



DataWorks DataWorks公共云合集

文档版本: 20220712



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。

- 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。
- 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。
- 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。	會学者 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。	大) 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 <i>,</i> 不是 用户必须了解的内容。	⑦ 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 listinstanceid
[] 或者 [alb]	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.产品版本规格与资源组	11
1.1. DataWorks版本规格	11
1.1.1. DataWorks各版本详解	11
1.1.2. DataWorks增值版本到期降级功能变化清单	20
1.2. DataWorks资源组	22
1.2.1. DataWorks资源组概述	22
1.2.2. 公共资源组	25
1.2.2.1. 公共资源组概述	25
1.2.2.2 使用公共资源组	27
1.2.3. 独享数据集成资源组	27
1.2.3.1. 独享数据集成资源组概述	27
1.2.3.2. 新增和使用独享数据集成资源组	30
1.2.4. 独享调度资源组	40
1.2.4.1. 独享调度资源组概述	40
1.2.4.2. 新增和使用独享调度资源组	42
1.2.5. 独享数据服务资源组	50
1.2.5.1. 独享数据服务资源组	50
1.2.5.2. 新增和使用独享数据服务资源组	51
1.2.6. 自定义资源组	54
1.2.6.1. 自定义资源组概述	54
1.2.6.2. 新增和使用自定义调度资源组	55
1.2.6.3. 新增和使用自定义数据集成资源组	61
1.2.7. 通用参考: 修改资源组归属工作空间	69
1.2.8. 通用参考: 取消包年包月产品的自动续费	72
1.2.9. 通用参考: 按量付费转包年包月	74
1.3. 变更规格操作	76

1.4. 扩缩容操作	81
1.5. 扣费与欠费说明	83
1.6. 到期说明及续费操作	87
2.基础操作	91
2.1. 工作空间管理	91
2.1.1. 配置工作空间	91
2.1.2. MaxCompute高级配置	106
2.1.3. 数据源管理	108
2.1.4. 工作空间模式升级	108
2.2. 工作空间规划与管理	112
2.2.1. 规划工作空间	112
2.2.2. 修改工作空间	116
2.3. 权限管控	118
2.3.1. DataWorks权限管控概述	118
2.3.2. 管控权限详情: 细分RAM Policy	119
2.3.3. 使用权限详情: 用户与角色	128
2.3.3.1. 用户、角色与权限概述	128
2.3.3.2. 角色及成员管理: 空间级	137
2.3.3.3. 角色及成员管理: 全局级	143
2.3.3.4. 附录: 预设角色权限列表(空间级)	146
2.3.4. 规划与配置角色权限	152
3.运维中心	154
3.1. 运维中心概述	154
3.2. 查看运维大屏	160
3.3. 实时任务运维	168
3.3.1. 实时计算任务运行与管理	168
3.3.2. 实时同步任务运行与管理	170
3.4. 周期任务运维	175

3.4.1. 周期任务运维概述	175
3.4.2. 查看并管理周期任务	184
3.4.3. 查看周期实例	188
3.4.4. 执行补数据并查看补数据实例	195
3.4.5. 执行测试并查看测试实例	207
3.4.6. DAG图功能介绍	213
3.4.7. 强制重跑下游	220
3.5. 手动任务运维	225
3.5.1. 运行并管理手动任务	225
3.5.2. 查看并管理手动实例	229
3.6. 智能监控	231
3.6.1. 智能监控概述	231
3.6.2. 智能基线	236
3.6.2.1. 智能基线概述	237
3.6.2.2. 基线管理	243
3.6.2.3. 基线实例	250
3.6.2.4. 事件管理	253
3.6.2.5. 附件	254
3.6.2.5.1. 合理配置基线承诺时间和预警余量	254
3.6.3. 规则管理	256
3.6.3.1. 自定义规则	256
3.6.3.2. 值班表	263
3.6.4. 自动运维	268
3.6.5. 报警信息	271
3.6.6. 常见问题	273
3.7. 智能诊断	274
3.8. 资源运维	279
3.9. 引擎运维	282

3.10. 操作历史	285
3.11. 常见问题	286
3.11.1. 数据未正常产出或任务未运行如何处理	286
3.11.2. 找不到任务实例如何处理	287
4.发布中心	289
4.1. 发布中心概述	289
4.2. 创建并配置发布环境	293
4.3. 创建发布包	300
4.4. 创建及发布发布包:跨云	301
4.5. 查看发布包列表	305
5.开放平台	308
5.1. 开放平台概览	308
5.2. 开放API(OpenAPI)	309
5.3. 开放事件(OpenEvent)	310
5.3.1. OpenEvent概述	310
5.3.2. 开启消息订阅	312
5.3.3. 附录:DataWorks服务关联角色	314
5.3.4. 附录:发送给EventBridge的消息格式	316
5.4. 扩展程序(Extensions)	342
5.4.1. 扩展程序概述	342
5.4.2. 准备工作	345
5.4.3. 开发部署扩展程序	348
5.4.4. 测试应用扩展程序	354
5.4.5. 附录	357
5.4.5.1. 参考: 扩展程序参数配置	357
5.4.5.2. 参考: 扩展程序选项配置	360
5.4.5.3. 应用示例:消息处理代码DEMO	368
5.4.5.4. 应用示例:数据开发过程触发事件检查	371

5.4.5.5. 应用示例: 运维中心触发事件检查	375
6.数据建模	270
6.1. DataWorks智能数据建模	379 379
6.1.1. 概述	379
6.1.2. 数仓规划	382
6.1.2.1. 数仓规划概述	382
6.1.2.2. 业务分类	384
6.1.2.3. 数仓分层	387
6.1.2.3.1. 创建数仓分层	387
6.1.2.3.2. 配置数仓分层检查器	391
6.1.2.3.3. 查看并管理数据分层	394
6.1.2.4. 公共层	395
6.1.2.4.1. 数据域	396
6.1.2.4.2. 业务过程	399
6.1.2.5. 应用层	402
6.1.2.5.1. 数据集市	402
6.1.2.5.2. 主题域	404
6.1.2.6. 维度管理	406
6.1.2.7. 建模空间	408
6.1.3. 数据标准	410
6.1.3.1. 数据标准概述	410
6.1.3.2. 字段标准	411
6.1.3.3. 标准代码	415
6.1.3.4. 度量单位	418
6.1.3.5. 命名词典	419
6.1.4. 维度建模	422
6.1.4.1. 维度建模概述	422
6.1.4.2. 创建维度表	423

6.1.4.3. 创建明细表	425
6.1.4.4. 创建汇总表	428
6.1.4.5. 创建应用表	430
6.1.4.6. 发布与管理表	432
6.1.4.7. 数据落标	440
6.1.4.8. 使用FML语句配置并管理数据表	442
6.1.4.9. 导入与导出表配置	446
6.1.4.10. 附录: 使用画板工具创建表	447
6.1.4.11. 逆向建模	448
6.1.4.12. 模型开发	452
6.1.5. 数据指标	454
6.1.5.1. 数据指标概述	454
6.1.5.2. 原子指标	455
6.1.5.3. 修饰词	460
6.1.5.4. 时间周期	464
6.1.5.5. 派生指标	467
7.HoloStudio	476
7.1. 概述	476
7.2. HoloStudio快速入门	476
7.3. 绑定交互式分析实例	482
7.4. 新建数据库	488
7.5. 子账号使用DataWorks	492
7.6. SQL Console	494
7.7. PG管理	497
7.7.1. 创建及管理内部表	497
7.7.2. 创建及管理外部表	501
7.8. 数据开发	504
7.8.1. 概述	504

7.8.2. Hologres开发: 周期性调度	505
7.8.3. 一键同步MaxCompute表结构	512
7.8.4. 一键同步MaxCompute数据	514
7.8.5. 一键上传本地文件	516
7.9. 终端	519

1.产品版本规格与资源组

1.1. DataWorks版本规格

1.1.1. DataWorks各版本详解

DataWorks版本包括基础版和增值版本两大类,其中增值版本包括标准版、专业版、企业版和旗舰版,您可以 根据自身的业务场景,选择合适的版本进行购买。本文为您介绍DataWorks各版本的推荐应用场景和功能差 异。

DataWorks各版本购买注意事项

- DataWorks的基础版不收取版本功能费用, DataWorks增值版的版本功能支持包年包月的计费方式进行计费, 计费详情可参见计费逻辑说明。
- 同一地域在相同时间段内, 仅支持开通购买一个DataWorks增值版本。已购买的版本到期释放后, 才可以继续在该地域购买其它版本。如果您希望到期前变更使用的DataWorks版本, 可进行变配操作升级版本。
- 如果您购买使用的是DataWorks增值版,如果因为没有及时续费,增值版本到期后会降级为基础版。降级后版本功能会变化,详情可参见DataWorks增值版本到期降级功能变化清单。

推荐应用场景

DataWorks基础版和增值版各版本的主要功能和推荐应用的场景如下表。

DataWorks版本	推荐应用场景说明
基础版	开通DataWorks按量付费后,即可免费获得DataWorks基础版。您可以通过该版本 完成基本的数据上云、数据开发与调度生产、简单的数据治理工作。
增值版-标准版	基于DataWorks基础版,增加更专业的数据治理功能,同时增强智能编辑器、数据 集成的功能。
增值版-专业版	基于DataWorks标准版,增加更专业的数据安全功能,同时增强数据治理、数据服 务的功能,提供产品级别化的数据治理、数据安全解决方案。
增值版-企业版	基于DataWorks专业版,您可以获得更加智能化的数据资产治理能力与更加人性化 的移动运维能力。
增值版-旗舰版	包括企业版全部功能,并支持自定义产品控制台、自定义Logo、自定义皮肤,提 供元数据回流服务,让您快速创造属于自己的数据中台。

您可以观看DataWorks增值版本详解,对DataWorks增值版本进行详细了解。

各版本功能支持详情

模块

模块	功能点	基础版	标准版	专业版	企业版	旗舰版	部署地域 (Region)
基础	计算引擎实例 数量 ? 说明 工作空间 中可绑定 引擎数量 的限制。	仅支持绑 定1个计 算引擎实 例。	仅支持绑 定1个计 算引擎实 例。	实引只定其型引支 时擎允个个同计最绑 人个同计最绑 5个。	实引只定其型引定数 时擎允个个同计可意 人人同计可意 人。	实引只定有 实引实了。 实引学 分 个 人 同 计 可 第 任 。 数 。 次 令 、 、 、 、 、 、 、 、 、 、 、 、 、 、 、 、 、 、	全部 DataWorks部 署地域
	计算引擎实例 类型	不支持绑 定 Analytic DB for Postgre SQL、 Analytic DB for MySQL计 算引擎实 例。	无限制	无限制	无限制	无限制	
	换肤	支持	支持	支持	支持	支持	
	GetProject	不支持	不支持	不支持	支持	支持	
	自定义角色, 详情请参见 <mark>角</mark> 色及成员管 理:空间级。	不支持	不支持	不支持	支持	支持	
	实时同步	支持	支持	支持	支持	支持	华东1(杭州) 华东2(上海) 华北2(北京) 华北3(张家 口) 华市1(深圳) 西南1(成都)
数据集成							

> 文档版本: 20220712

模块	功能点	基础版	标准版	专业版	企业版	旗舰版	部署地域 (Region)
	离线同步	支持	支持	支持	支持	支持	
	整库迁移和批 量上云	支持	支持	支持	支持	支持	全部 Dat aWorks部
	独享数据集成 资源	支持	支持	支持	支持	支持	署地域 华东2(上 海)金融云
	同步资源监控	不支持	规划中	规划中	规划中		华北2(政务 云)
	自定义资源组	不支持	不支持	支持(限 定个数 <=5)	支持	规划中	
	ODPS SQL	支持	支持	支持	支持	支持	
	代码搜索	不支持	支持	支持	支持	支持	
	多人协作	支持	支持	支持	支持	支持	
	控制节 点:do- while、for- each、归并节 点、分支节 点和赋值节点	不支持	支持	支持	支持	支持	
	独享调度资源	支持	支持	支持	支持	支持	
	业务流程	支持	支持	支持	支持	支持	
	手动业务流程	支持	支持	支持	支持	支持	
	发布管理	支持	支持	支持	支持	支持	全部 DataWorks部
	组件管理	不支持	支持	支持	支持	支持	署地域
	MaxCompute 函数	支持	支持	支持	支持	支持	
DataStudio					2		

模块	功能点	基础版	标准版	专业版	企业版	旗舰版	部署地域 (Region)
	自定义调度资 源组	不支持	不支持	不支持	支持	支持	华东1(杭 州) 华东2(上 海) 华京) 华京) 华口) 华山) 年圳) 西都) 中国(香港)
	自定义节点	不支持	不支持	不支持	支持	支持	全部 DataWorks部 署地域
	通过向导模式 生成API	支持	支持	支持	支持	支持	
	通过脚本模式 生成API	支持	支持	支持	支持	支持	
	注册API	支持	支持	支持	支持	支持	
	查看数据表	支持	支持	支持	支持	支持	
数据服务	数据类型	基础关系 型数据 库。 具体如 下: PolarDB- X (DRDS)、 Analytic DB for MySQL 2.0 (AD S)、 MySQL 2.0 (AD S)、 MySQL Server、 Postgre SQL和 Oracle。	 关系型 系据素系 型库 Table Store (OTS) 和 Mono goDB 。 Holog res 	 关系型 系据系 非型库 Table Store (OTS))和 Mono goDB 。 Holog res 	 关系型 系据素系 型库 Table Store (OTS) 和 Mono goDB 。 Holog res 	 关系型 数据库 非型数 Table Store (OTS))和 Mono goDB 。 Holog res 	全部 DataWorks部 署地域

模块	功能点	基础版	标准版	专业版	企业版	旗舰版	部署地域 (Region)
	API授权	支持	支持	支持	支持	支持	
	API调用SDK	支持	支持	支持	支持	支持	
	过滤器	不支持	不支持	暂不支持 新建 <i>,</i> 改 造中	暂不支持 新建 <i>,</i> 改 造中	暂不支持 新建 <i>,</i> 改 造中	
	函数	不支持	不支持	暂不支持 新建,改 造中	暂不支持 新建,改 造中	暂不支持 新建,改 造中	华东2(上 海)
	服务编排	不支持	不支持	不支持	暂不支持 新建,改 造中	暂不支持 新建,改 造中	
	质量总览	支持	支持	支持	支持	支持	
数据质量	按表配置监控 规则	支持	支持	支持	支持	支持	全部 Dat <i>a</i> Works部 署地域
	动态阈值	不支持	不支持	不支持	支持	支持	
	自定义数据质 量报告	不支持	不支持	不支持	支持	支持	
	规则模板库	不支持	不支持	不支持	支持	支持	
	周期任务	支持	支持	支持	支持	支持	
	报警信息	支持	支持	支持	支持		
	基线预警	不支持	支持	支持	支持	支持	
	事件告警	不支持	支持	支持	支持	支持	全部
运维中心	成环报警	支持	支持	支持	支持	支持	ーロー DataWorks部 署地域
	孤立节点报警	支持	支持	支持	支持	支持	
	自定义规则	支持	支持	支持	支持	支持	
	智能诊断	不支持	不支持	支持	支持	支持	
	值班表	不支持	不支持	支持	支持	支持	
	搜索数据-全 部对象	支持	支持	支持	支持	支持	
	工作空间详情	支持	支持	支持	支持	支持	

DataWorks公共云合集·产品版本规格 与资源组

模块	功能点	基础版	标准版	专业版	企业版	旗舰版	部署地域 (Region)
	引擎详情- MaxCompute project	支持	支持	支持	支持	支持	
	表详情-基础 信息:权限、 业务、技术信 息	支持	支持	支持	支持	支持	
	表详情-明细 信息:字段、 分区、变更	支持	支持	支持	支持	支持	
	表详情-产出 信息:耗时、 任务	支持	支持	支持	支持	支持	
	表详情-血缘 信息:表	不支持	支持	支持	支持	支持	
*** +모 나나 도기	表详情-使用 记录	支持	支持	支持	支持	支持	全部
数据 地图	表详情-使用 说明	支持	支持	支持	支持	支持	DataWorks部 署地域
	表详情-数据 预览	支持	支持	支持	支持	支持	
	类目管理	支持	支持	支持	支持	支持	
	个人资产-收 藏功能	支持	支持	支持	支持	支持	
	我的数据	支持	支持	支持	支持	支持	

Dat aWorks

模块	功能点	基础版	标准版	专业版	企业版	旗舰版	部署地域 (Region)
安全中心	 平台安全诊断 数据访问控制 数据使用诊断 敏感数据管理 	支持	支持	支持	支持	支持	华京) 华东2(上 东2(上 海) 华州) 华川) 本和 (杭 州) 年川 (茶 (〇 〇 〇 〇 〇 〇 〇 〇 〇 〇 〇 〇 〇 〇 〇 〇 〇
	计算引擎审批 策略	不支持	不支持	不支持	支持	支持	华东1(杭 州)
	数据服务审批 策略	不支持	不支持	不支持	支持	支持	华东2(上 海)
审批中心	数据集成任务 审批策略	不支持	不支持	不支持	支持	支持	华北2(北 京) 华北3(张家 口) 华南1(深 圳) 西南1(成 都) 新加坡
	分级管理	不支持	支持	支持	支持	支持	
	字段扫描	不支持	支持	支持	支持	支持	
	数据访问记录	不支持	支持	支持	支持	支持	
	资产大盘	不支持	支持	支持	支持	支持	华东1(杭
	自定义内容扫 描	不支持	不支持	支持	支持	支持	州) 华东2(上 海)
	风险识别大盘	不支持	不支持	支持	支持	支持	// 华北2(北
数据保护伞	动态脱敏	不支持	不支持	部分支持	支持	支持	京)

	自生成数据识						而志1 (成
	別模型	不支持	不支持	不支持	支持	支持	都) 新加坡
	内置扫描模型	不支持	不支持	不支持	支持	支持	
	创建并管理样 本库	不支持	不支持	不支持	支持	支持	
	文件管理	支持	支持	支持	支持	支持	
	导入功能	支持	支持	支持	支持	支持	
	表格编辑功能	支持	支持	支持	支持	支持	
	透视功能	表格数据	MaxCom pute表、 数据服务 和MySQL 数据源。	支持	支持	支持	全部
数据分析	图表功能	5个	23个	大于等于 27个	大于等于 27个	大于等于 27个	DataWorks部 署地域
	分享功能	不支持	分享给指 定人,仅 具有只读 权限,分 享数 <=10。	分享给指 定人, 读 有只读 4 编辑权 限, 分享 数 <=50。	分享给指 定人或所 有人,读 有只读或 编辑权 限,分享 数 <=100。	支持	
	SQL查询下载 功能	不支持	支持下载, 万行。				
	任务上云	支持	支持	支持	支持	支持	
							A #2

迁移助手 模块	功能点	基础版	标准版	专业版	企业版	旗舰版	DataWorks部 習者地域 署他域 (Region)		
	DataWorks迁 移	 租提10号的机 导大30号动发不户供次出试会 入小MB 动发不30MB 目交:持 	 租提10号的机 导大30号动发不户供次出试会 入小MB 动发不30MB 百交:持 	• 租提10导的机 导大无 导动发支户供次出试会 入小上 入提布持内 包用。 包:限 自交:	• 租免出数上 导大无 导动发支户费包:限 入小上 入提布持内导个无 包:限 自交:	• 租免出数上 导大无 导动发支户费包:限 入小上 入提布持内导个无 包:限 自交:			
	数仓规划	-				您需要开通			
DataWorks	数据标准	DataWorks	DataWorks的增值版本支持开通使用智能数据建模。智能数据建能数据建						
智能数据建 模	维度建模	健使用包牛包月计费模式,根据您在智能数据建模产品内创建的 后,才可以 对象数量进行计费。您可以根据实际需求,选购智能数据建模产 用该产品功 品的不同如格							
	数据指标	ער היארים איזיאיי	Ι Π΄ ο				能,叶间谓参 见智能数据建 模计费标准。		
	OpenAPI	 可用版本 可用地域 华北3(3 (香港) (孟买) 京)、澳 	: DataWorks : 华东1(杭 朱家口)、北京 、亚太东南1 、美国(弗吉 大利亚(悉尼	企业版和旗舰 、华东2(京政务云、华系 (新加坡)、E 尼亚)和德国)、马来西亚	版,详情请参 上海)、华南 F2(上海)金 D度尼西亚(雅 (法兰克福)、 (吉隆坡)。	见 <mark>DataWorks[;]</mark> 1(深圳)、华 融云、西南1([加达)、美国 、英国(伦敦)	各版本详解。 注12(北京)、 成都)、中国 (硅谷)、印度 、日本(东		
	OpenEvent概 述	 目前OpenEvent面向DataWorks企业版和旗舰版用户公测,华北2(北京1(杭州)、华南1(深圳)、华东2(上海)、华北3(张家口)、华南1圳)、西南1(成都)、美国(硅谷)、美国(弗吉尼亚)、德国(法兰日本(东京)、中国(香港)、新加坡地域支持该功能。 OpenEvent通过EventBridge实现消息的订阅与消费,所以您需要先开通EventBridge产品服务,关于EventBridge产品介绍详情请参见:什么是事EventBridge。 仅支持阿里云主账号或者有AliyunDataWorksFullAccess权限的RAM用户户可以直接进行界面相关配置,关于DataWorks权限说明,详情请参见指分权限管控:自定义RAM Policy。 							
开放平台		J							

模块	功能点	基础版	标准版	专业版	企业版	旗舰版	部署地域 (Region)
	扩展程序概述	 目前仅面 州)、华 国(硅 港)、新 仅支持以直 详情: 纽 	i向DataWorks 东2(上海)、)、美国(弗 加坡地域支持 I里云主账号或 接进行界面相 分RAM Policy	企业版和旗舰 华北3(张家 吉尼亚)、德 该功能。 者有AliyunDa 关配置,关于 。	版用户公测, 口)、华南1 国(法兰克福 taWorksFullA Dat <i>a</i> Works权	华北2(北京) (深圳)、西南)、日本(东京 ccess权限的R 限说明 <i>,</i> 详情计	、华东1(杭 引(成都)、美 京)、中国(香 AM用户权限的用 请参见 <mark>管控权限</mark>

1.1.2. DataWorks增值版本到期降级功能变化清单

您购买使用DataWorks增值版本后,如果因为没有及时续费等一些原因导致增值版本DataWorks到期,则到期 后增值版DataWorks会降级为基础版DataWorks,对您使用的DataWorks功能可能会有影响。本文为您介绍 DataWorks增值版本到期降级后功能变化清单。

模块	功能点	到期降级变化		
数据集成	自定义资源组	不可用,且使用自定义资源组的任务调度将置为 失败。		
	全文检索	功能停用。		
	自定义节点	功能停用、关闭节点入口、已有自定义节点运行 失败。		
DataStudio	组件管理	功能停用。		
	流程控制	功能停用、任务失败。		
	高级API	功能停用、请求接口报错。		
MaxCompute配置包	SQL组件节点	功能停用、关闭节点入口。		
	DAG开发和SQL&DAG互转	功能停用、仅能以只读方式查看DAG、不可编 辑,但可以转换为SQL。		
	DAG组件组	功能停用。		
StreamStudio	执行计划图形化	功能停用、仅能以只读方式查看。		
	本地调试 (实时数据预览)	市坐信田		
	高级运维-任务智能诊断	功能停用。		
	高级API	功能停用、请求接口报错。		
	数据类型	功能停用、已有接口调用报错。		
	过滤器	功能停用、过滤器失效、调用报错。		

	功能点	到期降级变化			
	Lambda	功能停用、Lambda可见但不可编辑和调用。			
	服务编排	功能停用、工作流可见但不可编辑和调用。			
	高级API	功能停用、调用报错。			
数据质量	动态阈值	功能停用、无法创建新的动态阈值规则,已创建 的规则不受影响。			
	高级API	功能停用、请求接口报错。			
	基线预警	无法进入基线管理、基线实例页面,自定义规则 无法选择基线规则,不产生基线实例。			
运维中心	事件报警	无法进入基线管理、基线实例页面,自定义规则 无法选择基线规则,产生基线实例。			
医维干心	高级API	功能停用、请求接口报错。			
	值班表	DataWorks会为您保留原有的值班表数据。当重 新升级到专业版或更高版本后,原有的值班表功 能可以正常使用。			
	详情-血缘信息:表、字段、影响分 析	功能停用、相关信息提示需升级到高版本方可使 用。			
数据地图	表详情-字段变更:影响分析、邮件 通知	功能停用、已有邮件发送记录仍存在但不可继续 使用。原功能点不可用,并提示您购买高版本。			
	高级API	功能停用、请求接口报错。			
	多引擎	功能停用、不显示已有的多引擎信息。			
数据保护伞	所有功能	功能停用、无法进入数据保护伞页面。			
	透视功能				
实时分析	图表功能	功能显示为灰色不可田			
	分享功能				
	维表编辑: MaxCompute表				
	数据存储优化-MaxCompute表优化				
	数据计算优化-MaxCompute任务优 化				
资源优化	数据采集优化-同步任务优化	功能停用,无法进入资源优化页面。			
	管理员视角资源优化				
	管理白名单				

模块	功能点	到期降级变化
	DataWorks迁移,租户内免费导出包 个数	降低为10个。
迁移助手	DataWorks导入包大小	减少为30M。
	DataWorks导入时自动提交、自动发 布	将不再支持自动提交、发布。

1.2. DataWorks资源组

1.2.1. DataWorks资源组概述

DataWorks支持公共资源组、独享资源组和自定义资源组,本文为您介绍各类资源组属性说明与选择指导。

背景信息

- 开通DataWorks时,DataWorks默认为您提供按量付费的公共资源组,同时,支持您购买包年包月独享资源 组,或者升级DataWorks增值版本,将您的IDC机器作为DataWorks任务运行的相关资源组,详情请参 见:资源组介绍。
- 不同类型的资源组用于任务不同的执行阶段,本文以离线任务调度为例,为您介绍DataWorks离线任务下发机制,与任务运行过程中使用的资源组。详情请参见:任务下发机制。
- 任务同一个执行阶段可选择不同类型的资源组,例如,离线任务调度可选择包年包月独享调度资源组,或 按量付费公共调度资源组,数据集成离线任务执行可选择包年包月独享数据集成资源组,或DataWorks增值 版本支持的自定义数据集成资源组,关于资源组选择详情请参见:资源组选择。

费用说明

各类资源组的费用说明详情请参见: 计费逻辑说明。

使用限制

仅阿里云主账号和同时拥有AliyunBSSOrderAccess和AliyunDataWorksFullAccess策略的RAM用户可以进行资源组购买操作。

资源组介绍

DataWorks包括独享资源组、公共资源组和自定义资源组,各资源组根据任务调度、数据集成、数据服务等不同的使用场景,细分为不同类型的资源组,包括:独享调度资源组、独享数据集成资源组、独享数据服务资源 组、公共调度资源组、公共数据服务资源组、自定义调度资源组、自定义数据集成资源组。其中,DataWorks 的调度资源组用于保障任务的调度运行,数据集成资源组用于保障数据集成任务中的数据传输。数据服务资 源组用于保障API的调用。。

资源组类 计费方式 资源组描述 3 型 4	调度资源	数据集成	数据服务
	组	资源组	资源组

DataWorks公共云合集·产品版本规格 与资源组

Dat aWorks

资源组类 型	计费方式	资源组描述	调度资源 组	数据集成 资源组	数据服务 资源组
独享资源 组	包年包月	 属于平台托管类型的资源组,购买后将 拥有该资源组的独立使用权。 支持通过资源组绑定工作空间的操作, 实现各个工作空间的资源隔离。 支持扩容、缩容、升配或规格变更等灵 活配置。 	独享调度 资源组	独享数据 集成资源 组	独享数据 服务资源 组
公共资源 组	按量付费	开通DataWorks时,阿里云DataWorks会 自动为用户提供公共资源组(公共数据服 务资源组和公共调度资源组),公共资源 组为DataWorks租户共享,任务运行高峰 期可能出现等待资源的情况。	公共调度 资源组	-	公共数据 服务资源 组
		如果您有富余的服务器资源,也可将此部 分资源作为Dat <i>a</i> Works中任务运行的资源 组使用。			
自定义资 源组	-	⑦ 说明 DataWorks增值版本才可以使用自定义资源组。自定义调度资源组和自定义数据集成资源组要求的版本不同。详情请参见:DataWorks各版本详解。	自定义调 度资源组	自定义数 据集成资 源组	-

