购买资源包的用户，将在次日 00:00 扣除前一天的资源消耗量。

监控资源包

监控资源包具体内容如下：

资源包名称	应用实例（个）	日志摄入量（GB）	资源包定价（元）		有效期
			非金融区	金融区	
初级监控资源包	864000	18000	61560	123120	1 年
中级监控资源包	4320000	90000	246240	492480	1 年
高级监控资源包	2880000	180000	369360	738720	1 年
黄金监控资源包	5960000	360000	615600	1231200	1 年

链路资源包

链路资源包具体内容如下：

资源包名称	span 数（百万）	资源包定价（元）		有效期
		非金融区	金融区	
初级链路资源包	36000	10800	21600	1 年
中级链路资源包	360000	90000	180000	1 年
高级链路资源包	3600000	720000	1440000	1 年

10.2.4. 计费案例

本文介绍按量计费和资源包计费相关案例，方便您根据实际业务需求选择最适合的付费方式。

> 说明
> 以下内容仅供参考，实际收费以账单为准。

案例 1 监控资源包 VS. 各监控项计量

案例背景

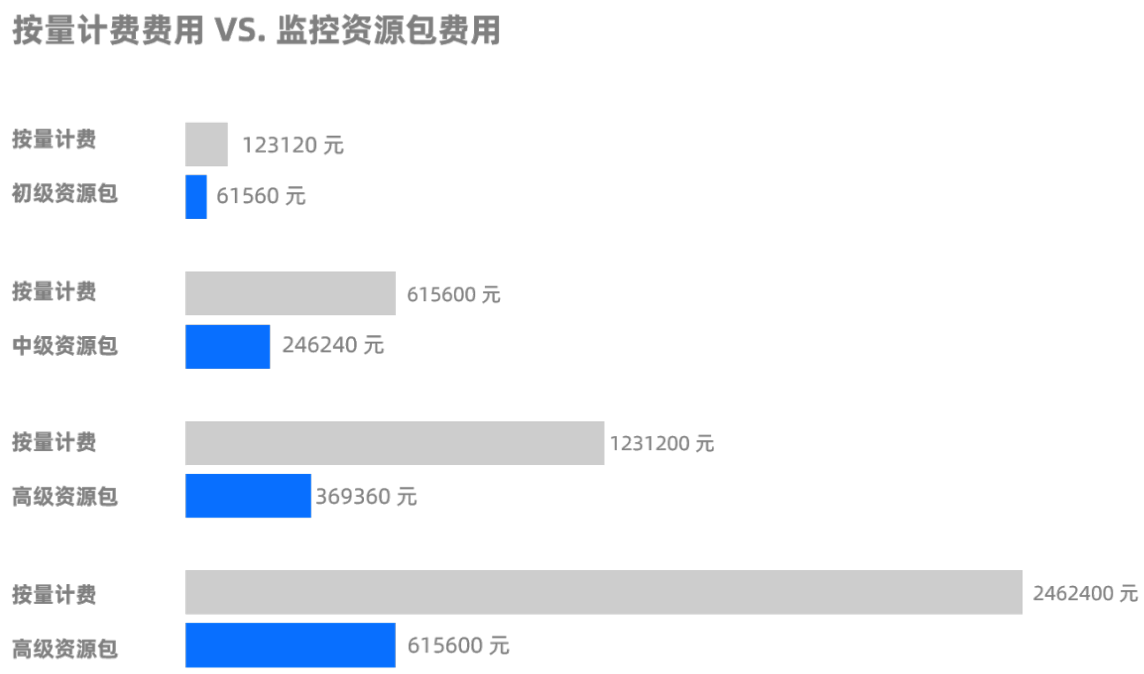
用户 A 在控制台想要购买监控相关功能项。用户想计算下按量付费和资源包哪个更优惠。

- 所在区域：非金区
- 时长：1 年

计费分析

预付费		按量付费		金额（元）	
资源包	资源包明细	计费项	计费项明细	资源包抵扣【1】	按量付费
初级监控资源包	• 应用实例：86400 个 • 日志摄入量：18000 GB	• 应用实例 • 日志监控数据	• 应用：100 个（每小时） • 日志量：50 GB（每天）	61560	123120 **计费逻辑**： 总费用 = 34560（应用每小时： 100）+88560（日志量每天：50）= 123120
中级监控资源包	• 应用实例：4320000个 • 日志摄入量：90000 GB		• 应用：500 个（每小时） • 日志量：250 GB（每天）	246240	615600 **计费逻辑**： 总费用 = 172800（应用每小时：500）+442800（日志量每天：250GB）= 615600
高级监控资源包	• 应用实例：2880000 个 • 日志摄入量：18000 GB		• 应用：1000 个（每小时） • 日志量：500 GB（每天）	369360	1231200 **计费逻辑**： 总费用 = 345600（应用每小时：1000）+885600（日志量每天：500GB）= 1231200
黄金监控资源包	• 应用实例：5960000 个 • 日志摄入量：360000 GB		• 应用：21000 个（每小时） • 日志量：1000 GB（每天）	615600	2462400 **计费逻辑**： 总费用 = 691200（应用每小时：2000）+1771200（日志量每天：1000GB）= 2462400

> 说明
> **资源包抵扣【1】**：所示价格为非金区价格。金区价格统一为非金区价格的 2 倍。



案例 2 链路资源包 VS. 链路请求量计量

案例背景

用户 B 在控制台想要购买链路相关功能项。用户想计算下按量计费和资源包哪个更优惠。

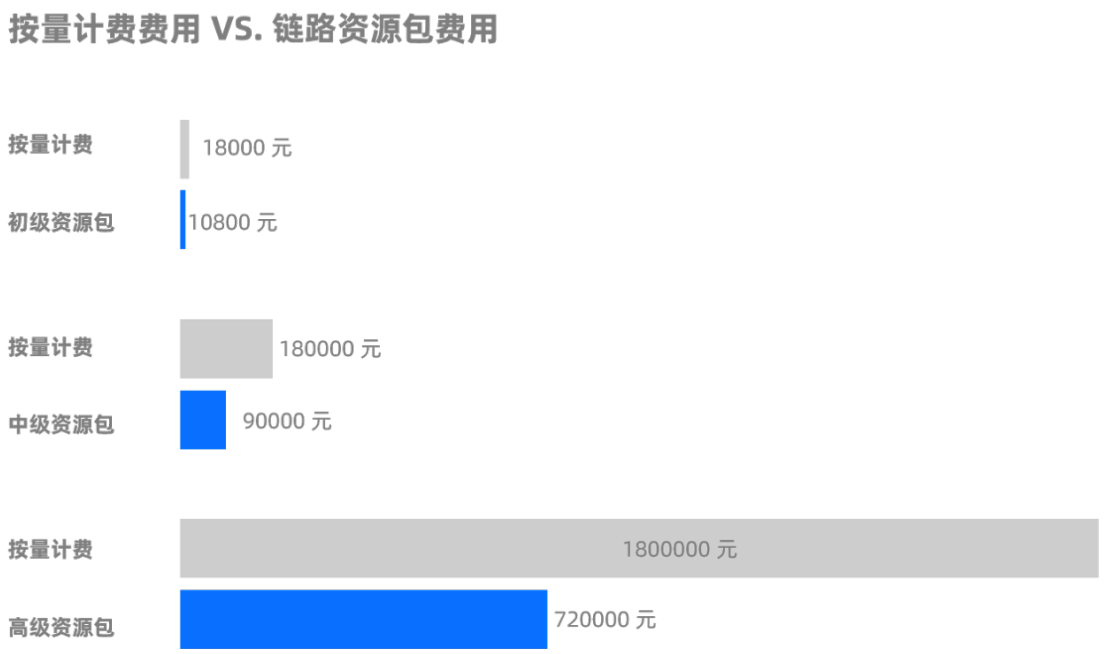
- 所在区域：非金区
- 时长：1 年

计费分析

预付费		按量付费		金额（元）	
资源包	资源包明细（span 数（百万））	计费项	计费项明细（span 数（百万））	资源包抵扣【2】	按量付费
初级链路资源包	总计：36000	分布式链路：span 数	每日数量：100 百万	10800	18000
中级链路资源包	总计：360000		每日数量：1000 百万	90000	180000
高级链路资源包	总计：3600000		每日数量：10000 百万	720000	1800000

说明

资源包抵扣【2】：所示价格为非金区价格。金区价格统一为非金区价格的 2 倍。



10.2.5. 欠费停服说明

欠费停服指服务因账号欠费而自动停止，处于欠费停服状态的服务不计费。本文说明欠费停服相关注意事项。

> 注意
> 当服务欠费后有停止服务的风险，系统会提醒或通知您，请及时续费，避免对您的服务造成影响。

欠费说明

如果您账号的可用额度（含阿里云账号余额、代金券、优惠券等）小于待结算的账单，即视为账号欠费。

停服说明

- 账号欠费后停止计费，并且欠费 24 小时后停止服务，数据默认保留 15 天。
- 续费后不会恢复已清理的数据，并且不会自动恢复服务，您必须在控制台手动开启服务。

10.3. 快速开始

10.3.1. 创建自定义监控

本文主要介绍如何创建自定义监控并进行监控配置。

> 说明
> 过滤多个 LDC 或 IDC 或单机，您需要通过逗号分隔，目前暂时不支持正则填写。

前提条件

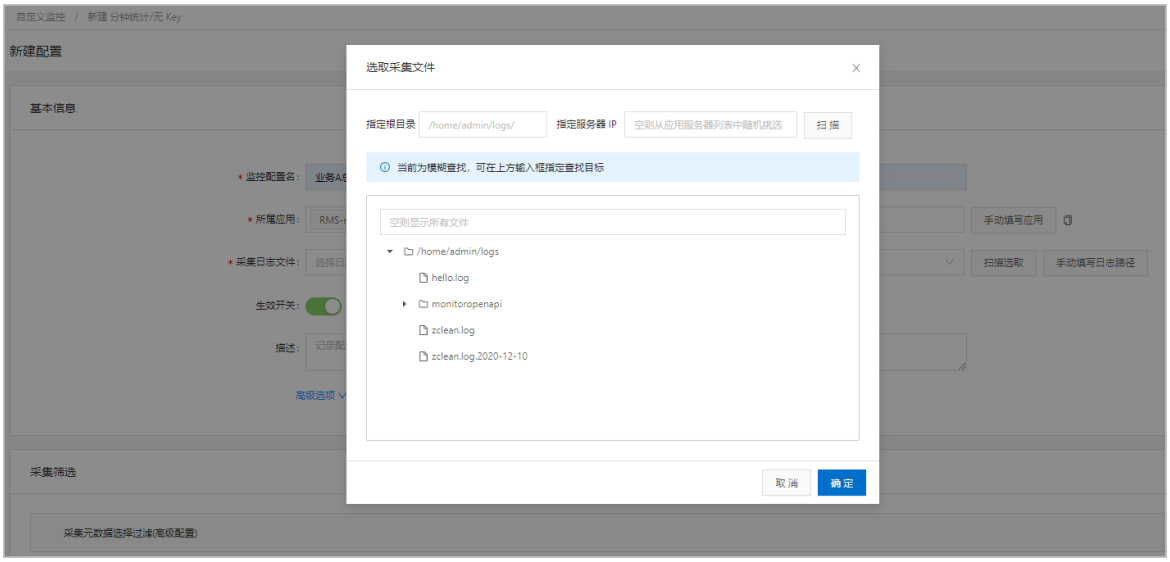
- 通过 CAFE 成功发布应用，并在应用监控中查看到该应用。
- 待创建应用的业务日志符合日志规范，并且具有日志数据。
- 熟悉待创建应用的业务日志的列值字段含义。

操作步骤

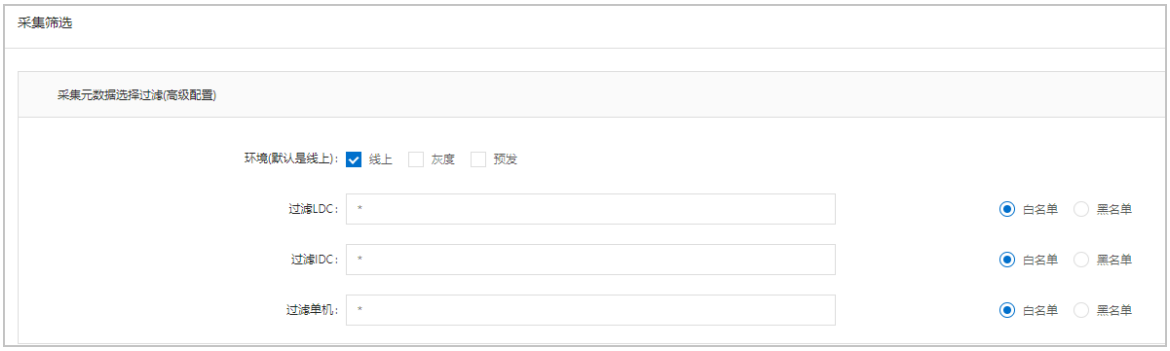
1. 进入控制台，单击左侧导航栏中的 **自定义监控**。
2. 在 **自定义监控** 页面，单击 **新增**。
3. 在 **常用** 页签下，单击 **分钟统计/无 Key**，进入分钟统计/无 Key 监控产品配置页面。



4. 在配置页面，填写 **监控配置名称**，选择所属应用及采集日志文件。
   选择 **采集日志文件** 时，如果列表中没有需要的日志文件，可单击 **扫描选取**，打开 **选取采集文件** 页面，页面会给出二个系统默认的文件路径，双击路径名打开进行日志文件选取。
   如果默认的文件路径中，没有找到需要的路径，可通过填写 **指定根目录** 或 **指定服务器** 来查找需要的日志文件。查找时，注意 **指定根目录** 或 **指定服务器 IP** 这二项只需填写其中一个进行查找。



5. 在采集筛选区域，您可以通过选择 **线上**、**灰度**、**预发环境**，或者通过过滤 LDC 逻辑机房、IDC 物理机房、单机，进行个性化的服务器采集定制。

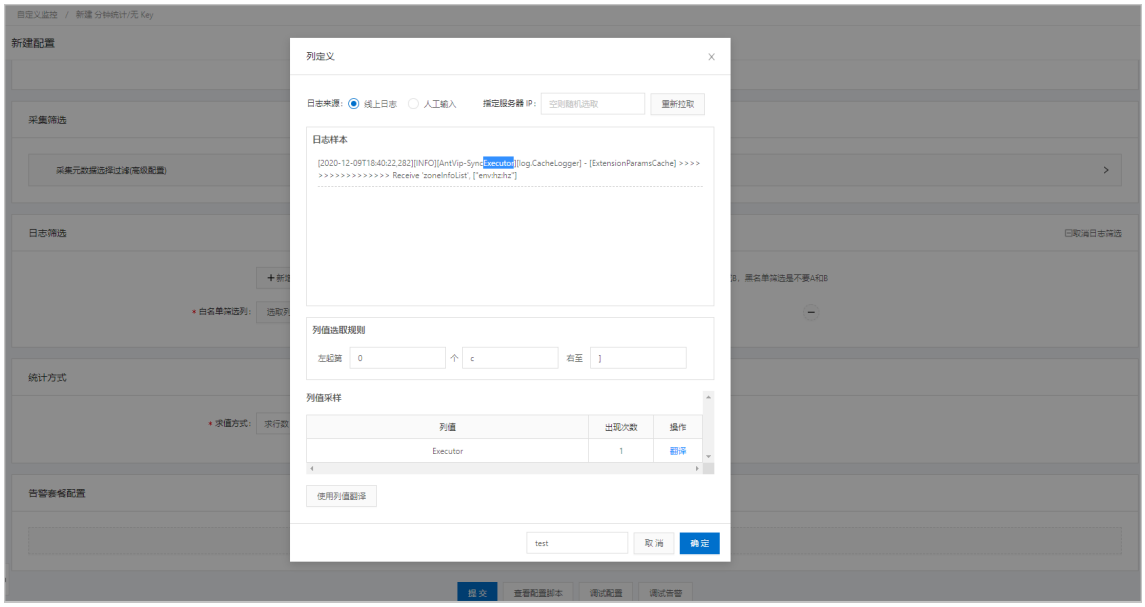


> **说明**
> 过滤多个 LDC 或 IDC 或单机，您需要通过逗号分隔，目前暂时不支持正则填写。

6. 在日志筛选区域，您可以通过配置 **新增白名单筛选** 或 **新增黑名单筛选** 进行日志关键字筛选（白名单相当于 SQL 里的 `[Where Field =?]` 语义，黑名单相当于 SQL 里的 `[Where Field < >?]` 语义）。



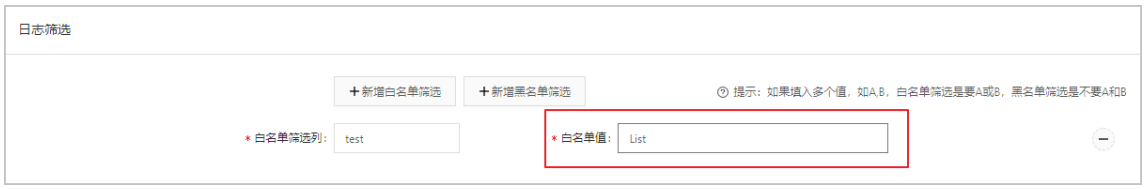
   - **白名单筛选列**
     单击 **新增白名单规则**，新增一条白名单筛选列，在 **列定义** 页面配置“白名单筛选列”位置。您可以在日志样本中选中目标一个关键词，系统会根据该关键词的位置自动生成一个 **列值选取规则**，并在下方的 **列值采样** 表格中显示根据这一选取规则采样得到的列值，完成后，在最下方填入列名，单击 **确定**，完成白名单列的选取。



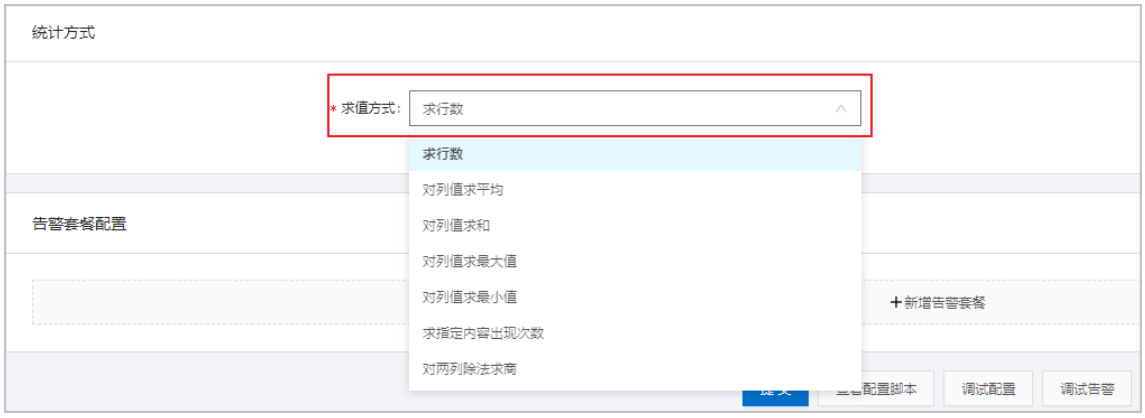
     左起右至分隔符类型如下：
     - 可见符号（推荐使用），如 , . / { } + 等等。
     - 字符串，如：“error”，“错误”等，中英文可见字符串，做为分隔符。
     - 不可见符号（一般不推荐使用），如空格等。一般空格做为分隔符的，建议选中关键字后，自动生成空格分隔。也可以用 ASCII 码，实现隐型符号的分隔，写法如：char(1,2)。

   - **白名单列值**

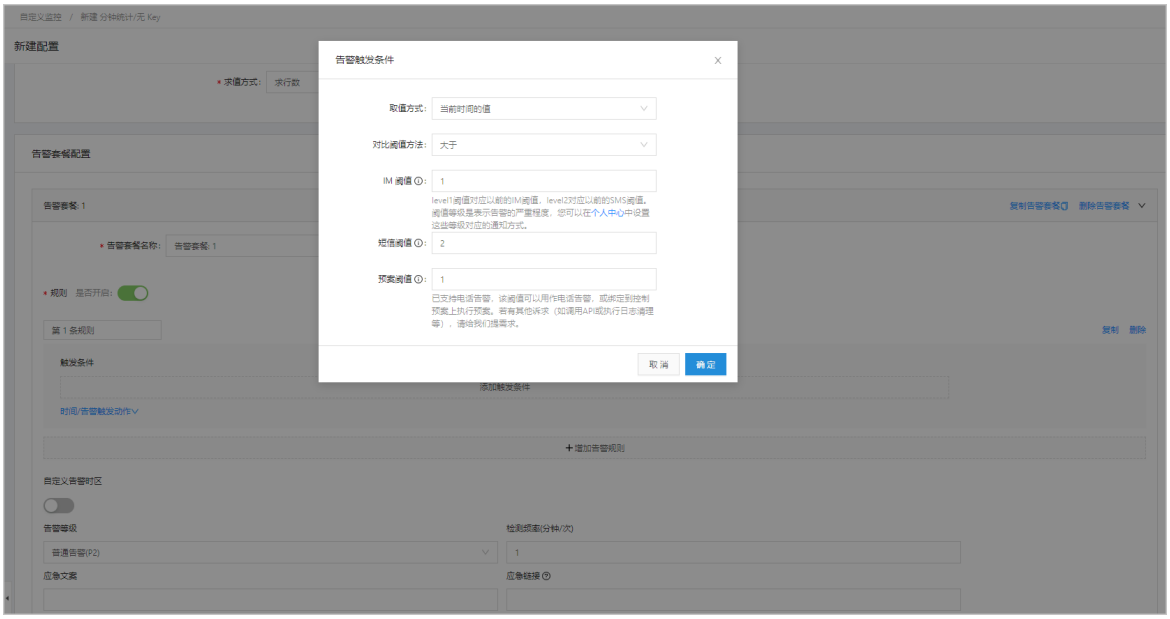
同一个列值位的白名单，支持多个值的选取，可以用“,”分隔，如：`aaaa,bbbb,cccc`。多个“白名单筛选列”之间的关系为“且”。白名单与黑名单为且的关系，多个黑名单之间是或的关系。



7. 在 **统计方式** 区域选择日志进行计算的方式。**分钟统计 / 无 Key** 类型产品提供以下七种方式：
   - 求行数（单位时间内的日志行数）。
   - 对列值求平均（需在 **求平均列** 区域选取列值）。
   - 对列值求和（需在 **求和列** 区域选取列值）。
   - 对列值求最小值（需在 **最小值** 区域选取列值）。
   - 对列值求最大值（需在 **最大值** 区域选取列值）。
   - 求指定内容出现次数 （需在 **有效列值** 区域填写内容）：统计逻辑为是，指定内容一行中出现多次，计多次。
   - 对两列除法求商（需在 **这个数（被除数）**、**除以（除数）** 这两个区域分别选取列值）：统计逻辑是，将 1 分钟内日志二个数值分别求和，然后相除。多用于统计日志，如耗时总量除以总计数。
     文中示例选择 **求行数** 进行计算。

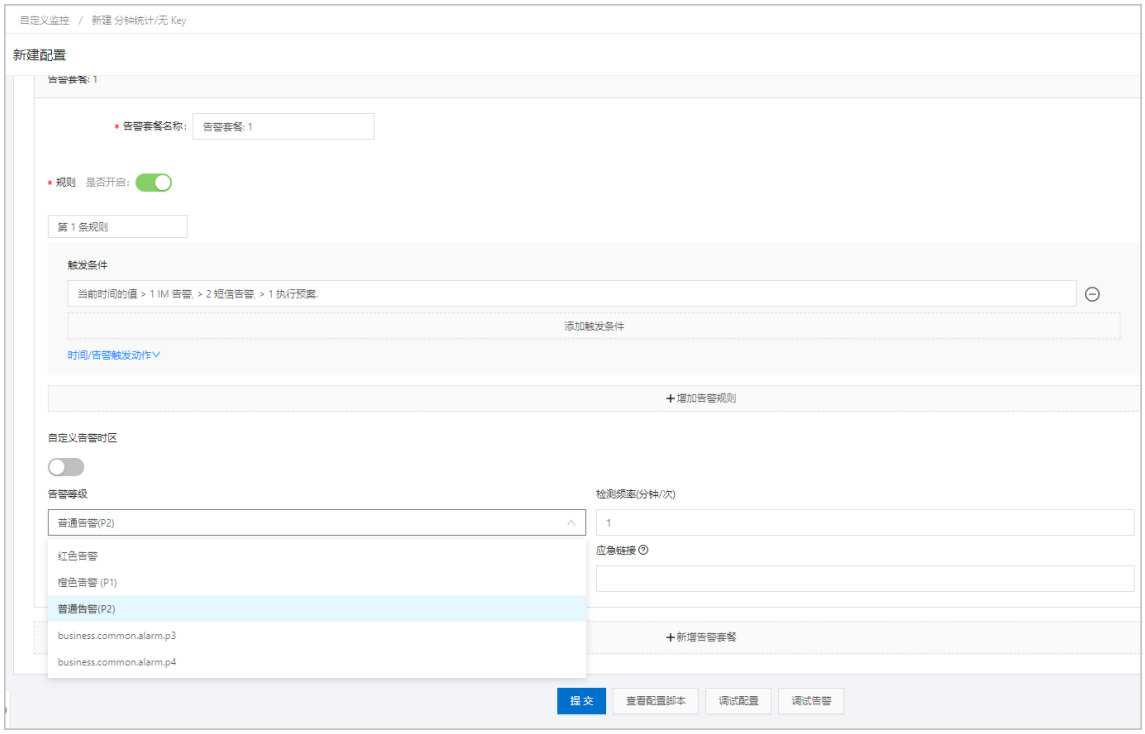


8. 单击 **+ 新增告警套餐**，再单击 **+ 增加告警规则** > **添加触发条件**，在弹出的 **告警触发条件** 配置页面，选取 **取值方式**（指告警规则的计算统计方式）与 **对比阈值方式**，并填写 **短信阈值**，最后单击 **确定** 完成条件配置。



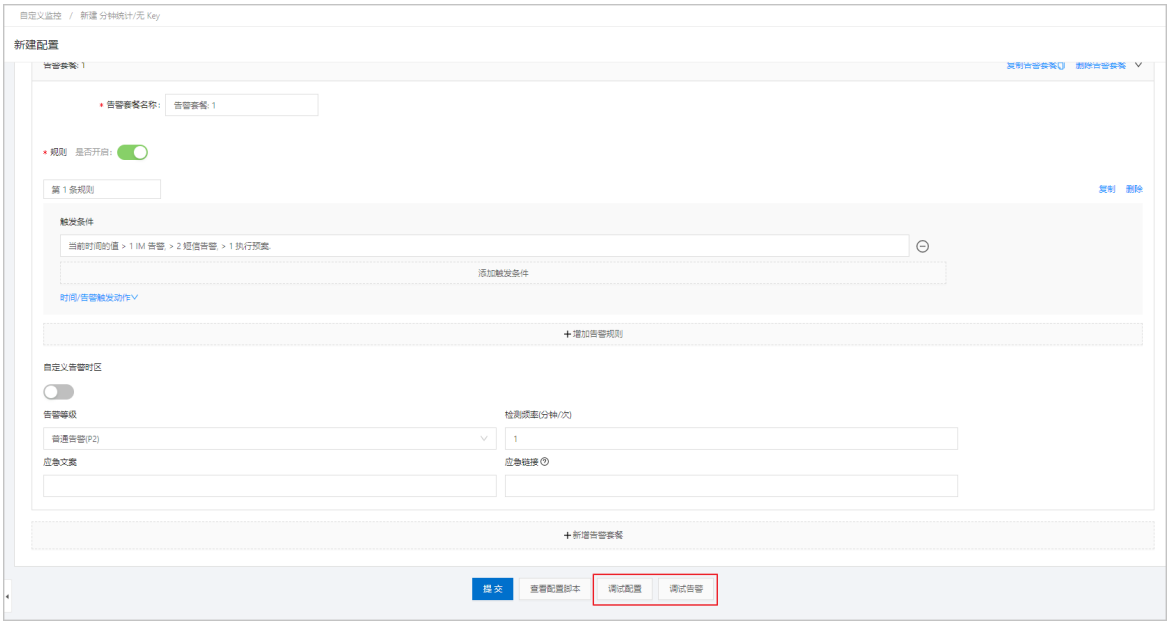
> **说明**
> **告警配置** 指设置一定的告警规则，可通过手机短信方式接收故障告警。RMS 提供可视化告警配置，支持多种告警统计规则、支持多条件多套告警配置、并可按不同时间段设置告警。

- **告警等级**
  告警等级 选项，用于告警的管理，设定不同风险等级，方便监控。**告警等级**会在 **总览**页面的 **告警信息**中做分类统计，以及会在告警接收文案中出现告警指标的等级标注。目前提供红色告警、橙色告警（P1）、普通告警（P2）、普通告警（P3）、普通告警（P4）。文中示例选择默认等级 **普通告警（P2）**。

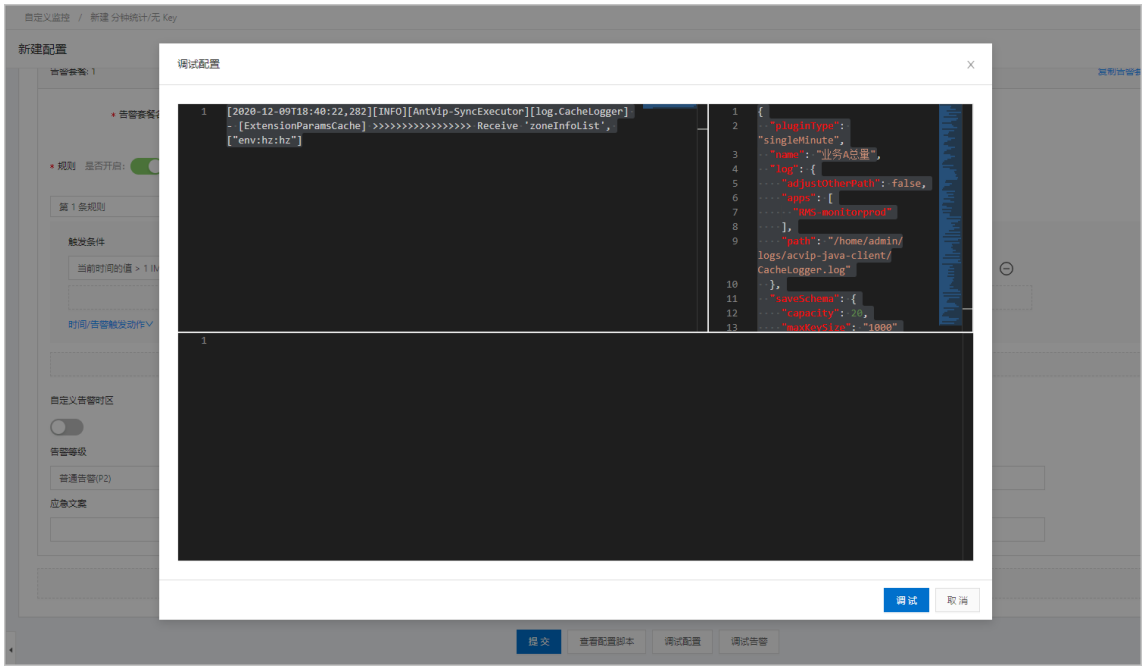


- **检测频率（分钟/次）检测频率（分钟/次）** 选项，指检测告警规则的间隔时长。如：填写为 5，表示规则每 5 分钟启动一次；填写 60，表示规则每个整点启动一次。下图为默认值 1 分钟启动一次。

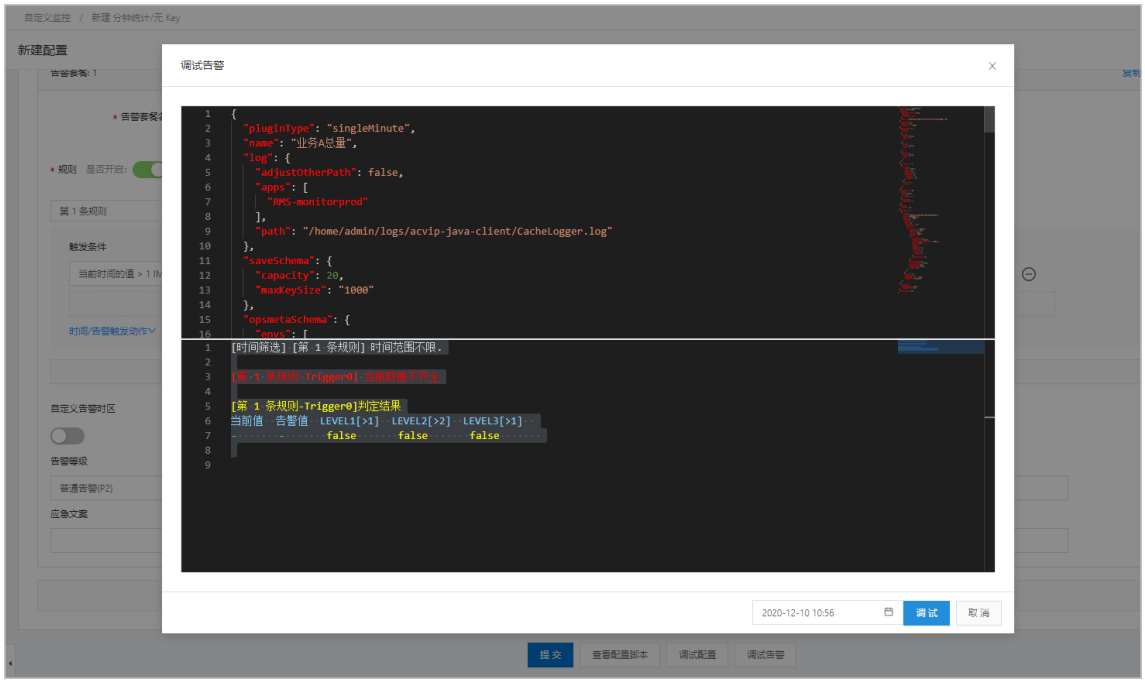
9. 单击监控产品配置页面的最下方 **调试配置** 和 **调试告警**，对配置进行自我检查。



- **调试配置** 对话框打开后，单击 **调试** 进行调试。



- **调试告警** 页面打开后，单击 **调试** 进行调试。



调试完成后，回到配置页面，单击配置页面最下方的 **提交**，便完成了一个自定义监控产品的配置，并生成了一个数据源。

10.3.2. 创建告警规则

本文主要介绍如何创建告警规则以及如何进行告警配置。

前提条件

- 通过 CAFE 成功发布应用，并在应用监控中可查看到该应用。
- 如给自定义监控创建告警规则，请先创建自定义监控。
- 如需通过钉钉、短信等发送告警消息，请提前配置 通知组。
- 秒级无 Key、秒级多 Key 由于计算消耗资源过多，暂不支持创建告警。

自定义监控告警配置

下面针对如下监控类型，介绍自定义监控告警配置的操作链路。

- 分钟统计/无 Key
- 分钟统计/多 Key
- 常用服务指标
- 单笔数据 Top
- 分钟统计 Top

分钟统计/无 Key

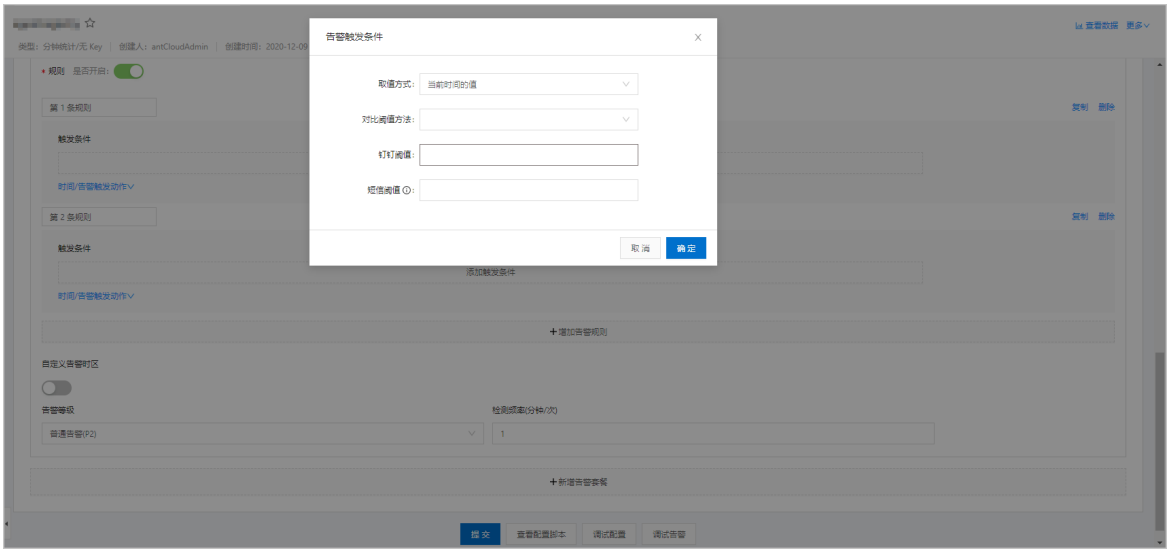
1. 进入控制台，单击左侧导航栏中的 **自定义监控**。
2. 在 **自定义监控** 页，选择 **分钟统计/无 Key** 类型的数据源，在 **操作** 列表中 **编辑**，打开配置页面，在最下方的 **告警配置** 进行配置。



在 **告警套餐配置** 区域进行新增告警套餐：



3. 单击 **增加告警规则**，打开 **告警触发条件** 页面后，选择 **取值方式**、**对比阈值方法**，填写 **IM 阈值**、**短信阈值** 和 **预案阈值**，单击右下方 **确定**，完成告警条件的增加。



> **说明**
>
> - **告警等级**：用于告警的管理，设定不同风险等级，方便监控。目前提供 **红色告警** 、**橙色告警**、**普通告警** 三种等级。文中示例选择默认值 **普通告警**。
> - **检测频率（分钟/次）**：检测告警规则的间隔时长。上图为默认值 1 分钟。
> - **钉钉阈值**：当指标触发该阈值时，会发送告警消息到已配置的钉钉群。
> - **短信阈值**：当指标触发该阈值时，会以短信方式发送告警消息给已配置的用户。



分钟统计/多 Key

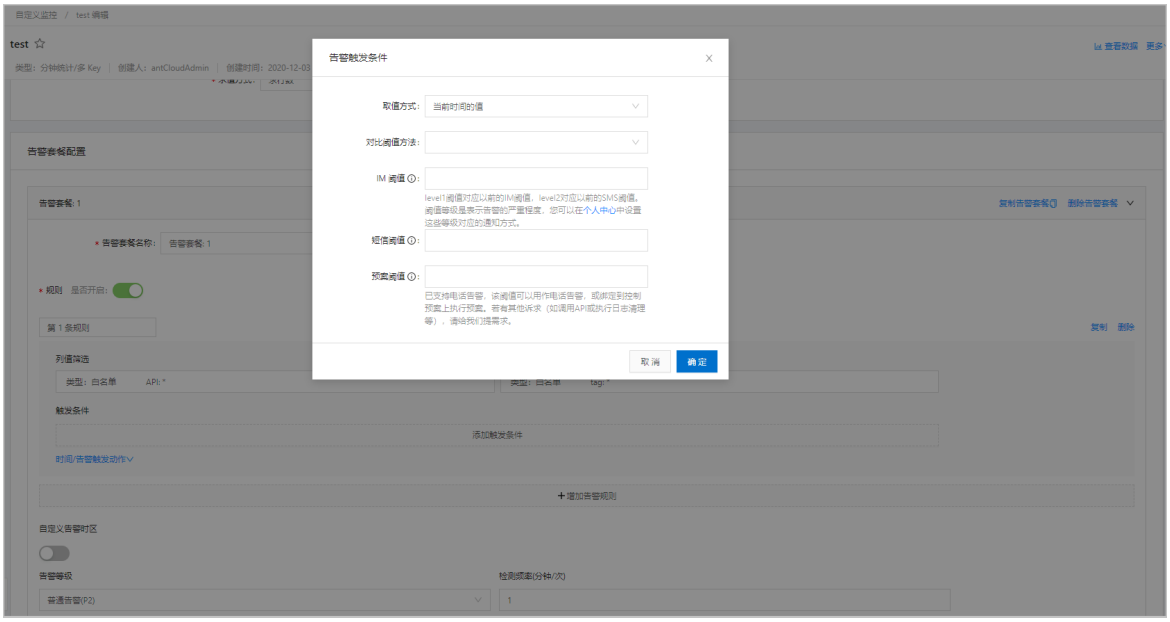
1. 在 **自定义监控** 页，选择 **分钟统计/多 Key** 类型的数据源，在 **操作** 列表中单击 **编辑**，打开配置页面，在最下方的 **告警配置** 进行配置。



在 **告警套餐配置** 区域进行新增告警套餐：



2. 单击 **增加告警规则**，打开 **告警触发条件** 页面后，选择 **取值方式** 与 **对比阈值方法**，填写 **IM 阈值**、**短信阈值** 和 **预案阈值**，单击右下方 **确定**，完成告警条件的增加。



带 **列值分组** 时，可对不同列值进行设置告警规则。所有 Key 需统一规则时，列值筛选填写 `*` 。

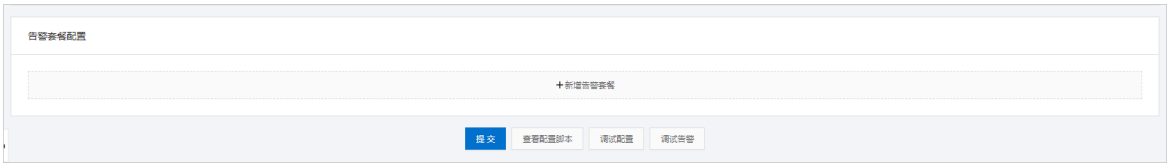


常用服务指标

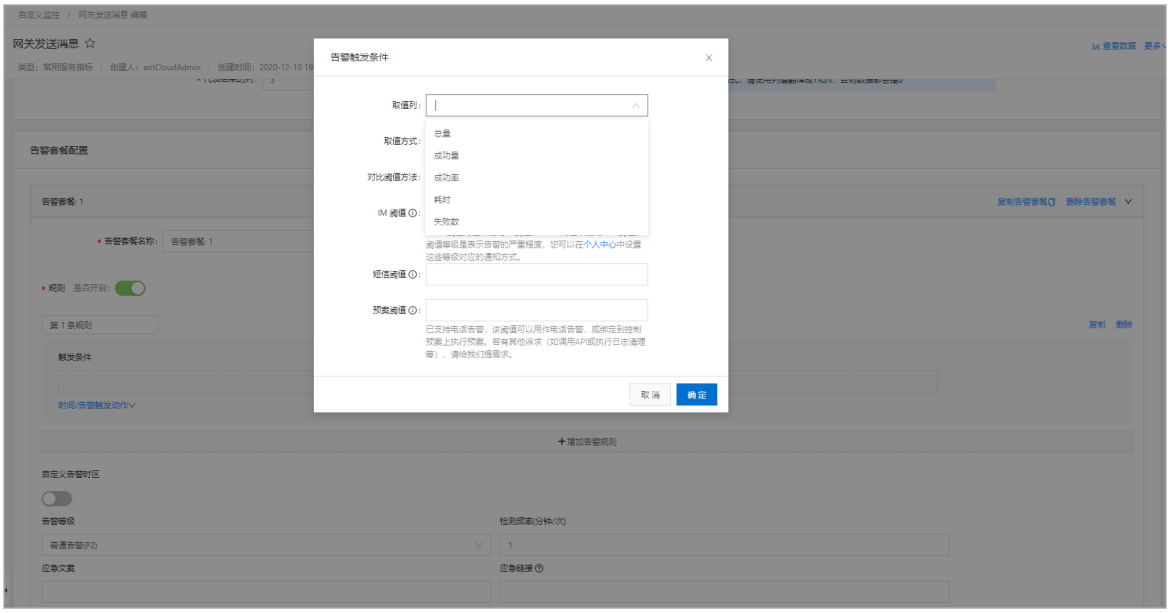
1. 在 **自定义监控** 页面，选择 **常用服务指标** 类型的数据源，在 **操作** 列表中单击 **编辑**，打开配置页面，在最下方的 **告警配置** 进行配置。



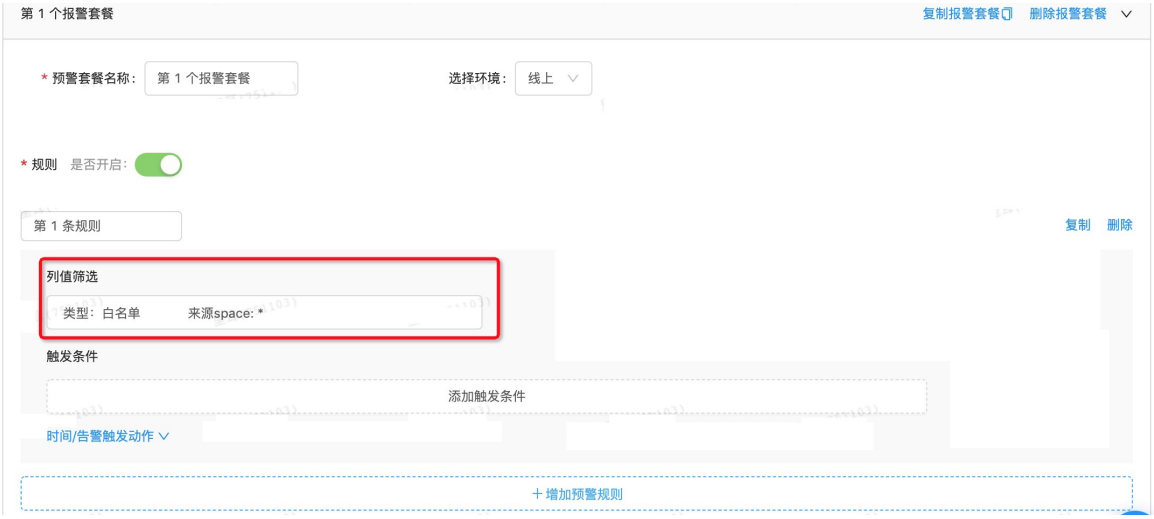
在 **告警套餐配置** 区域进行新增告警套餐：



2. 单击 **增加告警规则**，打开 **告警触发条件** 页面后，选择 **取值列**、**取值方式** 与 **对比阈值方法**，填写 **IM 阈值**、**短信阈值** 和 **预案阈值**，单击右下方 **确定**，完成告警条件的增加：



带 **列值分组** 时，可对不同列值进行设置告警规则。所有 Key 需统一规则时，列值筛选填写 *****。



单笔数据 Top

在 **自定义监控** 页面，选择 **单笔数据 TOP** 类型的数据源，在 **操作** 列表中单击 **编辑**，打开配置页面，在最下方的 **告警配置** 进行配置。以下配置含义为：排序列的值大于 10 的记录超过 3 个时报警。报警内容为日志内容，最多只发送 2 条日志内容。



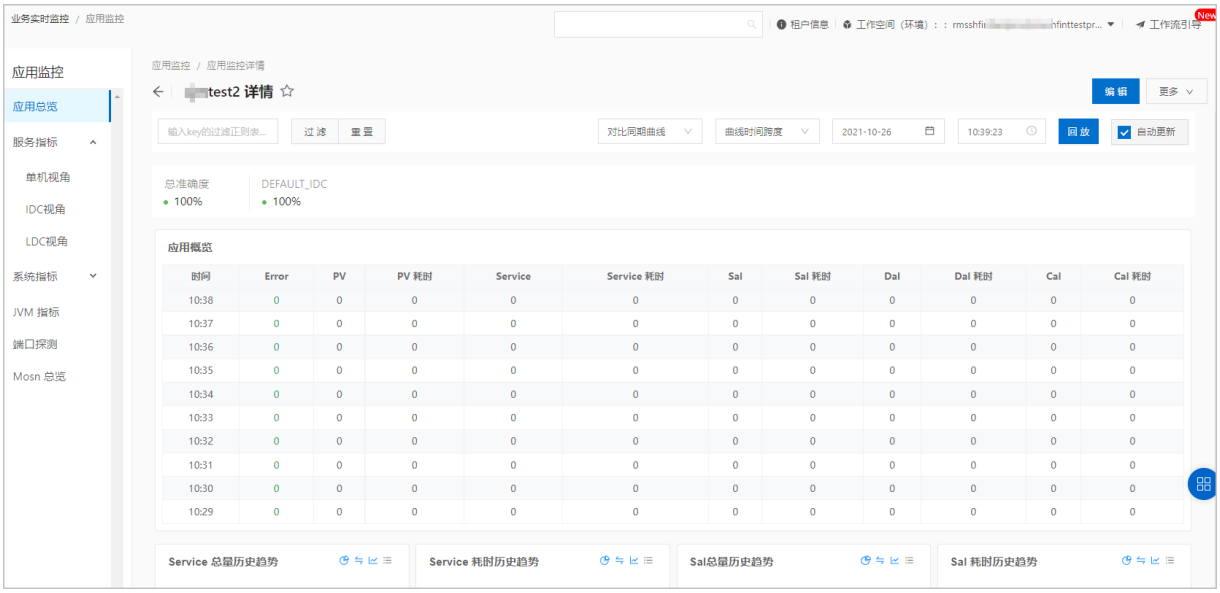
分钟统计 Top

在 **自定义监控** 页面，选择 **分钟统计 Top** 类型的数据源，在 **操作** 列表中单击 **编辑**，打开配置页面，在最下方的 **告警配置** 进行配置。以下配置含义为：列值大于 10 的 Key 的数量超过 3 个时，系统就会报警。报警内容为日志内容，最多只发送 2 条日志内容。

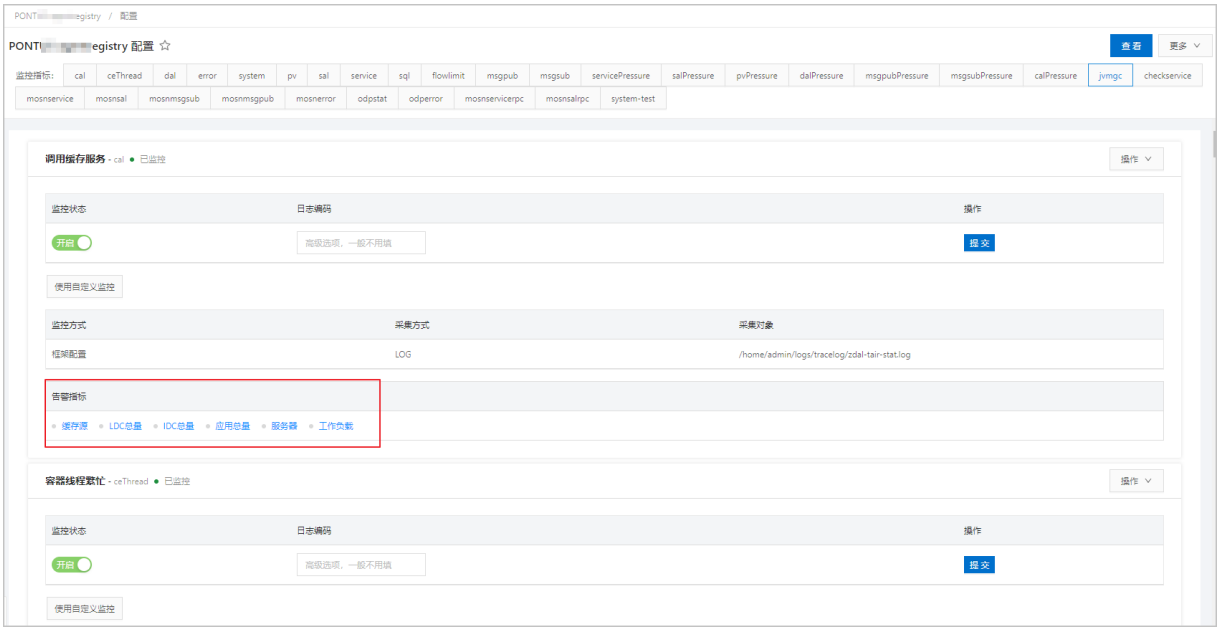


应用监控告警配置

进入监控主界面，搜索相应应用，然后进入应用主界面。



单击 **编辑** 进入应用监控日志及告警配置页面。

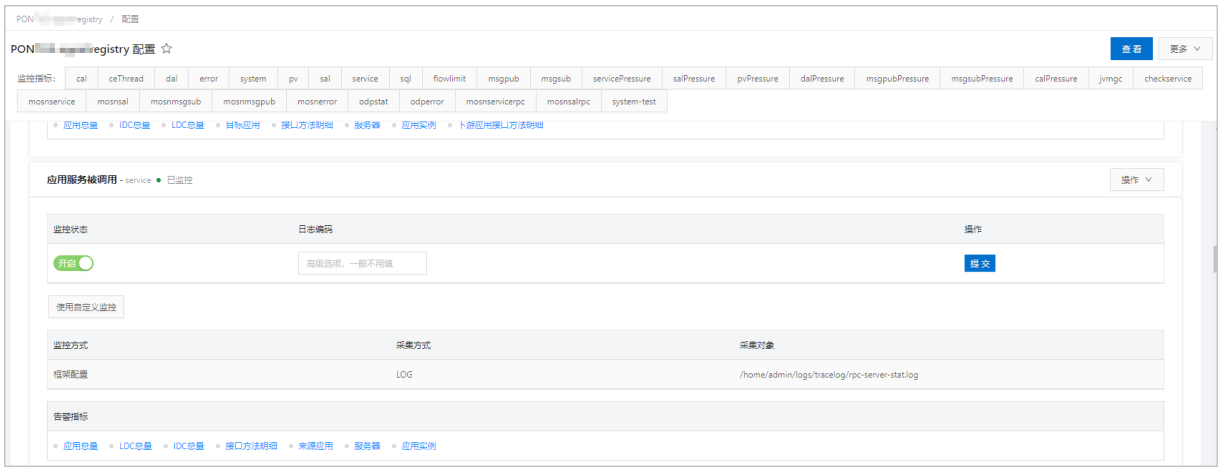


不同的配置方法如下：

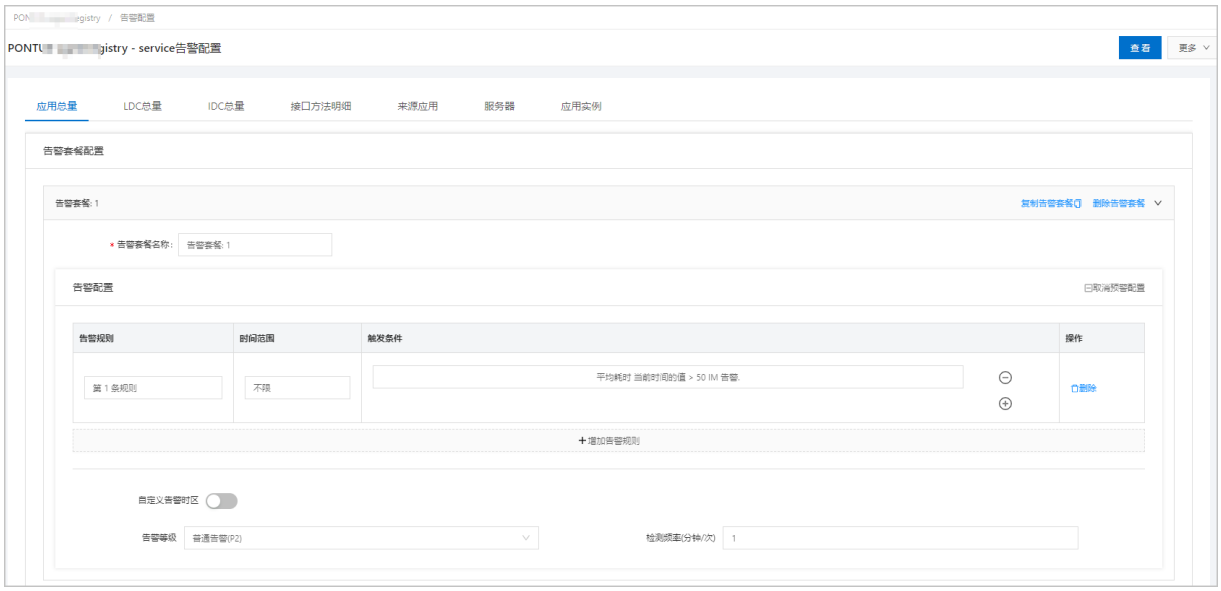
- 应用业务级指标配置
- 系统指标告警配置
- 端口检测告警配置

应用业务级指标配置

可以是 Error 错误、PV、Service、SAL、DAL、CAL。可按 **缓存源**、**LDC 总量**、**IDC 总量**、**应用总量** 等分别设置告警。单击下方的 **告警类型**，进入 **告警配置** 页面。例如：Service 告警配置入口。



单击 **应用总量** 后，进入 **告警配置** 页面，单击触发条件中的 **增加触发条件**，增加触发条件：

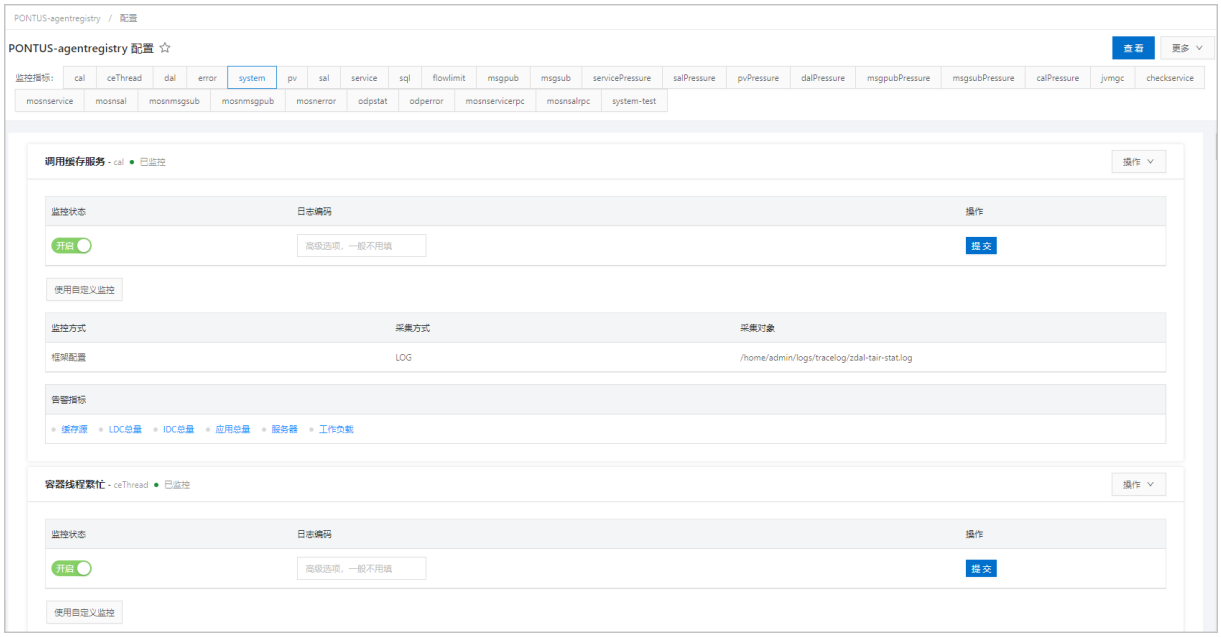


打开 **告警触发条件** 页面后，选择 **取值方式** 与 **对比阈值方法**，填写 **阈值**，单击右下方 **确定**，完成告警条件的增加。



系统指标告警配置

进入应用监控配置界面，单击上方导航栏 **System** 标签。

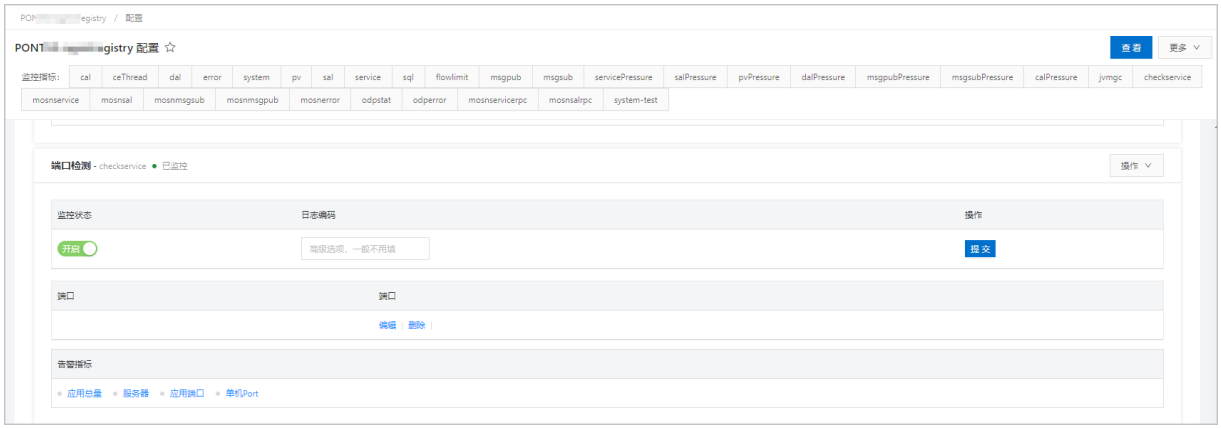


单击 **新增告警规则**，进入 **系统指标告警规则** 页面，选择**告警指标**、**对比阈值方法** 及 **短信阈值**，最后单击 **确定** 完成配置。



端口检测告警配置

进入应用监控配置界面，单击上方导航栏 **Port** 标签。



单击端口列中的 **编辑**，进入端口检测规则页面，配置端口号，最后单击 **确定** 完成配置。



10.3.3. 订阅告警

本文主要介绍订阅告警的方式以及如何取消订阅。

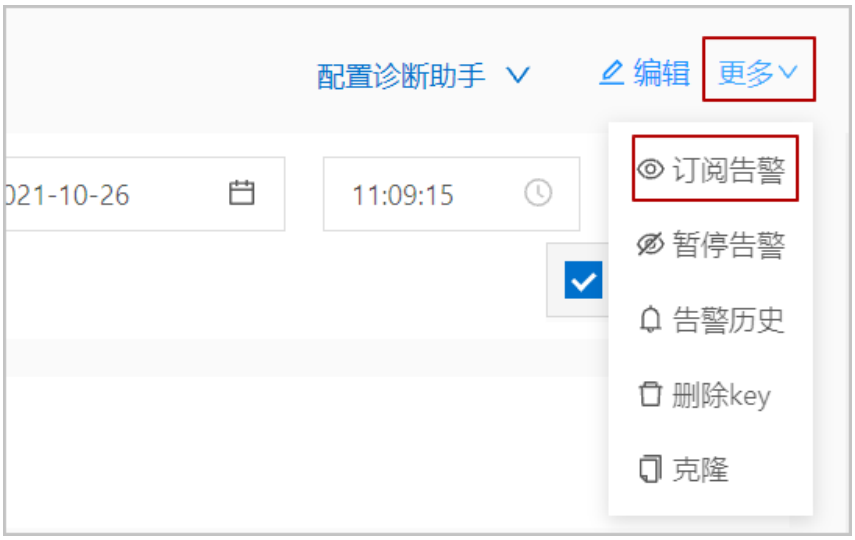
前提条件

- 已创建告警规则。
- 已创建通知组，并将联系人添加到通知组中。

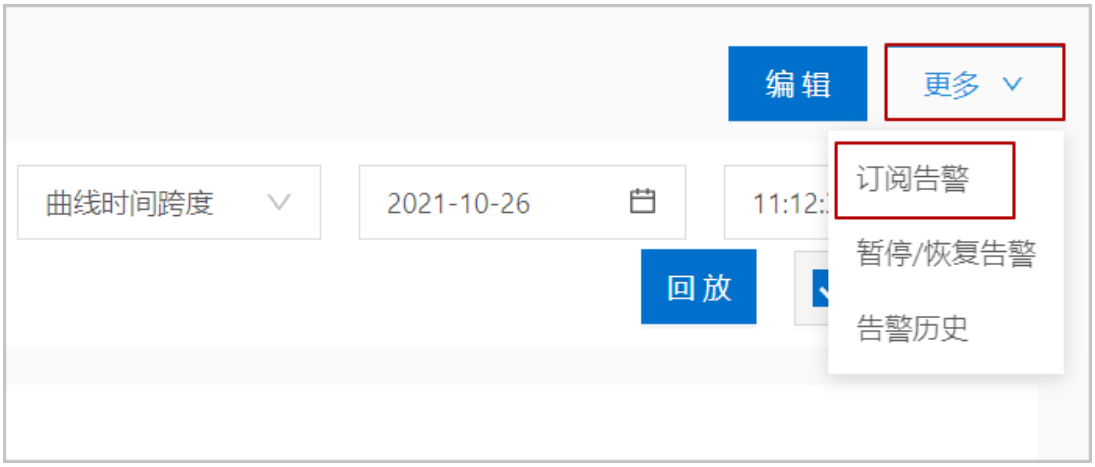
- 已配置钉钉或短信通知方式。

订阅告警入口

- 订阅自定义监控告警
  打开 **自定义监控配置** 页面，单击右上角 **更多** > **订阅告警** 打开订阅页面。



- 订阅应用监控告警
  通过搜索应用订阅告警或者使用搜索框搜索应用，打开某个应用详情页面，单击右上角 **更多** > **订阅告警** 打开订阅页面。



- 按监控文件夹订阅告警
  在文件夹区域，单击 **操作** > **订阅告警** 打开订阅页面。



配置告警信息

打开 **订阅告警** 界面可以添加订阅关系，添加订阅关系需要指定订阅人、订阅的告警等级、以及是否开启通知降频。

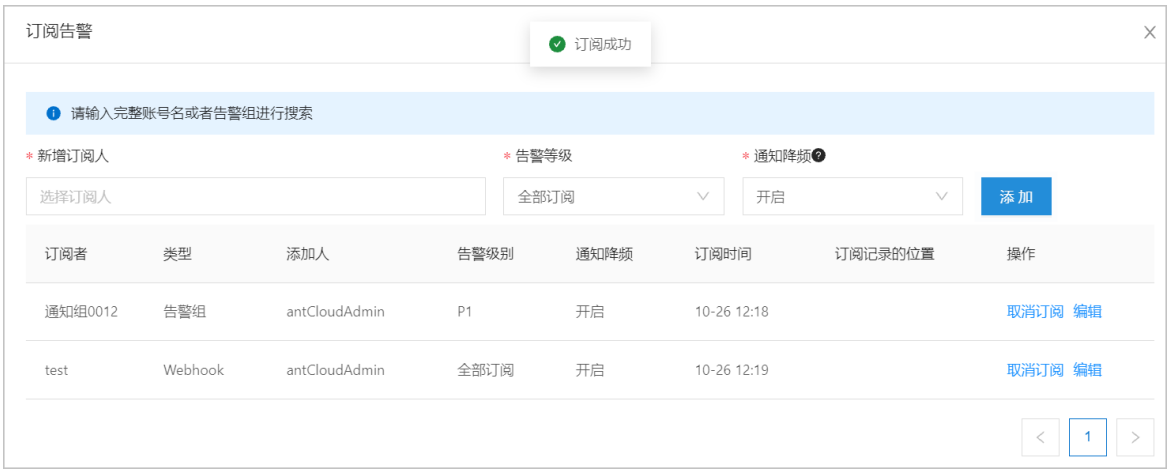
1. 设置订阅人、告警等级、通知降频开关。



- **新增订阅人**：选择对应的订阅人。当前仅支持订阅给告警联系人。如需新建告警联系人，请参见 创建联系人。
- **告警等级**：可以选择不同的告警等级（**全部订阅**、P0、P1、P2、 P3、 P4）进行订阅。
  告警等级 与告警套餐中的 **告警等级** 对应，存在一些特殊的自定义插件或者应用指标在配置告警时无法指定告警等级，此时默认告警等级为 **P2**。
- **通知降频**：可以选择订阅的告警是否进行降频，降频开关打开后会开启告警通知降频策略。
  当配置的规则持续每分钟都触发告警时会触发通知降频策略。