任务下发机制

DataWorks的离线任务通过调度资源组下发至各引擎或服务器资源上执行。例如,DataWorks通过调度资源组将任务下发至MaxCompute引擎上执行。数据集成离线同步任务通过调度资源组将任务下发至数据集成任务执行资源上执行。



资源组选择

以下通过机器资源归属、网络、任务执行时效性要求、适用场景、收费方式等维度为您对比展示各资源组的现 状。方便您根据自身需求选择合适的资源组。

● 调度资源组

类别	公共调度资源组	独享调度资源组	自定义调度资源 组
机器资 源归属	由DataWorks维护,是所有租户竞争使 用的共享资源。	由DataWorks维护,是每个租户独占使 用的计算资源。	由您自己维护, 是属于您的IDC机 器。
网络	支持阿里云内部数据源之间的访问连 通,数据源处于以下网络环境时,不支 持使用: • 数据源处于公网环境中,但数据源有 白名单设置,限制了未知IP的访问。 • 数据源处于阿里云VPC网络环境中。	可通过网络配置连通复杂的网络环境。 详情请参见: <mark>独享调度资源组概述</mark> 。	根据您自身机器 所处的环境决 定。
任务执 行时效 性要求	DataWorks用户共享,无法保障时效 性。	租户独占,可以控制资源组执行的并发 数,并保障任务按时执行。	根据您自身机器 所处的环境决 定。
适用场 景	由于租户共同拥有资源的使用权,所以 任务量多时,容易造成空间内任务阻 塞。仅适用于低并发、低频率的调用场 景。	租户可以控制资源组执行的并发数,并 支持对资源组进行扩缩容,规格变更等 操作,满足任务量大,且对任务时效性 要求高的场景。	自定义开发场 景。
收费	按量付费	包年包月	DataWorks版本 按月使用收费。
推荐指 数	*	****	*
选择指 引	 • 时效性要求:公共调度资源组所有阿里:时效要求,建议您选择独享调度资源组, • 访问非公网且有白名单限制的环境需求。 • 公 任务量需求:日调度任务较多时,您可 	云DataWorks用户共享,如果您的任务有 。 : 公共资源组不支持访问非公网并且有白 建议您选择使用独享调度资源组。 以选择包年包月独享调度资源组。	

• 数据集成资源组

类别	独享数据集成资源组	自定义数据集成资源组
机器资源归属	由DataWorks维护,是每个租户独占使用的 计算资源。	由您自己维护,是属于您的IDC机器。
网络	 拥有公网访问能力。 通过网络配置可以访问任意网络环境下的 数据库。	根据您自身机器所处的环境决定。

类别	独享数据集成资源组	自定义数据集成资源组
支持的数据源	全部数据源。	全部数据源。
任务执行时效性要 求	租户独占,可以控制资源组执行的并发数, 并保障任务按时执行。	根据您自身机器所处的环境决定。
适用场景	大量、重要的生产任务。	 如果您已有计算资源,可以对接阿里云重复使用,无需重新购买。 需要同步的数据源全部在IDC内。
收费	包年包月	Dat aWorks版本按月使用收费。
推荐指数	****	*
选择指引	 • 时效性要求:在数据集成任务高并发执行且无法错峰运行的情况下,企业需要独享的计算资源来保障数据快速、稳定地传输,此时您可选择使用独享数据集成资源组。 • 访问内网环境的需求:独享数据集成资源组拥有公网访问能力,如果您的数据库在公网内,独享数据集成资源组支持您走公网进行同步。部分数据库由于其自身特性,无法通过公网直接同步,具体详情请参见各个数据源与插件说明文档:支持的数据源与读写插件。 • 访问复杂环境的需求:当需要访问复杂网络环境下的数据库时,请选择包年包月独享数据集成资源组。关于网络解决方案详情请参见:配置资源组与网络连通。 • 实时同步需求:当进行数据实时同步时,您需要使用独享数据集成资源组。 	

● 数据服务资源组

公共数据服务资源组为租户共享资源。在对数据服务API进行高并发、高频率的接口调用并需要及时返回数据结果的情况下,企业需要专有的计算资源来保障数据服务的可用性和稳定性。此种场景下建议您选择 DataWorks的独享数据服务资源组。

1.2.2. 公共资源组

1.2.2.1. 公共资源组概述

开通DataWorks时, DataWorks自动为您提供公共资源组(包括公共调度资源组、公共数据集成(调试)资源 组、公共数据服务资源组),可用于数据开发、运行测试等操作。公共资源组被多个租户共享,高峰期租户间 可能会存在抢占资源的情况。本文为您介绍公共资源组的概要信息。

适用场景

公共资源组建议在任务量相对较少、对任务时效性要求不高的场景下使用。

使用限制

公共资源组由多个租户共享,在高峰期无法保证稳定的资源分配。

? 说明

- 公共资调度源组最多支持40个调度任务同时运行,但在任务执行高峰期(通常为00:00~09:00)会存在资源抢占情况,公共资调度源组的并发任务数可能会小于该数值。
- 公共数据服务资源组无法满足高频率、高并发的调用。

如果您希望保障任务运行资源独立且充足,建议开通使用独享资源组,具体如下表。

资源组类型	描述	参考链接
独享调度资源组	在任务高并发执行且无法错峰运行的情况下,需要专有的 计算资源来保障任务定时被调度运行时,建议使用 DataWorks的独享调度资源组。	独享调度资源组计费说 明:包年包月
独享数据集成资源组	在数据集成任务高并发执行且无法错峰运行的情况下,企 业需要专有的计算资源组来保障数据快速、稳定地传输 时,建议使用DataWorks的独享数据集成资源组。	独享数据集成资源组计费 说明:包年包月
独享数据服务资源组	在调用数据服务API需要高QPS和SLA保障的情况下,企业 需要保障调用成功。同时,满足高并发、高频率的接口调 用并及时返回结果数据时,建议使用DataWorks的独享数 据服务资源。	独享数据服务资源组计费 说明:包年包月

计费与相关操作

(1) 计费说明

⑦ 说明 购买DataWorks后, DataWorks会为您提供公共资源组,无需您单独购买。

根据公共资源组实例费用、同步进程费用等计费项收费。公共资源组开通时支持按量付费的计费模式,计费详 情可参见:

- 公共调度资源组计费说明: 按量付费
- 公共数据集成(调试)资源组:按量计费
- 公共数据服务资源组计费说明: 按量付费

(2) 扣费与欠费说明

DataWorks针对不同类型公共资源组的扣费及欠费处理情况存在差异,详情请参见扣费与欠费说明。

使用公共资源组

了解上述公共资源组的适用场景及计费详情后,您可以在数据集成、数据开发等任务中选择合适类型的资源组 使用。具体使用说明,详情请参见使用公共资源组。

网络连通方案

DataWorks资源组本质上为一组阿里云ECS实例,在运行数据集成、数据开发等任务时,需保障资源组与数据 源之间的网络是连通的,且不会因为白名单等特殊安全访问设置阻断资源组与数据源之间的网络连通性。

● 网络连通

公共资源组支持与阿里云内部数据源之间的访问连通,数据源所处的网络环境及其连通性支持情况如下。

公共调度资源组

- 若要访问公网地址,需要在工作空间配置界面配置沙箱白名单,配置沙箱白名单后仍无法访问,请使用独享调度资源组。
- 仅支持访问无白名单访问控制的地址。访问有白名单访问控制的地址,或访问VPC地址,需要使用独享 调度资源组。

⑦ 说明 建议所有访问外部地址的场景,均使用独享调度资源组。使用独享调度资源组,详情请参见<u>独享调度资源组概述</u>。

公共数据服务资源组

数据源所处的网络环境与资源组的连通性支持情况如下。

数据源网络环境	是否支持访问
数据源处于公网环境	支持
数据源处于经典网络	支持
数据源处于VPC网络	不支持

• 白名单设置

公共调度资源组为您提供任务安全沙箱功能,限制未知IP访问资源组。如果您的业务场景涉及访问资源组, 需要将访问资源组的IP添加至安全沙箱白名单中。详细操作请参见安全设置。

1.2.2.2. 使用公共资源组

DataWorks为您提供公共资源组,以便您直接创建和运行任务。

由DataWorks自身维护和创建的资源组即公共资源组,说明如下:

- 公共资源组是一个公共资源池,不保证任务可以及时得到资源并被调度执行。如果您需要任务得到高保证 的运行,请使用独享资源组,详情请参见新增和使用独享数据集成资源组、新增和使用独享调度资源组。
- 打通网络时,如果您需要公共资源组机器的相应信息,请参见场景示例: ECS自建数据库的安全组配置和添加白名单。

? 说明

- 在保证其它任务没有使用公共资源组的前提下,一个数据集成任务的公共资源组最多支持25个并发。在实际执行任务时,资源可能会被其它任务抢占,所以最大并发数未必达到25个并发,并发数以实际运行情况为准。
- 公共资源组不支持调整内存大小,您可以调整并发数。

内存的计算方式为: 并发数*512MB。

1.2.3. 独享数据集成资源组

1.2.3.1. 独享数据集成资源组概述

在数据集成任务高并发执行且无法错峰运行的情况下,需要专有的计算资源组来保障数据快速、稳定的传输时,建议您选择使用DataWorks的独享数据集成资源组。本文为您概要介绍独享数据集成资源组。

适用场景

- 数据集成任务量大、且对任务时效性要求高。
- 需要灵活调整资源组大小。
- 需要访问公网地址、VPC、IDC。
- 需要通过IP白名单进行访问控制。

? 说明

独享数据集成资源组保障的是数据同步实例可以同时运行的总线程数量,而非可以同时运行的数据同步实例本身。如果您需要保障可以同时运行的数据同步实例的数量,请购买独享调度资源,详情请参见购买指 引。

使用限制

- 独享数据集成资源组的计费方式为包年包月,在资源组到期前不支持提前删除、释放资源组,资源组到期 后会按照规定的时间停服、释放。
- 独享数据集成资源组不支持跨地域使用。例如,华东2(上海)地域的独享数据集成资源组,只能给华东2(上海)地域的工作空间使用。
- 独享数据集成资源组不支持访问阿里云经典网络。如果您的数据源处于经典网络环境中,建议您将数据源 迁移至专有网络中。
- 同一独享资源组下同规格资源数购买上限为20个,若需要更多,请提交工单联系我们升级。

性能指标

规格	离线同步最大并发线 程数	单表实时同步最大任 务数	整库多表实时同步最 大任务数	分库分表实时同步最 大任务数
4c8g	8	3	3	不支持
8c16g	16	6	6	1
12c24g	24	9	9	1
16c32g	32	12	12	2
24c48g	48	18	18	3

? 说明

- 单表和整库多表实时同步最大任务:此数量仅做参考,因为实时同步任务个体配置差异较大,实际执行时不一定可以达到最大的任务数。具体请根据实际执行后任务资源占用情况来做判断。
- 分库分表实时同步最大任务:当实时同步任务的源端配置了2个或2个以上数据库中的表,则至少 需要13个并发才可以运行起来。即至少需要8c16g规格的资源组。

规格选择及费用预估

您可以使用DataWorks独享资源组计算器预估所需规格及相关费用。预估规格及相关费用时,可进入 DataWorks的运维中心 > 运维大屏,查看资源运行的实例数、并发情况等信息作为参考。查看运维大屏,详 情请参见查看运维大屏。

计费与相关操作

(1) 计费说明

独享数据集成资源组的计费方式为包年包月,您可以根据自身需求选择合适的规格进行购买,不同规格的计费 详情请参见:<u>独享数据集成资源组计费说明:包年包月</u>。

(2) 扩容与缩容

您可以根据业务需求选择购买独享资源组,指定购买资源组内机器规格、机器数量,购买完成后,您可以对资 源组进行扩容或缩容操作,扩缩容操作步骤和注意事项请参见扩缩容操作。

(3) 规格变更(升降配)

如果当前独享资源组规格不符合实际需要,您可以针对该资源组进行规格变更。变更当前资源组的规格,资源 组内所有机器的规格将统一变更,规格变更操步骤和注意事项请参见变更规格操作。

(4) 续费、停服与释放资源组

您可以在独享资源组临近到期时进行续费操作,或等待资源组到期后停服并自动释放资源。详情请参见<mark>到期说</mark>明及续费操作。

使用独享数据集成资源组

了解上述独享数据集成资源的规格计费详情后,您可以选择购买合适的独享数据集成资源组并在数据集成任务 中使用。从购买至使用,您需进行的主要操作流程如下:

- 1. 购买独享资源组(创建订单)。
- 2. 基于订单创建资源组。
- 3. 绑定归属工作空间。
- 4. 为独享资源组绑定专有网络。
- 5. 添加白名单。
- 6. 应用示例: 独享数据集成资源组。

操作详情请参见:新增和使用独享数据集成资源组。

网络连通方案

独享数据集成资源组与其他独享资源组类似,本质上为一组阿里云ECS实例,在数据集成任务运行时,需保障 资源组与数据源之间的网络是连通的,且不会因为白名单等特殊安全访问设置阻断资源组与数据源之间的网络 连通性。

独享数据集成资源组购买后,需绑定一个VPC网络,后续您可以根据数据源所在的网络环境情况来选择网络连通方案。详情请参见:配置资源组与网络连通。

• 网络连通方案

数据源网络环境	网络连通方案
数据源处于公网环境中	数据源与独享资源组绑定的VPC网络直连即可连通。
数据源处于VPC网络中,且与独享资 源组在 同一 地域	建议独享资源组与数据源绑定同一VPC,绑定后系统自动为您添加一条路由至 整个VPC的路由,您绑定VPC的任意一个交换机,独享资源组与数据源的网络 都可连通。
数据源处于VPC网络中,且与独享资 源组在 不同 地域	使用高速通道或VPN等将网络连通数据源与独享资源组绑定的VPC网络,并添 加一条路由指向目标数据库IP,保障两个网络间网络可达。

数据 源网络 坏 境	网络连通万案

数据源处于IDC网络中	使用高速通道或VPN等将网络连通数据源与独享资源组绑定的VPC网络,并添 加一条路由指向目标数据库IP,保障两个网络间网络可达。
数据源处于经典网络中	独享资源组处于阿里云VPC网络,如果数据源处于经典网络环境中,则数据源 与独享资源组间网络无法连通,建议您将数据源迁移至VPC网络中。

• 白名单配置

如果数据源设置有白名单限制访问地址时,您需要将独享资源组的访问地址添加至数据源的访问白名单 中,避免访问受限。详情可参见使用独享数据集成资源组执行任务需要在数据库添加的IP白名单。

1.2.3.2. 新增和使用独享数据集成资源组

您可以根据业务场景需求选择购买合适规格的独享数据集成资源组,用于保障数据快速、稳定的传输。购买独 享资源组后,在实际使用前您还需进行网络绑定、白名单等配置操作。本文为您介绍从购买至使用独享数据集 成资源组的全流程与步骤详情。

前提条件

购买和使用独享数据集成资源组前:

你需要先了解独享数据集成资源组的规格性能(不同规格能同时跑多少任务)、计费类型(不同规格怎么收费)等细节,提前根据业务场景规划需购买的资源组规格和购买时长,详情请参见:独享数据集成资源组计费说明:包年包月。

⑦ 说明 您可以使用DataWorks独享资源组计算器预估所需规格及费用。预估规格及费用时,可进入 DataWorks的运维中心 > 运维大屏,查看资源运行的实例数、并发情况等信息作为参考。查看运维大 屏,详情请参见查看运维大屏。

由于执行数据同步任务流程的要求,需要保证资源组可以访问数据源(来源数据源和目标数据源)所在的网络,且不会因白名单限制等原因不能访问数据源。因此,您需要在保证网络连通的前提下,再使用数据集成,如果网络不通,会导致数据同步任务无法执行。独享数据集成资源组与数据源间的网络连通方案和白名单注意事项,详情请参见:独享数据集成资源组概述。

操作流程

从购买至使用独享数据集成资源组,您需进行的主要操作流程如下:

- 1. 购买资源组(创建订单)
- 2. 基于订单创建资源组
- 3. 绑定归属工作空间
- 4. 绑定专有网络
- 5. 添加白名单
- 6. 测试网络联通
- 7. 使用独享数据集成资源组

购买资源组(创建订单)

- 1. 登录阿里云账号并进入DataWorks独享资源组购买页面。
- 2. 在购买页面根据业务需要配置相应的参数,配置要点如下。
 - **地域和可用区**:选择后续需要使用此独享资源组的地域。

⑦ 说明 独享数据集成资源组不支持跨地域使用。例如,华东2(上海)地域的独享资源,只能 给华东2(上海)地域的工作空间使用。

○ 独享资源类型:选择为独享数据集成资源组。

其他配置项,例如计费周期等参数,可根据业务规划自行配置即可。

⑦ 说明 同一独享资源组下同规格资源数购买上限为20,若需要更多,请提交工单联系我们升级。

3. 单击**立即购买**,根据界面提示进行付款等操作后即可完成购买独享数据集成资源组。

完成购买后,您可以在我的订单中查看到此订单及订单号详情。

三 (-) 阿里云	✿ 工作台 ♥ 华东2 (上海) ∨	Q 搜索	费用 工单 ICP 备案 企业 支持 App
DataWorks	DataWorks / 资源组列表	充值	
概览页	资源组列表	订单	
工作空间列表	如直姿隔组 小土漆福组	费用账单	
资源组列表		续费管理	
报警配置	创建调度资源组 创建集成资源组 创建服务资源组 请指 ^	約入搜索关键词 Q 用户中心	

后续您可根据订单号将此次购买的独享数据集成资源组添加到DataWorks控制台。

基于订单创建资源组

- 1. 登录DataWorks控制台。
- 2. 单击左侧导航栏资源组列表,在独享资源组页签下单击创建集成资源组页面,配置各项参数。

创建独享资源组	\times
① 独字调度资源、独字数据集成资源用途不同,详情参考:帮助文档	
资源组类型: ○ 独享调度资源组 ● 独享数据集成资源组	
* 资源组名称:	
请输入资源组名称	
* 资源组备注:	
请输入资源组备注	
*订单号: 购买 当前地域:华东2(上海),请购买此地域的资源组	
请选择订单号	\sim

参数	描述
资源组类型	选择添加的独享资源组类型。 独享资源组包括 独享调度资源组 和 独享数据集成资源组 ,分别适用于通用任务调度 和数据同步任务专用。
	设置资源组的名称,此名称需在租户内唯一,请避免重复,否则会导致在后续进行确 定操作时报错。
资源组名称	 ⑦ 说明 租户即主账号,一个租户(主账号)下可以有多个用户(子账号)。
资源组备注	对资源组进行简单描述。
	在下拉框中选择已购买的独享资源组的订单号。
订单号	⑦ 说明 如果还需要继续购买独享资源组,也可单击订单号后的 购买 继续购买,完成购买后回到此页面再将新购的独享资源组添加至DataWorks控制台。

3. 单击**确定**, DataWorks即开始初始化此独享资源组,当资源组状态为运行中后,独享资源组即已添加至控制台。

⑦ 说明 独享资源组的初始化需要约20分钟,请您耐心等待其状态更新为运行中。

独享资源组添加至控制台后,您需要为资源组绑定工作空间,才可以在任务的配置中选择使用此独享资源组。

绑定归属工作空间

独享数据集成资源组需要绑定归属的工作空间,才可以在对应的工作空间下选择使用该独享资源组。一个独享 数据集成资源组可以分配给多个工作空间使用,但不可跨地域使用,例如,上海地域的独享数据集成资源组仅 可绑定上海地域的工作空间使用,不可绑定其他地域的工作空间。绑定归属工作空间的操作如下。

- 1. 登录DataWorks控制台。
- 2. 在资源组列表的独享资源组页签下,单击相应资源组后的修改归属工作空间。
- 3. 在修改归属工作空间单击对应工作空间后的绑定,即可将独享资源组绑定至此工作空间。

绑定专有网络

独享资源部署在DataWorks托管的专有网络(VPC)中,与其他网络环境不连通。使用独享资源时,您需进行 网络配置,为独享资源绑定一个可与数据源连通的VPC,通过此VPC实现与数据源网络连通。绑定专有网络的 操作如下。

注意 4c8g类型的独享数据集成资源组最多支持绑定2个专有网络,其他规格的独享数据集成资源组最多支持绑定3个专有网络。

- 1. 登录DataWorks控制台。
- 2. 在资源组列表的独享资源组页签下,单击相应资源组后的网络设置,进入专有网络绑定页面。
 - 绑定前,请首先使用阿里云主账号进行RAM授权(仅主账号有权限),让DataWorks拥有访问您的云资源

= (-)阿里云			Q 搜索			费用	工单	备案	企业	支持与服务	ş D_	Ų.	Ä	â	简体中文	0
			概览 工作的	2间列表 资	源列表	计算引擎列	表									
返回 资源组: 请选择	V														新增组	
资源名称 类型	1	可用区	专有网络		交换机			安全组			状态			操作		
				没有数据	89											
		无访问权	限					×								
		您尚未授权	DataWorks系统默认角色	,需要您 <mark>使用主账号</mark>	登录去RAM授	<mark>叹</mark> ,然后刷新页面	5. 刷新									
								关闭								
		_														
■ (-) 阿里云			Q 搜索				费	用工	单省	备案 企业	支持与	家服务	2	<u>,</u>	j⊟ į	简体中文
一云资源访问授权																
温馨提示:如需修	改角色权限,请前往	主RAM控制台角色管	理中设置,需要注意的	是,错误的配置可能	导致DataWork	s无法获取到必须	要的权限。							×		
DataWorks 下方是系统创建的 AliyunDa 报话: Data 权限描述: J	青求获取访问您 的可供DataWorks使 taWorksAcces Works默认使用此角 用于DataWorks角色	民 <mark>云资源的权限</mark> 用的角色,授权后, ssingENIRole 色来访问您在其他云 的授权策略, 包含弹	DataWorks拥有对您云 一一品中的资源 性网卡(ENI)的部分权限	资源相应的访问权限	ł.									~		
				同意授权	取消											

的权限。您可以通过<mark>云资源访问授权</mark>页面进行授权。也可以在主账号首次次进入管控后弹出的界面弹框中 进行授权。

3. 绑定专有网络VPC。

i. 单击**专有网络绑定**页面左上方的**新增绑定**,在**新增专有网络绑定**对话框中,配置各项参数,不同 网络环境下各参数的配置说明如下。

参数说明如下:

参数	配置说明(数据源与独享资源同 账号同地域)	配置说明(数据源与独享资源在不同账号或不同地 域)						
专有网络	如果您的数据源与独享资源组在 同一个阿里云账号下,建议配置 为数据源所在的VPC。 如果不在同一个阿里云账号下, 则与不在同一地域场景一致。	如果您的数据源与独享资源不在同一地域,例如,数 据源不在阿里云VPC网络环境中,您可单击创建专有 网络,为独享资源组创建一个VPC。创建完成后这里 配置为新建的VPC或选择已经与目标数据库网络打通 的VPC。						
可用区	选择数据库所在可用区。	选择已经与目标数据库网络联通的可用区。						
	专有网络配置为数据源所在VPC 时,建议选择与数据源绑定的交 换机。							
交换机	⑦ 说明 绑定数据源所在 VPC后,绑定VPC下任意一个 交换机,会自动添加路由至 整个VPC网段,实现独享数 据集成资源组在该VPC下网 络可达。	选择已经与目标数据库网络联通的交换机,若没有可 用交换机,可单击 创建交换机 为独享资源组创建交 换机。创建完成后这里配置为创建的交换机。						
安全组	安全组指定了独享资源实例需要允许或禁止哪些公网或内网的访问。您可根据业务需求选择已有的安全组,或单击 创建安全组 为独享资源实例创建安全组。创建安全组的详细参数 配置可参见 <mark>添加安全组规则</mark> 。							

ii. 单击**确定**,完成绑定VPC操作。

⑦ 说明 如果数据源和独享资源组不在同一个地域,或不在同一个阿里云账号下,则您需要绑定专有网络后,再添加路由规则指向目标数据库IP地址。

4. (可选) 配置Host。

如果您的数据源无法通过IP直接访问,例如,数据源通过Host的域名托管,通过Host域名直接被外部访问时,您需要参考以下步骤配置Host,否则在添加数据源时,使用数据源的Host域名进行连通性测试时会失败。

i. 单击**Host配置**,在Host配置页面左上方单击**新增**,在**新增域名Host配置**对话框中配置各项参数。 参数说明如下。

参数	配置说明						
IP地址	配置为数据源的实际IP地址。						
	配置为数据源对外提供访问服务的Host域名。如果有多个Host域名时,可换行依 次添加。						
Host域名	⑦ 说明 此处的域名需包含数字、字母、连字符(-)、点(.),且必须 以字母开头,以字母或者数字结尾。						

- ii. 如果有多个IP地址需要添加,可继续单击新增继续添加。
 - ? 说明
 - 新增的Host配置中, IP、域名不能与之前配置的Host中的IP或域名重复。
 - 一个Host配置中, IP与域名为1对多的关系, 即IP可以对应多个Host域名, 但是同一个 Host域名只能指向一个IP。
- 5. (可选) 配置DNS。

如果您的数据源无法通过IP直接访问,例如,数据源通过负载均衡域名直接被外部访问,通过内部域名解析服务器将域名解析至实际数据源IP地址时,您需要参考以下步骤配置DNS,否则在添加数据源时,使用数据源的Host域名进行连通性测试时会失败。

⑦ 说明 如果同一个域名同时配置了Host和DNS,系统会优先根据Host的访问规则访问数据源。

i. 单击**DNS配置**,在DNS配置页面左下角单击**添加**,配置完成DNS各项参数后单击**保存**。参数说明如下。

参数	配置说明
	非必配项。如果数据源有统一的一级域名,可在此处配置为数据源对外提供访问 的域名的一级域名。
Domain	例如,数据源1对外的访问或名为domain1.example.com,数据源2对外的访问 域名为domain2.example.com,此处建议配置为example.com。
	⑦ 说明 此处的域名需包含数字、字母、连字符(-)、点(.),且必须 以字母开头,以字母或者数字结尾。
NameServer	配置为数据源提供域名解析的解析服务的IP地址。如果有多个域名解析服务器时,可换行依次添加。

ii. 如果您需要修改之前配置的DNS,您可单击左下角的修改。

添加白名单

独享数据集成资源组和数据源同可用区时,如果同专有网络、交换机的情况下仍未连通网络,则可能是因为数据源有白名单限制。请添加以下内容至数据源的白名单列表中。

• 使用内网连通数据源与独享资源组:添加独享资源组绑定专有网络时的交换机网段至数据源白名单中。

您需要为独享资源组绑定专有网络后,进入DataWorks的**资源组列表 > 独享资源组**页面,单击相应资源组 后的**网络设置**,进入**专有网络绑定**页面查看**交换机网段**。

😑 (-) 阿里云 🗧) Iffe	当 🧧 华东2(上)	j) ~			Q 搜索	费用 工单	ICP 备案 企业 支	诗 App		₽′ @	简体 👩
DataWorks		DataWorks / 资源组列目	ξ.								产品动态	帮助文档
概范		+	×									
工作空间列表		专有网络绑定	DNS配置 Host配置									
报警配置	^	新端鄉定										
报警资源		资源组名称	专有网络	交通典机	交换机网段	弹性网卡ip	安全组			豪作		
报警联系人												
开放平台 New		类型 源组				A 100 K 10 K 10 K				解除绑定 自	定义路由	
计算引擎列表	^	可用区: cn- shanghai-f										