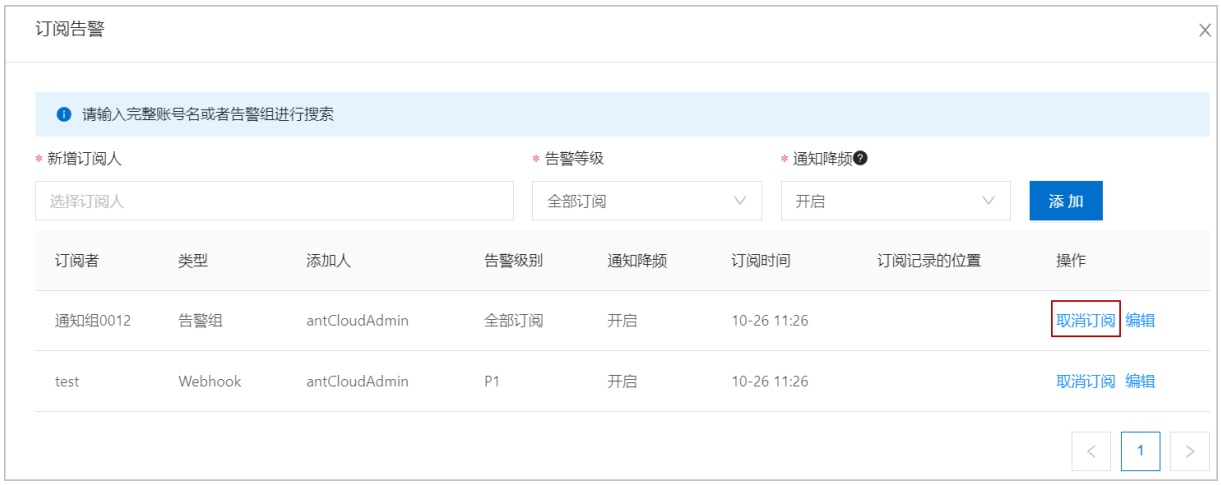
具体降频策略为：每隔 1 分钟通知一次，持续 3 分钟；每隔 2 分钟通知一次，持续 4 分钟；每隔 5 分钟通知一次，持续 25 分钟；每隔 10 分钟通知一次，持续 50 分钟，每隔 30 分钟通知一次，持续 150 分钟；直至每隔 1 个小时通知一次。

2. 单击 **添加** 完成告警订阅。
添加成功的订阅关系会以列表的形式展现。



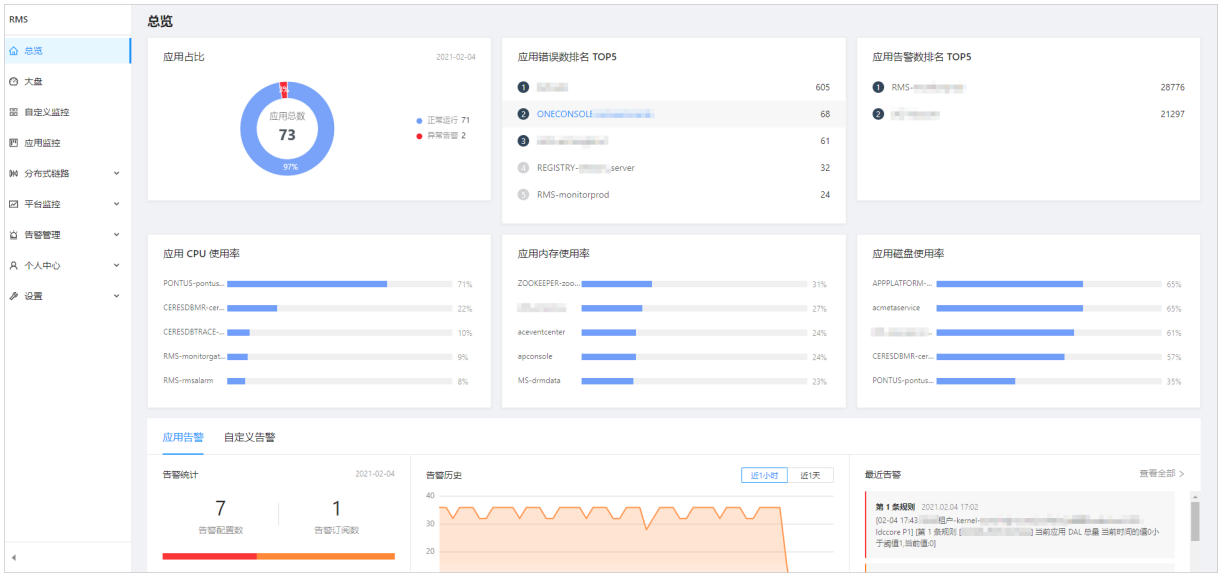
取消订阅

打开 **订阅告警** 界面查看已添加的订阅关系，选择要取消的订阅关系。单击 **操作** 列中的 **取消订阅** > **确认**。



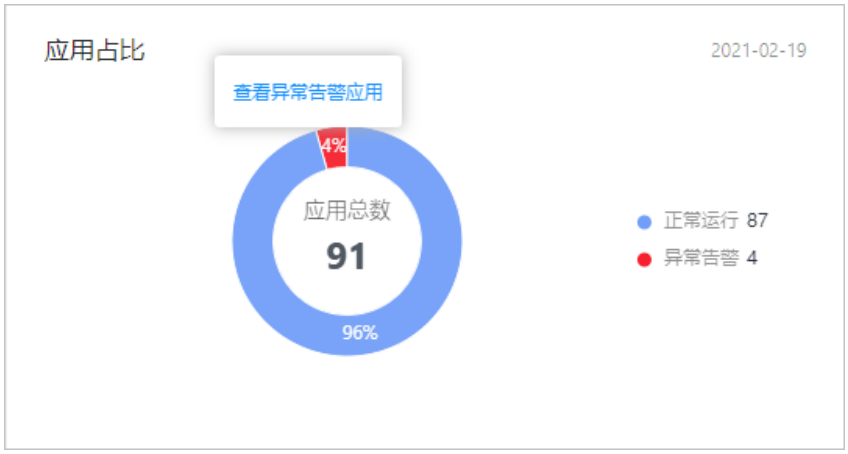
10.4. 总览

总览 页面提供了应用、资源、告警的全链路、多层次、一站式运维界面。分别展示了六种数据信息卡片及应用告警及自定义告警展示区，让运维人员能够快速直观地获知应用及告警情况，进而根据告警信息进行进一步的排查。



应用占比

此卡片主要展示正常运行应用及异常告警应用的占比情况。



- 当前 workspace 下具有实例的应用数
- 正常运行的应用数目
- 异常告警的应用数目
- 正常运行的应用数占应用总数比重
- 异常告警的应用数占应用总数比重
- 每隔一分钟系统自动刷新一次数据

> 说明
> 关于产生异常告警的应用数量的统计，当前仅统计应用关联的应用监控内的告警，不包括自定义监控内告警。

更多操作

查看异常告警应用

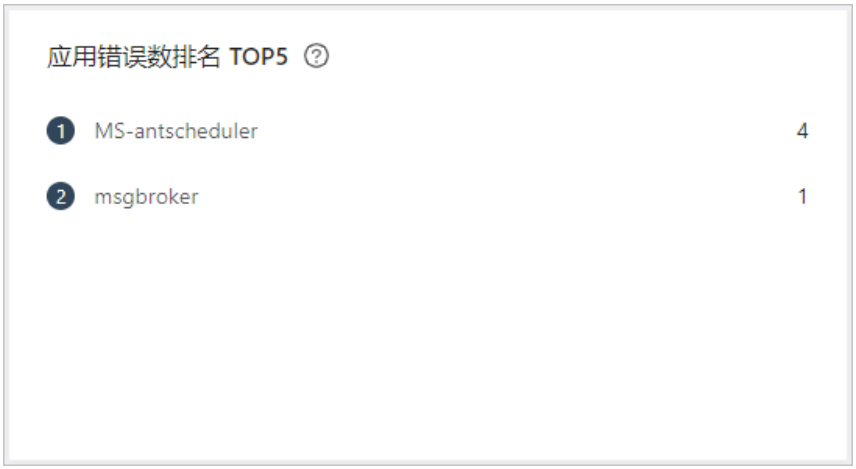
1. 将鼠标光标悬停在应用占比饼图区域上，将弹出 **查看异常告警应用** 提示窗口。
2. 单击 **查看异常告警应用**，将打开 **异常告警应用** 页。
3. 在 **异常告警应用** 页，可进行如下操作：

- 查看触发告警的应用名称及应用对应的告警数量。
- 单击应用名称，即可跳转至应用监控的 **告警历史** 页。

应用错误数排名

此卡片主要展示当前租户下应用错误数前 5 名的排名情况。

> 说明
> 应用错误数是指应用相关的 Error 服务指标的统计数据。



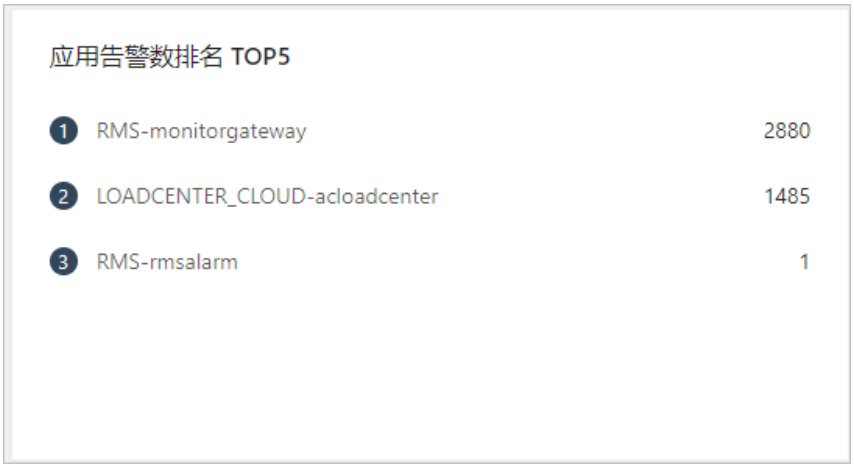
- 错误数排名前 5 的应用名称
- 应用对应的错误数量
- 每隔一分钟系统自动刷新一次数据

更多操作

单击应用名称，即可跳转到 **应用监控详情** 页。

应用告警排名

此卡片主要展示当前租户下应用告警数前 5 名的排名情况。



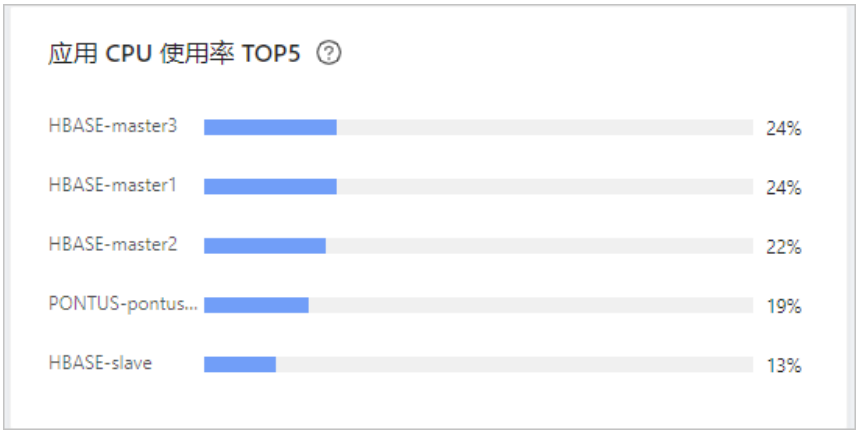
- 告警数排名前 5 的应用名称
- 应用对应的告警数量
- 每隔一分钟系统自动刷新一次数据

更多操作

单击应用名称，即可跳转到应用监控的 **告警历史** 页。

应用 CPU 使用率

此卡片主要展示当前租户下 CPU 使用率前 5 名的应用排名情况。



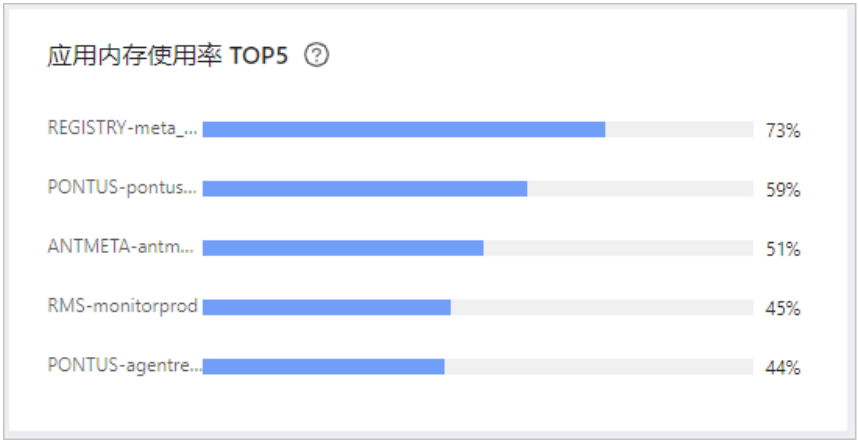
- CPU 使用率排名前 5 的应用名称
- 应用对应的 CPU 使用率
- 每隔一分钟系统自动刷新一次数据

更多操作

单击应用名称，即可跳转到应用监控 **系统指标** 页。

应用内存使用率

此卡片主要展示当前租户下内存使用率前 5 名的应用排名情况。



- 内存使用率排名前 5 的应用名称
- 应用对应的内存使用率
- 每隔一分钟系统自动刷新一次数据

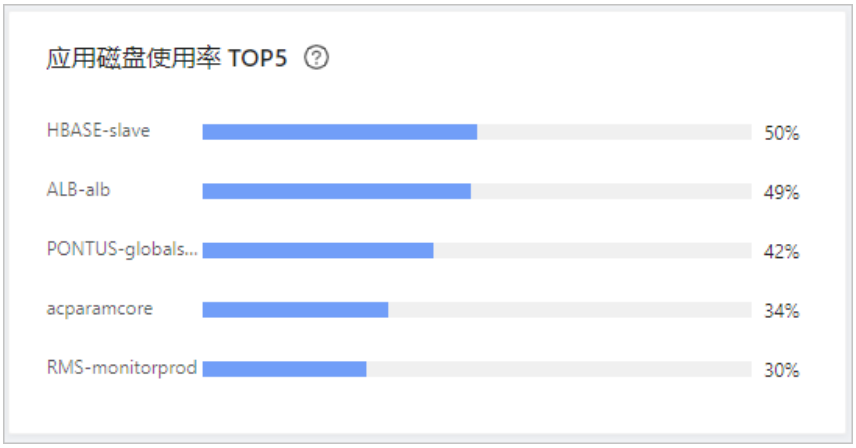
更多操作

单击应用名称，即可跳转到应用监控 **系统指标** 页。

应用磁盘使用率

此卡片主要展示当前租户下磁盘使用率前 5 名的应用排名情况。

> 说明
> 部分站点由于 Docker 版本差异，导致容器磁盘指标无法获取。



- 磁盘使用率排名前 5 的应用名称
- 应用对应的磁盘使用率
- 每隔一分钟系统自动刷新一次数据

更多操作

单击应用名称，即可跳转到应用监控 **系统指标** 页。

告警展示区

该区域主要展示 **应用告警** 或 **自定义告警** 的告警相关统计信息，每隔一分钟系统会自动刷新一次数据。

- 告警统计
- 告警历史
- 最近告警

告警统计

该区域主要展现当前租户下近一天内应用告警或自定义告警的告警统计情况。

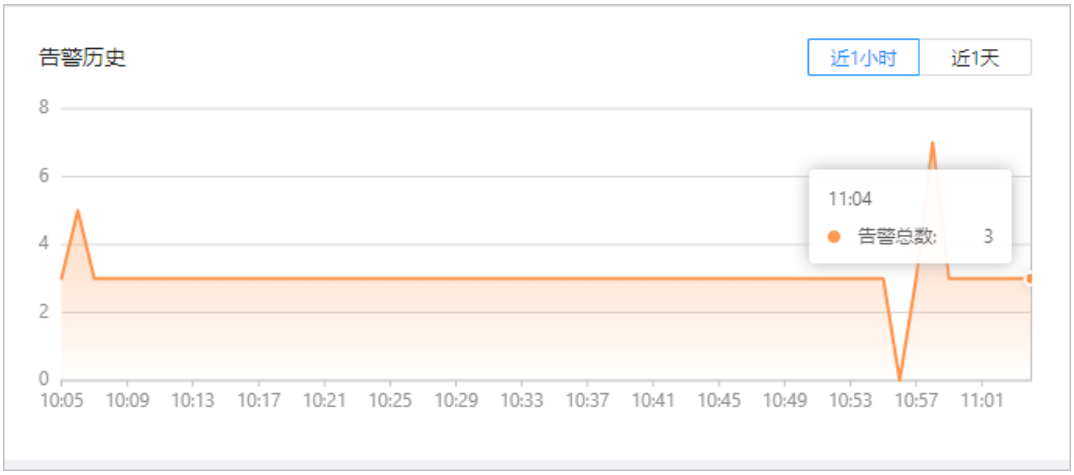


- **告警配置数**：应用告警或自定义告警的配置总数。
- **告警订阅数**：应用告警或自定义告警的订阅总数。
- **红色告警**：应用告警或自定义告警的红色告警总数。
- **橙色告警**：应用告警或自定义告警的橙色告警总数。
- **普通告警**：应用告警或自定义告警的普通告警总数。

告警历史

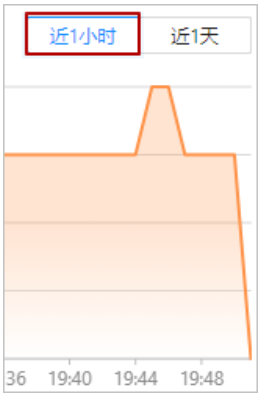
该区域主要展现应用告警或自定义告警的告警总数分布情况。

- 横轴代表各时间点。
- 纵轴代表应用告警或自定义告警的告警总数。
- 趋势图最右侧的时间点对应的纵轴值即最新时间点应用告警或自定义告警的告警总数。



更多操作

- 单击 **近 1 小时**，可查看近 1 小时告警的趋势图。趋势图每 1 分钟一个点，趋势图每一个点的值表示 1 分钟内应用告警或自定义告警的告警总数。



- 单击 **近 1 天**，可查看近 1 天告警的趋势图。趋势图每 1 小时一个点，趋势图每一个点的值表示 1 小时内应用告警或自定义告警的告警总数。
- 单击趋势图的某一个时间点，即可打开 **告警历史详情** 页。



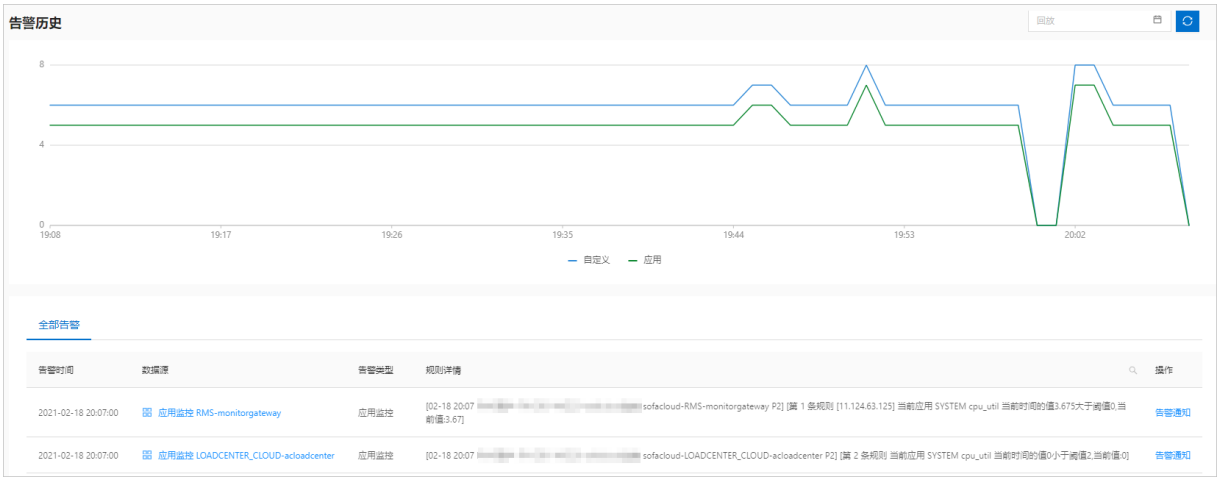
最近告警

该区域主要展示最近触发的前五条告警信息，如告警等级、告警时间、告警信息等。



更多操作

单击 **查看全部**，跳转到 **告警历史** 页。



10.5. 大盘

大盘是将已配置好的数据源，集中在一个页面上以报表形式呈现。根据具体的监控需求，通过对大盘模板的个性化编辑，即可实现个性化监控的可视化展示。

关键大盘 和 **我的大盘** 都是从 **大盘** 模板中进行创建的，您可以根据监控目的进行个性化配置。



- 单击大盘标签右侧的操作选项 **加入关键大盘**，将该大盘加入 **关键大盘**。
- 单击大盘标签右侧的操作选项 **加入收藏**，将该大盘加入 **我的大盘**。

> 说明
> 在自定义监控页面，单击 **大盘** 类型的监控右侧选项，选择 **收藏** 后，会添加到 **我的大盘**，其它监控产品选择 **收藏** 后，会添加到 **个人中心** 的 **收藏夹** 中。

10.6. 自定义监控

10.6.1. 自定义监控概述

自定义监控中新增的监控产品，实际上是提供了对应的监控产品模板，通过模板的配置，可以实现对自定义监控指标的可视化呈现。

自定义监控支持对业务日志的格式进行灵活切割，并按维度聚合数据。基于应用业务日志提供一套完整的业务实时分析及预警能力，使用监控像使用 SQL 一样简单。主要使用场景是个性化的、非标的业务监控，比如交易量等指标。

将一个业务日志，进行筛选指定列值作为指标，并进行采集展示告警的流程如下：



举个例子，有业务日志样例如下，希望通过监控获取每一分钟各商品各自创建了多少笔交易这一数据。

```
2012-11-11 11:11:11,2950211004,衣服,交易,创建,7,Y,
2012-11-11 11:11:12,2950211005,衣服,交易,创建,8,Y,
2012-11-11 11:11:13,2950211006,食品,交易,创建,9,Y,
2012-11-11 11:11:14,2950211007,家居,交易,创建,11,Y,
2012-11-11 11:11:15,2950211008,食品,交易,创建,7,N,
……
```

日志格式说明如下：

- 以 `,` 为分隔符，将关键字段进行切分，实际语义为：

  `日志打印时间,订单id,商品,服务名,方法名,耗时,结果` 。

- **数据统计**：由于监控系统面对的是海量数据，在获取结构化数据之后，需要对原始数据进行数据统计。统计方式可以有很多种，比如 **求行数、求和、求平均、最大/最小值** 等统计方法。
- **预警配置**：可在统计数据上定义类型丰富的预警规则，包括 **绝对值、同比、环比、最近 N 分钟求和** 等，支持定义多项规则和逻辑运算。

10.6.2. 监控类型概述

自定义监控提供数据统计的监控产品类型，您可以配置个性化的监控项。本文介绍监控产品的类型及部分应用示例。

监控类型

监控产品的类型以统计指标为依据进行划分，主要分为常用、高级类型。具体可参考下表。

名称	类别	应用场景	备注
分钟统计/无 Key	常用	统计单业务量每分钟的数据。 通过筛选日志固定位置的关键字，统计单个指标分钟级的数据，如统计某个接口在一分钟内被调用的总量等。	支持日志关键字筛选
分钟统计/多 Key	常用	统计 N 个维度的 N 个 Key 业务量每分钟的数据，如统计系统各个接口分别被调用的总量。 通过配置日志的 Group By 维度，统计各维度各 Key 数据，等同于 SQL 中 Group By 的功能。	支持日志关键字筛选
大盘	常用	将监控项通过灵活的配置，以报表及走势图形式展现。	只做报表展现，不支持告警配置。所有告警设定在配置监控项时完成。

名称	类别	应用场景	备注
常用服务指标	常用	统计单业务或 N 个维度的 N 个 Key 业务的总量、成功量、失败量、成功率、平均耗时 5 个统计项。	支持日志关键字筛选
分钟统计 Top	常用	统计维度的多个 Key 的每分钟的数据，并对数值排序，展现前 N 个最高数值的 Key，如统计平均耗时前 10 的错误码。通过计算多维度数值在每分钟内的大小，再将数值进行排序，展现前 N 个 Key 及其数值。	-
单笔数据 Top	常用	对单条日志的数值维度进行排序，查看最高 N 个单笔业务，如统计耗时前 10 的单笔订单。	-
匹配查找告警	高级	统计日志中固定位置的关键字出现次数，达到一定量后进行告警。 例如：日志中，打印日期后出现 `Error` 关键字就发送告警。	• 只有告警功能，无数据展现。 • 告警内容为符合告警条件的日志内容。 • 可自定义告警文案。
归档统计	高级	对一个秒级数据源，按小时、天、周、月统计数值。	只做统计，不做告警使用。

应用示例

示例一

假设存在下述日志：

```
2012-11-1111:11:11粤A123XX,广州北收费站,G25,广州,￥50
2012-11-1111:11:12粤A123XX,广州北收费站,G26,广州,￥50
2012-11-1111:11:13粤A123XX,广州北收费站,G27,广州,￥50
2012-11-1111:11:14粤A123XX,广州北收费站,G27,广州,￥50
2012-11-1111:11:15粤A123XX,广州北收费站,G26,广州,￥50
……
```

- **分钟统计/无 Key**：要从上面这批日志里了解在 2012-11-11 11:11 这一分钟，广州北收费站共计通过多少车辆，使用分钟统计/无 Key，通过“广州北收费站”关键字筛选，统计日志行数，就可以可统计出广州北收费站的分钟级数据。
- **分钟统计/多 Key**：要从上面这批日志里了解在 2012-11-11 11:11 这一分钟，每个收费站各通行了多少辆车，使用分钟统计/多 Key 对收费站这个维度进行 Group By 统计，就可以统计出每个收费站的分钟级数据。
- **分钟统计 Top**：在统计收费站车辆通行数量的场景里，全中国有几十万个收费站，我们并不需要每个收费站、每分钟的历史数据（因为绝大部分时间里大部分的收费站都是空闲的，数据都是 0，全部统计记录会浪费很多成本）。假设某交警部门想要在节假日期间分析出当前流量最大的收费站，以便做出交通警力调配。这种场景可以使用分钟统计 Top 监控功能，在上例里只需要万分之一的成本，就能实现此需求。
- **匹配查找告警**：要从上面这批日志里排查车牌为“粤 A12345”这辆车的去向，只要它出现在任何一个收费站就报警。使用匹配查找告警，通过车牌“粤 A12345”关键字筛选，就能对这辆车进行监控。

示例二

假设存在下述日志：

```
2012-11-1111:11:11粤 A123XX,广州北收费站,G25,广州,￥50,2340ms,Y
2012-11-1111:11:12粤 A123XX,广州北收费站,G26,广州,￥50,3209ms,Y
2012-11-1111:11:13粤 A123XX,广州北收费站,G27,广州,￥50,4200ms,Y
2012-11-1111:11:14粤 A123XX,广州北收费站,G27,广州,￥50,5700ms,Y
2012-11-1111:11:15粤 A123XX,广州北收费站,G26,广州,￥50,8500ms,Y
……
```

常用服务指标：要从日志中统计每分钟每个收费站收费多少车辆、成功收费多少车辆、多少车辆收费失败、平均收费时长多少，使用常用服务指标对收费站这个维度进行 Group By，将收费时长维度配成耗时，再配置结果维度，即可统计出每个收费站收费 **总量、成功量、失败量、成功率、收费时长** 等的分钟数据。

示例三

假设存在下述日志：

```
2012-11-1111:11:11粤A123XX,广州北收费站,G25,广州,￥50,收费员-张三
2012-11-1111:11:12粤A123XX,广州北收费站,G26,广州,￥50,收费员-李四
2012-11-1111:11:13粤A123XX,广州北收费站,G27,广州,￥50,收费员-王五
2012-11-1111:11:14粤A123XX,广州北收费站,G27,广州,￥50,收费员-李四
2012-11-1111:11:15粤A123XX,广州北收费站,G26,广州,￥50,收费员-张三
……
```

单笔数据 Top：假设某交警部门正在调查乱收费现象，想知道每分钟里收费最高的前 100 笔记录。在这前 100 笔记录中，收费金额超过 ￥500 就要介入调查。在这种场景下，不要使用 Group By 对收费员维度进行统计，因为这样会产生几十万、几百万的无效数据长期占用资源。使用单笔数据 Top 不仅能用极低的成本找到前 100 笔记录，而且能查询到每一笔记录的详情。

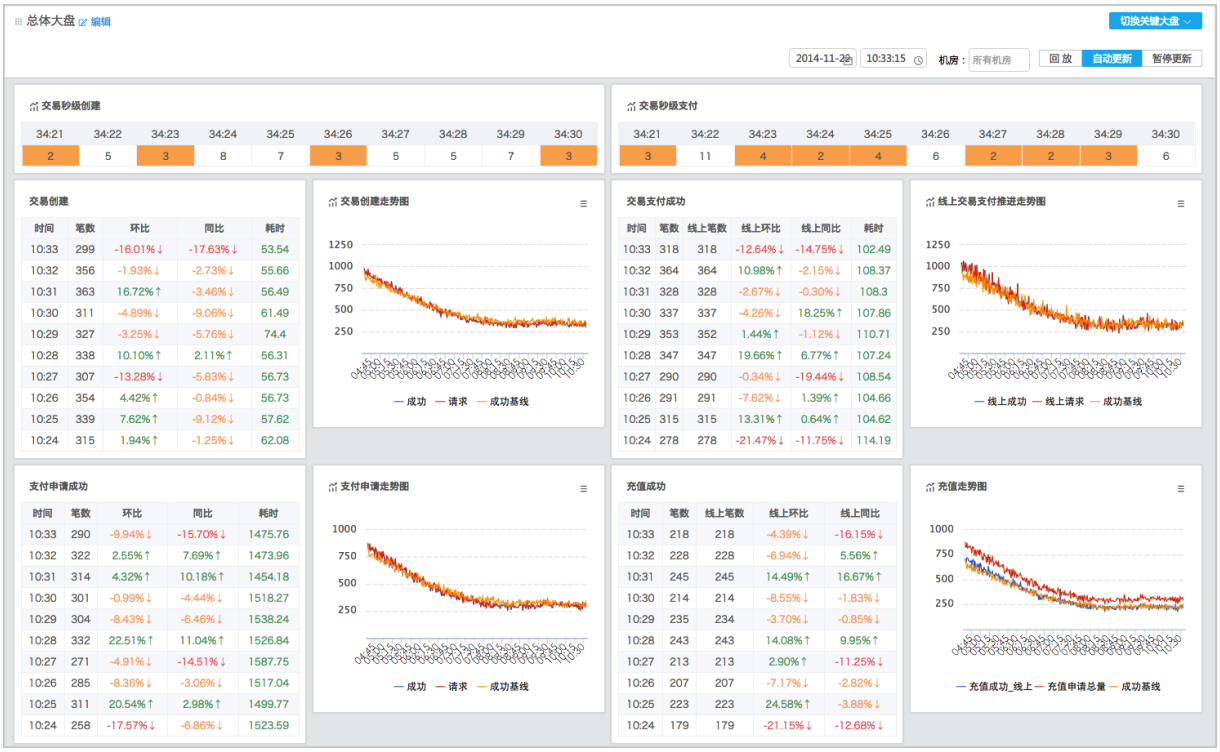
10.6.3. 创建大盘

大盘是将已配置好的数据源，集中在一个页面上以报表形式呈现。大盘只用于数据展现，没有预警功能。配置大盘时，可根据业务需要、数据源类型，选择不同的报表类型。

目前，实时监控支持以下报表类型，具体配置方法，详见各报表类型说明。

- 最近 10 分钟表格
- 趋势图
- 当前分钟多 Key 竖铺表格
- 多分钟多 Key 竖铺表格
- 最近 10 分钟多 Key 表格
- 秒级表格
- 秒级多 Key 表格
- Key 横竖 MAP 表格（高级）
- 常用业务表格
- 应用错误 Top
- Top 数据源系列表格
- 应用监控相关报表
- 数据归档报表

下文介绍如何通过自定义监控配置大盘。下图是一个业务的 **总体大盘** 示例。



操作步骤

1. 进入控制台，单击左侧导航栏中的 **自定义监控**。
2. 在 **自定义监控** 页面，单击 **新增**。
3. 在 **常用** 标签页下，单击 **大盘**，进入大盘的配置页面。



4. 输入大盘名称，在 **大盘布局** 中，单击 **新增报表**，开始配置报表。大盘是由报表组成的，配置大盘就是完成一个个报表的配置，一个报表主要由以下几部分组成：
   - **报表名称**：必填，自定义报表名称。
   - **报表外部链接**：可选。报表的辅助 URL。设置完成后，将展示在报表名右侧。
   - **宽度**：必填，报表宽度。每一行为 12 个宽度，填写 12 就是一行显示 1 个报表，填写 6 就是一行显示 2 个报表，以此类推。在总体大盘中，第一行的 2 个秒级报表的宽度均为 6 。



- **高度**：可选。报表的高度，单位为像素。如果不填写，系统将根据内容自适应高度。
- **报表类型**：选择报表展现的样式。
- **数据源**：选择报表类型后，会出现 **数据源** 选项。报表数值列的数据来源，支持多选。单击 **数据源**，弹出数据源选择页面，选中需要添加的数据源。



- **数值列**：可选。报表的列内容的配置，根据不同的报表类型确定是否需要填写。
  单击 **增加数值列** 或 **智能生成数值列** 添加数值列，其中部分数据源支持智能生成数值列数值添加，单击 **智能生成数值列** 后会自动配置数值列。
  如下图 **交易创建** 报表是 **最近 10 分钟表格** 的报表类型，由 5 个数值列组成。**交易创建走势图** 报表是 **趋势图** 报表类型，由 3 个数值列组成。



单击 **增加数值列**，在弹出的数值列窗口中配置数值列内容。配置完毕后，单击 **确定** 保存。

- **一级表头**：用于将几列数值统一标示名称，如下图 **总体** 就是一级表头。



关键渠道

	总体			[illegible]			
时间	请求	成功	成功率	请求	成功	成功率	占比
16:21	740	617	83.38%	441	361	81.86%	59.59%
16:20	775	640	82.58%	464	404	87.07%	59.87%
16:19	701	621	88.59%	441	399	90.48%	62.91%
16:18	753	640	84.99%	460	400	86.96%	61.09%
16:17	713	598	83.87%	441	382	86.62%	61.85%
16:16	747	654	87.55%	457	402	87.96%	61.18%
16:15	745	620	83.22%	451	404	89.58%	60.54%
16:14	756	658	87.04%	483	435	90.06%	63.89%
16:13	802	668	83.29%	510	444	87.06%	63.59%
16:12	757	647	85.47%	447	399	89.26%	59.05%

- **列名**：用于描述数值的名称。如上图的“请求”、“成功”等。
- **取值函数**：用于配置数值的计算方法。这里提供了以下 14 种取值函数，可以直接显示，也可以配置二个数据源计算得出数值。



- **数值格式化**：指设定数值的小数位数（如只取小数点后 2 位等、百分号形式等）。

- **告警函数**：是指设定数值在报表上用不同色，体现数据告警情况。**橙色阈值** 或 **红色阈值** 表示参数值超出或低于该阈值时，报表上的数值显示橙色或红色。



下图报表中，环比及同比数值列就是配置了报警函数后的效果。



交易创建

时间	笔数	环比	同比	耗时
11:15	314	-3.98%↓	-9.77%↓	57.12
11:14	327	0.31%↑	8.28%↑	59.83
11:13	326	-2.10%↓	-2.40%↓	52.69
11:12	333	-12.37%↓	-6.98%↓	54.7
11:11	380	7.04%↑	18.38%↑	60.49
11:10	355	7.25%↑	5.65%↑	62.13
11:09	331	-2.65%↓	15.73%↑	55.1
11:08	340	5.26%↑	8.97%↑	53.41
11:07	323	-1.52%↓	-4.44%↓	56.46
11:06	328	0.61%↑	3.14%↑	53.98

- **诊断链接**：输入异常数据的诊断页面链接。设置完成后，在大盘报表数据右边会出现问号图标，单击图标后可跳转到链接页面。

- **展示时间长度**：展示时间区间，单位为秒。若不填写，则为默认时间长度。如“最近 10 分钟表格”报表类型，默认显示 10 分钟的数据，如果填写 300，显示 5 分钟（即 300 秒）的数据。
- **展示时间粒度**：展示时间粒度，单位为秒。若不填写，则为默认时间粒度。

5. （可选）配置完报表后，可根据需要执行如下操作：
   - 复制：单击 **复制此报表** 即可复制一个。
   - 调试：单击 **调试预览报表** 即可调试并预览报表配置效果。
   - 删除：单击 **删除此报表**。
   - 新增：单击 **新增报表** 配置更多报表。
6. 配置完毕后，单击 **提交** 完成大盘配置。

下面介绍不同的报表类型的样式及配置操作。

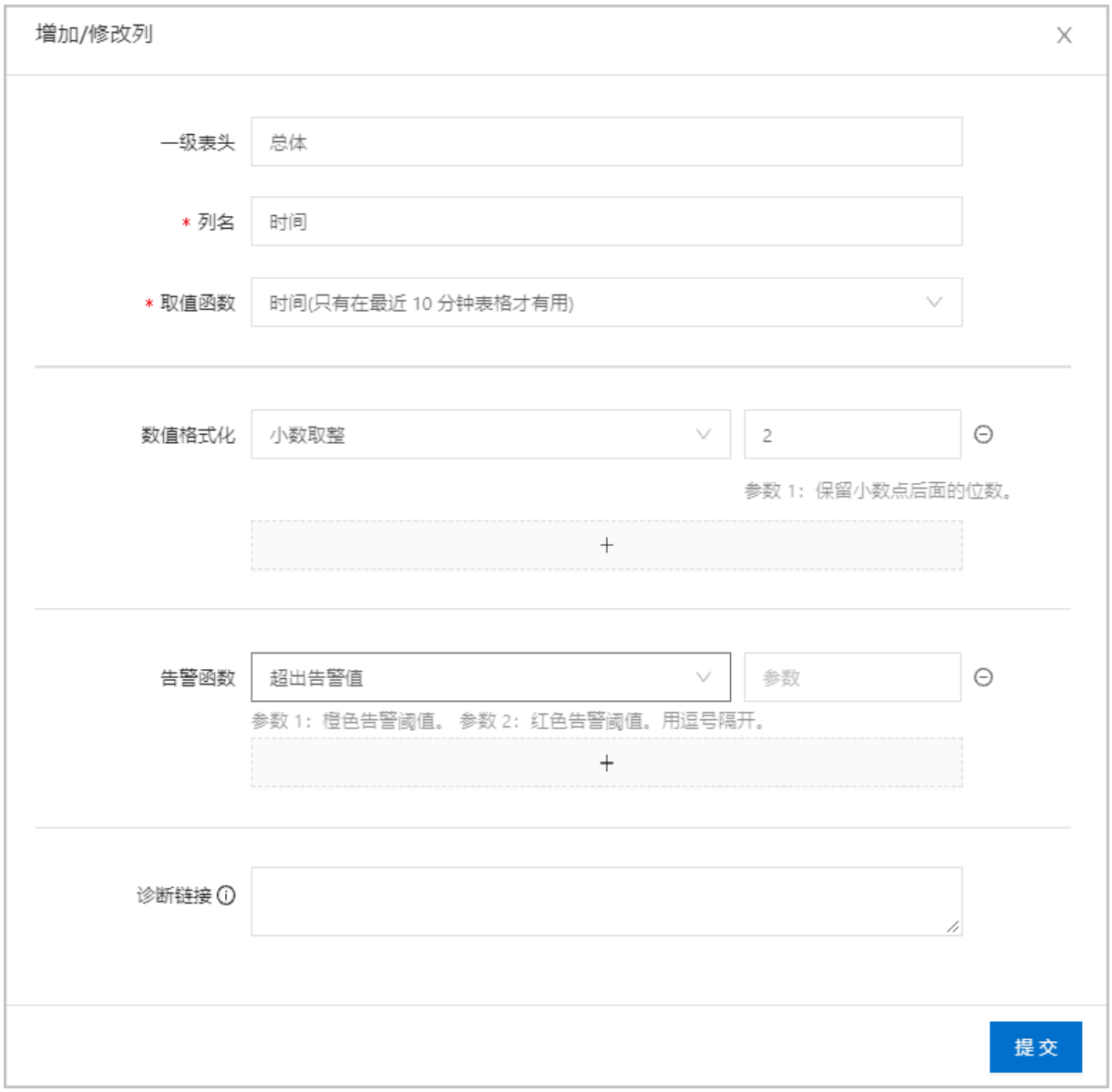
最近 10 分钟表格

最近 10 分钟表格的样式为报表左侧为时间列，仅显示 10 分钟的数据。支持配置 分钟统计/无 Key、分钟统计/多 Key、常用服务指标 等类型的数据源。下图为最近 10 分钟表格报表样式。

交易创建

时间	笔数	环比	同比	耗时
11:15	314	-3.98%↓	-9.77%↓	57.12
11:14	327	0.31%↑	8.28%↑	59.83
11:13	326	-2.10%↓	-2.40%↓	52.69
11:12	333	-12.37%↓	-6.98%↓	54.7
11:11	380	7.04%↑	18.38%↑	60.49
11:10	355	7.25%↑	5.65%↑	62.13
11:09	331	-2.65%↓	15.73%↑	55.1
11:08	340	5.26%↑	8.97%↑	53.41
11:07	323	-1.52%↓	-4.44%↓	56.46
11:06	328	0.61%↑	3.14%↑	53.98

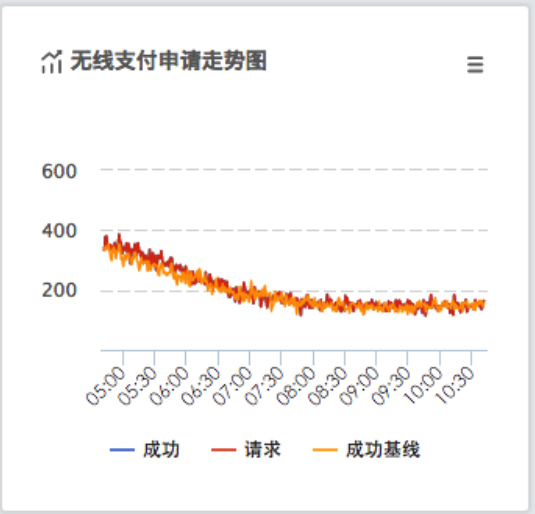
配置最近 10 分钟表格时，需要注意 **时间** 列必须配置，以下为数值列的配置页面：



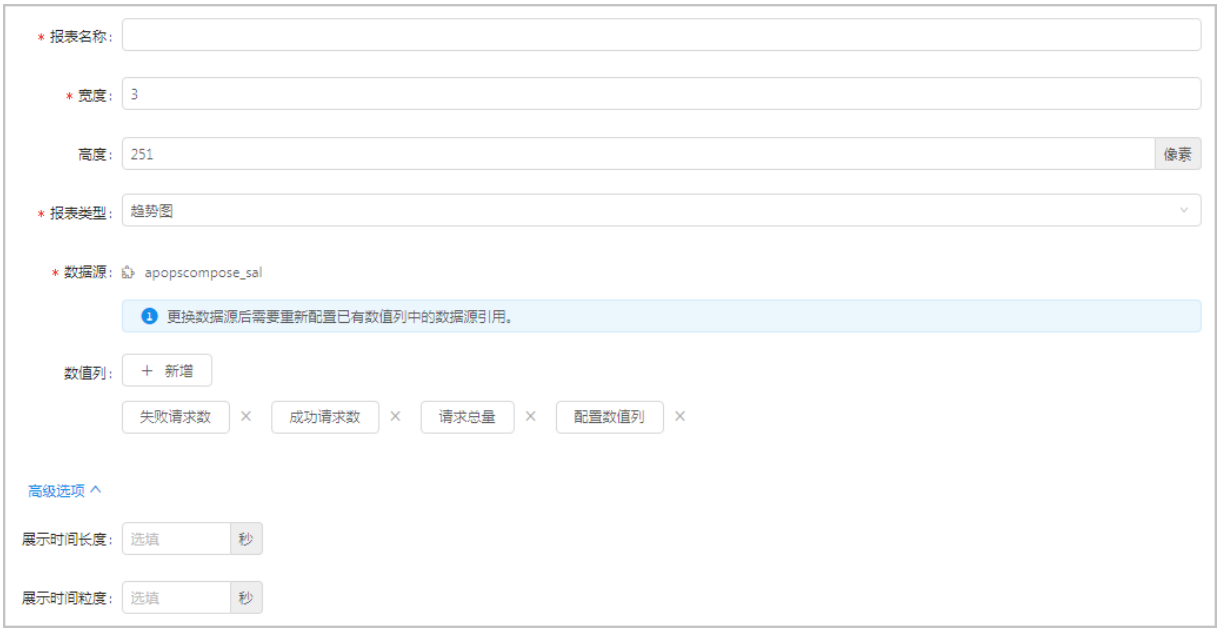
以上具体各项的配置说明，可参考 配置操作。

趋势图

趋势图支持配置 分钟统计/无 Key、分钟统计/多 Key、常用服务指标 、秒级统计/无 Key 等类型的数据源。趋势图的展现样式如下图所示。



趋势图可支持配置多个走势线，只需在配置页面中添加数值列，就可实现多走势线的展示。上图中就是配置了 3 个数值列，显示出 3 条走势线。趋势图配置页面见下图。



上图数据源选择的是常用服务指标的数据源，以下为此类数值列的配置页面：



配置常用服务指标类型数据源的数值列时，必须填写 **数据源中的 Key**。这里的 key 指常用服务指标统计的 **总量**、**成功量**、**成功率**、**总平均耗时** 以及 **失败量**。



业务D的分析量

	11:52					11:51				
接口	总量	成功量	成功率	总平均耗时	失败量	总量	成功量	成功率	总平均耗时	失败量
gets	158	158	100.00%	1.42	0	180	180	100.00%	1.65	0
touch	132	132	100.00%	0.06	0	157	157	100.00%	0.03	0
cas	14	14	100.00%	1.5	0	12	12	100.00%	1.42	0
delete	11	11	100.00%	0.09	0	9	9	100.00%	0	0
get	8	8	100.00%	1.88	0	7	7	100.00%	1.43	0

以上具体各项的配置说明，可参考 配置操作。

当前分钟多 Key 竖铺表格和多分钟多 Key 竖铺表格

当前分钟多 Key 竖铺表格和多分钟多 Key 竖铺表格支持分钟统计/多 Key 或常用服务指标带列值分组的数据源的展现。

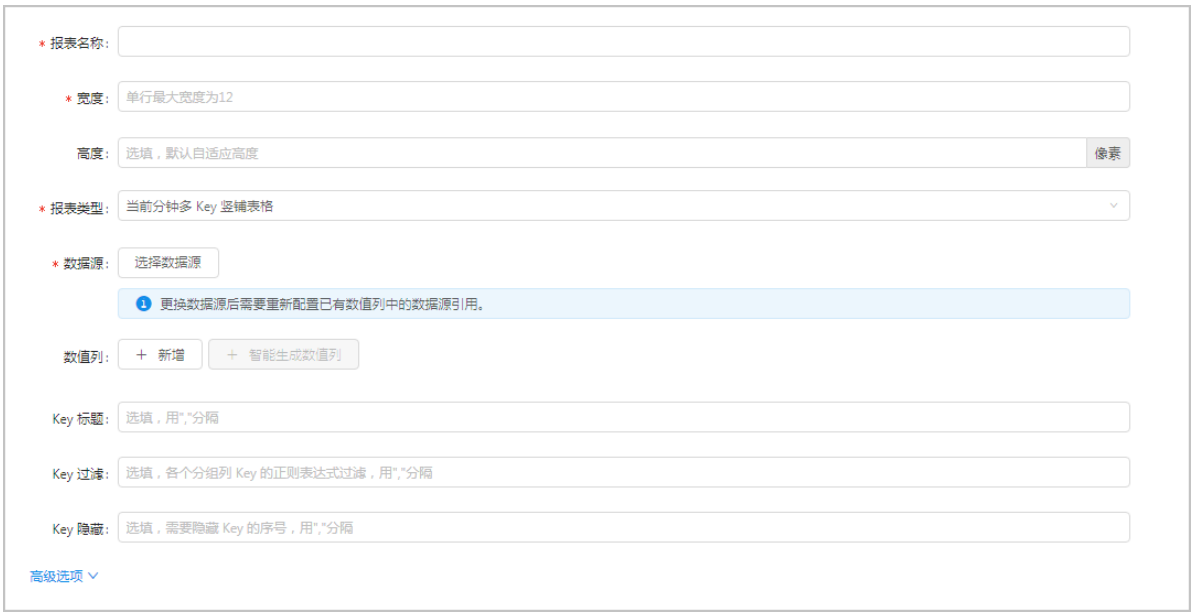
- 当前分钟多 Key 竖铺表格
  下图为当前分钟多 Key 竖铺表格报表样式。



业务D (14:00)

接口	业务D的分析量总量	业务D的分析量成功量	业务D的分析量成功率	业务D的分析量平均耗时	业务D的分析量失败量
gets	116	116	100%	1.56	0
touch	111	111	100%	0.03	0
cas	5	5	100%	1.4	0
get	5	5	100%	1.2	0
delete	5	5	100%	0	0

以下为当前分钟多 Key 竖铺表格报表配置页面。配置页面上 **Key 标题** 指在大盘中数据源 **列值分组** 的列名，以上表样就是在 **Key 标题** 上配置了 **接口**。



- 多分钟多 Key 竖铺表格
  下图为多分钟多 Key 竖铺表格报表样式。



业务D

	14:11					14:10					14:09					14:08					14:07				
接口	总量	成功量	成功率	平均耗时	失败量	总量	成功量	成功率	平均耗时	失败量	总量	成功量	成功率	平均耗时	失败量	总量	成功量	成功率	平均耗时	失败量	总量	成功量	成功率	平均耗时	失败量
gets	107	107	100%	1.42	0	92	92	100%	1.65	0	113	113	100%	1.62	0	59	59	100%	1.64	0	202	202	100%	1.63	0
touch	97	97	100%	0.04	0	89	89	100%	0.03	0	109	109	100%	0.05	0	54	54	100%	0.04	0	192	192	100%	0.03	0
cas	7	7	100%	1.86	0	3	3	100%	1.67	0	3	3	100%	1.33	0	4	4	100%	1	0	8	8	100%	2	0
delete	6	6	100%	0	0	3	3	100%	0	0	3	3	100%	0.67	0	6	6	100%	0	0	7	7	100%	0.14	0
get	5	5	100%	1.8	0	3	3	100%	1.67	0	3	3	100%	1.33	0	4	4	100%	1.25	0	6	6	100%	2.17	0

多分钟多 Key 竖铺表格的配置页面可参考当前分钟多 Key 竖铺表格。

具体各项的配置说明，可参考 配置操作。

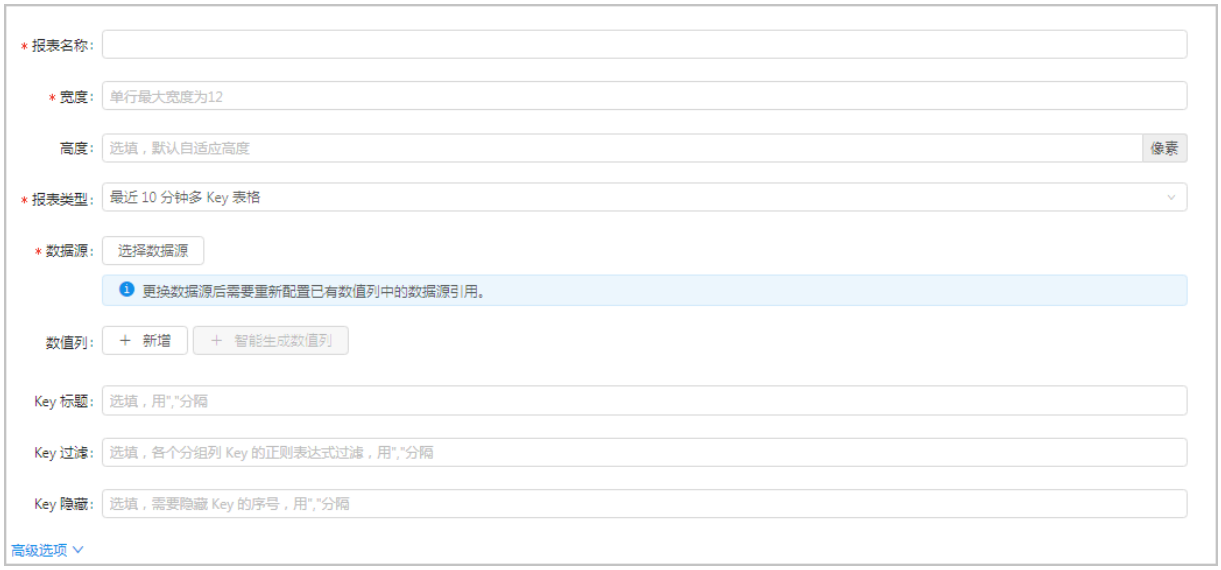
最近 10 分钟多 Key 表格

最近 10 分钟多 Key 表格是类似于多分钟多 Key 竖铺表格的转置显示，即：时间放在左列，列值分组中的 Key 放到报表上第一行显示。可支持分钟统计/多 Key 或常用服务指标带列值分组的数据源配置。下图为最近 10 分钟多 Key 表格报表样式：

业务B最近10分钟多key报表

时间	cas		delete		get		gets		touch	
time	流量	环比	流量	环比	流量	环比	流量	环比	流量	环比
18:21	8	60.00%↑	8	60.00%↑	8	60.00%↑	164	78.26%↑	154	77.01%↑
18:20	5	66.67%↑	5	66.67%↑	5	66.67%↑	92	84.00%↑	87	85.11%↑
18:19	3	-25.00%↓	3	-25.00%↓	3	-25.00%↓	50	-24.24%↓	47	-24.19%↓
18:18	4	-20.00%↓	4	-20.00%↓	4	-20.00%↓	66	3.13%↑	62	5.08%↑
18:17	5	0.00%↓	5	-16.67%↓	5	0.00%↓	64	-48.39%↓	59	-50.42%↓
18:16	5	400.00%↑	6	500.00%↑	5	400.00%↑	124	342.86%↑	119	340.74%↑
18:15	1	-75.00%↓	1	-75.00%↓	1	-75.00%↓	28	-53.33%↓	27	-51.79%↓
18:14	4	0.00%↓	4	0.00%↓	4	0.00%↓	60	-37.50%↓	56	-39.13%↓
18:13	4	-20.00%↓	4	-20.00%↓	4	-20.00%↓	96	20.00%↑	92	22.67%↑
18:12	5	-28.57%↓	5	-28.57%↓	5	-28.57%↓	80	-31.03%↓	75	-31.19%↓

表格配置方法与多分钟多 Key 竖铺表格类似，配置页面如下。



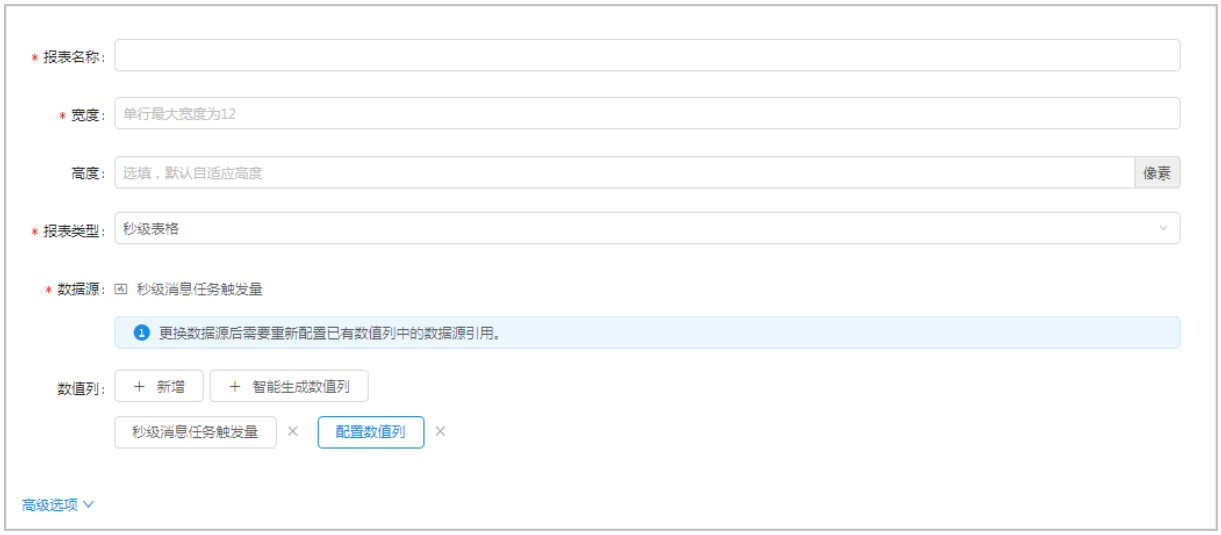
以上具体各项的配置说明，可参考 配置操作。

秒级表格

秒级表格仅支持秒级统计/无 Key 数据源的展现。秒级表格的报表样式及数据源配置页面如下图所示。



秒级表格的配置页面如下所示。秒级统计/无 Key 数据源支持智能生成数据列。这里的 **报表名**、**宽度**、**高度**、**数据源**、**数值列** 等配置类同于 最近 10 分钟表格 中的操作。



秒级多 Key 表格

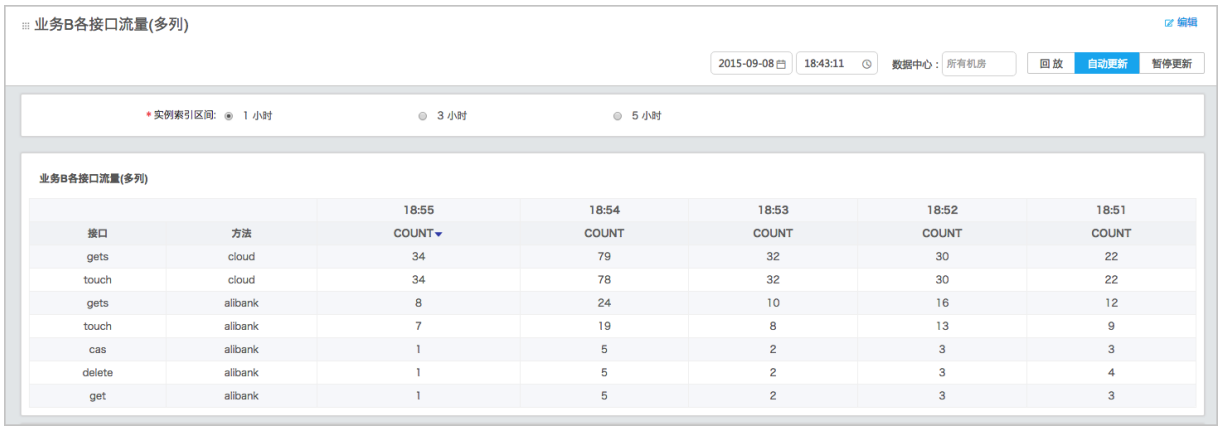
秒级多 Key 表格仅支持 秒级统计/多 key 数据源的展现。其报表配置方法与分钟多 Key 表格类似，配置页面如下。



Key 横竖 MAP 表格（高级）

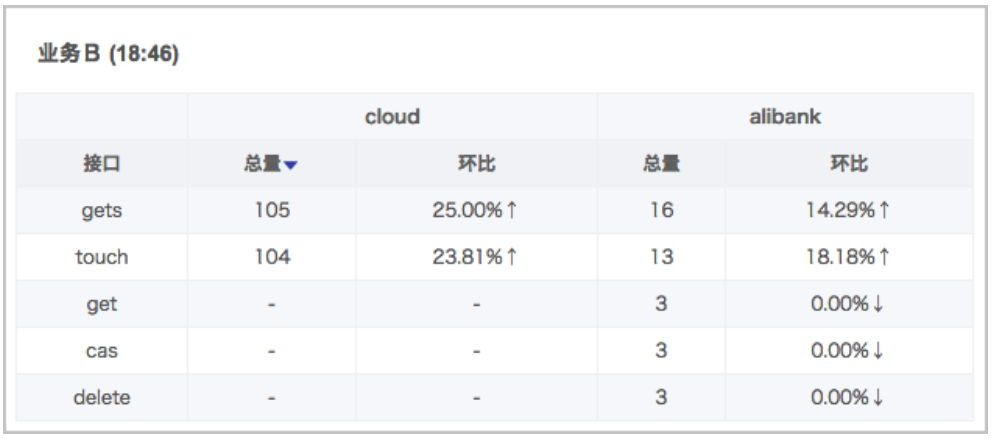
Key 横竖 MAP 表格（高级）是分钟统计/多 Key 和常用服务指标带列值分组类型的数据源，更高级的个性化报表设置。

下图是一个分钟统计/多 Key 类型的数据源表格。



		18:55	18:54	18:53	18:52	18:51
接口	方法	COUNT	COUNT	COUNT	COUNT	COUNT
gets	cloud	34	79	32	30	22
touch	cloud	34	78	32	30	22
gets	alibank	8	24	10	16	12
touch	alibank	7	19	8	13	9
cas	alibank	1	5	2	3	3
delete	alibank	1	5	2	3	4
get	alibank	1	5	2	3	3

可以通过 Key 横竖 MAP 表格（高级）将 **方法** 列在左侧隐去，并将 **方法** 列中的 Key 显示到报表上方（即报表第一行）。样式如下：



业务B (18:46)

	cloud		alibank	
接口	总量	环比	总量	环比
gets	105	25.00%↑	16	14.29%↑
touch	104	23.81%↑	13	18.18%↑
get	-	-	3	0.00%↓
cas	-	-	3	0.00%↓
delete	-	-	3	0.00%↓

具体的配置方法如下：

1. 隐藏左侧 **方法** 列：在 **Key 隐藏** 中设定，填写需要隐藏列的序号，比如：隐藏第一列则填写 `0`，隐藏第二列则填写 `1`，隐藏第一列和第三列则填写 `0,2`。
2. 将 **方法** 列名配置到报表上方：通过数值列配置页面中的 **Key 值补充** 项配置。如下图所示，新数值列是显示 **方法** 的其中一个 Key “cloud” 的数值，所以填入 `.,cloud`，`.` 表示 **接口** 列的 Key 不做筛选。按照此步骤依次配置其他方法即可。



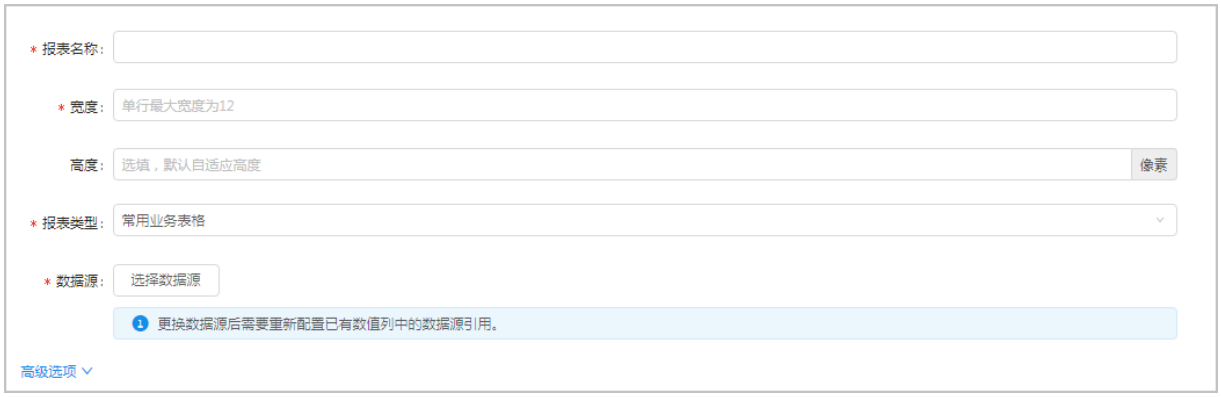
常用业务表格

常用业务表格只支持常用服务指标不带列值分组的数据源的展现。下图为此类型的报表样式及配置页面。

业务D

时间	总量	成功量	成功率	失败量	总平均耗时
16:43	285	285	100.00%	0	0.92
16:42	218	218	100.00%	0	0.82
16:41	210	210	100.00%	0	0.82
16:40	208	208	100.00%	0	0.82
16:39	270	270	100.00%	0	0.83
16:38	222	222	100.00%	0	0.8
16:37	306	306	100.00%	0	0.83
16:36	178	178	100.00%	0	0.91
16:35	262	262	100.00%	0	0.86
16:34	148	148	100.00%	0	0.91