• 使用公网连通数据源与独享资源组:添加独享资源组的EIP地址至数据源白名单中。

	▲ 工作台 ₩ 华东2 (上海) ∨	_	Q 提家。		要用	工单 ICP 备案 企业	支持 App 🔄 🚨 🙀 🕐	简体
DataWorks	DataWorks / 资源组列表	独享资源组		×			产品动态	帮助文
概范	独享资源组 公共资源组 自定义资源组	资源组名称:						
工作空间列表		资源组ID:	Carl Inc. (Carl					c
资源组列表		资源组类型:	THE REAL PROPERTY AND A DESCRIPTION OF A					
报警配置	资源组名称/备注	订单号:		1	虛	使用率	操作	
报警资源	22	实例名称:	Carry and A researched			0%	查看信息 E维助手 扩容 缩容 续费 网络设置 修改归属工作空间	
报當联系人		资源数量:	2			7.81%	查看信息 扩容 缩容 续费 网络设置	
开放平台 New		安全组;	e de la companya de la				音看信息 运堆助手 扩容 缩容 续费	
计算引擎列表	^	VPC:				0%	网络设置修改归属工作空间	
MaxCompute	00000	EIP#84E: @				0%	查看信息 运缴助手 扩容 缩容 续费 网络设置 修改归属工作空间	
Graph Compute		网段: 主网卡IP:	The second se			0%	查看信息 运筹助手 扩容 缩容 续要 网络设置 修改归属工作空间	1
noogies	and the second s	环境配置:	点击查看详情			0%	查看信息 运输助手 扩容 缩容 续费 网络设置 修改归属工作空间	
	32			关闭		62.5%	查看信息 扩容 缩容 续费 网络设置 修改归属工作空间	
							查看信息 运生助手 扩容 缩容 续费	

测试网络联通

以上网络配置完成之后,您需要通过以下方式测试资源组与数据源之间的网络连通性。

- 1. 单击相应工作空间后的:图标,单击工作空间配置,在工作空间配置对话框中,单击更多设置,进入工 作空间配置页面。
- 2. 在左侧导航栏单击数据源管理,进入数据源管理页面。
- 3. 单击相应数据源操作列的编辑。
- 4. 在数据集成页签下,单击相应资源组名称后的测试连通性,连通状态为可连通时,表示连通成功。
| 编辑 | | | | | | | | | |
|----|--|---|--------------------|------------------------|-------------|---|--|--|--|
| | 资 | 源组连通性: 数据集成 数据服务 | 子 任务调度 | | | | | | |
| | 如果数据同步时使用了此数据源,那么就需要保证对应的资源组和数据源之间是可以联通的。请参考资源组的详细概念和网络解决方案。 | | | | | | | | |
| | → 新建独字数据集成资源组 | | | | | | | | |
| | | 独享数据集成资源组名称 | 连通状态
(点击状态查看详情) | 测试时间 | 操作 | | | | |
| | | Second Weakley Hills | ⊘可连通 | 2022/06/06
16:59:16 | 测试连通性 | | | | |
| | | ******* | ⊘可连通 | 2021/12/15
11:35:09 | 测试连通性 | | | | |
| | | ***** | ⊘可连通 | 2022/06/06
16:59:21 | 测试连通性 | | | | |
| | | 1992.0 | ⊘可连通 | 2022/01/18
16:10:08 | 测试连通性 | | | | |
| | | 车通性 | | | C 刷新 更多选项 | Ē | | | |
| | 1 注意手 | 顼 | | 原公共/ | 自定义资源组已移至此处 | | | | |
| | 如果测试 | 不通,可能的原因为: | | | | | | | |
| | 1. 数据》
2. DataV | 车没有启动,请确认已经止常启动。
Works无法访问数据库所在网络,请确 | 制保网络已和阿里云打通。 | | | | | | |
| | 3. DataWorks被数据库所在网络防火墙禁止,请添加白名单。 | | | | | | | | |
| | 4. 数据库域名无法被正确解析,请确认域名可以被正常解析访问。 | | | | | | | | |
| | | | | | | | | | |
| | 完成 | | | | | | | | |

⑦ 说明 如果网络不通,您可以单击自助排查解决,使用诊断工具进行诊断网络联通异常。若您需要进一步了解各网络环境下数据源与独享资源组的网络联通,详情请参见:配置资源组与网络连通。

5. 单击**完成**。

使用独享数据集成资源组

独享数据集成资源组创建并配置完成后,您可以通过如下方式切换任务使用的资源组。

操作环境 支持 作

操作环境	支持的切换操 作	界面入口			
		进入运维中心 > 周期任务界面设置。 勾选需要修改资源组的任务,单击底部菜单栏的修改数据集成资源组,即可批 量修改。			
切换生产环境 资源组	批量切换				
		THE T LEFT OF THE T			

操作环境	支持的切换操 作	界面入口
		进入DataStudio页面设置。 ● 单个任务切换 进入目标节点的编辑页面,在右侧导航栏的数据集成资源组配置中即可修改。 2 醒時 2 醒時 3 麗敏
		数据来源 ○
		・数据源名称 地域 す C 数据原名称 第选择 ● 数据原名称 第选择 ● 第选择 ● 第选择 ● 第选择 ● ● 数据原名称 第选择 ● ● ● ● ● 数据原名称 第选择 ●
切换开发环境 资源组	 单个任务切换 批量切换 	別式主通性 希望治学数据集成资源组 別式主通性 予想の方言
		● 批量切换
		单击 国图标,勾选需要修改资源组的任务,单击底部菜单栏的修改数据集成 资源组,即可批量修改。
		2 Annual and an

1.2.4. 独享调度资源组

1.2.4.1. 独享调度资源组概述

在任务高并发执行且无法错峰运行的情况下,需要专有的计算资源来保障任务定时被调度运行时,建议您选择 使用DataWorks的独享调度资源组。本文为您概要介绍独享调度资源组。

适用场景

- 对任务时效性要求高,避免公共资源组资源争抢导致任务延时。
- 需要灵活调整资源组大小。
- 需要访问公网地址、VPC、IDC。
- 需要通过IP白名单进行访问控制。
- DataWorks某些任务类型仅支持独享资源组,包括EMR、CDH、ADB等。

使用限制

- 独享调度资源组的购买计费类型为包年包月,在资源组到期前不支持提前删除、释放资源组,资源组到期 后会按照规定的时间停服、释放。
- 独享调度资源组不支持跨地域使用。例如,华东2(上海)地域的独享调度资源组,只能给华东2(上海) 地域的工作空间使用。
- 同一独享资源组下同规格资源数购买上限为20,若需要更多,请提交工单联系我们升级。

性能指标

规格	最大并发运行实例数
4c8g	16
8c16g	32
12c24g	48
16c32g	64
24c48g	96

规格选择及费用预估

您可以使用Dat aWorks独享资源组计算器预估所需规格及相关费用。预估规格及相关费用时,可进入 Dat aWorks的运维中心 > 运维大屏,查看资源运行的实例数、并发情况等信息作为参考。查看运维大屏,详 情请参见查看运维大屏。

计费与相关操作

(1) 计费说明

独享调度资源组的计费方式为包年包月,您可以根据自身需求选择合适的规格进行购买,不同规格的计费详情 请参见<u>独享调度资源组计费说明:包年包月。</u>

(2) 扩容与缩容

您可以根据业务需求选择购买独享资源组,指定购买资源组内机器规格、机器数量,购买完成后,您可以对资 源组进行扩容或缩容操作,扩缩容操作步骤和注意事项请参见扩缩容操作。

(3) 规格变更(升降配)

如果当前独享资源组规格不符合实际需要,您可以针对该资源组进行规格变更。变更当前资源组的规格,资源 组内所有机器的规格将统一变更,规格变更操步骤和注意事项请参见变更规格操作。

(4) 续费、停服与释放资源组

您可以在独享资源组临近到期时进行续费操作,或等待资源组到期后停服并自动释放资源。详情请参见到期说明及续费操作。

使用独享调度资源组

了解上述独享调度资源的规格计费详情后,您可以选择购买合适的独享调度资源组并在数据集成、数据开发等任务中使用。从购买至使用,您需进行的主要操作流程如下:

- 1. 购买独享资源组。
- 2. 添加资源组至DataWorks控制台。
- 3. 绑定工作空间。
- 4. (可选)为独享资源组绑定专有网络。
- 5. (可选)添加白名单。
- 6. 应用示例: 独享调度资源组。

操作详情请参见新增和使用独享调度资源组。

网络连通方案

独享调度资源组与其他资源组类似,本质上为一组阿里云ECS实例,在进行数据开发等任务运行时,需保障资源组与数据源之间的网络是连通的,且不会因为白名单等特殊安全访问设置阻断资源组与数据源之间的网络连通性。

⑦ 说明 如果独享调度资源组的应用过程中,不涉及与数据源的交互,则不涉及独享调度资源组与数据 源间的网络连通问题,独享调度资源组购买后即可直接使用,无需考虑网络的相关影响。

独享调度资源组购买后,需绑定一个VPC网络,后续您可以根据数据源所在的网络环境情况来选择网络连通方 案。

• 网络连通方案

数据源网络环境	网络连通方案
数据源处于公网环境中	数据源与独享资源组绑定的VPC网络直连即可连通。
数据源处于VPC网络中,且与独享资 源组在同一地域	建议独享资源组与数据源绑定同一VPC,绑定后系统自动为您添加一条路由至 整个VPC的路由,您绑定VPC的任意一个交换机,独享资源组与数据源的网络 都可连通。
数据源处于VPC网络中,且与独享资 源组在 不同 地域	使用高速通道或VPN等将网络连通数据源与独享资源组绑定的VPC网络,并添 加一条路由指向目标数据库IP,保障两个网络间网络可达。
数据源处于IDC网络中	使用高速通道或VPN等将网络连通数据源与独享资源组绑定的VPC网络,并添 加一条路由指向目标数据库IP,保障两个网络间网络可达。

数据源网络环境	网络连通方案
数据源处于经典网络中	独享资源组处于阿里云VPC网络,如果数据源处于经典网络环境中,则数据源 与独享资源组间网络无法连通,建议您将数据源迁移至VPC网络中。

● 白名单配置

如果数据源设置有白名单限制访问地址时,您需要将独享资源组的访问地址添加至数据源的访问白名单中,避免访问受限。详情可参见使用独享数据集成资源组执行任务需要在数据库添加的IP白名单。

1.2.4.2. 新增和使用独享调度资源组

您可以根据业务场景需求选择购买合适规格的独享调度资源组,用于任务的调度运行。购买独享资源组后,在 实际使用前您还需进行网络绑定、白名单等配置操作。本文为您介绍从购买至使用独享调度资源组的全流程与 步骤详情。

前提条件

购买和使用独享调度资源组前:

 您需先了解独享调度资源组的规格性能(不同规格能同时跑多少任务)、计费类型(不同规格怎么收费)等细节,提前根据业务场景规划需购买的资源组规格和购买时长,详情请参见独享调度资源组计费说明: 包年包月。

⑦ 说明 您可以使用DataWorks独享资源组计算器预估所需规格及费用。预估规格及费用时,可进入 DataWorks的运维中心 > 运维大屏,查看资源运行的实例数、并发情况等信息作为参考。查看运维大 屏,详情请参见查看运维大屏。

 (可选)如果独享调度资源组在后续的应用中可能涉及与数据源的交互,例如,在调度场景下,需要通过 Shell访问自建数据库或某个内部地址;或者使用了EMR、CDH这类必须要使用独享调度资源组的引擎时,您 还需了解不同场景下,独享调度资源组与数据源间或引擎间的网络连通方案和白名单注意事项,详情请参 见独享调度资源组概述。

⑦ 说明 如果您暂无连通数据库实例的需求,只是为了缓解因公共调度资源组的资源紧张,导致任务延迟的问题,则无需关注本文涉及的网络相关问题。您可以购买任意可用区的独享调度资源组,且无需进行网络相关配置操作。

操作流程

从购买至使用独享调度资源组,您需进行的主要操作流程如下:

- 1. 购买资源组(创建订单)
- 2. 基于订单创建资源组
- 3. 绑定归属工作空间
- 4. (可选)绑定专有网络
- 5. (可选)添加白名单
- 6. (可选)测试网络联通
- 7. 应用示例: 独享调度资源组

购买资源组(创建订单)

- 1. 登录阿里云账号并进入DataWorks独享资源组购买页面。
- 2. 在购买页面根据业务需要配置相应的参数, 配置要点如下。
 - **地域和可用区**:选择后续需要使用此独享资源组的地域。

⑦ 说明 独享调度资源组不支持跨地域使用。例如,华东2(上海)地域的独享资源,只能给华东2(上海)地域的工作空间使用。

○ 独享资源类型:选择为独享调度资源组。

其他配置项,例如计费周期等参数,可根据业务规划自行配置即可。

⑦ 说明 同一独享资源组下同规格资源数购买上限为20,若需要更多,请提交工单联系我们升级。

3. 单击**立即购买**,根据界面提示进行付款等操作后即可完成购买独享调度资源组。

完成购买后,您可以在我的订单中查看到此订单及订单号详情。

☰ €Э阿聖云	✿ 工作台 ② 华东2 (上海) ∨ Q	搜索 费用 工单 ICP 备案 企业 支持 App
DataWorks	DataWorks / 资源组列表	充值
概览页	资源组列表	2)単 发票
工作空间列表	独享资源组 公共资源组	费用账单
资源组列表		续费管理
报警配置	创建调度资源组 创建集成资源组 创建服务资源组 请输入搜索	影关键词 Q 用户中心

后续您可根据订单号将此次购买的独享调度资源组添加到DataWorks控制台。

基于订单创建资源组

- 1. 登录DataWorks控制台。
- 2. 单击左侧导航栏资源组列表,在独享资源组页签下单击创建调度资源组页面,配置各项参数。

创建独享资源组					
计 独享调度资源、独享数据集成资源用途	不同,详情参考:帮助文档				
 资源组类型: ④ 独享调度资源组 ④ 独享规度策源组 	送资源组				
请输入资源组名称					
* 资源组备注:					
请输入资源组备注					
* 订单号: 购买 当前地域: China (Sha	nghai),请购买此地域的资源组				
清选择订单号					
参数	描述				
	选择添加的独享资源组类型。				

独享资源组包括**独享调度资源组**和**独享数据集成资源组**,分别适用于通用任务调度 和数据同步任务专用。

资源组类型

参数	描述			
	设置资源组的名称,此名称需在租户内唯一,请避免重复,否则会导致在后续进行确 定操作时报错。			
资源组名称	 ⑦ 说明 租户即主账号,一个租户(主账号)下可以有多个用户(子账号)。 			
资源组备注	对资源组进行简单描述。			
	在下拉框中选择已购买的独享资源组的订单号。			
订单号	⑦ 说明 如果还需要继续购买独享资源组,也可单击订单号后的购买继续购买,完成购买后回到此页面再将新购的独享资源组添加至DataWorks控制台。			

3. 单击**确定**, DataWorks即开始初始化此独享资源组,当资源组状态为运行中后,独享资源组即已添加至控制台。

⑦ 说明 独享资源组的初始化需要约20分钟,请您耐心等待其状态更新为运行中。

独享资源组添加至控制台后,您需要为资源组绑定工作空间,才可以在任务的配置中选择使用此独享资源组。

绑定归属工作空间

独享调度资源组需要绑定归属的工作空间,才可以在对应的工作空间下选择使用该独享资源组。一个独享调度 资源组可以分配给多个工作空间使用,但不可跨地域使用,例如,上海地域的独享调度资源组仅可绑定上海地 域的工作空间使用,不可绑定其他地域的工作空间。绑定归属工作空间的操作如下。

1. 登录DataWorks控制台。

- 2. 在资源组列表的独享资源组页签下,单击相应资源组后的修改归属工作空间。
- 3. 在修改归属工作空间单击对应工作空间后的绑定,即可将独享资源组绑定至此工作空间。

(可选)绑定专有网络

如果独享调度资源组在后续的应用中可能涉及与数据源的交互,例如,在调度场景下,需要通过Shell访问自 建数据库或某个内部地址;或者使用了EMR、CDH这类必须要使用独享调度资源组的引擎时,您还需进行以下 绑定专有网络、白名单的相关配置。

独享资源组部署在DataWorks托管的专有网络中,您需要绑定自己的专有网络后,独享资源组才可与其他数据 源连通。绑定专有网络的操作如下。

↓ 注意 4c8g类型的独享调度资源组最多支持绑定2个专有网络,其他规格的独享调度资源组最多支持 绑定3个专有网络。

1. 登录DataWorks控制台。

2. 在资源组列表的独享资源组页签下,单击相应资源组后的专有网络绑定,进入绑定页面。

绑定前,请首先使用阿里云主账号进行RAM授权(仅主账号有权限),让DataWorks拥有访问您的云资源 的权限。

(-) 阿里云		Q 搜索	费用	工单	备案	企业	支持与服务	۶	<u>0</u> •	Ä	简体中;
13	云资源访问授权										
	晶馨提示:如需修改角色权限,请前往RAM控制台角色管理中设置,	需要注意的是,错误的配置可能导致DataWorks无法获取到必要的	汉限。						×		
	DataWorks请求获取访问您云资源的权限 下方是系统创建的可供DataWorks提用的角色,使权后,DataWorks拥有对您云资源相应的访问权限。										
	AliyunDataWorksAccessingENIRole 描述: DataWorka默认使用此角色来访问您在其他云产品中的 权限描述:用于DataWorks角色的既权策略,包含弹性两卡EN	流之源 贝约兰部分 化2限							~		
		同意授权									

- 3. 绑定专有网络VPC。
 - i. 单击**专有网络绑定**页面左上方的**新增绑定**,在**新增专有网络绑定**对话框中,配置各项参数,不同 网络环境下各参数的配置说明如下。

参数说明如下:

参数	配置说明(数据源与独享资源同 账号同地域)	配置说明(数据源与独享资源在不同账号或不同地 域)					
专有网络	如果您的数据源与独享资源组在 同一个阿里云账号下,建议配置 为数据源所在的VPC。 如果不在同一个阿里云账号下, 则与不在同一地域场景一致。	如果您的数据源与独享资源不在同一地域,例如,数 据源不在阿里云VPC网络环境中,您可单击 创建专有 网络,为独享资源组创建一个VPC。创建完成后这里 配置为新建的VPC或选择已经与目标数据库网络打通 的VPC。					
可用区	选择数据库所在可用区。	选择已经与目标数据库网络联通的可用区。					
	专有网络配置为数据源所在VPC 时,建议选择与数据源绑定的交 换机。						
交换机	⑦ 说明 绑定数据源所在 VPC后,绑定VPC下任意一个 交换机,会自动添加路由至 整个VPC网段,实现独享数 据集成资源组在该VPC下网 络可达。	选择已经与目标数据库网络联通的交换机,若没有可 用交换机,可单击 创建交换机 为独享资源组创建交 换机。创建完成后这里配置为创建的交换机。					
安全组	安全组指定了独享资源实例需要允许或禁止哪些公网或内网的访问。您可根据业务需求选 择已有的安全组,或单击 创建安全组 为独享资源实例创建安全组。创建安全组的详细参数 配置可参见 <mark>添加安全组规则</mark> 。						

ii. 单击**确定**,完成绑定VPC操作。

⑦ 说明 如果数据源和独享资源组不在同一个地域,或不在同一个阿里云账号下,则您需要绑定专有网络后,再添加路由规则指向目标数据库IP地址。

4. (可选) 配置Host。

如果您的数据源无法通过IP直接访问,例如,数据源通过Host的域名托管,通过Host域名直接被外部访问时,您需要参考以下步骤配置Host,否则在添加数据源时,使用数据源的Host域名进行连通性测试时会失败。

i. 单击**Host 配置**,在Host 配置页面左上方单击**新增**,在**新增域名Host 配置**对话框中配置各项参数。 参数说明如下。

配置说明				
配置为数据源的实际IP地址。				
配置为数据源对外提供访问服务的Host域名。如果有多个Host域名时,可换行依 次添加。				
⑦ 说明 此处的域名需包含数字、字母、连字符(-)、点(.),且必须 以字母开头,以字母或者数字结尾。				

- ii. 如果有多个IP地址需要添加,可继续单击新增继续添加。
 - ? 说明
 - 新增的Host配置中, IP、域名不能与之前配置的Host中的IP或域名重复。
 - 一个Host配置中, IP与域名为1对多的关系, 即IP可以对应多个Host域名, 但是同一个 Host域名只能指向一个IP。
- 5. (可选) 配置DNS。

如果您的数据源无法通过IP直接访问,例如,数据源通过负载均衡域名直接被外部访问,通过内部域名解 析服务器将域名解析至实际数据源IP地址时,您需要参考以下步骤配置DNS,否则在添加数据源时,使用 数据源的Host域名进行连通性测试时会失败。

⑦ 说明 如果同一个域名同时配置了Host和DNS,系统会优先根据Host的访问规则访问数据源。

i. 单击**DNS配置**,在DNS配置页面左下角单击**添加**,配置完成DNS各项参数后单击**保存**。参数说明如下。

参数	配置说明
	非必配项。如果数据源有统一的一级域名,可在此处配置为数据源对外提供访问 的域名的一级域名。 例如,数据源1对外的访问域名为domain1.example.com,数据源2对外的访问 域名为domain2.example.com,此处建议配置为example.com
Domain	⑦ 说明 此处的域名需包含数字、字母、连字符(-)、点(.),且必须 以字母开头,以字母或者数字结尾。
NameServer	配置为数据源提供域名解析的解析服务的IP地址。如果有多个域名解析服务器时,可换行依次添加。

ii. 如果您需要修改之前配置的DNS,您可单击左下角的修改。

(可选)添加白名单

独享调度资源组和数据源同可用区时,如果同专有网络、交换机的情况下仍未连通网络,则可能是因为数据源 有白名单限制。请添加以下内容至数据源的白名单列表中。

• 使用内网连通数据源与独享资源组:添加独享资源组绑定专有网络时的交换机网段至数据源白名单中。

您需要为独享资源组绑定专有网络后,进入DataWorks的资源组列表 > 独享资源组页面,单击相应资源组 后的**专有网络绑定**,查看**交换机网段**。

≡	(-) 阿里云	â In	治 🧧 华东2(」	[海] ~			Q 搜索	费用	工単	ICP 备案	企业	封支	Арр	۶.,	٥.	∀່ @	简体 🌔
Data	Works		DataWorks / 资源组列	利表												产品动态	帮助文档
概范			←	×													
I/F8 3033	空间列表 组列表		专有网络绑定	DNS配置 Host配置													
报警	記置	^	新增將定														
Ħ	要警察課		资源组名称	专有网络	交換机	交换机网段	弹性网卡ip		安全组				3	榮作			
折	國醫联系人																
开放3	平台 New		类型 源组										1	网络螺纹	自定	义路由	
计算	目攀列表 AaxCompute	^	可用区: cn- shanghai-f														

• 使用公网连通数据源与独享资源组:添加独享资源组的EIP地址至数据源白名单中。

日()阿里云	☆ 工作台 ● 华东2 (上海) ∨	_	Q	变欢	费用 工年	1 ICP 备案 企业	支持 App 🖸 🛕 👾 🕜	简体(
DataWorks	DataWorks / 资源组列表	独享资源组		×			7*583.5/s	g 帮助文林
概范	独享资源组 公共资源组 自定义资源组	资源组名称:						
工作空间列表		资源组D:	Concentration of the Concentra					c
资源组列表		资源组类型:	and the second sec					
报警配置	2)原垣名称/畜注	订单号:			堂	使用率	操作	
报警告簿	22	实例名称: 实例规格:	And the second se			0%	查看信息 武士助手 扩容 缩容 续 网络设置 修改归属工作空间	#
报警联系人		资源数量:	2			7.81%	查看信息 扩容 缩容 续费 网络设 修改归属工作空间	*
开放平台 New 计算引擎列表	_ ·	VPC:	 Parameters 			0%	查看信息 运维助手 扩容 缩容 续 网络设置 修改归属工作空间	类
MaxCompute	2002	VSwitch: EIP地址: @				0%	重看信息 运输助手 扩容 缩容 续 网络设置 修改归属工作空间	费
Graph Compute Hologres		网段: 主网卡IP:				0%	查看信息 运堆助手 扩容 结容 续 网络设置 修改归属工作空间	æ (
	And Statements	环境配置:	点击直翻洋情			0%	查看信息 运爆助手 扩容 缩容 续 网络设置 修改归属工作空间	æ
	12			关闭		62.5%	查看信息 扩容 缩容 续费 网络设修改归属工作空间	z
							查看信息 运堆助手 扩容 缩容 续	费 🗸

(可选)测试网络联通

以上网络配置完成之后,您需要通过以下方式测试资源组与数据源之间的网络连通性。

- 1. 单击相应工作空间后的:图标,单击工作空间配置,在工作空间配置对话框中,单击更多设置,进入工作空间配置页面。
- 2. 单击相应数据源操作列的编辑。
- 3. 在任务调度页签下,单击相应资源组名称后的测试连通性,连通状态为可连通时,表示连通成功。

E	议/店/凉				
	* 用户名:				
	*密码: …				
	资源组连通性: 数据集成	赵据服务 任务调度			
0	如果任务调度时使用了此数据 络解决方案。	暑源,那么就需要保证对应的资	5源组和数据源之间是可以联通的	。请参考资源组的问	羊细概念和网
	资源组名称	类型	连通状态 (点击状态查看详情)	测试时间	操作
	100	独享调度资源组	⊘可连通	2022/06/06 17:19:52	测试连通性
		独享调度资源组	⊘可连通	2022/06/06 17:19:57	测试连通性
	1.01.0 10110	独享调度资源组	⊘可连通	2022/06/06 17:20:00	测试连通性
	刘试连通性				
日注	注意事 项				
如果测	则试不通, 可能的原因为:				
1. 数	数据库没有启动,请确认已经	正常启动。			
2. D)ataWorks无法访问数据库所得	王网络,请确保网络已和阿里云	打通。		
3. D)ataWorks被数据库所在网络 物据左域タ天法被正确解析	历火墙禁止,请添加 日名甲。 唐确认诚么可以被正受解场前	a		
4. 3	xuna)++∞%rh2/U/A10X111WH8941/1,	1104 171#8(+h-11,XII)/9 (*) FW6W (#)	-Jo		
					完成

源组与网络连通。

4. 单击完成。

应用示例: 独享调度资源组

- 修改在DataStudio页面测试任务使用的资源组
 - i. 在DataStudio页面,双击打开相应节点的编辑页面。
 - ii. 单击工具栏中的回图标。

iii. 在参数对话框中,选择测试运行使用的调度资源组。

iv. 单击**确定**。

- 在DataStudio修改使用的调度资源组
 - i. 在DataStudio面板,双击打开相应节点的编辑页面。
 - ii. 单击右侧的调度配置,在资源属性区域选择任务调度运行时需要使用的资源组。详情请参见配置资源 属性。