常用业务表格特用于常用服务指标无列值分组类型的数据源，所以此类报表无需配置数值列，数值列默认为 6 列（如上图）。如果只需要显示其中几列，可以用最近 10 分钟表格类型来配置。



以上具体各项的配置说明，可参考 配置操作。

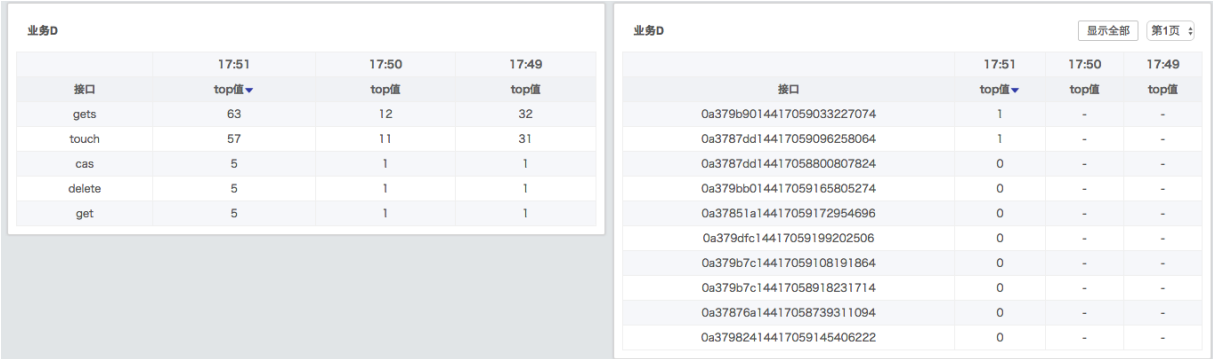
应用错误 Top

应用错误 Top 报表用于展现站内错误量最高的前 10 个应用。此类型报表无需配置数据源和数值列。
报表样式如下：

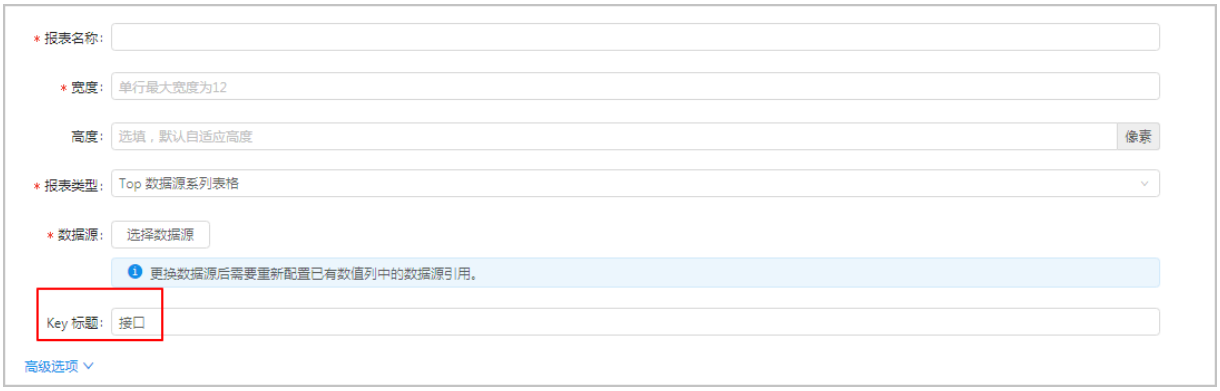


Top 数据源系列表格

Top 数据源系列表格支持单笔数据 Top 与分钟统计 Top 两种数据源的展现，下图为此类型表格样式。

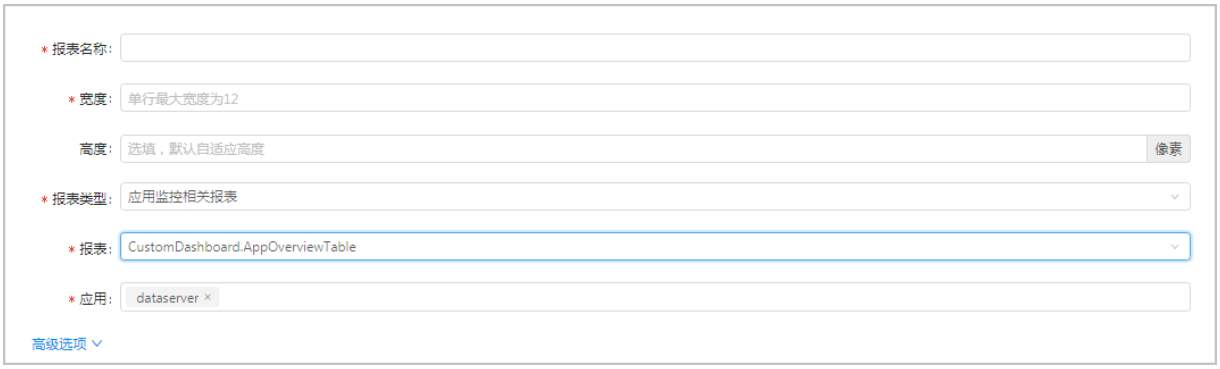


Top 数据源系列表格的配置页面上 **Key 标题** 指在大盘中数据源 **列值分组** 的列名，必填。如上图所示，**Key 标题** 上配置了“接口”。配置页面如下。



应用监控相关报表

将应用监控相关报表自定义配置到大盘里，配置页面如下。



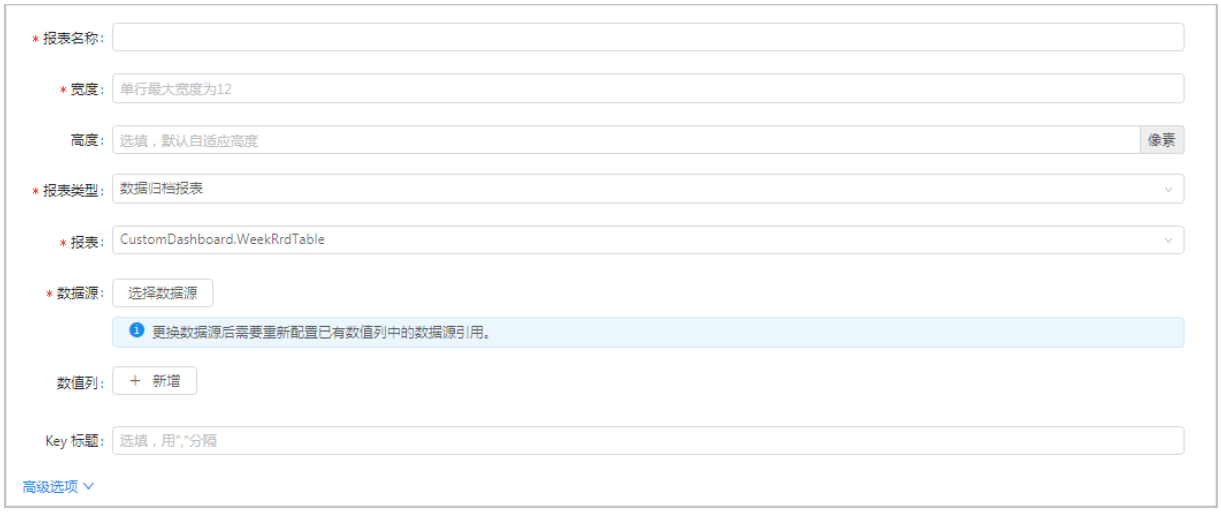
数据归档报表

数据归档报表支持归档统计产品的展现，表格样式如下图所示：



数据归档报表的配置页面如下所示：

> **注意**
> 配置数据归档报表时，务必选取 **报表**。其它配置，参考 常用业务表格。



10.6.4. 分钟统计/无 Key

本文以添加 **分钟统计/无 Key** 为例，介绍如何快速配置自定义监控。

背景信息

分钟统计/无 Key 是通过对日志固定位置的关键字筛选，统计单个指标数据，如统计某个接口被调用的总量等。

更多监控类别应用场景，请参见 监控类型概述。

配置自定义监控主要包括如下的步骤：

1. 选择监控产品
2. 填写基本信息
3. 采集筛选
4. 配置日志筛选
5. 选择统计方式
6. 配置告警
7. 配置诊断

一、选择监控产品

1. 进入控制台，单击左侧导航栏中的 **自定义监控**。
2. 在 **自定义监控** 页面，单击 **新增**。
3. 在 **常用** 标签页下，单击 **分钟统计/无 Key**，进入分钟统计/无 Key 监控产品配置页面。

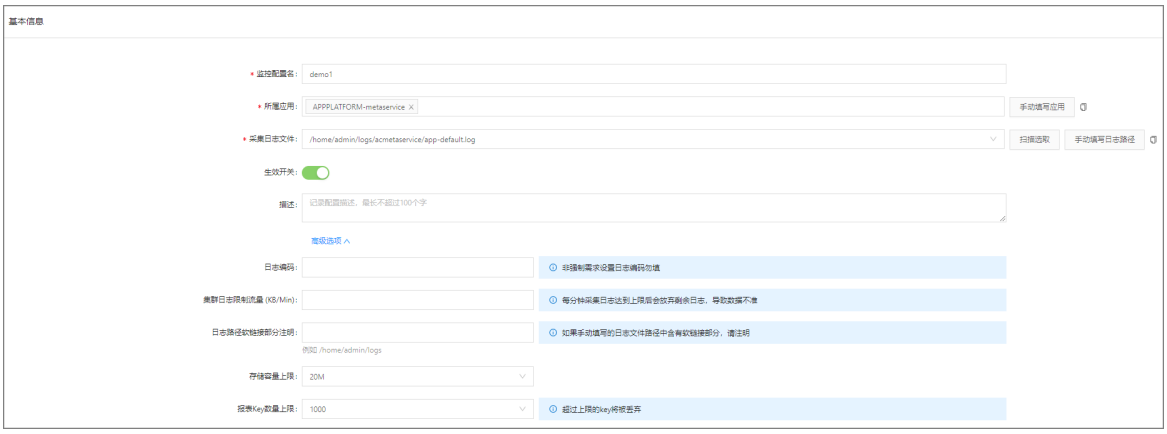


二、填写基本信息

在 **基本信息** 区域中，配置以下信息：

- **监控配置名**：输入自定义的监控配置名称。
- **所属应用**：选择该监控产品所属的应用名称。
- **采集日志文件**：选择日志文件。提供以下采集日志的方式：
  - **扫描路径**：如果列表中没有需要的日志文件，完成以下步骤。默认 **扫描路径** 方式：
    a. 单击 **扫描路径** 打开 **选取采集文件** 页面。该页面显示两个系统默认的文件路径。
    b. 双击路径名打开进行日志文件选取。如果默认的文件路径中，没有找到需要的路径，填写 **指定根目录** 或 **指定服务器 IP** 来查找需要的日志文件。查找时，只需要填写 **指定根目录** 或 **指定服务器** 进行查找。
    c. 单击选中需要的日志文件名，单击 **确认**，完成日志文件选取。
  - **手动填写日志路径**：手动填写日志路径。注意路径规则。
- **生效开关**：选择是否开启 **生效开关**，默认开启。
- （可选）单击 **高级选项**，配置如下信息：
  - **日志编码**：编码为系统自动识别的，一般支持 GBK、UTF-8 等。当有编码不识别时，可以在此处配置填写。
  - **集群日志限制流量** ：集群日志流量限值，单位为 KB/Min。每分钟采集日志达到上限后会放弃剩余日志，导致数据不准。
  - **日志路径软链接部分注明**：例如 `/home/admin/logs` 。如果手动填写的日志文件路径中含有软链接部分，请注明。
  - **存储容量上限**：选择容量上限。

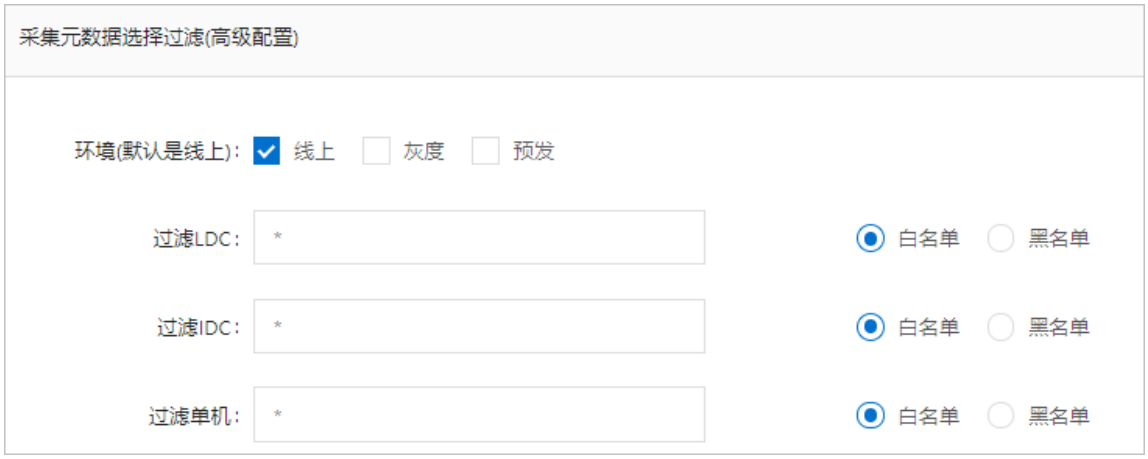
- **报表 Key 数量上限**：超过上限的 key 将被丢弃。



三、采集筛选

采集元数据默认环境为 **线上**。灰度和预发环境为可选环境。同时，支持单独统计单个或多个 LDC（逻辑机房）、IDC（物理机房）、单机。

> **说明**
> 如需过滤多个 LDC 或 IDC 或单机，需通过逗号分隔。目前暂时不支持正则填写。



四、配置日志筛选

日志筛选 提供 **新增白名单筛选** 或 **新增黑名单筛选** 两种方式进行日志关键字筛选：

- **白名单** 相当于 SQL 里的 `[Where Field =?]` 语义。
- **黑名单** 相当于 SQL 里的 `[Where Field < >?]` 语义。

默认关闭。如需筛选日志关键字，可开启日志筛选开关。以新增白名单筛选规则为例，完成以下步骤：

1. 单击 **新增白名单筛选**，新增一条白名单筛选列。
2. 在 **白名单筛选列**，单击 **选取列值**。
3. 在 **列定义** 对话框，配置白名单筛选列的位置。

4. 用鼠标在日志样本中选中目标一个关键词，系统会根据该关键词的位置自动生成一个 **列值选取规则**，并在下方的列值采样表格中显示根据这一选取规则采样得到的列值。



5. 在最下方填入列名，单击 **提交**，完成白名单列的选取。

> **说明**
>
> - 若要进行一些方便的统计归类，以及编码与名称的映射，可使用 **列值翻译** 功能。
> - 当翻译量较大时，可以将之前配置过的类似翻译规则转化为脚本并复制到当前脚本配置框中，根据具体情况进行调整。

6. 填写 **白名单值**。

> **说明**
>
> 同一个列值位的白名单，支持多个值的选取，可以用逗号（,）分隔，如 `aaaa,bbbb,cccc` 。多个 **白名单筛选列** 之间的关系为 **且**。

五、选择统计方式

分钟统计/无 Key 产品提供以下统计方式对日志进行计算：

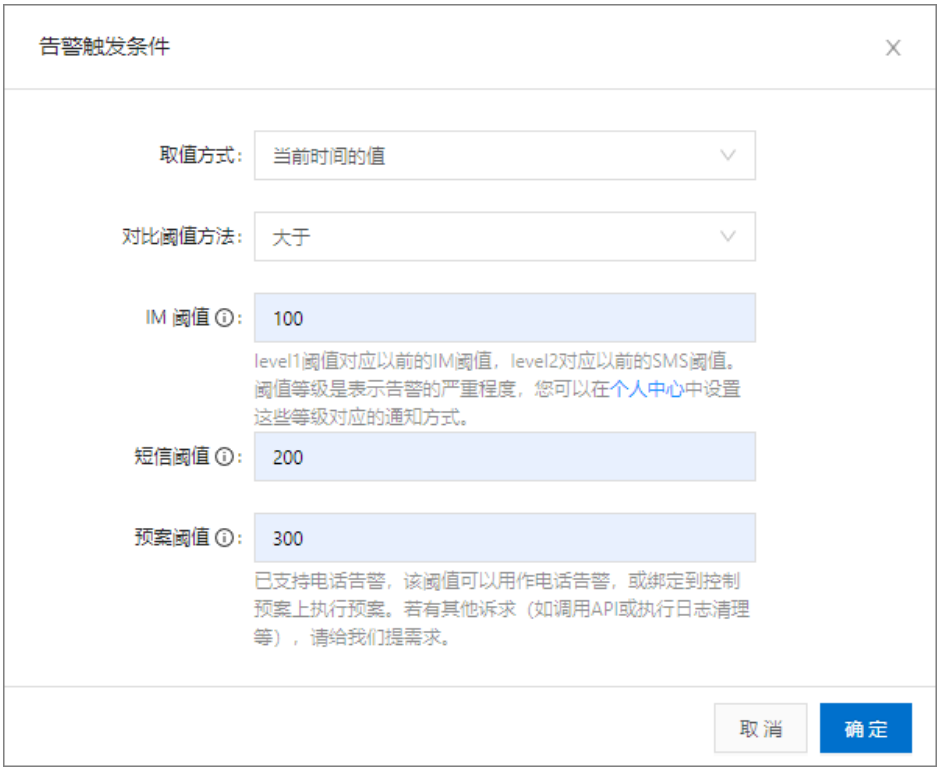
- **求行数**：单位时间内的日志行数。
- **对列值求平均**：需选取计算数值维度，如耗时、金额等。
- **对列值求和**：需选取计算数值维度，如耗时、金额等。
- **对列值求最大值**：需选取计算数值维度，如耗时、金额等。
- **对列值求最小值**：需选取计算数值维度，如耗时、金额等。

本步骤中，选择 **求行数** 进行计算。

六、配置告警

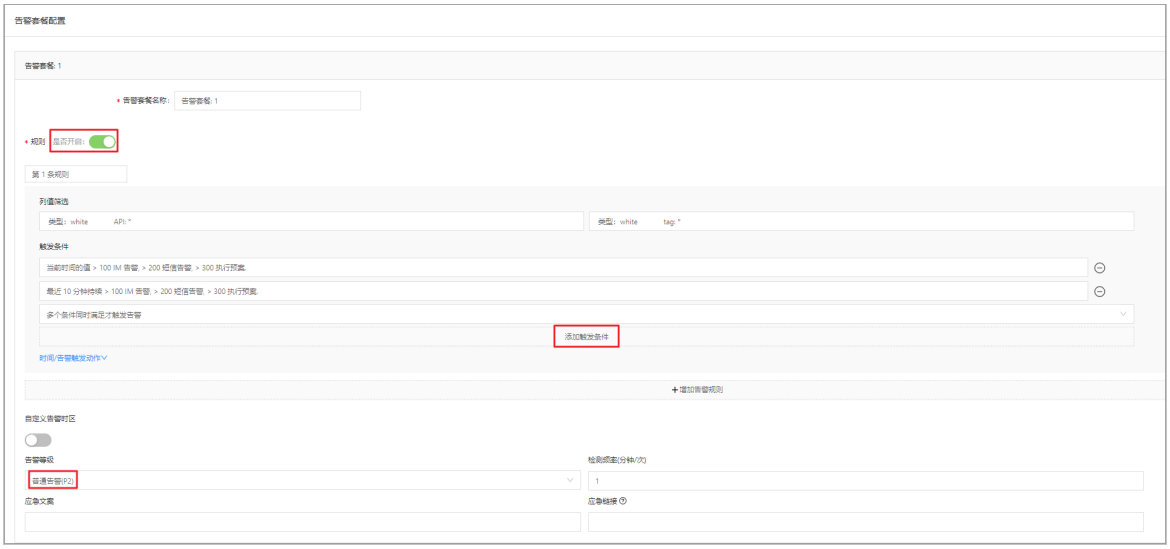
告警配置 指设置一定的告警规则，可通过手机短信方式接收故障告警。实时监控提供可视化告警配置，支持多种告警统计规则、支持多条件多套告警配置、并可按不同时间段设置告警。

1. 单击 **新增告警套餐**。告警规则开关默认开启。
2. 单击 **增加告警规则**。可以设置告警时间、触发动作等参数。
3. 单击 **新增触发条件**，配置取值方式（即告警规则的计算统计方式）、对比方法、各类告警渠道的阈值。



4. 如需多个触发条件，可单击 **新增触发条件** 进行配置。如果设有多个触发条件，需选择 **满足全部条件触发** 还是 **满足任意条件触发**。
5. 设置 **告警等级**、**检测频率**，并根据实际情况确定是否要 **自定义告警时区**。告警等级会在 **总览** 页面的 **监控告警** 区域中做分类统计，同时告警文案中会显示告警指标的等级标注。

> 说明
> 实时监控提供 **红色告警（P0）**、**橙色告警（P1）**、**普通告警（P2）**、**普通告警（P3）**、**普通告警（P4）** 五种等级。在本步骤中，选取默认等级 **普通告警（P2）**。



6. 如需配置多条告警规则，重复上述步骤即可。
7. 配置完毕后，单击 **提交** 完成该自定义监控产品的配置，并生成一个数据源。

七、配置诊断

当上述步骤完成，监控产品已经基本配置完成。单击所配置的监控产品，可进入可视化监控详情页面。单击右上角 **配置诊断助手** 可以对上述步骤的配置进行诊断，主要诊断反馈包括：

- 问题原因
- 诊断结果
- 处理建议

10.6.5. 分钟统计/多 Key

本文以添加 **分钟统计/多 Key** 为例，介绍如何快速配置自定义监控。

背景信息

分钟统计/多 Key 是统计 N 个维度的 N 个 Key 业务量每分钟的数据，如统计系统各个接口分别被调用的总量。通过配置日志的 Group By 维度，统计各维度各 Key 数据，等同于 SQL 中 Group By 的功能。

更多监控类别应用场景，请参见 监控类型概述。

配置自定义监控主要包括如下的步骤：

1. 选择监控产品
2. 填写基本信息
3. 配置日志筛选
4. 设置列值分组
5. 选择统计方式
6. 配置告警
7. 配置诊断

一、选择监控产品

1. 进入控制台，单击左侧导航栏中的 **自定义监控**。
2. 在 **自定义监控** 页面，单击 **新增**。
3. 在 **常用** 标签页下，单击 **分钟统计/多 Key**，进入分钟统计/多 Key 监控产品配置页面。

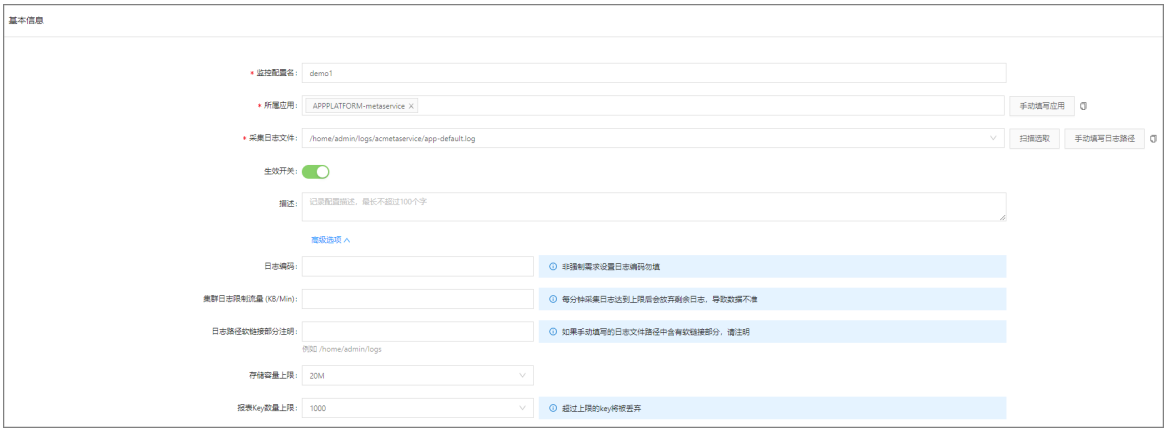


二、填写基本信息

在 **基本信息** 区域中，配置以下信息：

- **监控配置名**：输入自定义的监控配置名称。
- **所属应用**：选择该监控产品所属的应用名称。
- **采集日志文件**：选择日志文件。提供以下采集日志的方式：
  - **扫描路径**：如果列表中没有需要的日志文件，完成以下步骤。默认 **扫描路径** 方式：
    a. 单击 **扫描路径** 打开 **选取采集文件** 页面。该页面显示两个系统默认的文件路径。
    b. 双击路径名打开进行日志文件选取。如果默认的文件路径中，没有找到需要的路径，填写 **指定根目录** 或 **指定服务器 IP** 来查找需要的日志文件。查找时，只需要填写 **指定根目录** 或 **指定服务器** 进行查找。
    c. 单击选中需要的日志文件名，单击 **确认**，完成日志文件选取。
  - **手动填写日志路径**：手动填写日志路径。注意路径规则。
- **生效开关**：选择是否开启 **生效开关**，默认开启。
- （可选）单击 **高级选项**，配置如下信息：
  - **日志编码**：编码为系统自动识别的，一般支持 GBK、UTF-8 等。当有编码不识别时，可以在此处配置填写。
  - **集群日志限制流量** ：集群日志流量限值，单位为 KB/Min。每分钟采集日志达到上限后会放弃剩余日志，导致数据不准。
  - **日志路径软链接部分注明**：例如 `/home/admin/logs` 。如果手动填写的日志文件路径中含有软链接部分，请注明。
  - **存储容量上限**：选择容量上限。

- **报表 Key 数量上限**：超过上限的 key 将被丢弃。



三、配置日志筛选

日志筛选 提供 **新增白名单筛选** 或 **新增黑名单筛选** 两种方式进行日志关键字筛选：

- **白名单** 相当于 SQL 里的 `[Where Field =?]` 语义。
- **黑名单** 相当于 SQL 里的 `[Where Field < >?]` 语义。

默认关闭。如需筛选日志关键字，可开启日志筛选开关。以新增白名单筛选规则为例，完成以下步骤：

1. 单击 **新增白名单筛选**，新增一条白名单筛选列。
2. 在 **白名单筛选列**，单击 **选取列值**。
3. 在 **日志筛选** 对话框，配置白名单筛选列的位置。
   日志筛选 提供 **新增白名单筛选** 或 **新增黑名单筛选** 两种方式进行日志关键字筛选：
   - **白名单** 相当于 SQL 里的 `[Where Field =?]` 语义。
   - **黑名单** 相当于 SQL 里的 `[Where Field < >?]` 语义。
4. 用鼠标在日志样本中选中目标一个关键词，系统会根据该关键词的位置自动生成一个 **列值选取规则**，并在下方的列值采样表格中显示根据这一选取规则采样得到的列值。



5. 在最下方填入列名，单击 **提交**，完成白名单列的选取。

> **说明**
>
> - 若要进行一些方便的统计归类，以及编码与名称的映射，可使用 **列值翻译** 功能。
> - 当翻译量较大时，可以将之前配置过的类似翻译规则转化为脚本并复制到当前脚本配置框中，根据具体情况进行调整。

6. 填写 **白名单值**。

> 说明
> 同一个列值位的白名单，支持多个值的选取，可以用逗号（,）分隔，如 `aaaa,bbbb,cccc` 。多个 **白名单筛选列** 之间的关系为 **且**。如果填入多个值，如 `A,B` ，白名单筛选是要 A 或 B，黑名单筛选是不要 A 和 B。



四、设置列值分组

列值分组即数据的 group by，指明最终生成的数据按什么做分组，RMS 目前支持按 label 分组，即按运维维度做 group by。例如，按机房 IDC 做分组。

> 说明
> 列值分组 Group By 中必须配置一个维度，分组列可配多列。

您可以直接勾选运维维度列作为分组列，若需要自定义分组列，单击 **新增分组列 > 选取**，在弹出的 **日志筛选** 弹框中定义列值选取规则。
用鼠标在日志样本中选中一个目标关键词，系统会根据该关键词的位置自动生成一个列值选取规则，并在下方的列值采样表格中显示根据这一选取规则采样得到的列值。

五、选择统计方式

分钟统计/无 Key 产品提供以下统计方式对日志进行计算：

- **求行数**：单位时间内的日志行数。
- **对列值求平均**：需选取计算数值维度，如耗时、金额等。
- **对列值求和**：需选取计算数值维度，如耗时、金额等。
- **对列值求最大值**：需选取计算数值维度，如耗时、金额等。
- **对列值求最小值**：需选取计算数值维度，如耗时、金额等。

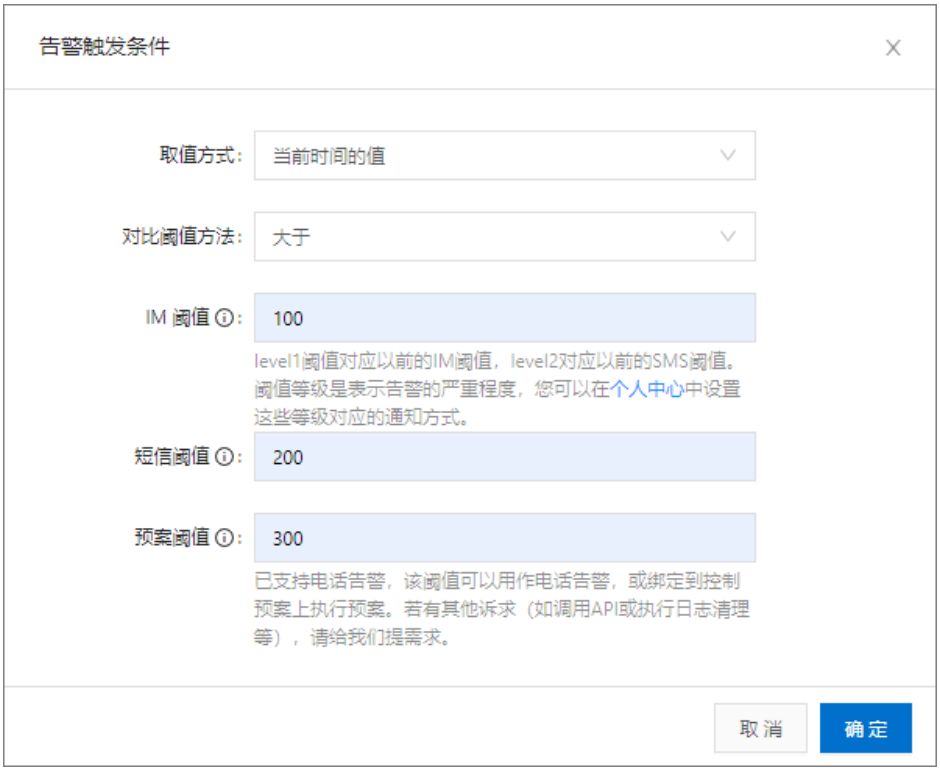
本步骤中，选择 **求行数** 进行计算。

六、配置告警

告警配置 指设置一定的告警规则，可通过手机短信方式接收故障告警。实时监控提供可视化告警配置，支持多种告警统计规则、支持多条件多套告警配置、并可按不同时间段设置预警。

1. 单击 **新增告警套餐**。 告警规则开关默认开启。
2. 单击 **增加告警规则**。
3. 在 **列值筛选** 列的输入框中，输入白名单或黑名单的取值。

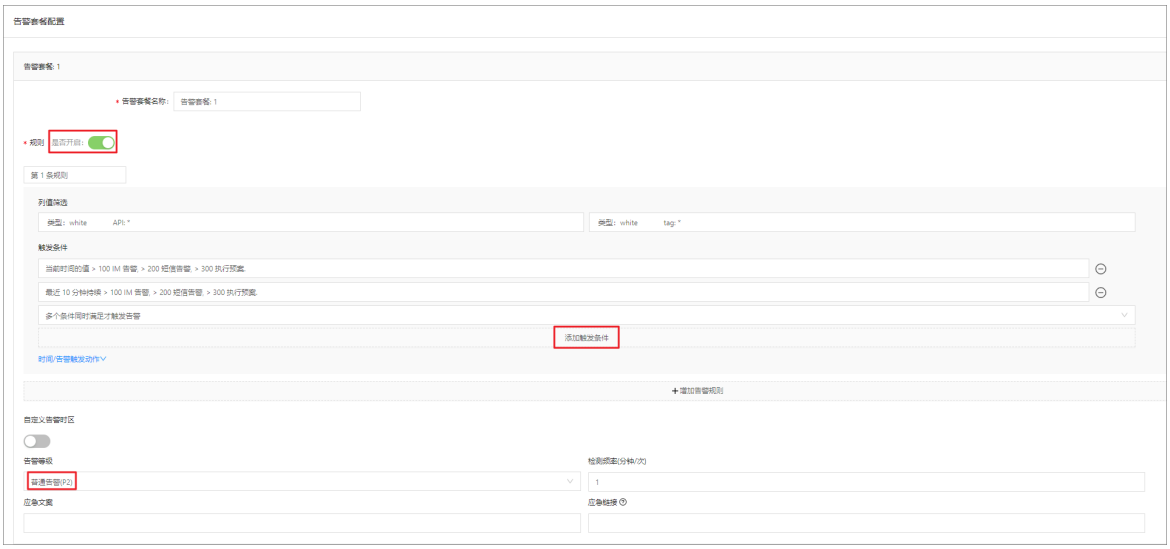
4. 单击 **新增触发条件**，配置取值方式（即告警规则的计算统计方式）、对比方法、各类告警渠道的阈值。



5. 如需多个触发条件，可单击 **新增触发条件** 进行配置。如果设有多个触发条件，需选择 **满足全部条件触发** 还是 **满足任意条件触发**。

6. 设置 **告警等级**、**检测频率**，并根据实际情况确定是否要 **自定义告警时区**。告警等级会在 **总览** 页面的 **监控告警** 区域中做分类统计，同时告警文案中会显示告警指标的等级标注。

> **说明**
> 实时监控提供 **红色告警**、**橙色告警（P1）**、**普通告警（P2）**、**普通告警（P3）**、**普通告警（P4）** 五种等级。在本步骤中，选取默认等级 **普通告警（P2）**。



7. 如需配置多条告警规则，重复上述步骤即可。

8. 配置完毕后，单击 **提交** 完成该自定义监控产品的配置，并生成一个数据源。

七、配置诊断

当上述步骤完成，监控产品已经基本配置完成。单击所配置的监控产品，可进入可视化监控详情页面。单击右上角 **配置诊断助手** 可以对上述步骤的配置进行诊断，主要诊断反馈包括：

- 问题原因
- 诊断结果
- 处理建议

10.6.6. 配置常用服务指标

常用服务指标 用于快速配置单业务或 N 个维度的 N 个 Key 业务的总量、成功量、失败量、成功率、平均耗时等 5 个统计项。

应用场景

要从日志中统计每分钟每个收费站收费多少车辆、成功收费多少车辆、多少车辆收费失败、平均收费时长多少，使用常用服务指标对收费站这个维度进行 Group By，将收费时长维度配成耗时，再配置结果维度，即可统计出每个收费站收费总量、成功量、失败量、成功率、收费时长等的分钟数据。

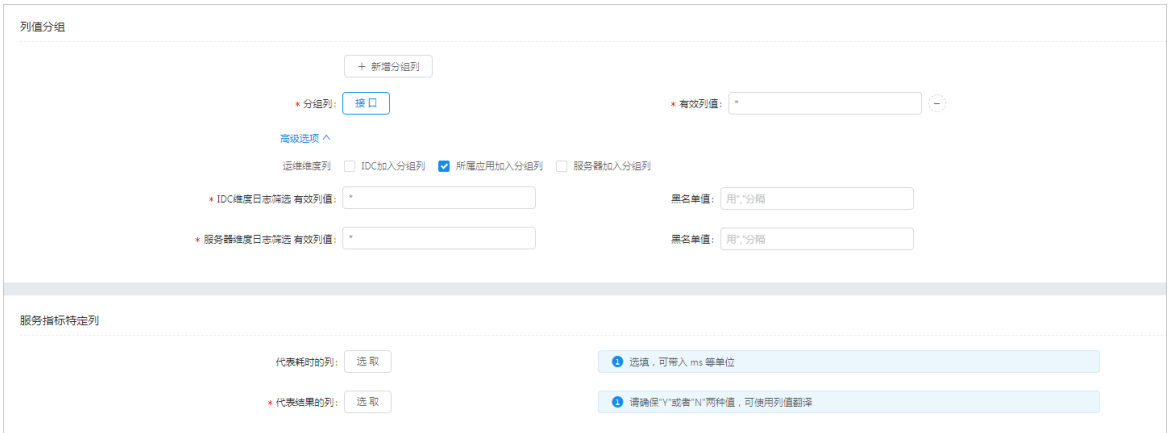
日志内容：

```
2012-11-1111:11:11粤 A123XX,广州北收费站,G25,广州,￥50,2340ms,Y
2012-11-1111:11:12粤 A123XX,广州北收费站,G26,广州,￥50,3209ms,Y
2012-11-1111:11:13粤 A123XX,广州北收费站,G27,广州,￥50,4200ms,Y
2012-11-1111:11:14粤 A123XX,广州北收费站,G27,广州,￥50,5700ms,Y
2012-11-1111:11:15粤 A123XX,广州北收费站,G26,广州,￥50,8500ms,Y
……
```

操作步骤

1. 登录实时监控产品控制台，从左侧导航栏选择 **自定义监控 > 新建 > 新建监控产品**。
2. 在 **常用** 标签下，单击 **常用服务指标**。
3. 配置基本信息。在 **基本信息** 区域，填写 **监控配置名**、**所属应用**、**采集日志文件** 等信息。
4. 配置日志筛选。您可以选择配置白名单或黑名单，具体步骤参考 配置自定义监控。
5. （可选）配置 **列值分组**。选取列值分组的维度，产品将按各维度各 Key 统计数值。

> **说明**
> 列值分组中必须配置一个维度，分组列可配多列。



6. 设置 **服务指标特定列**。常用服务指标可以快速配置总量、成功量、成功率、失败量和平均耗时等 5 个指标数据，所以计算这些量时，需要配置 **代表耗时的列** 和 **代表结果的列**。

- **代表结果的列**：必填，即 Y 的位置，如果日志内容不是“Y”，可通过列值翻译将代表成功的字符串译成“Y”。
- **代表耗时的列**：选填，即 N 的位置，如果日志内容不是“N”，可通过列值翻译将代表失败的字符串译成“N”。若日志中无耗时或不需要统计耗时，可选 **不需要统计耗时**。

> 说明
> - 若要进行一些方便的统计归类，以及编码与名称的映射，可使用 **列值翻译** 功能。
> - 当翻译量较大时，可以将之前配置过的类似翻译规则转化为脚本并复制到当前脚本配置框中，根据具体情况进行调整。

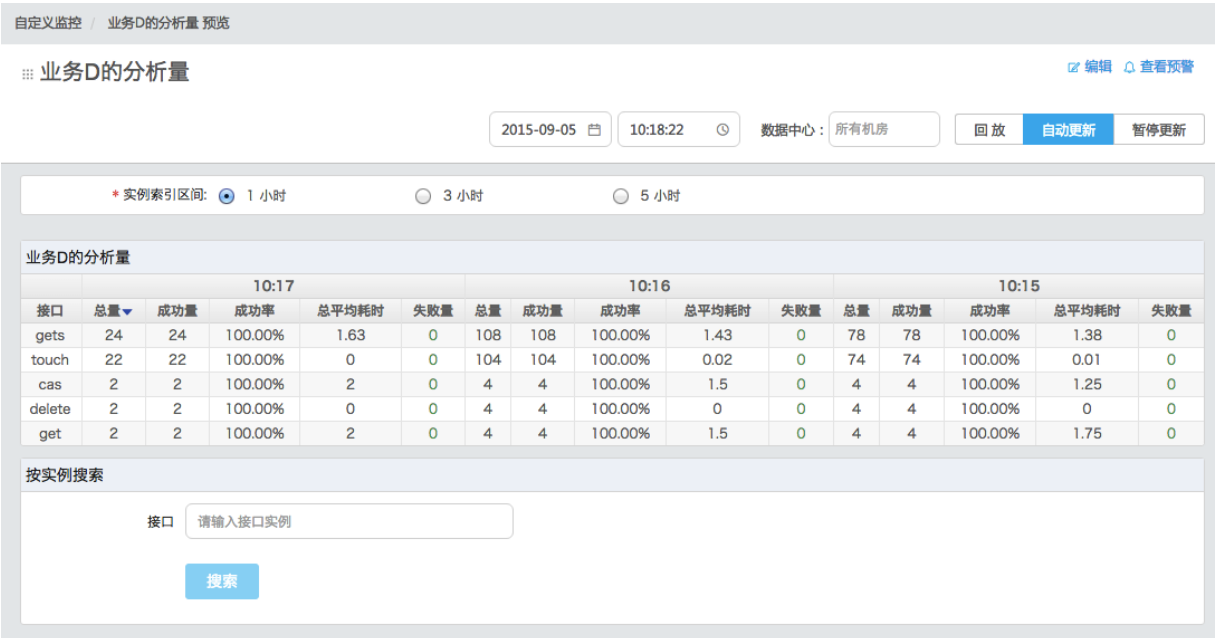
7. 预警配置。具体步骤参考 配置预警。
8. 单击 **提交** 完成配置。

查看数据

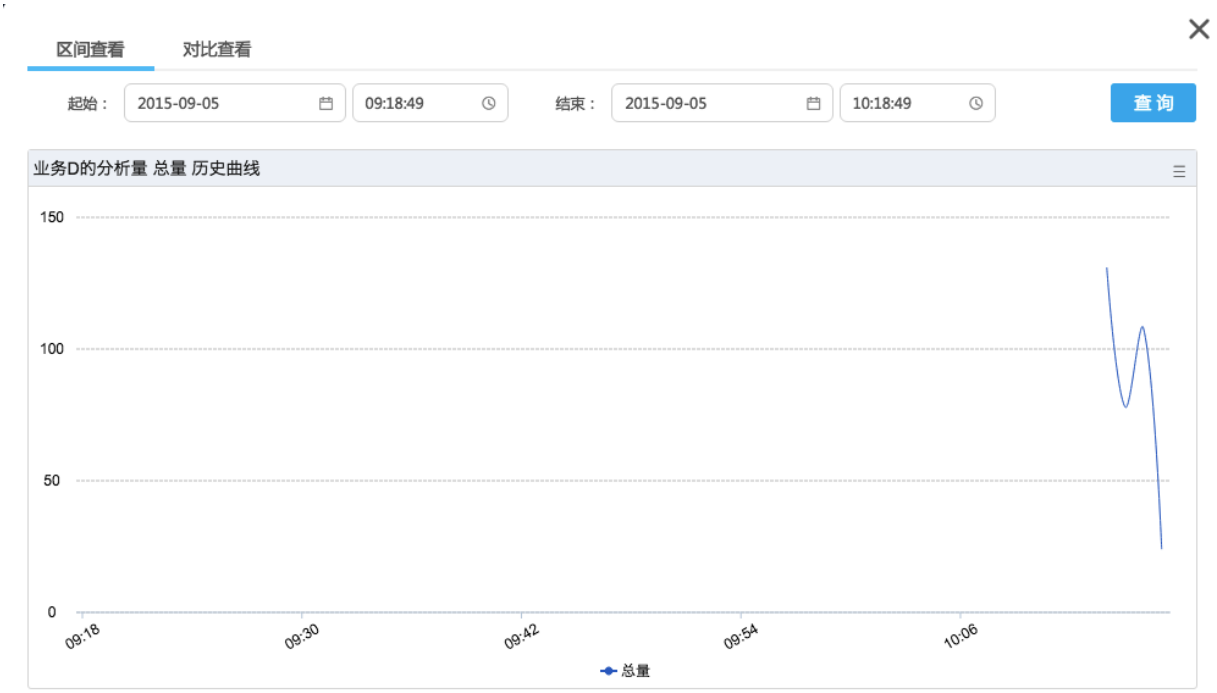
单击左侧导航栏中的 **自定义监控**，然后选择您新增的监控项 **业务 D 的分析量**，可查看统计值。

```
<divclass="note notice note-notice"id="note-ucj-4xg-ybf">
<divclass="note-icon-wrapper"><iclass="icon-note note"></i></div>
<divclass="note-content"><strong>说明</strong>：配置完后，约 2 至 3 分钟后可以查看统计数据。</div>
</div>
```

预览数据报表样例：



如果该监控项配置了 **列值分组**，注意预览数据报表页面中 **实例索引区间** 的 **1 小时**、**3 小时** 及 **5 小时**分别指您当前所在区域时间（如北京时间等）的近 1 小时内、3 小时内或 5 小时内。例如，如果选择 **实例索引区间** 为 **1 小时**，则 “业务 D 的分析量”报表数据就只显示该区间时间内数值大于 0 的接口。
如果要查看当前时间的 5 小时前才有流量的接口的数据，可以在 **按实例搜索** 中的 **接口** 输入框内输入接口名称，进行搜索。单击报表中的数值（如：单击上图 `gets` 接口的总量），可以查看各接口更长的历史数据走势。



10.6.7. 单笔数据 Top

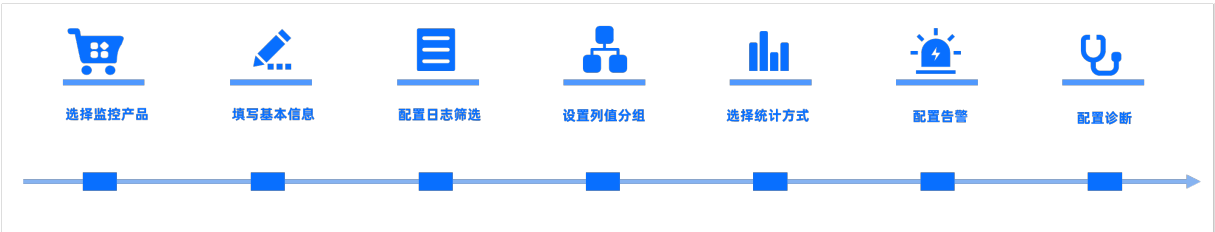
本文以添加 **单笔数据 Top** 为例，介绍如何快速配置自定义监控。

背景信息

单笔数据 Top 是通过对单条日志的数值维度进行排序，查看最高 N 个单笔业务，如统计耗时前 10 的单笔订单。

更多监控类别应用场景，请参见 监控类型概述。

配置自定义监控主要包括如下的步骤：



1. 选择监控产品
2. 填写基本信息
3. 配置日志筛选
4. 设置列值分组
5. 选择统计方式
6. 配置报警
7. 配置诊断

一、选择监控产品

1. 进入控制台，单击左侧导航栏中的 **自定义监控**。
2. 在 **自定义监控** 页面，单击 **新增**。
3. 在 **常用** 页签下，单击 **单笔数据 Top**，进入单笔数据 Top 监控产品配置页面。

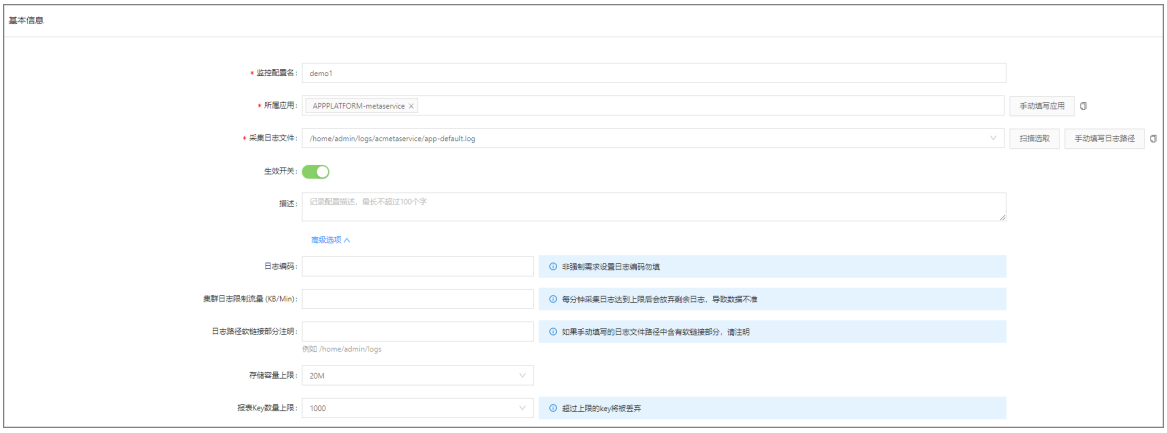


二、填写基本信息

在 **基本信息** 区域中，配置以下信息：

- **监控配置名**：输入自定义的监控配置名称。
- **所属应用**：选择该监控产品所属的应用名称。
- **采集日志文件**：选择日志文件。提供以下采集日志的方式：
  - **扫描路径**：如果列表中没有需要的日志文件，完成以下步骤。默认 **扫描路径** 方式：
    a. 单击 **扫描路径** 打开 **选取采集文件** 页面。该页面显示两个系统默认的文件路径。
    b. 双击路径名打开进行日志文件选取。如果默认的文件路径中，没有找到需要的路径，填写 **指定根目录** 或 **指定服务器 IP** 来查找需要的日志文件。查找时，只需要填写 **指定根目录** 或 **指定服务器** 进行查找。
    c. 单击选中需要的日志文件名，单击 **确认**，完成日志文件选取。
  - **手动填写日志路径**：手动填写日志路径。注意路径规则。
- **生效开关**：选择是否开启 **生效开关**，默认开启。
- （可选）单击 **高级选项**，配置如下信息：
  - **日志编码**：编码为系统自动识别的，一般支持 GBK、UTF-8 等。当有编码不识别时，可以在此处配置填写。
  - **集群日志限制流量** ：集群日志流量限值，单位为 KB/Min。每分钟采集日志达到上限后会放弃剩余日志，导致数据不准。
  - **日志路径软链接部分注明**：例如 `/home/admin/logs` 。如果手动填写的日志文件路径中含有软链接部分，请注明。
  - **存储容量上限**：选择容量上限。

- **报表 Key 数量上限**：超过上限的 key 将被丢弃。



三、配置日志筛选

开启日志筛选开关，以新增白名单筛选规则为例，完成以下步骤：

1. 单击 **新增白名单筛选**，新增一条白名单筛选列。
2. 在 **白名单筛选列**，单击 **选取列值**。
3. 在 **日志筛选** 对话框，配置白名单筛选列的位置。
   日志筛选 提供 **新增白名单筛选** 或 **新增黑名单筛选** 两种方式进行日志关键字筛选：
   - **白名单** 相当于 SQL 里的 `[Where Field =?]` 语义。
   - **黑名单** 相当于 SQL 里的 `[Where Field < >?]` 语义。
4. 用鼠标在日志样本中选中目标一个关键词，系统会根据该关键词的位置自动生成一个 **列值选取规则**，并在下方的列值采样表格中显示根据这一选取规则采样得到的列值。



5. 在最下方填入列名，单击 **确定**，完成白名单筛选列的选取。

> 说明
>
> - 若要进行一些方便的统计归类，以及编码与名称的映射，可使用 **列值翻译** 功能。
> - 当翻译量较大时，可以将之前配置过的类似翻译规则转化为脚本并复制到当前脚本配置框中，根据具体情况进行调整。

6. 填写 **白名单值**。

> 说明
> 同一个列值位的白名单，支持多个值的选取，可以用逗号（,）分隔，如 `aaaa,bbbb,cccc` 。多个 **白名单筛选列** 之间的关系为 **且**。如果填入多个值，如 `A,B` ，白名单筛选是要 A 或 B，黑名单筛选是不要 A 和 B。



四、设置列值分组

列值分组即数据的 group by，指明最终生成的数据按什么做分组。若需要自定义分组列，单击 **新增分组列** > **选取列值**，在弹出的 **列定义** 弹框中定义列值选取规则。用鼠标在日志样本中选中一个目标关键词，系统会根据该关键词的位置自动生成一个列值选取规则，并在下方的列值采样表格中显示根据这一选取规则采样得到的列值。

五、选择统计方式

单笔数据 Top 产品提供 **最大排名数** 统计方式对日志进行计算。本步骤中，保持默认排名数 10 进行计算。

用于排序的列：单击 **选取列值**，再弹出框中选取用区排序的列。

六、配置报警

Top 报警配置 指设置一定的报警规则，可通过手机短信方式接收故障报警。实时监控提供可视化报警配置，支持大于排序值阈值的统计项数量的报警统计规则、支持多条件多套报警配置、并可按不同时间段设置预警。



1. 为不同的报警方式设置排序值阈值及统计项阈值。
2. 配置完毕后，单击 **提交** 完成该自定义监控产品的配置，并生成一个数据源。

七、配置诊断

当上述步骤完成，监控产品已经基本配置完成。单击所配置的监控产品，可进入可视化监控详情页面。单击右上角 **配置诊断助手** 可以对上述步骤的配置进行诊断，主要诊断反馈包括：

- 问题原因
- 诊断结果
- 处理建议

10.6.8. 匹配查找告警

本文以添加 **匹配查找告警** 为例，介绍如何快速配置自定义监控。

背景信息

匹配查找告警 是对固定位置关键字进行告警，可发送包含该关键字的日志内容为告警内容，也可提取部分日志内容作为告警内容的监控产品。

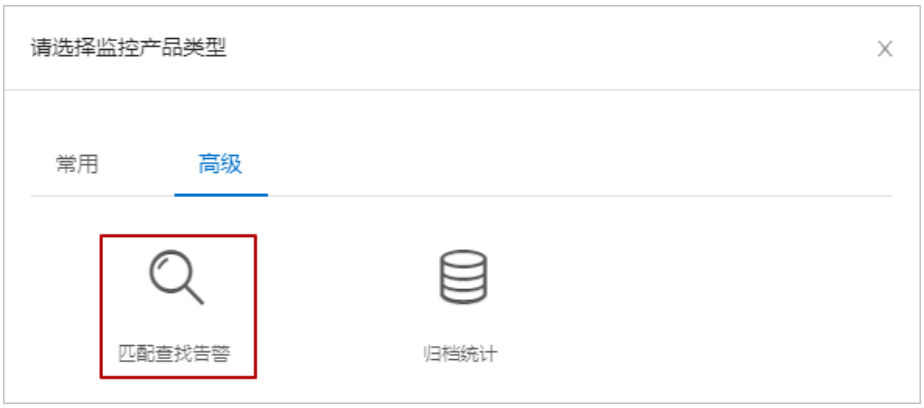
> 说明
> **匹配查找告警** 仅提供告警功能，不支持数据统计及查看数据功能。

更多监控类别应用场景，请参见 监控类型概述。
配置自定义监控主要包括如下的步骤：

1. 选择监控产品
2. 填写基本信息
3. 采集筛选
4. 配置日志匹配
5. 自定义文案
6. 配置告警
7. 配置诊断

一、选择监控产品

1. 进入控制台，单击左侧导航栏中的 **自定义监控**。
2. 在 **自定义监控** 页面，单击 **新增**。
3. 在 **高级** 页签下，单击 **匹配查找告警**，进入匹配查找告警监控产品配置页面。



二、填写基本信息

在 **基本信息** 区域，完成以下信息配置：

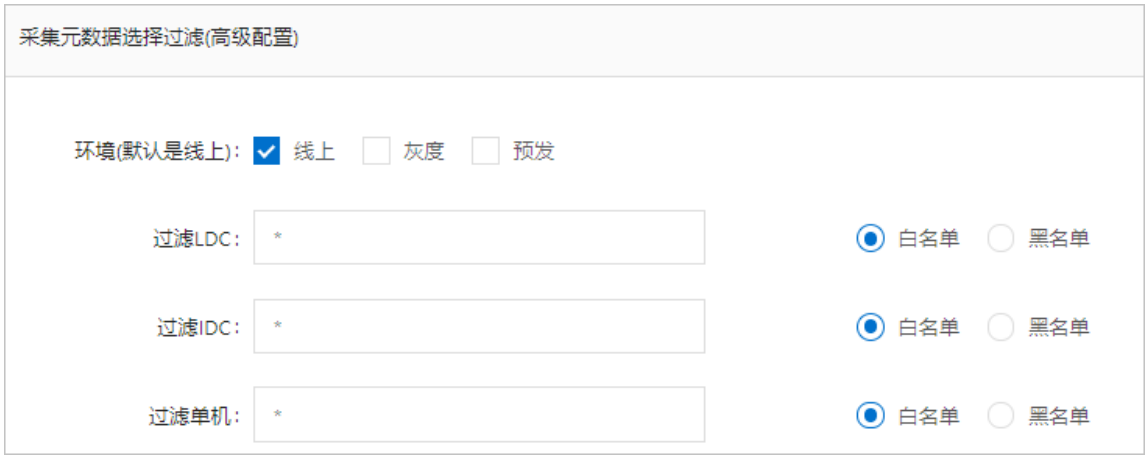
- **监控配置名**：输入自定义的监控配置名称。
- **所属应用**：选择该监控产品所属的应用名称。
- **采集日志文件**：选择日志文件。提供以下采集日志的方式：
  - **扫描选取**：如果列表中没有需要的日志文件，完成以下步骤。默认 **扫描选取** 方式：
    a. 单击 **扫描选取** 打开 **选取采集文件** 页。
    b. 双击路径名打开进行日志文件选取。
    如果默认的文件路径中没有找到需要的路径，填写 **指定根目录** 或 **指定服务器 IP** 来查找需要的日志文件。查找时，只需要填写 **指定根目录** 或 **指定服务器** 进行查找。
    c. 选中需要的日志文件名，单击 **确认**，完成日志文件选取。
  - **手动填写日志路径**：输入日志路径。注意路径规则。
- **生效开关**：选择是否开启 **生效开关**，默认开启。
- **描述**：选填。输入关于自定义配置的描述。
- （可选）单击 **高级选项**，配置如下信息：
  - **日志编码**：编码为系统自动识别的，一般支持 GBK、UTF-8 等。当有编码不识别时，可以在此处配置填写。

- **集群日志限制流量** ：集群日志流量限值，单位为 KB/Min。每分钟采集日志达到上限后会放弃剩余日志，导致数据不准。
- **日志路径软链接部分注明**：例如 `/home/admin/logs` 。如果手动填写的日志文件路径中含有软链接部分，请注明。
- **存储容量上限**：选择容量上限。
- **报表 Key 数量上限**：超过上限的 Key 将被丢弃。

三、采集筛选

采集元数据默认环境为 **线上**。灰度和预发环境为可选环境。同时，支持单独统计单个或多个 LDC（逻辑机房）、IDC（物理机房）、单机。

> 说明
> 如需过滤多个 LDC 或 IDC 或单机，需通过逗号分隔。目前暂时不支持正则填写。



四、配置日志匹配

在日志匹配区域，可通过开启 **是否需要日志匹配** 配置 **新增白名单匹配规则** 或 **新增黑名单匹配规则** 进行日志关键字匹配。

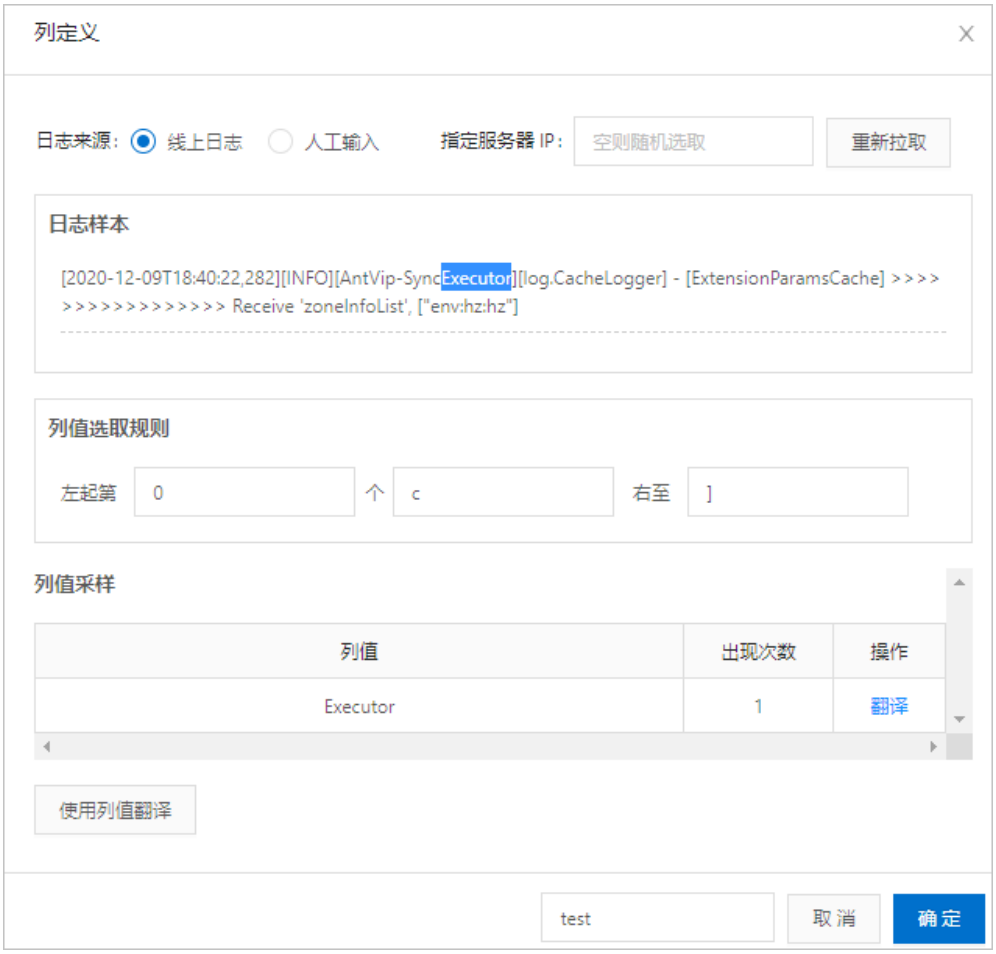
> 说明
> - 白名单相当于 SQL 中的 `[Where Field =?]` 语义，黑名单相当于 SQL 中的 `[Where Field < >?]` 语义。
> - 当配置了多个白名单匹配列时，它们之间的逻辑关系为 **且**。
> - 当配置了白名单匹配列和黑名单匹配列时，它们之间的逻辑关系为 **且**。
> - 当配置了多个黑名单匹配列时，它们之间的逻辑关系为 **或**。

以 **新增白名单匹配规则** 为例，具体步骤如下：

1. 单击 **新增白名单规则**，新增一条白名单匹配列。



2. 单击 **选取取列值**，在弹出的 **列定义** 页配置白名单匹配列位置。
   您可以在日志样本中选中目标一个关键词，系统会根据该关键词的位置自动生成一个 **列值选取规则**，并在下方的 **列值采样** 表格中显示根据这一选取规则采样得到的列值。



在最下方填入列名，单击 **确定**，完成白名单列的选取。

3. 配置 **白名单值**。
   同一个列值位的白名单，支持多个值的选取，可以用英文逗号（,）分隔，如：`aaaa,bbbb,cccc` 。

五、自定义文案

您可以选择 **使用默认告警文案** 或者 **使用自定义告警文案** 来配置告警内容文案。
当不需要整条日志内容作为告警内容时，可用正则方式提取部分日志内容作为告警内容。



六、配置告警

匹配查找告警 的告警内容为整条日志内容。在 **告警配置** 区域中，配置以下信息：

- **告警标题**：输入自定义的告警标题。
- **告警内容附带最大日志条数**：发送整条日志的最多条数。
- **触发 IM 告警最小匹配日志条数**：触发 IM 告警的条件。
- **触发短信告警最小匹配日志条数**：触发短信告警的条件。
- **告警等级**：支持 **普通报警（P2）**、**橙色报警（P1）**、**红色报警（P0）**。
- **应急文案**：自定义应急文案。
- **应急链接**：输入应急链接。应急链接和文案会随告警消息一并发出。

七、配置诊断

当上述步骤完成，监控产品已经基本配置完成。单击所配置的监控产品，可进入可视化监控详情页面。单击右上角 **配置诊断助手** 可以对上述步骤的配置进行诊断，主要诊断反馈包括：

- 问题原因
- 诊断结果
- 处理建议

10.6.9. 归档统计

本文以添加 **归档统计** 为例，介绍如何快速配置自定义监控。

背景信息

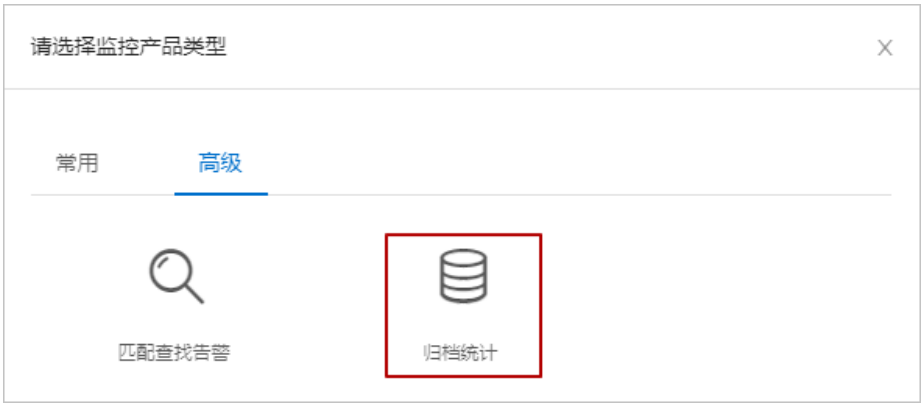
归档统计 是用于时间单位归档的监控产品，方便将平台分钟级数据源按小时、天、星期归档。