×	调度配置			调
	22 15 F2 M			度配
	資源唐任 —			罿
	调度资源组	公共调度资源组	▲ 查看水位	

• 在运维中心修改使用的调度资源组

- i. 在运维中心的左侧导航栏,单击周期任务运维 > 周期任务。
- ii. 单击页面中间的箭头展开任务列表后, 单击相应周期任务后的更多 > 修改资源组。

6	🏶 运维中心	5		•							& Dat	taStudio	್ಷ 📕	e ka ke
¢	运输大屏													
\$\$\$	实时任务运维	~	搜索:	节点名称/节点ID	Q 节点类型:	请选择节点类型	> 责任人: 请选择责任	入 🗸 🕹	市点 🗌 今日修改的节点	1 暫停 (冻结)	节点			
a	周期任务运输	^											〇周新「展	干搜索
	周期任务			名称		节点ID	修改日期 11	任务类型	责任人	调度类型	资源组 🎧	操作		
	周期实例			100000000		700003259535	2020-04-28 10:46:40	ODPS_SQL	-	日调度	公共调度资料	DAG	测试 补数据 ▼	更多 🔻
	补数据实例			-14170.00		700002944540	2020-04-28 10:46:39	ODPS_SQL	Benefit Bart	日调度	公共调度资源	DAG	暂停 (冻结)	-
	15-34/142/			1000		700002944539	2020-04-28 10:35:15	ODPS_SQL	manufacture and	日调度	公共调度资料	DAG	1代夏(1軒/休) 查看实例	-
	apainter and					700002345286	2020-04-27 10:31:24	ODPS_SQL	-	小时调度	xc_schedule	DAG	添加报警 修改责任人	-
JAL	HINGING L			- 34.544		700002345289	2020-04-27 10:31:24	ODPS_SQL	dependence of the second	小时调度	xc_schedule	DAG	添加到基线	
						700002563306	2020-04-24 18:32:08	ODPS_SQL	Basello, Brazil	日调度	公共调度资料	DAGI	修改资源组 配置质量监控	-
						700002554758	2020-04-23 16:24:57	ODPS_SQL	100.000	小时调度	公共调度资料	DAGI	查看血缘	-
													上下游	

↓ 注意 虚节点、业务流程、机器学习等任务类型不支持修改资源组,请勿选择。

如果您需要批量修改资源组,请在**周期任务**页面,选中需要修改的周期任务,单击页面下方的修改资 **源组**。

⑤ 🏶 运维中心	10	1 ¹⁰⁰	~						& Data	Studio 💐 🛑
🕒 运维大屏										
🐝 实时任务运维	~	投業: 节点名称/节点ID	Q 节点类	型: 请选择节点类型	→ 責任人: 消洗择	责任人 🗸 🗸	我的节点 今日惨改的节点	暫停 (冻结)	节点	
口 周期任务运维	^									〇刷新 展开搜索
周期任务		- 名称		节点ID	修改日期 11	任务类型	责任人	调度类型	资源组 🎧	操作
周期实例			存2	700003259535	2020-04-28 10:46:40	ODPS_SQL	1004103-00102	日调度	公共调度资源	DAG图 测试 补数据 ▼ 更多 ▼
計数据实例		• • • • • • • • •	<i>a</i>	700002944540	2020-04-28 10:46:39	ODPS_SQL	00000A_0000	日调度	公共调度资源	DAG圈 测试 补数据 ▼ 更多 ▼
Q 手动杆务运输		-	Ø	700002944539	2020-04-28 10:35:15	ODPS_SQL	meaning mean	日调度	公共调度资)	DAG图 测试 补数据 ▼ 更多 ▼
▲ 智能监控	Ú.	1.11.101		700002345286	2020-04-27 10:31:24	ODPS_SQL	mounds.mout	小时调度	xc_schedule	DAG图「测试丨补数据 ▼「更多 ▼
		1.00.000		700002345289	2020-04-27 10:31:24	ODPS_SQL	000000.0000	小时调度	xc_schedule	DAG图「测试丨补数据 ▼ 更多 ▼
		-		700002563306	2020-04-24 18:32:08	ODPS_SQL	Second Second	日调度	公共调度资料	DAG图 测试 补数据 ▼ 更多 ▼
		-		700002554758	2020-04-23 16:24:57	ODPS_SQL	10000	小时调度	公共调度资)	DAG圈 测试 补数据 ▼ 更多 ▼
				700002564164	2020-04-23 16:23:59	数据集成	dependencie and	小时调度	同步资源组:	DAG图 测试 补数据 ▼ 更多 ▼
		1.789.08		700003250117	2020-04-23 16:20:28	ODPS_SQL	0000000.00000	小时调度	公共调度资料	DAG图 测试 补数据 ▼ 更多 ▼
		4,78		700002967416	2020-04-23 15:41:06	ODPS_SQL	meaning meaning	小时调度	公共调度资料	DAG图丨测试丨补数据 ▼丨更多 ▼
				700002967477	2020-04-23 15:39:13	ODPS_SQL	depends_dens2	小时调度	公共调度资料	DAG图丨测试丨补数据 ▼丨更多 ▼
				700003249371	2020-04-23 11:31:10	ODPS_SQL	000005.0000	小时调度	公共调度资源	DAG图 测试 补数据 ▼ 更多 ▼ ↓
	•	添加报警 修改表	F人 修改资源组	添加到基线	香停 (冻结)	解冻) 下线节点				< 1 2 3 >

iii. 在修改资源组对话框中,选择相应的调度资源组,单击确定。

1.2.5. 独享数据服务资源组

1.2.5.1. 独享数据服务资源组

在调用数据服务API需要高QPS和SLA保障的情况下,企业需要使用独享的数据服务资源来保障调用成功。同时,独享数据服务资源组能够满足高并发、高频率的接口调用并及时返回结果数据。

使用限制

- 目前独享数据服务资源组仅支持在华东1(杭州)地域和华东2(上海)地域使用。性能指标请参见独享数据服务资源组计费说明:包年包月。
- 目前独享数据服务资源组仅支持同一个地域VPC网络下的数据源。

适用场景

数据服务API调用频率高,高并发且需要及时返回数据结果的场景,请使用独享数据服务资源组。

计费说明

独享数据服务资源组的计费方式为包年包月,您可以根据自身业务需求选择合适的规格进行计费,详情请参见<u>独享数据服务资源组的计费标准</u>。

网络连通

独享数据服务资源需要访问VPC环境下的数据源时,需要绑定该数据源所在的VPC,绑定VPC后会自定义路由 到整个专有网络,实现独享资源组在整个专有网络内网络可达。

其他使用指引

扩容说明

您可以根据业务需求,在DataWorks管理控制台中对每个独享数据服务资源订单中的资源配置进行升级。您 仅需支付升配资源从当前时间点至原订单到期时间点之间的差价。

? 说明

- 升配仅会根据已购订单升级资源规格,不会增加资源组数量。
- 执行升配后,新配置将在升级订单支付后的10分钟内生效。
- 续费说明

您可以选择在独享数据服务资源订单到期前、或到期后的7天内,为该订单内所有资源统一续费。如果您在 到期前未续费,到期后将立即停止该订单内资源的服务。

- 到期说明
 - 停服

默认在距离您所选购的独享数据服务资源到期前的7天、3天和1天,分别向您的阿里云主账号绑定的手机和邮箱等发送到期通知。如果您在到期前未续费,则到期后将立即停止该订单内资源的服务。

○ 释放

独享数据服务资源到期后,将为您继续保留7天,如果7天内您仍未完成续费,则该实例将被释放。

您可以根据业务需求,在Dat aWorks管理控制台中对每个独享数据服务资源订单中的资源配置进行升级。您仅需支付升配资源从当前时间点至原订单到期时间点之间的差价。

1.2.5.2. 新增和使用独享数据服务资源组

您可以根据业务场景需求选择购买合适规格的独享数据服务资源组,以保障高并发和高频率的API调用。本文为您介绍从购买至使用独享数据服务资源组的全流程与步骤详情。

前提条件

购买和使用独享数据服务资源组前:

- 你需要先了解独享数据服务资源组的规格性能(不同规格能支持的QPS)、计费类型(不同规格怎么收费)等细节,提前根据业务场景规划需购买的资源组规格和购买时长,详情请参见:独享数据服务资源组计费 说明:包年包月。
- 购买独享数据服务资源组时需要进行网络配置,您可以为独享资源组绑定一个可与目标数据源连通的VPC和 该VPC下对应可用区中的V-Switch,实现自动添加路由至整个VPC网段,从而保障独享数据服务资源组与该 VPC环境中的数据源之间的网络连通性。查看数据源所在网络环境的信息如何查看数据源的网络环境?。
- 在发起API调用请求时,可能会涉及到与不同网络环境下数据源之间的交互,因此你还需要了解独享数据服务资源组和不同网络环境下数据源之间的网络连通方案,详情请参见:网络连通与白名单配置。

操作流程

从购买至使用独享数据服务资源组,您需进行的主要操作流程如下:

- 1. 购买独享数据服务资源组
- 2. 绑定归属工作空间
- 3. 网络连通与白名单配置
- 4. 网络联通测试
- 5. 应用示例: 独享数据服务资源组

购买独享数据服务资源组

- 1. 登录DataWorks管理控制台。
- 2. 单击左侧导航栏**资源组列表**,在**独享资源组**页签下单击**创建服务资源组**,进入购买页面。在购买页面 根据业务需要配置相应的参数,配置要点如下。
 - 地域和可用区:选择后续需要使用此独享资源组的地域。

? 说明

- 可用区是在同一地域下的不同物理区域,可用区之间内网互通。
- 独享资源组不支持跨地域使用。例如,华东2(上海)地域的独享资源,只能给华东2(上海)地域的工作空间使用。
- 专有网络(VPC): 下拉列表为您列出当前登录账号下的所有专有网络,请选择需要跟独享资源组所 处的专有网络进行绑定的目标专有网络。
- 交换机(V-Switch):下拉列表会为列出您选择的目标专有网络下的所有交换机实例,请按实际需要 进行选择。
- 独享数据服务资源:根据需求选择独享数据服务资源组的规格。

其他配置项,例如计费周期等参数,可根据业务规划自行配置即可。

第击**立即购买**,根据界面提示进行付款等操作后,DataWorks即开始初始化此独享资源组,当资源组状态为运行中后,独享资源组即已添加至控制台。

⑦ 说明 独享资源组的初始化需要约20分钟,请您耐心等待其状态更新为运行中。

独享资源组添加至控制台后,您需要为资源组绑定工作空间,才可以在API的配置中选择使用此独享资源组。

绑定归属工作空间

独享数据服务资源组需要绑定归属的工作空间,才可以在对应的工作空间下选择使用该独享资源组。一个独享 数据服务资源组可以分配给多个工作空间使用,但不可跨地域使用,例如,上海地域的独享数据服务资源组仅 可绑定上海地域的工作空间使用,不可绑定其他地域的工作空间。绑定归属工作空间的操作如下。

- 1. 登录DataWorks管理控制台。
- 2. 在资源组列表的独享资源组页签下,单击相应资源组后的修改归属工作空间。
- 3. 在修改归属工作空间单击对应工作空间后的绑定,即可将独享资源组绑定至此工作空间。

网络连通与白名单配置

独享数据服务资源组与数据源之间进行数据服务时,需要确保独享数据服务资源组与不同网络环境数据源之间的连通性,独享数据服务资源组包括公网、阿里云VPC网络、IDC网络等场景的网络连通方案,不同连通场景下需要添加资源组白名单到数据库的白名单配置中。详情请参见:网络连通与白名单配置。

网络联通测试

以上网络配置完成之后,您需要通过以下方式测试资源组与数据源之间的网络连通性。

- 1. 单击相应工作空间后的:图标,单击工作空间配置,在工作空间配置对话框中,单击更多设置,进入工 作空间配置页面。
- 2. 在左侧导航栏单击数据源管理,进入数据源管理页面。
- 3. 单击相应数据源操作列的编辑。
- 4. 在数据服务页签下,单击相应资源组名称后的测试连通性,连通状态为可连通时,表示连通成功。

编辑	数据源					×				
	* 用户名:					-				
	*认证选项: 💿 无认证 🔵 SSL认证									
备库设置: 🌒 ?										
	资源组连通性: 数据集成	数据服务任务调度				*				
i \$	如果数据服务使用了此数据源	9,那么就需要保证公共资源组;	和数据源之间是可以联通的。							
	资源组名称	类型	连通状态 (点击状态查看详情)	测试时间	操作					
	1	独享资源组	ਂ⊙可连通	2022/06/06 17:23:52	测试连通性	н				
	公共	资源组	未测试		测试连通性	н				
	试连通性					Ш				
 注意事项 如果测试不通,可能的原因为: 1. 数据库没有启动,请确认已经正常启动。 2. DataWorks无法访问数据库所在网络,请确保网络已和阿里云打通。 3. DataWorks被数据库所在网络防火墙禁止,请添加 白名单。 4. 数据库域名无法被正确解析,请确认域名可以被正常解析访问。 										
					完成	取消				

⑦ 说明 独享数据服务资源组在购买时要选择需要进行连通的网络环境,若您的数据库存在白名单 访问控制,您可添加对应的白名单,详情请参见:网络连通与白名单配置。

5. 单击**完成**。

应用示例: 独享数据服务资源组

⑦ 说明 当前仅数据服务的生成API可以使用独享数据服务资源组。

1. 在数据服务面板的目录树中双击API的名称打开编辑页面。

2. 单击右侧的**服务资源组,方案**选择**独享服务资源组**,并在独享服务资源组下拉框选择API调用时需要使 用的资源组,详情请参见:通过向导模式生成API和通过脚本模式生成API。 与资源组



1.2.6. 自定义资源组

1.2.6.1. 自定义资源组概述

DataWorks支持自定义资源组(包括自定义调度资源组和自定义数据集成资源组),如果您有富余的服务器资源,可将此部分资源作为DataWorks中任务运行的资源组使用。本文为您介绍自定义资源组的概要信息。

适用场景

自定义资源组适用于如下场景:

- 访问IDC网络环境:进行数据同步时,需要访问自己的IDC,该场景可以使用自定义资源组。
- 减少迁移工作量:如果本地已经有任务运行,您可以选用自定义资源组,直接调度任务在自己的服务器上,减少脚本迁移的工作量。

使用限制

自定义资源组的限制情况如下表。

资源组类型	版本限制	可执行任务限制	删除操作限制
自定义调度资源组	需要开通DataWorks企业版及以上版本,并 且 <mark>提交工单</mark> 开放白名单。	仅支持执行Shell任 务。	不支持删除自定义调 度资源组。
自定义数据集成资 源组	需要开通DataWorks专业版及以上版本。	仅支持执行数据集成 任务。	不涉及。

计费说明

购买对应增值版本DataWorks后,即可使用自定义资源组功能,自定义资源组不另外收费。DataWorks增值版 本的费用,详情请参见<mark>版本服务计费说明</mark>。 ⑦ 说明 使用自定义数据集成资源组后,执行数据集成任务时不会收取相关费用;产生的公网流量,也 不会在DataWorks中收费。

使用自定义资源组

使用自定义资源组的主要操作流程如下:

- 自定义调度资源组
 - i. 开通DataWorks企业版及以上版本,并且提交工单开放白名单。
 - ii. 新增自定义调度资源组。
 - iii. 资源组添加并初始化服务器。

iv. 使用自定义调度资源组,即修改任务运行的资源组为自定义调度资源组。 操作详情请参见新增和使用自定义调度资源组。

自定义数据集成资源组

- i. 购买云服务器ECS。
- ii. 查看ECS主机名和内网IP地址。
- iii. 开通8000端口。
- iv. 开通DataWorks专业版及以上版本。
- v. 新增并配置自定义数据集成资源组,包括添加服务器、安装Agent、检查连通性。
- vi. 使用自定义数据集成资源组,即修改任务运行的资源组为自定义数据集成资源组。

操作详情请参见新增和使用自定义数据集成资源组。

1.2.6.2. 新增和使用自定义调度资源组

DataWorks支持创建自定义调度资源组,以保证任务的调度灵活性和数据同步的及时性。本文为您介绍如何新 增并使用自定义调度资源组。

前提条件

• 购买DataWorks

仅DataWorks企业版及以上版本支持使用自定义调度资源组。您需要提前购买DataWorks企业版及以上版本,购买详情请参见版本服务计费说明。

• 已有自有服务器资源

自定义调度资源组为您提供增加并发调度任务数量的能力。您需要购买自有服务器资源,才能使用自定义 调度资源组。本文以ECS服务器为例,为您介绍自定义调度资源组如何添加自有服务器资源并使用,详情请 参见资源组添加并初始化服务器。购买ECS服务器请进入ECS售卖页。

背景信息

调度资源组用于执行或分发调度系统下发的任务,DataWorks的调度资源组包括公共调度资源组、独享调度 资源组和自定义调度资源组。对于单个用户来说,公共资源组中的调度能力不能无限量承载不断增加的并发 调度任务量。如果调度的任务瞬时并发数量过多,则会出现资源等待的情况,导致任务调度实际运行时间延 迟。如果您想将已购买的ECS实例用于DataWorks任务调度,可以将其添加为DataWorks自定义调度资源组。

⑦ 说明 默认情况下,您提交的调度任务均会运行在DataWorks提供的公共调度资源组上。

自定义资源组和独享资源组的运维管理差异如下:

- 自定义调度资源组:机器完全可控、可登录访问,需要您自行运维。
- 自定义数据集成资源组:运维完全托管于系统,系统提供可用保障,无需担心机器服务等问题。

使用限制

自定义调度资源组仅支持Shell类型的任务使用。

操作流程

使用自定义调度资源组的操作流程如下:

- 1. 新增自定义调度资源组
- 2. 资源组添加并初始化服务器
- 3. 使用自定义调度资源组

自定义资源组在使用、配置文件、命令等方面的常见问题和解决方案,详情请参见自定义资源组。

新增自定义调度资源组

- 1. 进入自定义资源组页面。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击资源组列表, 选择目标地域。
 - iii. 在资源组列表页面,单击自定义资源组。
- 2. 新增自定义调度资源组。
 - i. 在自定义资源组页签下, 单击右上方的新增调度资源。
 - ii. 在新增调度资源对话框中, 输入资源名称并选择归属工作空间。

新增调度资源		×
* 资源名称: * 归属工作空间:	请输入资源名称 请输入归属工作空间 ~	
	确定	取消

自定义调度资源组配置完成后即可供其所归属的工作空间内的任务使用。如果后续您需要修改资源组 归属的工作空间,可参见通用参考:修改资源组归属工作空间。

iii. 单击确定,新增自定义调度资源组完成。

资源组添加并初始化服务器

自定义调度资源组创建完成后,您需要为资源组添加需要的服务器,并初始化服务器。

1. 添加服务器。

资源组列	表			
独享资源组	公共资源组	自定义资源组		
输入透露名称/服务				******
10				銀作
				新装装打扮化 新装装装饰 中次注意
				网络新拉州化 网络新香港 中文出版
			管理服务器	
			新増減度资源 3 ×	ESECUL EXECT FACE
			● 同志兴型: ○ 協典同語 ② 专有用語	网络新订加化 网络新香港 使次日常
			* ECS UUID: 資格入UUID, 利服防器合称 * 初期: 初期: 初期: 加速防治法	服务器订出化 服务器管理 修改日期
			ROM MAR	医弗莱尔特化 医异肠管理 使次日期
(Incomparison) - (58	
		经典网络		服装器初始化 服装器管理 博改注意

- i. 在自定义资源组页签下,单击目标资源组操作列的服务器管理。
- ii. 在管理服务器对话框中, 单击增加服务器。
- iii. 在新增调度资源对话框中, 配置各项参数。

参数	描述
网络类型	包括经典网络和专有网络。
	⑦ 说明 经典网络仅支持华东2(上海)地域,建议您选择网络类型为专有网络。
服务器名称	输入真实的主机名称,不支持自定义。您可以在ECS上执行 hostname 命令,获取服务器名称。
	⑦ 说明 仅当网络类型选择经典网络时,需要配置该参数。
	您可以在ECS上执行 dmidecode grep UUID 命令,获取ECS UUID。
ECS UUID	⑦ 说明 仅当网络类型选择专有网络时,需要配置该参数。
+0 99 UD	
心品で	

iv. 单击**确定**,成功添加服务器。

2. 初始化服务器。

i. 在自定义资源组页签下,单击目标资源组操作列的服务器初始化。

ii. 您需要登录ECS服务器,参考服务器初始化对话框中的操作初始化服务器。

服务器初始化	×
新添加阿里云ECS实例,请按照如下步骤操作来进行初始化准备: step1: SSH登录ECS实例,保持在root用户下。 step2:执行命令:wget https://di-agent-image-cn-shanghai.oss-cn-shanghai.aliyuncs.com/install.sh - no-check-certificate step3:执行命令:shinstall.sh -user_name=z - password: image-cn-shanghai -time_zone=GMT+8 step4:稍后在添加服务器页面,单击刷新按钮,观察服务状态是否转为"正常"状态。 step5:请开通服务器的8000端口。	
我知道	7

初始化完成后,您可以开始使用新增的自定义调度资源组。

使用自定义调度资源组

自定义调度资源组创建并配置完成后,您可以通过如下方式切换任务使用的资源组为自定义调度资源组。

? 说明

- 自定义调度资源组仅支持Shell类型的任务使用。
- 自定义调度资源组仅支持其所归属工作空间中的任务使用。如果您需要修改资源组归属的工作空间,详情请参见通用参考:修改资源组归属工作空间。

操作 界面入口		
---------	--	--

Dat aWorks

操作环境	支持的切换操作	界面入口
切换生产环境资源组	 单个任务切换 批量切换 	<section-header><text><text><text><image/><section-header><text><text></text></text></section-header></text></text></text></section-header>

DataWorks公共云合集·产品版本规格 与资源组

操作环境	支持的切换操作	界面入口
切换开发环境资源组	 单个任务切换 批量切换 	<section-header></section-header>

操作环境	支持的切换操作	界面入口
切换任务调试资 源组	仅支持单个任务 切换	进入DataStudio > 节点编辑页面设置。 单击 ◎图标,即可切换节点使用的调度资源组。 **********************************

切换资源组的详细操作步骤,请参见通用参考:修改资源组归属工作空间。

1.2.6.3. 新增和使用自定义数据集成资源组

使用自定义数据集成资源组运行数据同步任务,能够帮助您解决DataWorks公共资源组与您的数据源无法连通的问题,实现更高速的传输。本文为您介绍如何新增并使用自定义数据集成资源组。

前提条件

• 购买DataWorks

仅DataWorks专业版及以上版本支持使用自定义数据集成资源组。您需要提前购买DataWorks专业版及以上版本,购买详情请参见版本服务计费说明。

• 已有自有服务器资源

使用自定义数据集成资源组之前,您需要已有自有服务器资源。本文以使用ECS服务器示例,为您介绍使用 自定义数据集成资源组之前与自有服务器相关的准备工作,详情请参见准备工作。

使用限制

• 权限限制

仅拥有**空间管理员**角色权限的用户支持新增或修改自定义数据集成资源组。如果指定用户需要新增或修改 自定义数据集成资源组,则可授权该用户为**空间管理员**角色,详情请参见角色及成员管理:空间级。

● 与ECS服务器相关的限制

- 自定义数据集成资源组所在的ECS服务器的时间与当前互联网时间差必须在2分钟之内,否则会导致部署 的资源组服务请求接口超时,服务异常,无法执行任务。
- 一台ECS机器只能添加于一个自定义数据集成资源组,每个资源组只能选择一种网络类型。
- 使用范围限制
 - 自定义数据集成资源组只能给当前工作空间作为数据同步资源组使用,不会显示在资源组列表页面。
 - 自定义数据集成资源组不支持给手动业务流程的数据同步节点使用。

准备工作

1. 购买云服务器ECS。

购买ECS云服务器的具体操作请参见购买ECS云服务器。

- ? 说明
 - 。请使用CentOS 6、CentOS 7或Aliyun OS操作系统。
 - 如果您添加的ECS需要执行MaxCompute引擎任务或数据同步任务,则需要检查当前ECS的 Python版本是否为Python2.6或2.7(Cent OS 5的版本为Python 2.4,其它OS自带Python 2.6 以上版本)。
 - 请确保ECS具有访问公网的能力,您可以使用 ping www.aliyun.com 命令查看连通性。
 - 建议ECS的配置为8核16G。
- 2. 查看ECS主机名和内网IP地址。

登录云服务器ECS控制台,在实例页面,查看购买的ECS主机名和IP地址。主机名和IP地址用于后续自定义数据集成资源组添加ECS服务器。

云服务器 ECS	云服务器 ECS / 实例						实例使用须知
板 流	实例						
事件	(別次例 识別搜索 · 选择实例属性项搜索, B	和 《 · · · · · · · · · · · · · · · · · ·				105	「「「「「「」」」」 (1) (1) (1) (1) (1) (1) (1) (1) (1) (1)
自助问题排查	— 实例D/名称	标签 监控 可用区 🛛	IP地址	状态 🔽	网络类型 🔽 配置	付薨方式 ♡	操作
发送命令/文件 (云助手) 应用管理 🚥		♥ ♥♥ 凸 青島可用区:	1000	●日停止	专有网络 1vCPU1GiB (I/O优化) ecs.un4.small 1Mbps	按量 2020年1月17日 19:57 创建	管理 升降配 更改实例规格 更多 マ
实例与摘象 へ		應码 续费 按量付靠特包年包月	释放设置 更多。			共有1条,有	顶显示: 20 • 条 《 (1) 》
案例							
18:9							

3. 开通8000端口。

开通8000端口便于您读取日志,定位问题。

? 说明

- 如果您的ECS是VPC专有网络类型,则无需开通8000端口。下述步骤仅适用于经典网络。
- 如果您发现alisatasknode日志中有超时报错信息 response code is not 200,通常是因为 某个时间段访问服务接口不稳定的异常导致。如果异常持续时间小于10分钟,自定义资源组 服务器依然可以正常服务。您可以查看日志 /home/admin/alisatasknode/logs/heartbeat.l
 og 进行确认。
- i. 进入安全组页面。

进入ECS服务器的安全组列表页面。

ii. 单击目标安全组操作列的配置规则。

iii. 开通8000端口。

在**安全组规则 > 入方向**页签,单击**手动添加**,设置端口范围为8000/8000,授权对象为数据集成的固定IP。

使用前缀列表提高安全	追规则管理的效率, <u>查</u>	潘最佳实践 🚥					
安全組成別 安全組成別 安全組内案所列展 安全組内弾性の中	基本信息 安全组ID/名称: 安全组关型: 描述: 资源组:	普通安全组	ng yang disebut yang		网络: 相方法道照明: 相方正道 Q型: 标签:	的目的時間	
Q	访问规则	1. 号入安全组成列 3. 出方向 (快速反加) ()	, 号出 ②健康检查 編入調口或者授权対象进行推索	」 【 米由井	v		
	授权策略	优先级 ①	协议类型 自定义 TCP	9月日范閣 ① → 目的: 8000 ×	授 权対象 ○ • 遭: 输入cidr、安全指id	描述 或者前部列表:d	操作 (保存) 预流 删除
	⊘ 允许	110	自定义 TCP	10.00	A	Lynn contact on	影於
	⊘ 允许	110	自定义 TCP	the second	Research 1	1.000-0.000-0.0	2019
	⊙ 允许	110	全部 ICMP(IPv4)	100.00	B	1,000	2079

iv. 单击保存,成功开通8000端口。

操作流程

使用自定义数据集成资源组的主要操作流程如下:

- 1. 自定义数据集成资源组使用前准备工作
- 2. 新增并配置自定义数据集成资源组
- 3. 使用自定义数据集成资源组

自定义数据集成资源组配置完成后,可能出现的常见问题及解决方案,可参见<mark>常见问题</mark>;后续对资源组的相关 管理操作,可参见<mark>管理自定义数据集成资源组</mark>。

新增并配置自定义数据集成资源组

- 1. 进入数据集成页面。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 选择工作空间所在地域后,单击相应工作空间后的数据集成。
- 2. 进入自定义资源组管理。

在数据集成页面左侧导航栏,单击配置选项 > 自定义资源组,进入自定义资源组管理页面。

注意 该页面不显示公共资源组,仅显示您添加的自定义数据集成资源组。

- 3. 创建并配置自定义数据集成资源组。
 - i. 在自定义资源组管理页面,单击右上方的创建自定义资源组。

ii. 配置资源组基本信息。

a. 在创建自定义资源组 > 创建资源组页签, 输入资源组名称。

⑦ 说明 数据源名称必须以字母、数字、下划线(_)组合,且不能以数字和下划线(_) 开头。

- b. 单击下一步。
- iii. 资源组添加服务器。

创建自定义资源组		×
创建资源组	添加服务器 安装Agent 检查连通	
* 网络 <u>类型</u> : 服务器1	● 阿里云经典网络 🔷 专有网络 🛛 🤡	
*服务器名称:	请输入真实主机名称,不可自定义	
* 机器IP:	请输入内网机器IP	
* 机器CPU (核) :		
* 机器内存 (GB) :		
添加服务器		
	上一步	步

参数	描述
	包括阿里云经典网络和专有网络。
网络类型	⑦ 说明 除华东2(上海)地域支持经典网络外,其它地域仅支持专有网络。
	输入 服务器名称 。您可以登录ECS,执行 hostname 命令,获取服务器名称。
服务器名 称	⑦ 说明 仅当网络类型 选择阿里云经典网络时,需要配置该参数。

参数	描述		
ECS UUID	输入ECS UUID。您可以登录ECS,执行 dmidecode grep UUID 命令,获取ECS UUID 。		
	⑦ 说明 仅当网络类型 选择专有网络时,需要配置该参数。		
机器IP	请输入内网机器IP。		
机器 CPU(核)	建议自定义资源组机器CPU配置至少为4核。		
机器内存 (GB)	建议自定义资源组机器内存配置至少为8 GB RAM和80 GB磁盘。		

- b. 单击下一步。
- iv. 安装Agent。
 - a. 根据安装Agent对话框中的提示,依次操作。

创建资源		加服务器	安装Agent	检查连通	
		_			
gent5-dill安全的	H生Linux机器上,添加B	加與个服勞體都需要初	题代。		
口果是新添加林	几器, 清按照如下步骤	最作			
tep1: SSH登	录ECS服务器,保持在	oot用户下;			
tep2: 执行命	+ wget https://di-ag	the local	and the second second	s.com/install.sh -no-	
heck-certifica	te -O install.sh 复制				
tep3: 执行命	🕸: sh install.sh -user	_name=z			
assword=		-enable_uuid=true	-region_id=cn-shangh	ai-bucket=di-agent-	
-		+8deploy_locale=zh_	CN 复制		
step4: 稿后在	添加服务器页面,点击	刷新按钮,观察服务状	犬态是否转为可用状态。		
step5: 请开通	服务器的8000端口。	(专有网络类型可以不持	打开此端口)		
			E-	-# T-#	
~					
? 说明	月				
			ᅀᄜᄖᄲᆃᇔ	五千がわた 仮	高玉 <u>七</u>
	如未执行 ins	stall.sh	之时 出 宿 虬 需	安里新执行,恣	需安住 install.sh
	一个目录下先	执行 rm -rf	install.sh	命令,删除已约	至生成的文件,再执辞
	tall sh				
	0				
	初始化界面对	于每个用户的]命令都不一样	,请根据自己的	初始化界面执行相关
	< 0				
	~ 0				

b. 单击下一步。

v. 检查连通性。

在检查连通页签,单击刷新,确认服务器的当前状态。

- 当服务器状态为可用时,表示资源组和服务器连通,您可以正常使用资源组。
- 当服务器状态为不可用时,表示资源组和服务器未连通,您需要根据实际情况排查处理。
- 4. 单击完成,成功创建并配置自定义数据集成资源组。

新增并配置完成自定义数据集成资源组后,可能出现的问题与解决方案,详情请参见常见问题。

使用自定义数据集成资源组

自定义数据集成资源组创建并配置完成后,您可以通过如下方式切换任务使用的资源组为自定义数据集成资源 组。

⑦ 说明 仅离线同步任务支持切换自定义数据集成资源组。



操作环境	支持的切换操作	界面入口
切换开发环境资源组	 单个任务切换 批量切换 	<section-header><section-header><section-header><section-header></section-header></section-header></section-header></section-header>

管理自定义数据集成资源组

自定义数据集成资源组创建并配置完成后,您可以在自定义资源组管理页面,查看资源组的网络类型、绑定的服务器等信息,或对目标资源组执行服务器初始化、管理、删除等操作。

≡	① DataWorks 数	昭集成 🧖 🗸				ር 🗷 🔍 🐬 🚃 🗄
00	首页	自定义资源组管理 请输入资源组名称进行搜索				创建自定义资源组
۲	数据源	帝秦祖名称	网络带型	服金編	攝作	
Ġ	同步任务				22/7 12 million (c) and 10 million	
ø	任务运维	xc_datax_demoz	TE-FQ-PS-FR		SCHWHURTIG BIE SDIF	
野	副普选项 ^	xc	-		服务器初始化 管理 删除	
	认证文件管理	test]_1			服务器初始化 管理 删除	
	实时同步DDL消息处理。	xhb_test			服务器初始化 管理 删除	
	THE COUNTY	xc_dataxs			服务器初始化 管理 删除	

管理:用于查看资源组所绑定服务器的IP、状态、资源使用率等信息。您也可以修改、删除资源组所绑定的服务器,或为该资源组添加新的服务器。具体操作,详情请参见新增并配置自定义数据集成资源组的添加服务器步骤。

? 说明

- 资源使用率不为 0% 时,表示该资源组当前的服务器中存在运行中的任务。
- 资源组添加新服务器后,必须执行**服务器初始化**操作。
- 服务器初始化:资源组添加服务器后,需要执行初始化操作。

单击**服务器初始化**,按照所示步骤进行初始化。

服务器初始化	×
Agent只能安装在Linux机器上,添加的每个服务器都需要初始化。	
如果是新添加机器,请按照如下步骤操作	
step1: SSH登录ECS服务器,保持在root用户下;	
step2:执行命令: 复制	
step3: 执行命令:	
	1
step4: 稍后在添加服务器页面,点击刷新按钮,观察服务状态是否转为可用状态。	
step5:请开通服务器的8000端口。(专有网络类型可以不打开此端口)	
×	闭

● 删除: 单击删除, 即可删除目标资源组。

⑦ 说明 DataWorks不支持删除存在运行任务的资源组,执行删除操作前,您需要确保目标资源组中不存在运行状态的任务。

您可以进入**运维中心 > 周期任务运维 > 周期任务**页面,通过资源组名称进行筛选,查看资源组中任务的运行状态,详情请参见查看并管理周期任务。

常见问题

新增并配置完成自定义数据集成资源组后,如果添加的ECS服务器一直处于停止状态,则可能会出现下图中的问题。

"alisatasknode.log" 3937L, 445471C	3937,2-9	Bot
24 more		
. java:32)		
at com.alibaba.alisa.common.util.PropertyUtils.get	Property(Prope	rtyUtils
at java.util.ResourceBundle.getString(ResourceBund	le.java:407)	
at java.utii.ResourceBunale.getUbject(ResourceBuna	le.java:450)	
va.util.PropertyResourceBundle, key alisa.node.host.name		<u> </u>
Caused by: java.util.MissingResourceException: Can't find	resource for b	undle ja
at com.alibaba.alisa.node.server.StartUn.main(Star	tlln. java:24)	
nit>(ClassPathXmlApplicationContext.java:93)		
at org.springframework.context.support.ClassPathXm	lApplicationCon	ntext. <i< td=""></i<>
nit>(ClassPathXmlApplicationContext.java:139)	* *	
at org.springframework.context.support.ClassPathXm	lApplicationCon	ntext. <i< td=""></i<>
h(AbstractApplicationContext, java:425)		
at org.springframework.context.support.AbstractApp	licationContex	t.refres
BeanFactoruInitialization(AbstractApplicationContext.java:	895)	
at org.springframework.context.support.AbstractApp	licationContex	t.finish
nreInstantiateSingletons(DefaultListableBeanFactory, java:5	85)	actory.
at org springframework beans factory support Defau	ltListableBean	Factoru
(AbstractBeanFactory java: 192)	ac the anit ac tory	. ge ebean
all (nustratiucaliratiory, java-200) at any enningframework beans factory sunnant Abstr	actBeanFactoru	getBean
at org.springiramework.beans.iactorg.support.Hostr	actbeanractory	.αυαειβε
y.getSingleton(DeraultSingletonbeannegistry.java.222)	actBeanFactory	doCatPa
at ory.springiramework.beans.iactory.support.beiau	rtsing reconbea	inegistr
at ong spningfpameuonk beans factory support Defau	1+SingletonBeau	Regista

您需要按照如下步骤排查处理:

- 1. 切换至Admin账号。
- 2. 执行 hostname -i 命令, 查看Host的绑定情况。
- 3. 执行 vim/etc/hosts 命令, 添加IP地址和主机名。
- 4. 刷新页面服务状态,查看ECS服务器是否注册成功。
 - 如果刷新后ECS服务器为启用状态,则问题已解决。
 - 如果刷新后ECS服务器仍为停止状态,请按如下步骤重启alisa。
 - a. 切换至Admin账号。
 - b. 执行 /home/admin/alisatasknode/target/alisatasknode/bin/serverctl restart 命令重启 alisa。

⑦ 说明 命令中涉及您的访问密钥,请勿轻易泄露。

1.2.7. 通用参考: 修改资源组归属工作空间

独享资源组与自定义调度资源组在购买、新增配置完成后即可用于所属工作空间内的任务使用。如果您想修改 资源组归属的工作空间,您可根据实际场景,参考本文进行修改操作。

背景信息

对于已绑定工作空间的资源组,如果您想修改资源组的归属工作空间的话,您可以在资源组列表页面解绑资源 组之前绑定的工作空间,重新绑定。解绑操作可参见本文的<mark>解绑工作空间</mark>部分。 如果资源组正在被已发布的任务使用中,您无法直接将资源组与归属的工作空间解绑,您需要先修改任务使用的资源组,确保资源组没有被已发布的任务使用后,再解绑。修改各类任务使用的资源组入口如下。



具体操作可参见修改任务使用的资源组和修改实时同步任务使用的资源组。

修改任务使用的资源组

如果资源组正在被已发布的任务使用中,您无法直接将资源组与归属的工作空间解绑,您可以在运维中心页面 修改各类任务使用的资源组,其中:

- 实时同步不支持在运维中心页面修改,其他任务均支持在运维中心页面修改。实时同步任务修改使用资源 组的操作可参见修改实时同步任务使用的资源组。
- 手动离线任务不支持批量修改,其他任务均支持批量修改。

周期性任务与手动任务的操作类似,下面以周期任务为例,为您介绍详细的操作步骤。

e	运输大屏				2							
\$\$\$	实别任务运输	~	搜索:	节点名称/节点D Q	节点类型: 可选择	节点类型	✓ 责任人: d	dataworks_demo2 V	编度资源组 xc_test ~	🛃 我的节点] 今日修改的节点	 1 暂停 (冻结) 节点
1	同期任务运输	^	- 14	立节点 🗌 过期节点								
	周期任务											C 刷新「展开搜索
	周期实例		3	名称	市市	ID	修改日期 11	任务类型	责任人	V0.92.743.003	调度资源组 5	操作
	MUSEM			xc_权限测试_test	7000	002563368	2020-11-11 14:06:0	03 ODPS SQL	dataworks_demo2	日调度	xc_test	DAGES I SHUC I ANSOLA I DESS
Q	手动任务运输	~		xc_mongoreader_mongowriter	7000	002531653	2020-09-18 09:59:4	45 离线同步	dataworks_demo2	日调度	xc_test	DAG語「現版」科教語「更多
*	1968年19	~		xc_流程参数_数据集成	7000	002567740	2020-09-18 09:59:4	45 高线同步	dataworks_demo2	日调度	xc_test	DAG题丨测试丨补数据丨更多
ų,	121630-INF			xc_createTbale	7000	002570212	2020-09-18 09:59:4	45 ODPS SQL	dataworks_demo2	日褐度	xc_test	DAG题丨例试丨补数据 丨更多
				xc_mysql_向导模式	7000	002570284	2020-09-18 09:59	45 离线同步	dataworks_demo2	日调度	xc_test	DAG語 詞版 补数据 更多
				xc_pyodps部件外发	7000	002572329	2020-09-18 09:59:4	45 Py0DPS 2	dataworks_demo2	小时调度	xc_test	DAG题丨例试丨补数器 丨更多
				xc_ots_reader	7000	002573897	2020-09-18 09:59.4	45 高线同步	dataworks_demo2	日调度	xc_test	DAGES STAL HINDER BLS
				xc_赋值节点下游_shell	7000	002472181	2020-09-18 09:59:4	45 Shell	dataworks_demo2	日调度	xc_test	DAG器「例は「計数据 更多
				xc_赋值节点_上游	7000	002604960	2020-09-18 09:59:4	45 赋值节点	dataworks_demo2	日调度	xc_test	DAG图 测试 补数器 更多
			4	- 4								
			193	责任人 修改编座资源组	修改数据集成资源		\$ (汤结) [续3	夏(解冻) 夏多			< 1/1 >	每页显示: 40 🗸 🗸

- 1. 进入运维中心页面。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏,单击工作空间列表。
 - iii. 在对应工作空间的操作列单击进入运维中心。
- 2. 单击周期任务运维 > 周期任务, 进入周期任务页面。
- 在顶部筛选节点、资源信息后,选择待修改的任务,单击底部的修改调度资源组或修改数据集成资源组。
 组。
- 4. 在弹出的页面中,根据界面提示选择新的资源组,单击确认,完成资源组修改。

修改完成后,您可将资源组绑定到其他工作空间上。

修改实时同步任务使用的资源组

修改实时同步任务使用的独享数据集成资源时,您需要在**批量操作-数据开发**页面中修改,支持批量修改。

≡	n DataWorks DataStudio	•								9节点配置 🕹 任务发	艾布 の 跨項目!	的难 🖉	运输中心 众 团	<i>थ</i> , 🕈 🕴
Ø	ыяжия A Bala Straws	○ 批量操作	E数据开发 ×											⊚ ≡
*			光波 的数											历史操作记录
Q	> ND決方室 日 1997年1999 日 1997年1999 日 1997年1999 日 1997年1999 日 1997年1999 日 1997年1999 日 1997年1999日日 1997年1997年1997日日 1997年1997年1997日日 1997年1997年1997日日 1997年1997年1997年1997年1997年1997年1997年1997	1972-	連续λ 节点名称/ID		T-688-	tiate	1		建洗探心名法程				建选择事任人	
©		调度资源	 請选择環胞資源组 		引搬突型: 浙江				请选择引整实例		× 200		通過拝数振集成资源は	1 V
â			i请选择数据来源类型			2			请选择数据去向关键					~
≡			请选择节点提交状态											
₽			400	2679	11/32-2014		18/45 A	2188-879	21882104	412	17149-22-25141		Asiamitta IN	ARTHUR IN
fx		-		9044	32551042		Jan 12 A	Dete	51013003		PERIODICIPAL		2020 11 22	2020 11 22
亩				实时同步				Integration		未提交	公共調度資源	8a	18:03:22	18:03:22
			2001	实时同步	Section 2.	_hour_20		Data Integration		未提交	公共調度資源	HE	2021-05-12 14:54:22	2021-05-12 14:54:22
			AAAA	家时间步	-		-	Data Integration		未提交	公共調度資源	80	2020-08-12 16:15:44	2020-08-12 16:15:44
			asaMc	实时同步				Data Integration		未提交	公共調度資源	89	2021-04-26 16:58:20	2021-04-26 16:58:20
				实时同步	-	.mysql_to		Data Integration		未提交	公共調度資源	He	2020-06-03 11:00:52	2020-06-03 11:00:52
				实时同步	access*	olo_rdsm		Data Integration		未提交	公共調度资源	Re	2020-06-10 11:25:49	2020-06-10 11:25:49
				实时同步				Data Integration		未提交	公共調度资源	HE	2020-03-23 11:23:26	2020-03-23 11:22:46
				家时同步	-	务流程		Data Integration		未盛交	公共調度资源	880.	2020-06-01 16:00:04	2020-06-01 16.00.04
				实时同步				Data Integration		未提交	公共調度資源	Re	2021-02-07 21:19:10	2021-02-07 21:19:10
~				实时同步				Data Integration		日发布	公共調度資源	He	2021-03-03 20:12:20	2021-03-03 20:10:16
				家时同步		停改调整 停改数据	201068 E#462010568 5	Data Integration		已发布	公共調度資源	R80	2021-02-24 20:21:14	2021-02-24 20:17:16
				实时同步		修改调整		Data Integration		未提交	公共調度資源	題	2020-11-02 17:05:02	2020-11-02 17:05:02
				家时同步		停改调整	tan	Data Integration		已发布	公共調度资源	Ria	2020-10-28 00:21:16	2020-10-27 20:38:50
				实时同步	-	特改调题 特 特改数据	EEEEDEDTED KARE	Data Integration		未提交	公共調度資源	Re	2021-04-29 14:35:32	2021-04-29 14:35:32
۲						85 8								· · · ·
≡		已选择2个	选择全部(242) 提交	传改责任人	使改引擎实例	更多	<u> 4 </u>		4	上一页] 2	3 4 …	- 13	下一页 > 1/13	到第二页 梁定

- 1. 进入DataStudio页面。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 在对应工作空间的操作列单击进入数据开发。

2. 单击数据开发工具栏的 式 按钮,进入批量操作-数据开发页面。

3. 筛选节点类型为实时同步,选择待修改的任务,单击底部的更多 > 修改数据集成资源组。

4. 在弹出的页面中,根据界面提示选择新的资源组,配置其他相关信息,单击**确认**,完成资源组修改。 修改完成后,您可将资源组绑定到其他工作空间上。

解绑工作空间

如果独享资源组和自定义调度资源组没有被已发布的任务使用,您可以在控制台的资源组列表解绑工作空间, 并重新绑定工作空间即可。

⑦ 说明 如果无法直接解绑资源组的工作空间,说明该资源组正在被已发布的任务使用,您可以先修改任务使用的资源组,将任务运行在其他资源组上,修改完成后再解绑工作空间。

Ξ	(一)阿里云 华东1 (杭)	H) -		Q 搜索文档、控制台、API、解决方案和资源	费用 工单 备案	企业 支持 App 🖸 🗳 🗑 简体 🕚
Data	Works	DataWorks / 资源组列表	修改归属工作空间		×	产品动态 帮助文档
15.07			dogfood0904002	未占用	师定 🔺	
MACO		独享资源组 公	dataworks_project_0827	朱占用	绑定	
1 12	2间列表	创建独享资源组	dogfood0909	未占用	绑定	c
建設相	的表	100-100-101-07-00-142.2+	DWLabClone_2381696	朱占用	绑定	1947
报警告	記録	new 0112	DWLabClone_9538265	未占用	绑定	
开放平	P台	new_0112	DWLabClone_3539762	朱占用	绑定	修改扫描工作空间
11303	華列表 ^	eileen_test_ziyuanzu eileen_test_ziyuanzu	DWLabClone_9185996	未占用	郷建	查看信息 运输助手 扩容 缩容 续费 网络设置 修改归属工作空间
м	axCompute	zhenshui_batch_group	DWLabClone_0156001	未占用	鄉定	查看信息 扩容 缩容 续奏 网络设置
G	raph Compute	zhenshui_batch_group	DWLabClone_4738169	未占用	郷定	修改归属工作空间
н	MaxCompute Graph Compute Hologres	zhenshui_real_group zhenshui_real_group	DWLabClone_7527960	未占用	绑定	查看信息 扩容 缩容 续费 网络设置 2. 修改归属工作空间
		kjk	DWLabClone_8857868	未占用	绑定	查看信息 扩容 缩容 续费 网络设置
		jk	test_delete_project001	未占用	網定	修改归屬工作空间
		wengzi_test_resgroup 独享资源组创建	xd_test_mc5_1	朱占用	鄉定	查看信息 运输助手 扩容 缩容 续费 网络设置 修改归屋工作空间
			zhenshui_test_hz	已占用 3	NU6	8
					取消	Ţ

- 1. 主账号登录DataWorks控制台。
- 2. 在左侧导航栏,单击资源组列表,并选择相应地域。
- 3. 单击相应资源组后的修改归属工作空间,在新页面中单击解绑。

1.2.8. 通用参考: 取消包年包月产品的自动续费

如果您开通的包年包月产品后续无需再使用,则可以配置相应产品取消自动续费或退订。本文为您介绍取消自 动续费及退订的相关操作。

使用限制

仅开通自动续费功能的包年包月产品支持取消自动续费。

取消包年包月产品的自动续费

1. 进入续费管理页面。

登录DataWorks控制台,在顶部导航栏右侧,单击费用>续费管理,进入续费管理页面。

2. 取消目标产品的自动续费。

(-) 阿里云	a Ins				Q 搜索		興用 工单	ICP 餐宴 全业	支持 App 🖬	0 H ())简体
	续赛管理						其他产品续费:	18489 183	云虚拟主机 企业	EMI MERE	静助文档
	您可能想了解:如何读者实例?如何自动读着? 1.已过期的资源:新读费周期从读费当前时间开始计算 2.对于配置了自动读着,且9天内到期的实例,如果想	【:已过期的实例,不允许设置/修改自动续费。 否试修改自动续奏周期或设置为不续要或设置为不自动映要,1	可能会因实例已自动	读要完成导致当前读要关型5	迎来生效的情况。						
## 全部菜单项 >	突例10 、多个时间以适号分隔	秋志: 运行中、已过期 >	搜索						云服券	AECS85(5-3)10	160 2 ()
资金管理	时间: 全部 7天内到期 15天内到期 30天内;	日期 2个月内到期 3个月内到期	1								
收支明细 ▼	产品: 全部 DataWorks(現享资源(包年包月)25 地域: 全部地域	实时1 nk版 7 DetaWorks谭值版本(包年包月)7	云服务器ECS(包月	B) 5 EAS专尾机器预付费	3 关系型数据库 3						
所单管理 月所单概近	手动续费 2 自动读言 19 多期不续费										
账单详情	严恶	实例10/实例名称	地域	案例状态	例计时	付農方式	开始/结束时间	续赛调期			操作
成本管理 预算管理 Nee	DataWorks没享武康(包平包月)	1945.1949.00000000	华东2 (上海)	运行中	9天	包年包月	2021-03-24 13-26-06 2022-02-25 00:00:00	5 1个月	续赛 伊改自动快速		STIRE.
成本分析	DataWorks投享资源(包年包月)	-	华东2 (上海)	运行中	10天	包年包月	2020-09-25 16:11:34	4 1个月	续费 修改自动续获	8 77 988 1588	【手动续费
分账管理 财务单元	DateWorks投算更漂(包年包月)	2010/01/01/01/01/01	至]]积7个3 9 为了報約	(天教) (夏好地进行资源管理,我)]体旧会为您发送到期期	1 韓信息,但是会在挑	躍中对不嗓震项进行标	识.	续要 修改自动成功	8 不续题 恢复	《手动续奏
▼	DataWorks投享资源(包年包月)		9 9	2週不续費、交例到期待机ま	#且只会問題一次,忽而	可以在停机前随时停改	149 2.		续奏 修改自动续	₽ 不喷费 恢复 [:]	(#10)
我的订单 续费答理	DateWorks投享资源(包年包月)	CONTRACTOR AND	\$ 实例ID/3	9. 朝廷议里为于两个体数 实例名称	-31	到期时间	/	街计时	续赛 修改自动续制	8 不续器 恢复	(手动续费
★ 卡券管理	DataWorks投享资源(包年包月)		4		1.000	2022-02-25	5 00:00:00	9天	续赛 停改自动续	₩ 不 球器 恢复	【手动续奏
提近弊管理 (P-c-的管理	DateWorks投享资源(包年包月)	CONTRACTOR OF STREET,	3				御认	取消	续奏 修改自动映	A 不读器 恢复	(手动疾病
- i. 在续费管理页面,您可以选择DataWorks独享资源(包年包月)或DataWorks增值版本(包年 包月)产品类别,单击自动续费,即可查看对应类别下已开通自动续费功能的产品列表。
- ii. 单击目标产品操作列的不续费,在弹出的到期不续费对话框单击确认,即可取消该产品的自动续费 功能。

⑦ 说明 为了帮助您更好地进行资源管理,取消自动续费后,系统依旧会为您发送到期提醒信息,但是仅会提醒一次,并且在提醒中对不续费项进行标识。实例到期后将会停机,您可以在停机前随时修改该设置。

取消指定独享资源组的自动续费

- 1. 获取资源组的实例名称。
 - i. 登录DataWorks控制台,在左侧导航栏单击资源组列表,默认进入独享资源组页面。
 - ii. 单击目标资源组操作列的查看信息,获取资源组的实例名称。

	工作会 🔤 华东2 (上海) 🗸				Q 建汞		鹅用 工单 ICP 餐館	企业支持 App 🖾 🕻 🗑 180年 🚱
DataWorks	DataWorks / 资源组列表							产品动态 制能文档
概道	被享受原相 公共资源组 自定义资源组							
工作空间转换	· 新建水素系统的 · 新建水素系统的 · 新建水素系统的 · 风							C
変通相列表 ジンドラ シン	资源组名称/做注	英型	8	to 7	95(\$905)B)	歌廳	使用車	操作
开放平台 四四	and the second s	数编程务法源	il ~	/运行中	2022-03-20		0%	查看信息 升配 续费 修改归雕工作空间
计算引撃列表 ン 数据結集成(地合一体)	Statistics and	调度洗滞运		/ 遗行中	2022-03-10	2	0%	重要信息 运输助手 扩容 燒容 埃爾 网络设置 仲次归属工作空间 变更成格
	COMM.	投享资源组 宗语记文处	×	运行中	2022-03-23	1	0%	查看信息 远线划手 扩容 燒容 烘费 网络设置 伸改归属工作空间 交更规格
	17 anno	资源组D: 资源组类型: 数据服务资源组		运行中	2022-03-15			查看信息 扩容 综合 统备 网络设置 修改旧雕工作空间 变更现物
	an and a second s	订单号: 实例名称:		运行中	2022-03-14	2	0%	查看信息 运续起承 扩容 熔容 续费 网络设置 传改归属工作空间 变更成格
		实例规格: VPC:		启动中	2022-01-24		0%	升配 娘蠢 傳改扫描工作空间
	surfatures.	VSwitch:		运行中	2022-03-12	2	0%	查看信息 扩容 缩容 续奏 网络设置 传改旧属工作空间 实现现情
	Concession of the second		关闭	运行中	2022-03-04	1	0%	重要信息 扩容 编容 续奏 网络设置 修改归属工作空间

2. 进入续费管理页面。

登录DataWorks控制台,在顶部导航栏右侧,单击费用>续费管理,进入续费管理页面。

3. 取消资源组的自动续费功能。

三(一)阿里云	☆ 工作台	Q. 提訳	第月 工単 ICP 敏震 企业 支持 App 🖾 🗘 🗑 第体 🌔
Illouino			Normann: Irann me fidhigh glann in Mark
	ロシリング、シングの構成をひ、シングの構成をし、シングロング、 1. 已は期的支援、新規機関制以構築制成が開始が同分抽片算; 已は期的支付,不允许必要/律政目的構 2. 対于配置了自动情景,且9天内到期的支付,如果必要法律政目的情景問題成及量为不供表成交付	8. 重为不自动映器,可能会因实例已自动映真完成导致当能映真地型变更未主效的情况。	
₩ 全部菜单项 >	実例の - 通行中、日辺期	2	云級务職和CS的统一到明日设置()
首页 汚合管理	时间: 全部 7天内到期 15天内到期 30天内到期 2个月内到期 3个月内到期		
收支明细	产品: 金部 DetaWorks版革定语(包年包月) 25 实时计算Flink版 7 DetaWorks潜道版	: (包年包月) 7 二 云陽朱器ECS(包月) 5 EAS专権机器技行義 3 二 共長型設備率 3 11000000	
账单管理 月账单概返	地域:全部地域 手动映器 <u>自动映画</u> 1 王JJ和不依要		
账单洋情	产品 实例10/实例名称	地域 实例状态 倒计时 付费方式	开始//选来时间 续腾闯期 操作
成本管理 预算管理 New	DataWorks限享定源(包年包月)		2019-12-13 14:39:23 3个月 地面 超双码记录器 不成面 3 印度用
成本分析	透中 0 项 批量续要 传改自动续要 恢复为手动使费 设置为不续要	马出惊颤: 到期不续费 ×	每页型示: 20 ~ 共有1条 〈上一页 1 下一页 〉
分账管理		为了和世界并指由于我用意思。我们这些自分世际发展间围建度里。但是会们出现中方不得是也许可杀。 如果你必要不确定,如果你的同时,却能只会相关一次,也可以是可能成都可考虑的问题。	(
財务单元 ▼		UTITOR WEREAMEN	
订单管理		Exhibitin Exhibitin 2022 04-27 00 00 00 70所	
我的订单		- 10.00	

i. 在续费管理页面的实例ID区域, 输入步骤1获取的资源组实例名称, 单击搜索, 显示目标产品信息。

ii. 单击目标产品操作列的不续费,在弹出的**到期不续费**对话框单击确认,即可取消该产品的自动续费 功能。

⑦ 说明 为了帮助您更好地进行资源管理,取消自动续费后,系统依旧会为您发送到期提醒信息,但是仅会提醒一次,并且在提醒中对不续费项进行标识。实例到期后将会停机,您可以在停机前随时修改该设置。