> 说明
> 归档统计 仅支持数据统计功能，不提供告警功能。

更多监控类别应用场景，请参见 监控类型概述。

一、选择监控产品

1. 进入控制台，单击左侧导航栏中的 **自定义监控**。
2. 在 **自定义监控** 页面，单击 **新增**。
3. 在 **高级** 页签下，单击 **归档统计**，进入归档统计监控产品配置页面。



二、填写基本信息

在 **基本信息** 区域，完成以下信息配置后，单击 **提交**。

- **产品名称**：必填。自定义产品名称。本例中输入 **常用服务指标的归挡统计_Demo**。
- **数据源**：单击 **数据源**，在弹出框中选择归档数据源类型。
- **归档周期**：支持 **小时**、**天**、**星期**。本例中选择 **星期**。

> **说明**
> - 如果选择按 **天** 归档，会同时默认选中按小时归档。
> - 如果选择按 **星期** 归档，会同时默认选中按小时归档和按天归档。

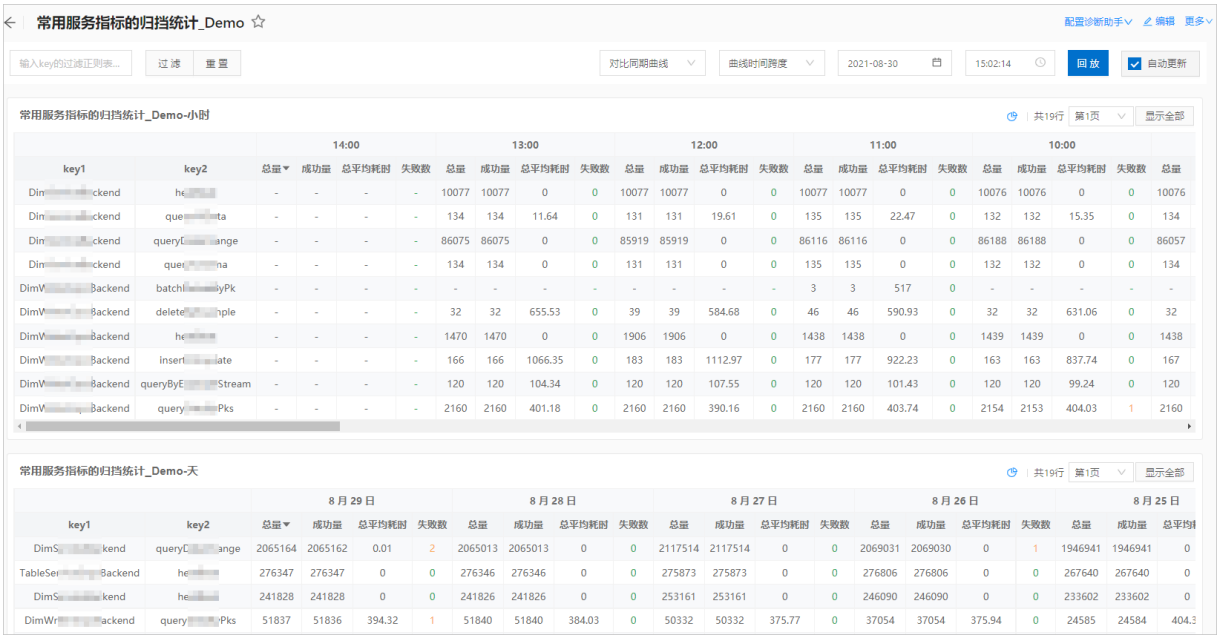
- **自定义索引**：可选。输入自定义索引。
- **存储容量上限**：可选。从下拉列表中选择存储容量上限值。
- **报表 Key 数量上限**：必填。从下拉列表中选择报表 Key 数量上限值。

> **说明**
> - 如果需要统计小时、天、星期三个统计时间的值，此处 Key 上限需要是数据源实际 Key 量的 3 倍。
> - 如果报表 Key 数量超过所选的上限，超过上限的 Key 将被丢弃。



后续操作

监控产品基本配置完成后，单击所配置的监控产品，进入可视化监控详情页面，可查看统计数据。



10.6.10. 文件夹操作

在自定义监控页，您还可以进行以下操作：

- 通过文件夹对监控指标进行分组管理。
- 通过批量移动对分组后的内容进行快速移动。
- 通过排序，更改监控内容的展示方式。
- 通过告警历史，可以查看监控项的告警历史信息。
- 通过文件搜索，可以对自定义监控页面的文件进行模糊搜索。

添加文件夹

1. 登录控制台，从左侧导航栏选择 **自定义监控 > 添加文件夹**。
2. 在 **新建文件夹** 窗口，填写文件夹名，单击 **确定** 完成添加。

合并文件夹

1. 在 **自定义监控** 页，选择目标文件夹，单击 **操作 > 合并内容到另个文件夹**。



2. 选择监控项要移至的目标文件夹，单击 **下一步**。
3. 确认目标文件夹信息无误后，单击 **确认**。



删除文件夹

选择目标文件夹，单击 **操作**，然后单击 **删除** > **确定**。

> 说明
> 仅空文件夹才可以删除。若文件夹下包含有监控配置，需先将监控配置删除。

批量移动
对分组后的文件夹进行快速移动和重新分组。

排序
可以更改自定义监控页面内容的展示方式。目前支持根据名称、类型进行排序。

告警历史
可以查看目标监控项的告警历史信息，包括告警时间、数据源、告警类型、规则详情等信息。在告警历史页面，还可以自定义设置搜索的时间范围。

文件搜索
可以对自定义监控页面的文件进行模糊查询。

10.7. 应用监控

10.7.1. 应用监控概述

应用监控提供应用分析能力，通过一个入口总览分析应用相关的所有监控数据。当应用通过 SOFAStack CAFE（Cloud Application Fabric Engine）云应用引擎部署成功后，RMS 会自动监控应用，并以各种维度和视角的监控大盘展示，满足应用研发或运维人员对应用的监控，进而保证应用运行稳定性。

背景信息
RMS 应用监控支持容器应用和经典应用，并将应用的指标由内而外分为如下维度。

维度	示例	说明
服务指标	Error、Service、SAL、CAL、DAL	• **框架配置**： 基于 SOFA/SOFABoot 框架开发的应用，PV、Service、SAL、SQL、Cal、CE Thread 等监控项会自动基于默认框架下的日志路径采集指标，只需开启监控即可。 具体框架日志路径可参考 服务指标。 • **自定义监控配置**： ○ 对于非 SOFA/SOFABoot 框架开发的应用：需要为每个监控项配置采集日志路径和列值，比如 Error、Dal。 ○ 对于基于 SOFA/SOFABoot 框架开发的应用：如果不希望使用默认日志路径监控的，也可以切换为 **自定义监控配置**。
系统指标	CPU、内存、磁盘	容器应用的系统指标通过 asar 采集主机上容器的指标是容器的指标，经典应用的系统指标通过 asar 采集主机上进程的指标。

维度	示例	说明
JVM 指标	GC、线程	Java 应用启动后，默认会在 `/tmp/hsperfdata_${username}` 目录下以该进程的 id 为文件名新建文件，并在该文件中存储 jvm 运行的相关信息，其中 userName 为当前的用户名，`/tmp/hsperfdata_${username}` 目录会存放该用户所有已经启动的 Java 进程信息。 RMS 的 JVM 指标通过解析 `/tmp/hsperfdata_${username}/${pid}`（例如 `/tmp/hsperfdata_admin/1234`）来实现。
端口探测	/	端口探活是从 Agent 发起对应用服务的端口检测。 例如，某一服务暴露 12200 端口，正常情况下通过运行 `ping` 命令即可连通，但如果服务出现宕机，则该端口无法连通。 通过端口探测，可周期性判断应用服务的存活状态，从而保证应用可提供对外服务。

为了保证应用高可用，一般是多实例部署或是单元化部署，不同应用实例（单机）可能会分布在不同的 IDC（Internet Data Center）机房以及不同 LDC（Logic Data Center）逻辑机房，所以 RMS 将每个维度的指标数据都分成三个视角（即单机视角、IDC 视角、LDC 视角）进行聚合，从 IDC、LDC 或单机的空间分布和时间分布上对数据进行对比分析。

功能入口

1. 登录实时监控产品控制台。
2. 在左侧导航栏上，单击 **应用监控** 进入应用监控页面。
   应用监控页面提供 **我的应用** 和 **全部应用** 两个类别的应用汇总信息。目前仅展示在当前 workspace 下具有实例的应用。
   - 全部应用：在 CAFE 中创建并发布应用后，RMS 会将所有发布的应用同步到 **全部应用** 中。
   - 我的应用：在 **全部应用** 页签下，选择某个应用，单击 **更多** > **收藏应用** 即可将该应用添加到 **我的应用**。

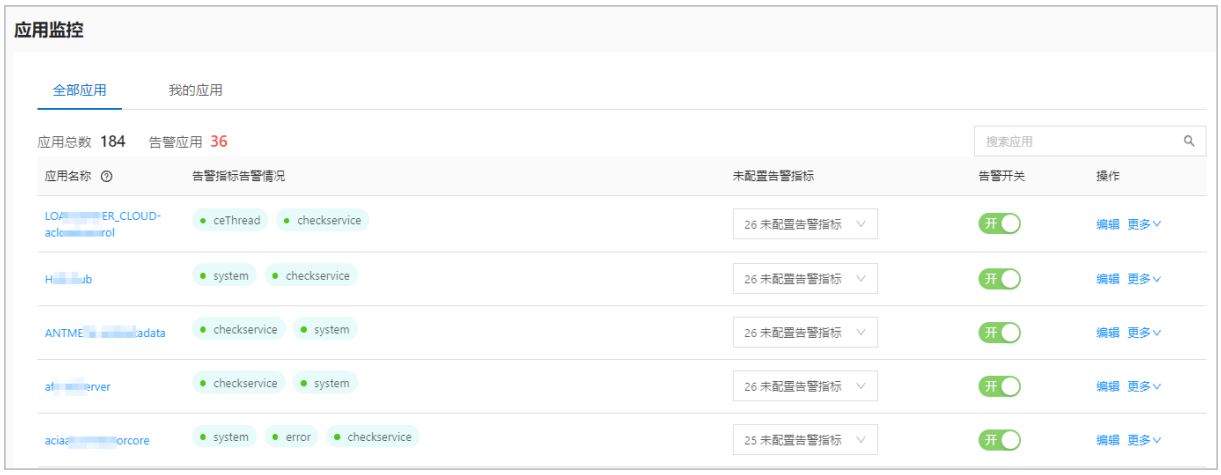
全部应用

单击 **全部应用** 页签，可以查看以下信息：

类别	说明
概述信息	• 应用总数 • 告警应用数

类别	说明

列表信息	• 应用名称
	• 告警指标告警情况
	• 告警规则数
	• 未配置告警指标
	• 告警开关
	• 操作：
	  ○ 编辑：单击 **编辑** 会跳转到应用的配置界面。
	  ○ 更多：收藏应用、订阅告警、查看告警历史、查看告警规则、暂停告警。



我的应用

单击 **我的应用** 页签，可以查看以下信息：

类别	说明
概述信息	• 关注总数 • 健康应用数 • 异常应用数 • 健康度：健康应用占总应用的比例。
列表信息	• 应用名称 • 告警指标告警情况 • 告警规则数 • 未配置告警指标 • 告警开关 • 操作：订阅告警、查看告警历史、查看告警规则。

在该页面上，您还可以执行如下操作：

- 批量添加关注：支持批量添加关注应用。
- 批量取消关注：支持批量取消对部分应用的关注。

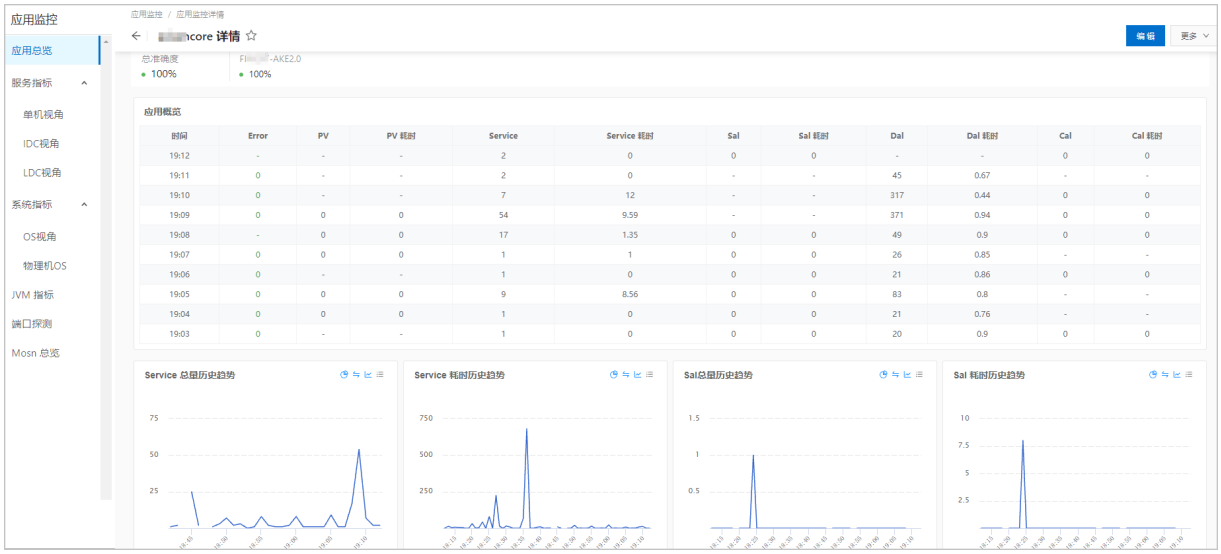
- 复制关注的应用：可以复制关注的应用，便于后续处理，例如发送给好友。



10.7.2. 应用总览

应用监控提供了应用分析能力，将应用相关的 Error、Service、SAL、CAL、DAL、OS、JVM、CE Thread、MOSN、端口等服务指标，和 CPU 、内存、磁盘等系统指标进行聚合透出，从 **IDC/LDC/单机** 的空间分布和时间分布上进行对比分析，一个入口总览分析应用相关的所有监控数据。主要监控指标如下：

- Error 系统错误量
- PV 页面访问量
- Service 被外部调用的服务量
- SAL 调用外部的服务量
- DAL 数据库访问
- SQL 耗时监测
- 系统指标监控
- 可用性（端口）
- MOSN 指标



设置查询条件

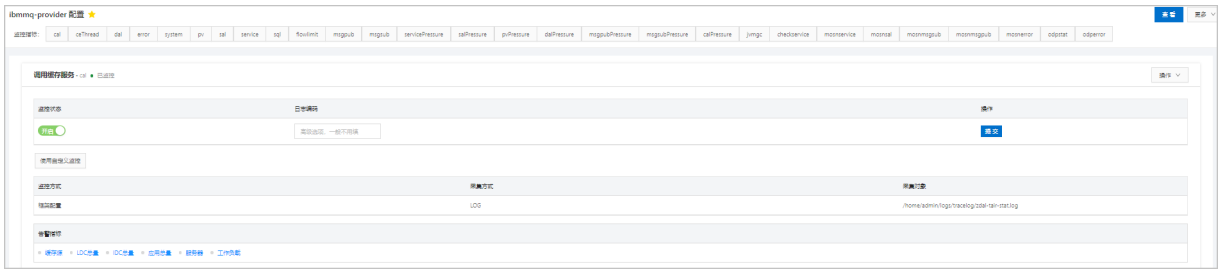
支持设置如下查询条件。

- 输入或重置过滤的正则表达式。
- 选择对比曲线和时间跨度。

- 回放：回放指定时间段的数据。
- 自动更新：定时刷新页面数据。

编辑总览页

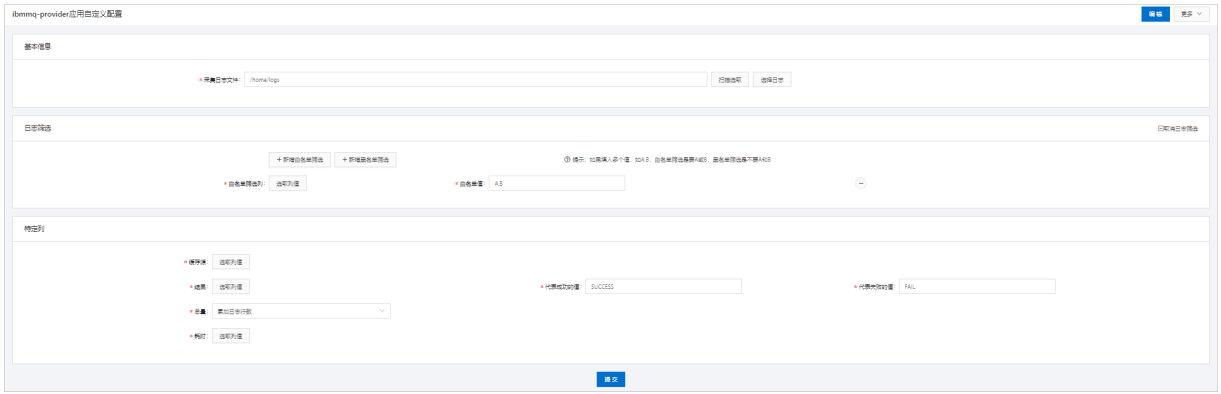
单击 **编辑**，可进行自定义配置和添加告警套餐，示例如下：



> 说明
> 系统会根据日志的编码自动适配，一般不用手动填写 **日志编码** 。

自定义配置

自定义配置示例如下：



主要配置的指标包括：

- 采集日志文件
- 新增白名单或黑名单
- 缓存源
- 结果
- 总量
- 耗时

主要的监控指标详情见下表：

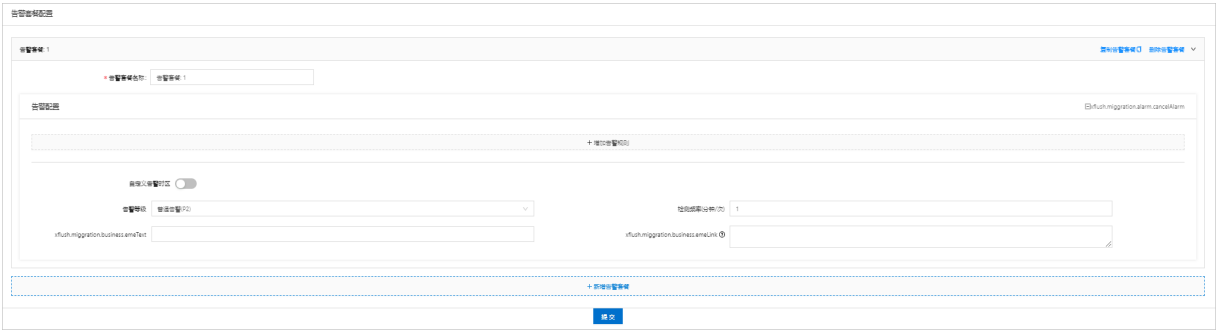
指标名称	指标说明
调用缓存服务-cal	用来监控应用访问缓存的相关指标。配置要求如下： • SOFA 框架中没有默认日志。 • 自定义配置中可以使用自己设定的日志文件，要求日志文件中必须有：数据源、结果、耗时、计数。

指标名称	指标说明
调用 DB 服务-dal	数据库访问。用来监控应用访问数据库的相关指标，配置要求如下： • SOFA 框架中默认使用 `/home/admin/logs/tracelog/zdal-db-stat.log` 。 • 自定义配置中可以使用自己设定的日志文件，要求日志文件中必须有：数据源、库、表、结果、耗时、计数。
Error-error	系统错误量。用来监控应用中错误日志数量，统计错误变化趋势。 在应用监控中，可直接单击错误查看相关日志。配置要求如下： • SOFA 框架中默认使用 `/home/admin/logs/stderr.log` 或 `/home/admin/logs/{app}/common-error.log` 日志文件。 • 自定义配置中可以使用自己设定的日志文件，对文件中的行数按分钟进行计数。
系统监控-system	用来监控应用本身所在虚拟机或 ECS 的资源情况。 包括 CPU、Load、Memory 等情况。配置要求如下： • 无需配置日志。 • 打开开关即可，监控 Agent 会自动获取对应数据。
PV-pv	页面访问量。用来监控应用中页面被调用次数相关指标。配置要求如下： • SOFA 框架中默认使用 `/home/admin/logs/tracelog/sofa-mvc-stat.log` 。 • 自定义配置中可以使用自己设定的日志文件，要求日志文件中必须有：url、结果、耗时、访问计数这几项指标。
调用外部服务-sal	用来监控应用调用其他服务的指标。配置要求如下： • SOFA 框架中默认使用 `/home/admin/logs/tracelog/sofa-mvc-stat.log` 。 • 自定义配置中可以使用自己设定的日志文件，要求日志文件中必须有：机房、目标应用、接口、方法、结果、耗时、计数。
应用服务被调用-service	被外部调用的服务量。用来监控应用中服务被调用情况。配置要求如下： • SOFA 框架中默认使用 `/home/admin/logs/tracelog/sofa-mvc-stat.log` 。 • 自定义配置中可以使用自己设定的日志文件，要求日志文件中必须有：机房、目标应用、接口、方法、结果、耗时、计数。
SQL 耗时检测-sql	用来监控应用访问数据库的 SQL 执行情况，包括查找慢 SQL。配置要求如下： • SOFA 框架中默认日志：`/home/admin/logs/tracelog/zdal-db-digest.log` 。 • 自定义配置中可以使用自己设定的日志文件，要求日志文件中必须有：数据库、SQL、耗时、事务ID。

指标名称	指标说明
Flow Limit -flowlimit	流量监控指标。
消息发布-msgpub	消息发布监控指标。
消息订阅-msgsub	消息订阅监控指标。
JVM 监控-jvmgc	JVM 监控指标。
端口检测-checkservice	端口检测监控指标。
MOSN 应用服务被调用-mosnservice	MOSN 应用服务被调用时的监控指标。 说明 mosnservice 即 Mesh Sidecar 指标。
MOSN 调用外部服务-mosnsal	MOSN 调用外部服务时的监控指标。
MOSN 消息订阅-mosnmsgsub	MOSN 消息订阅时的监控指标。
MOSN 消息发布-mosnmsgpub	MOSN 消息发布时的监控指标。
MOSN Error-mosnerror	MOSN 错误信息监控指标。
OB Proxy 指标-odpstat	OB（OceanBase）代理监控指标。
ODP 异常-odperror	ODP（数据访问代理，Open Database Proxy）异常监控指标。
MOSN Metric 数据	MOSN Metric 数据监控指标。
ibmmqput-ibmmqput	ibm 消息 put 请求的监控指标。
ibmmqget-ibmmqget	ibm 消息 get 请求的监控指标。

告警套餐

单击告警指标中某个具体指标，可以添加告警套餐，示例如下：



主要配置的告警套餐信息包括：

- 告警套餐名称
- 自定义告警时区
- 告警等级

不同的监控指标可以应用的告警条件不同，监控指标与告警指标的对应关系如下表所示：

监控指标	告警条件
调用缓存服务-cal	缓存源、LDC 总量、IDC 总量、应用总量、服务器、应用实例
调用 DB 服务-dal	应用总量、LDC 总量、IDC 总量、数据源、服务器、应用实例
Error-error	应用总量、LDC 总量、IDC 总量、服务器、应用实例
系统监控-system	应用总量、LDC 总量、IDC 总量、服务器、应用实例、serviceidc、单机IP
PV-pv	应用总量、LDC 总量、IDC 总量、URL 总量、服务器、应用实例
调用外部服务-sal	应用总量、IDC 总量、LDC 总量、目标应用、接口方法明细、服务器、应用实例、下游应用接口方法明细
应用服务被调用-service	应用总量、LDC 总量、IDC 总量、接口方法明细、来源应用、服务器、应用实例
SQL 耗时检测-sql	应用 Top
Flow Limit -flowlimit	应用总量、LDC 总量、IDC 总量、服务器
消息发布-msgpub	应用总量、LDC 总量、IDC 总量、Topic Eventcode 总量、服务器、应用实例
消息订阅-msgsub	应用总量、LDC 总量、IDC 总量、Topic Eventcode 总量、服务器、应用实例
JVM 监控-jvmgc	单机、应用总量、机房、逻辑机房、应用实例
端口检测-checkservice	应用总量、服务器、应用端口、单机 Port

监控指标	告警条件
MOSN 应用服务被调用-mosnservice	应用总量、LDC 总量、IDC 总量、服务器、接口方法明细
MOSN 调用外部服务-mosnsal	应用总量、IDC 总量、LDC 总量、服务器、下游应用接口方法明细
MOSN 消息订阅-mosnmsgsub	应用总量、LDC 总量、IDC 总量、服务器、Topic Eventcode 总量
MOSN 消息发布-mosnmsgpub	应用总量、LDC 总量、IDC 总量、服务器、Topic Eventcode 总量
MOSN Error-mosnerror	应用总量、LDC 总量、IDC 总量、服务器
OB Proxy 指标-odpstat	应用总量、IDC 总量、Idc、单机、data source, sql type、错误码总量
ODP 异常-odperror	应用总量、LDC 总量、IDC 总量、服务器
MOSN Metric 数据	无
ibmmqput-ibmmqput	应用总量、队列维度
ibmmqget-ibmmqget	应用总量、队列维度

订阅告警

在该页面，可以新增订阅告警，也可以对已有的订阅告警进行编辑。



添加订阅告警时，需配置如下项：

- **订阅人**：选择订阅人。
- **告警等级**：支持 **全部订阅**、**P0**、**P1**、**P2**、**P3**、**P4**。
- **通知降频**：包含 **开启** 和 **关闭** 两个开关。默认开启。若是持续发生的告警，按照每隔 1 分钟、2 分钟、5 分钟、10 分钟、30 分钟、60 分钟的频率进行降级通知。其中每隔 1 分钟、2 分钟通知两次，每隔 5 分钟通知三次，每隔 10 分钟、30 分钟通知五次，直到每隔一个小时通知一次。

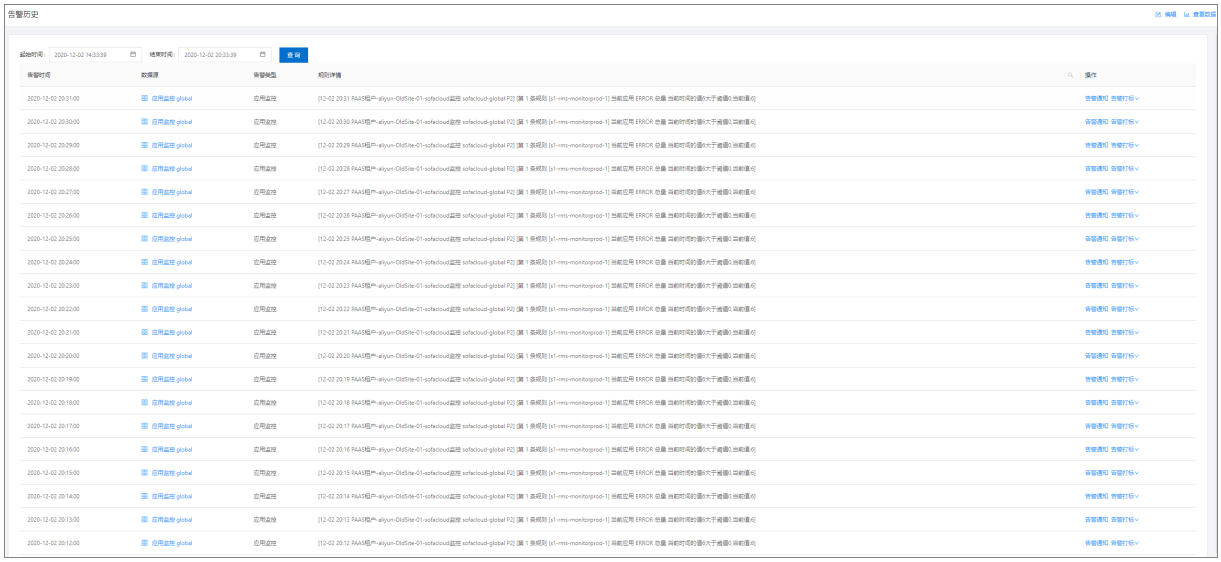
暂停/恢复告警

在该页面，可以对告警暂停时间及关闭原因进行选择，示例如下：



告警历史

在该页面，可以添加查询告警历史，示例如下：



10.7.3. 服务指标

应用监控提供了应用分析能力，服务指标将应用相关的 Error、Service、SAL、CAL、DAL 等服务指标进行聚合透出，从 **IDC（机房）/LDC（单元化）/单机** 的空间分布和时间分布上进行对比分析，一个入口总览分析应用相关的所有监控数据。

单击控制台左侧边栏 **应用监控**，选择 **我的应用** 或 **全部应用**，单击某个应用后，会进入应用监控详情页面。

从服务维度展示应用监控结果时，您可以分别从 **IDC（机房）/LDC（单元化）/单机** 三个视角配置和监测以下监控指标：

指标名称	说明	单位	配置要求

指标名称	说明	单位	配置要求

cal	缓存访问，用来监控应用访问缓存的相关指标。	次	SOFA 框架中默认使用
			`/home/admin/logs/tracelog/zdal-tair-stat.log`
			日志文件。自定义配置中可以使用自己设定的日志文件，要求日志文件中必须有：数据源、结果、耗时、计数。
			样例日志为：
			`2015-04-19 21:56:16.627,Test,Tair,1,0,Y,T 2015-04-19 21:56:16.628,Test,Tair,1,20,N,F`
			样例日志格式如下：
			• 日志打印时间
			• 当前应用名
			• 逻辑 Tair 名
			• 本段时间内的请求数量
			• 本段时间内的请求总耗时
			• 结果信息（Y/N）
			• 全链路压测标志（T/F）

error	应用错误。 用来监控应用中错误日志数量，统计错误变化趋势。 应用监控中可直接单击错误查看相关日志。	次	SOFA 框架中默认使用
			`/home/admin/logs/stderr.log`
			或
			`/home/admin/logs/{app}/common-error.log`
			日志文件。自定义配置中可以使用自己设定的日志文件，对文件中的行数按分钟进行计数。

指标名称	说明	单位	配置要求
pv	页面访问量。用来监控应用中页面被调用次数。	次	SOFA 框架中默认使用 `/home/admin/logs/tracelog/sofa-mvc-stat.log` 。自定义配置中可以使用自己设定的日志文件，要求日志文件中必须有 **url**、**结果**、**耗时** 和 **访问计数** 几项值。 样例日志为： `2014-09-01 00:03:22.559,tbapi,http://tbapi.alipay.com/trade/batch_payment.htm,GET,2,11,Y,F` 样例日志格式如下： • 日志打印时间 • 当前应用名 • 请求 URL • 请求方法 • 本段时间内请求次数 • 本段时间内的请求总耗时 • 请求结果（1 开头和 2 开头的结果码表示成功，302 表示的重定向算成功，其他代表失败） • 压测标记

指标名称	说明	单位	配置要求

service	应用服务被调用，用来监控应用中服务被调用情况。在一个面向服务架构（SOA）中，系统之间有复杂的服务依赖关系。Service 表示某个系统所暴露的服务被其他系统调用的统计数值。	次	SOFA 框架中默认使用 `/home/admin/logs/tracelog/rpc-server-stat.log`。自定义配置中可以使用自己设定的日志文件，要求日志文件中必须有：来源应用、接口、方法、结果、耗时、计数。
			样例日志为：
			`2014-06-19 17:14:02.186,client,client,com.alipay.cloudenginetest.services.SofaApiWebReferenceLocalFalseTrService:1.0,service_method,1,7,Y,T,GZ00B`
			`2014-06-19 17:15:02.190,client,client,com.alipay.cloudenginetest.services.SofaApiWebReferenceLocalFalseTrService:1.0,service_method,1,12,Y,F,GZ00B`
			样例日志格式如下：
			• 日志打印时间
			• fromApp
			• toApp（即 currentApp）
			• 服务名
			• 方法名
			• 本段时间内的被调用次数
			• 本段时间内的请求处理总耗时
			• 结果（Y/N）
			• 全链路压测标志（T/F）
			• 调用者 Zone

指标名称	说明	单位	配置要求
sal	调用外部的服务（ Service Access Layer，SAL），用来监控应用调用其他服务的指标。在一个面向服务架构（SOA）中，系统之间有复杂的服务依赖关系。SAL 表示某个系统调用其他系统暴露的服务的统计数值。	次	SOFA 框架中默认使用 `/home/admin/logs/tracelog/rpc-client-stat.log` 。自定义配置中可以使用自己设定的日志文件，要求日志文件中必须有：机房、目标应用、接口、方法、结果、耗时、计数。 样例日志为： `2014-06-19 17:14:02.186,client,client,com.alipay.cloudenginetest.services.SofaApiWebReferenceLocalFalseTrService:1.0,service_method,1,79,Y,T,RZ00B 2014-06-19 17:09:02.146,client,,com.alipay.core.service.lazy.LazyPublisherWsService:1.0,service_method,1,26,N,F,RZ00B` 样例日志格式如下： • fromApp（即 currentApp） • toApp • 服务名 • 方法名 • 本段时间内调用次数 • 本段时间内的调用总耗时 • 结果（Y/N） • 全链路压测标志（T/F） • 目标 Zone
ceThread	容器线程繁忙 running 状态的线程数	个	应用日志（自定义日志文件） 配置于 `CE Thread 容器线程繁忙` 。样例日志为： `2014-07-10 00:01:01,316 [CE-RES-MONITOR] - (0,1,79,80,) 2014-07-10 00:01:31,316 [CE-RES-MONITOR] - (0,0,80,80,)` 格式为：blockQueue active idle poolSize

指标名称	说明	单位	配置要求
dal	数据库服务（Data Access Layer，DAL），用来监控应用访问数据库的相关指标。大部分应用系统会涉及数据库访问，DAL 表示某个系统对数据库访问调用的统计数值。	次	SOFA 框架中默认使用 `/home/admin/logs/tracelog/zdal-db-stat.log` 。每一分钟打印一次统计日志。自定义配置中可以使用自己设定的日志文件，要求日志文件中必须有：数据源、库、表、结果、耗时、计数。 样例日志为： `2015-05-11 20:03:30.487,Test,DataSourceName,LogDatabaseName,tableName,mysql,2,27,N,F 2015-05-11 20:03:30.487,Test,DataSourceName,LogDatabaseName,tableName,mysql,1,0,Y,` 样例日志格式如下： • 日志打印时间 • 当前应用名 • 数据源名称 • 逻辑库名称 • 表名 • 数据库类型 • 本段时间内的请求数量 • 本段时间内的请求总耗时 • 结果信息（Y/N） • 全链路压测标志（T/F）
sql	SQL 耗时监测 ，用来监控应用访问数据库的 SQL 执行情况，包括查找慢 SQL。		SOFA 框架中默认日志： `/home/admin/logs/tracelog/zdal-db-digest.log` 。自定义配置中可以使用自己设定的日志文件，要求日志文件中必须有：数据库、SQL、耗时、事务 ID。
flowlimit	限流	次	框架配置日志。文件采集路径 `/home/admin/logs/{app}/guardian/guardian-limit-stat.log` 。

指标名称	说明	单位	配置要求
msgpub	消息推送	次	框架配置日志。文件采集路径 `/home/admin/logs/tracelog/msg-pub-stat.log` 。 样例日志为： `2018-12-07 16:21:40.168,,TP-openmq-chengyi-1,MQ-PRODUCER-GROUP,8,307,Y,F` 样例日志格式如下： • 日志打印时间 • currentApp • topic • pubGroupId • 本段时间内的消息发送数量 • 本段时间内的消息发送 • 总耗时 • 结果信息（Y/N） • 全链路压测标志（T/F）
msgsub	消息订阅	次	框架配置日志。文件采集路径 `/home/admin/logs/tracelog/msg-sub-stat.log` 。 样例日志为： `2018-12-07 16:23:54.187,,TP-openmq-chengyi-1,S_openmq_chengyi_0,16,67,N,FF` 样例日志格式如下： • 日志打印时间 • currentApp • topic • subGroupId • 本段时间内的消息接收数量 • 本段时间内的消息接收 • 总耗时 • 结果信息（Y/N） • 全链路压测标志（T/F）
servicePressure	压测应用服务被调用	次	框架配置日志，该监控项不支持自定义。文件采集路径 `/home/admin/logs/tracelog/shadow/rpc-server-stat.log` 。

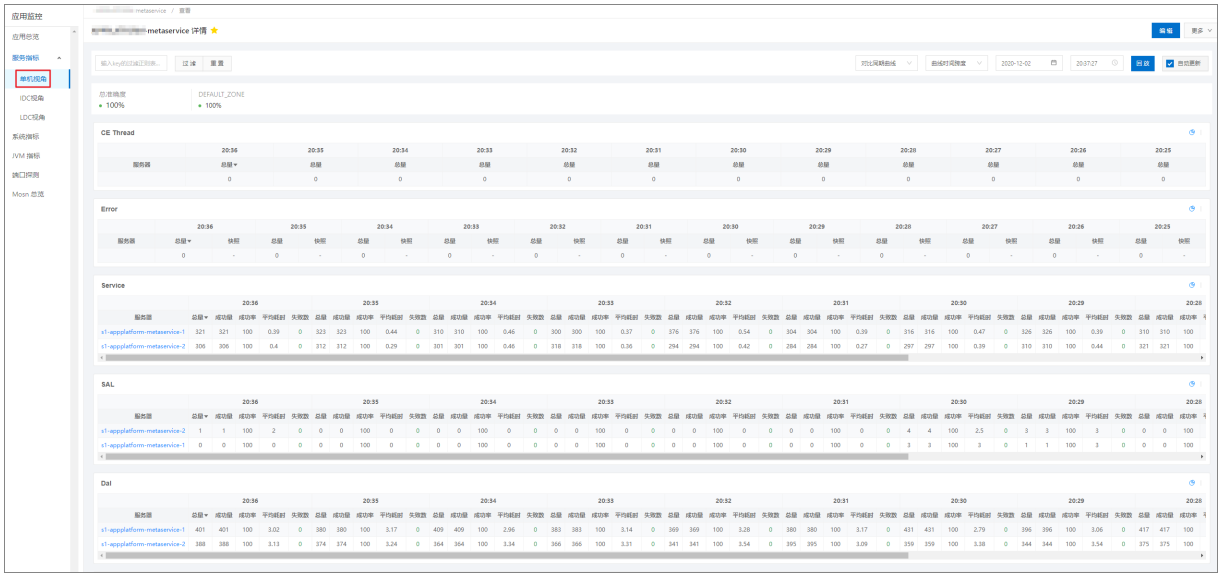
指标名称	说明	单位	配置要求
salPressure	压测调用外部服务	次	框架配置日志，该监控项不支持自定义。文件采集路径 `/home/admin/logs/tracelog/shadow/rpc-client-stat.log` 。
pvPressure	压测 PV	次	框架配置日志，该监控项不支持自定义。文件采集路径 `/home/admin/logs/tracelog/shadow/sofa-mvc-stat.log` 。
dalPressure	压测调用 DB 服务	次	框架配置日志，该监控项不支持自定义。文件采集路径 `/home/admin/logs/tracelog/shadow/zdal-db-stat.log` 。
calPressure	调用缓存服务	次	框架配置日志，该监控项不支持自定义。文件采集路径 `/home/admin/logs/tracelog/shadow/zdal-tair-stat.log` 。
msgpubPressure	压测消息发布	次	框架配置日志，该监控项不支持自定义。文件采集路径 `/home/admin/logs/tracelog/shadow/msg-pub-stat.log` 。
msgsubPressure	压测消息订阅	次	框架配置日志，该监控项不支持自定义。文件采集路径 `/home/admin/logs/tracelog/shadow/msg-sub-stat.log` 。

指标名称	说明	单位	配置要求
mosnservice	MOSN 应用服务被调用	次	框架配置日志。文件采集路径 `/home/admin/logs/tracelog/mosn/rpc-server-stat.log` 。 插件类型为：SPM 日志格式如下： • 日志打印时间 • fromApp • toApp（即 currentApp） • 服务名 • 方法名 • 本段时间内的被调用次数 • 本段时间内的请求处理总耗时 • 结果（Y/N） • 全链路压测标志（T/F） • 调用者 Zone
mosnsal	MOSN 调用外部服务	次	框架配置日志。文件采集路径 `/home/admin/logs/tracelog/mosn/rpc-client-stat.log` 。 插件类型为：SPM 日志格式如下： • 日志打印时间 • fromApp（即 currentApp） • toApp • 服务名 • 方法名 • 本段时间内调用次数 • 本段时间内的调用总耗时 • 结果（Y/N） • 全链路压测标志（T/F） • 目标 Zone
mosnmsgsub	MOSN 消息订阅	次	框架配置日志。文件采集路径 `/home/admin/logs/tracelog/mosn/msg-sub-stat.log` 。

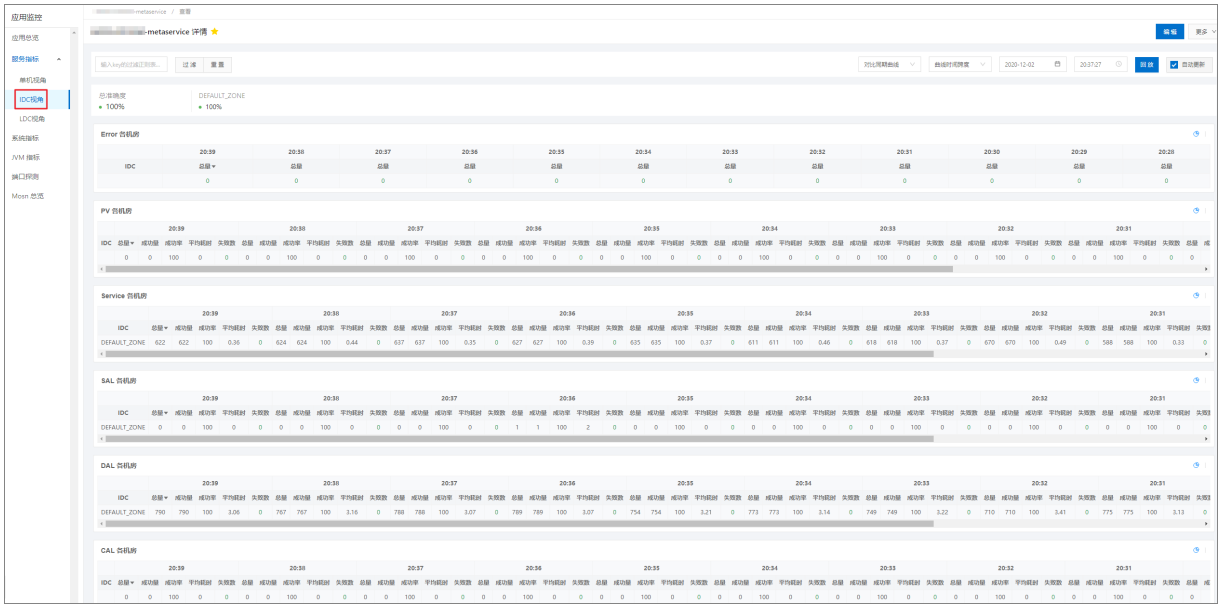
指标名称	说明	单位	配置要求
mosnmsgpub	MOSN 消息发布	次	框架配置日志。文件采集路径 `/home/admin/logs/tracelog/mosn/msg-pub-stat.log` 。 插件类型为：SPM 日志格式如下： • 日志打印时间 • currentApp • topic • eventCode • pubGroupId • 本段时间内的消息发送数量 • 本段时间内的消息发送总耗时 • 结果信息（Y/N） • 全链路压测标志（T/F）
mosnerror	MOSN 的错误数	次	可选择自定义配置或框架配置日志，框架配置日志的文件采集路径为：`/home/admin/logs/mosn/default.log` 、`/home/admin/logs/mosn/antvip.error.log` 、`/home/admin/logs/mosn/drm.error.log` 、`/home/admin/logs/mosn/endpoint.error.log` 、`/home/admin/logs/mosn/registry.error.log` 、`/home/admin/logs/mosn/zoneclient.error.log` 、`/home/admin/logs/mosn/alert.mirror.log` 。 插件类型为：ERROR
mosnErrorCode	MOSN 的错误码	次	可选择自定义配置或框架配置日志，框架配置日志的文件采集路径为：`/home/admin/logs/mosn/alert.default.log` 。 插件类型为：MM

指标名称	说明	单位	配置要求
odpstat	ob proxy 指标	次	框架配置日志，该监控项不支持自定义。文件采集路径 `/home/admin/logs/obproxy/log/obproxy_stat.log` 。 插件类型为：MM 日志格式如下： • 日志打印时间 • 当前应用名 • 数据源（trade_ds）名称，（预留，1.x 固定一个，2.x 为逻辑库的名字） • 物理库（master00）名称，换成 shema 信息（cluster:tenant:database） • 数据库类型 （OB/RDS） • SQL 类型（CRUD） • 执行结果（success/failed） • 错误码（succ时为空） • 总请求数量 • 30 ms ~ 100 ms 请求数量 • 100 ms ~ 500 ms 请求数量 • 大于 500 ms 请求数量 • 执行总耗时（ms，包括内部 sql） • 预执行时间 • 数据库执行时间
odperror	ODP 异常	次	框架配置日志，该监控项不支持自定义。文件采集路径 `/home/admin/logs/obproxy/log/obproxy_error.log` 。 插件类型为：ERROR

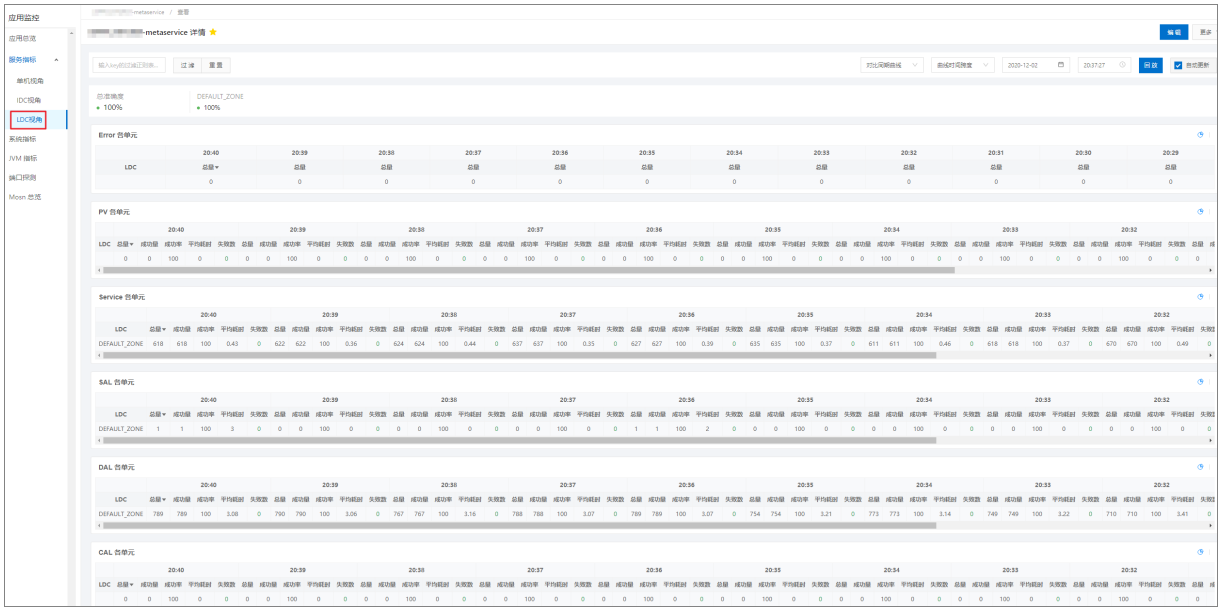
单机视角



IDC 视角



LDC 视角



10.7.4. 系统指标

系统指标用来监控应用本身所在虚拟机或 ECS 的资源情况。 包括 CPU、Load、Memory 等情况。您无需配置日志，打开开关即可，监控 agent 会自动获取对应数据。

单击控制台左侧边栏 **应用监控**，选择 **我的应用** 或 **全部应用**，单击某个应用后，会进入应用详情页面。本文主要介绍应用详情在可视化展示时，对应的系统指标详情。关于监控信息的自定义配置等更多详情，请参见 应用总览。



RMS 支持的系统指标说明如下：

指标名称	说明
cpu_total_cores	容器核心数
cpu_util	内存使用率，由 `user+sys+nice+guset+hirq+sirq` 组成，能反应当前系统 CPU 的利用率。

指标名称	说明
cpu_sys	内核态时间比例
cpu_steal	被偷走的 CPU 时间的占比
cpu_wait	io wait 时间的比例
cpu_user	用户态时间的比例
load_load1	1 min 内 load 指标
load_load5	5 min 内 load 指标
load_load15	15 min 内 load 指标
mem_util	内存使用率
mem_used	total - buff - cache - free 的内存的量，特别注意不包含 cache。
mem_total	总的内存
traffic_bytin	网卡 rx 方向的流量大小，单位 bytes/s。
traffic_bytout	网卡 tx 方向的流量大小，单位 bytes/s。
traffic_pktin	网卡 rx 方向的 pps。
traffic_pktout	网卡 tx 方向的 pps。
traffic_pktdrp	出入方向的丢包数和。在容器内（采用 veth pair）这个数据一般为 0。
traffic_pkterr	出入方向的错误包数和。在容器内（采用 veth pair）这个数据一般为 0。
tcp_active	active open/s，主动发起连接的每秒次数。
tcp_lisove	listen overflow，这种是由于 listen accept 队列满导致被丢弃的数量。
tcp_AtmpFail	AttemptFails，发起连接失败的每秒次数，比如半连接队列满导致 syn 包被丢弃等。
tcp_CurrEstab	Current Established connection，当前 TCP 连接数。
tcp_pasive	passive open/s，被动生成连接的每秒次数。
tcp_retran	retransmit ratio，根据重传 segment 和总的 segment，计算出来的重传率。
udp_idgm	主机受到的 udp 包的个数/s 与 odgm 一起来监控主机的 udp 报文的流量。
udp_odgm	主机发送的 udp 包的个数/s。

指标名称	说明
udp_idmerr	udp 错误包的数目，包括 udp csum 错误等。一般用来监控 udp 通信的异常。
udp_noport	未知 udp 端口号的包个数/s，主机侧没有进程监听在该端口导致。
partition_util	使用的磁盘空间的比率，单位 %。
partition_bfree	空闲的空间，单位 byte。该挂载点下剩余可用的磁盘空间。
partition_bused	使用的空间，单位 byte。已经使用的磁盘空间。
partition_btotl	总共的磁盘空间，单位 byte。该挂载点总共可用的磁盘空间。
partition_ifree	文件系统可用的 inode 个数，ext4 文件系统有总的 inode 的使用个数限制。
partition_itotl	总的可用的 inode 的个数。
partition_iutil	使用的 inode 个数占总的可使用个数的占比，总体使用率，单位 %。
io_wio	写 IO 每秒次数
io-rio	读 IO 每秒次数
io_rbytes	读 IO 每秒 byte 数
io_wbytes	写 IO 每秒 byte 数
io_rqsize	request sector/IO ，每个 IO 请求的大小，单位为 KB。
io_qusize	queue size，当前 IO 队列中的IO请求数量。
io_await	average response time for I/O requests，每个 IO 从发起到结束平均消耗时间，单位 ms。
io_svctm	service time，实际下发到设备到返回的时间，不包含队列中等待的时间。

10.7.5. JVM 指标

单击控制台左侧边栏 **应用监控**，选择 **我的应用** 或 **全部应用**，单击某个应用后，会进入应用详情页面。本文主要介绍应用详情在可视化展示时，对应的 JVM指标信息。

> 说明
> 关于监控信息的自定义配置等操作，请参见 应用总览。

下文是 RMS 提供的 JVM 监控指标，但产品界面只将最常用的指标进行图形化展示。



GC

指标名称	说明
ygc_count	ygc 次数
ygc_time	ygc 总耗时，单位为秒
fgc_count	fgc 总次数（CMS 不准）
fgc_time	fgc 总耗时（CMS 不准），单位为秒
tlab_alloc	在 tlab 分配的总大小

Runtime

指标名称	说明
thread_started	线程（被 start 过的）数量
thread_live	存活的线程的数量
thread_daemon	daemon 线程数量
thread_live_peak	活的线程的峰值数量
safepoints	safepoints 次数
safepoint_time	safepoint 时间
safepoint_sync_time	safepoint sync 阶段花费的总时间
application_time	应用运行总时间
parks	park 的次数

指标名称	说明
notificcations	notify 的次数
mon_extant	待补充
inflations	锁膨胀的次数
futile_wakeups	待补充
deflations	待补充
contented_lock_attemps	待补充

eden 区

指标名称	说明
eden_used	Eden 区使用量
eden_capacity	当前 Eden 区容量，可以继续申请扩大
eden_max	Eden 区总容量，无法再扩

Old 区

指标名称	说明
old_used	Old 区使用量
old_capacity	当前 Old 区容量，可以继续申请扩大
old_max	Old 区总容量，无法再扩

metaspace 区

metaspace 区指标很多，绝大多数情况下，关注 meta_used/meta_capacity/meta_util/meta_rutil 即可。另外 chunks 相关指标必须将 jdk 升级到 8.7.12_fp1 版本及以上。

指标名称	说明
meta_used	Metaspace 的使用量
meta_capacity	当前 Metaspace 的容量，可以继续申请扩大
meta_util	meta_used/meta_capaticy，不考虑碎片的使用率
meta_rutil	meta_chunks_used/meta_capacity，考虑了碎片的使用率
compressedclass_used	包含在 sun.gc.metaspace.used 中，表示 compressed class 区域

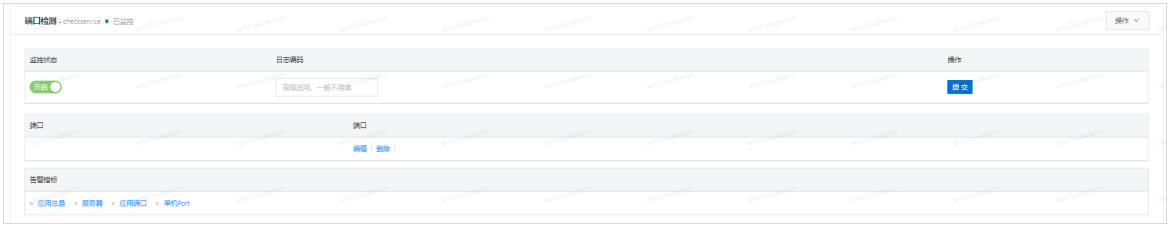
指标名称	说明
compressedclass_capacity	包含在 sun.gc.metaspace.capacity 中，表示 compressed class 区域
meta_chunks_free	Metaspace 空闲 Chunks 的总容量
meta_chunks_used	Metaspace 使用中的 Chunks 的总容量，接近 MaxMetaspace 时触发 FGC
compressedclass_chunks_free	包含在 com.alibaba.metaspace.freeChunksTotal 中，表示 compressed class 区域
compressedclass_chunks_used	包含在 com.alibaba.metaspace.usedChunksTotal 中，表示 compressed class 区域

10.7.6. 端口探测

端口探活是从 agent 发起对应用服务的端口检测。例如，一个服务暴露 12200 端口，正常情况下是可以 ping 通，但如果服务宕机了，那么这个端口就不通了。通过端口探测可周期性的判断应用服务的存活状态，以此保证应用可提供对外服务。本文主要介绍应用详情在可视化展示时，如何自定义监测端口。

操作步骤

1. 单击控制台左侧边栏 **应用监控**，选择 **我的应用** 或 **全部应用**，单击某个应用后，会进入应用详情页面。
2. 在应用详情页面，选择 **端口探测**，单击右上角 **编辑** 按钮，进入配置详情页面，选择 **端口检测**，可以自定义待监控端口，示例如下：



3. 单击 **端口检测** 中的 **编辑** 链接，即可在弹出页面填入待监测端口，示例如下：



10.7.7. MOSN 总览

单击控制台左侧边栏 **应用监控**，选择 **我的应用** 或 **全部应用**，单击某个应用后，会进入应用详情页面。本文主要介绍应用详情在可视化展示时，对应的 MOSN 指标详情。

> 说明
> 关于监控信息的自定义配置等更多详情，请参见 应用总览。

RMS 支持的 MOSN 指标说明如下：

指标名称	说明	单位	配置要求
mosnservice	MOSN 应用服务被调用	次	框架配置日志。文件采集路径 `/home/admin/logs/tracelog/mosn/rpc-server-stat.log` 。 压测日志路径 `/home/admin/logs/tracelog/shadow/mosn/rpc-server-stat.log` 。 插件类型为：SPM > 说明 > SPM 全称为 Service Performance Metric，用于监控常用服务指标。 日志格式如下： • 日志打印时间 • fromApp • toApp（即 currentApp） • 服务名 • 方法名 • 本段时间内的被调用次数 • 本段时间内的请求处理总耗时 • 结果（Y/N） • 全链路压测标志（T/F） • 调用者 Zone

指标名称	说明	单位	配置要求

mosnsal	MOSN 调用外部服务	次	框架配置日志。文件采集路径
			`/home/admin/logs/tracelog/mosn/rpc-client-stat.log`
			。
			压测日志路径
			`/home/admin/logs/tracelog/shadow/mosn/rpc-client-stat.log`
			。
			插件类型为：SPM
			说明
			SPM 全称为 Service Performance Metric，用于监控常用服务指标。
			日志格式如下：
			• 日志打印时间
			• fromApp（即 currentApp）
			• toApp
			• 服务名
			• 方法名
			• 本段时间内调用次数
			• 本段时间内的调用总耗时
			• 结果（Y/N）
			• 全链路压测标志（T/F）
			• 目标 Zone

mosnmsgsub	MOSN 消息订阅	次	框架配置日志。文件采集路径 `/home/admin/logs/tracelog/mosn/msg-sub-stat.log` 。

指标名称	说明	单位	配置要求
mosnmsgpub	MOSN 消息发布	次	框架配置日志。文件采集路径 `/home/admin/logs/tracelog/mosn/msg-pub-stat.log` 。 插件类型为：SPM **说明** SPM 全称为 Service Performance Metric，用于监控常用服务指标。 日志格式如下： • 日志打印时间 • currentApp • topic • eventCode • pubGroupId • 本段时间内的消息发送数量 • 本段时间内的消息发送总耗时 • 结果信息（Y/N） • 全链路压测标志（T/F）

指标名称	说明	单位	配置要求
mosnerror	MOSN 的错误数	次	可选择自定义配置或框架配置日志，框架配置日志的文件采集路径为：`/home/admin/logs/mosn/default.log`、`/home/admin/logs/mosn/antvip.error.log`、`/home/admin/logs/mosn/drm.error.log`、`/home/admin/logs/mosn/endpoint.error.log`、`/home/admin/logs/mosn/registry.error.log`、`/home/admin/logs/mosn/zoneclient.error.log`、`/home/admin/logs/mosn/alert.mirror.log`。 插件类型为：ERROR **说明** ERROR 用于监控 Error 指标。

指标名称	说明	单位	配置要求
mosnErrorCode	MOSN 的错误码	次	可选择自定义配置或框架配置日志，框架配置日志的文件采集路径为：`/home/admin/logs/mosn/alert.default.log`。 插件类型为：MM **说明** MM 全称为 multiMinute，用于监控分钟统计/多Key。

10.8. 分布式链路

10.8.1. 分布式链路概述

分布式链路（原 SOFAStack 中间件中的 分布式链路分析产品）会逐步融合到实时监控中，实现监控和链路分析一体化，方便用户对问题进行跟踪定位，对应用性能、调用链路进行实时分析。

背景信息

分布式链路是面向分布式架构、微服务 （比如 Spring Cloud、SOFAStack、Service Mesh）架构等云原生架构，提供应用可观测性的一种金融级解决方案。帮助运维人员、开发人员和架构师轻松应对复杂的大规模微服务架构下的应用及服务之间的复杂调用关系、性能指标、出错信息与关联日志，从而实现故障根因分析、服务治理、应用开发调试、性能管理、性能调优、架构管控、故障定责等运维开发工作。

- 全链追踪：能够深入应用、服务、数据库、消息，捕获性能异常，识别出现故障的组件及服务。
- 易于使用：可以与 SOFAStack 上的应用进行无缝对接，用户的业务代码无需任何修改即可轻松接入，实现性能可视化与问题分析。
- 扩展性强：遵循业界 OpenTracing 标准，同时支持各类主流的编程框架与数据库等。

分布式链路核心概念

下表介绍分布式链路核心概念，方便查看和索引。

名词	说明
应用拓扑	拓扑是对应用间调用关系和依赖关系的可视化展示。 链路拓扑图中，每个图标元素代表一个节点，具体图标说明可参考 拓扑图元素说明。
Trace ID	代表唯一一次请求的 ID，此 ID 一般由集群中第一个处理请求的系统产生，并在分布式调用下通过网络传递到下一个被请求系统。
Span	代表本次请求的完整信息，包括调用是否成功、调用类型、调用耗时等。其中最核心是 Span ID，代表了本次请求在整个调用链路中的位置或层次。

名词	说明
满意度（Apdex）	满意度（Application Performance Index，简称 Apdex）即对应用的每次响应计时，并划定区间，将用户对应用的使用感受定义为三个等级： • 满意（0-T） • 可容忍（T-4T） • 不满意（> 4T） 说明 • Apdex 是由 Apdex 联盟开发的用于评估应用性能的工业标准，具体可参考 Apdex 官方文档。 • Apdex 定义了应用响应时间的最优门槛为 T（即 Apdex 阈值）。

10.8.2. 应用拓扑

10.8.2.1. 概述

应用拓扑能持续地自动发现您整个应用层级中的依赖关系，实时以端到端的方式展示应用架构。

- **全链路信息展示**：展示应用程序及其关联内部、外部服务系统的响应时间、吞吐量和状态，同时显示了各个服务之间的相互影响。如果一项服务中断，您可以立即看到其他服务所受到的影响。
- **后端服务性能管理**：快速、持续地监控应用性能，让您在第一时间了解应用问题。
- **实时运行状态**：通过监控黄金指标（吞吐量，响应时间， 错误率）、可视化的应用的性能监控描述，快速了解每个应用的运行状态。

应用拓扑 页面以可视化方式展现了全局范围内运行的应用，在指定时间段内的性能数据和拓扑结构。通过应用拓扑页面，可以进行以下操作：

- 指定展示时间段
- 对应用模糊查询
- 选择展示方式：
  - 图形拓扑
  - 列表拓扑

10.8.2.2. 查看应用拓扑图

应用拓扑 页面分别以列表和拓扑图的形式展示当前监控的所有业务系统及其相关的性能指标数据。默认情况下，系统以拓扑图的形式展示。

拓扑概述

拓扑是对应用间调用关系和依赖关系的可视化展示。链路拓扑图中，每个图标元素代表一个节点，即链路中实际调用的对象，每个箭头代表一个调用关系。支持跨应用调用，即拓扑图可展示不同应用的服务间调用关系。
单击节点连线箭头即可查看节点调用信息。
各节点间的连线颜色反应了应用间的调用耗时情况，可以非常直观地看到应用与应用之间的调用状态。

拓扑图元素说明

拓扑是对应用间调用关系和依赖关系的可视化展示。链路拓扑图中，每个图标元素代表一个节点，即链路中实际调用的对象，支持拖拽以移动各图标元素。关于各元素说明可参考下表。

元素	说明
	用户，代表当前节点为 USER。 **说明** 无法通过单击该节点图标来查看指标数据等信息。
	当前 workspace 应用，支持查看节点的监控概览信息，包括请求数、响应时间、错误率。具体指标说明请参见 指标说明。
	其他 workspace 应用，不支持查看节点的监控概览信息。
	MQ（消息队列），代表当前节点为 MQ 应用，不支持查看节点的监控概览信息。
	DB（数据库），代表当前节点为 DB 应用，不支持查看节点的监控概览信息。
	Cache（缓存），代表当前节点为缓存应用，不支持查看节点的监控概览信息。
	应用节点连线，表示两个应用节点间的调用关系，具有方向性。 • 当前 workspace 应用：支持查看连线监控概览信息，包括请求数、满意度、响应时间、错误率、协议。单击连线上的连接点或者箭头，可查看两个应用节点间链路的调用详情，如请求量、平均耗时等指标信息。具体指标说明请参见 指标说明。 • 其他 workspace 应用：不支持查看连线监控概览信息。 • DB：不支持查看连线监控概览信息。 • MQ：不支持查看连线监控概览信息。

指标说明

RMS 分布式链路使用不同的拓扑图标颜色进行标识，蓝色表示应用健康、黄色表示应用出现异常情况、红色表示应用出现错误，方便用户迅速聚焦到非健康应用，快速定位异常链路，完成故障排查。
此外，将鼠标悬在代表应用的节点或者节点连线上，您可以看到监控概览信息，比如请求数、错误率等。具体指标说明可参考下表。

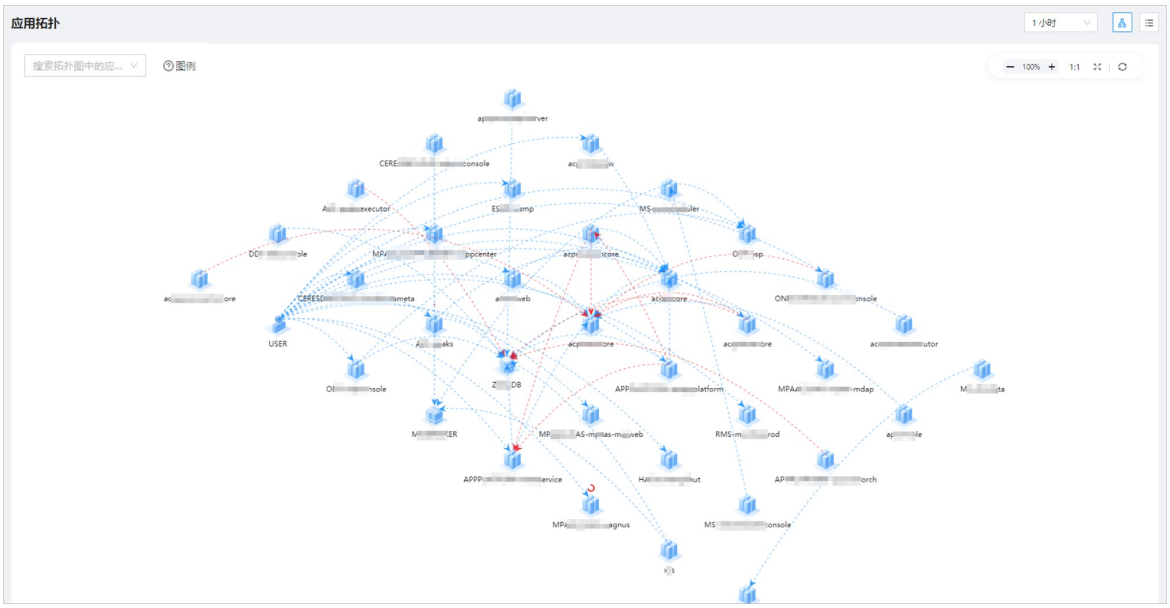
指标项	说明

指标项	说明
健康状态	请求错误率和耗时满意度的阈值进行聚合计算。健康状态如下： • **健康**：根据错误率和耗时满意度评估，应用处于健康状态，即请求错误率和耗时满意度均未超出阈值。 • **异常**：请求错误率未超出阈值，但耗时满意度超出阈值。 • **错误**：请求错误率超出阈值，耗时满意度超出阈值。
请求数	当前应用相关链路的请求总数，单位为次。
响应时间	当前应用相关链路的所有请求的响应时间，单位为 ms。
错误率	当前应用相关链路的错误率（总错误数 / 总请求数），单位为 %。
协议	当前应用相关链路的协议类型。
平均耗时	当前应用相关链路的所有请求的平均耗时，单位为 ms。

查看应用拓扑图

应用拓扑图展示当前应用为核心关联的链路。在 **应用拓扑** 页上，您可以更加直观地看到应用的上下游节点以及与它们的调用关系，从而更快速地找出应用的性能瓶颈。

1. 在左侧导航栏上，单击 **分布式链路** > **应用拓扑**。
2. 在拓扑图页上查看以下信息：



   - 选定时间内的应用调用关系拓扑图。

> **说明**
>
> - 单击右上角时间选择框，可设置时间范围。
> - 将鼠标光标悬停在 **图例** 上，可查看拓扑图元素图例信息。

- 选定时间内应用的请求数、响应时间、错误率。

  > 说明
  > 将鼠标光标悬停在应用节点图标上，即可查看应用的监控概览信息，可参考 拓扑图元素说明、指标说明。

- 选定时间内应用的节点调用信息。

  > 说明
  > - 单击节点连线上的连接点或者箭头，可查看两个应用节点间链路的调用详情，如请求量、平均耗时等指标信息。详情请参见 查看节点调用信息。
  > - 单击 **实时链路** 列的 **查询**，即可跳转至分布式链路的 **链路查询** 页。

其他操作

您可以通过拓扑页面右上角的各个按钮对拓扑图进行如下相关操作。

操作	说明
搜索拓扑图中的应用	从下拉列表中选择目标应用，将打开应用详情面板，您可以查看目标应用在指定时间范围内多项性能指标数据。 
以图形形式展示拓扑	单击 。
以列表形式展示拓扑	单击 。
配置链路采样	单击 。相关操作请参见 链路采样配置。
放大或缩小拓扑图	单击 即可等比例缩小拓扑图。 单击 即可等比例放大拓扑图。
恢复至默认大小	单击 1:1，拓扑图将恢复到默认的大小，即标准大小 100%。

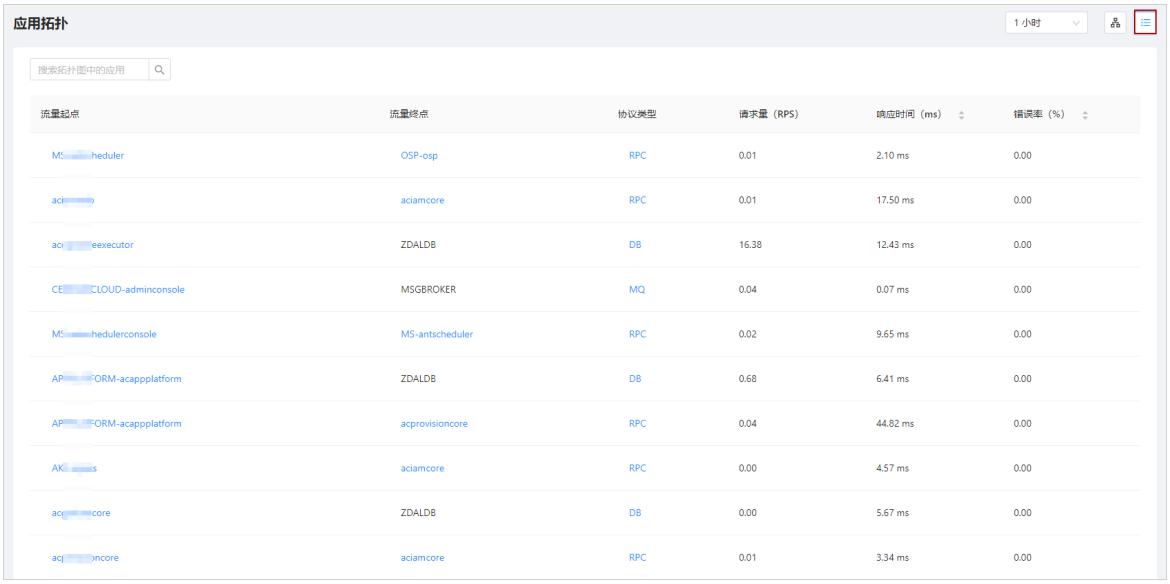
操作	说明
全屏展示拓扑图	单击 。 进入全屏模式后，再次单击该图标即可退出全屏模式。
设置时间查询范围	单击右上角的时间选择框，然后单击一个预设的时间范围，例如 **1 小时**。

10.8.2.3. 查看应用列表

在 **应用拓扑** 页，单击右上角列表图标，可切换至应用列表视图。您可以查看指定统计时段内的各个应用的概览信息，包括流量起点、流量终点等信息。

操作步骤

1. 在左侧导航栏上，单击 **分布式链路 > 应用拓扑**。
2. 单击列表图标切换至应用列表视图。
3. 查看指定统计时段内的各个应用的概览信息，包括流量起点、流量终点、协议类型、请求量（RPS）、响应时间（ms）以及错误率（%）。



更多操作

在应用列表页，您还可以执行如下操作：

操作项	说明
设置查询时间范围	单击右上方时间选择框，选择时间范围。默认值为 **1 小时**。支持的最短时间间隔为 **15 分钟**，最长时间间隔为 **7 天**。
查看应用性能详情	单击流量起点名称或流量终点名称，即可查看应用性能详情。详见 查看应用性能详情。
查看节点调用信息	单击协议类型，即可查看节点调用信息。详见 查看节点调用信息。

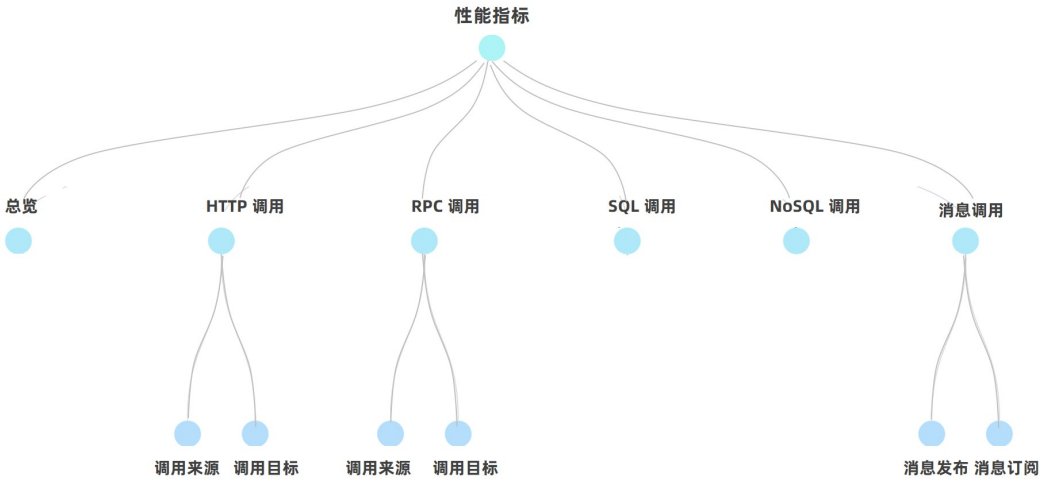
10.8.2.4. 查看应用性能详情

本文介绍如何查看应用性能详情，了解应用在指定时间范围内的性能指标数据。

背景信息

分布式链路追踪提供链路性能分析功能，具体如下：

- 展示所选节点的黄金指标统计（比如吞吐量、平均响应时、错误率、满意度等）趋势和指标明细，并默认提供 HTTTP 调用、RPC 调用、缓存调用、SQL 调用等节点调用类型指标统计。



- 支持基于应用 > 上下游应用 > 接口等逐层下钻分析，建立从底层至上层间的数据关联信息，从而深度分析分布式场景下影响应用性能的问题根因。若发现某个接口调用异常，可跳转链路查询界面，按照相关参数查询链路。

功能入口

1. 在左侧导航栏上，单击 **分布式链路 > 应用拓扑**。
2. 单击列表图标切换至应用列表视图。



3. 在应用列表页，单击流量起点名称或流量终点名称，进入应用性能详情页。



总览

应用概览信息。**总览** 页签下的指标根据该应用节点作为 Server 的请求数据统计而来，与拓扑节点显示指标一致，包括应用的请求数、错误率、平均时间和失败数。
参数说明可参考下表。

参数名	说明	单位
请求次数	每分钟请求次数。	次
错误率	每分钟的错误率。	%
平均耗时	每分钟的平均耗时。	ms
失败次数	每分钟的失败请求次数。	次

HTTP 调用

当前应用通过 HTTP 调用其他应用的指标数据。

调用来源

显示指标趋势和来源应用明细，包括调用来源应用所选应用的 URL、调用方法等信息。
参数说明可参考下表。

参数名	说明	单位
URL	调用来源应用所选应用的 URL。	/
调用方法	所选应用调用来源应用接口的调用方法。	/
来源应用名称	所选应用调用来源应用名称。	/
请求次数	所选应用调用来源应用接口请求数总数。	次
错误率	来源应用接口请求所选应用的错误率（错误总数 / 总请求数）。	%
平均耗时	所选应用调用目标应用接口平均耗时总数。	ms
失败次数	所选应用调用目标应用接口失败总数。	次

调用目标

显示指标趋势和目标应用明细，包括目标应用调用所选应用的 URL、调用方法等信息。
参数说明可参考下表。

参数名	说明	单位
URL	目标应用调用所选应用的 URL。	/
调用方法	目标应用接口调用所选应用的调用方法。	/
目标应用名称	调用应用的名称，即应用列表中所单击的应用名称。	/
请求次数	目标应用接口请求所选应用的请求数总数	次

参数名	说明	单位
错误率	目标应用接口请求所选应用的错误率（错误总数 / 总请求数）。	%
平均耗时	目标应用接口请求所选应用的平均耗时总数。	ms
失败次数	目标应用接口请求所选应用的失败总数。	次

RPC 调用

当前应用通过 RPC 调用其他应用的指标数据。

调用来源

显示指标趋势和来源应用明细，包括来源应用调用所选应用的接口、调用方法等信息。
参数说明可参考下表。

参数名	说明	单位
接口	来源应用调用所选应用的接口。	/
调用方法	来源应用接口调用所选应用的调用方法。	/
来源应用名称	调用应用的名称，即应用列表中所单击的应用名称。	/
请求次数	来源应用接口请求所选应用的请求数总数	次
错误率	来源应用接口请求所选应用的错误率（错误总数 / 总请求数）。	%
平均耗时	来源应用接口请求所选应用的平均耗时总数。	ms
失败次数	来源应用接口请求所选应用的失败总数。	次

调用目标

显示指标趋势和目标应用接口明细，包括所选应用调用目标应用的接口、调用方法等信息。
参数说明可参考下表。

参数名	说明	单位
接口	所选应用调用目标应用的接口。	/
调用方法	所选应用调用目标应用接口的调用方法。	/
目标应用名称	所选应用调用目标应用名称。	/
请求次数	所选应用调用目标应用接口请求数总数。	次
错误率	来源应用接口请求所选应用的错误率（错误总数 / 总请求数）。	%
平均耗时	所选应用调用目标应用接口平均耗时总数。	ms
失败次数	所选应用调用目标应用接口失败总数。	次

SQL 调用

调用数据库的指标。用来监控应用访问数据库的相关指标。**SQL 调用** 页面显示了指标趋势和数据源明细，包括来所选应用调用的数据源、逻辑库等信息。
参数说明可参考下表。

参数名	说明	单位
数据源	所选应用调用的数据源。	/
逻辑库	所选应用调用的逻辑库。	/
数据库类型	所选应用调用数据库的类型。	/
请求次数	所选应用调用数据库的请求数总数。	次
错误率	来源应用接口请求所选应用的错误率（错误总数 / 总请求数）。	%
平均耗时	所选应用调用数据库的平均耗时总数。	ms
失败次数	所选应用调用数据库的失败总数。	次

NoSQL 调用

显示指标趋势和消息订阅明细，包括所选应用调用的 NoSQL 类型、调用方法等信息。
参数说明可参考下表。

参数名	说明	单位
NoSQL 类型	所选应用调用的 NoSQL 类型。	/
调用方法	所选应用调用 NoSQL 的方法。	/
请求次数	所选应用调用 NoSQL 的请求数总数。	次
错误率	所选应用调用 NoSQL 的错误率（错误总数 / 总请求数）。	%
平均耗时	所选应用调用 NoSQL 的平均耗时总数。	ms
失败数	所选应用调用 NoSQL 的失败总数。	次

消息调用

消息发布

消息发布的指标。**消息发布** 页签显示了指标趋势和数据源明细，包括所选应用调用的消息发布的消息主题（Topic）、请求次数等信息。
参数说明可参考下表。

参数名	说明	单位
Topic	所选应用调用的消息发布的消息主题。	/
请求次数	所选应用发布消息的请求数总数。	次
错误率	所选应用发布消息的错误率（错误总数 / 总请求数）。	%

参数名	说明	单位
平均耗时	所选应用发布消息的平均耗时总数。	ms
失败次数	所选应用发布消息的失败总数。	次

消息订阅

消息订阅的指标。**消息订阅** 页签显示了指标趋势和数据源明细，包括所选应用调用的消息订阅的消息主题（Topic）、请求次数等信息。
参数说明可参考下表。

参数名	说明	单位
Topic	所选应用调用的消息订阅的消息主题（Topic）。	/
请求次数	所选应用调用消息订阅的请求数总数。	次
错误率	所选应用调用消息订阅的错误率（错误总数 / 总请求数）。	%
平均耗时	所选应用调用消息订阅的平均耗时总数。	ms
失败次数	所选应用调用消息订阅的失败总数。	次

10.8.2.5. 查看节点调用信息

本文介绍如何通过应用列表页面查看节点调用信息。

操作步骤

1. 在左侧导航栏上，单击 **分布式链路 > 应用拓扑**。
2. 单击列表图标切换至应用列表视图。
3. 在应用列表页，单击协议类型名称，将打开节点调用信息页。
   在该页面，您可以查看应用节点调用详情，包括指标趋势和调用明细。



更多操作

在 **节点调用信息** 页，还可以执行如下操作：

- 在 **指标趋势** 区域，选择对应的调用方式，将按所选的调用方式呈现对应指标趋势图。
- 单击 **实时链路** 列的 **查询**，即可进入 **链路查询** 页。在该页面，您可以查询关联的 Trace ID。

10.8.3. 链路查询

10.8.3.1. 查询调用链路

RMS 支持按多维查询项检索调用链，包括 HTTP、RPC、DB、MQ 等不同类型的调用方式，可实现按各场景自定义 Tag（比如交易号）查询对应的调用链路集合。本文介绍如何通过控制台查询调用链路。

操作步骤

1. 在左侧导航栏上，单击 **分布式链路 > 链路查询**。
2. 在 **链路查询** 页，设置查询项。
   搜索项都是链路（Trace）字段，具体查询项可参考下表。

参数类型	说明
本地应用名	选择本地应用名。支持模糊搜索。 **说明** **本地应用名** 即应用类型为 Client 的应用名。
远端应用名	选择远端应用名。支持模糊搜索。 **说明** **远端应用名** 即应用类型为 Server 的应用名。
Trace ID	链路的唯一标识。输入Trace ID。
调用结果	**成功** 或 **失败**。
耗时大于	调用的耗时大于指定毫秒数。
起止时间	您可以选择预设的时间范围，或者输入自定义的时间范围。 ○ 单击页面右上角的时间选择框，然后单击一个预设的时间范围，例如 **最近 15 分钟**、**最近 30 分钟**、**最近 1 小时**。 ○ 如果没有符合需求的预设时间范围，则在日历中选择开始和结束时间，或者在文本框内手动输入，并单击 **确定**。

参数类型	说明
调用方式	支持 ALL、HTTP、RPC、DB、MQ 或 NoSQL。选择不同的调用方式，将展开更多查询条件，具体如下： ○ 选择 HTTP：新增 **调用方法**、URL 这两个查询项。 说明 查询时 URL 字段受数据库限制，字符串长度超过 91 时将被忽略。 ○ 选择 RPC：新增 **调用方法**、**服务名** 这两个查询项。 说明 查询时 **服务名** 字段受数据库限制，字符串长度超过 91 时将被忽略。 ○ 选择 DB：新增 **插件类型**、**数据库**、**数据源**、**数据类型** 这四个查询项。 ○ 选择 MQ：新增 **消息类型**、topic 这两个查询项。 ○ 选择 NoSQL：新增 NoSQL 类型、**调用方法** 这两个查询项。
自定义业务 Tag	输入自定义的业务 Tag。仅支持 JSON 格式。

3. 单击 **查询**。
   在查询结果列表中，查看链路的 Trace ID、调用时间、响应时长、本地应用名称及 IP 地址、远程应用名称及 IP 地址。

后续操作

在查询结果列表中，单击任一链路的 Trace ID，即可进入调用链路详情页，查看链路详情。相关操作请参见 查看链路详情。

10.8.3.2. 查看链路详情

如需获取链路详情，可单击链路查询结果中的 Trace ID，进入链路详情页。链路详情页提供链路图、拓扑图、时序图三种展现形式。您不仅可以查看到调用链路的细节信息，例如应用类型、调用方式，还能通过上下游的拓扑图，理解该链路中的调用关系。

功能入口

1. 在左侧导航栏上，单击 **分布式链路** > **链路查询**。
2. 在 **链路查询** 页，设置查询项。详情请参见 查询调用链路。
3. 在查询结果列表中，单击目标链路的 Trace ID 进入链路详情页。
   - 链路图：链路详情页默认显示该链路的链路图。详情请参见 链路图展示。
   - 拓扑图：单击拓扑图图标查看链路上下游调用拓扑关系。详情请参见 拓扑图展示。
   - 时序图：单击时序图图标查看应用间发送消息的时间顺序，以及多个应用之间的动态协作。详情请参见 时序图展示。

链路图展示

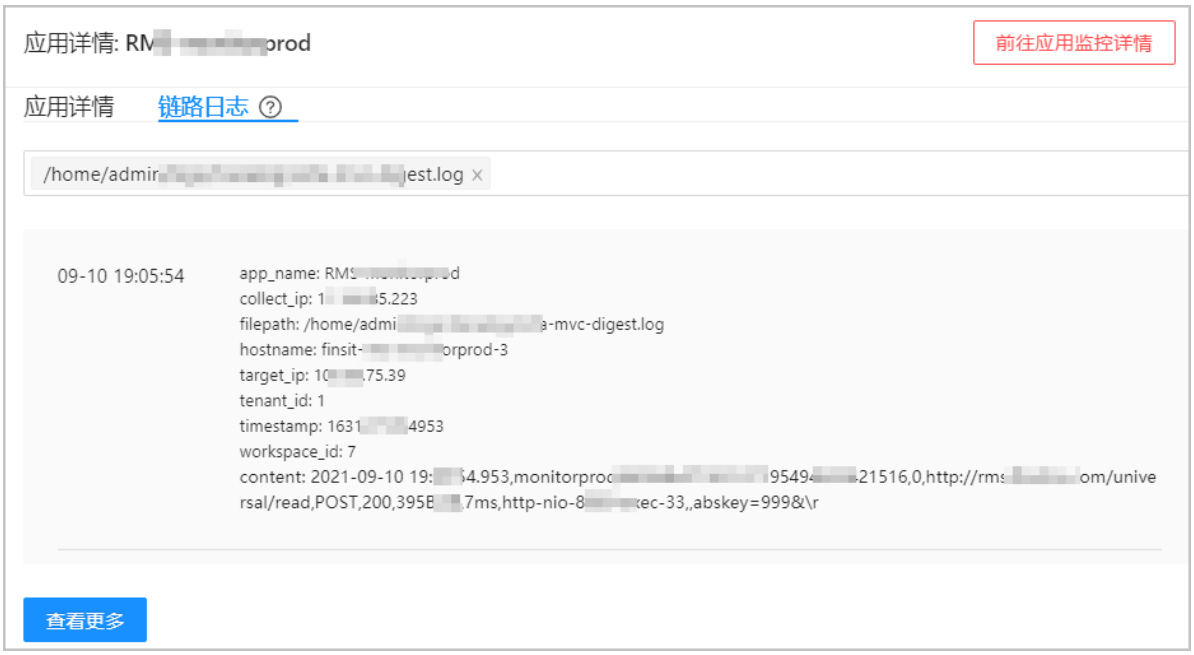
链路详情页默认显示该链路的链路图，您可以直观地获取到此调用链路的应用名、IP 地址、调用类型、调用状态（成功/失败）、服务信息、单元化信息、耗时等。
单击目标调用链路，您可以更直观地看到详细的调用树状图。通过该树状图可以轻松找到整条链路中，哪些调用出现失败、哪些调用的耗时过长等。

若设置调用满意度阈值，在调用树状图中使用不同颜色对不同区间的值进行标识（如下图所示），方便您快速发现应用性能问题。



单击应用名进入应用详情页，您可以看到更多链路的细节信息。

- 单击 **链路日志** 页签，查看日志信息。选择目标日志路径，将显示查询到的日志信息。

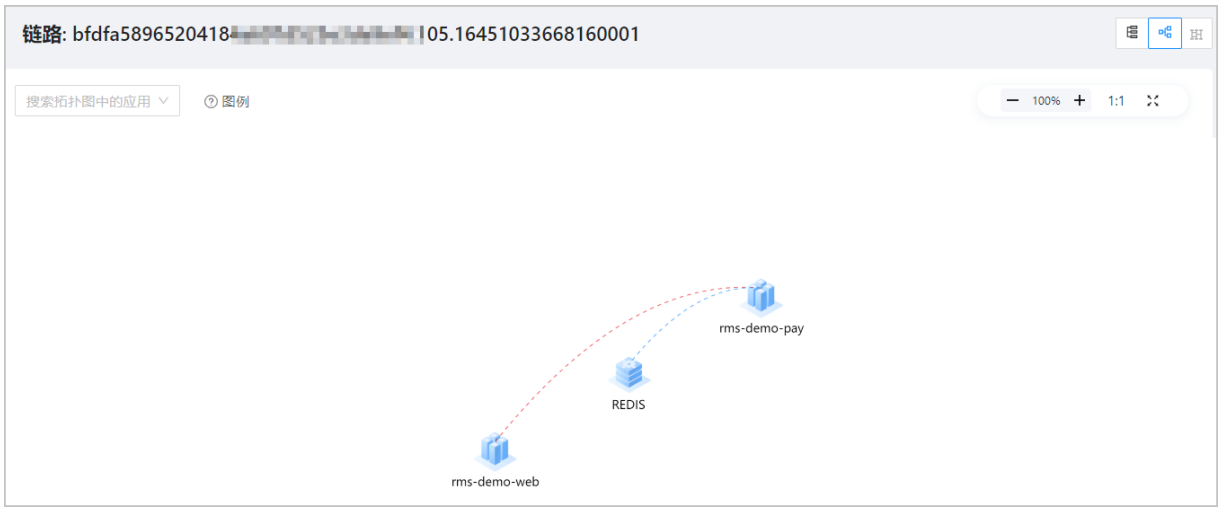


单击 **查看更多** 跳转到 **日志查询** 页。在 **日志查询** 页，将自动填充应用 Trace ID、应用名等参数。系统根据这些参数自动查询对应的日志信息。

- 单击右上角 **前往应用监控详情**，将跳转至 RMS 应用监控页面，显示更为详细的应用监控指标，比如系统指标和 JVM 指标等。

拓扑图展示

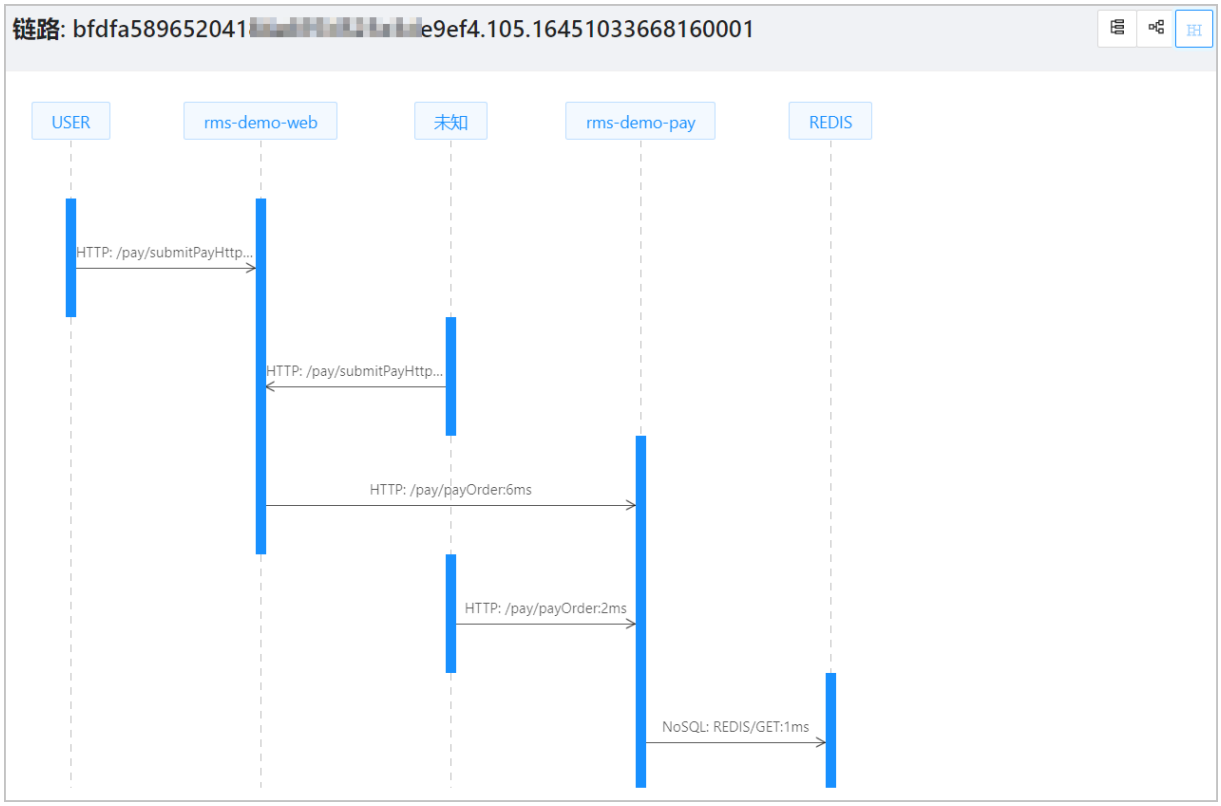
单击拓扑图图标，即可切换至单链路拓扑图展示页，查看该链路的上下游调用拓扑关系（如下图所示）。如果将鼠标光标悬浮于应用节点图标上，还可以查看各节点或调用关系间的性能数据，如请求量、响应时间和错误率。



时序图展示

单击时序图图标，即可切换至单链路时序图展示页，查看应用间的完整调用关系。

- 应用名：按请求发送时间顺序展示应用名。
- 调用标识：发起请求和接受请求的标识。
- 调用连线：展示调用的服务信息和耗时。



10.8.3.3. 链路上报

在 RMS 3.0 中，分布式链路追踪系统支持 SOFATracer 的主动上报模式。

操作步骤

如需使用该模式，请执行如下操作：

1. 引入依赖。

```
<dependency>
    <groupId>com.alipay.sofa</groupId>
    <artifactId>tracer-enterprise-sofa-boot-starter</artifactId>
</dependency>
```

2. 增加配置。

```
com.alipay.sofa.tracer.remote=dst
com.alipay.sofa.tracer.dst.acvip=false
com.alipay.sofa.tracer.dst.url=http://{{monitor-gateway-lb}}/zipkin/api/v1/spans
```

> 说明
> 关于 monitor-gateway-lb 列表，参见如下示例。
>
> 公有云-杭州金区生产环境：{cn-hangzhou-domainName}
> 即： com.alipay.sofa.tracer.dst.url=http://{cn-hangzhou-domainName}/zipkin/api/v1/spans

10.8.4. 分布式链路配置

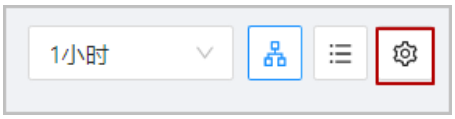
分布式链路模块提供自定义链路配置功能，方便用户快速查看和定位问题，比如应用节点错误率是否过高、应用的响应时间是否能接受等。

链路采样配置

链路采样配置支持链路关键字段采集配置、链路采样率配置、链路采样率配置。

功能入口

1. 在左侧导航栏上，单击 **分布式链路 > 应用拓扑**。
2. 单击配置图标，进入链路采样配置页面。



链路关键字段采集配置

RMS 默认采集链路每个 span 的全字段，如需节约存储资源，可开启开关，将仅采集部分关键字段，比如调用路径、调用返回码等。

链路采样率配置

为进一步降低计算和存储成本，RMS 支持链路采样率配置。通过为 workspace 设置链路服务端采样率，按照配置进行采集链路请求数据，而不是跟踪每个请求。

如需配置链路采样率，开启开关，输入采样率后保存即可。比如输入 **70%**，表示仅采集 70% 的链路服务端数据。

错误链路全采样

如需查询追踪每一笔错误调用请求，开启错误链路全采集，将全部错误链路全采样，其他类型链路按照上述采样率进行采集存储。

10.9. 告警管理

10.9.1. 通知组管理

对各类告警规则设置通知组后，一旦告警触发，告警消息可以及时发送给通知对象。

创建通知组

1. 在左侧导航栏上，单击 **告警管理** > **通知组**。
2. 单击 **新增通知组**，在弹出的页面上填写以下信息。
   - **通知组名称**：必填。输入通知组的名称。
   - **联系人**：从下拉列表中选择对应联系人。
     如需创建新的联系人，请参见 创建联系人。
   - **钉钉机器人**：输入自定义机器人的 webhook 地址后，单击 **添加钉钉机器人** 即可。

     > 说明
     > 有关如何获取自定义机器人 webhook，请参见 获取自定义机器人webhook。

3. 单击 **保存**。

更多操作

通知组创建完成后，您还可以在通知组列表页面执行更多操作。具体如下所示。

- 如需查看通知组，请单击 **告警管理** > **通知组管理** 进入通知组列表页。
- 如需搜索通知组，请在页面右上角的输入栏，输入通知组的查询关键字，并单击搜索图标。
- 如需编辑通知组，请单击操作栏中的 **编辑**，并在通知组对话框中编辑相关信息，然后单击 **保存**。
- 如需删除通知组，请单击操作栏中的 **删除**，然后在弹出的提示对话框中单击 **确定**。

> 说明
> 删除后，所有订阅该通知组的告警规则，都将自动取消该通知组的订阅信息。

10.9.2. webhook 管理

设置 webhook 告警后，您可以将告警通知以指定方式发送到自定义的 webhook 地址中。您可以在 **webhook 管理** 页创建、修改和删除 webhook。

创建 webhook

1. 在左侧导航栏上，单击 **告警管理** > **webhook 管理**。
2. 单击 **新增 webhook**，在 webhook 对话框中输入配置信息。
   基本参数描述如下所示。

参数	说明
Webhook 名称	必填，自定义 Webhook 名称。
请求方式	默认为 POST，不能修改。 URL：webhook 地址，不可超过 100 个字符。

参数	说明
请求头	非必填，设置请求头，不可超过 200 个字符。您可以添加其他 Header 信息或 Param 信息。 **说明** ○ 默认请求头为 `Content-Type: text/plain; charset=UTF-8` 。 ○ Header 和 Param 总数不能超过 6 个。
Access Key & Secret Key	非必填。webhook 对应的 Access Key 和 Secret Key，用于校验签名。
全局订阅	默认关闭。开启后，将订阅当前工作空间下所有告警信息。 **说明** ○ 如果开启开关，用户无需手动订阅告警信息即可接收到当前工作空间下所有告警信息。 ○ 如果关闭开关，则需要手动订阅告警信息。详情请参见 订阅信息入口。

3. 单击 **保存**。

执行结果

- 创建完成后，该 webhook 将显示在 webhook 列表中。



- 如果在创建 webhook 过程中开启 **全局订阅**  ，该 webhook 将显示在 **告警订阅** 页。



相关操作

- 如需查看 webhook，请单击 **告警管理** > **webhook 管理** 进入 webhook 列表页。
- 如需搜索 webhook，请在页面右上角的输入栏，输入 webhook 的查询关键字，并单击搜索图标。
- 如需编辑 webhook，请单击操作栏中的 **编辑**，并在编辑 webhook 对话框中编辑相关信息，然后单击 **保存**。
- 如需删除 webhook，请单击操作栏中的 **删除**，然后在弹出的提示对话框中单击 **确定**。

> **说明**
> 删除后，该 webhook 订阅的所有告警规则都将自动取消订阅。

10.9.3. 告警历史

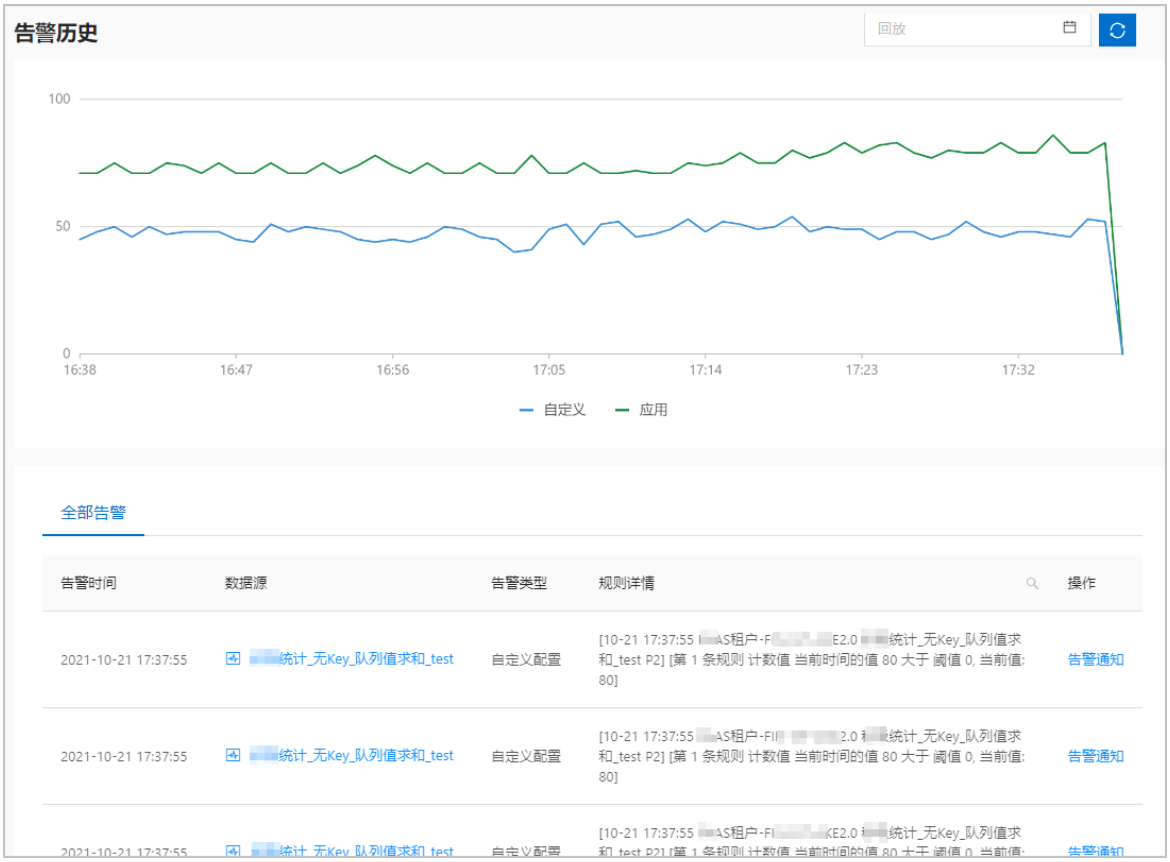
RMS 提供对应用和自定义监控的多指标联合告警，通过将告警规则订阅到某个通知组，实现通过钉钉、短信、电话等方式将告警消息通知到用户，并可在告警历史中查看告警的具体信息，并将告警信息和监控数据打通，实现关联跳转，帮助运维人员通过告警消息快速的定位到具体问题，并做进一步排查。
RMS 支持的告警触发条件如下：

- 当前时间的值
- 当前时间的环比
- 当前时间的值与昨日同比
- 当前时间的值与上周同比
- 最近 N 分钟求和
- 最近 N 分钟求平均
- 最近 N 分钟求和与上 N 分钟求和环比
- 最近 N分钟求和与昨天同比

告警历史

除了在不同的监控类型页面查看告警历史，您还可以在 **告警管理** > **告警历史** 页面中集中地查看告警历史。

1. 登录控制台。
2. 在左侧导航栏，单击 **告警管理** > **告警历史**。
3. 查看告警历史页面。
   - **告警历史**：显示在过去某一段时间内的不同资源类型的告警数据，以曲线图表示。
     默认显示最近一小时的告警历史数据。如要修改，单击右上角的数据显示间隔，编辑告警历史数据显示方式，包括显示的历史时间、区间、页面刷新频率。
   - **全部告警**：显示不同告警类型的详细情况，包括告警时间、规则详情、告警类型等，并且可以查看告警通知。



10.9.4. 订阅管理

订阅管理主要提供告警订阅信息的可视化展示，订阅信息主要分为下述两个维度：

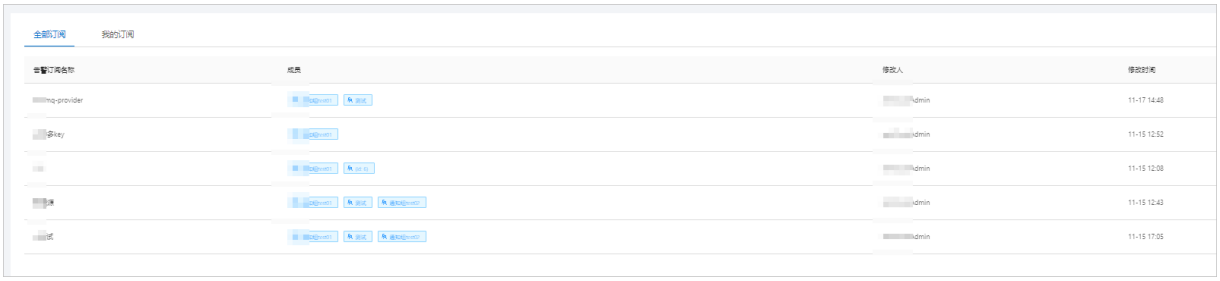
- **全部订阅**：查看当前业务域的告警订阅情况。
- **我的订阅**：查看当前登录用户的告警订阅情况。

订阅内容

订阅信息主要包括：

- 告警订阅名称
- 成员
- 修改人
- 修改时间
- 操作

订阅信息界面示意图



订阅信息入口

全部订阅入口

在自定义监控页面，单击所添加的监控产品右侧 **操作** > **订阅告警**，并在弹出的页面上添加订阅告警信息后，所配置的告警订阅信息会出现在 **全部订阅** 中。

全部订阅入口



我的订阅入口

进入左侧导航栏 **应用监控**，选择 **我的应用** 或 **全部应用** 中的某个应用，在应用详情页，选择右上角 **更多** > **订阅告警**，在弹出的添加订阅告警页面上，将新增订阅人配置为自己的账号，则配置的告警订阅信息会同时出现在 **我的订阅** 和 **全部订阅** 中。

我的订阅入口





10.9.5. 告警联系人

联系人支持通过短信、邮件接收告警通知。您可以在实时监控控制台上配置告警联系人。

创建联系人

1. 在左侧导航栏上，单击 **告警管理** > **告警联系人**。
2. 在 **联系人管理** 页，单击右上角的 **新增联系人**。

3. 在 **创建联系人** 对话框中，完成联系人配置。

配置项	说明
联系人名称	输入联系人名称信息。本例中输入 Contact_Test。 **说明** 名称应包含 2~40 个字符，必须以大小写字母或中文开头，可包含数字、 . 、下划线。
手机号	根据实际需求，填写联系人手机号码。 **说明** ○ 手机号码、邮箱至少填写一项。 ○ 每个手机号码只能用于一个联系人。
邮箱	根据实际需求，填写联系人电子邮箱地址。 **说明** ○ 手机号码、邮箱地址至少填写一项。 ○ 每个邮箱地址只能用于一个联系人。

4. 单击 **确定**。

执行结果

创建完成后，名为 Contact_Test 的联系人将显示在联系人管理列表中。



相关操作

- 如需将联系人添加到通知组，可前往通知组管理页进行设置，详情请参见 通知组管理。
- 如需修改联系人信息，请单击操作栏中的 **编辑**，并在 **编辑联系人** 对话框中修改相关信息，然后单击 **确定**。
- 如需删除联系人，请单击操作栏中的 **删除**，然后在弹出的提示对话框中单击 **确定**。

说明

删除后，所有添加过该联系人的通知组，都将自动删除该联系人的信息。

10.9.6. 设置钉钉报警机器人

设置钉钉报警机器人后，您可以通过设定的钉钉群接收报警通知。本文介绍如何配置钉钉报警机器人。

步骤一：新增通知组

1. 登录控制台。
2. 在左侧导航栏上，单击 **告警管理** > **通知组**。
3. 单击 **新增通知组**，在弹出的页面上填写以下信息后，单击 **保存**。
   - **通知组名称**：必填。输入通知组的名称。
   - **钉钉机器人**：输入自定义机器人的 webhook 地址。

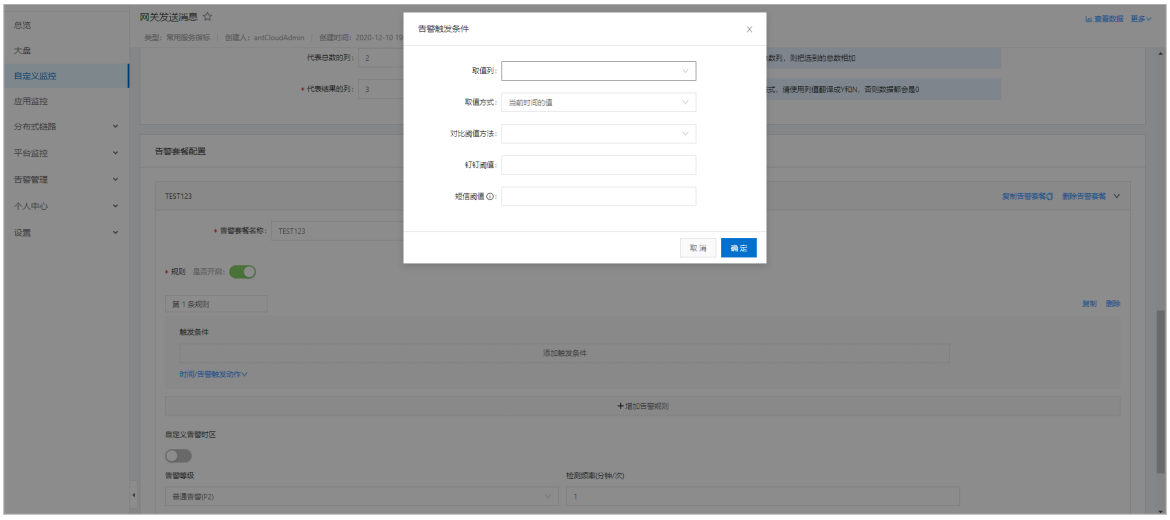
> 说明
> 有关如何获取自定义机器人 webhook，请参见 获取自定义机器人webhook。

步骤二：配置监控报警

1. 在 **自定义监控** 页，选择 **常用服务指标** 类型的数据源，在 **操作** 列表中单击 **编辑**。

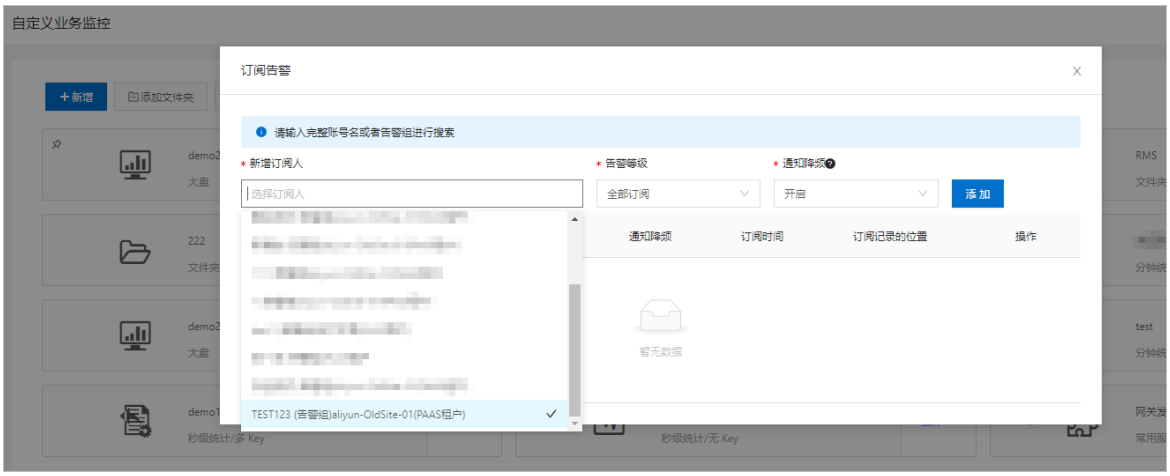


2. 在 **预警套餐配置** 区域新增告警套餐，然后单击 **增加预警规则**。



3. 单击所添加的监控产品右侧 **操作** > **订阅告警**。
4. 在弹出的 **订阅告警** 页面上添加订阅预警信息。
   - **新增订阅人**：选择的订阅人信息，从下拉列表中选择对应的订阅人。
   - **告警等级**：用于告警的管理，可以选择不同的等级。目前提供 **全部订阅**、**P0**、**P1**、**P2**、 **P3**、 **P4** 五种选项。本例选择 **全部订阅**。
   - **通知降频**：可以选择订阅的预警是否进行降频，降频开关打开后会开启预警通知降频策。

> **说明**
> 具体降频策略如下：
> 每隔 1 分钟通知一次，持续 3 分钟；每隔 2 分钟通知一次，持续 4 分钟；每隔 5 分钟通知一次，持续 25 分钟；每隔 10 分钟通知一次，持续 50 分钟每隔 30 分钟通知一次，持续 150 分钟；直至每隔 1 个小时通知一次。



5. 相关信息设置完毕后，单击 **添加**。



添加成功的订阅将以列表的形式展现。

验证您的操作

操作至此，您已成功设置一个钉钉机器人报警。当报警触发时，您将在设置接收报警的钉钉群中收到一则报警通知，例如：



10.10. 个人中心

10.10.1. 收藏夹

个人中心是对当前用户的收藏和配置操作，比如当前用户收藏的大盘、应用和监控项等，以及通知给当前用户的告警和订阅，帮助用户只关心和自己相关的工作和内容，避免其他复杂数据和功能的干扰。其次用户可对根据个人情况配置时区、代理等功能。

在收藏夹中所提供的信息分为下述几类：

- 大盘
- 应用
- 监控项

大盘

进入自定义监控后，对于 **大盘** 类型的监控产品，单击产品右侧 **操作** > **加入收藏** 可以添加到收藏夹中的 **大盘**。示例如下：

添加收藏示例



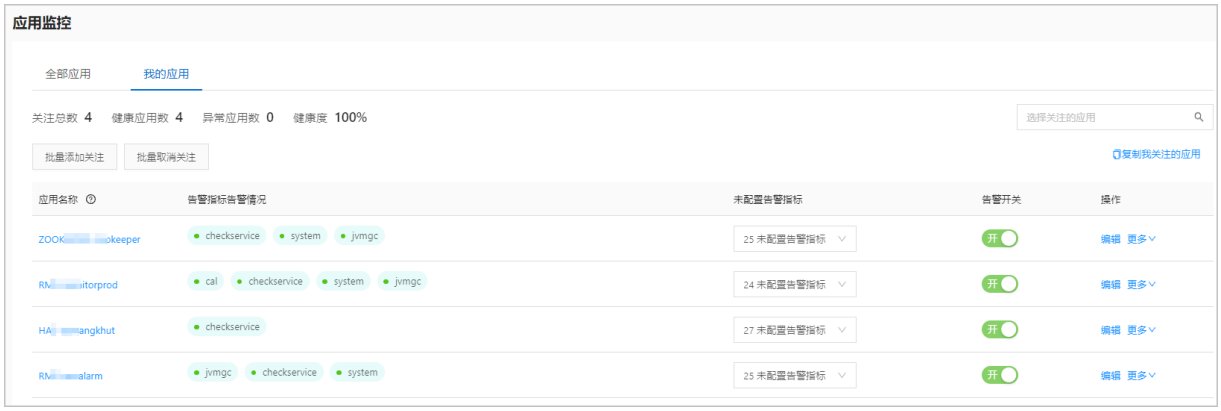
收藏夹大盘界面示例



应用

这部分功能是对 **应用监控 > 我的应用** 内容的复用，更多详情，请参见 应用监控概述。

收藏夹应用界面示例



监控项

进入自定义监控后，选择非 **大盘** 类型的监控产品，单击右侧 **操作 > 加入收藏**，可以将监控产品添加到收藏夹。

收藏夹监控项界面示例



10.10.2. 告警和订阅

通过 **告警和订阅** 页面，用户可以查询到的信息主要包括：

- 告警历史
- 业务告警订阅
- 应用监控告警订阅
- 我的通知组

告警通知和订阅界面示例



告警历史

以柱状图形式展示告警历史：

- 柱形高低代表通知总量。点击柱形，可以查看订阅的告警历史。主要信息包括：
  - 告警时间
  - 告警类型
  - 数据源
  - 告警详情
- 默认告警区间：最近一小时。
- 可以通过右上角的回放功能，浏览指定时间的告警历史数据。

应用监控告警订阅

在自定义监控页面，单击所添加的监控产品右侧 **操作 > 订阅告警**，并在弹出的页面上添加订阅告警信息后，所配置的告警订阅信息会出现在 **应用监控告警订阅** 中。

订阅告警入口



应用监控告警订阅

进入左侧导航栏 **应用监控**，选择 **我的应用** 或 **全部应用** 中的某个应用，在应用详情页，选择右上角 **更多 > 订阅告警**，在弹出的页面上添加订阅告警信息后，所配置的告警订阅信息会出现在 **应用监控告警订阅** 中。

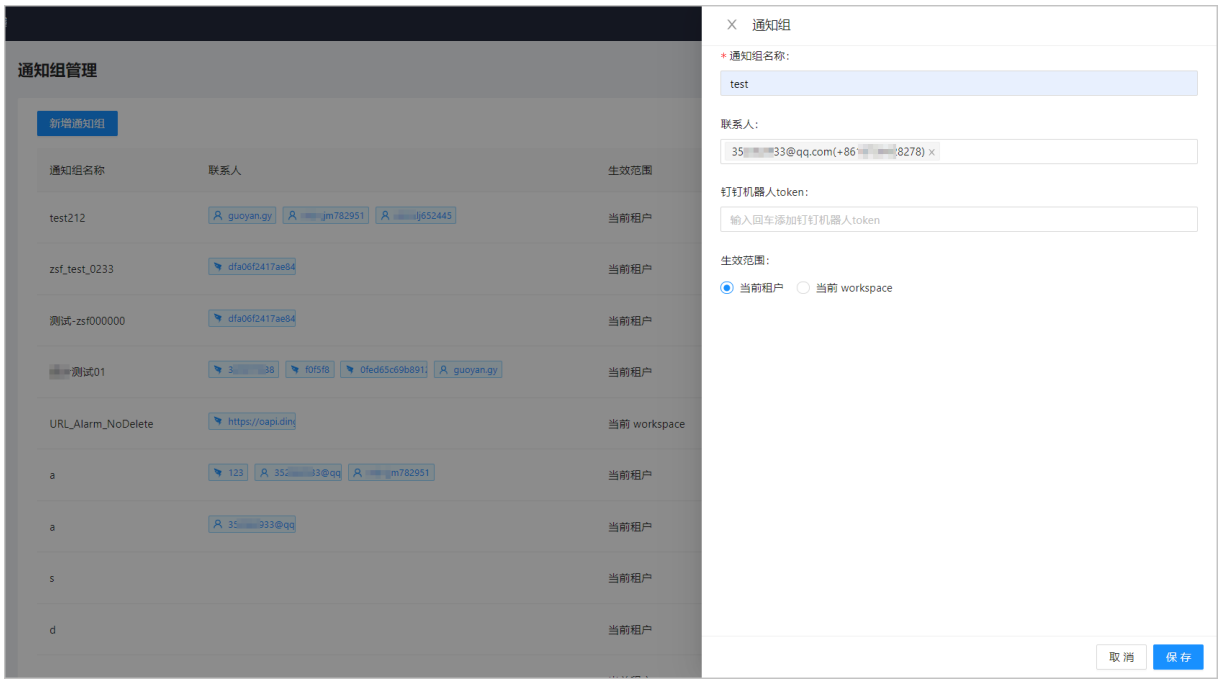
应用监控告警订阅入口



我的通知组

进入左侧导航栏 **告警管理**，选择 **通知组管理**，单击 **新增通知组**，将当前租户添加到通知组后，在 **个人中心 > 告警和订阅 > 我的通知组** 中就可以看到通知组的信息。

我的通知组入口



10.10.3. 个人设置

个人设置，主要用来编辑和查看当前登录用户的相关监控和报警信息，单击右上角 **编辑** 按钮，可以对下述几个部分进行编辑：

- 基础设置
- 报警渠道
- 代理

个人设置界面示例



基础设置

主要包括下述内容：

- **产品时区**：监控系统显示的时间和指标项所用时区。
- **产品语言**：监控系统使用的语言版本。

报警渠道

主要包括 3 个 Level 的报警渠道及对应的报警开关。报警渠道支持钉钉、短信、邮件、电话等通知方式。

代理

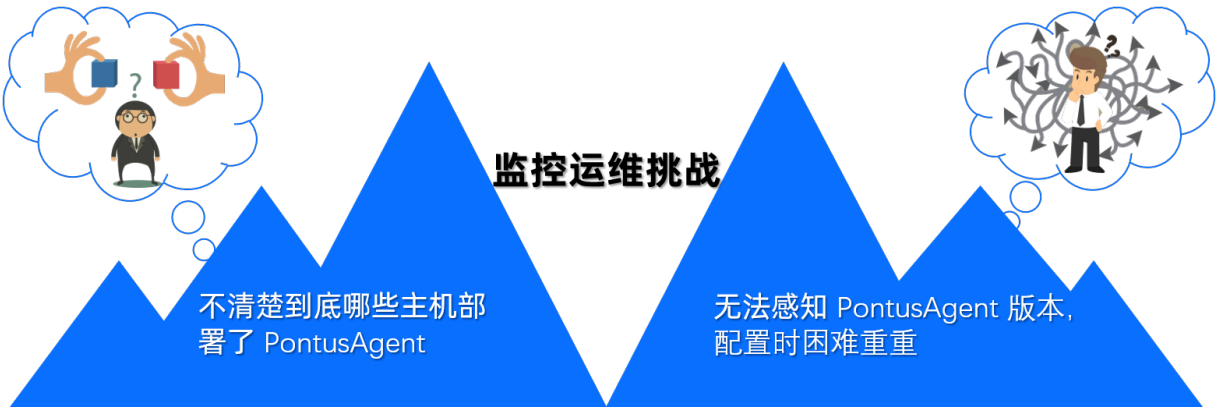
支持通过设置代理的方式，将报警通知以及对应的处理，在指定时间内转交代理人。

10.11. 设置

10.11.1. Agent 管理

RMS 3.0 提供全局设置管理功能，帮助运维人员可通过界面白屏自定义配置，以及更好的排查监控问题，比如哪些资源对象被监控，采集多少个监控配置项，采集、存储时间等是否可修改。

监控运维遇到挑战



- 运维人员不清楚到底哪些主机部署了 PontusAgent，需要在各节点上通过命令查看。
- 用户无法感知 PontusAgent 版本，可能存在一个集群中有多个版本，需要手动进行升级。

Agent 管理帮您解决



RMS 提供 PontusAgent 查看和管理功能，帮助您全面掌握主机上的 PontusAgent 部署情况、PontusAgent 版本信息等。

PontusAgent 状态

PontusAgent 状态说明详见下表。

状态	说明
健康	为主机安装 PontusAgent 且状态正常运行。
未安装	主机未安装 PontusAgent。
安装中	正在为主机安装 PontusAgent。 说明 安装 PontusAgent 预计需要 3 分钟左右，请耐心等待。

状态	说明
安装失败	主机 PontusAgent 安装失败。
异常	主机 PontusAgent 功能异常。
时钟偏差	服务运行的机器与监控服务器时钟偏差大于 3s。

部署 PontusAgent

RMS 自动部署容器应用所在节点的 PontusAgent。对于经典应用所在节点的 PontusAgent，如果存在未安装或者安装失败的情形，可通过 Agent 管理进行部署。部署完成后，可查看主机的 PontusAgent 状态。

1. 在左侧导航栏上，单击 **设置 > Agent 管理**。
2. 在 **Agent 管理** 页，选择应用类型：
   - **经典**：经典应用运行在 ECS 上，RMS 通过定时任务检查 ECS 是否安装 PontusAgent，如果无则自动安装。
   - **容器**：容器应用运行在 K8s 集群上，RMS 通过 Daemonset 方式自动部署，保证所有新增主机都会有 PontusAgent。
3. 选择主机节点，单击 **部署 > 确定**。

> 说明
>
> - 对于 **容器** 类型的应用，RMS 会自动部署容器应用所在节点的 PontusAgent。
> - 部署预计需要 3 分钟左右，请耐心等待。



开启告警

Agent 管理模块现提供异常告警能力，当 Agent 异常时，比如未安装、失联等情况出现时，系统会针对异常 Agent 提供对应的告警通知，让运维人员快速感知到异常问题。

1. 左侧导航栏上，单击 **设置 > Agent 管理**。
2. 在 **Agent 管理** 页，选择应用类型 **容器** 或 **经典**。
3. 单击 **全局开启告警**。

订阅告警

成功开启告警后，可以根据需要添加订阅。添加订阅后，后续的告警消息即可通过设定的方式（例如钉钉机器人或邮件）进行发送。

1. 在 **Agent 管理** 页，单击 **订阅告警**。
2. 在 **订阅** 面板，选择通知对象。

3. 单击 **确定**。

关闭告警

如需关闭告警，单击 **全局关闭告警**。

10.12. 常见问题

一般使用常见问题

- 数据源类
- 配置和接入类
- 告警类

数据源类

日志支持哪些时间格式？

支持如下时间格式：

- **normal**：最常见的时间格式 2016-12-13 10:00:14
- **apache**：apache 时间格式 08/May/2017 01:13:52
- **_long**：完整时间戳，精确到毫秒，格式 1476064860000
- **shortlong**：时间戳，精确到秒，格式 1476064860
- **monthfirst**：月份在行首，且不包括年份 May 08 01:13:52
- **revert**： normal 格式的月份在前 05-25-2017 18:28:42
- **monthfirst2**：月份在行首，且包含年份 May 08 2017 01:13:52

国际服务器日志打印后无法统计，时区怎么设置？

国际服务器的日志打印时区需要和容器时区相同，不能日志打印是中国时区，而容器是 UTC 时区。

日志通过软链接是否可以采集？

可以。需要明确在配置里勾选是软链接。只支持一层软链。

应用监控不能识别 dal 框架，怎么办？

因为 dal 框架识别逻辑是，通过扫描应用工程里 zdal jar 包来确定版本号，扫描不到就无法识别。

批量移动数据源，对引用这个数据源的大盘有没有影响？

没有影响，大盘通过数据源本身进行数据拉取，与数据源所在的文件夹位置无关。

系统监控中 Ifout 与 Ifin 的单位

数据展现的单位 Btye，报警规则这里是 KB 单位。

配置和接入类

非标应用怎么接入 RMS？

- **标准 SOFA 应用**：上线后自动接入。
- **非标准 SOFA 应用**：只要在 CAFE 上发布上线，产品就会进行接入，只是展示的监控项目有限。

黑/白名单里各项之间的关系是 and 还是 or？

白名单是要 A 或 B，黑名单是不要 A 和 B。

黑白名单支持正则吗？

两者都是精确匹配，不支持正则。如需正则，请使用列值翻译。

配置 key 出现乱码，该怎么办？

现象：配置 key 出现乱码，如下图所示：



错误码

			19:47	19:46	19:45	19:44	19:43	19:42
错误码	错误码含义	错误码场景(服务接口)	COUNT	COUNT	COUNT	COUNT	COUNT	COUNT
AE0310266035	实际金额大于资产金额错误	0266	0	0	0	0	0	0
AE0310268151	ÏÎ³ÌÖ´ÐÐ³¬Ê±Ò³£	0268	0	0	0	0	0	0
AE0310268151	线程执行超时异常	0268	0	0	0	0	0	0
AE0310268151	������������������	0268	0	0	0	0	0	0
AE0310268438	¶¶ÈÏÞÈ ¨ ×ÉÑ¯³¬Ê±	0268	0	0	0	0	0	0
AE0310268438	额度限权咨询超时	0268	0	0	0	0	0	0
AE0310268438	������������������	0268	0	0	0	0	0	0
AE0310271003	决策请求参数为空	0271	0	0	0	0	0	0
AE0510266114	收款资产可用性校验异常	0266	0	0	0	0	0	0
AE0510266114	������������������������	0266	0	0	0	0	0	0

出现的乱码的原因，一般情况是日志编码与服务器编码不统一。

解决方案：进入 **编辑** 页面，**基本信息** 中打开高级选项，然后直接填写日志编码，如下图所示：



告警类

短信告警未收到

可以尝试如下方法：

1. 访问 报警明细看板，查看短信是否发出。
2. 是否在个人页面关闭了短信渠道。
3. 有时候是手机问题，重启手机便会收到短信。

配置了告警钉钉群没有收到信息，为什么？

钉钉群现在设有安全设置，匹配关键字才会通知，请前往钉钉群设置。



告警页面暂停后，消息是否还会继续在消息队列收到？

会继续在消息队列中收到。

11.资金安全监控
11.1. 产品简介
11.1.1. 概述

在金融、电商等行业，企业每天都在运转着大量资金业务，一旦由于系统原因或人为因素导致资金数据相关问题，若无法在第一时间发现问题，就会为企业带来巨大的经济损失。
资金安全监控产品旨在帮助您发现资金损失的风险。该产品可以通过执行用户定义的规则，按每天、每小时等多种时效策略，发现资金类数据问题，及时向您发出告警。您可以在第一时间收到告警，根据异常数据排查问题，分析原因，进而解决问题。

资金安全监控产品提供公有云和专有云两种不同的部署模式，支持客户在不同环境中对自身业务进行自定义核对检查。

优势

资金安全监控产品具有以下优势：

- 与业界成熟数据平台产品深度融合，提供从上层应用场景到底层基础资源的全面覆盖。
- 多种核对模式和频率，满足各类用户对于资损风险监控时效性的不同需求。
- 除传统的短信、邮件通知外，提供与用户平台对接的通知渠道，可为核心业务提供即时监控、即时应急保障。
- 提供资损风险监控覆盖的度量功能，并配备专家咨询服务，将多年积累的资损防控经验赋能给用户。

功能特性

- **数据节点**
  您可以在数据节点列表中统一管理资金业务表，通过节点命名、添加描述的方式详述表的业务语义，以便其他用户识别、引用。
- **核对点**
  您可以通过新建核对点（包含编写核对规则），来进行资金业务表或表与表之间的核对检查。核对点负责人可以进行编辑删除、上线下线等操作，当核对点为已上线状态时，系统会根据设定的核对类型进行周期性检查。您可以查看每个周期的核对规则运行情况。
- **核对场景**
  您可以将业务梳理为核对场景树（最多三级），核对点可与对应的核对场景绑定，便于从业务视角查看特定场景下的监控能力和异常情况。
- **通知订阅**
  您可自行订阅某个核对点上下线和修改更新、核对任务执行失败、执行发生异常告警等信息，选择邮件、钉钉的方式通知您。您也可以对多个核对点进行批量订阅。
- **异常列表**
  您可以查看核对规则粒度的异常情况，其中包括一次核对任务运行时发现的差异数据详情，每次核对任务最多展示差异数据的10 条抽样。
- **异常处理**
  您可以对异常记录处理结论，包含是否误报、是否资损以及备注等信息，关闭异常。异常状态变为已处理，并将以上处理结论通知核对点异常告警的订阅人员，便于及时复盘。
- **权限控制**
  首次使用资金安全监控产品，主账号用户必须先在阿里云的 DataWorks 平台上使用 MaxCompute 计算引擎创建用于核对的工作空间，并对资金安全监控账号进行授权。另外，如果您不是核对点或异常的负责人，则无权上下线该核对点或关闭该异常。