1.2.9. 通用参考: 按量付费转包年包月

开通DataWorks后,默认享有DataWorks按量付费的基础版相关功能服务,您可使用该类服务执行任务。同时,DataWorks支持包年包月的增值版本及资源组服务。后续当您需要更丰富的功能时,可转换按量付费为包年包月。

背景信息

• 按量付费方式介绍

开通DataWorks后,默认拥有按量付费的公共资源组使用权,任务类型及资源使用情况如下表。

任务类型	所用资源组	资源组介绍及计费说明
调度任务	公共调度资源组	 公共资源组概述 公共调度资源组计费说明:按量付费
数据集成任务	公共数据集成(调试)资 源组	 公共资源组概述 公共数据集成(调试)资源组:按量计费
调用数据服务API	公共数据服务资源组	 公共资源组概述 数据服务公共数据服务资源组计费说明:按量付费

? 说明

- 除上述服务外,DataWorks还包含其他产品计费类型,部分产品仅支持通过按量付费方式提供服务,您可根据业务需求合理选择,详情请参见计费逻辑说明。本文以版本产品服务和资源组为例进行服务升级说明。
- DataWorks基础版支持的产品能力,详情请参见各版本功能支持详情。

● 按量付费升级为包年包月说明

按量付费升级为包年包月服务,本质为前期使用按量付费方式执行任务,后续购买包年包月服务并以该方 式执行任务。您可参考本文进行服务升级及使用:

- o 升级为包年包月服务
- · 包年包月服务使用指导
- 费用情况说明

两种方式的计费详情则按照各自的使用情况进行计费。具体费用请以实际费用账单为准,查看账单,详情 请参见查看账单详情。

升级为包年包月服务

产品服务与任务运行使用的资源费用除默认的按量付费方式外,还可以根据对应产品能力与业务需要选择包年 包月的增值版本与资源组服务。不同类型任务可选择升级为相应的包年包月方式执行,具体如下表。

 ⑦ 说明 购买包年包月服务时您需要具
 有AliyunBSSOrderAccess及AliyunDataWorksFullAccess权限。授权详情请参见管控权限详情: 细 分RAM Policy。

产品及任务类型	升级说明	产品介绍及计费说明
版本服务	从基础版升级为增值版。 各版本支持的产品能力存 在差异,您可根据业务需 要选择对应的增值版本。	各版本功能支持详情。
调度任务	从按量付费的公共调度资 源组升级为包年包月的独 享调度资源组。	 独享调度资源组概述 独享调度资源组计费说明:包年包月
数据集成任务	从按量付费的公共调数据 集成(调试)资源组升级 为包年包月的独享数据集 成资源组。	 独享数据集成资源组概述 独享数据集成资源组计费说明:包年包月
调用数据服务API	从按量付费的公共数据服 务资源组升级为包年包月 的独享数据服务资源组。	• 独享数据服务资源组• 独享数据服务资源组计费说明:包年包月

包年包月服务使用指导

产品服务及资源组升级后,您可根据如下指导使用。

使用指导	说明	参考链接
开通增值版本	增值版本的开通为地域级别,购买完成后,当前地域下所 有工作空间均可使用增值版本的产品能力。	增值版本产品能力详情
资源组网络绑定	独享资源组购买完成后,若存在特殊环境网络访问需求, 您需要为资源组进行网络绑定。	 独享数据集成资源组:网络连通方案 独享调度资源组:网络连通方案 独享数据服务资源组:网络连通方案。
	资源组绑定工作空间 :资源组购买完成后,您需要为资 源组绑定所需工作空间。	 独享数据集成资源组: 绑定归属工作空间 独享调度资源组: 绑定归属工作空间 独享数据服务资源组: 绑定归属工作空间。

使用指导	说明	参考链接
	 修改资源组:资源组配置完成后,您需要为目标任务切换新购买的资源组。 修改任务调度及运行资源组 切换任务执行时使用的资源组 您可以切换开发环境及生产环境的调度任务、数据集成任务执行时使用的资源组为新购的资源组。 	
使用资源组	⑦ 说明 切换数据集成任务使用的资源组时,请先确认切换后的资源组与数据库网络可连通。若资源组切换后由于网络问题导致任务失败,请参考配置资源组与网络连通对资源组进行网络配置。	
	 切换任务配置时的默认资源组 您需要进入数据开发界面,在设置 > 调度设置中修改任务使用的默认资源组,修改后,后续新建的任务 你默认使用该资源组。 	 修改任务调度及运行资源组 切换任务配置时的默认资源组 修改数据服务API调用资源组
	⑦ 说明 此处修改后,已创建任务现有的资源组配置不会改变。	
	 修改数据服务API调用资源组 您可以在数据服务API调用时将资源组修改为新购的包 年包月资源组。 	

1.3. 变更规格操作

如果当前资源组使用率过高,需要并发执行的任务数较大,当前资源组不能满足实际需求,您可以针对该资源 组进行升配;若当前资源组规格较大超过业务实际需要,您可以针对该资源组进行降配,以节省资源费用。本 文为您介绍如何对资源组进行规格变更操作。

规格变更介绍

规格变更是对当前产品进行升配或降配,其相关介绍如下表。

变更类型	变更内容	变更范围	费用说明	生效时间

变更类型	变更内容	变更范围	费用说明	生效时间
升配	仅会根据 已购级资, 知级格, 加 、资 源 组 数 量。	变更当前 资源组的 规格,资	您需要支付新增资源数量在升配生效时间点至原订单到期 时间点之间的差价。具体费用请以实际费用账单为准,查 看账单,详情请参见 <mark>查看账单详情</mark> 。	新配置将 在规格变 更订单支
降配	仅会根据 已购级级保证。 一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个	病 机器的 规格将统 一变更。	DataWorks会退回从降配生效时间点至原订单的到期时间 点之间的差价。具体费用请以实际费用账单为准,查看账 单,详情请参见 <mark>查看账单详情</mark> 。具体费用请以实际费用账 单为准,查看账单,详情请参见 <mark>查看账单详情</mark> 。	付后的20 分钟左右 生效。

使用限制

● 资源组限制

仅运行中状态的独享数据集成资源组、独享调度资源组支持规格变更操作。

• 权限限制

仅阿里云主账号,以及同时拥有AliyunDataWorksFullAccess、AliyunBSSOrderAccess权限策略的 RAM用户可执行规格变更操作。

• 产品限制

按量付费产品不支持规格变更。

• 订单操作限制

若独享调度资源组和独享数据集成资源组存在未完成的扩缩容、规格变更订单,则当再次执行扩缩容、规格变更操作时,会有弹窗提示您前往订单中心处理该订单,您需要将未完成的订单处理完之后才能继续执 行扩缩容、规格变更操作。

包年包月产品规格变更支持情况

包年包月产品规格变更支持情况如下表。√表示支持, ×表示不支持。

资源类型	升配	降配	说明
独享数据集成资源组	V	\checkmark	资源组内每个机器规格,可能影响当前单个任务 可配置的最高并发数。 如果您的同步任务数据量大,运行时间长,则可 通过如下方式缩短运行时长: 1.扩大资源组内机器规格,调整单个机器可同 时执行的并发数。 2.调大单个同步任务的并发数。
独享调度资源组	\checkmark	\checkmark	不涉及。
独享数据服务资源组	\checkmark	×	不涉及。

资源类型	升配	降配	说明
DATABLAU数据建模	\checkmark	×	仅支持对DAT ABLAU的并发登录用户数进行升 配,详情请参见 <mark>DAT ABLAU数据建模升配</mark> 。

规格变更操作步骤

独享调度资源组和独享数据集成资源组变更规格操作类似,本文以独享调度资源组的变更规格操作为例,为您 介绍如何变更资源组规格。

- 1. 进入资源组列表页。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击资**源组列表**, 选择目标地域。
- 2. 变更规格。
 - i. 在独享资源组页签, 鼠标悬停至目标资源组操作列的: 图标, 单击变更规格。

ii. 变更准备。

资源组变更						×
1 変更准	<u>S</u>	2 设置规格	3	支付		执行变更
支付完成后 起任务报警	,系统将挂起资源组并终止所有运行在该资 ; 3. 部分生产任务可能无法自动重跑,比如	;源组上的任务,然后执行资源组规格变更。]实时同步任务可能无法从停止位点继续,此	请务必确认影响并进行相关设置。 比类情况须手动进行任务启动/重跑	造成的影响包括但不限于: 1. 运行在该资 。; 4. 降配导致任务延迟。	源组上的任务被终止; 2.	. 可能引
资源组名称: 运行在该资源组上的	的任务,变更规格时会被自动终止: 调度	任务 (1)	英型: 独和	享调度资源组		G
任务来源 🖓	名称/ID	责任人 ₽	工作空间	开始运行时间	操作	
		• • • • • • • • • • • • • • • • • • •	(1000,000)	2022-03-29 00:13:07	查看 3 下一步	取消

a. 在变更准备页签,确认操作影响。

⑦ 说明 变更操作会对您的任务产生一定影响,请您仔细确认。确认后,决定是否继续执行变更操作。

变更操作产生的影响如下。

操作影响	描述
终止任务	变更操作需要终止正在运行的任务。DataWorks会统计当前资源组上正在运行的任务数及任务类型,您可以在 变更准备 界面查看并确认终止当前任务是否会对业务产 生影响。同时支持您基于任务责任人,任务所在工作空间等进行过滤目标任务。
引发报警	若被终止的任务存在报警规则,此次操作可能会触发报警。
部分任务无法自 动重跑	变更操作引发的任务终止,可能会导致部分任务可无法自动重跑。例如,实时同步 任务可能无法从停止位点继续运行,此类情况需要您手动启动任务或重跑任务并指 定同步位点,详情请参见 <mark>通用配置流程</mark> 。
降配导致任务延 迟	如果您进行的是降配操作,则会导致任务运行延迟。

b. 选择是否自动重跑被终止的任务,并勾选我已知晓。

选择是,则变更完成后DataWorks会自动重跑被终止的生产任务;选择否,则变更完成后 DataWorks不会自动重跑被终止的任务。

? 说明

- 生产任务包括:生产环境运维中心的实时同步任务、实时计算任务、周期调度任务和补数据任务,不包括DataStudio中临时运行的测试任务。
- 被平台终止的周期调度任务和在变更后自动重跑的周期调度任务将在实例操作记录
 界面分别产生操作记录。

c. 单击下一步。

iii. 设置规格。

在新规格下拉列表选择合适的规格,勾选服务协议,单击去支付。

⑦ 说明 资源组的规格变更,支持升配与降配,不同规格支持执行的任务数不同,详情请参考独享调度资源组不同规则支持的任务数及独享数据集成资源组不同规则支持的任务数。

资源组变更				×
✓ 变更准备	2 设置规格		3 支付	4 执行变配
 调度资源组的最 	大并发量等于规格乘以数量,等于可同时运行的任务数,更多信息请查看文档	ĺ.		
资源组名称:	100.000 (0.000)	英型:	44110	
实例ID:	type product means and product on the	到期时间:	2022年4月22日 10:41:56	
可用地域:	2.8	数量:	- F	
当前规格:	4c8g			
* 新规格:	16 vCPU 32 GiB 🗸 🗸 🚺			
*服务协议:	✓ DataWorks独享资源(包年包月)服务协议			3
				总配置费用 CNY 0 去支付 取消

iv. 支付订单。

单击前往支付页,在订单支付页面最终确认**订单概要**与**订单详情**,选择支付方式后单击**立即支** 付,根据界面提示完成支付。

资源组变更			×
変更准备	✓ 设置规格	3 支付	4 执行变配
	等待支付 支付完成后,系统将挂起资源组并终止所有运行在读述	t 騎往支付页 > 克源组上的任务,然后执行资源组织推变更。	
			取消

⑦ 说明 如果您取消支付,此次规格变更操作将不生效,不会对任务与资源组造成影响。如果 支付未成功,建议您进入订单中心取消该笔订单,避免对后续资源组扩容、缩容、变更规格操作 造成影响。

v. 执行变配。

支付成功后, DataWorks将自动终止正在该资源组上运行的任务、进行资源组规格变更、并且重跑被终止的生产任务。

资源组变更				×
→ 变更准备		→ 设置规格	文 支付	- 4 执行变更
				100%
	✓ 桂記溶源组			
	✓ 终止任务			
	✔ 应用新规格			
	变更成功			
				关闭

变更操作需要一定时间,若您不小心关掉当前变更页面,您可以在**资源组列表**页面,将鼠标悬停至 该资源组操作列的国标,单击变更规格,查看规格变更进度。

1.4. 扩缩容操作

如果当前资源组使用率过高,需要并发执行的任务数较大,但该资源组不能满足实际需求,您可以针对该资源 组进行扩容;当资源组长时间闲置不使用,您可以针对该资源组进行缩容,以节省资源费用。本文为您介绍如 何对资源组执行扩缩容操作。

扩缩容介绍

资源组扩缩容的相关介绍如下表。

操作类 型	变更内 容	内存及网络说明	费用说明	操作说 明	生效时 间
扩容操 作	增源相格 加组同资 个 数。	 不会升级资源CPU 与内存配置。 不会增加绑定的 VPC网络数量。 	您需要支付新增资源数量在扩容生效时间点至 原订单到期时间点之间的差价。具体费用请以 实际费用账单为准,查看账单,详情请参见 <mark>查</mark> <mark>看账单详情</mark> 。	无需停 止正在 运行的 任务。	订付后统执 ^向 单成,需行握
缩容操 作	减源相格的数。 少组同资个 。	不涉及	DataWorks会退回从缩容生效时间点至原订 单的到期时间点之间的差价。具体费用请以实 际费用账单为准,查看账单,详情请参见 <mark>查看 账单详情</mark> 。具体费用请以实际费用账单为准, 查看账单,详情请参见查看账单详情。	需停在的务可行操 要止运任,以缩作 了执容。	☆作约 30分时 ,心 30分时 ,心 。

⑦ 说明 同一独享资源组下同规格资源数上限为20,若需要更多,请提交工单联系我们升级。

使用限制

• 资源组限制

- **仅独享数据集成资源组**及独享调度资源组支持扩缩容操作。
- 仅运行中状态的资源组支持扩缩容操作。
- 权限限制

仅支持同时具有AliyunDataWorksFullAccess和AliyunBSSOrderAccess权限策略的用户执行扩缩容操作。

扩容操作

本文以独享调度资源组为例,为您介绍如何进行资源组扩容,独享数据集成资源组的扩容操作类似。

- 1. 进入资源组列表页面。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击资源组列表, 选择目标地域。
- 2. 在独享资源组页签, 鼠标悬停至目标资源组操作列的三图标, 单击扩容。
- 3. 在变配页面,调整资源数量并勾选服务协议。

? 说明	扩容操作	F只能增加资源数	牧量。			
变配						
-	CONTRACTOR OF		1111-1164			
	当前配置					
	实例名称:		独享调度资源: 4 vCPU 8 GiB	独享资源类型	: 独享调度资源	资源数量:3
	地域: 华东2 (上海)					
	当前到期时间: 2022年5月	14日 00:00:00				
	温馨提示:	请选择您要扩(缩)容的资源数量。	规格与新购时一致。			
	资源数量	- 4 +				
	服务协议	✓ DataWorks独享资源(包年包月)	服务协议			

4. 单击立即购买,进行资源扩容。

? 说明

- 每日凌晨(00:00~08:00)是扩容高峰期,扩容时间可能会被延长1~2小时。
- 中国(香港)、亚太东南1(新加坡)、印度尼西亚(雅加达)、美国(硅谷)、印度(孟买)、美国(弗吉尼亚)、德国(法兰克福)、英国(伦敦)、日本(东京)、澳大利亚(悉尼)、马来西亚(吉隆坡)和阿联酋(迪拜)等地域扩容时间整体较长,请您耐心等待

· 资源组扩容时,您可以在资源组列表页面,将鼠标悬停至资源组的状态,查看扩容进度。

资源组列表		10301	9730391	war30391			
独享资源组 公共资源组	自定义资源组						
创建调度资源组 创建集成资	^{[源} 组 创建服务资源组 计	「输入搜索关键词	扩容中 (复制环境 (0%)	WE 30397			WETHOUGH
资源组名称/备注	类型 ☑	状态 ♡	资源扩容	数;量	使用率	操作	
10.000.000	数据集成资源组	◎ 更新中 〈	资源注册	4		网络设置 修改	如月属工作空间 扩容 :
included of processing	数据集成资源组	① 创建失败	2022-06-21	1		网络设置 修?	牧归属工作空间 扩容 :

 资源组扩容完成后,您可以在资源组列表页面,单击该资源组操作列的查看信息,查看扩容后的资源 数量。

缩容操作

本文以独享调度资源组为例,为您介绍如何进行资源组缩容,独享数据集成资源组的缩容操作类似。

⑦ 说明 缩容操作可能会导致任务延迟,请您谨慎评估后再进行操作。

- 1. 进入资源组列表页面。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击资**源组列表**, 选择目标地域。
- 2. 在独享资源组页签, 鼠标悬停至目标资源组操作列的 ::图标, 单击缩容。
- 3. 在降配页面,调整资源数量并勾选服务协议。
 - ? 说明
 - 缩容操作只能减少资源数量。
 - 当资源组中只有一个资源时,无法执行缩容操作。

降配				
		PRODUCTION, THE CONSTRUCTOR,		
	当前配置			
	实例名称: 地域: 华东2 (上海)	拉琴调度资源:4 vCPU 8 GB	独掌创源关型: 独掌调度资源	资源数量: 3
	当前到期时间: 2022年5月 [;]	14日 00:00:00		
	温馨提示:	请选择您要扩(喻)森的资源数量,规格与新和时一致。		
	资源数量	- 2 +		
	服务协议	✔ DataWorke投掌资源(包年包月)服务协议		

- 4. 单击立即购买,进行资源缩容。
 - 资源组缩容时,您可以在资源组列表页面,将鼠标悬停至资源组的状态,查看缩容进度。
 - 资源组缩容完成后,您可以在资源组列表页面,单击该资源组操作列的查看信息,查看缩容后的资源 数量。

1.5. 扣费与欠费说明

DataWorks针对按量付费产品提供了产品侧独有的扣费及欠费行为。本文为您介绍扣费及欠费的相关说明。

使用限制

仅按量付费产品涉及DataWorks产品侧提供的扣费及欠费行为。

扣费与欠费说明

按量计费产品的扣费、欠费说明如下表。

产品类型	计费周 期	扣费说明	欠费说明	延停权益
公共调度资 源组	按天计费	 计费系统在每天00:00开始,清算前一个自然日内运行成功的实例数量,并根据收费标准中的价格区间进行扣费。收费标准,详情请参见公共调度资源组计费说明:按量付费。 每次仅扣除一个区间的日费用,不会叠加扣费。 	 账号已欠费,但未超过24小时 为您延续24小时的服务时间,并在 欠费后的第12小时和第23小时发送 通知提醒您尽快续费。 账号欠费超过24小时 服务会被停用。此时如果您没有购 买公共资源调度组资源包,则未开 始运行的实例无法启动,已开始运 行的实例仍可维持运行至结束。 欠费后充值说明 欠费后24小时内充值,服务不会被 停用。 	
公共数据集 成资源组	按天计费	 计费系统在每天00:00开始,清算前一个自然日内运行成功的实例数量,并根据收费标准中的价格区间进行扣费。收费标准,详情请参见公共数据集成(调试)资源组:按量计费。 每次仅扣除一个区间的日费用,不会叠加扣费。 	 账号已欠费,但未超过360小时 为您延续360小时的服务时间,并在 欠费后的第192小时、288小时和第 336小时发送通知提醒您尽快续费。 账号欠费超过360小时 服务会被停用。此时无法启动未开 始运行的实例,已开始运行的实例 仍可维持运行至结束。 欠费后充值说明 欠费后360小时内充值,服务不会被 停用。 	

产品类型	计费周 期	扣费说明	欠费说明	延停权益
公共数据服 务资源组	按小时 计费	计费系统在每小时开始时,计 算前一小时的消耗费用并生成 账单,从阿里云主账户余额中 扣除实际消费的金额。收费标 准,详情请参见数据服务公共 数据服务资源组计费说明:按 量付费。	 账号已欠费,但未超过24小时 为您延续24小时的服务时间,并在 欠费后的第12小时和第23小时发送 通知提醒您尽快续费。 账号欠费超过24小时 服务会被停用,此时您将无法正常 调用已发布的API。 欠费后充值说明 欠费后24小时内充值,服务不会被 停用。 	
数据集成公 网流量计费	按小时 计费	计费系统会在每小时开始时, 计算前一小时的消耗费用并生 成账单,从阿里云主账户余额 中扣除实际消费的金额。收费 标准,详情请参见 <mark>公网流量计</mark> 费说明。	 账号已欠费,但未超过24小时 为您延续24小时的服务时间,并在 欠费后的第12小时和第23小时发送 通知提醒您尽快续费。 账号欠费超过24小时 服务会被停用,此时无法启动未开 始运行的实例,已开始运行的实例 仍可维持运行至结束。 欠费后充值说明 欠费后24小时内充值,服务不会被 停用。 	

阿延即费生提度续务延常停留里停当的欠供或使的停计的天云权按资费一时用权期费权量提益量源后定长云益间。益负供,付发,额继服。正延额世

DataWorks公共云合集·产品版本规格

与资源组

产品类型	计费周 期	扣费说明	欠费说明	及不定入员 总额的上 延停权益 限。您延停 的额度或时
		 计费系统在每天00:00开 始,清算前一个自然日内运 行成功的实例数量,并根据 收费标准中的价格区间进行 扣费。收费标准,详情请参 见数据质量计费说明。 每次仅扣除一个区间的日费 用,不会叠加扣费。 	 账号已欠费,但未超过24小时 为您延续24小时的服务时间,并在 	时在历因月并多参停 风松盖。 (1) (1) (1) (1) (1) (1) (1) (1) (1) (1)
数据质量	按天计费	 说明 运行成功指规则 本身运行成功, 由于数据质量则 整或阳塞,也会 按运行成功进行 计费。 当天数据质量规 则实例超过 5,000后达到按 量计费上限,后 续规则实例将停 止运行。如果需 要更大的日调度 实例数量,请提 交工单咨询。 	 为恣延续24小时的服务时间,并在 欠费后的第12小时和第23小时发送 通知提醒您尽快续费。 账号欠费超过24小时 服务会被停用,此时您无法启动新 的数据质量规则实例,已启动的规则实例不受影响。 欠费后充值说明 欠费后24小时内充值,服务不会被 停用。 	

DataWorks公共云合集·产品版本规格 与资源组

产品类型	计费周 期	扣费说明	欠费说明	延停权益
智基 生妻	按天计费	 计费系统在每天00:00开始,清算前一个业务日期的基线实例数量,并根据收费标准中的价格区间进行扣费。收费标准,详情请参见基线实例计费说明。 每次仅扣除一个区间的日费用,不会叠加扣费。 ④ 行对小时任务的小时基线,按照1条基线实例计费。 ● 针对小时任务的小时基线,按照1条基线实例计费。 ● 开启状态的基线数量超过100后达到按量计费上限,将无法生成更多的基线实例数量,请提交工单咨询。 	 账号已欠费,但未超过24小时 为您延续24小时的服务时间,并在 欠费后的第12小时和第23小时发送 通知提醒您尽快续费。 账号欠费超过24小时 服务会被停用,此时您无法生成新 的基线实例,已启动的基线实例则 不受影响。 欠费后充值说明 欠费后24小时内充值,服务不会被 停用。 	
智能监控: 报警短信与 报警电话计 费	按天计 费	不涉及	不涉及	
OpenAPI: 调用量	按天计 费	不涉及	如果您的账号欠费,服务将会被立即 停止,为避免影响你的业务,建议您 及时续费。	
自定义调度 资源组	不涉及			
自定义数据 集成资源组	א <i>ו עו</i> ידי <i>ו</i> א			

1.6. 到期说明及续费操作

当您使用的包年包月产品即将到期,如果需要延长产品的到期时间,则可以对该产品进行续费。如果未能及时续费,则相应产品可能会被停止服务或停机释放。本文为您介绍包年包月产品的到期说明及续费操作。

使用限制

- 仅包年包月产品涉及到期说明及续费操作,按量付费产品不涉及。
- 仅支持同时具有AliyunDataWorksFullAccess和AliyunBSSOrderAccess权限策略的用户执行续费操

作。

产品到期说明

包年包月产品到期后, 会被停止服务或停机释放, 具体如下表。

⑦ 说明 资源组商品到期将影响任务的调度和运行,请及时续费。

产品类型	到期通知	停服说明	释放说明		
		到期前未续费,则到期后将立即停止各增值版本的功能,并将其降级为基 础版。			
DataWorks増值 版本(标准版、 专业版、企业版 和旗舰版)		 ? 说明 ● 降级可能对您使用的 值版本到期降级后功 值版本到期降级功能 ● 降级后,如果您仍需 增值版本(包年包月) 	DataWorks功能有影响,DataWorks增 能变化清单,详情请参见 <mark>DataWorks增</mark> <mark>变化清单</mark> 。 使用增值版本,请再次登录 <mark>DataWorks</mark>)售卖页重新购买。		
独享数据集成资 源组		到期前未续费,则到期后将立	 到期后,将为您继续保留7天,如果7 天内您仍未完成续费,则该实例将被 释放 		
独享调度资源组		即停止订单内资源的服务。	λ≠ <i>1</i> ,0 x 。		
独享数据服务资 源组	到期前的第7天、 第3天和第1天, 分别向您的阿里云 主账号绑定的手 机、邮箱、站内信 发送到期通知	⑦ 说明 如果您还需 要使用该资源组,则请 尽快续费,续费详情请 参见续费情况介绍。	 ⑦ 说明 如果您还需要使用该资源组,则请尽快续费,续费详情请参见续费情况介绍。 释放资源组的前一天,将向您的阿里云主账号绑定手机、邮箱等发送释放相联语如 		
	汉 达到州 远 州。		旋 胜 进 刈。		
		 到期前未续费,则到期后DAT 法通过本地客户端和Web页面 DAT ABLAU。 被释放后,DAT ABLAU中已保 	「ABLAU相关功能将立即停止使用(即您无 氩端访问DAT ABLAU),同时,系统会释放 R存的数据模型不会被删除。		
DAT ABLAU数据 建模		⑦ 说明 如果您需要删入DataStudio > 设置 >见批量删除DataBlau DDM	除DAT ABLAU中已过期的数据模型,请进 安全设置与其他 页面删除,详情请参 <mark>数据模型</mark> 。		
		• 被释放后,如果您仍需要使用]DATABLAU, 请重新购买。		
		 ⑦ 说明 重新购买后, 继续使用。购买DAT ABLAU 标准。 	原有DATABLAU中已保存的数据模型仍能 J,详情请参见 <mark>DATABLAU数据建模计费</mark>		