应用场景

- **业务护航**
  您可以阶段性地或定期梳理涉及资金的核心业务链路。通过配置核对点，全面覆盖业务资金链路；同时系统将会按照核对规则，对各数据做相互核对，或对数据内容做逻辑检查，以监控业务资损风险。
- **变更风险检查**
  业务变更发布前，您可以添加变更业务表的核对点，或者变更业务表与关联表的核对点，确保变更上线后不存在资损风险监控盲点。同理，技改类变更发布前，您也可以针对变更点，新增核对点，或修改原有核对点。
- **历史数据扫雷**
  批量检查存量业务的历史数据是否已存在异常，及时排查问题、分析原因，修复漏洞、追回损失以及避免更大的资损。
- **数据质量监控**
  数据缺漏时，也可能间接引发资损。您可以通过配置核对点，检查数据完整性，对数据质量作监控，及时发现异常数据。

11.1.2. 功能特性

数据源配置管理

您可以在产品中添加多种不同类型的数据源配置，每个数据来源配置代表一个核对系统在后续核对工作中可以用来读取数据的数据来源配置，当前支持如下类型的数据源配置：

- **OSS**：OSS 文件系统的读取配置，需要您输入访问信息，endpoint 和 bucket 信息等。
- **FTP**：从 FTP 系统中读取文件的配置，需要您输入 FTP 服务器访问信息。
- **DB**：从 DB 中读取数据的配置，需要您输入数据库访问配置信息，建议使用备库来进行核对。
- **离线**：离线数据仓库访问的配置信息。

数据源管理

您可以在产品中添加多种不同类型的数据源，每个数据源代表一个核对系统在后续核对工作中可以用来读取数据的数据来源，当前支持如下类型的数据源：

- **文件**：指定一个可以从 OSS、FTP 数据源配置中可以读取的文件，配置字段分割，解码方式等信息。
- **DB**：指定一个可以从 DB 数据源配置中访问的 DB 表信息、配置分库分表、关联表信息等，也支持通过 SQL 直接筛选数据。

数据节点管理

您可以通过新建数据节点来添加和管理不同类型的核对数据节点，可以完成如下的映射：

1. 将离线数据库中的一张表映射为一个核对数据节点，通过为数据节点添加名称以及关联核对点来更方便的进行核对数据节点的管理。
2. 将文件映射为一个核对数据节点，给出文件中需要参与核对的字段，排序和散列等信息。
3. 将数据库表映射为一个核对数据节点，给出表中需要参与核对的字段，排序和散列等信息。

所有在后续核对中需要使用到的文件，DB 表和离线表都应该被创建为核对点。

核对点管理

您可以通过创建核对点的方式来实现对待核对数据模型的管理。通过向核对点中添加多个节点，并在核对点中创建对应的核对规则来维护一个完整的核对模型。在核对点中，可以关联对应的核对场景，指定负责人，核对类型和延迟执行时长。
针对同一个数据模型（例如：从订单表到交易表的核对）所创建的核对规则，都应该被放到一个核对点中进行管理，方便后续的维护工作。
在创建核对点时需要选择该核对点所支持的核对类型，当前支持如下几种核对类型：

- **文件到文件核对**：在文件和文件之间进行数据比对核对。
- **文件到 DB 核对**：在文件和 DB 数据之间进行对比核对
- **DB 到 DB 核对**：在 DB 表和表之间进行对比核对。

- **离线核对**：通过在离线数据中执行 SQL 核对规则来进行数据校验。

核对规则管理

您可以通过创建规则，来进行资金业务表之间的核对检查。规则负责人可以对规则进行编辑、下线、上线等操作，当规则为已上线状态时，系统会根据该规则设定的核对频率，进行周期性检查。您可以查看每个周期的核对运行情况。

核对模板管理和核对模板导入

您可以通过创建核对模板的方式来实现核对数据模型的重用。例如：在不同业务中可能都牵涉到汇总表和流水表的核对，我们可以创建对应的核对模版和规则，在不同业务中进行导入重用。
创建核对模版的方式和创建核对点的方式基本一致，在核对模板下，也一样可以添加对应的规则。
在创建核对点的时候，可以选择通过导入核对模板的方式来实现重用。在导入时，需要对数据节点进行替换，更改核对规则名称。导入完成后，对应的规则和数据节点都会被自动添加到对应的数据节点中。

核对场景

在核对平台中，您可以手工添加和维护核对场景并在创建核对点的时候指定核对场景。通过这种方式来实现对核对节点的分类维护。
核对场景支持父子节点的管理，来实现对多级业务的映射。

通知订阅

您可以订阅某个规则的状态、异常、任务相关通知，指定通知方式及通知对象。

核对异常告警

当系统发现差异数据，并创建异常事件时，会通过短信、邮件、平台告警的方式通知您。

查看差异数据

您可以查看一次核对任务运行时发现的差异明细数据，一次最多展示 10 条抽样的差异明细数据。

关闭异常

您可以对异常事件记录原因，便于后续复盘。定位原因并解决问题后，您可以关闭异常来更新异常的状态，便于相关人员知晓异常的处理进展。

批量关闭误报的异常

因规则 SQL 不合理，或业务跨分区导致未核对到完整业务数据时，可能会是误报异常。您可以在异常列表页中勾选多个异常事件，点击批量关闭，标记误报并备注信息。

- **误报**：指因为 SQL 错误、数据错误、或系统异常等原因导致的和规则预期不符的情况。
- **非误报**：指系统发现的差异数据确实为异常现象，符合规则预期的情况。

11.1.3. 应用场景

业务护航

您可以阶段性地或定期梳理涉及资金链路的核心业务。可通过配置核对规则，对各类数据作相互核对，或对数据内容作逻辑检查，系统将按规则监控业务的资损风险。

变更风险检查

业务变更发布前，您可以添加变更业务表与关联表的核对规则，或者变更业务表数据的检查规则，确保变更上线后不存在资损风险监控盲点。同理，技改类变更发布前，您也可以针对变更点，新增核对规则，或修改原有核对规则。

历史数据扫雷

批量检查存量业务的历史数据是否已存在差异，可以及时分析差异原因，修复漏洞，追回资损。

数据质量监控

数据缺漏时，也可能间接引发资损。您可以配置核对规则，以检查数据完整性，对数据质量作监控，及时发现故障。

11.1.4. 基础术语

术语	说明
资金安全监控	SOFAStack 的一款风险防控类型的产品，能帮助您发现资金损失的风险。该产品可以通过执行您自定义的规则，以小时频率，准实时等多种时效策略，发现资金类数据问题，向您发出告警。您可以在第一时间收到告警，根据异常数据排查问题，分析原因，进而解决问题。
规则	资金安全监控产品中的规则，也叫核对规则，是数据核对的最小单位。可结合分类来管理。规则主要由 ODPS 表和 SQL 脚本构成，另外，您还可设置一些规则相关信息。规则的核对逻辑由 SQL 体现。
任务	资金安全监控产品中的任务（Task）指核对任务，即核对规则周期性执行的任务。每个周期有一组任务实例，每次执行对应一个任务实例。
任务周期	也叫核对周期，指核对任务的调度周期。如核对频率为天核对，就是一天执行一次，例如 2018.5.29 执行的核对任务对应周期 20180529；如核对频率为小时核对，就是一小时执行一次，例如 2018.5.29 日 10 点执行的核对任务对应周期 2018052910。
异常	核对任务运行中发现数据有差异，系统会创建异常事件（Issue）。 注意：任务重跑时可能会重复产生差异。
差异数据	也叫差异，指核对任务执行时（执行 SQL）发现的不匹配核对规则的数据。您可以在异常详情里查看在某个核对周期中发现的抽样差异数据。
差异数量	核对任务执行时发现的差异数据的总数。 注意：您在异常详情页看到的差异数量，是在当次任务周期中发现的差异数据总数，但是在差异数据（抽样数据），仅展现最多 10 条抽样的差异数据。
核对频率	核对频率有：按天核对，小时核对、5 分钟核对和自定义核对。 • 小时核对：一小时执行一次核对。 • 天核对：一天执行一次核对。 • 5 分钟核对：5 分钟执行一次核对。 • 自定义核对：根据用户配置的 cron 表达式触发核对执行。

11.2. 快速入门

11.2.1. 离线核对

本文以 **离线核对** 为例，引导您快速上手资金安全监控产品的主要操作链路。
总体流程如下图所示：



准备工作

在使用本产品前，您需要完成以下准备工作。后续步骤中用到的核对数据主要来自于数据仓库平台，因此首先需要将数据导入数据仓库平台。

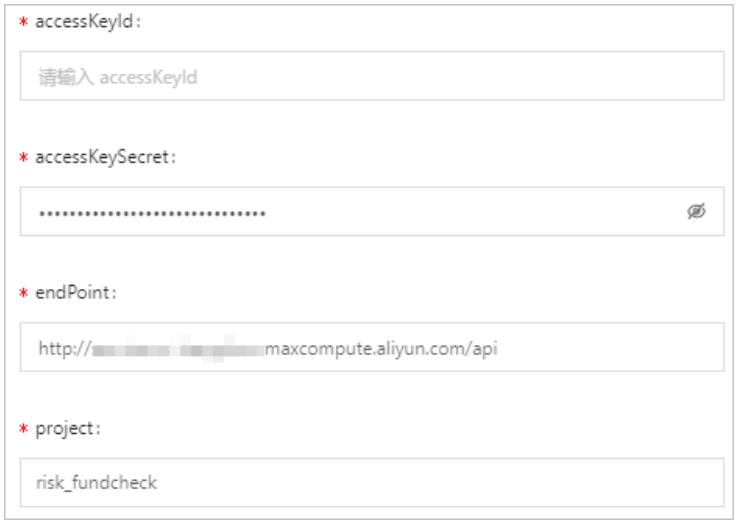
- 导入数据
- 创建数据同步任务

> 说明
> 上述内容仅适用于公有云环境。专有云请前往实际使用的数据仓库进行数据准备工作。

步骤一：新建数据源配置

进入资金安全监控控制台后，需创建数据源配置，以便将数据源配置和数据源关联。

1. 登录资金安全监控控制台，在左侧导航栏单击 **数据资源配置**。
2. 在 **数据源配置** 页面上，单击 **新建数据源配置**。
3. 在 **新建数据源配置** 页面上，完成以下数据源配置。
   - **数据源配置名称**：数据源配置的名称，租户内唯一。本例中输入“odps_risk_fundcheck”。
   - **类型**：支持 **OSS**、**FTP**、**DB**、**离线** 四种类型。本例中选择 **离线**。
   - 访问授权信息：需填写 **accessKeyId**、**accessKeySecret**、**endPoint** 和 **project**（本例中输入“risk_fundcheck”）。



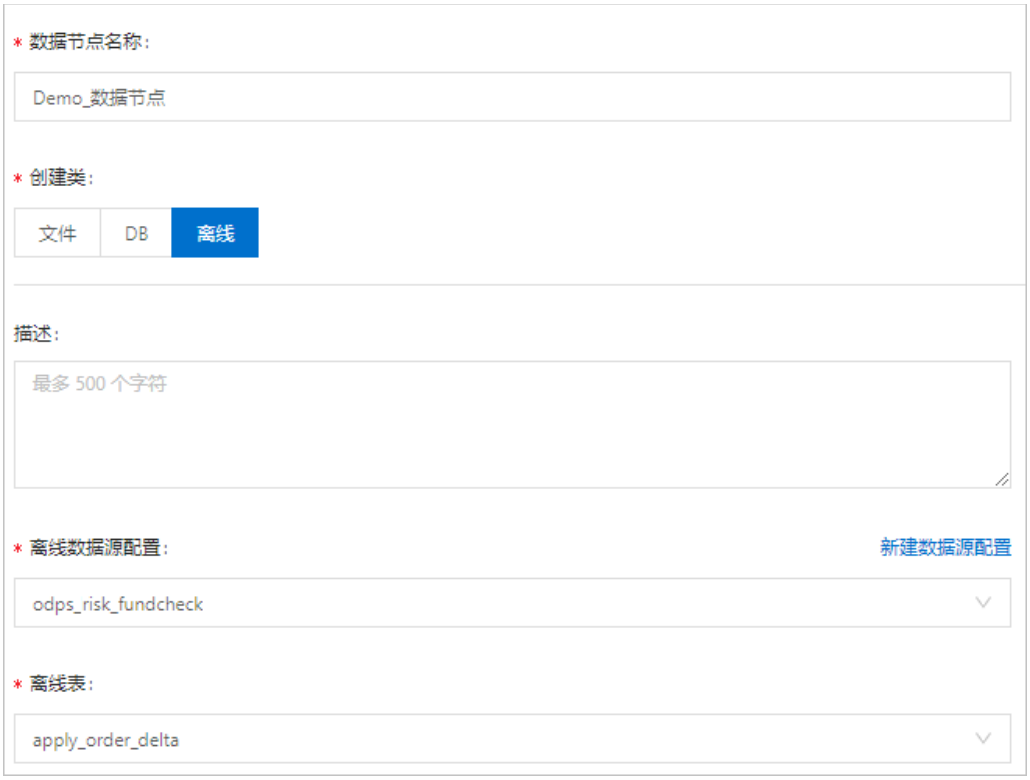
> 说明
> 关于 accessKeyId 和 accessKeySecret 的获取方式，请参见 获取 AccessKey。

4. 配置完成后，单击 **授权**。
   上述操作完成后，可在数据源配置列表页面，看到刚创建的数据源配置。

步骤二：新建数据节点

为了统一管理资金业务表，便于其他用户识别、引用，您可以将待检查的资金业务表定义为数据节点。

1. 登录资金安全监控控制台。
2. 单击左侧导航栏中的 **数据节点**，进入 **数据节点** 页面。
3. 单击 **新建数据节点**，在 **新建数据节点** 页面上，填写以下信息：



- **数据节点名称**：节点的名称，租户内唯一。本例中输入“Demo_数据节点”。
- **创建类**：支持 **文件**、**DB**、**离线**。本例中选择 **离线**。
- **描述**：选填，节点的备注信息，最大长度为 500 个字符。
- **离线数据源配置**：选择 **odps_risk_fundcheck**。
- **离线表**：选择用于核对的 ODPS 表（即选择 DataWorks 中通过数据采集在线表得到的离线表），例如“apply_order_delta”。

4. 单击 **提交** 完成数据节点的创建。
   上述操作完成后，可在数据节点列表页面，看到刚创建的节点 **Demo_数据节点**。

步骤三：新建核对点

核对点用做进行资金业务表或表与表之间的核对检查，找到异常数据。

1. 登录资金安全监控控制台，在左侧导航栏单击 **核对点列表**。
2. 单击 **新建核对点**，进入 **新建核对点** 页面。
3. 在 **新建核对点** 页面中，填写以下信息。
   - **基本信息**：
     - **核对点名称**：核对点的名称，租户内唯一。本例中为“Demo_核对点”。
     - **核对场景**：选择某一选定的核对场景。如果尚未创建核对场景，可单击右侧的 **新建核对场景**。详情请参见 创建核对场景。
     - **描述**：选填，核对点的备注信息，长度不能超过 500 个字符。
     - **负责人**：在租户下本产品的用户名单中选择，可选多人。
   - **核对配置**：
     - **核对方式**：支持 **文件-文件**、**文件-DB**、**DB-DB** 和 **离线**。本例中选择 **离线**。
     - **数据节点**：在已创建的数据节点中选择，可选多个。本例中选择 **Demo_数据节点**。

- **核对类型**：选择执行周期，提供 **小时核对** 和 **天核对** 两种选项。本例中选择默认值 **天核对**。
- **延迟执行时长**：指定每个周期的核对执行时间，默认值为 6 小时。

4. 配置完成后，单击 **提交**。

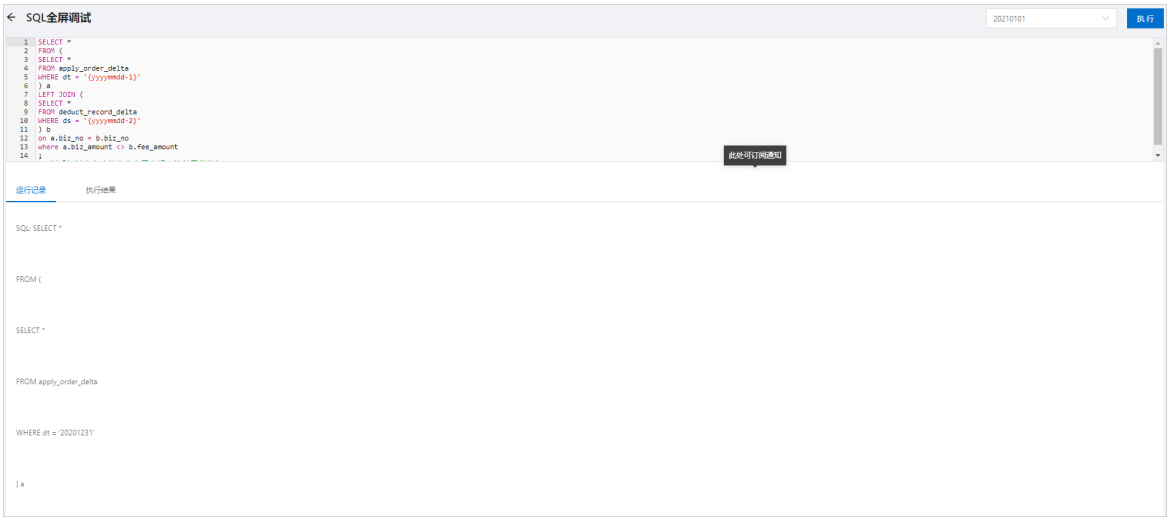
步骤四：新建核对规则

资金安全监控基于规则进行资金业务表之间的核对检查，要执行核对任务，需要先创建规则。

1. 在左侧导航栏上，单击 **核对点列表**。
2. 进入 **核对点列表** 页面后，单击已创建的核对点名称，本例中为 **Demo_核对点**。
3. 进入 **核对点详情** 页面，单击页面下方的 **新建核对规则** 按钮。
4. 在 **新建核对规则** 页面，填写如下配置信息：
   - **规则名称**：核对规则的名称，租户内唯一。
   - **描述**：规则的备注信息，最大长度为 500 个字符。
   - **节点信息**：数据节点的信息。
   - **规则文本**：规则的核对逻辑，按 SQL 语法编写（一般采用日期字段包含宏的格式，且需带有分区信息 `dt` 、 `hh` 或其他分区字段，便于周期性获取相应时间值），即在数据节点（也就是离线表）之间关联 `JOIN` ，指定不符合核对预期的条件，以便过滤异常数据。本例中输入：

```
SELECT *
FROM (
SELECT *
FROM apply_order_delta
WHERE dt ='{yyyymmdd-1}'
) a
LEFT JOIN (
SELECT *
FROM deduct_record_delta
WHERE ds ='{yyyymmdd-2}'
) b
on a.biz_no = b.biz_no
where a.biz_amount <> b.fee_amount
;
--过滤订单表和账单表中交易金额不等的异常数据
```

5. 填写完上述信息后，单击 **提交** 即可完成核对规则的创建。
6. （可选）对 SQL 进行 全屏调试，确保规则文本的准确性。

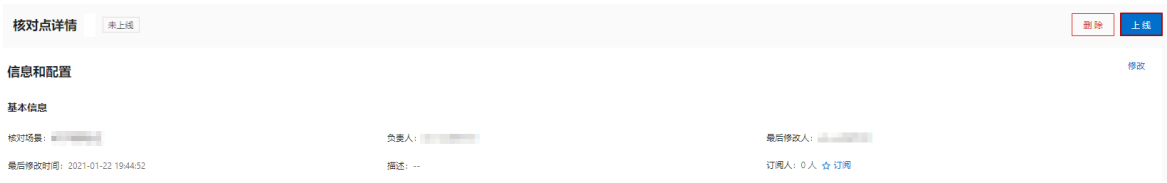


步骤五：上线核对点

仅状态为 **已上线** 的核对点才会进行监控，即周期性地运行核对任务，并在发现差异数据时创建异常。

> 说明
> 仅核对点负责人才拥有上下线权限。

1. 进入 **核对点详情** 页面，单击右上方 **上线** 按钮。



2. 核对点进入发布验证，状态变为 **上线中**。核对点成功上线后，核对点状态将变为 **已上线**。成功上线后，将按照指定的延迟执行时长，周期性执行核对任务，具体执行情况可在 **核对点详情** 页面的运行记录中查看。



步骤六：发现并处理异常

当核对点上线运行发现异常时，系统会按照核对规则粒度创建异常，记录在 **异常列表** 页面中。

> 说明
> 如果您 订阅 了核对点“执行发生异常告警”的通知内容，当异常创建时，系统会给您发送消息通知。

1. 在左侧导航栏上，单击 **异常列表**，进入 **异常列表** 页面。

2. 选择要处理的异常，单击其 ID，进入 **异常详情** 页面。



3. 在 **差异数据** 页面，单击 **详情** 查看异常详情数据，进行相应排查。



4. 单击 **关闭异常** 按钮，在 **关闭异常** 面板上填写如下信息，并单击 **提交**。
   - **是否误报**：
     - **误报**：例如 SQL 错误、数据错误或系统异常等，导致与核对预期不符的情况。
     - **非误报**：判定后确定是异常现象。
   - **是否资损**：
     - **有资损**：根据异常数据排查问题并分析原因，确定发生资金损失。
     - **无资损**：未出现资金损失。
   - **异常原因描述**：长度不超过 500 个字符。可以详细写明判断为误报或非误报、有资损或无资损的原因，便于后续复盘，也可以方便其他用户了解异常的原因。

11.2.2. 文件核对

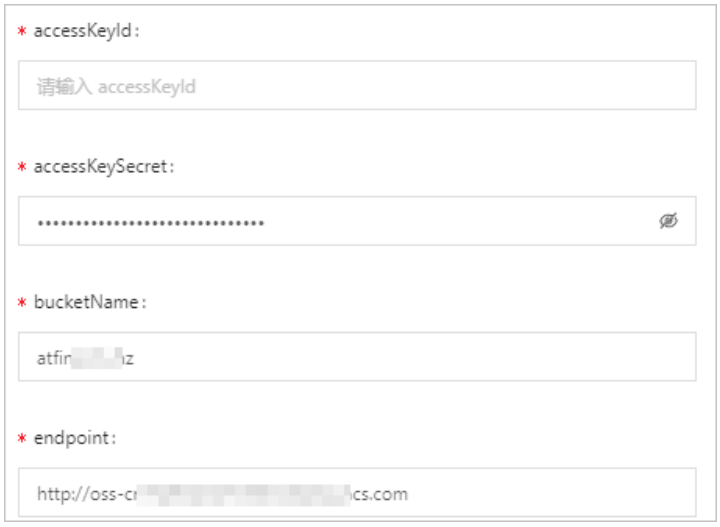
本文以 **文件核对** 为例，引导您快速上手资金安全监控产品的主要操作链路。
总体流程如下图所示：



步骤一：新建数据源配置

进入资金安全监控控制台后，需创建数据源配置，以便将数据源配置和数据源关联。

1. 登录资金安全监控控制台，在左侧导航栏单击 **数据资源配置**。
2. 在 **数据源配置** 页面上，单击 **新建数据源配置**。
3. 在 **新建数据源配置** 页面上，完成以下数据源配置。
   - **数据源配置名称**：数据源配置的名称，租户内唯一。本例中输入“Demo_数据源配置”。
   - **类型**：支持 OSS、FTP、DB、离线 四种类型。本例中选择 OSS。
   - 访问授权信息：需填写 accessKeyId、accessKeySecret、bucketName 和 endpoint。



> 说明
> 关于 accessKeyId 和 accessKeySecret 的获取方式，请参见 获取 AccessKey。

4. 配置完成后，单击 **授权**。
   上述操作完成后，可在数据源配置列表页面，看到刚创建的数据源配置。

步骤二：新建数据源

1. 登录资金安全监控控制台，在左侧导航栏单击 **数据源**。
2. 在 **数据源** 页面上，单击 **新建数据源**。
3. 在 **新建数据源** 页面上，输入以下配置信息：
   - **数据源名称**：数据源的名称。租户内唯一。本例中输入“Demo_数据源”。
   - **创建类**：提供 **文件** 和 DB 两个选项。本例中选择 **文件**。
   - **文件类型**：提供 OSS 和 FTP 两个选项。本例中选择 OSS。
   - **文件路径**：文件存放路径，支持正则表达式。本例中输入“test/file.txt”。
   - **数据源配置**：选择 Demo_数据源配置。如果尚未创建数据源配置，可单击右侧的 **新建数据源配置**。详情请参见 创建数据源配置。
   - **文件属性**：可选择 **表格按行录入** 或 **JSON 导入**。本例中选择 **表格按行录入**。
   - **文件编码**：文件编码类型，支持 UTF-8 和 GBK。
   - **文件协议**：用于定义文件的解析方式，目前支持 SP、DE 和 FUND 三种类型。
     - SP：简单协议布局模板（Simple Protocol）。
     - DE：蚂蚁集团自主提供的一种文件交互格式。

- FUND：国标基金数据协议。

- **分隔符**：字段间的分隔符。
- **字段**：配置以下参数。如需配置多条，单击 **新增属性** 即可。



4. 配置完毕后，单击 **提交** 完成创建。
5. 重复上述步骤 1-4 再创建一个数据源，本例中为“Demo_数据源 2"。

步骤三：新建数据节点

为了统一管理资金业务表，便于其他用户识别、引用，您可以将待检查的资金业务表定义为数据节点。

1. 登录资金安全监控控制台。
2. 单击左侧导航栏中的 **数据节点**，进入 **数据节点** 页面。
3. 单击 **新建数据节点**，在 **新建数据节点** 页面上，填写以下信息：



- **数据节点名称**：节点的名称，租户内唯一。本例中输入“Demo_数据节点”。
- **创建类**：支持 **文件**、**DB**、**离线**。本例中选择 **文件**。
- **描述**：节点的备注信息，选填。长度为 1-500 个字符。
- **数据源**：选择 **Demo_数据源**。如果尚未创建数据源配置，可单击右侧的 **新建数据源** 来创建。
- **参与核对字段**：需要参与核对的字段。对于 **文件** 是指文件属性中的字段。对于 **DB** 是指 SQL 中查询出来的字段。

  > 说明
  > 如果有多个需要参与核对的字段，则中间用英文逗号（,）隔开。

- **排序字段**：用于指定对当前数据进行排序所依据的字段。可从输入的核对字段中选择。
- **取数策略**：可选择 **每次取 1 行** 或 **取固定字段**。
- **散列字段**：用于数据分片时进行散列计算的字段。可以与排序字段保持一致。
- **散列位数**：自定义的散列位数。

- **高级配置**：默认保持关闭。
- **是否汇总核对**：默认保持关闭。

4. 配置完成后，单击 **提交**。
5. 重复上述步骤 1-4 再创建一个数据节点，本例中为“Demo_数据节点 2”。
   上述操作完成后，可在数据节点列表页面，看到刚创建的节点 **Demo_数据节点**、**Demo_数据节点 2**。

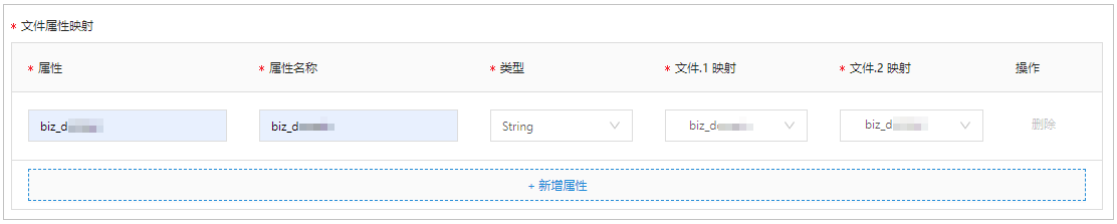
步骤四：新建核对点

核对点用做进行资金业务表或表与表之间的核对检查，找到异常数据。

1. 登录资金安全监控控制台，在左侧导航栏单击 **核对点列表**。
2. 单击 **新建核对点**，进入 **新建核对点** 页面。
3. 在 **新建核对点** 页面中，填写以下信息。
   - **基本信息**：
     - **核对点名称**：核对点的名称，租户内唯一。本例中为“Demo_核对点”。
     - **核对场景**：选择某一选定的核对场景。如果尚未创建核对场景，可单击右侧的 **新建核对场景**。详情请参见 创建核对场景。
     - **描述**：选填，核对点的备注信息，长度不能超过 500 个字符。
     - **负责人**：在租户下本产品的用户名单中选择，可选多人。
   - **核对配置**：
     - **核对方式**：支持 **文件-文件**、**文件-DB**、**DB-DB** 和 **离线**。本例中选择 **文件-文件**。
     - **文件 1**：用于核对的数据节点，选择 **Demo_数据节点** 。
     - **文件 2**：用于核对的数据节点。选择 **Demo_数据节点 2**。

       > **注意**
       > **文件 2** 所选的数据节点不能与 **文件 1** 所选的数据节点相同。

     - **文件属性映射**：配置属性参数（见下图）。如需配置多条，单击 **新增属性** 即可。

     - **触发条件**：支持 **定时任务** 和 **OpenAPI**。本例中选择 **定时任务**。
     - **定时任务**：事件触发规则执行的周期，本例中输入 `0 0 16 * * ?` ，表示每天 16 点触发。
     - **执行日历**：默认按 **自然日** 执行规则。
     - **延迟告警（分钟）**：在定时任务设置的时间点开始执行规则，如果在延迟告警配置的时长内没有执行完成，则触发告警通知。本例中选择 **300** 分钟。
4. 配置完成后，单击 **提交**。

步骤五：新建核对规则

资金安全监控基于规则进行资金业务表之间的核对检查，要执行核对任务，需要先创建规则。

1. 在左侧导航栏上，单击 **核对点列表**。

2. 进入 **核对点列表** 页面后，单击已创建的核对点名称，本例中为 **Demo_核对点**。
3. 进入 **核对点详情** 页面，单击页面下方的 **新建核对规则** 按钮。
4. 在 **新建核对规则** 页面中，填写如下配置信息：
   - **基本信息**
     - **规则名称**：核对规则的名称，租户内唯一。本例中输入“Demo_核对规则”。
     - **核对方式**：支持 **明细核对**、**汇总核对**、**自校验核对** 三种方式。
       - **明细核对**：端到端明细一一核对。
       - **汇总核对**：针对统计字段做汇总核对。
       - **自校验核对**：单文件针对文件头和文件体对应字段做汇总核对。
     - **描述**：规则的备注信息，选填，最大长度为 500 个字符。
   - **核对配置**
     - **节点信息**：展示数据节点的信息。
     - **规则表达式**：需配置相关参数（见下图）。如需配置多条，单击 **添加一行** 即可。



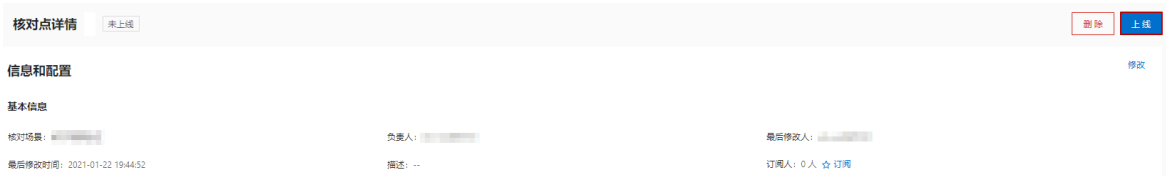
     - **高级配置**：默认保持关闭。
5. 填写完上述信息后，单击 **提交** 即可完成核对规则的创建。
6. （可选）对创建的核对规则执行 全屏调试，确保规则文本的准确性。

步骤六：上线核对点

仅状态为 **已上线** 的核对点才会进行监控，即周期性地运行核对任务，并在发现差异数据时创建异常。

> **说明**
> 仅核对点负责人才拥有上下线权限。

1. 进入 **核对点详情** 页面，单击右上方 **上线** 按钮。



2. 核对点进入发布验证，状态变为 **上线中**。核对点成功上线后，核对点状态将变为 **已上线**。成功上线后，将按照指定的延迟执行时长，周期性执行核对任务，具体执行情况可在 **核对点详情** 页面的运行记录中查看。



步骤七：发现并处理异常

当核对点上线运行发现异常时，系统会按照核对规则粒度创建异常，记录在 **异常列表** 页面中。

> **说明**
> 如果您 订阅 了核对点“执行发生异常告警”的通知内容，当异常创建时，系统会给您发送消息通知。

1. 在左侧导航栏上，单击 **异常列表**，进入 **异常列表** 页面。
2. 选择要处理的异常，单击其 ID，进入 **异常详情** 页面。



3. 在 **差异数据** 页面，单击 **详情** 查看异常详情数据，进行相应排查。



4. 单击 **关闭异常** 按钮，在 **关闭异常** 面板上填写如下信息，并单击 **提交**。
   - **是否误报**：
     - **误报**：例如 SQL 错误、数据错误或系统异常等，导致与核对预期不符的情况。

- **非误报**：判定后确定是异常现象。

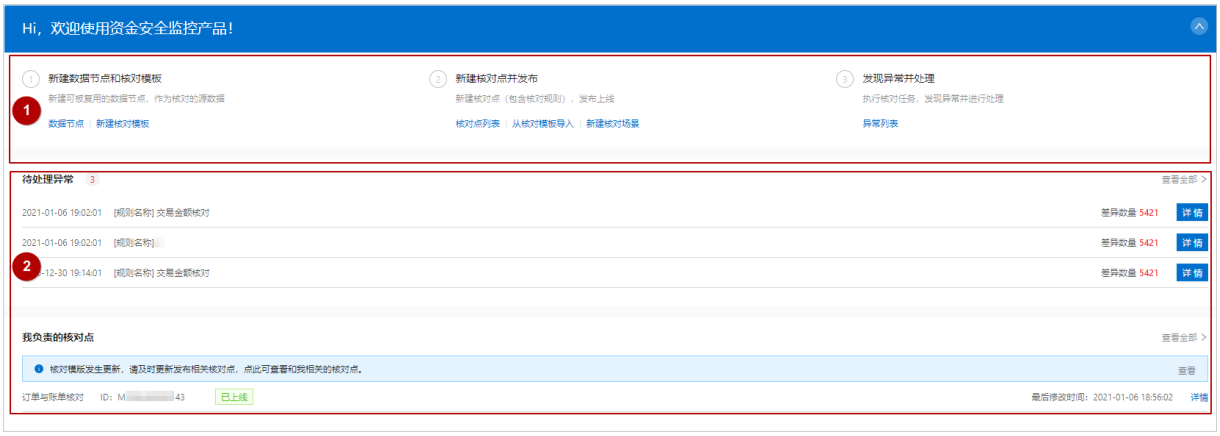
- **是否资损**：
  - **有资损**：根据异常数据排查问题并分析原因，确定发生资金损失。
  - **无资损**：未出现资金损失。

- **异常原因描述**：长度不超过 500 个字符。可以详细写明判断为误报或非误报、有资损或无资损的原因，便于后续复盘，也可以方便其他用户了解异常的原因。

11.3. 概览

资金安全监控旨在帮助您发现资金损失的风险。作为产品首页，概览页旨在让您了解产品全貌，从而对产品使用过程有一个直观的体验。此外，您还可以根据异常数据排查问题、分析原因，进而修复漏洞、避免资损。
该页面息包括以下内容：

- 操作引导区（图示中 ①）
- 工作区（图示中 ②）



操作引导区

登录资金安全监控控制台后，您将在 **概览** 页面上方将看到产品使用的核心步骤：**新建数据节点和核对模板** > **新建核对点并发布** > **发现异常并处理**。您可以单击该区域的 **数据节点**、**核对点列表**、**新建核对场景**、**异常列表** 进入对应页面，也可以单击该区域右侧的 **收起**，隐藏步骤区。

> 说明
> 新建核对模板为可选步骤，您可以根据需要自行选择。

工作区

工作区占概览页面大部分区域，便于您直观地看到需要处理的事项。

- 首次使用资金安全监控产品，您会看到关于新建核对点的提示。可通过单击 **新建核对点** 进入 新建核对点 页面来创建，也可以单击 **从核对模板导入** 来创建。
- 当已新建核对点或核对点执行发现异常时，工作区将展示 **待处理异常** 和 **我负责的核对点** 两部分信息。
  - **待处理异常**：展示您当前负责的核对点执行发生异常的概况，包含创建时间、规则名称和差异数量。单击右侧的 **详情** 可进入异常详情 页面，单击右上方的 **查看全部** 可进入所有待处理的 异常列表 页面。
  - **我负责的核对点**：展示当前负责的核对点的概况，包含核对点名称、核对点 ID、核对点状态和最后修改时间。单击右侧的 **详情** 可以进入 核对点详情 页面，单击 **查看全部** 可进入所有我负责的 核对点列表 页面。

11.4. 核对点

11.4.1. 管理核对点

本文为您介绍如何创建和编辑核对点。完成创建后，您可以在 **核对点列表** 页面，查看当前租户下的所有核对点。

核对点列表支持以下筛选方式：

- **核对点范围**：全部核对点 、我负责的核对点、我订阅的核对点。
- **核对场景**：全部核对场景，某一选定的核对场景。
- **指定条件**：核对点名称（支持模糊匹配）、核对点 ID、规则名称（支持模糊匹配）、规则 ID、数据节点名称（支持模糊匹配）。
- **核对类型**：全部、小时核对、天核对。
- **最后修改时间**：按升序或降序排序。
- **状态**：全部、未上线、上线中、已上线。

创建核对点

以下两种方法您可任选其一来创建核对点。

通过新建核对点进行创建

1. 登录资金安全监控控制台，在左侧导航栏单击 **核对点列表**。
2. 单击 **新建核对点**，进入 **新建核对点** 页面。
3. 在 **新建核对点** 页面中，填写以下信息。
   - **基本信息**：
     - **核对点名称**：核对点的名称，租户内唯一。由 1-100 个英文或中文字符组成，可包含数字或下划线（_）。
     - **核对场景**：选择某一选定的核对场景。如果尚未创建核对场景，可单击右侧的 **新建核对场景**。详情请参见 创建核对场景。
     - **描述**：选填，核对点的备注信息，长度不能超过 500 个字符。
     - **负责人**：在租户下本产品的用户名单中选择，可选多人。

       > **说明**
       > 若负责人已使用邮件通知的订阅功能，则默认为负责人订阅该核对点的邮件通知，如需修改请在 **核对点详情** 页面中修改订阅。

   - **核对配置**：
     - **核对方式**：支持 **文件-文件**、**文件-DB**、**DB-DB** 和 **离线**。
       根据选择的核对方式进行相应的配置，具体如下。

- 选择 **文件-文件**、**文件-DB** 或 **DB-DB** 时，需填写如下内容（以 **文件-文件** 为例）：
  - **文件 1**：用于核对的数据节点。如果数据节点还未创建，可直接单击右侧的 **新建数据节点** 。详情请参见 创建数据节点。
  - **文件 2**：用于核对的另一个数据节点。所选的数据节点不能与 **文件 1** 所选的数据节点相同。
  - **文件属性映射**：配置以下参数（见下表）。如需配置多条，单击 **新增属性** 即可。

参数名称	说明
属性	属性字段
属性名称	属性的名称
类型	字段的类型。支持如下类型： ■ String ■ BigDecimal ■ Date ■ Money ■ Integer
文件 1 映射	映射字段
文件 2 映射	映射字段

> **说明**
> 如果在第一个属性字段中填入 MatchID，则两个数据节点将按照第一个属性映射的字段进行数据匹配。

  - **触发条件**：支持 **定时任务** 和 **OpenAPI**。选择 **定时任务** 时，需设置 **定时任务**，即事件触发规则执行的周期，支持 CRON 表达式（CRON 语法格式：`秒 分 小时 日期 月份 星期`）。例如：`0 0 16 * * ?` 表示每天 16 点触发。
  - **执行日历**：默认按 **自然日** 执行规则。
  - **延迟告警（分钟）**：在定时任务设置的时间点开始执行规则，如果在延迟告警配置的时长内没有执行完成，则触发告警通知。
- 选择 **离线** 时，需填写如下内容：
  - **数据节点**：可选择多个已创建的数据节点。如果数据节点还未创建，可直接单击右侧的 **新建数据节点** 。详情请参见 创建数据节点。
  - **核对类型**：
    - **小时核对**：一小时执行一次核对。
    - **天核对**：一天执行一次核对。
  - **延迟执行时长**：可指定每个周期的核对执行时间。可根据离线数据同步完成时间指定，例如：小时核对一般在整点后 15 分钟左右数据同步完成，则可填 15。小时核对类型默认选择整点后 15 分钟执行核对，天核对类型默认选择 0 点后 6 小时执行核对。

4. 填写完上述信息后，单击 **提交**。
   上述操作完成后，可在核对点列表页面，看到刚创建的核对点。

通过核对模板进行创建

> **注意**
> 本方法仅适用于已创建核对模板的场景。若没有，请先 创建核对模板。

1. 登录资金安全监控控制台，在左侧导航栏单击 **核对点列表**。
2. 在 **核对模板列表** 页面上，从 **新建核对点** 下拉列表中选择 **从核对模板导入**。



3. 在 **导入核对模板** 页面上，选择所需的核对模板，或者通过输入核对模板名称来选择。



4. 在 **新建核对点信息**，填写如下配置信息：



   - **核对点名称**：自定义的核对点名称，由中英文或数字组成，长度为 1-100 个字符。

- **核对场景**：系统会根据所选核对模板，自动选择对应的核对场景。也可以选择其他核对场景。
- **数据节点**：在已创建的数据节点中选择，只能选择一个。
- **核对规则名称**：系统会根据所选的核对模板，自动选择对应的核对规则名称。也可以自行填写规则名称。
- **负责人**：在租户下本产品的用户名单中选择，可选多人。

5. 填写完上述信息后，单击 **提交**。

查看核对点详情

单击核对点列表中核对点名称，进入 **核对点详情** 页面。核对点详情页面包含 **操作**、**信息和配置**、**核对规则** 和 **运行记录**。

- **操作**：
  - **删除**：删除当前核对点。如果该核对点已上线，需要先下线，否则无法删除。
  - **上线**：单击 **上线**，核对点进入发布验证，状态变为 **上线中**。验证需要时间，发布失败会通知已订阅 **核对点修改和上下线** 通知内容的用户。核对点成功上线后，核对点状态变为 **已上线**。
  - **下线**：单击 **下线**，核对点状态变为 **未上线**，不执行核对。
  - **发布更新**：核对点上线后，负责人对核对点的核对配置或核对规则进行了修改，单击 **发布更新** 替换线上版本。

> **说明**
> - 仅核对点负责人才可进行删除、上线和发布更新操作。
> - 只有状态为 **已上线** 的核对点才会进行监控，即周期性地运行核对任务，并在发现差异数据时创建异常。

- **信息和配置**：主要包含以下两部分内容。
  - 基本信息（图示中 ①）
  - 核对配置（图示中 ②）



- **核对规则**：可查看该核对点的所有核对规则。
  - 单击 **新建核对规则**，进入 **新建核对规则** 页面。

- 单击核对规则的 **展开**，展示规则文本。
- 单击规则 ID 后的 **修改**，进入 **修改核对规则** 页面。
- 单击规则文本右侧的 **调试**，进入 **SQL全屏调试** 页面。

- **运行记录**： 可查看该核对点的所有核对规则的运行情况。

修改核对点

1. 登录资金安全监控控制台，在左侧导航栏单击 **核对点列表**。
2. 单击要修改的核对点名称，进入 **核对点详情** 页面。
3. 单击 **信息和配置** 右侧的 **修改** 操作，进入 **修改核对点** 页面。
4. 根据需要修改相应信息，然后单击 **提交**。

> **说明**
> 仅核对点的负责人才可以修改该核对点。

11.4.2. 创建核对规则

资金安全监控基于规则进行资金业务表之间的核对检查，要执行核对任务，您需要先创建规则。目前，资金安全监控支持文件与文件核对、文件与 DB 核对、DB 与 DB 核对、离线核对，可以创建不同核对频率的规则，包括小时核对、天核对和文件核对等。

离线核对规则

1. 在左侧导航栏上，单击 **核对点列表**。
2. 进入 **核对点列表** 页面后，单击已创建的离线核对点名称。
3. 进入 **核对点详情** 页面，单击页面下方的 **新建核对规则** 按钮。



4. 在 **新建核对规则** 页面中，填写如下配置信息：
   - **规则名称**：由中文、英文或数字组成，长度为 1-100 个字符，租户内唯一，本例中填写“交易金额核对”。
   - **描述**：规则的备注信息，选填，最大长度为 500 个字符。
   - **节点信息**：数据节点的信息。
   - **规则文本**：规则的核对逻辑，按 SQL 语法编写（一般采用日期字段包含宏的格式，便于周期性获取相应时间值），即在数据节点（也就是离线表）之间关联 JOIN，指定不符合核对预期的条件，以便过滤异常数据。本例中输入：

```
SELECT *
FROM (
SELECT *
FROM apply_order_delta
WHERE dt ='{yyyymmdd-1}'
) a
LEFT JOIN (
SELECT *
FROM deduct_record_delta
WHERE ds ='{yyyymmdd-2}'
) b
on a.biz_no = b.biz_no
where a.biz_amount <> b.fee_amount
;
--过滤订单表和账单表中交易金额不等的异常数据
```

> **说明**
>
> - 一个核对规则中仅能填写一个 SQL 语句。
> - 您可先对 SQL 进行调试后再提交规则，详见 SQL 全屏调试 操作说明。

5. 填写完上述信息后，单击 **提交**。

文件与文件核对规则

由于文件与文件核对、文件与 DB 核对、DB 与 DB 核对规则的创建方式类似，本部分将介绍文件与文件核对规则的创建方式。

1. 在左侧导航栏上，单击 **核对点列表**。
2. 进入 **核对点列表** 页面后，单击 **文件-文件** 类型的核对点名称。
3. 进入 **核对点详情** 页面，单击页面下方的 **新建核对规则** 按钮。
4. 在 **新建核对规则** 页面中，填写如下配置信息：
   - **基本信息**
     - **规则名称**：核对规则的名称，租户内唯一。由 1-100 个英文或中文字符组成，可包含数字或下划线（_）。
     - **核对方式**：支持 **明细核对**、**汇总核对**、**自校验核对** 三种方式。
       - **明细核对**：端到端明细一一核对。
       - **汇总核对**：针对统计字段做汇总核对。
       - **自校验核对**：单文件针对文件头和文件体对应字段做汇总核对。
     - **描述**：规则的备注信息，选填，最大长度为 500 个字符。
   - **核对配置**
     - **节点信息**：数据节点的信息。

- **规则表达式**：需配置相关参数（见下图）。如需配置多条，单击 **添加一行** 即可。



- **高级配置**：默认保持关闭。如果开启，需输入规则表达式，详情请参见 高级配置。

5. 填写完上述信息后，单击 **提交**。

11.4.3. 编辑核对规则

本文介绍修改核对规则的标准操作流程。您可以在核对点详情页编辑核对规则，修改规则基本信息、规则表达式等信息。

操作步骤

> 说明
> 仅核对规则负责人才能修改核对规则。

1. 在 **核对点详情** 页面，单击核对规则区域的 **修改** 。
2. 进入 **修改核对规则** 页面，根据需要修改相应信息。
3. 修改完毕后，单击 **提交**。

11.4.4. SQL 全屏调试

调试功能用于检查 SQL 的正确性。

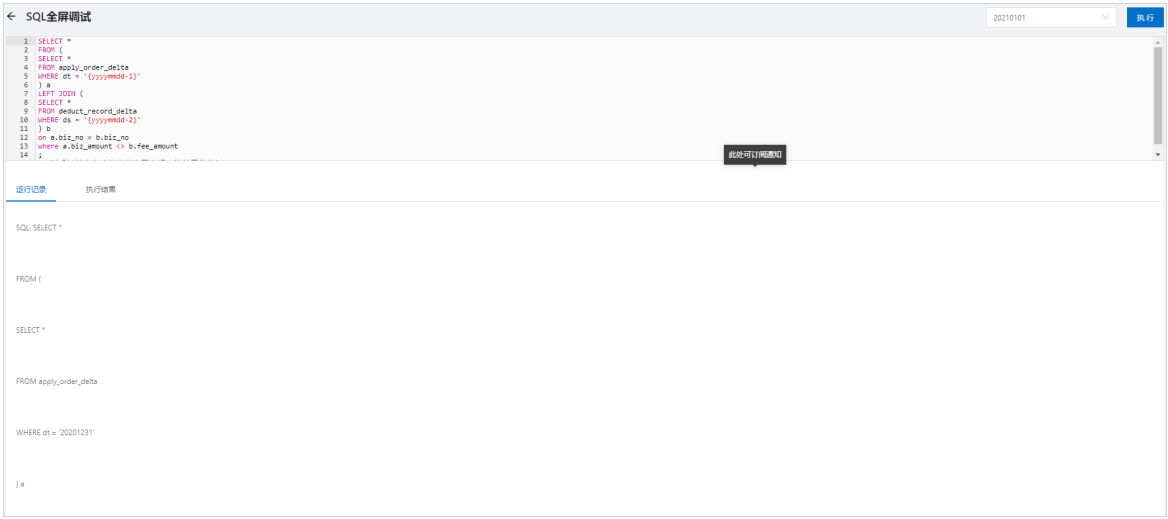
前提条件

已创建核对规则。

操作步骤

1. 在 **核对点详情** 页面，单击核对规则区域的 **调试** 。

2. 进入 **SQL 全屏调试** 页面，选择调试周期，单击 **执行** 即可完成调试操作 。



> **说明**
> 在 **SQL 全屏调试** 页面可查看运行记录。若运行成功，可查看执行结果。

11.4.5. 高级配置

在 **创建文件核对规则** 的过程中，您可以在设置核对配置时进行以下高级配置。

> **说明**
> **高级配置** 仅适用于创建文件核对规则的场景。对于离线核对规则，未提供该选项。

规则表达式

如果在设置规则表达式时开启 **高级配置** 开关，您可以使用 **qlExpress** 或 **Groovy** 自定义较为复杂的表达式。

- 若选择 **qlExpress**，示例如下：
  - 引用数据源或文件时，需要使用数据节点名称。
  - 多个规则用英文逗号 [,] 隔开。
  - 获取数据节点上的数据。使用时须使用字段对应的映射字段。

```
DCD0191204k29o26m1.fileamount < DCD019120997hcfyh1.payamount , DCD0191204k29o26m1.fileamount
<= DCD019120997hcfyh1.payamount
```

- 若选择 **Groovy**，示例如下：

> **注意**
> **模型** 一词非字面义，本文特指 **映射字段**，即字段映射关系。以 `核对模型,一左一右` 为例，是指会根据映射字段分别为两个数据节点生成两个模型。

```
import com.alibaba.common.lang.StringUtil

import com.alibaba.fastjson.JSONObject
import org.slf4j.LoggerFactory;
import com.alipay.finorm.common.util.LoggerConstants;
import com.alipay.finorm.common.util.LoggerUtil;
import com.alipay.finorm.common.facade.enums.RuleResultErrorCode
import com.alipay.finorm.common.facade.sdk.GuardianContext
import com.alipay.finorm.common.facade.model.immune.RuleResult

def ruleExecute(rule, immuneModelList) {

    def LOGGER = LoggerFactory.getLogger(LoggerConstants.BIZ_GUARDIAN);

    def ruleResult = new RuleResult()
    ruleResult.setRuleId(rule.ruleId)
    ruleResult.setRuleName(rule.ruleName)
    ruleResult.setIsSuccess(true)

    if (immuneModelList==null){
       return ruleResult
    }

    //上下文 map
    def tmModel = GuardianContext.getBizEventContext();

    //核对模型,一左一右,可以通过模型id等关键信息判断具体谁左谁右
    def model1 = immuneModelList.get(0);
    def model2 = immuneModelList.get(1);

    //获取模型上的数据,明细核对时候,是一行数据的fieldName; 汇总核对时候,是自定义的汇总字段fieldName;
    Map<String, Object> kv1 = model1.getFieldName2Value();
    Map<String, Object> kv2 = model2.getFieldName2Value();
    //汇总数值  ;  明细核对时候,可以获取到明细数值
    Object id_sum1 = kv1.get("fileamount");
    Object id_sum2 = kv2.get("payamount");

    //可以统一转为bigdecimal的方式处理

    def isSuccess = id_sum1 > id_sum2;

    // 核对结果
    ruleResult.setIsSuccess(isSuccess);
    // 如果失败,失败的原因可以在这里放进去
    ruleResult.setRuleErrorCode(" fileamount :" + id_sum1 +"  [class:"+id_sum1.class + "] vs p
ayamount :" + id_sum2+"  [class:"+id_sum2.class + "]" )

    打印日志
    LoggerUtil.info(LOGGER, " ###groovy###  immuneModelList :" + JSONObject.toJSONString(immun
eModelList) + "   ruleResult :" + JSONObject.toJSONString(ruleResult));

    return ruleResult

}
```

11.5. 异常列表

当系统发现差异数据后，会创建异常事件。您可以在 **异常列表** 页面，查看当前租户下的所有异常。
目前支持按照以下条件筛选异常：

- **异常范围**：全部异常、我订阅的异常、待处理异常。
- **核对场景**：全部核对场景，某一选定的核对场景。
- **指定条件**：规则名称（支持模糊匹配）。

异常详情

单击异常列表中的异常 ID，进入 **异常详情** 页面。异常详情页面包含 **处理结论**、**基本信息** 和 **差异数据**。

- **处理结论**： 处理结论仅在异常状态为 **已关闭** 时可见。



- **基本信息**：
  - **负责人**：异常的负责人即核对点的负责人。
  - **核对点**：执行发生该异常对应的核对点名称，单击即可进入 **核对点详情** 页面。
  - **核对规则**：执行发生该异常对应的核对规则名称，单击即可进入 **核对点详情** 页面。
  - **创建时间**：异常的创建时间。
  - **核对场景**：核对点绑定的核对场景名称。
  - **处理人**：关闭异常的人。
  - **关闭时间**：异常的关闭时间。
  - **处理时长**：异常的关闭时间减去异常的创建时间得到的时长。
  - **差异数量**：核对任务执行时发现的差异数据的总数。



- **差异数据**：核对任务执行时发现的不符合核对规则的数据，单击 **详情** 进入 **差异数据详情** 页面。如果您没有该异常涉及表的查询权限，则提示“无该异常涉及表的查询权限，请联系 MaxCompute 项目管理员开通”。

差异数据详情

需具有该异常涉及表的查询权限，才可进入 **差异数据详情** 页面，查看核对任务执行时发现的与核对预期不符的数据，每个核对任务最多展示差异数据的 10 条抽样。

数据列表支持按照任务周期的筛选方式。

关闭异常

> **说明**
> 仅核对点负责人才可以执行关闭异常操作。

1. 单击异常列表中的异常 ID，进入 **异常详情** 页面。



2. 单击 **关闭异常** 按钮，在 **关闭异常** 面板上填写如下信息，并单击 **提交**。



- **是否误报**：

- **误报**：例如 SQL 错误、数据错误或系统异常等，导致与核对预期不符的情况。
- **非误报**：指系统发现的差异数据确实为异常现象，符合核对预期的情况。

- **是否资损**：
  - **有资损**：根据异常数据排查问题、分析原因，确定发生资金损失。
  - **无资损**：虽然发生了异常，但是没有发生资金损失。
- **异常描述**：可以详细写明判断为误报或非误报、有资损或无资损的原因，便于后续复盘，也可以方便其他用户了解异常的原因。

11.6. 核对模板

通过 **核对模板** 模块，您可以将常用的核对点信息保存为模板，便于快速创建同类别的核对点，从而高效地实现了模板的快速创建与后期管理，大幅简化了核对点的创建过程。

创建核对模板

1. 登录资金安全监控控制台，在左侧导航栏单击 **核对模板**。
2. 在 **核对模板列表** 页面，单击 **新建核对模板**。
3. 在 **新建核对模板** 页面，输入以下配置信息。
   - **基本信息**
     - **核对模板名称**：由 1-100 个英文或中文字符组成，可包含数字或下划线（_），租户内唯一。
     - **核对场景**：选择核对场景，用于为核对模板绑定所属的业务场景。如果尚未创建核对场景，请单击右侧 **新建核对场景**，具体步骤请参见 创建核对场景。
     - **描述**：选填，核对模板的备注信息，最大长度为 500 个字符。
     - **负责人**：在租户下本产品的用户名单中选择，可选择多位负责人。
   - **核对配置**

- **核对方式**：支持 **文件-文件**、**文件-DB**、**DB-DB** 和 **离线**。
  - 选择 **文件-文件**、**文件-DB** 或 **DB-DB** 时，需填写如下内容（以 **文件-文件** 为例）：
    - **文件 1**：选择用于核对的数据节点，只能选择一个。如果数据节点还未创建，可直接单击右侧的 **新建数据节点** 。详情请参见 创建数据节点。
    - **文件 2**：选择用于核对的数据节点，只能选择一个。所选的数据节点不能与 **文件 1** 所选的数据节点相同。
    - **文件属性映射**：配置相关参数（见下图）。如需配置多条，单击 **新增属性** 即可。



    - **触发条件**：支持 **定时任务** 和 **OpenAPI**。选择 **定时任务** 时，需设置以下参数。
      选择 **定时任务** 时，需设置以下参数。**定时任务**：事件触发规则执行的周期，支持 CRON 表达式（CRON 语法格式：`秒 分 小时 日期 月份 星期`）。例如：`0 0 16 * * ?` 表示每天 16 点触发。
    - **执行日历**：默认按 **自然日** 执行规则。
    - **延迟告警（分钟）**：在定时任务设置的时间点开始执行规则，如果在延迟告警配置的时长内没有执行完成，则触发告警通知。
  - 选择 **离线** 时，需填写如下内容：
    - **数据节点**：在已创建的数据节点中选择，可选择多个。如果尚未创建数据节点，请单击右侧 **新建数据节点**，具体步骤请参见 创建数据节点。
    - **核对类型**：选择执行周期，当前提供 **小时核对** 和 **天核对** 两种选项。默认类型是 **天核对**。
    - **延迟执行时长**：指定每个周期的核对执行时间，默认值为 6 小时。

4. 单击 **提交** 即可完成核对模板创建。

后续操作

您可以依照创建好的核对模板新增核对规则，详情请参见 创建核对规则。

其他操作

> **说明**
> 仅核对模板负责人可进行修改或删除操作。

- 修改核对模板：在 **核对模板详情** 页面，单击 **修改**，并根据需要修改相应信息，单击 **提交**。
- 删除核对模板：在 **核对模板详情** 页面，单击 **删除**，在弹出的信息提示框中，单击 **是**。

11.7. 数据节点

本文为您介绍如何创建数据节点。完成创建后，您可以在 **数据节点** 页面，查看当前租户下的所有数据节点。节点列表支持按照数据节点名称进行筛选（支持模糊匹配），还支持按类型（**文件**、**DB**、**离线**）进行筛选。

创建数据节点

1. 登录资金安全监控控制台，在左侧导航栏单击 **数据节点**。
2. 在数据节点列表页，单击 **新建数据节点**。

3. 在 **新建数据节点** 页面，填写以下基本信息：
   - **数据节点名称**：节点的名称，租户内唯一。长度为 1-100 个字符，可包含中英文、数字和下划线（_）。
   - **创建类**：支持 **文件**、DB、**离线**。一旦完成创建，该项不能修改。
   - **描述**：节点的备注信息，选填。长度为 1-500 个字符。
4. 根据上一步选择的 **创建类**，进行相应的配置。
   - **文件/DB**
     - **数据源**：选择对应的数据源配置，也可通过输入数据源配置名称关键字来选择。如果尚未创建数据源配置，可单击右侧的 **新建数据源配置**。详情请参见 创建数据源配置。
     - **参与核对字段**：需要参与核对的字段。对于 **文件** 是指文件属性中的字段。对于 DB 是指 SQL 中查询出来的字段。

       > 说明
       > 如果有多个需要参与核对的字段，则中间用英文逗号（,）隔开。

     - **排序字段**：用于指定对当前数据进行排序所依据的字段。可从输入的核对字段中选择。
     - **取数策略**：可选择 **每次取 1 行** 或 **取固定字段**。
       - **每次取 1 行**：核对时从排序后的数据中两边各取一条数据，然后进行比较。
       - **取固定字段**：将值相同的数据一次取出。

         > 注意
         > 该固定字段必须与排序字段相同。

     - **散列字段**：用于数据分片时进行散列计算的字段。当核对的数据量较大时，可选择参与核对字段的某个字段，针对某几位进行散列，读取到的数据就会按照散列后的值散列到不同分片中，支持后续并发排序和核对，从而提升核对效率。

       > 说明
       > 散列字段可以与排序字段保持一致。

     - **散列位数**：自定义的散列位数。例如 4-6，表示取 4、5 两位。
     - **高级配置**：默认保持关闭。如果开启，则需要填写过滤表达式。

       > 说明
       > 关于 **过滤表达式**，即对数据进行逐条过滤处理，支持 Groovy 语法。

- **是否汇总核对**：默认保持关闭。如果开启，需配置相关参数（见下图）。如需配置多条，单击 **新增属性** 即可。



- **离线**
  - 离线数据源配置：选择对应的离线数据源配置，也可通过输入数据源配置名称关键字来选择。如果尚未创建离线数据源配置，可单击右侧的 **新建数据源配置**。详情请参见 创建数据源配置。
  - 离线表：选择用于核对的 ODPS 表。

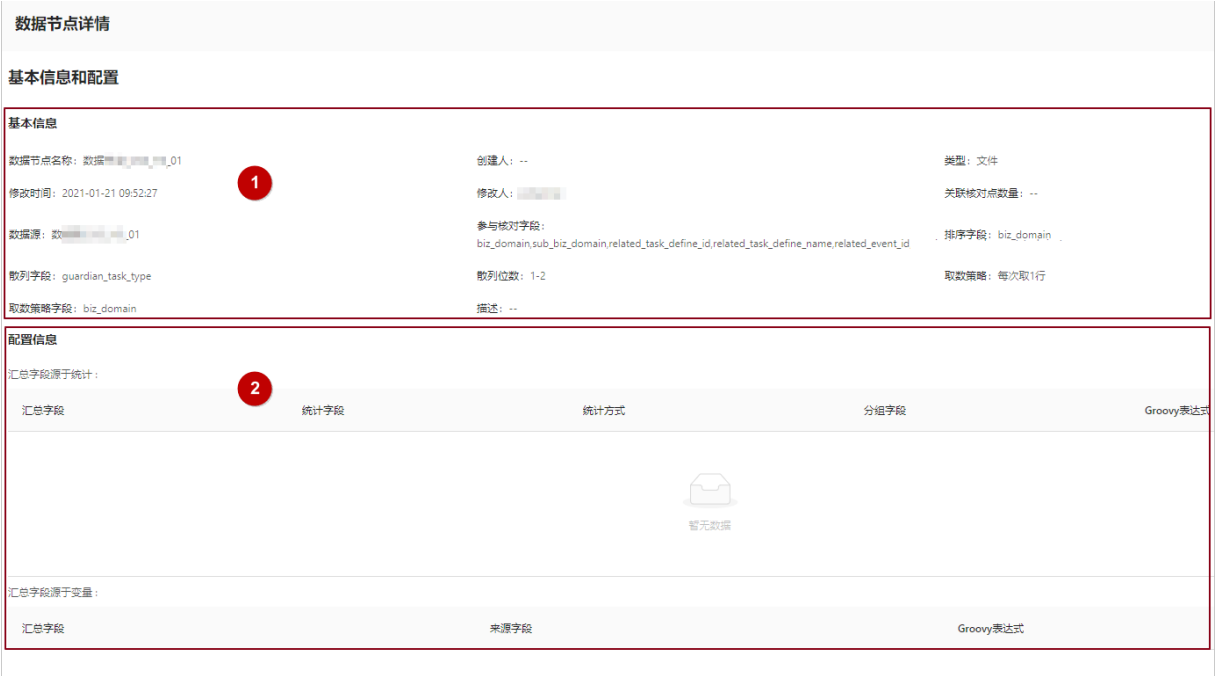
> **说明**
> ODPS 表支持模糊匹配，如果匹配不到，建议您前往 DataWorks 平台确认项目名、表名。

5. 配置完成后，单击 **提交**。
   上述操作完成后，可在数据节点列表页面，看到刚创建的节点。

查看数据点详情

单击数据节点列表内数据节点名称，进入 **数据节点详情** 页面，即可查看数据节点详情，主要包含以下两部分内容：

- 基本信息（图示中 ①）
- 配置信息（图示中 ②）



修改数据节点

1. 单击数据节点列表操作列的 **修改**，进入 **修改数据节点** 页面。
2. 根据需要修改节点信息。
3. 完成后单击 **提交** 保存修改。修改后的节点信息将在关联的核对点中同步生效。

11.8. 数据源

11.8.1. 数据资源配置

登录资金安全监控控制台后，需创建数据源配置，将数据源配置和数据源关联。本文介绍如何创建数据源配置。

创建数据源配置

1. 登录资金安全监控控制台，在左侧导航栏单击 **数据资源配置**。
2. 在 **数据源配置** 页面上，单击 **新建数据源配置**。
3. 在 **新建数据源配置** 页面上，完成以下数据源配置。
   - **数据源配置名称**：数据源配置的名称。租户内唯一，长度为 1-100 个字符，可包含中英文、数字、下划线（_）和短横线（-）。
   - **类型**：支持 **OSS**、**FTP**、**DB**、**离线** 四种类型。可根据您的实际业务需求自行选择。类型一旦选择确认后，将不能更改。
   - 访问授权信息：针对上述不同类型，需要填写的访问授权信息有所不同，具体如下。
     - **OSS**：需填写 accessKeyId、accessKeySecret、bucketName 和 endpoint。

       > 说明
       > 关于 accessKeyId 和 accessKeySecret 的获取方式，请参见 获取 AccessKey。

     - **FTP**：需填写 accessKeyId、accessKeySecret、hostName 和 port。
     - **DB**：需填写 **数据库名称**、**用户名**、**密码**、host 和 port。

- **离线**：需填写 accessKeyId、accessKeySecret、endPoint 和 project。

4. 配置完成后，单击 **授权**。
   上述操作完成后，可在数据源配置列表页面，看到刚创建的数据源配置。

编辑数据源配置

您可在 **数据源配置** 页编辑数据源配置，修改数据源配置名称和访问授权信息。
具体操作步骤如下：

1. 登录资金安全监控控制台，在左侧导航栏上单击 **数据资源配置**。
2. 在数据源配置列表中选择相应的数据源配置，单击 **操作** 列下的 **修改**。
3. 进入 **修改数据源配置** 页面，根据需要修改相应信息。
4. 修改完成后，单击 **授权**。

> 说明
> 数据源配置创建后，类型无法修改。

11.8.2. 创建数据源

本文为您介绍如何创建数据源，您可以添加文件类数据源，也可以针对 DB 类创建数据源。

创建文件类数据源

1. 登录资金安全监控控制台，在左侧导航栏单击 **数据源**。
2. 在 **数据源** 页面上，单击 **新建数据源**。
3. 在 **新建数据源** 页面上，输入以下配置信息：
   - **数据源名称**：数据源的名称。租户内唯一，长度为 1-100 个字符，可包含中英文、数字、下划线（_）和短横线（-）。
   - **创建类**：选择 **文件**。一旦完成创建，该项不能修改。
   - **文件类型**：提供 OSS 和 FTP 两个选项。
   - **文件路径**：文件存放路径，支持正则表达式。
   - **数据源配置**：选择对应的数据源配置，也可通过输入数据源配置名称关键字来选择。如果尚未创建数据源配置，可单击右侧的 **新建数据源配置**。详情请参见 创建数据源配置。
   - **文件属性**：可选择 **表格按行录入** 或 **JSON 导入**。本例中选择 **表格按行录入**。

     > 说明
     > 若选择 JSON 导入，只需进行以下配置即可完成数据源的创建：
     > - 文件模板：提供 中登 01、中登 02、中登 03、中登 04、中登 05、中登 06、中登 07 七个模板选项。可根据业务需求自行选择。
     > - JOSN 数据：无需手动输入，系统会根据所选的 **文件模板** 自动匹配 JSON 数据。

   - **文件编码**：文件编码类型，支持 UTF-8 和 GBK。
   - **文件协议**：文件协议定义了文件的解析方式，目前支持 SP、DE 和 FUND 三种类型。详情请参见下表。

协议类型	说明	适用场景

协议类型	说明	适用场景
SP	简单协议布局模板（Simple Protocol）	■ 文件没有文件头和列头 ■ 文件只有列头和文件内容
DE	蚂蚁集团自主提供的一种文件交互格式	适用于第一行为文件头、第二行为列头、从第三行往后都是文件内容的情况
FUND	国标基金数据协议	适用于前面每行都是头，后面是文件内容的情况，且文件内容是按长度进行分割的情况

- **分隔符**：字段间的分隔符。
- **字段**：配置以下参数。如需配置多条，单击 **新增属性** 即可。



  - **分类**：支持 body（文件体）和 head（文件头）两种类型。
  - **字段名称**：字段的名称。
  - **描述**：关于字段的简要说明。
  - **类型**：支持 String、BigDecimal、Date、Money、Integer。
  - **长度**：字段的长度。

4. 配置完毕后，单击 **提交** 完成创建。

创建 DB 类数据源

1. 登录资金安全监控控制台，在左侧导航栏单击 **数据源**。
2. 在 **数据源** 页面上，单击 **新建数据源**。
3. 在 **新建数据源** 页面上，输入以下配置信息：
   - **数据源名称**：数据源的名称。租户内唯一，长度为 1-100 个字符，可包含中英文、数字、下划线（_）和短横线（-）。
   - **创建类**：选择 **DB**。一旦完成创建，该项不能修改。
   - **是否分库分表**：是否支持 分库分表。本例中选择 **否**（即，不支持分库分表）。

     > 说明
     > 若选择 **是**，需填写 **分库分表规则**，该规则由 Groovy 脚本实现。

   - **数据源配置**：选择对应的数据源配置，也可通过输入数据源配置名称关键字来选择。如果尚未创建数据源配置，可单击右侧的 **新建数据源配置**。详情请参见 创建数据源配置。
   - **关联表**：选择对应的关联数据表。也可通过输入数据表名称的关键字来选择。
   - **SQL 表达式**：通过输入 SQL 表达式获取 DB 需要参与核对的数据。

11.9. 核对场景

本文为您介绍如何创建核对场景。完成创建后，您可以在 **核对场景** 页面，查看当前租户下的所有核对场景。场景列表支持按照核对场景名称的筛选方式（支持模糊匹配）。

创建核对场景

1. 登录资金安全监控控制台，在左侧导航栏单击 **核对场景**。
2. 在 **核对场景** 页面上，单击 **新建核对场景**，进入 **新建核对场景** 页面。



2.在 **新建核对场景** 面板，填写如下信息，然后单击 **提交**。



- **名称**：核对场景的名称，由 1-100 个英文或中文字符组成，可包含数字、下划线（_）或短横线（-）。租户内唯一。
- **上一级核对场景**：选择上一级核对场景，便于对核对场景进行分类。

  > 说明
  > 如果不选择，则默认当前新增场景为一级核对场景。

- **描述**：核对场景的备注信息，选填，长度不能超过 500 个字符。

修改核对场景

1. 单击核对场景列表操作列的 **修改**，进入 **修改核对场景** 页面。



2. 根据需要修改核对场景信息。仅支持修改 **名称** 和 **描述**。



3. 完成后单击 **提交** 保存修改。修改后的核对场景信息将在绑定的核对点中同步生效。

删除核对场景

单击核对场景列表操作列的 **删除** 即可删除核对场景。若该核对场景已绑定核对点，需要先解除绑定，否则无法删除。

> 说明
> 仅核对点负责人才可以进行删除操作。

11.10. 订阅

为了方便您及时查看核对点或任务的变化及异常情况，资金安全监控平台提供了订阅功能。您可以设置订阅方式和订阅内容，例如，如果设置了通过邮件通知异常，当系统发现异常时，将向您发送告警邮件。收到告警后，单击 **详情** 即可前往资金安全监控平台处理异常。

可以订阅如下三种类型的通知：

- **核对点通知**：核对点状态（上下线或修改）变更时触发通知。
- **异常通知**：核对任务执行出现异常时触发通知。
- **任务通知**：核对任务执行结果为失败时触发通知。

订阅成功后，您可以在 **核对点列表** 页面选择 **全部核对点** > **我订阅的核对点** 过滤出您已订阅的核对点。下面将介绍批量订阅和单个订阅的具体操作步骤。

批量订阅

1. 在核对点列表页面，勾选多个核对点，单击列表上方的 **批量订阅**，从下拉列表中选择 **批量订阅**。
2. 进入 **订阅** 页面，完成以下配置后，单击 **提交**。



i. 选择通知方式（可多选）。

- **邮件通知**：可指定邮件通知的对象。
  - 首次使用订阅功能时，您需要输入 Email 地址，订阅成功后，Email 地址将与该系统登录的账号建立绑定关系。
  - 再次使用订阅功能时，系统默认展示已绑定的 Email 地址。若修改，所有已订阅通知的 Email 地址将同步修改。
- **钉钉通知**：配置钉钉机器人信息。支持配置多个钉钉机器人。
  - **机器人名称**：自定义的名称，便于用户识别。
  - **webhook 地址**：系统需要此信息才能发送钉钉通知。

    > 说明
    > 有关如何获取自定义机器人 webhook，请参见 获取自定义机器人 webhook。

  - **安全密钥**：机器人安全设置页面，加签栏下面显示的以 `SEC` 开头的字符串。

ii. 选择通知内容（可多选）。

- 核对点修改和上下线。
- 核对任务执行失败。
- 执行发生异常告警。

单个订阅

1. 在 **核对点列表** 页面，单击核对点名称，进入 **核对点详情** 页面。
2. 在 **信息和配置** 页面，单击 **订阅**，完成相关配置后，单击 **提交**。具体配置说明请参见 批量订阅。

11.11. 常见问题

为什么核对规则一直处于运行中的状态？

核对规则运行是将编写好的 SQL 规则提交至 MaxCompute 去执行，所以一直处于运行中状态可能有两种情况：

- MaxCompute 的资源调度比较紧缺，从而使得您提交的 SQL 规则需要排队执行。
- SQL 规则比较复杂导致执行缓慢，您耐心等待即可。

权限相关问题

角色	授权 MaxCompute 产品权限	删除数据节点权限	编辑删除和上线下线核对点权限	关闭异常权限
阿里云主账号用户	√	√	仅核对点负责人才能操作。如需操作，需联系负责人添加为核对点负责人。	仅核对点负责人才能操作。如需操作，需联系负责人添加为核对点负责人。
阿里云子账号用户	×	×	仅核对点负责人才能操作。如需操作，需联系负责人添加为核对点负责人。	仅核对点负责人才能操作。如需操作，需联系负责人添加为核对点负责人。

相关工具使用问题

- DataWorks 平台使用相关问题
- SQL 代码编码原则与规范

12.高可用管理

12.1. 产品简介

12.1.1. 什么是高可用管理平台

高可用管理平台（High Availability Service，HAS）是基于蚂蚁多年技术风险防控理论和实践而衍生出的产品，是蚂蚁分布式架构 SOFAStack 的重要组成部分。
高可用管理平台覆盖了应用运行风险事件事前、事中、事后的全流程管理。

- **事前**：通过应用巡检、故障诊断以及和监控平台的联动，实现应用运行风险的主动发现。
- **事中**：通过故障诊断、应急预案、容灾切换实现风险事件快速定位和恢复。
- **事后**：通过风险定级、回溯，实现风险事件的闭环管理。有效提升 IT 技术风险防御水平，保证业务健康、持续、稳定运行。

12.1.2. 产品优势

蚂蚁技术风险管理体系方法论+平台工具完整落地
通过高可用管理平台工具，可以输出蚂蚁多年积累点的技术风险防控内容，并在此基础上助力用户建设符合用户实际情况的技术风险防控体系，提升用户整体技术风险防控水平。
技术风险防控效率有效提升
通过高可用管理平台技术风险防控能力，自动化、标准化、例行化日常运维，降低操作复杂度，运维结果清晰可见，实现风险事件的闭环管理。

- 通过日常巡检主动感知业务运行风险，在对业务造成影响前及时处理。
- 自动化故障诊断和标准化应急预案快速定位并恢复故障，减少因故障引起的业务中断时间。
- 故障演练主动检验应用高可用能力。
- 支持蚂蚁产品双中心容灾切换，满足监管合规需求。

技术风险防控内容库快速更新
阿里云、蚂蚁技术风险团队基于域内、域外技术风险防控经验，共建日常巡检、故障诊断、应急预案内容库，用户可享受最新的技术风险防控内容。

12.1.3. 产品架构

高可用管理平台 HAS 在蚂蚁 SOFA 中间件以及 Café 应用发布平台的基础上，为用户应用及蚂蚁应用提供日常巡检、风险管理、应急预案、故障演练等技术风险管理能力，满足用户应用巡检、故障处置、应急管理、故障演练等多种应用运维场景。

系统架构

高可用管理平台 HAS 的系统架构包括：

- 风险管理模块是 HAS 的核心组件，实现风险事件的汇集以及风险事件处理的流转。
- 日常巡检、业务监控、故障诊断发现风险事件，并将其上报至风险管理中心，由风险管理中心进行统一调度，如触发故障诊断、应急预案等。
- 运行总览集中展示整个平台的运行状态、风险情况。

部署架构

HAS 部署时，需要依赖以下底层服务：

- 数据库 RDS 或 OceanBase
- 负载均衡 SLB 或 ALB

12.1.4. 功能特性

高可用管理平台 HAS 定位于 IT 技术风险防控。产品主要功能包括风险管理、日常巡检、故障诊断、应急预案、故障演练等。

风险管理

风险管理 是高可用管理平台核心，是风险事件汇集、处置的中枢平台，具体如下。

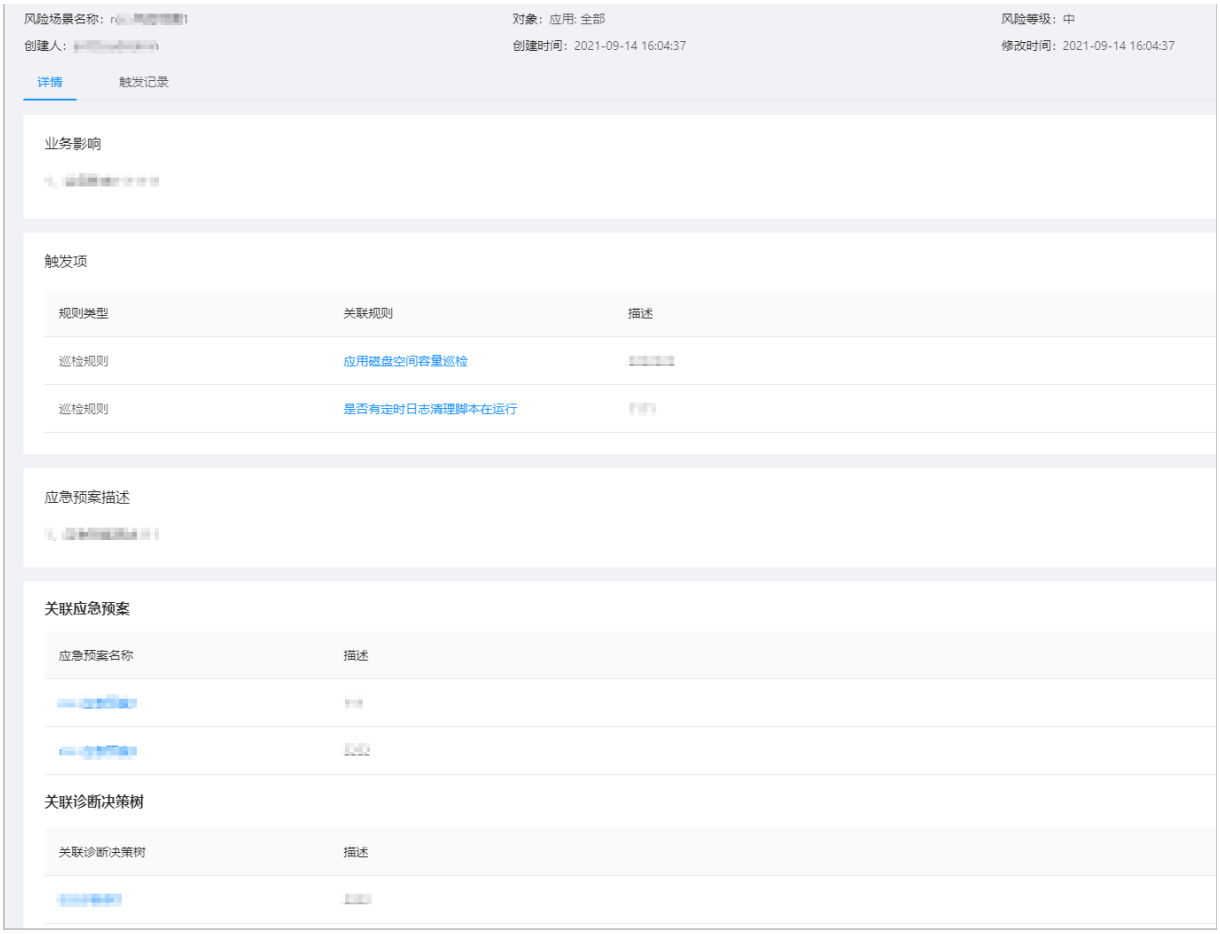
风险事件

- **风险事件汇集**：负责将监控、巡检、诊断产生的风险或告警信息进行汇总。
- **风险事件处置**：支持直接在风险事件列表中，对风险事件进行处理。在处理风险事件时，风险管理能够推荐可执行的应急预案，支持直接触发执行。



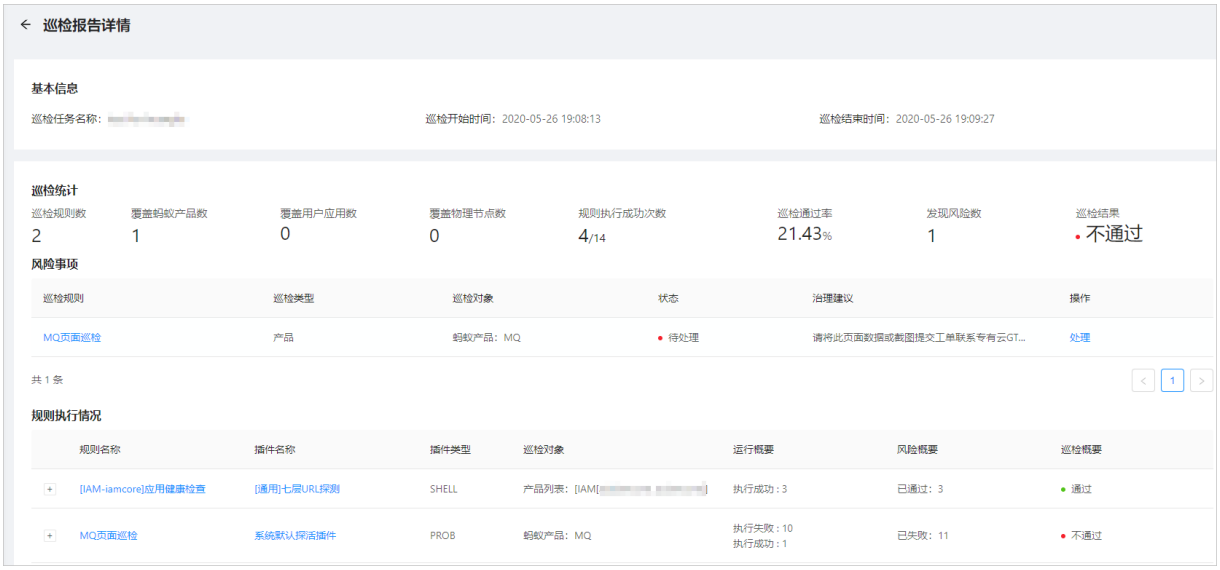
风险场景

风险场景 是针对特定风险事件进行集中化处理的模块，风险场景中包含了处理风险事件所需要诊断决策树、应急预案、业务影响等信息。目前应急场景升级后，需要将风险场景和应急响应联动，所以需要添加更多属性。



日常巡检

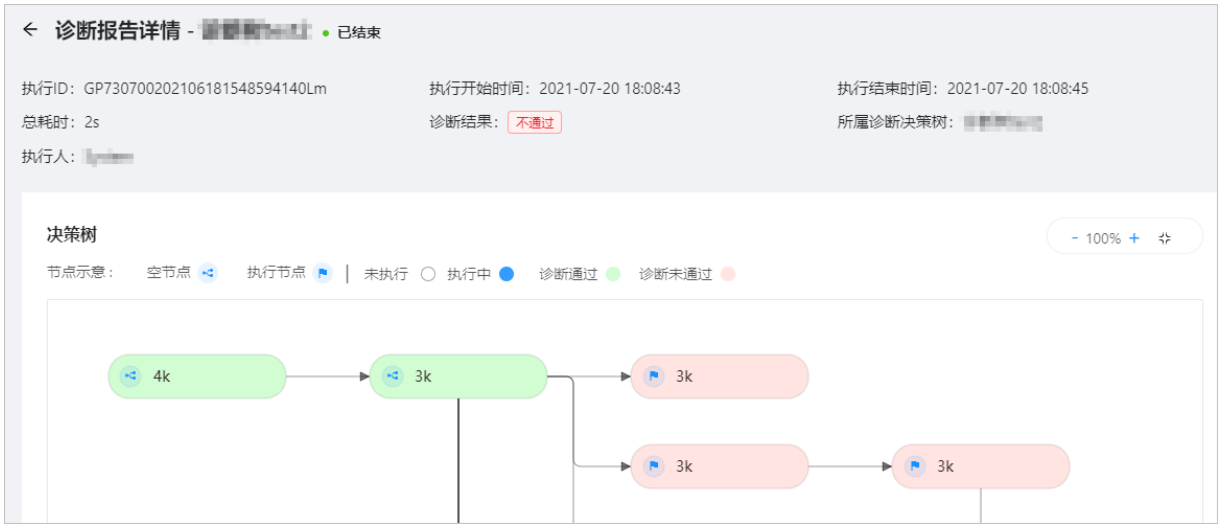
日常巡检 是高可用管理平台最常用的功能。通过日常巡检功能，可以例行化、自动化地对系统稳定性、可用性进行巡查，并将巡检结果实时同步推送至指定的钉钉群中，便于运维人员第一时间了解应用风险；同时支持生成巡检报告，供运维人员统一归档。巡检插件支持多种类型，包括 python、shell、自动化测试镜像、页面探活等。用户可以根据应用情况自定义巡检插件。同时，高可用管理平台也提供了蚂蚁内部及各个用户长期使用过程中沉淀的巡检规则，开箱即用。



故障诊断

故障诊断 的核心能力是将在运维人员头脑中或在故障排查手册中的经验、排查过程通过工具平台进行沉淀和展示。

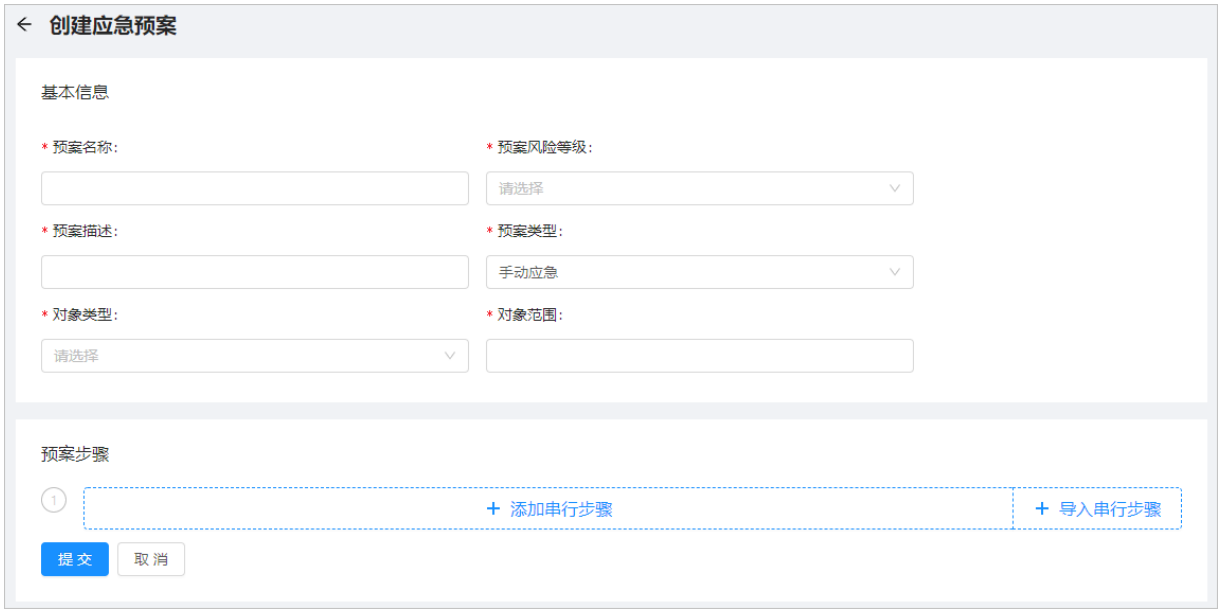
运维人员通过决策树方式，图形化地编排故障诊断过程、设计排查顺序。继而在风险事件发生时，将例行化、程式化、标准化的排查过程，通过故障决策树自动执行，并直接反馈诊断结果。通过故障诊断平台，能够极大地缩短故障排查时间。同时，屏蔽了不同运维人员在故障排查时的经验和技能差异，实现故障的快速定位。



应急预案

应急预案 提供了应用运维原子操作的编排能力，如应用重启、应用摘流、数据库切换、物理服务器重启等操作。

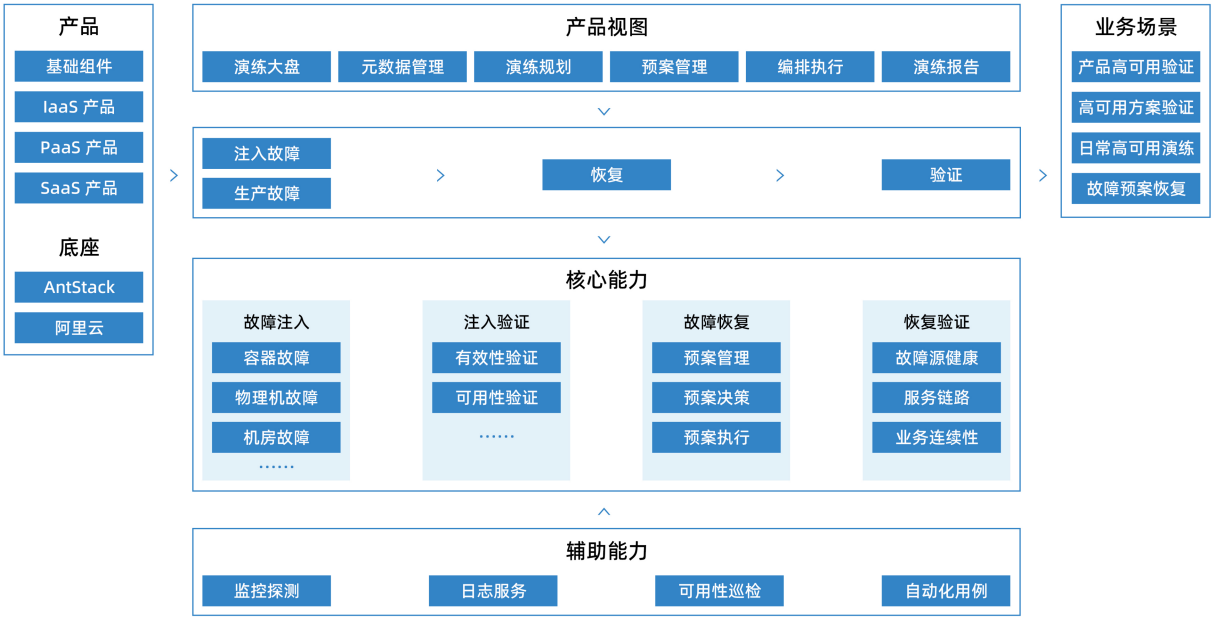
运维人员可以根据常见故障场景的处理过程，选择需要的原子能力进行编排组合，形成可执行的应急预案。当风险事件发生时，风险事件中心会推荐可执行的应急预案，供运维快速选择并自动化执行，从而通过标准化处理流程，实现故障快速恢复。



故障演练

故障演练 提供了故障注入能力，通过演练平台主动触发故障，以此观测应用软件的高可用性。

故障演练平台支持触发 CPU 利用率升高、内存利用率升高、内存占用、网络丢包、容器宕机、物理机宕机等常见故障，并针对故障制定出详细的演练和恢复计划，保证用户能够有计划地测量和观测应用高可用能力。



12.1.5. 应用场景

本文主要介绍高可用管理平台 HAS 的应用场景。

日常风险防控

在日常的运维场景中，通过多功能模块的联动使用，自动化执行日常运维脚本，实现定期可控的日常巡检运维；同时不断更新优化日常巡检、故障诊断、应急预案等内容的建设，不断丰富和完善应用技术风险防控体系，简化日常应用运维操作。

故障演练

为不断提升产品高可用能力，通过高可用管理平台的故障演练模块，设计并规划演练计划和恢复方案，继而在演练过程中不断发现、解决容灾预案存在的问题。以此，降低产品使用过程中故障发生概率，提高故障恢复效率，进而实现产品高可用性的有效提升。

机房级容灾

- **同城双活**
  同一个城市，建设两个机房环境，两地距离 50 km 以内，万兆光纤专线互连，业务应用层面可以两个机房同时提供业务服务，当一个机房故障，不影响另外一个机房业务使用。
- **异地主备**
  满足容灾需求，两地不同城市分别建设两个机房，一主一备，两地距离超过 1000 km，主机房承载业务流量，备机房无业务流量，只做备用机房使用。当主机房故障，可以切换流量到备机房快速恢复业务，等主机房故障恢复以后，再回切流量到主机房。
- **两地三中心**
  两地三中心，也称为同城双活加异地主备方案，即上文提及的同城两机房做双活部署，外加一个异地机房只做备份，不承载任何业务流量，基本只做冷备数据使用。最大程度地保障了数据的高可用备份。
- **LDC 单元化（异地多活）**
  LDC 单元化架构是可以实现异地多活和高并发场景的架构体系，LDC（Logic Data Center）逻辑数据中心是相对于传统的 IDC（Internet Data Center）提出的。逻辑数据中心所表达的中心思想是无论物理结构如何的分布，整个数据中心在逻辑上是协同和统一的。主要适用于大型互联网公司在线交易系统支持，比如淘宝、支付宝、携程等。

12.1.6. 基础术语

AKE

容器引擎（Ant Financial Kubernetes Engine，AKE）是将底层物理资源按照计算、网络、存储等进行切分和抽象的容器引擎。AKE 通过使用 Kubernetes 和 Docker 技术将整个物理资源进行池化，向上层服务提供按量使用的计算、网络和存储资源。

ALB

负载均衡（Ant Financial Load Balancer，ALB）是将访问流量根据转发规则分发到后端多台后端服务器的流量分发控制服务。通过流量分发扩展应用系统对外的服务能力，通过消除单点故障提升应用系统的可用性。

IAM

蚂蚁科技身份访问管理（Identity and Access Management，IAM）控制台是管理成员、分配权限、管理身份源、查看操作记录的平台。

OceanBase

OceanBase 是阿里巴巴与蚂蚁科技独立自主研发的一款分布式关系数据库产品，融合传统关系数据库和分布式系统的优势，具备高可用、高性能、高可扩展性，在功能上兼容 MySQL 等特点，在通用硬件上提供金融级高可用的数据库服务。

RPO

数据恢复点目标（Recovery Point Objective，RPO），以时间为单位，即在灾难发生时，系统和数据必须恢复的时间点要求。RPO 标志系统能够容忍的最大数据丢失量。系统容忍丢失的数据量越小，RPO 的值越小。

RTO

恢复时间目标（Recovery Time Objective，RTO），以时间为单位，即在灾难发生后，信息系统或业务功能从停止到必须恢复的时间要求。RTO 标志系统能够容忍的服务停止的最长时间。系统服务的紧迫性要求越高，RTO 的值越小。

容灾预案

指包含容灾步骤的可执行预案。

页面探活

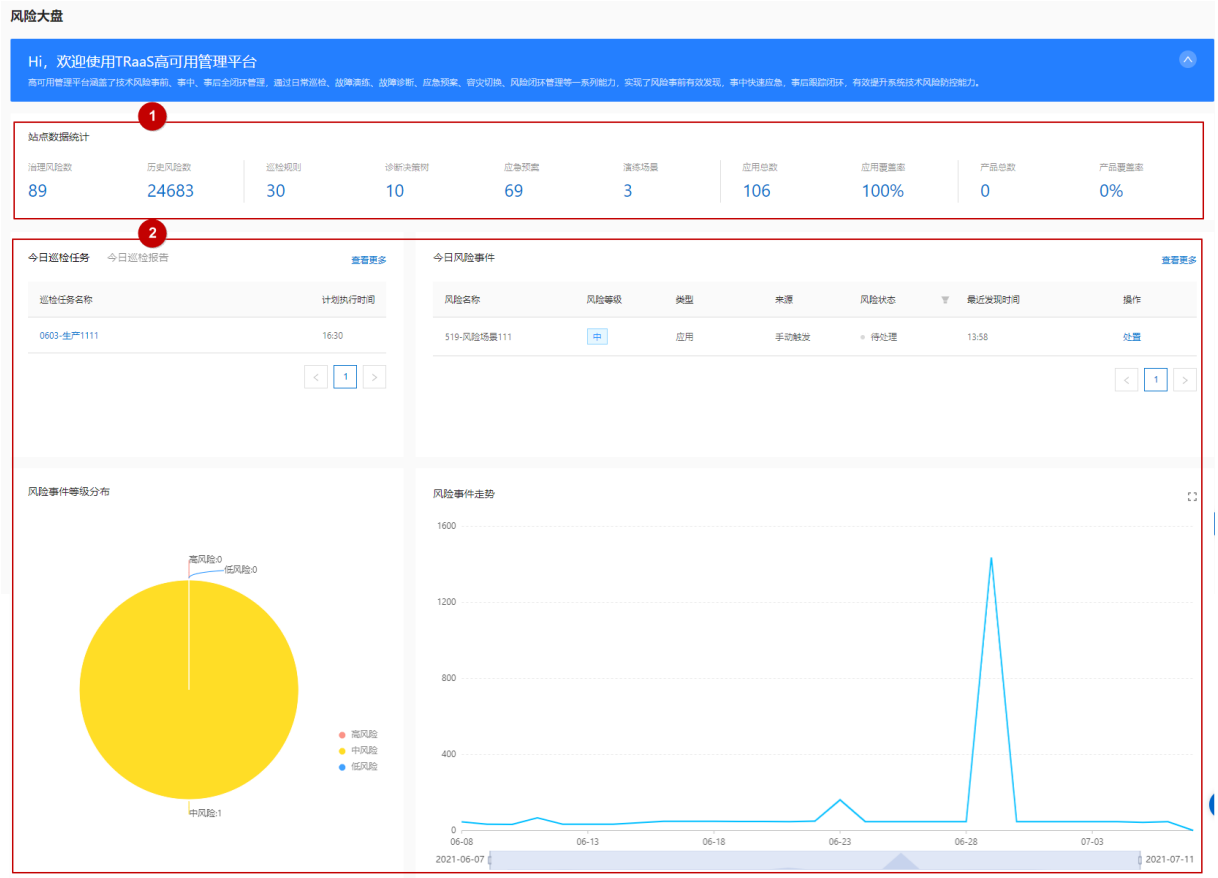
指通过浏览器打开巡检页面来判断页面存活情况。高可用容灾平台除了支持无需登录的静态页面探活外，还支持需要登录态的页面探活，也支持匹配页面的内容或元素来确定页面是否已渲染成功。在高可用容灾平台上，可以将页面探活配置成巡检任务以定时巡检页面。

云游

云游是蚂蚁科技的一站式专有云规划、交付、运维平台，管理着专有云从诞生到落地的整个生命周期。

12.2. 总览

您可以在高可用管理平台产品控制台 **总览** 页面查看站点数据统信息和风险统计信息。



- 站点数据统计（图示中 ①）：展示了治理风险数、历史风险数、巡检规则、诊断决策树数量、应急预案、演练场景、应用总数、应用覆盖率、产品总数以及产品覆盖率。单击统计数据可直接跳转至相应的业务详情列表。
- 风险统计（图示中 ②）：
  - **今日巡检任务**：展示了今日巡检任务。单击 **查看更多**，可直接跳转至巡检任务列表。
  - **今日巡检报告**：展示了今日巡检任务的结果及报告。单击目标巡检报告名称，可直接跳转至巡检报告详情；单击 **查看更多**，可直接跳转至巡检报告列表。
  - **今日风险事件**：展示了今日各个状态（全部、待处理、处理中、已关闭）的风险事件统计数量。单击 **查看全部**，可直接跳转至风险事件列表。
  - **风险事件等级分布**：展示了风险事件等级（高、中、低）分布。
  - **风险事件走势**：展示了每日风险事件数量的走势情况。

12.3. 风险管理

12.3.1. 风险事件

风险管理是高可用管理平台的核心能力，负责汇总所有监控、巡检产生的风险或告警信息，是风险事件汇集、处置、闭环管理的中枢平台。
在 **风险事件** 页面中，展示了当前环境的所有风险事件及其概览信息，包括风险名称、风险状态、对象类型、摘要等。

在风险事件列表中，如果风险事件没有及时处理，会产生大量的重复低风险事件。目前，**风险管理** 模块提供将风险聚合的能力，能够将相同风险事件聚合一起。您可以选择单个处理或批量一次处理。

新建风险事件

1. 登录高可用管理平台控制台。
2. 在左侧导航栏上，单击 **风险管理** > **风险事件**。
3. 单击 **新建风险事件**，在 **新建风险事件** 面板，配置以下信息：

参数	是否必填	说明
风险名称	必填	自定义的风险名称。
风险等级	必填	用于风险的管理，设定不同风险等级，方便进行风险监控。目前提供 **高** 、**中** 、**低** 三种等级 。
描述	必填	自定义的风险描述。
对象类型	必填	可选择 **用户应用** 或 **其他**。
对象名称	必填	选择/输入对象名称。

4. 单击 **确认**。

处置风险事件

未关联风险场景

风险事件未关联风险场景时，应急分析、预案推荐无数据。

操作步骤

1. 在左侧导航栏上，单击 **风险管理** > **风险事件**。
2. 选择目标 **待处理** 状态的风险，单击 **操作** 列下的 **处置**，进入风险详情页面。
3. 处置风险事件，具体操作步骤如下：
   - 在 **应急过程** 页签下单击 **添加应急步骤**，目标风险状态变为 **处理中**。
   - 在 **应急响应** 页签下单击 **立即响应**，目标风险状态变为 **处理中**。
4. 处置完成后，在风险事件详情页面单击 **完成应急**，即可完成风险处理闭环，风险事件状态变为 **已关闭**。

关联风险场景

风险事件可以手动关联风险场景，目前不会触发诊断执行，预案可以手动执行。风险管理模块会推荐可执行的应急预案。

操作步骤

1. 在左侧导航栏上，单击 **风险管理** > **风险事件**。
2. 选择目标 **待处理** 状态的风险，单击 **操作** 列下的 ··· > **关联风险场景**。
3. 选择风险场景。如需新增风险场景，请参见 创建风险场景。
4. 选择目标 **待处理** 状态的风险，单击 **操作** 列下的 **处置**，进入风险详情页面。
5. 处置风险事件，具体操作步骤如下：
   - 在 **应急响应** 页签下单击 **立即响应**，目标风险状态变为 **处理中**。
   - 在 **应急过程** 页签下单击 **添加应急步骤**，目标风险状态变为 **处理中**。
   - 在 **预案推荐** 页签下单击执行预案，目标风险状态变为 **处理中**。
6. 处置完成后，在风险事件详情页面单击 **完成应急**，即可完成风险处理闭环，风险事件状态变为 **已关闭**。

查看风险事件详情

风险事件详情页主要展示风险事件的基本信息、风险信息、执行对象、执行参数、执行详情以及执行日志。

操作步骤

1. 在左侧导航栏上，单击 **风险管理 > 风险事件**。
2. 选择目标风险事件，单击 **操作** 列下的 **处置/查看**，查看聚合风险详情：
   - **风险场景信息**：包括风险名称、对象类型、事件来源、风险等级、发现次数、关联的风险场景、风险首次发现时间、最近发现时间、摘要、应急响应时间、应急完成时间以及风险事件详情。
   - **应急分析**：风险事件关联了诊断决策树时，该模块将展示自动触发的诊断决策树中诊断异常的信息。
   - **应急响应**：包括未响应和已响应的成员。
   - **应急过程**：展示了当前风险应急处理的流程与进展（发现风险 > 应急过程 > 应急完成）以及应急处理记录。
   - **预案推荐**：风险事件关联了应急预案时，该模块将展示根据诊断情况推荐的可执行应急预案，并支持快速执行。

> 说明
> 您也可以展开风险事件，单击 **操作** 列下的 **详情** 查看单次风险事件详情。

12.3.2. 风险场景

风险场景 模块直观展示了风险涉及的产品及风险等级等基本信息，将预案场景的定义、应用范围、前置检查、预案执行等信息进行了关联，并提供场景的创建、编辑、触发等多种能力。

创建风险场景

1. 登录高可用管理平台控制台。
2. 在左侧导航栏上，单击 **风险管理 > 风险场景**，进入 **风险场景** 页面。
3. 单击 **新建风险场景**，在 **新建风险场景** 面板，配置以下信息：

参数	是否必填	说明
风险场景名称	必填	自定义的风险场景名称。
风险等级	必填	用于风险的管理，设定不同风险等级，方便进行风险监控。目前提供 **高**、**中**、**低** 三种等级 。
对象类型	必填	可选择 **用户应用** 或 **其他**。 ○ 如果选择 **用户应用**，则需要选择对应的应用范围。 ○ 如果选择 **其他**，则需要输入对应的对象名称。
业务影响	必填	自定义对业务的影响。例如，**影响核心交易系统**。
应急预案描述	必填	自定义的应急预案描述。
关联应急预案	选填	手动选择应急预案。如需新增应急预案，请参见 创建应急预案。
关联诊断决策树	选填	手动选择诊断决策树。如需新增诊断决策树，请参见 新建诊断决策树。

参数	是否必填	说明
添加触发项	选填	可选 **巡检规则** 或 **监控规则**。 ○ 如果选择 **巡检规则**，则需要选择对应的关联规则。如需新增巡检规则，请参见 新建规则。 ○ 如果选择 **监控规则**，则需要输入对应的关联规则。

4. 配置完成后，单击 **提交**。

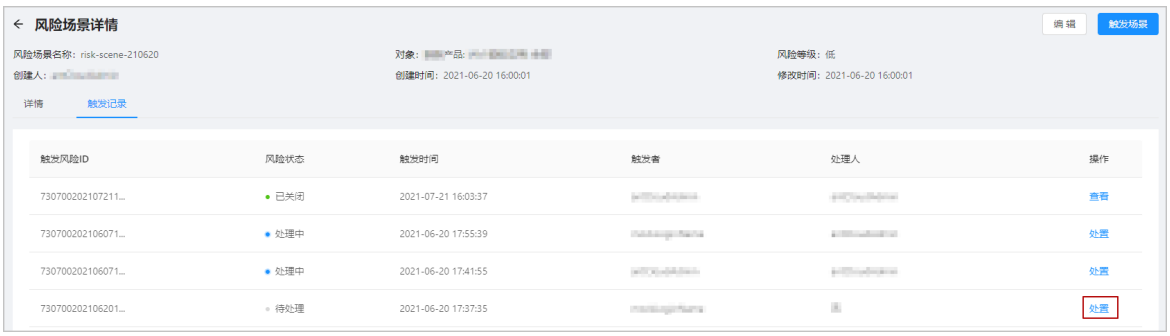
执行风险场景

1. 在左侧导航栏上，单击 **风险管理** > **风险场景**，进入 **风险场景** 页面。
2. 选择目标风险场景，单击 **操作** 列下的 **触发场景** > **确认**。



> **说明**
> 您也可以在 **风险场景** 页面，单击目标风险场景名称进入风险场景详情页，单击右上角 **触发场景** > **确认** 触发风险场景。

3. 在 **风险场景详情** 页面，选择目标风险，单击 **操作** 列下的 **处置**。



> **说明**
> 有关如何处置风险的详细步骤，请参见 处置风险事件。

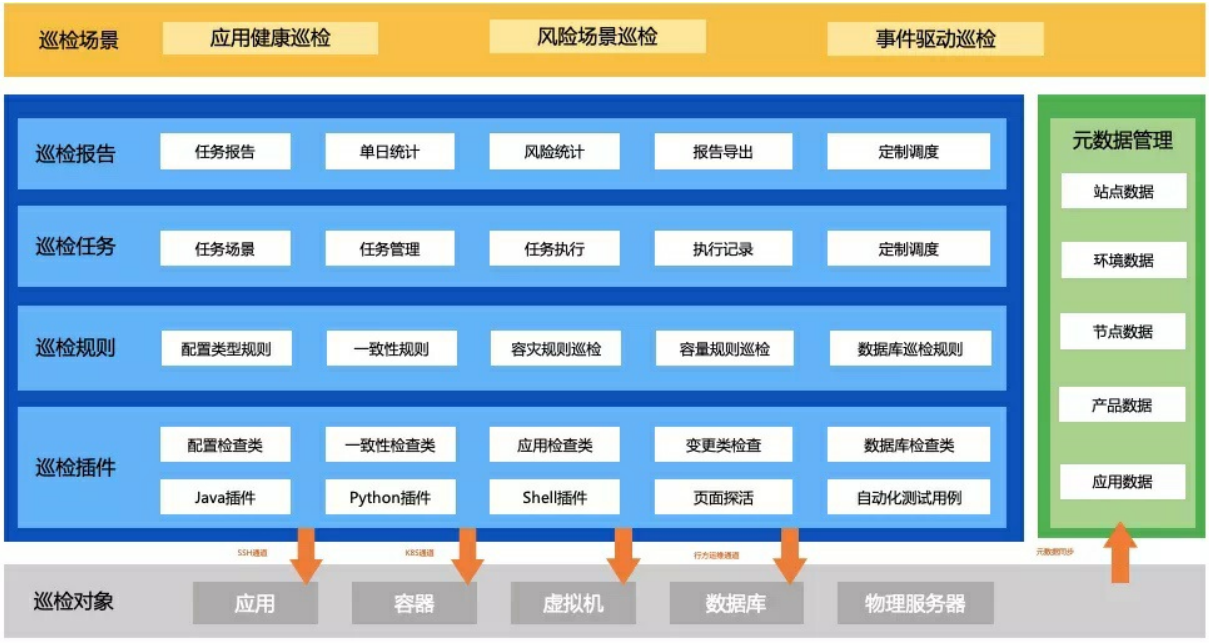
相关操作

- 如需修改风险场景信息，请单击 **操作** 列下的 **编辑**，在修改风险场景页修改相关信息，然后单击 **确认**。
- 如需删除风险场景信息，请单击 **操作** 列下的 **删除**，然后在弹出的提示对话框中单击 **确认**。

12.4. 日常巡检

12.4.1. 日常巡检概览

日常巡检是被高频使用的功能，通过巡检规则/报告 + 巡检插件的体系，从而实现“去人为化”操作。相对人工巡检需消耗1到2个小时，脚本化、自动化的巡检方式可将时间压缩至 10 到 20 分钟。同时，在此过程中生成的可视化巡检报告可用于进一步分析和存档。而且，支持定期可控巡检，便于日常运维和掌握站点的健康信息，大大减轻运维人员的排查问题负担和时间。



12.4.2. 巡检插件

巡检插件是巡检平台最底层的模型，日常巡检类插件为事先配置，也可以通过控制台添加。

插件概述

在左侧导航栏上，单击 **日常巡检 > 巡检插件** 进入插件列表页查看当前环境的所有插件。您可以根据插件类型、插件状态来筛选插件，或直接输入插件名称进行搜索。

单击目标插件名称，即可查看该插件的详细信息，包括插件基本信息（如插件 ID、名称、类型、状态、标签等）、参数设置、以及脚本内容。脚本内容因插件类型而异。

目前，日常巡检支持以下几种类型的插件：

- **Shell/Python 脚本**：该类插件运行在插件所被挂载的规则所指定的目标机器或容器中。当需要在目标机器或容器发起的较为基础的巡检时，可以使用此类插件。例如，磁盘容量检查，进程是否存在等。
- **自动化测试镜像**：支持自动激活云游测试镜像，通过事先在云游上配置好测试镜像完成巡检目标。
- **页面探活**：通过浏览器打开巡检页面来判断页面存活情况。高可用管理平台除了支持无需登录的静态页面外，还支持需要登录态的页面探活，也支持匹配页面的内容或元素来确定页面已渲染成功。在高可用管理平台上，可以将页面探活配置成巡检任务以定时巡检页面。

创建插件

您可以在 **日常巡检 > 巡检插件** 中根据应用情况自定义巡检插件。

操作步骤

1. 登录高可用管理平台控制台。
2. 在左侧导航栏上，单击 **日常巡检 > 巡检插件**。
3. 单击 **新建插件**，在 **新建插件** 面板，配置以下信息。单击 **确定**。

配置项	说明
插件名称	填写插件名称。
描述	输入关于插件的描述。
插件类型	可选择 PYTHON 脚本 或 Shell 脚本。
标签	选择一个或多个标签，便于查询和统计。

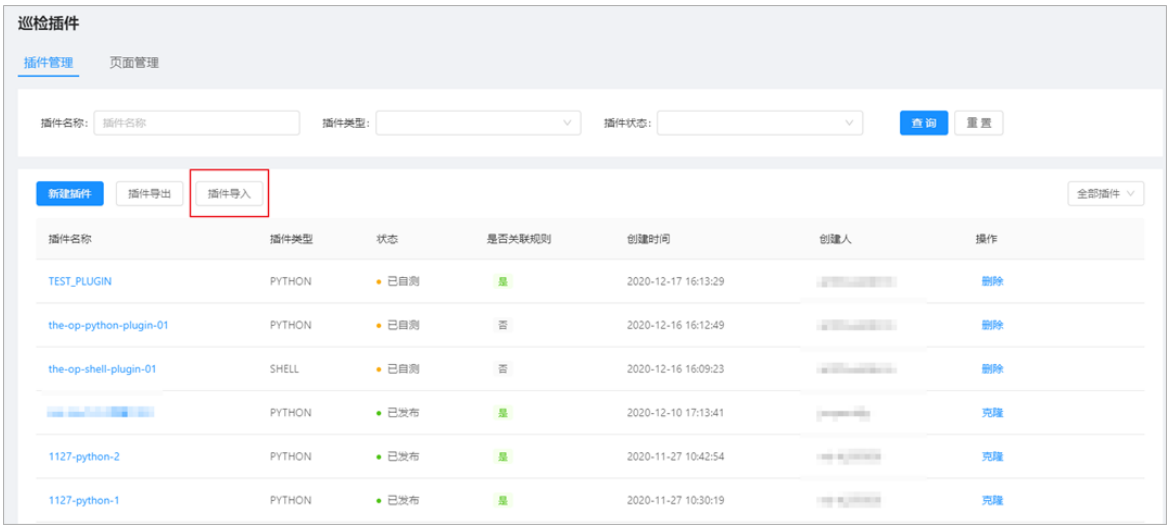
4. 单击 **确定**。

导入插件

如果您想快速创建或应用巡检插件，可以使用插件导入功能。例如，希望将蚂蚁技术风险团队基于域外技术风险防控经验沉淀的巡检插件运用到某类应用场景。

操作步骤

1. 在左侧导航栏上，单击 **日常巡检 > 巡检插件**。
2. 在 **巡检插件** 页面，单击 **插件导入**。



3. 从本地选择需要导入的巡检插件，完成插件的导入。



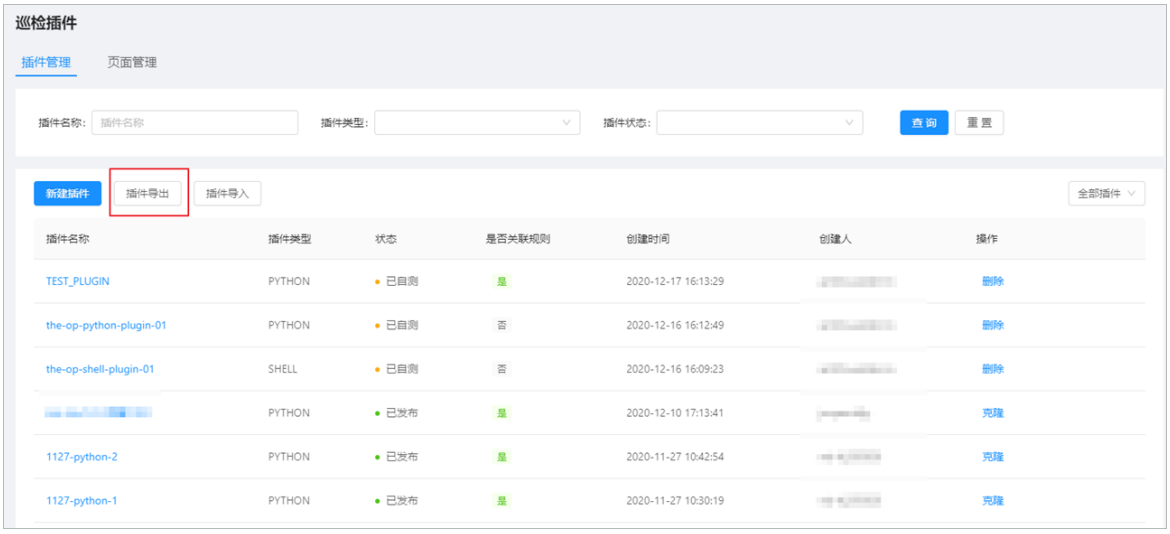
导出插件

如果您想在本地备份巡检插件，可以使用插件导出功能。

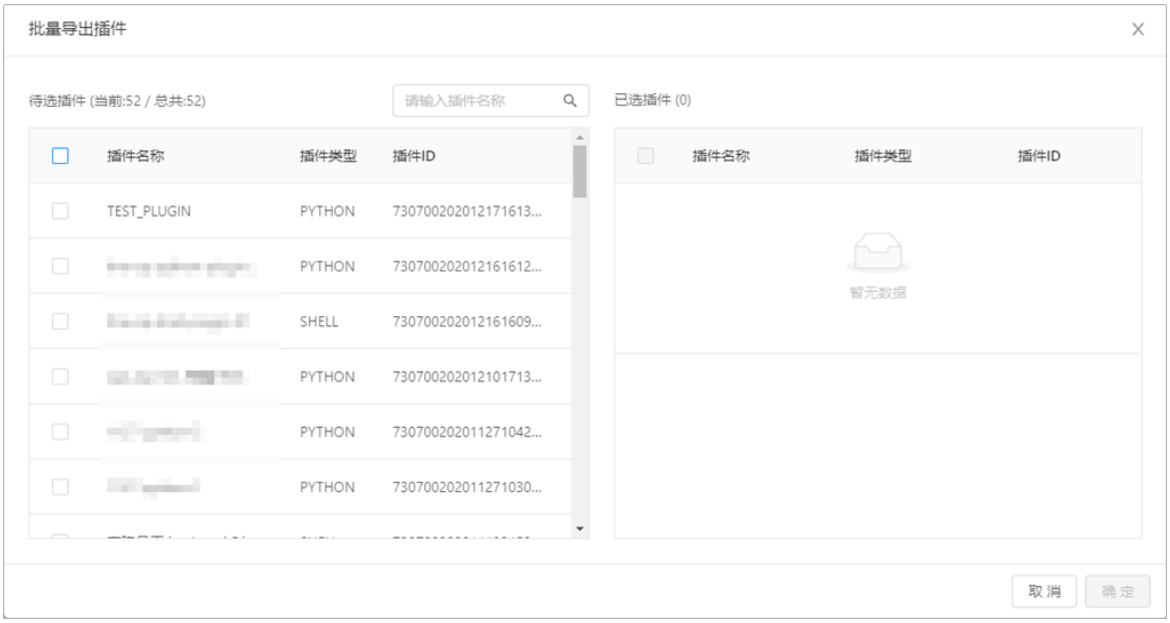
操作步骤

1. 在左侧导航栏上，单击 **日常巡检 > 巡检插件**。

2. 单击 **插件导出**。



3. 在弹出的 **批量导出插件** 页面上，勾选所需的插件，然后单击 **确定** 即可。



参数限定

高可用管理平台的 **巡检插件** 模块提供参数注入能力，即可以在插件实例化时，通过自定义插件参数的方式进行输入。如果输入的参数不正确，极大可能会对应用造成破坏，因此需要对传入的参数进行严格管控和限定。

前提条件

已创建插件。

操作步骤

1. 在左侧导航栏上，单击 **日常巡检 > 巡检插件**。

2. 在 **巡检插件** 页面，单击已创建的插件名称（本例中为 TEST_PLUGIN）。



3. 新增参数设置。

   i. 在插件详情页面的参数设置区域，单击 **新增一条**。



   ii. 在展开的对话框中，完成以下参数（见下表）配置后，单击 **保存**。

参数	说明
参数名	必填项。 参数的名称。具有唯一性。
参数限制	必填项。 正则表达式。默认值为 `^[0-9a-zA-Z.]+$` 。
参数描述	必填项。 关于参数的说明。
参数示例	必填项。 参数值。
信息加密	选填项。 选择参数是否需要加密。

其他操作

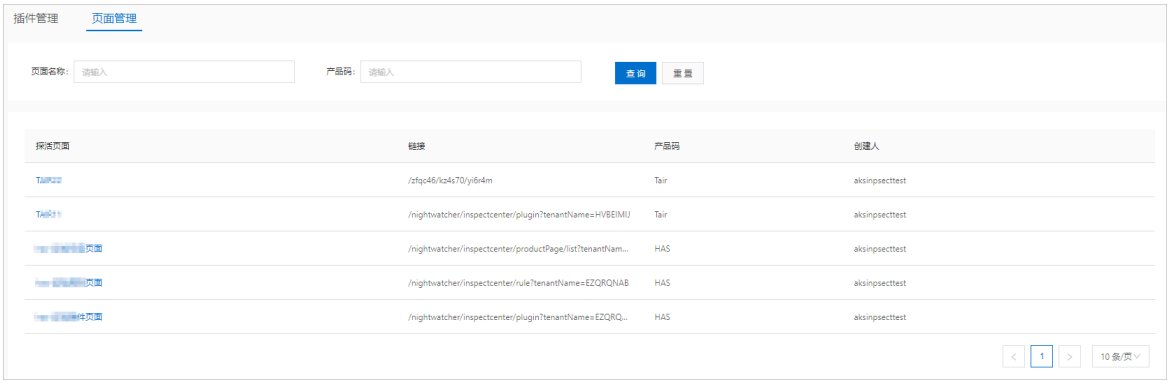
- 新增参数设置：如需再新增一个参数限定，单击 **新增一条** 即可继续添加。
- 删除参数设置：单击右侧的删除图标，在弹出的信息提示框中，单击 **是**。

查看探活页面

在高可用管理平台上，可以将页面探活配置成巡检任务以定时巡检页面。日常巡检内的探活页面均为事先配置，不支持通过控制台添加。

操作步骤

1. 在左侧导航栏上，单击 **日常巡检 > 巡检插件**。
2. 在 **巡检插件** 页面，单击 **页面管理** 页签。
3. 在探活页面列表页面，查看当前环境的所有探活页面。



4. 在探活页面列表中，单击目标页面名称即可打开探活页面详情页。
   在该页面上可查看探活页面配置详情，包括页面名称、URL、对应产品码、是否启用内容匹配、匹配内容（如启用内容匹配）、是否需要登录、cookie 信息、自定义探活时长（如自定义时长）等信息。

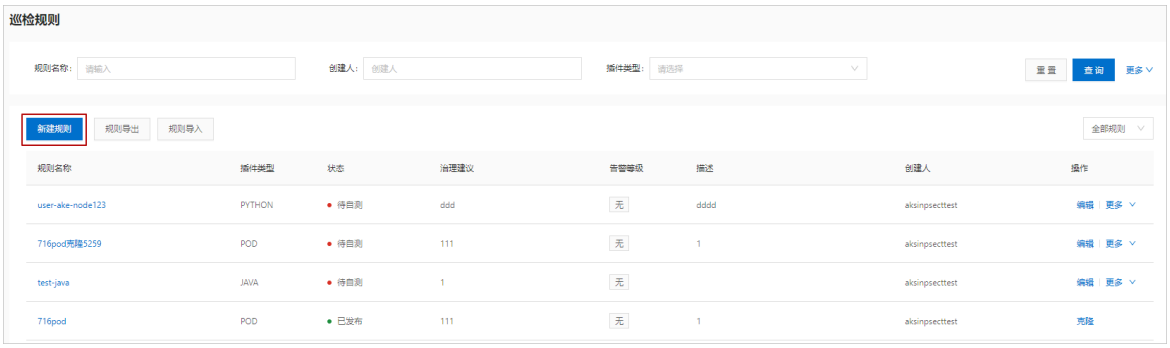
12.4.3. 巡检规则

巡检规则是巡检插件的实例，关联巡检脚本。规则中的实例化参数需要保证与环境无关，支持通过云游渲染参数。

在左侧导航栏上，单击 **日常巡检 > 巡检规则** 进入规则列表页查看当前环境的所有规则。您可以根据规则 ID、规则名称、插件类型、创建者来筛选插件。

新建规则

1. 登录高可用管理平台控制台。
2. 在左侧导航栏上，单击 **日常巡检 > 巡检规则**。
3. 单击 **新建规则**，然后在 **新建规则** 对话框配置规则信息：



参数说明如下：

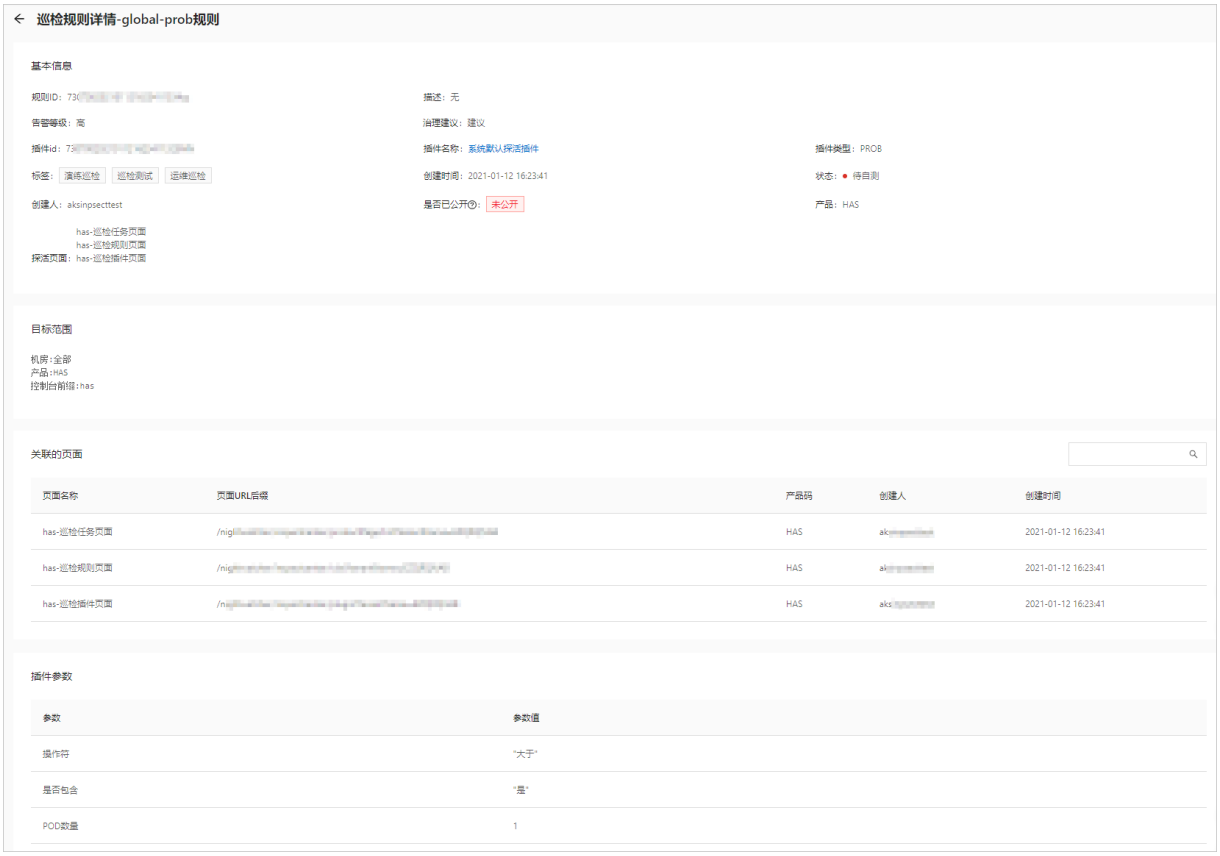
参数	是否必填	参数说明
规则名称	必填	填写插件名称。
描述	选填	输入关于规则的描述。
告警等级	必填	可选择 **高**、**中**、**低** 三个等级。
治理建议	必填	输入治理建议。

参数	是否必填	参数说明
插件类型	必填	可选择 **PYTHON 脚本**、**Shell 脚本**、**系统插件**。各类型插件说明详见插件概述。 ○ 选择 **PYTHON 脚本** 或 **Shell 脚本** 时需要添加规则目标，用于校验用户的应用操作权限。 ■ **业务应用**：仅支持业务租户的用户操作。 ■ **自定义服务器**：操作对象无限制。 ○ 选择 **系统插件** 时需要添加操作符、目标数量以及应用，用于查询应用 POD 数。
关联插件	必填	选择插件。如需新增插件，请参见 新建插件。
标签	选填	选择一个或多个标签，便于查询和统计。

4. 单击 **确认**。

查看规则详情

在规则列表中，单击规则名称进入规则详情页查看规则详情，包括规则基本信息、规则适用的目标范围、插件参数。



- **基本信息**：展示当前巡检规则的基本信息，包括规则 ID、执行状态、创建者、告警等级等。
- **目标范围**：规则的巡检范围，包括规则作用的机房、目标类型、产品等。
- **关联的页面**：仅页面探活类型的规则会展示关联的页面列表。提供规则关联的页面名称、URL 后缀、对应的产品码等信息。
- **插件参数**：展示插件的脚本入参。

其他操作

- 导入规则：单击 **规则导入**，从本地选择需要导入的巡检规则，完成规则的导入。
- 导出规则：单击 **规则导出**，在弹出的 **规则导出** 页面上，勾选目标规则，然后单击 **导出**。
- 克隆规则：单击目标规则操作列下的 **克隆**，在弹出的信息提示框中，单击 **是**。
- 删除规则：单击目标规则操作列下的 **更多** > **删除**，在弹出的信息提示框中，单击 **是**。

12.4.4. 巡检任务

巡检任务是巡检规则的聚合载体，并且提供修改规则的执行范围以及调度方式的功能。您可通过巡检任务模块对巡检任务进行统一的管理操作，包括创建、编辑、复制、删除、执行、调度任务等。