续费情况介绍

包年包月产品续费情况如下表。

产品类型	续费说明	续费操作	费用详情
DataWorks増值 版本(标准版、 专业版、企业版 和旗舰版)	您可在增值版本到期前进行续费。如果在到 并将其降级为基础版。 ⑦ 说明 ● 降级可能对您使用的DataWorks 变化清单,详情请参见DataWor ● 降级后,如果您仍需使用增值版。 卖页重新购买。	期前未续费,则到期后将立即停止 功能有影响,DataWorks增值版本 <mark>'ks增值版本到期降级功能变化清单</mark> 本,请再次登录DataWorks增值版	各增值版本的功能, 到期降级后功能 。 <mark>本(包年包月)</mark> 售
独享数据集成资 源组	 续费范围 续费是为订单内的所有资源统一续 		
独享调度资源组	费。	独享调度资源组、独享数据集	
独享数据服务资 源组	 续费时间 订单到期前或到期后的7日内可执行续 费操作。 未续费影响 如果您在订单到期前未续费,则到期 后该订单内的资源将立即停止服务。 	成资源组、独享数据服务资源 组的续费操作类似,本文以独 享调度资源组为例,为您介绍 如何进行资源组续费。详情请 参见 <mark>续费操作</mark> 。	具体费用请以实际 费用账单为准 <i>,</i> 查
DAT ABLAU数据 建模	到期前续费,则可以继续正常使用 DATABLAU。 续费费用的计算公式: 续费费用 = 单 价 * 需要使用的时长 * 付费并发登录 用户数 。 ? 说明 单价、需要使用的时 长、付费并发登录用户数的计算逻 辑请参考DATABLAU数据建模计费 标准。	直接进入 <mark>购买</mark> 页面,按购买指 引操作即可。	看账单,详情请参 见 <mark>查看账单详情</mark> 。

续费操作

- 1. 进入资源组列表页面。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击**资源组列表**, 选择目标地域。
- 2. 在独享资源组页签, 鼠标悬停至目标资源组操作列的 图标, 单击续费。
- 3. 在续费页面,选择计费周期并勾选服务协议。

续费											
🖕 en ser en e											
	当前配置										
	实例名称:			3	e享调度资源: 4	VCPU 8 GiB				独享资源类型:独享调度资源	资源数量:1
	地域:华东2 (上海)										
	当前到期时间,2022年5日22	₽ 00-00-00									
	=H03304043143. 20224-0/320	Ц 00.00.00									
	计费周期	1个月	2个月	3个月	4个月	5个月	6个月	更多时长	•		
								7个月			
	到期时间: 2022年6月23日	00:00:00						8个月			
								9个月			
	817 Apr 14, 200	DataWark	、油膏漆液 (根午	an Fili Redation	N			1年			
	HR751014X	Datawork	80549-141/8 (8 <u>1</u> 4	-0673 / BR951974	×			2年			

4. 单击**立即购买**,完成续费操作。

续费成功后,您可以继续使用相关产品。同时,您还可以针对包年包月产品的续费方式进行相关管理操 作,详情请参见<mark>通用参考:管理包年包月产品的续费方式</mark>。

2.基础操作 2.1.工作空间管理 2.1.1.配置工作空间

您可以在工作空间配置页面,对当前工作空间的属性进行管理和配置。您还可以绑定MaxCompute、E-MapReduce、Hologres、AnalyticDB for PostgreSQL、AnalyticDB for MySQL及ClickHouse等类型的计算引擎。

进入工作空间配置

- 1. 登录DataWorks控制台。
- 2. 在左侧导航栏,单击工作空间列表。
- 3. 进入工作空间配置页面。

您可以通过以下两种方式进入工作空间配置:

单击相应工作空间后的工作空间配置。在工作空间配置对话框中,单击更多设置,进入工作空间配置页面。

工作空间配置	
基本信息	
工作空间名称	
显示名	
*模式	简单模式(单环境)
描述	
高级设置	
* 启动调度周期 💿	π
* 能下载Select结果 📀	π
更多设置	关闭

○ 单击相应工作空间后的进入数据开发。在数据开发页面,单击右上角的工作空间管理<
 ▲ 入工作空间配置页面。

\$	💸 DataStudio		~ ~					@任务发布	∂ 运维	中心 🔍 💎
		₽₽С⊕	SELECTO	1 × Sq se	lect	dw_u	user_info_all_d			工作空间管理
a	< ▲ □ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	ע≘	1 c 2	(S) (C) odps sql	******	•••••••••••	%			
Q	Sq select 我锁闭	筐 07-04 15:35 我销定 07-04 16:15								
©				********	*****	******	********	*********	******	*********
ê										

4. 根据自身需求,您可以在该页面进行基本属性、安全设置和计算引擎信息等配置。

基本属性

Dat aWorks

⑤ DataWorks	
 ● 权限列表 工作空间D: ▲ MaxCompute高级配置 ▲ 数级紧管理 水 改振紧管理 	创建日期 · 2018-07-13 15:38:12 模式: 标准模式 负责人:
显示名 (Ø	雁丞:
参数	描述
工作空间ID	当前工作空间的ID。
工作空间名称	当前工作空间的名称,仅支持字母或者数字(必须字母开头),不区分大小写。它 是该工作空间的唯一标识,创建后无法修改。
	工作空间包括正常、 已删除、初始化、初始化失败、手动禁用、删除中、删除 失败、欠费冻结、更新中和更新失败等状态。
状态	 ⑦ 说明 如果工作空间新建失败,状态为初始化失败。您可以重试新建操作。 正常状态的工作空间可以被管理员禁用。禁用后该工作空间所有功能 无法使用,数据保留,已经提交的任务正常执行。 被禁用的工作空间可以通过恢复,将工作空间重新置于正常状态。
显示名	当前工作空间的显示名称,用于标识工作空间。支持字母、数字或中文,可以修 改。
创建日期	当前工作空间的创建日期,中国站以东八区为准,无法变更。
	包括简单模式和标准模式。
模式	⑦ 说明 DataWorks简单模式和标准模式工作空间的配置不同。DataWorks标准模式 的工作空间需要分别配置生产环境和开发环境的参数。 如果当前的工作空间为简单模式,阿里云主账号可以进行工作空间模式升级 的操作,将简单模式升级为标准模式,升级入口仅阿里云主账号可见,详情请 参考:工作空间模式升级,本文以标准模式的工作空间为例。
负责人	当前工作空间的所有者,拥有删除、禁用工作空间的权限,并且该身份无法变更。
描述	当前工作空间的描述信息,用于备注工作空间的相关信息,可以编辑。支持128位 中文、字母、符号或数字。

安全设置

6	DataWorks	•		<i>ع</i> 🖡
	=			
۵	工作空间配置	安全设置		
25	成员管理	能下载select结果:	允许子账号变更自己的节点责任人:	
0	权限列表	沙箱白名单(配置ohell任务可以访问的IP地址或域名):		添加沙楠白名单
*	MaxCompute高级配置	IP地址	端口	攝作
*	数据源管理		没有数据	

描述			
设置是否能够下载数据开发中查询的数据结果。如果关闭,则无法下载select的数 据查询结果。			
⑦ 说明 仅空间管理员权限可以开启或关闭下载select结果功能。			
设置是否允许子账号变更自己的节点责任人。			
设置Shell任务运行在默认资源组上,也可以直接访问的IP(此处白名单可以配置IP 和域名)。			
⑦ 说明 必须填写可以被访问到的公网地址或域名。如果是内部服务,建 议使用独享资源保证网络可达,详情请参见独享资源组模式。			

添加沙箱白名单的操作如下:

- 1. 在安全设置区域,单击添加沙箱白名单。
- 2. 在添加沙箱白名单对话框中, 输入地址和端口。
- 3. 单击**确定**。

绑定MaxCompute计算引擎

- 1. 在**计算引擎信息**区域,单击**MaxCompute**,查看当前工作空间下,所有MaxCompute计算引擎的信息。
- 2. 单击增加实例。
- 3. 在新增MaxCompute实例对话框中,配置各项参数。

新增MaxCompute实例		×
* 新增方式:	创建MaxCompute项目	
* 实例显示名称:	请输入实例显示名称	
* 地域:	华东1 (杭州)	
* 付费模式:	🔹 按量付费 💿 包年包月 去购买 💿 开发者版本 😨	
* Quota组:	请选择・・・	
* MaxCompute数据类型:	2.0数据类型(推荐)	
* 是否加密:	○ 不加密 ○ 加密	
* MaxCompute项目交称。	法检) 项日之农	
* MaxCompute项目名称:	请输入项目名称	
访问身份:	请选择	
开发环境		
MaxCompute项目名称:	请输入项目名称	
访问身份:	任务执行者	
如当前登录执行创建MaxCom MaxCompute Super_Administ	pute项目的账号为RAM子账号,为方便管理,该子账号将被加入至 rator角色(标准模式下仅开发环境项目)。	
如当前登录执行创建MaxCom MaxCompute Super_Administ	pute项目的账号为RAM子账号,为方便管理,该子账号将被加入至 rator角色(标准模式下仅开发环境项目)。 确定 耳	以消

参数	描述
新增方式	您可以创建MaxCompute项目或者绑定已有MaxCompute项目。
实例显示名称	实例名称长度需要控制在3~28个字符,仅支持字母开头,仅包含字母、下划线 (_)和数字。
地域	默认当前工作空间所在的地域。
付费模式	包括按量付费、包年包月和开发者版本。 ⑦ 说明 标准模式下不支持添加开发者版本的实例。
Quota组	Quota用来实现计算资源和磁盘配额。
MaxCompute数据类型	包括2.0数据类型(推荐)、1.0数据类型(面向已有使用1.0数据类型用 户)和Hive兼容类型(面向Hive迁移用户),详情请参见:数据类型版本说 明。
是否加密	根据实际情况选择是否需要加密当前实例。

参数	描述
生产环境	生产环境下的配置项包括MaxCompute项目名称和访问身份: MaxCompute项目名称:当前DataWorks工作空间底层使用的MaxCompute生产环境的项目名称。 访问身份:支持您通过阿里云主账号、阿里云子账号或阿里云RAM角色角色访问MaxCompute。
T + 	开发环境下的配置项包括MaxCompute项目名称和访问身份: MaxCompute项目名称:当前DataWorks工作空间底层使用的 MaxCompute开发环境的项目名称。
开反 环境	⑦ 说明 该MaxCompute项目是计算和存储资源。
	○ 访问身份 :默认是 任务执行者 ,不可以修改。

4. 单击确定。

创建完成后,您可以将相应的计算引擎**设置为默认实例**,您还可以单击右上角的**解绑**按钮,解绑对应的 引擎实例。

? 说明

只有空间管理员才可以执行解绑操作。

解绑后DataWorks各模块将产生如下影响:

- 调度任务:基于该引擎运行的任务(实例)即将失败,建议您通过DataStudio中的任务列表 单击**转移到可用的引擎实例**修改引擎实例后执行提交、发布,以恢复任务运行。
- 数据集成:涉及该引擎的同步任务将运行失败,建议您及时到DataSstudio更换相关任务的数据源。
- 数据服务: 涉及该引擎的数据服务AP将调用失败, 建议您及时更换数据服务相关数据源。
- 数据分析:涉及该引擎的查询将不可用,建议您及时更换相关数据源。
- 数据地图、资源优化、综合治理、安全中心将不再展示该引擎的相关信息。

绑定E-MapReduce计算引擎

○ 注意

- DataWorks已暂停绑定E-MapReduce计算引擎,如您有相关需求,请提交工单申请。
- E-MapReduce任务实例仅支持运行在独享调度资源组上。
- 1. 在**计算引擎信息**区域,单击E-MapReduce,查看当前工作空间下,所有E-MapReduce计算引擎的信息。
- 2. 单击增加实例。
- 3. 在新增EMR集群对话框中, 配置各项参数。

DataWorks为您提供**快捷模式**和**安全模式**两种绑定EMR引擎的模式,以实现不同类型的企业、安全要求场景。您可以基于**快捷模式**快速开展各类数据的工作,也可以基于**安全模式**实现更具安全性的数据权限

管理。

新增EMR集群				×
* 实例显示名称:				*
* 地域:	华东1 (杭州) 华东2 (上海) 华北2 (北京)	华南1 (深圳) 西南1 (成都	3)	
* 访问模式: 🕐	快捷模式			
	请确保EMR项目已关闭"安全模式",否则任务将运行失败			
生产环境		开发环境		- 1
基础信息		基础信息		
*调度访问身份 🍞 :	○ 阿里云主账号 ○ 阿里云子账号 温馨提醒: 请对访问身份账号、空间内相关开发人员 账号授权AllyunEMRDevelopAccess策略, 以确保 EMR类任务可以成功提交至EMR集群。	* 访问身份:	● 任务执行者 温馨提醒:请对访问身份账号、空间内相关开发人员 账号授权AliyunEMRDevelopAccess策略,以确保 EMR类任务可以成功提交至EMR集群。	
引擎信息		引擎信息		
* 集群ID 😗 :	请选择	* 集群ID:	请选择	
* 项目ID 🥝 :	清选择	* 项目ID:	请选择	
* YARN溶源队列·	default	* YARN资源队列:	default	-
			确定	取消

参数	描述
实例显示名称	自定义实例的名称。
地域	默认当前工作空间所在的地域。
访问模式	 当EMR计算引擎的绑定模式为快捷模式时,阿里云主账号或RAM用户在 DataWorks运行代码或自动调度任务,都只是下发代码至EMR集群,实际运 行的身份为集群内的Hadoop用户。 当EMR计算引擎的绑定模式为安全模式时,阿里云主账号或RAM用户在下 发代码的同时,EMR集群内会匹配对应的同名用户来运行任务。管理者可以 使用EMR集群内的Ranger组件对每个用户进行权限管控,最终实现不同阿里 云主账号、任务责任人或阿里云子账号在DataWorks上运行EMR任务时, 拥有对应不同数据权限的目的,进行更安全的数据权限隔离。详情请参见准 备工作:绑定EMR引擎。

参数	描述
调度访问身份	 如果访问模式选择快捷模式,当提交任务至调度系统后,DataWorks调度 系统自动运行任务时,您可以选择阿里云主账号或阿里云子账号提交代码 至EMR集群。 如果访问模式选择安全模式,当任务被提交、发布至生产环境 后,DataWorks调度系统自动运行任务时,您可以选择任务责任人、阿里 云主账号或阿里云子账号提交代码至EMR集群,同时该身份对应的Hadoop 用户将实际运行代码。
	 ⑦ 说明 • 仅生产环境涉及配置该参数。 • 为保证工作空间成员可以在DataStudio内正常运行EMR类的任务,请确保开发、管理员等相关角色拥 有AliyunEMRDevelopAccess权限策略。 • 如果您使用阿里云主账号运行任务,该账号天然拥 有AliyunEMRDevelopAccess权限策略。 • 如果您使用RAM用户运行任务,则需要授予该用 户AliyunEMRDevelopAccess权限策略,详情请参见为 RAM用户授权。
	在开发环境运行任务时,提交代码至EMR引擎所使用的身份。此处默认为 任务 执行者 。
访问身份	 ⑦ 说明 • 仅DataWorks标准模式的工作空间会显示该参数,并且仅开发环境涉及配置该参数。 • 任务执行者可以为阿里云主账号或RAM用户。 为保证工作空间成员可以在DataStudio内正常运行EMR类的任务,请确保开发、管理员等相关角色拥有AliyunEMRDevelopAccess权限策略。 • 如果您使用阿里云主账号运行任务,该账号天然拥
	有AliyunEMRDevelopAccess权限策略。 如果您使用RAM用户运行任务,则需要授予该用 户AliyunEMRDevelopAccess权限策略。
集群ID	从下拉列表中选择调度访问身份账户所在的EMR集群,作为任务的运行环境。
	从下拉列表中选择调度访问身份账户所在的EMR项目,作为任务的运行环境。
项目ID	⑦ 说明 如果EMR项目开启 访问模式 为安全模式,则无法被选择。
YARN资源队列	当前集群下的队列名称。如果无特殊需求,请输入default。

参数	描述
覆盖DataStudio任务队列映 射	 勾选覆盖DataStudio任务队列时:所有任务将按照当前指定的队列运行。 未勾选覆盖DataStudio任务队列时: 如果节点高级配置中配置了queue,则按照节点的queue配置运行。 如果节点高级配置中没有配置queue(自己删除了该配置),任务将按照当前界面指定的队列运行。 ⑦ 说明 如果您看不到覆盖DataStudio任务队列功能,可以提交工单联系我们进行升级。
Endpoint	EMR的Endpoint,不可以修改。
选择资源组	选择已与当前DataWorks工作空间配置网络连通性的独享调度资源组。如果您 没有可用的独享调度资源组则需要新建。新建独享调度资源组并配置网络连通 性,详情请参见新增和使用独享调度资源组。 选择后,单击 测试连通性 进行测试,以验证独享调度资源组和E-MapReduce实 例的网络连通性。验证后会对资源组进行初始化。 ⑦ 说明 当EMR集群配置变更时,需要对资源组重新进行初始化。

4. 单击**确定**。

创建完成后,您可以将相应的计算引擎设置为默认实例,并可以进行编辑。

绑定Hologres计算引擎

- 1. 在**计算引擎信息**区域,单击Hologres,查看当前工作空间下,所有Hologres计算引擎的信息。
- 2. 单击**绑定HologresDB**。
- 3. 在**绑定HologresDB**对话框中,配置各项参数。

绑定HologresDB						×
* 实例显示名称:						
生产环境			开发环境			
基础信息			基础信息			
* 访问身份 ? :	○ 阿里云主账号 ○ 阿里云子账号		* 访问身份 🤫 :	• 任务执行者		
引擎信息			引擎信息			
* Hologres实例名称:	请选择	~	* Hologres实例名称:	请选择	~	
* 数据库名称 🤫 :			* 数据库名称 🕐 :			
连通性测试			连通性测试			
测试连通性:	测试连通性		测试连通性:	测试连通性		
					确定	取消

参数	描述		
实例显示名称	实例的显示名称。		
访问身份	 DataWorks任务提交调度后,作为在Hologres引擎内执行代码的账号。包括阿里云主账号和阿里云子账号。 		
	⑦ 说明 仅生产环境涉及配置该参数。		
	 在开发环境运行任务时,此处默认为任务执行者。 		
Hologres实例名称	选择需要绑定的Hologres实例。		
数据库名称	输入在 SQL Console 中新建的数据库的名称。例如,testdb。		

- 4. 单击测试连通性。
- 5. 测试连通性通过后,单击确定。

绑定AnalyticDB for PostgreSQL计算引擎

- ↓ 注意
 - 您需要开通DataWorks标准版及以上版本,当前页面才会显示AnalyticDB for PostgreSQL。
 - AnalyticDB for PostgreSQL任务实例仅支持运行在独享调度资源组上。

1. 在计算引擎信息区域,单击AnalyticDB for PostgreSQL。

2. 单击增加实例。

标准模式的工作空间,开发环境和生产环境隔离。如果您使用的是标准模式的工作空间,则需要分别为开 发环境和生产环境添加实例。

3. 在新增AnalyticDB for PostgreSQL实例对话框中,配置各项参数。以下以标准模式的环境为您展示。

新增Analytic	DB for PostgreSQL实例			×
* 实例显示名 称:				
生产环境		开发环境		
基础信息		基础信息		
* 实例名称:	请选择	* 实例名称:	请选择	~
* 数据库名称:		* 数据库名称:		
* 用户名:		* 用户名:		
* 密码:		* 密码:		
连通性测试		连通性测试		
选择资源组:	请选择 🗸 测试连通性	选择资源组:	请选择 > 测试连通性	
	新建独享资源组网络连通性的帮助文档		新建独享资源组网络连通性的帮助文档	
				确定取消

参数	描述
实例显示名称	实例的显示名称,请注意名称需要保持唯一性。
实例名称	选择需要在工作空间中使用的AnalyticDB for PostgreSQL实例。
数据库名称	AnalyticDB for PostgreSQL实例下需要绑定使用的数据库的名称。
用户名	拥有数据库操作权限的Username。,详情请参见 <mark>创建数据库账号</mark> ,进入 AnalyticDB for PostgreSQL的 账号管理 页面获取。
密码	拥有数据库操作权限的Password。,详情请参见 <mark>创建数据库账号</mark> ,进入 AnalyticDB for PostgreSQL的 账号管理 页面获取。
连通性测试	AnalyticDB for PostgreSQL任务实例需要运行在独享调度资源组上,请在此处选择需要使用的独享调度资源组,详情请参见 <u>独享资源组模式</u> 。 选择后,单击 测试连通性 进行测试,以验证独享调度资源组和AnalyticDB for PostgreSQL实例的网络连通性。如果您没有可用的独享调度资源组,请单击 新 建独享资源组创建。

4. 测试连通性通过后,单击确定。

绑定AnalyticDB for MySQL计算引擎

< ○ 注意

- 您需要开通DataWorks标准版及以上版本,当前页面才会显示AnalyticDB for MySQL。
- AnalyticDB for MySQL任务实例仅支持运行在独享调度资源组上。
- 阿里云RAM子账号绑定引擎时,需要拥有引擎实例的DescribeDBClusters权限,授权详情可参考 AnalyticDB MySQL文档: RAM子账号和权限。
- 1. 在计算引擎信息区域,单击AnalyticDB for MySQL。

2. 单击增加实例。

标准模式的工作空间,开发环境和生产环境隔离。如果您使用的是标准模式的工作空间,则需要分别为开 发环境和生产环境添加实例。

3. 在添加AnalyticDB for MySQL实例对话框中,配置各项参数。以下以标准模式的环境为您展示。

添加AnalyticDB	for MySQL实例	×	
* 实例显示名称:			
生产环境		开发环境	
基础信息		基础信息	
* 实例名称 🅐 :	请选择	* 实例名称 🕥 : 请选择 🛛 🗸	
* 数据库名称:		* 数据库名称:	
* 用户名:		* 用户名:	
* 密码:		* 密码:	
连通性测试		连通性测试	
选择资源组:	请选择 > 测试连通性	选择资源组: 请选择 > 测试连通性	
	新建独享资源组网络连通性的帮助文档	新建独享资源组网络连通性的帮助文档	
		确 定 取消	

参数	描述
实例显示名称	实例的显示名称,请注意名称必须唯一。
实例名称	选择需要在工作空间中使用的AnalyticDB for MySQL实例。
数据库名称	AnalyticDB for MySQL实例下绑定的需要使用的数据库名称。
用户名	拥有数据库操作权限的Username。,详情请参见 <mark>数据库账号和权限,</mark> 进入 AnalyticDB for MySQL的 账号管理 页面获取。
密码	拥有数据库操作权限的Password。,详情请参见 <mark>数据库账号和权限</mark> ,进入 AnalyticDB for MySQL的 账号管理 页面获取。
连通性测试	AnalyticDB for MySQL任务实例需要运行在独享调度资源组上,请在此处选择 需要使用的独享调度资源组,详情请参见 <u>独享资源组模式</u> 。 选择后,单击 测试连通性 进行连通性测试,以验证独享调度资源组和 AnalyticDB for MySQL实例的网络连通性。如果您没有可用的独享调度资源 组,请单击 新建独享资源组 创建。

4. 测试连通性通过后,单击确定。

绑定CDH计算引擎

- 1. 在计算引擎信息区域,单击CDH。
- 2. 单击增加实例。

标准模式的工作空间,开发环境和生产环境隔离。如果您使用的是标准模式的工作空间,则需要分别为开 发环境和生产环境添加实例。

3. 在新增CDH引擎实例对话框中, 配置各项参数。

新增引擎实例时,可选择使用**快捷模式**或安全模式访问模式,安全模式可以实现不同云账号运行任务时的数据权限隔离。配置界面如下:

新增EMR集群					×
* 实例显示名称:					^
* 地域:	华东1 (杭州) 华东2 (上海) 华北2 (北京)	华南1 (深圳) 西南1 (成都	阝)	- 1
* 访问模式: 🕐	快捷模式	~			- 1
	请确保EMR项目已关闭"安全模式"	,否则任务将运行失败			- 1
生产环境			开发环境		- 1
基础信息			基础信息		- 1
* 调度访问身份 🥑 :	○ 阿里云主账号 ○ 阿里 温馨提醒:请对访问身份账号。 账号授权AllyunEMRDevelopAc EMR类任务可以成功提交至EM	云子账号 、空间内相关开发人员 c <mark>cess策略</mark> ,以确保 IR集群。	* 访问身份:	● 任务执行者 温馨提醒:请对访问身份账号、空间内相关开发人员 账号授权AllyunEMRDevelopAccess策略,以确保 EMR类任务可以成功提交至EMR集群,	1
引擎信息			引擎信息		
* 集群ID 😗 :	请选择	~	* 集群ID:	请选择	
* 项目ID 🕐 :	请选择	~	* 项目ID:	请选择	
* YARN溶源队列·	default		* YARN资源队列:	default	•
					取消

参数	描述
实例显示名称	自定义的实例显示名称,请注意名称必须唯一。
访问模式	 快捷模式:该访问模式使用便捷,多个云账号对应一个集群账号,多个账号均可访问同一个集群账号内的数据,无法实现不同云账号运行任务时的数据权限隔离。 安全模式:该访问模式允许您配置云账号与CDH集群账号的身份映射关系,实现不同云账号运行任务时的数据权限隔离。
选择集群	 如果访问模式选择快捷模式,则此处需要选择未开启Kerberos认证的CDH 集群。如果您当前无CDH集群,请先创建集群,请参见:在DataWorks中新 增CDH集群配置。 如果访问模式选择安全模式,则此处需要选择已开启Kerberos认证的CDH 集群。您可以查看CDH集群是否开启Kerberos认证(在工作空间管理页面, 单击CDH集群配置,找到需要查看的集群,单击修改,进入后下拉查看映 射配置中的认证类型)。如果您当前无CDH集群,请先创建集群。

参数	描述
	 快捷模式无认证方式,仅支持指定特定账号,建议使用admin或hadoop账号。该账号仅用于下发任务。 安全模式 您可以根据需求选择调度访问身份。该身份用于在任务提交调度后自动调度运行任务,并且需要配置云账号与CDH集群账号的身份映射,请参见配置访问身份的映射关系。您可以选择任务责任人、阿里云主账号和阿里云子账号。
访问身份	 ⑦ 说明 ● 仅生产环境涉及配置该参数。 ● 在DataStudio页面,运行任务所使用的身份均为当前已登录 云账号映射的集群访问身份。因此,除了需要为调度访问身 份配置身份映射外,建议为项目空间开发成员也配置身份映 射,避免页面运行任务失败。 ● 在开发环境运行任务时,此处默认为任务执行者。
	选择已与当前DataWorks工作空间配置网络连通性的独享调度资源组。如果您 没有可用的独享调度资源组,则需要新建。新建独享调度资源组并配置网络连
独享调度资源组	通性,详情请参见 <mark>新增和使用独享调度资源组</mark> 。 选择后,单击 测试连通性 进行测试,以验证独享调度资源组和CDH实例的网络 连通性。

4. 测试连通性通过后,单击确定创建计算引擎实例。

绑定ClickHouse计算引擎

绑定ClickHouse计算引擎之前,您需要完成以下准备工作:

- 已创建ClickHouse集群。
 - 您可以在E-MapReduce上创建ClickHouse集群,详情请参见创建集群。
 - 您可以在云数据库ClickHouse上创建ClickHouse集群。
- 已开通DataWorks并创建好对接使用ClickHouse的工作空间。对接使用ClickHouse的工作空间无需绑定计算引擎,在创建工作空间时可跳过选择引擎步骤,创建工作空间的其他步骤,详情请参见创建工作空间。
- 已开通独享调度资源组,并绑定至所用的DataWorks工作空间,详情请参见新增和使用独享调度资源组。