创建任务

1. 从左侧导航栏进入 **日常巡检** > **巡检任务** 页面，单击 **创建巡检任务**。
2. 在 **创建巡检任务** 页面，输入任务名称、任务描述，选择任务类型以及钉钉机器人，然后单击 **提交**，新建的任务将出现在任务列表中。
3. 为任务绑定巡检规则。如果不绑定，任务将无法执行。
   i. 单击目标任务名称，进入任务详情页。
   ii. 在任务详情页面的 **关联规则** 标签页，单击 **关联规则**。
   iii. 勾选要绑定的规则，单击 **确定** 即可。关联的规则将出现在规则列表中。
4. （可选）设置任务调度，一个任务可以设置多个调度。为巡检任务添加调度后，系统将自动进行定时任务调度。
   i. 单击 **任务调度** 标签页。
   ii. 单击 **创建调度**，在 **创建规则调度** 页面配置以下信息。
      - **调度方式**：
        - **按时间**：系统会在固定时间执行任务。例如，每天上午 8：00 点执行。
        - **按间隔**：系统会每隔一段时间执行任务。例如，每隔 8 个小时执行一次任务。



      - **开关**：设置是否开启调度。
   iii. 单击 **确定**。

查看任务详情

从左侧导航栏进入 **日常巡检 > 巡检任务** 页面，单击目标任务名称进入任务详情页。任务详情页主要展示巡检任务的基本信息、关联规则、调度信息以及执行记录。

- **基本信息**：包括任务名称、描述、关联的钉钉机器人、创建时间、创建者信息。
- **关联规则**：展示当前任务所关联的规则信息。
- **任务调度**：展示当前任务的调度信息，包括调度开始时间、调度方案、开关状态、创建者。
- **执行记录**：展示各个状态的任务执行记录，包括任务 ID、执行状态、状态及结果概要、任务创建时间。

编辑任务

编辑基本信息

1. 从左侧导航栏进入 **日常巡检 > 巡检任务** 页面。
2. 单击目标任务 **操作** 列下的 **编辑**，可对任务名称、描述以及关联的钉钉机器人进行修改。

修改关联规则

单击目标任务名称进入任务详情页，进入任务详情页面。在 **关联规则** 标签页上，进行以下规则操作：

- **添加关联规则**：单击 **关联规则**，勾选要绑定的规则，单击 **关联** 即可。关联的规则将出现在规则列表中。
- **解绑规则**：勾选要取消绑定的规则，单击 **解绑** 即可。
- **编辑绑定的规则**：单击目标规则 **操作** 列下的 **编辑参数**，按需对规则进行配置即可。



修改任务调度

1. 在 **任务调度** 标签页上，选择目标任务调度。
2. 单击 **编辑** 对当前调度进行修改。

执行任务

执行任务前，确保该任务关联的规则以及相关参数已设置好。

1. 从左侧导航栏进入 **日常巡检 > 巡检任务** 页面。
2. 单击目标任务 **操作** 列下的 **执行**。

克隆任务

日常巡检支持基于已有的任务创建新的巡检任务。

1. 从左侧导航栏进入 **日常巡检 > 巡检任务** 页面。
2. 选择要复制的任务，单击 **操作** 列下的 **··· > 克隆**。
   克隆出的新任务将复用原任务的规则和调度设置。

删除任务

1. 从左侧导航栏进入 **日常巡检 > 巡检任务** 页面。
2. 选择要删除的任务，单击 **操作** 列下的 ··· > **删除**。

批量添加任务调度

支持同时为多个巡检任务设置调度。

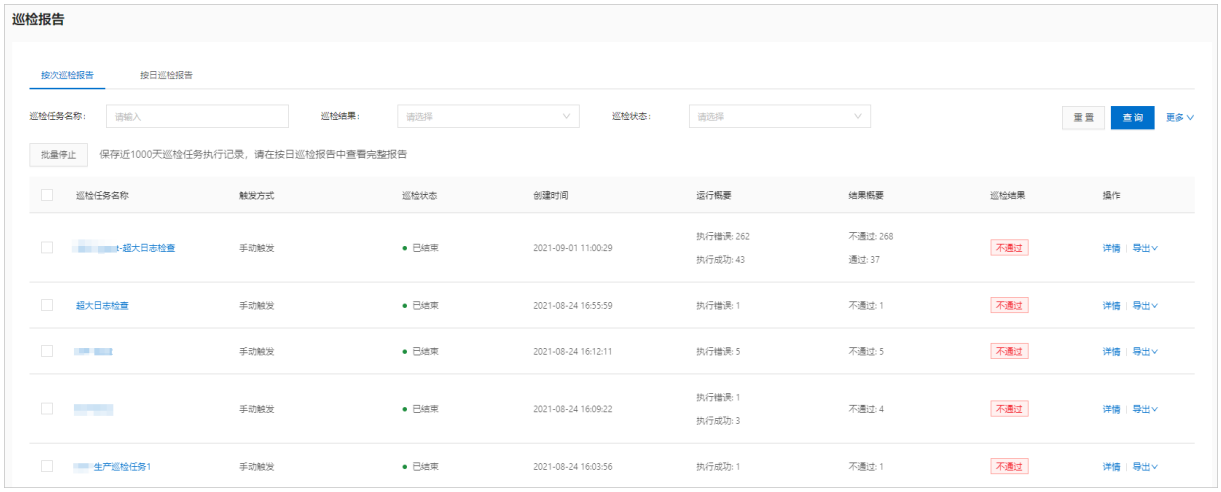
1. 从左侧导航栏进入 **日常巡检 > 巡检任务** 页面。
2. 勾选要添加调度的任务，单击 **批量添加调度。**
3. 设置调度方式、时间和开启状态后，单击 **确定** 即可。

12.4.5. 巡检报告

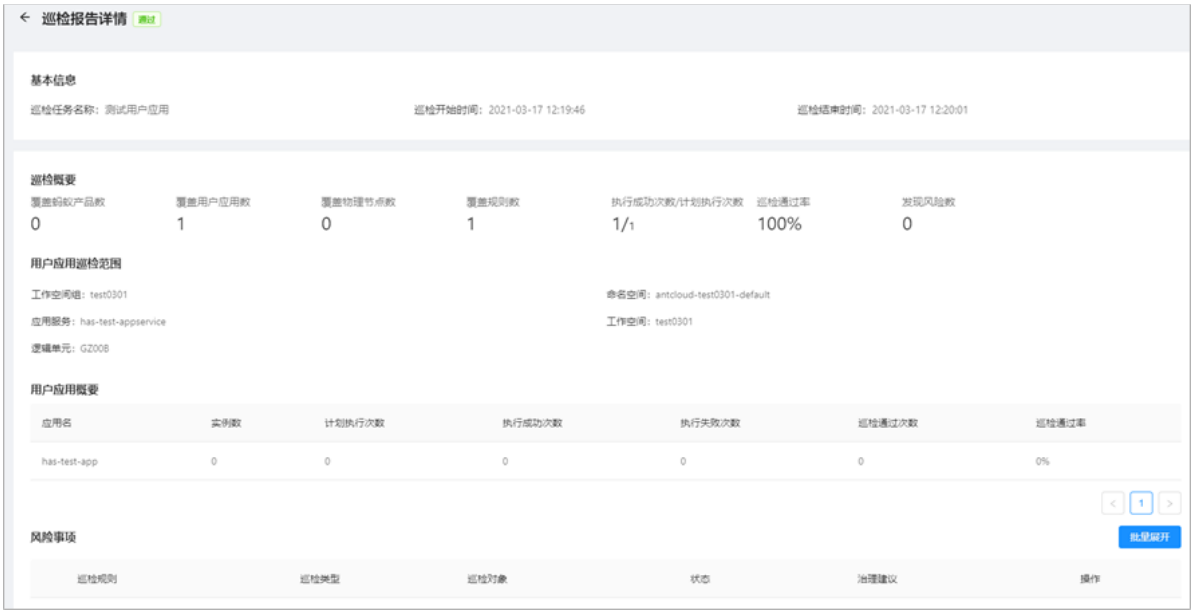
巡检报告展示日常巡检针对产品的整体巡检数据。在左侧导航栏上，单击 **日常巡检 > 巡检报告** 查看日常巡检按次与按日的巡检数据。

按次巡检报告

在 **按次巡检报告** 标签页下，展示了所有巡检任务执行记录。您可以根据巡检任务名称、巡检结果、巡检状态或触发方式来筛选巡检运行记录及其巡检数据。数据包括巡检任务名称、触发方式、巡检状态、创建时间、运行概要、执行结果概要以及巡检结果。



- 选择任意 **执行中** 的任务，单击 **批量停止** 按钮，可直接终止该巡检任务。
- 单击目标巡检任务名称即可跳转至巡检任务详情页。
- 单击 **操作** 列下的 **详情**，即可查看该巡检运行记录对应的详细巡检报告，具体如下图所示。

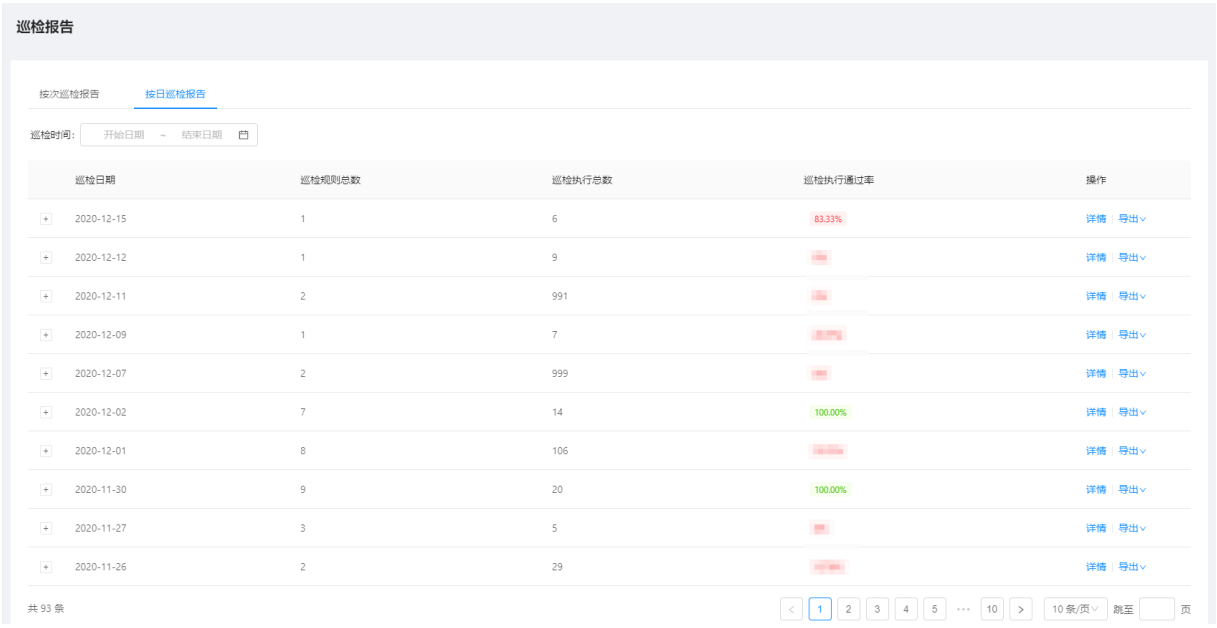


- 单击 **操作** 列下的 **导出**，即可导出该巡检报告（支持导出格式为 JSON）。

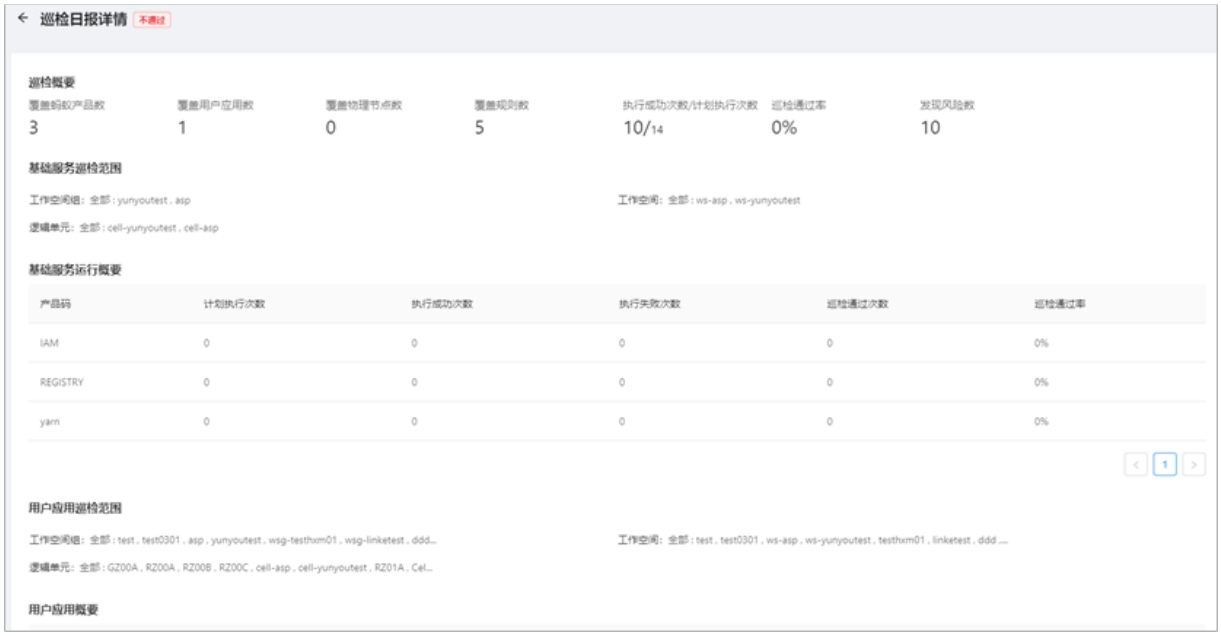
按日巡检报告

在 **按日巡检报告** 标签页下，展示了每天的巡检数据。您可以指定巡检时间范围，进行数据筛选。
展示的数据包括巡检日期、巡检规则总数、巡检执行总数、巡检执行通过率。

- **巡检规则总数**：当日执行的任务所涉及的规则总数。
- **巡检执行总数**：当日执行的任务总数。
- **巡检执行通过率**：巡检通过的任务数占执行任务总数的百分比。



单击 **操作** 列下的 **详情**，即可查看巡检报告详情，详情如下图所示。



单击 **操作** 列下的 **导出**，即可导出该巡检报告（支持导出格式为 JSON）。

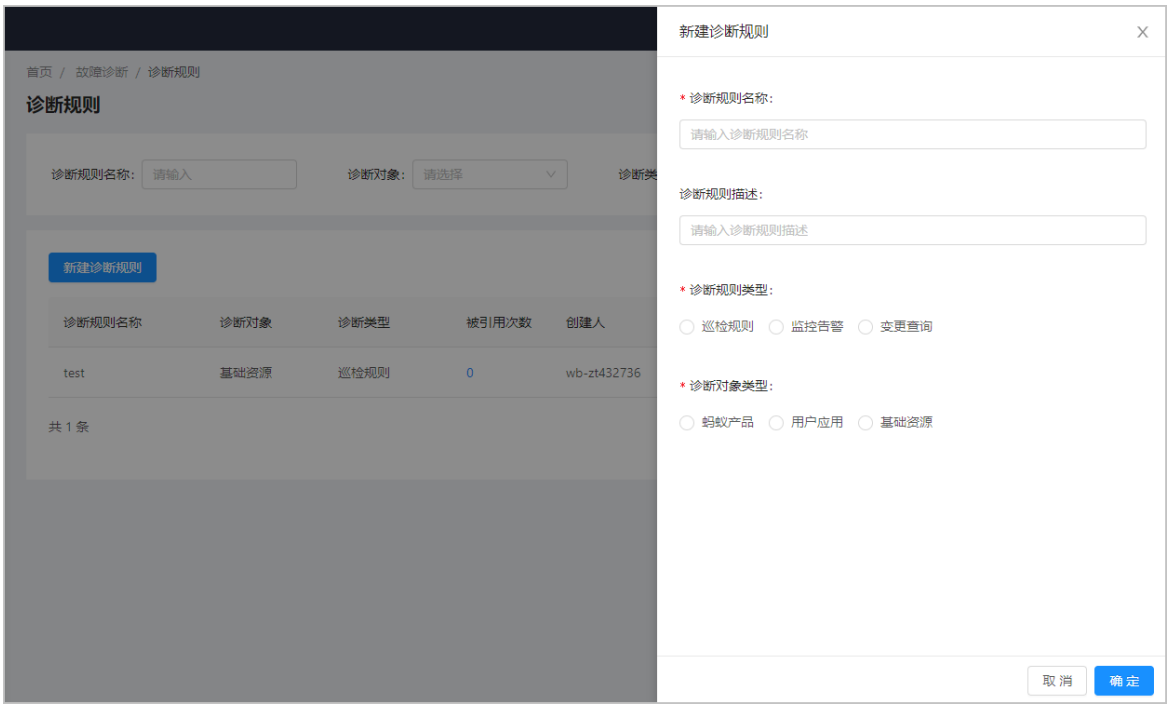
12.5. 故障诊断

12.5.1. 诊断规则

通过故障诊断平台，运维人员可以将诊断过程、排查顺序进行图形化编排与设计，即故障诊断决策树。而后，在实际故障发生时，即可执行自动化、标准化的故障排查，并直接输出诊断报告，反馈诊断结果。故障诊断功能有效提升了故障排查效率，实现故障的快速定位与处理。本节将向您介绍诊断规则的相关操作。
诊断规则是诊断节点的构成，诊断规则提供多种诊断能力，包括巡检、监控告警、变更查询。

新建诊断规则

1. 从左侧导航栏进入 **故障诊断** > **诊断规则** 页面，单击 **新建诊断规则**。
2. 在右侧新窗口中，输入诊断规则名称与描述，并根据选择的诊断规则类型，配置对象与应用范围等。



- **巡检规则**：选择诊断对象类型（蚂蚁产品/用户产品/基础资源）、产品范围（全部产品/部分产品）以及关联巡检规则。
- **监控告警**：选择诊断对象类型（蚂蚁产品/用户产品）、workspace、产品、应用范围以及查询时间范围。
- **变更查询**：选择诊断对象类型（用户产品）、workspace、应用范围以及查询时间范围。

3. 单击 **确定**。

编辑诊断规则

在 **故障诊断** > **诊断规则** 页面，单击目标规则 **操作** 列下的 **编辑**，在右侧编辑窗口中修改。

> 说明
> 除诊断规则类型不支持修改，其他信息均可修改。

删除诊断规则

在 **故障诊断** > **诊断规则** 页面，单击目标规则 **操作** 列下的 **删除** ，然后在弹出的提示对话框中单击 **确认**。

> 说明
> 已被节点引用的诊断规则暂不支持删除。

12.5.2. 诊断节点

诊断节点是诊断决策树的子节点，通过节点之间的关联，实现一棵完整的决策树。

新建节点

1. 从左侧导航栏进入 **故障诊断** >**诊断节点** 页面，单击 **新建节点**。
2. 在右侧新窗口中，配置新节点的信息，包括节点名称、描述，选择节点类型、超时时间及标签，然后单击 **确认**，新建的节点将出现在节点列表中。



3. 为该诊断节点绑定巡检规则，否则节点将无法执行。
   i. 单击节点右侧 **操作** 列下的 **详情** 按钮。
   ii. 在节点详情页面，单击 **绑定诊断规则**。
   iii. 在右侧窗口中，勾选要绑定的规则，单击 **确认**。

返回节点详情页面，关联的规则将出现在规则列表中。

克隆节点

在 **诊断节点** 页面，选择需要克隆的节点，单击其 **操作** 列下的 **克隆**，然后确定即可。

查看节点详情

在节点列表中，找到目标节点，单击其 **操作** 列下的 **详情** 按钮，即可进入节点详情页查看节点详情，包括节点基本信息（如插件名称、标签、描述等）、规则适用的目标范围、插件参数、执行记录。

1. 在 **故障诊断 >诊断节点** 页面，选择要查看的节点，单击 **操作** 列下的 **详情**，进入节点详情页。
2. 在节点详情页中，您可以查看到该诊断节点的基础信息与绑定规则。
   - **基础信息**：包括节点名称、类型、超时时间、创建人、创建时间以及节点描述等。
   - **绑定规则**：展示当前任务所绑定的诊断规则及其信息，包括规则名、诊断类型、诊断对象等。

编辑节点

编辑基本信息

在 **故障诊断 > 诊断节点** 页面，选择要修改的节点，单击 **操作** 列下的 **编辑** 按钮，可对节点名称、描述、超时时间以及标签进行修改。

修改绑定规则

单击目标节点右侧的 **详情** 按钮，进入节点详情页面。在规则列表中，可进行以下操作：

- **绑定规则**：单击 **绑定规则**，勾选要绑定的规则，单击 **确认** 即可。关联的规则将出现在规则列表中。
- **解绑规则**：勾选要取消绑定的规则，单击 **解绑** 即可。

- **编辑绑定的规则**：单击目标规则对应的 **编辑参数** 按钮，按需对规则配置进行即可。

删除节点

在 **故障诊断 > 诊断节点** 页面，选择要删除的诊断节点，单击 **操作** 列下的 **删除**，然后确定即可。

> 说明
> 无法删除已被诊断计划引用的节点。

12.5.3. 诊断决策树

您可以通过诊断决策树，编排故障排查的过程。对于已知的明确故障，可以根据诊断现象，编排诊断决策树，进而故障发生时执行，完成故障定位。对于未知的故障，您可以依据运维经验，编排出常见的排查路径，辅助快速故障定位。

新建诊断决策树

1. 从左侧导航栏进入 **故障诊断 > 诊断决策树** 页面，单击 **新建诊断决策树**。
2. 在右侧新窗口中，配置诊断决策树信息，包括决策树名称、描述，超时时间以及标签，然后单击 **确认**，新建的诊断决策树将出现在诊断决策树列表中。
3. 编辑诊断决策树，否则节点将无法执行。进入目标诊断决策树操作列下的 **详情** 进入的诊断计划详情页面，即可开始编辑决策树。
   - 单击 **添加初始节点**，在右侧窗口中，即可输入节点名称，选择之前创建好的诊断节点，单击 **确定**。
   - 单击添加的初始节点右侧的加号图标（⊕），即可在新窗口中添加初始节点后的下游执行节点。
   - 单击执行节点下方的加号图标（⊕），即可添加为该执行节点添加一个依赖节点。
   - 单击节点上的删除图标，即可将该节点删除。删除后，其所有子节点也将一并删除。



编辑诊断决策树

编辑基本信息

在 **故障诊断 > 诊断决策树** 页面，选择要修改的诊断决策树，单击 **操作** 列下的 **编辑** 按钮，可对计划名称、描述、超时时间以及标签进行修改。

编辑决策树

在 **故障诊断 > 诊断决策树** 页面，选择要修改的诊断计划，单击 **操作** 列下的 **详情** 按钮，进入 **编辑决策树** 页面，即可直接对决策树中各个节点进行增加、修改与删除操作。

执行诊断决策树

您可以通过以下任一方式执行诊断决策树：

- 在 **故障诊断 > 诊断决策树** 页面，选择要执行的诊断决策树，单击 **操作** 列下的 **执行** 按钮，然后确定即可。
- 在要执行的诊断决策树的详情页，单击决策树区域右侧的 **执行** 按钮，然后确定即可。

> 注意
> 仅支持手动执行故障诊断决策树。

删除诊断决策树

在 **故障诊断 > 诊断决策树** 页面，选择要删除的诊断决策树，单击 **操作** 列下的 **删除**，然后单击 **确定** 即可。

12.5.4. 诊断报告

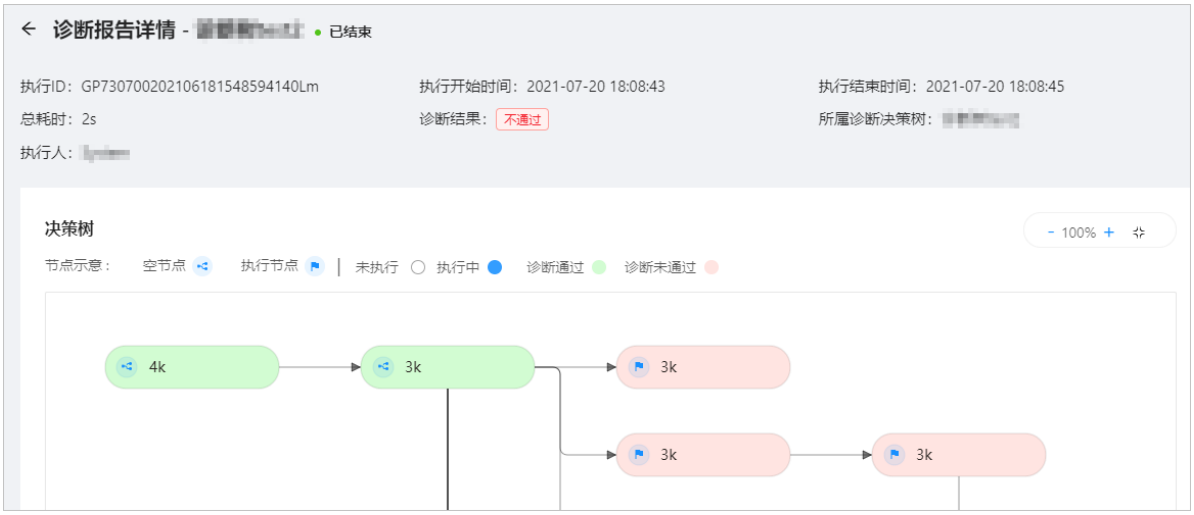
执行诊断决策树时，系统会生成对应的诊断报告，记录诊断的执行信息。你可以在 **诊断报告** 页面查看所有诊断决策树的执行记录，包括诊断决策树名称、触发方式、诊断状态、创建时间、状态概要、结果概要以及诊断结果。

在左侧导航栏上，单击 **故障诊断 > 诊断报告** 进入 **诊断报告** 页面，您可以根据诊断任务名称、诊断结果、诊断状态或触发方式来筛选诊断运行记录及其诊断数据。

在诊断任务列表中，选择目标决策树，单击 **操作** 列下的 **详情** 按钮，即可查看对应的诊断报告详情。

诊断报告详情包括两部分：

- **基本信息**：包括诊断决策树的执行 ID、执行开始与结束时间、总耗时、诊断结果、所属诊断计划以及执行人。
- **决策树**：在 **决策树** 区域，可以查看该诊断决策树的执行流程，以及各个节点的执行状态（未执行/执行中/诊断通过/诊断未通过）。



12.6. 应急预案

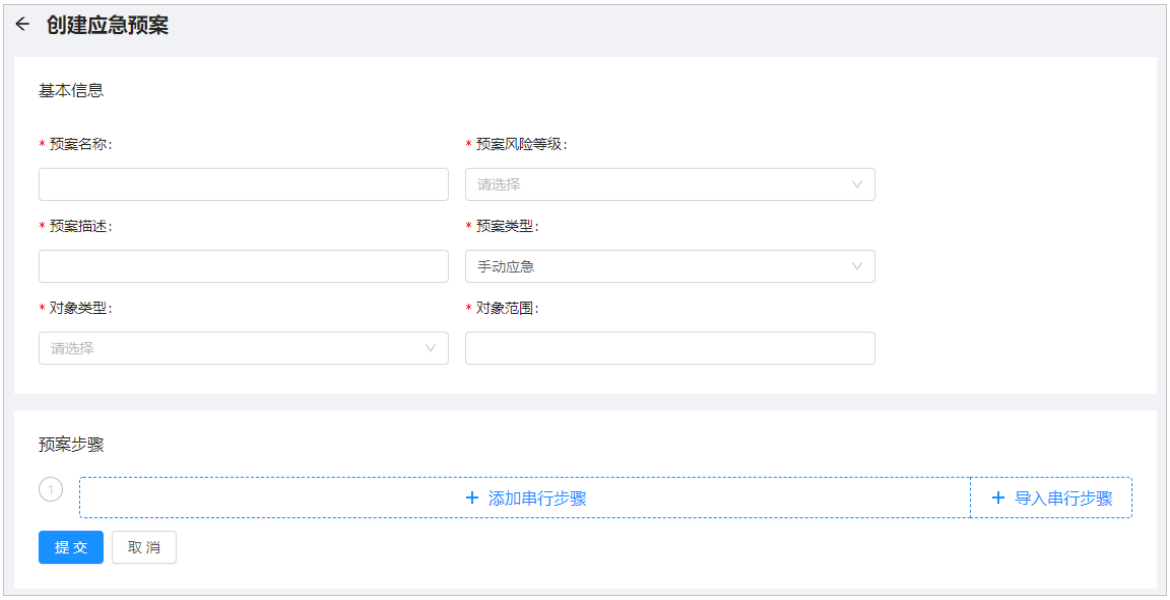
12.6.1. 预案管理

通过 **预案管理** 模块，您可以创建一个应急预案，直接执行或将其与风险事件进行关联执行。您还可以对所有的预案进行统一管理，包括查看、编辑、删除预案。

应急预案

创建应急预案

1. 登录高可用管理控制台。
2. 在左侧导航栏上，单击 **应急预案 > 预案管理**。
3. 在 **应急预案** 页签，单击 **创建应急预案**。
4. 在 **创建应急预案** 页面配置以下基本信息。



参数说明如下：

参数	参数说明
预案名称	输入预案名称。
预案风险等级	选择预案风险等级。共 P0~P4 5 个风险等级可选，P0 代表最高风险等级，P4 代表最低风险等级。
预案描述	输入预案描述。
预案类型	选择预案类型。 ○ **定时调度**：针对已知场景，定时执行的应急预案，通过钉钉通知到相关人员进行执行。 预案类型选择 **定时调度** 时需要配置定时通知策略和业务通知钉钉群，配置说明如下： ■ **开始时间**：必填，单击文本框设置定时通知开启时间。 ■ **结束时间**：必填，单击文本框设置定时通知结束时间。 ■ **通知策略**：必填，可选 **间隔** 或 **定时**，单击文本框填入时间。 ■ **通知状态**：默认为 **关闭** 状态，可手动打开/关闭通知开关。 ■ **业务通知钉钉群**：选填，选择需要接收通知的钉钉群，可多选。 ○ **手动应急**：对于通过巡检、诊断、监控上报的异常事件，通过风险管理模块进行展示，然后人工操作执行。
对象类型	选择预案的对象类型。 ○ **用户应用**：全部用户应用、指定用户应用。 ○ **蚂蚁产品**：全部蚂蚁产品、指定蚂蚁产品。 ○ **基础资源**：OB、ALB、AKE。

参数	参数说明
对象范围	输入预案的对象范围。 ○ 选择全部产品/应用：该预案可以和所有产品/应用巡检规则关联。执行的时候，针对报出风险的容器去重启。 ○ 选择指定产品/应用/资源：该预案只能和指定产品/应用/资源巡检规则关联。执行的时候，针对报出风险的容器去重启。

5. 在 **预案步骤** 区域配置串行步骤。
   i. 添加串行步骤。
   您可以选择添加串行步骤或导入串行步骤，操作步骤如下：
   - 添加串行步骤：单击 **添加串行步骤**，在右侧弹出的节点列表中选择原子操作，并根据界面提示配置原子操作信息。
   - 导入串行步骤：单击 **导入串行步骤**，在右侧弹出的预案列表中，从已有预案中导入预案步骤。

   > **说明**
   > 您可以添加多个串行步骤，各串行步骤将按照添加顺序依次执行。

   ii. （可选）添加并行步骤。
   您可以选择添加并行步骤或导入并行步骤，操作步骤如下：
   - 添加并行步骤：单击 **添加并行步骤**，在右侧弹出的节点列表中选择原子操作，并根据界面提示配置原子操作信息。
   - 导入并行步骤：单击 **导入并行步骤**，在右侧弹出的预案列表中，从已有预案中导入预案步骤。

   > **说明**
   > 您可以添加多个并行步骤，各并行步骤将按照添加顺序依次执行。

   iii. （可选）调整节点顺序。
   节点添加完毕后，您可以通过单击各节点右上方的 **上移** 或 **下移** 按钮进行顺序调整。
6. 单击 **提交**，返回至应急预案列表。
   此时，预案状态为 **可用**，您可以执行该应急预案。

执行应急预案

1. 在左侧导航栏上，单击 **应急预案** > **预案管理**。
2. 在 **应急预案** 页签，单击目标应急预案 **操作** 列下的 **更多** > **执行**。您也可以单击目标应急预案名称进入 **应急预案详情** 页面，单击页面右上角 **执行**。
3. 在 **执行应急预案** 页面，填写预案执行单名称，并选择目标机房。



4. 单击 **确定** 跳转至预案执行页面。
5. 单击 **执行**，系统将自动按步骤依次执行功能或资源的切换或恢复。

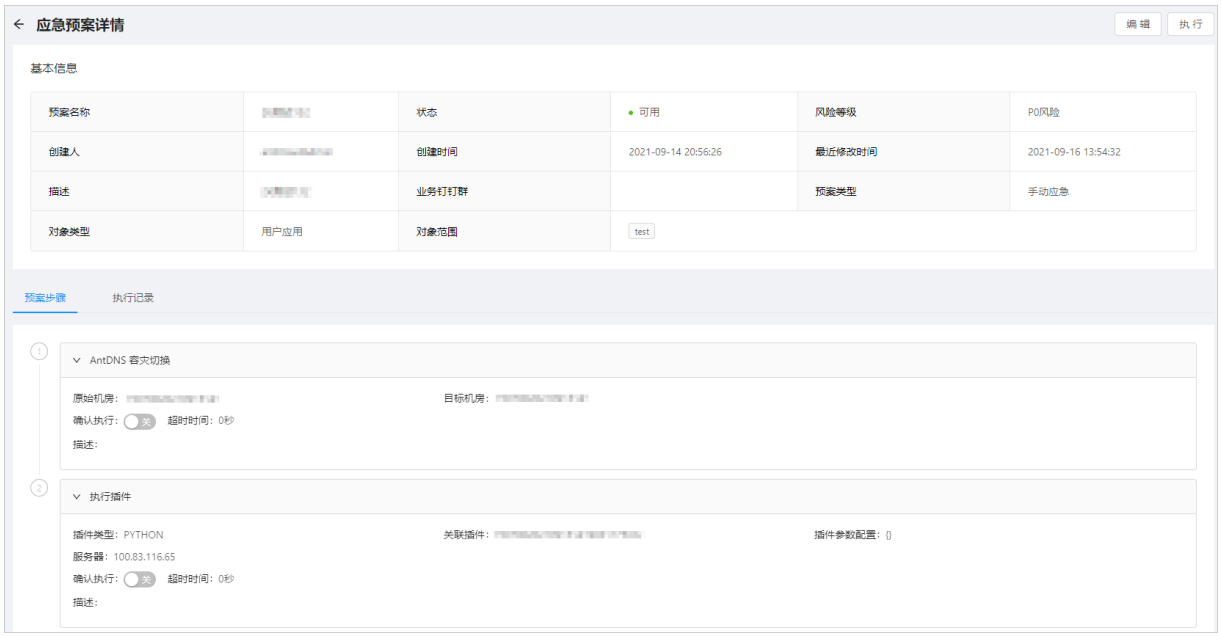
> 说明
> - 若无需执行该预案，您可以单击 **取消执行**。
> - 若在执行过程中出现错误，您可以查看失败步骤对应的日志，分析出错原因，待问题排查完成后，单击 **重试**。或者单击 **跳过** 忽略该步骤。

更多操作

除了创建和执行应急预案，您还可以进行编辑、克隆、删除预案以及查看预案详情操作。

查看应急预案详情

应急预案创建完成后，您可以在应急预案列表中，单击预案名称，进入预案详情页面查看详情。详情页中，展示了预案的基本信息、预案步骤以及详细的执行记录。





- **基本信息**：展示当前应急预案的基本信息，包括预案名称、状态、风险等级、创建人、创建时间、最近修改时间、描述、预案类型、对象类型以及对象范围。
- **预案步骤**：展示当前应急预案的执行步骤。
- **执行记录**：展示当前应急预案的执行记录，包括预案执行单名称、预案类型、对象、状态、创建人、开始时间以及操作人。

编辑应急预案

您可以通过以下任一方式编辑预案信息：

- 方法一：
  i. 在左侧导航栏上，单击 **应急预案** > **预案管理**。
  ii. 在 **应急预案** 页签，单击目标预案 **操作** 列下的 **编辑**，在 **编辑应急预案** 页面，修改预案信息。
  iii. 单击 **提交**。
- 方法二：
  i. 在左侧导航栏上，单击 **应急预案** > **预案管理**。
  ii. 在 **应急预案** 页签，单击目标应急预案名称进入 **应急预案详情** 页面，单击页面右上角 **编辑**，在 **编辑应急预案** 页面，修改预案信息。
  iii. 单击 **提交**。

克隆应急预案

预案管理模块支持克隆已有的应急预案，您可以在已有预案的基础上进行调整，生成为新的预案。
操作步骤说明如下：

1. 在左侧导航栏上，单击 **应急预案 > 预案管理**。
2. 在 **应急预案** 页签，单击目标预案 **操作** 列下的 **更多 > 克隆**。
   克隆的预案将出现在预案列表中，当前预案状态为 **编辑中**，不能被执行，需要在提交后方可执行。
3. 单击克隆的预案 **操作** 列下的 **编辑**，在 **编辑应急预案** 页面，修改预案信息，如名称、描述、风险等级，并对预案步骤进行设置、删除或顺序调整。
4. 单击 **提交** 使该预案生效。

删除应急预案

> 注意
> 预案一旦删除，不可恢复，务必谨慎操作。

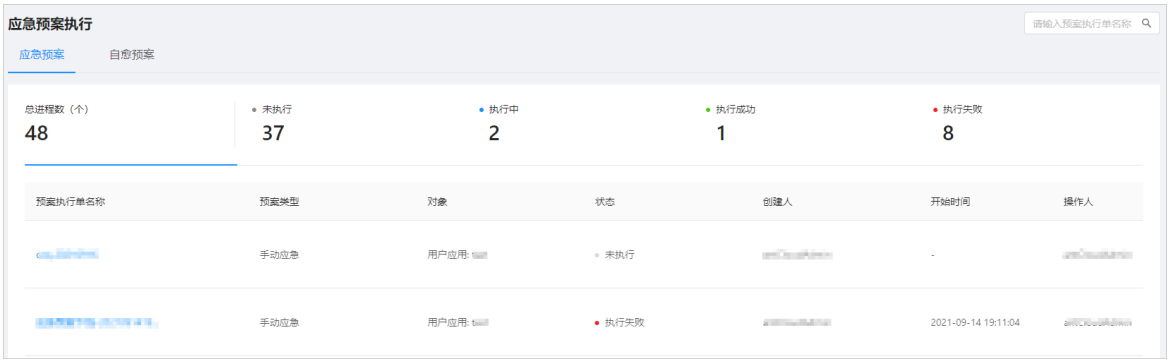
1. 在左侧导航栏上，单击 **应急预案 > 预案管理**。
2. 在 **应急预案** 页签，单击目标预案 **操作** 列下的 **更多 > 删除**，并在弹出的对话框中单击 **确定**。

12.6.2. 预案执行

预案执行时，系统会生成一个对应的预案执行单，记录预案的执行信息。

查看应急预案执行单

1. 登录高可用管理控制台。
2. 在左侧导航栏上，单击 **应急预案 > 预案执行**。
3. 选择 **应急预案** 页签。
   应急预案总览页面展示了当前环境的预案执行单总量、各种状态的（未执行、执行中、执行成功、执行失败、参数待确认）执行单数量以及预案执行单列表。所有生成的预案执行单（包括未执行的）都包含在列表中。



4. 单击目标预案执行单名称，进入 **应急预案执行详情** 页面。

> 说明
> 您也可以在列表右上方的搜索框中输入执行单名称来查找目标执行单。

应急预案执行详情页面包括以下 2 部分内容：

- **基本信息：** 展示当前预案的执行单元、切换状态、风险等级、开始/结束时间、耗时等信息。



- **执行信息**：展示当前预案各步骤的执行情况。对于执行失败的容灾步骤，您可以进行以下操作：
  - 单击 **重试** 再次执行该步骤。
  - 单击 **日志** 查看日志详情，分析出错原因。待问题排查完成后，单击 **重试** 重新执行。
  - 单击 **跳过** 忽略该步骤。



查看自愈预案执行单

1. 在左侧导航栏上，单击 **应急预案** > **预案执行**。
2. 选择 **自愈预案** 页签。
   自愈预案总览页面展示了当前环境的预案执行单总量、各种状态的（未执行、执行中、执行成功、执行失败）执行单数量以及预案执行单列表。所有生成的预案执行单（包括未执行的）都包含在列表中。

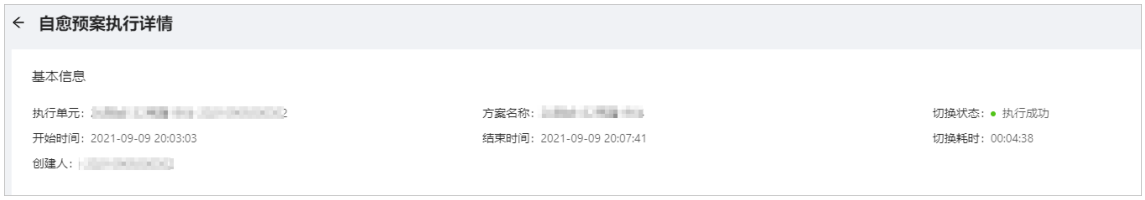


3. 单击目标预案执行单名称，进入 **自愈预案执行详情** 页面。

> **说明**
> 您也可以在列表右上方的搜索框中输入执行单名称来查找目标执行单。

自愈预案执行详情页面包括以下 2 部分内容：

- **基本信息**：展示当前预案的执行单元、切换状态、开始/结束时间、耗时等信息。

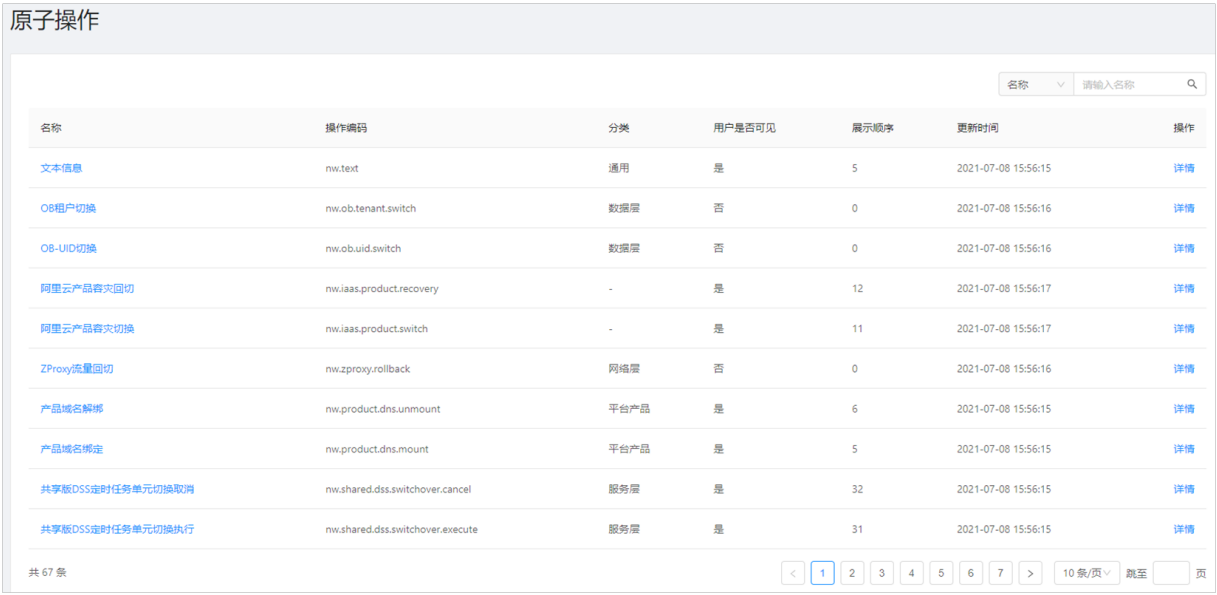


- **执行信息**：展示当前预案各步骤的执行情况。对于执行失败的步骤，您可以进行以下操作：
  - 单击 **重试** 再次执行该步骤。
  - 单击 **日志** 查看日志详情，分析出错原因。待问题排查完成后，单击 **重试** 重新执行。
  - 单击 **跳过** 忽略该步骤。

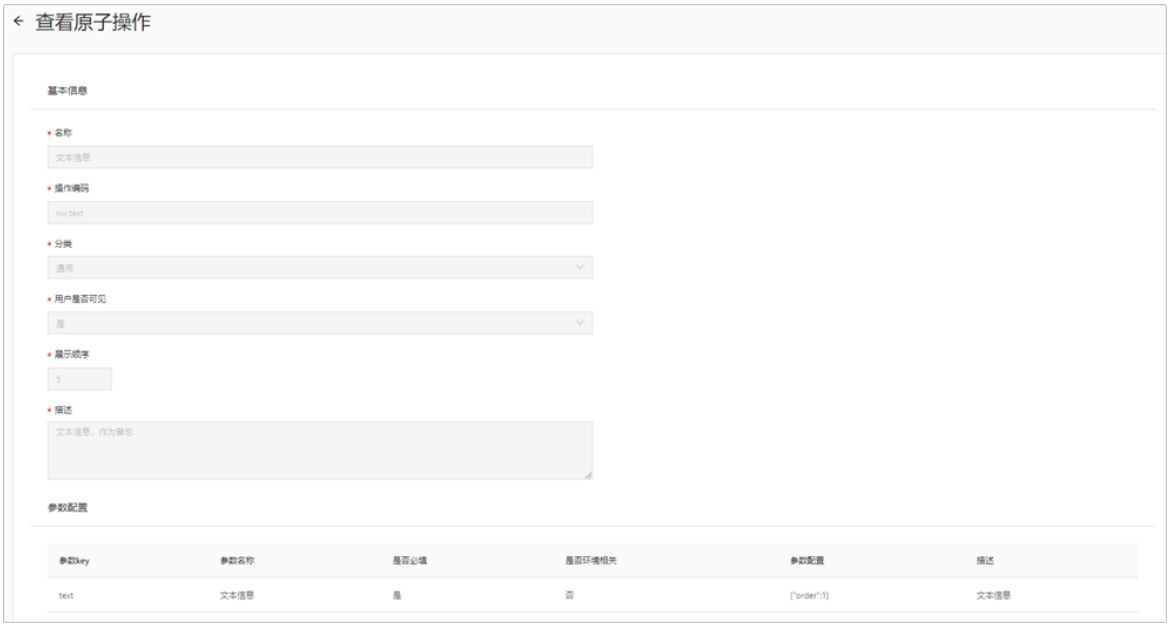
12.6.3. 预案原子操作

预案原子操作指最小单元的应急或容灾操作，您可以将其应用到应急预案或容灾预案中。

查看预案原子操作详情



1. 在左侧导航栏上，点击 **应急预案 > 预案原子操作**，进入容灾预案原子操作管理页。
2. 点击需要查看的容灾预案原子操作名称或对应的 **详情** 按钮，可进入原子操作详情页查看基本信息和参数配置信息。



12.7. 故障演练

12.7.1. 演练场景

故障演练是业务系统上线前必要的演练环节，旨在对应急预案的可行性，进一步完成应急预案，从而帮助提升产品、集群、机房的稳定性，减少故障的发生，提高故障应急效率，进而提升产品竞争力。
一个完整的故障演练过程为 **创建故障原子服务 > 创建故障演练场景 > 创建并发布演练方案 > 创建演练执行单并执行 > 查看演练报告 > 导出演练方案**。
演练场景模块允许您根据容灾需要，自行设计故障演练场景，并对已创建的演练场景进行管理。



创建演练场景

在设置故障演练场景的每一个步骤时，可以根据需要设置是否在演练时自动执行该步骤。
操作步骤如下：

1. 登录高可用管理控制台。
2. 在左侧导航栏上，单击 **故障演练 > 演练场景**。
3. 单击页面上的 **创建演练场景** 按钮，进入演练场景创建页面。
4. 完成基本信息配置后，单击 **下一步**。
   - **场景名称**：自定义的场景名称。必填。
   - **标签**：选择所属标签。
   - **场景描述**：输入关于故障场景的描述。

- **预期目标**：填写故障可能会造成的影响。
- **问题严重度**：选择问题严重度。

> 说明
> 关于问题严重程度，P0 代表严重程度最高，P4 代表严重程度最低。

- **关联的风险场景**：选择关联的风险场景。如需关联多个风险场景，单击 **添加** 即可。

5. （可选）完成前置检查项配置后，单击 **下一步**。
   i. 设置是否在演练时自动执行该步骤，系统默认自动执行。若不想要自动执行，将 **是否自动执行** 设为 **否**。
   ii. 添加检查项。支持通过巡检检查和通过测试镜像检查两种方式，可以添加多个检查项。
      - 如果选择 **通过巡检检查**，需进行以下配置：
        - **巡检规则**：执行检查的巡检规则。
        - **执行机房**：执行巡检规则的机房。
        - **期望结果**：期望巡检通过或失败。
      - 如果选择 **通过测试镜像检查**，需进行以下配置：
        - **镜像产品**：选择测试镜像。镜像产品必须与基本信息中选择的所属产品保持一致。
        - **期望结果**：期望巡检通过或失败。
6. 注入故障。完后以下配置后，单击 **下一步**。
   i. 设置是否在演练时自动执行该步骤，系统默认自动执行。若不想要自动执行，将 **是否自动执行** 设为 **否**。
   ii. 单击 **添加故障注入**，完成故障注入配置。
      - **故障节点类型**：支持 **蚂蚁产品**、**用户应用**、**AKE 宿主机节点** 或 **自定义服务器**。
        - 如果选择 **蚂蚁产品**，需进行以下配置。

配置项	说明
故障注入产品	选择注入故障的产品。
应用	选择注入故障的应用。
IP 列表	指定执行演练的 IP 地址，如果有多个，可以用逗号分隔。 > 说明 > 指定 IP 后，该演练场景不能在多个环境之间进行同步。
原子操作	根据实际业务需求选择原子操作。 > 注意 > 请确保集群中已安装故障注入组件。该组件可以通过 **容器应用服务（AKS）** > **集群管理** > **组件管理** 进行安装， 或通过 **云应用引擎 (Captain)** > **应用商店** 上传 故障注入组件产品模版 进行安装。

- 如果选择 **用户应用**，需进行以下配置。

配置项	说明
故障注入应用	选择注入故障的应用。支持选择多个应用。
工作空间组	选择注入故障应用所在的工作空间组。支持选择多个工作空间组。
命名空间	选择注入故障应用所在的命名空间。支持选择多个命名空间。
应用服务	选择一个或多个应用服务。
IP 列表	指定执行演练的 IP 地址，如果有多个，可以用逗号分隔。 **说明** 指定 IP 后，该演练场景不能在多个环境之间进行同步。
label	指定物理机的 label。
原子操作	根据实际业务需求选择原子操作。 **注意** 请确保集群中已安装故障注入组件。该组件可以通过 **容器应用服务（AKS）> 集群管理 > 组件管理** 进行安装， 或通过 **云应用引擎 (Captain) > 应用商店** 上传 故障注入组件产品模版 进行安装。

- 如果选择 **AKE 宿主机节点** 或 **自定义服务器**，需进行以下配置。

配置项	说明
IP 列表	指定执行演练的 IP 地址，如果有多个，可以用逗号分隔。 **说明** 指定 IP 后，该演练场景不能在多个环境之间进行同步。
原子操作	根据实际业务需求选择原子操作。 **注意** 请确保集群中已安装故障注入组件。该组件可以通过 **容器应用服务（AKS）> 集群管理 > 组件管理** 进行安装， 或通过 **云应用引擎 (Captain) > 应用商店** 上传 故障注入组件产品模版 进行安装。

- **原子操作设置：** 根据所选的原子操作设置对应的参数值。

7. （可选）故障恢复。完成以下配置后，单击 **完成**。

   i. 设置是否在演练时自动执行该步骤，系统默认不自动执行。若想要自动执行，将 **是否自动执行** 设为 **是**。

ii. 选择应急预案。

修改演练场景

可通过如下方式对演练场景执行修改操作：

- 在演练场景列表页，选择目标演练场景，单击 **操作** 列下的 **编辑**，即可编辑场景，后续操作同创建演练场景。
- 单击某个演练场景名称，进入演练场景详情页，然后单击页面右上方的 **编辑** 按钮，开始修改演练场景中的各个步骤。

克隆演练场景

克隆演练场景，指基于已有的演练场景复制出一个新的演练场景。
在演练场景列表页，选择目标演练场景，单击 **操作** 列下的 **更多 > 克隆**，克隆出一个新场景。然后根据需要，对克隆出来的演练场景配置进行修改。修改操作同编辑演练场景。

废弃演练场景

对于不再使用的演练场景，可以将其废弃。
在演练场景列表页，选择目标演练场景，单击 **操作** 列下的 **更多 > 废弃**，使该场景失效。废弃后的演练场景后，仍会显示在演练场景列表中，状态会变为已废弃。

> 说明
> 如果确定不再需要废弃后的演练场景，可将其删除。详情请参见 删除演练场景。

删除演练场景

> 注意
> 演练场景删除后，不可恢复，请谨慎操作。

可以通过如下方式来删除演练场景：
在演练场景列表页，选择目标演练场景，单击 **操作** 列下的 **更多 > 删除。**演练场景删除后，将不再显示在演练场景列表中。

更多操作

在演练场景列表页，您还可以基于目标演练场景快速创建演练方案，具体步骤如下：

1. 选择目标演练场景，单击 **操作** 列下的 **更多 > 创建演练方案。**
2. 在 **创建演练方案** 页面，只需填写演练方案名称和描述即可完成创建。详情请参见 创建演练方案。

12.7.2. 演练方案

演练方案指针对不同故障场景设计、编排的一套容灾演练计划。故障演练模块支持创建、编辑、发布、复用、删除、导出演练方案，以及查看演练记录，下面逐一进行操作说明。

创建演练方案

1. 在左侧导航栏上，单击 **故障演练 > 演练方案**，进入故障演练方案管理页面。
2. 单击页面上的 **创建演练方案** 按钮，进入演练方案创建页面。
3. 配置基本信息，包括方案名称、方案描述。
4. 添加演练场景，并设置演练场景的执行顺序。一个方案中可以包含多个演练场景。
5. 完成上述配置后，单击 **保存**，保存当前演练方案。或者，单击 **保存并创建执行单**，保存方案的同时创建相应的演练执行单。

查看演练方案详情

在故障演练方案管理页面上，单击目标演练方案卡片，进入演练方案详情页查看即可。演练方案详情主要包括基本信息、演练场景和演练记录三部分。

- **基本信息**：包括方案创建时间、创建人、最近修改时间、方案描述。
- **演练场景**：展示当前方案中涉及的演练场景信息。
- **演练记录**：仅执行过的演练方案才会有演练记录，包括演练执行单名称、演练开始时间、结束时间、耗时、执行详情、演练报告等。

编辑演练方案

在故障演练方案管理页面上，单击目标演练方案卡片上的 **更多 > 编辑**，开始编辑该方案，后续操作同创建演练方案。或者，单击目标演练方案卡片，进入演练方案详情页，然后单击页面右上方的 **编辑** 按钮，开始修改演练方案内容。

克隆演练方案

克隆演练方案，指基于已有的演练方案复制出一个新的演练方案。

在故障演练方案管理页面上，单击目标演练方案卡片上的 **更多 > 克隆**，复制出一个新方案。然后根据需要，对复用出来的演练方案进行修改。修改操作同编辑演练方案。

删除演练方案

在故障演练方案管理页面上，单击目标演练方案卡片上的 **更多 > 删除**，删除当前方案。或者，单击目标演练方案卡片，进入演练方案详情页，然后单击页面右上方的 **更多 > 删除** 按钮进行删除。

12.7.3. 演练执行

您可以在 **演练执行** 页面查看演练执行单总量、各执行状态下的执行单数据，执行单列表。

查看执行详情

1. 在左侧导航栏上，单击 **故障演练 > 演练执行**，进入故障执行单列表页面。
2. 单击目标执行单右侧的 **执行详情**，查看执行单的详细信息，包括执行状态（执行成功、执行失败、未执行、执行中、取消执行）、执行进程、基本信息。
   针对不同状态的执行单，进行不同的操作：
   - 对于 **执行中** 的执行单，可以取消执行。
   - 对于 **执行失败** 的执行单，可以重新执行容灾演练步骤。单击失败的步骤卡片，可以查看失败详情，从而进行问题排查。
   - 对于 **执行成功** 的执行单，可以返回演练方案查看演练场景和演练记录。
   - 对于 **已取消** 的执行单，可以重新执行容灾演练步骤。

12.7.4. 演练原子操作

故障演练原子操作指最小单元的故障。在 **故障演练 > 演练原子操作** 页面中，展示了平台上可用于故障演练的原子服务。可根据原子服务状态分类，查看当前环境中的故障原子操作信息，包括原子操作名称、操作编码、添加时间、状态、被引用次数等信息。

> 说明
> 仅已发布状态下的演练原子操作可引用到故障演练中的故障注入环节。已废弃的原子操作无法被引用。



12.8. 应急管理

通过 **应急团队管理** 模块，您可以创建一个应急团队，将其与风险场景进行关联执行。您还可以对所有的应急团队进行统一管理，包括查看、编辑和删除应急团队。

创建应急团队

前提条件

- 已添加风险场景名称。
  如需新增风险场景名称，请参见 创建风险场景。
- 已添加钉钉机器人。
  如需新增钉钉机器人，请参见 添加钉钉机器人。

操作步骤

1. 在左侧导航栏上，单击 **应急管理** > **应急团队管理**，进入 **应急管理团队** 页面。
2. 单击 **创建应急团队**，进入应急团队创建页面。

3. 在弹出的页面中输入 **应急团队名称**、**团队成员** 及 **描述**，添加关联 **风险场景名称** 、 **钉钉机器人**、**等级**。



4. 单击 **确定**。

查看应急团队

应急团队创建完成后，您可以在应急团队列表页查看团队详情，包括团队名称、团队成员、关联风险场景以及创建时间。



编辑应急团队

1. 在左侧导航栏上，单击 **应急管理 > 应急团队管理**，进入 **应急管理团队** 页面。
2. 在应急团队列表中，找到目标团队，单击其操作列下的 **编辑** 按钮，即可进入编辑页修改团队信息。

删除应急团队

> **注意**
> 应急团队一旦删除，不可恢复，务必谨慎操作。

1. 在左侧导航栏上，单击 **应急管理 > 应急团队管理**，进入 **应急管理团队** 页面。
2. 在应急团队列表中，找到目标团队，单击其操作列下的 **删除** 按钮，并在弹出的对话框中单击 **确定** 删除该应急团队。

12.9. 配置管理

12.9.1. 审批管理

审批管理 模块允许您基于风险防控方面的考量，自行创建审批流程，针对巡检插件的发布、规则的发布、修改和删除进行审批。

创建审批流程

1. 登录高可用管理控制台。
2. 在左侧导航栏上，单击 **配置管理 > 审批管理。**
3. 单击 **新建审批流程**，在右侧弹出的 **新建审批流程** 页面上，填写以下信息。



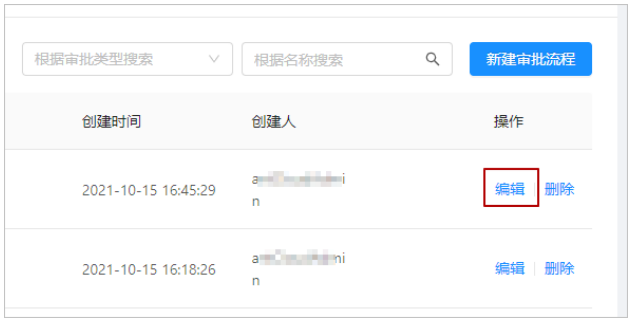
- **审批类型**：支持 **发布巡检插件**、**发布巡检规则**、**修改巡检规则**、**删除巡检规则** 四种类型。
- **审批流程名称**：输入自定义的审批流程名。
- **提交人选择**：提交人信息。
- **审批人选择**：节点信息、审批人信息。单击 **添加审批人** 即可添加多个审批人。

> 说明
> 多人审批时，一人同意即为通过，一人拒绝即为拒绝。

4. 单击 **确认**，新建的审批流程将显示在审批流程列表中。

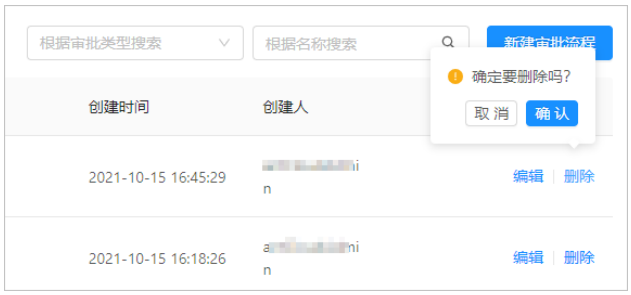
编辑审批流程

1. 在左侧导航栏上，单击 **配置管理 > 审批管理**。
2. 选择要编辑的审批流程，单击操作栏下方的 **编辑** 按钮，修改后，单击 **确认**。



删除审批流程

1. 在左侧导航栏上，单击 **配置管理 > 审批管理**。
2. 选择要删除的审批流程，单击操作栏下方的 **删除** 按钮，然后单击 **确认**。



搜索审批流程

1. 在左侧导航栏上，单击 **配置管理 > 审批管理**。
2. 在输入栏内，输入审批流程名称的关键字，单击搜索图标，搜索结果将自动显示在下方。
   您也可以从下拉框中选择所需的审批类型，按审批类型进行搜索。

我的审批

1. 在左侧导航栏上，单击 **我的审批**。
2. 在 **我的审批** 页面上，您可以查看以下三种状态的审批流：
   - **待处理**：需要由您进行审批的审批流。
   - **已处理**：您已审批的审批流。
   - **我发起的**：您已发起审批的审批流。

12.9.2. 基线管理

在 **基线管理** 模块，您可以拉取相关对接产品的基线数据，使高可用管理平台上的产品应用数据与其保持同步。从左侧导航栏进入 **配置管理 > 基线管理** 页面，单击页面右上方的 **数据同步** 按钮开始同步数据。
基线管理页面展示了当前租户下的 PaaS 产品、业务应用以及 K8s 基线信息，支持按产品码或分组名搜索应用。

PaaS 产品

对 **PaaS 产品** 按照产品码进行分类展示。单击产品码列前面的展开图标（ + ），查看该产品下的应用信息，包括应用名称、版本、状态、创建时间和更新时间。

业务应用

对业务应用按照应用分组进行分类展示。单击应用分组名列前面的展开图标（ + ），查看该分组下的应用信息，包括应用名称、版本、状态、创建时间和更新时间。

K8s 基线

展示了当前租户下所有 K8s 资源信息，包括 IaasMetaData 名称、命名空间、Pod、容器名、创建时间和更新时间。

业务应用分组

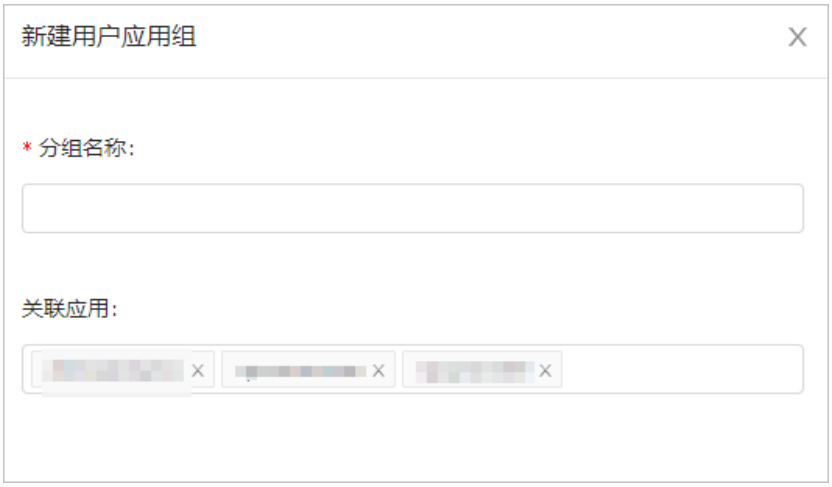
为方便您在面对若干应用时能快速选择同一业务域内的一组应用，基线管理模块推出了业务应用分组的功能。在新建巡检规则或者应急预案时，借助该功能，可轻松将设定的某类规则或配置项应用到某组应用组，免去了针对多个应用分别进行选择、配置等耗时费力的操作，从而达到快速批量选择和配置的目的。

新建用户应用组

1. 登录高可用管理控制台，从左侧导航栏选择 **配置管理 > 基线管理**。
2. 单击 **业务应用分组** 下方的 **新建业务应用分组**。



3. 在右侧弹出的 **新建用户应用组** 页面上，填写以下信息。



   - **分组名称**：必填项。输入自定义的分组名。
   - **关联应用**：从下拉列表中选择需要关联的应用，可以是多个应用。

4. 填写完上述信息后单击 **确认** 按钮，即可新增一个应用分组，成功增加的分组将以列表的形式展现。

12.9.3. 钉钉机器人

高可用管理平台支持添加与管理钉钉机器人。在巡检任务编辑时，添加钉钉机器人，可以将巡检结果在对应的钉钉群里进行推送通知。

添加钉钉机器人

1. 登录高可用管理控制台。
2. 在左侧导航栏上，单击 **配置管理** > **钉钉机器人管理**。
3. 单击 **添加钉钉机器人**。
4. 在右侧弹窗中，添加钉钉机器人信息，包括钉钉群名称、Token 连接、备注，同时也支持加签认证。



编辑钉钉机器人

编辑基本信息

在 **钉钉机器人管理** 页面，选择要编辑的钉钉群，单击 **操作** 列下的 **编辑**，即可在右侧的窗口中修改钉钉群名与备注信息。

更改认证方式

在 **钉钉机器人管理** 页面，选择要修改的钉钉群，单击 **操作** 列下的 **更改认证方式**，即可在弹出的窗口中修改 Token 按钮与加签认证配置。

查看使用详情

在 **钉钉机器人管理** 页面中，选择要查看使用详情的钉钉群，单击 **操作** 列下的 **使用详情**，即可在右侧的窗口中查看该钉钉群已关联的具体任务。