⑦ 说明 建议独享调度资源组与部署的ClickHouse集群绑定同一地域的同一个VPC网络。如果您需要 将独享调度资源组与部署的ClickHouse集群绑定于不同地域或不同的VPC网络,则需要执行较复杂的网 络连通流程,详情请参见配置资源组与网络连通。

- 1. 在计算引擎信息区域,单击ClickHouse。
- 2. 单击增加实例。

标准模式的工作空间,开发环境和生产环境隔离。如果您使用的是标准模式的工作空间,则需要分别为开 发环境和生产环境添加实例。

3. 在新增ClickHouse集群对话框中,配置各项参数。

您可以使用**连接串模式**或**EMR集群模式**连接ClickHouse集群,以下以标准模式的环境为您展示如何使用 两种模式新增ClickHouse集群。

新增ClickHouse集	群			×
* 实例显示名称:	请输入实例显示名称			^
*集群类型:	连接串模式 >			- 1
*访问模式:	快捷模式 🗸			- 1
生产环境		开发环境		-
集群信息		集群信息		- 1
* JDBC URL:	jdbc:clickhouse://xxx.xxx.xxx.xxx/	* JDBC URL:	jdbc:clickhouse://xxx.xxx.xxx.xxx/	- 1
访问身份		访问身份		- 1
* 用户名:	xiangcui	* 用户名:	请输入ClickHouse用户名	- 1
* 密码:		* 密码:	请输入ClickHouse用户密码	- 1
	建议使用SHA256分配密码		建议使用SHA256分配密码	- 1
网络连通性		网络连通性		
请 添加独享调	度资源组以实现DataWorks与ClickHouse集群的连通!	i 请添加独享调	度资源组以实现DataWorks与ClickHouse集群的连通!	
独享调度资源组:	请选择	独享调度资源组:	请选择 / 刷新 请您终考 出文档 对独享资源组讲行网络配置,如当前	•
			确定	取消

参数	描述			
实例显示名称	自定义的实例显示名称,请注意名称必须唯一。			
集群类型	选择连接串模式或EMR集群模式新增ClickHouse集群。			
访问模式	目前仅支持 快捷模式 。当ClickHouse计算引擎的绑定模式为快捷模式时, 阿里 云主账号或RAM用户在DataWorks运行代码或自动调度任务,都只是下发代码 至ClickHouse集群,实际运行任务的身份为配置项访问身份内所输入的用户。			
集群信息	 连接串模式需要配置JDBC URL为接入ClickHouse集群的JDBC连接串。 EMR集群模式选择需要绑定的EMR ClickHouse集群ID。 说明 如果您当前使用的是RAM用户登录,请先授予该用户Aliyun EMRDevelopAccess策略后再选择集群。授权RAM用户AliyunEMRDevel opAccess权限策略,详情请参见为RAM用户授权。 			
用户名	访问ClickHouse集群所使用的用户名。 如果您使用的是EMR(E-MapReduce)的ClickHouse集群,则可以登录 <mark>EMR管</mark> 理控制台进入 集群管理的集群服务 页面,查看目标集群的用户名,如下图配置 信息所示。			
密码	访问ClickHouse集群所使用的密码。 如果您使用的是EMR(E-MapReduce)的ClickHouse集群,则可以登录 <mark>EMR管</mark> 理控制台进入 集群管理的集群服务 页面,查看目标集群的密码,如下图配置信 息所示。			

参数	描述
独享调度资源组	选择已与当前DataWorks工作空间配置网络连通性的独享调度资源组。如果您 没有可用的独享调度资源组,则需要新建。新建独享调度资源组并配置网络连 通性,详情请参见 <mark>新增和使用独享调度资源组</mark> 。

EMR(E-MapReduce)的ClickHouse集群,配置信息如下图所示。

E-MapReduce 器 概念							0	
- 100 (T) (T)	品 集群管理 (⊙ 操作历史 ■ 事件列表	山 数据开发 → 元数据管理	◎ 監控大盘 ^{Beta}			✿ 系统管理 ∨	帮助 🗗
■ 目の離立には ck_test002 配置授業 ittel >		Q	服労電C画 全部 server-metrika client-	config server-users server-conf	fig	,		(秋子) 定义配置
 集群基础信息 品置范围 				users.default.networks.host		0		
 集 計 版 务 第 計 版 务 配 置 类 型 	記置 >		prof	iles.default.use_uncompressed_cache users.default.networks.host_regexp	0	0		
 ⁵⁵ Ganglia 基础配置 ✓ ZooKeeper 数据路径 	高级配置 日志路径	只读配置 日志相关		profiles.default.max_memory_usage	1000000000	0		
Wh ClickHouse 上述 ClickHouse 生態相当	数据相关 时间相关 (数据库相关编解码相关		users.default.profile	default	0		
 王もの決 OSS相当 引导操作 磁盘相当 	地址端口 P	为存配置 文件路径		users.default.networks.ip	1 /0	0	-	
ゆ 集群脚本 URL或U	Ц			users.default.password		20		
				users.default.quota	default	0		8
				profiles.default.load_balancing	random 每页显示: 20 50	2100 全部	< 1 > 3	共10条

4. 测试连通性通过后,单击确定。

2.1.2. MaxCompute高级配置

您可以通过工作空间管理中的MaxCompute高级配置操作,对当前工作空间的MaxCompute属性进行管理和配置。

进入配置工作空间页面后,单击左侧导航栏中的MaxCompute高级配置,即可进入MaxCompute高级配置页面。

MaxCompute高级配置页面包括基本设置和自定义用户角色两个模块。

基本设置

您可以在基本设置模块,对选择的MaxCompute项目进行安全配置。

G DataWorks	23	•	থ্
三 〇〇 工作空间配置	MaxCompute 项目选择:	开发环	
🏨 成员管理	基本设置	MaxCompute安全配置	
	自定义用户角色	使用ACL接权: 检查用户的访问权限时使用ACL	
✓ WaxComputeress(#)		允许对象创建者访问过象: 允许对象创建者该职、特征或删除自己创建的对象	
		允许对象创建者统权对象: 允许对象创建者自主地将自己创建的对象境权的该项目空间中的其他用户	
		項目空間数据保护: 用例工作空间的数据外流	
		子除号接号 : 显态开始RAM子陈马服务true	
		使用Policy接段: 检查用户的达词印刷时使用policy	
		启动影响起的达到 通过Lable可以达物影响现象	

MaxCompute安全设置:涉及底层MaxCompute相关的权限及安全设置,详情请参见安全参数列表。

配置	说明
使用ACL授权	激活或冻结该开关,相当于Owner账号在MaxCompute项目中执行 set CheckPermissionUsingACL=true/false 操作,默认激活。
允许对象创建者访问对象	激活或冻结该开关,相当于Owner账号在MaxCompute项目中执行 set ObjectCreatorHasAccessPermission=true/false 操作,默认激活。
允许对象创建者授权对象	激活或冻结该开关,相当于Owner账号在MaxCompute项目中执行 set ObjectCreatorHasGrantPermission=true/false 操作,默认激活。
项目空间数据保护	激活或冻结该开关,相当于Owner账号在MaxCompute项目中执行 set ProjectProtection=true/false ,默认冻结。
子账号服务	激活或冻结该开关,控制子账号可以访问或禁止访问该MaxCompute项目,默认激 活。
使用Policy授权	激活或冻结Policy授权机制,相当于Owner账号在MaxCompute项目中执行 set CheckPermissionUsingPolicy=true/false 操作,默认激活。
启动列级别访问控制	项目空间中的LabelSecurity安全机制默认关闭,项目Owner可以自行开启,相当于 Owner账号在MaxCompute项目中执行 Set LabelSecurity=true/false 。

自定义用户角色

您可以在自定义用户角色模块,对选择的MaxCompute项目进行用户角色的配置。

G DataWorks		× ع	
- 二作空间配置	MaxCompute 项目选择:	〒按环境 ────	
😹 成员管理	基本设置	自主义用户条色	新増角色
权限列表	自定义用户角色		
✓ MaxCompute高级配置		输入角色各称进行搜索	
▶ 数据源管理		角色名称 摄作	
		登着详情 成页管理	
		查看详情 成员管理 初限管理	
		重音详情 成页管理 初频管理 删除	
配置		说明	
角色名称		MaxCompute项目中的角色名称。	
操作		 查看详情:查看当前角色中包含的成员列表,以及当前角色对表或项目的权限。 成员管理:添加或删除当前角色中的成员。 权限管理:设置并管理当前角色对表或项目的权限,详情请参见权限列表。 删除:仅支持删除当前账号新建的角色。 	2

配置	说明	
新增角色	说明 单击右上角的新增角色,在新增角色对话框中填写角色名称,在待添加账号处包,添加的成员账号,单击>,将需要添加的账号移动至已添加的账号中,单击确定,加成功。 新增角色 ① ● 角色名称: ● 角色名称: ● 「● 角色名称: ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●	选需要 即可添 ×
	■ 1/5 项 ① 项 ④ 施定 取消	

⑦ 说明 此处自定义角色所设置的权限,将与原有的系统权限取并集。

2.1.3. 数据源管理

数据源主要用于数据集成过程中Reader和Writer的对象,您可以在数据源管理页面查看、新增及批量新增数据 源。

进入配置工作空间页面后,单击左侧导航栏中的数据源管理,即可进入数据源管理页面。

您可以在数据源管理页面,根据数据源类型、数据源名称等条件筛选需要查看的数据源。

⑦ 说明 新绑定的计算引擎实例都会产生的一个默认的计算引擎数据源,例如odps_first,命名格式为 0_regionId_引擎名称。

单击页面右上角的新增数据源,即可新增相应的数据源,详情请参见数据源配置。

单击页面右上角的**批量新增数据源**,即可批量新增相应的数据源,详情请参见批量新增数据源。

2.1.4. 工作空间模式升级

标准模式下,一个DataWorks工作空间对应两个MaxCompute项目,隔离开发环境和生产环境,增加任务的发 布流程,保证任务代码的正确性。

标准模式的优势

在DataWorks旧版本中,您创建的是一个DataWorks工作空间对应一个MaxCompute项目,即新版DataWorks中的简单模式。
简单模式的工作空间容易导致表权限不可控。例如,在简单模式下无法直接实现仅允许工作空间中的某些人查询部分表的场景。因为一个DataWorks工作空间对应一个MaxCompute项目,DataWorks的开发角色权限在MaxCompute项目下有全部表的操作权限,所以无法对表权限进行精准控制。此时您必须新建一个单独的DataWorks工作空间,使用工作空间隔离的方式,完成数据的隔离。

旧版DataWorks支持手动发布、绑定两个DataWorks工作空间。设置工作空间A为工作空间B的发布工作空间, 工作空间A即可接收工作空间B中发布过来的任务,无需开发代码。此时工作空间A成为类似生产环境的工作空间,工作空间B则是类似开发环境的工作空间。

绑定两个DataWorks工作空间存在一定的不足:工作空间A是一个正常的DataWorks工作空间,可以直接在数据开发模块中进行任务开发,导致(生产)环境的代码更新入口不唯一,容易影响整个开发流程。

针对上述问题,新版DataWorks推出了标准模式。标准模式的工作空间具有以下优势:

- 一个DataWorks工作空间对应两个MaxCompute项目,可以将开发和生产的计算引擎分开。工作空间成员只 拥有开发环境的权限,默认无权限操作生产环境的表,提升了生产环境数据的安全性。
- 标准模式下,数据开发界面默认操作开发环境的任务,只有发布后才会是生产环境的任务。标准模式可以 保证生产环境代码编辑入口的唯一性,并提升生产环境代码的安全性。
- 标准模式下,开发环境默认不进行周期调度,可以减少账号下的计算资源消耗,保障生产环境任务运行的资源。

标准模式对使用流程的影响

如图,标准模式"生产、开发隔离"的模式将影响数据模型设计、数据处理逻辑代码发布等流程。



工作空间模式升级

阿里云主账号可以通过下述操作,升级简单模式的工作空间为标准模式。

- 1. 登录DataWorks控制台,单击相应工作空间后的工作空间配置。
- 2. 单击工作空间配置对话框中的更多设置,即可进入工作空间配置页面。

您也可以进入数据开发页面,单击右上角的工作空间管理,进入工作空间配置页面。

😚 💸 DataStudio	•
₩ 数据开发	数据开发 2 歳 🕻 С 🕀 🖸
💼 组件管理	Q 文件名称/创建人 证
Q 临时查询	> 解决方案 器
④ 运行历史	~ 业务流程
- 手动业务流程	 ▲ > ⇒ 数据集成
市 回收站	> 🚺 MaxCompute
	> 🔅 EMR > 😂 数据服务
	> 🧮 算法
	> 🔕 適用
	> 易 自定义

3. 在工作空间配置页面,单击模式后的升级为标准模式。

6	DataWorks	·	
	=		
٢	工作空间配置	配置	
25	成员管理	基本属性	
9	权限列表	工作空间ID:76834	创建日期 : 2018-07-23 11:06:12
w.	MaxCompute高级配置	工作空间名称	模式:简单模式 升级为标准模式
*	数据源管理	状态正常	负责人:
		显示名 ()	描述: 🥑 🧭

4. 在升级为标准模式对话框中,输入开发环境下的MaxCompute项目名称,并勾选确认要升级此工作空间,单击确定。

注意 此处您需要新建一个简单模式工作空间作为开发环境下的MaxCompute项目,并可以 在工作空间配置 > 计算引擎信息页面,选择使用个人账号或工作空间所有者操作生产环境的数据。

5. 单击确认升级提示框中的确认。

完成上述操作后,即可返回工作空间配置页面,查看该工作空间的模式已显示为标准模式。

简单模式升级为标准模式的影响

简单模式升级为标准模式后,DataWorks会将原工作空间中的成员加入新建的MaxCompute开发项目中,并将 原项目的成员和角色都保留。但会撤销项目成员在生产项目的权限,只有项目所有者具有生产项目的权限。

例如,某公司在DataWorks上有一个工作空间A,单击**升级为标准模式**后,创建了一个开发环境的工作空间 A_dev。原来工作空间A中的成员、角色、表、资源都会在工作空间A_dev下进行创建(仅创建表,不会克隆 表数据)。原工作空间A下的成员A1(开发角色)、B1(运维角色),同时也会加入工作空间A_dev,并保留 角色权限。A项目会变成生产项目,A1和B1用户在工作空间A中的数据权限会被撤销,默认没有表的select和 drop权限,生产环境的数据会直接受到保护。

默认在DataStudio(数据开发)页面操作的MaxCompute项目是A_dev。如果在数据开发界面查询生产环境的数据,需要使用项目名.表名的方式。数据开发页面只能编辑A_dev环境的代码,如果需要更新工作空间A中的代码,只能通过A_dev提交任务至调度系统,发布至生产环境的方式进行更新。任务发布(审核)可以保障生产环境代码的正确性。

⑦ 说明 工作空间模式升级后,无法直接访问原工作空间的数据,需要申请角色权限。在数据开发界面 查询的表,默认是开发环境的表。如果要访问生产表,需要申请完角色权限后,使用项目名.表名的方式 访问。

工作空间升级到标准模式后,RAM用户之前的角色会被清除。如果您的代码中写了某个账号的密钥,可能 会产生无权限的报错。

影响行为示例:

假设当前绑定的引擎为MaxCompute引擎,	简单模式工作空间使用计算引擎的账号为阿里云主账号,	升级为标
准模式工作空间后RAM用户的操作行为如下	表所示。	

行为场景	原简单模式工作空间	现标准模式工作空间
操作资源 (table/resource/function)	直接操作即可成功。 ⑦ 说明 简单模式工作空间 直接使用阿里云主账号权限执行 任务,所以即便是RAM用户操 作,其同样与主账号有相等的权 限,由于阿里云主账号权限较 大,所以RAM用户可以直接操作 生产数据,无法做到权限控制。	 操作开发环境资源:RAM用户在数据开发界面可以通过命令直接操作开发环境资源。 操作生产环境资源:RAM用户在数据开发界面无法直接操作生产环境资源。 说明标准模式工作空间下RAM用户默认没有生产环境权限,生产表权限需要在安全中心走审批进行审批。
访问环境及账号问题	 简单模式仅生产环境,所有命令直接操作生产。 数据开发:使用阿里云主账号访问生产资源 生产运维中心:使用阿里云主账号访问生产资源。 	 数据开发:使用RAM用户、个人账号访问资源。默认访问/写入开发资源。 生产运维中心:使用生产环境指定账号访问/写入生产资源。

行为场景	原简单模式工作空间	现标准模式工作空间	
	projectname.tablename/resource /function	 访问开发资源: projectname_dev.tablename/re source/function 	
		 ⑦ 说明 数据开发 DataStudio界面默认访问开 发资源。即在sql中不带项目 名直接访问 table/resource/function 	
访问对应环境资源		 访问生产资源: projectname.tablename/resour ce/function 	
		⑦ 说明 生产运维中心默 认访问生产资源。即在sql中 不带项目名直接访问 table/resource/function	

⑦ 说明 关于简单模式与标准模式差异详情可参考文档: MaxCompute引擎在不同模式工作空间访问的 相关问题

2.2. 工作空间规划与管理

2.2.1. 规划工作空间

本文介绍DataWorks的工作空间体系,以及适用于不同场景的工作空间划分方案。

什么是工作空间

工作空间是DataWorks管理任务、成员,分配角色和权限的基本单元。工作空间管理员可以加入成员至工作空间,并赋予工作空间管理员、开发、运维、部署、安全管理员或访客角色,以实现多角色协同工作。

一个工作空间支持绑定MaxCompute、E-MapReduce和实时计算等多种类型的计算引擎实例。绑定引擎实例 后,即可在工作空间开发和调度引擎任务,并管理引擎中存储的数据。

在工作空间内部,提供"空间管理员"、"数据分析师"、"部署"、"开发"、"模型设计师"、"访客"、"项目所有者"、"运维"、"安全管理员"这几种成员角色设定,每种角色有不同的权限。

工作空间类型

DataWorks提供简单模式和标准模式两种类型的工作空间:

• 简单模式工作空间

简单模式工作空间包含的每一个计算引擎实例只能有一套环境,无法区分开发和生产环境,只能进行简单 的数据开发,且无法对数据开发流程和表权限进行强控制。

○ 优点:简单直观、迭代快,代码提交后,无需发布即可生效。

• 缺点:开发角色权限过大,可以随意删除本工作空间下的表,存在数据安全风险。

在简单模式工作空间中,任务实际的执行身份可以选择为"任务负责人"或"阿里云主账号",生成数据的 所有者也因此设定而不同。

• 标准模式工作空间

标准模式工作空间包含的每一个计算引擎实例都包含两套环境,分别为开发环境和生产环境。同时,可以 基于角色定义用户权限,开发人员仅能操作开发环境中的数据和任务脚本,而运维人员仅能操作生产环境 的任务脚本,所有人均无法直接访问生产环境的数据。各种角色相互监督相互保证,确保项目符合数仓管 理规范。

同时,标准模式的工作空间可以严格控制表权限,禁止随意操作生产环境表的行为,保证生产表的数据安全。

- 优点: 权限、流程规范化管理, 保障数据安全和代码安全。
- 缺点:需额外创建用于开发的计算引擎实例环境,权限和角色模型存在理解成本。

在标准模式工作空间中,开发、生产环境的任务执行身份有所不同:

- 开发环境:锁定为"任务负责人"。
- 。 生产环境:可以选择通过"阿里云主账号"或指定的"RAM账号"身份运行调度任务。

工作空间权限模型

DataWorks各主要模块针对工作空间的权限隔离设定有所不同:

功能模块	权限模型			
工作空间管理	不同的工作空间的是 完全隔离 的。 不同的工作空间可以有不同的管理员、不同的内部成员, 各工作空间拥有完全独立的成员 角色设定以及引擎实例的各项参数开关。 ⑦ 说明 所有工作空间的所有者账号均为阿里云主账号。			
数据开发(Dat <i>a</i> Studio)	 各工作空间之间的数据开发工作是完全隔离的。 工作空间之间的业务流程、任务节点独立开发,互不影响。 在工作空间内部: 仅"开发"、"管理员"角色的成员具备创建、编辑、删除任务节点的权限。 仅"开发"、"管理员"角色具备提交发布包的权限。 仅"运维"、"部署"、"管理员"具备正式发布的权限。 			
	说明 任务节点的调度依赖是可以跨工作空间配置的。			

功能模块	权限模型
任务运维	 各工作空间之间的任务运维是部分隔离的。 实时/周期/手动任务运维:在工作空间间隔离,并且只有在对应工作空间内具备"开发"、"运维"或"管理员"角色的成员才有权限操作。 运维大屏:工作空间间隔离,运维大屏显示工作空间内任务运行的汇总数据,任务运行情况一览便知。 智能监控:在工作空间间打通,智能监控可以设定跨工作空间的监控基线。仅阿里云主账号、工作空间管理员具备创建智能监控基线的权限。
数据地图	各工作空间之间的数据地图是 租户内共享 的。 在数据地图中,搜索和展示的范围均为当前租户、当前Region下所有工作空间的元数 据。
数据质量	各工作空间之间的数据质量是 完全隔离 的。 仅对应工作空间的"开发"、"运维"或"管理员"角色具备配置数据质量规则的权限。
数据服务	各工作空间之间的数据服务是 部分隔离 的。 各工作空间会共享服务分组的定义,但是注册或发布的服务API仅在当前工作空间可见。
数据保护伞	各工作空间之间的数据保护伞是 全局共享 的。 所有工作空间共享一套数据安全策略和敏感数据目录。若数据保护伞的"高级安全模 式"开启,则只有各工作空间的"安全管理员"具备操作数据保护伞的权限。

通用准则

无论采用何种维度规划创建工作空间,建议遵循以下通用准则:

- 工作空间命名方式统一,包含明确的业务含义任务节点名称包含工作空间名称缩写,便于快速识别。
- 工作空间应由主账号或"组织管理员"创建,并委派"工作空间管理员"。原则上任何人不应同时具备"开发"、"运维"角色。

工作空间规划实践

工作空间规划可按照公司部门、公司业务或数仓层次进行规划,或综合三种维度进行混合规划:

细分 按部门划分 按业务划分	按数仓层次划分
----------------	---------

DataWorks公共云合集·基础操作

细分	按部门划分	按业务划分	按数仓层次划分
划分依据	工作空间的划分可以与公司的 组织架构相一致。 例如:生产部、营销部、人力 资源部、财务部等。各工作空 间承载部门内部的数据开发需 求,管理各自的数据表。	工作空间的划分也可以根据具 体业务项目规划。 例如:"季度销售冲刺战 役"、"春季安全生产大检 查"或"高管驾驶舱报 表"等。各业务项目涉及多个 横向部门,对接多个业务系统 的数据,汇总加工,形成数据 产出。	按照数仓的层级结构划分工作 空间,每一层可以有独立的一 个或多个工作空间。 例如:"统一数据接入"、 "ODS层"、"数仓汇总 层"等。
适用场景	部门业务单一,部门内部人员 具备开发能力,数据共享场景 较少,单一部门即可完成端到 端业务开发。	业务优先的攻坚项目 <i>,</i> 多部门 联合项目。	大型数仓,企业数仓公共层, 数据中台。
优点	工作空间成员与组织架构一 致,人员最稳定,数据安全性 最高。同时计算、存储成本归 属清晰。	工作空间内业务专一,人员可 根据业务动态调整,数据链路 清晰,易运维。	数据架构清晰,共享便利,人 员开发技能要求单一,可根据 各层特性分配不同资源。
缺点	容易形成数据烟囱,数据重复 计算、重复存储,跨空间依赖 复杂,资源易争抢。	数据架构不清晰,各业务口径 不一致,工作空间内人员复 杂,数据安全风险高。	开发周期长,运维链路长,标 准模式下上层任务正式发布前 需要修改代码。
架构稳定性	****	***	****
人员灵活性	★☆☆☆☆	****	★★★★☆
业务复杂度	★★☆☆☆	★★★★☆	★★★☆☆
数据安全	****	★★☆☆☆	★★★☆☆
可运维性	★★☆☆☆	****	★★☆☆☆
数据共享	***	***	****

以上三种划分模式可以混合使用,以综合各自优点。一种常用的混合策略是整体按数仓层次划分,但各层内部 并非单一工作空间,而是进一步划分为多个工作空间。

- 数据接入层(STG):按应用系统划分,例如 "stg_营销系统"、 "stg_生产管理系统"等。
 - 任务节点:只有数据集成任务。
 - 数据表:只有原始数据,生命周期短。
 - 空间成员: 各应用系统的DBA。
 - 资源倾斜:数据集成资源组、存储空间。
- 数据清洗层(ODS):按部门划分,不同部门内数据统一口径,清洗掉不宜公开的数据,例如 "ods_人力资源部"、 "ods_生产部"等。
 - 任务节点:只有单一输入、单一产出的SQL任务。
 - 数据表: ODS层表。
 - 空间成员: 各部门委派的数据清洗人员。

- 资源倾斜:时间靠前的(例如0点~2点)的调度资源组、引擎计算资源。
- 数仓整合层(DW): 整合为一个统一的工作空间, 或按照业务域划分, 例如"dw_客户域"、"dw_商品 域"等。
 - 任务节点:只有多输入、单一产出的SQL任务。
 - 数据表: DW层事实表、维度表。
 - 空间成员:数据公共层专职开发人员。
 - 资源倾斜:中期(例如2点~5点)的调度资源组、引擎计算资源、存储空间(应对数据膨胀)。
- 标签数据层(TDM): 整合为一个统一的工作空间, 或按照业务对象划分。
 - 任务节点:只有多输入、单一产出的SQL任务。
 - 数据表:标签表。
 - 空间成员:数据公共层专职开发人员。
 - 资源倾斜:中晚期(例如5点~7点)的调度资源组、引擎计算资源、存储空间(应对数据膨胀)
- 应用数据层(ADS): 按业务划分, 针对各专项业务, 建立独立工作空间。
 - 任务节点: SQL任务、数据集成任务。
 - 数据表: 以满足业务场景为优先。
 - 空间成员:项目组成员。
 - 。 资源倾斜:晚期(例如7点~9点)的调度资源组、引擎计算资源、数据集成资源组。

2.2.2.修改工作空间

工作空间是DataWorks管理任务、成员,分配角色和权限的基本单元,进行DataWorks的各项任务前您需要先 规划并创建好工作空间。后续如果业务有变动,您也可以修改工作空间的配置。本文为您介绍如何修改工作空间。

背景信息

工作空间的配置主要包含基本配置和服务配置两大类。

● 基本配置

基本配置主要包括工作空间名称、显示名、模式、描述、启动调度周期和能下载Select结果几个参数, 其中:

- 工作空间名称:不支持修改。
- 模式:工作空间的模式包括简单模式(不区分开发环境和生产环境)和标准模式(开发环境和生产环境 隔离),主账号可将简单模式的工作空间升级为标准模式,标准模式的工作空间不支持修改为简单模 式。简单模式升级的操作可参见修改工作空间模式,更多工作空间的模式的介绍可参见工作空间模式升级。

其他参数的修改可参见修改工作空间基本配置。

• 服务配置

服务配置包括DataWorks绑定的引擎及引擎的详细配置。

- 如果您想绑定新类型的引擎,操作可参见修改工作空间的服务配置(新增引擎类型)。
- 如果您想修改已绑定的引擎的详细配置信息,或增加某种已绑定引擎的数量,操作可参见修改工作空间
 已绑定引擎的数量或引擎详情。

修改工作空间的服务配置(新增引擎类型)

如果您需要新增一种类型的引擎,可参考以下步骤操作。如果您想增加已绑定引擎的数量或修改已绑定引擎的 详细配置信息,您可参考修改工作空间已绑定引擎的数量或引擎详情进行操作。

- 1. 进入工作列表页。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏,单击工**作空间列表**。
- 2. 在右侧配置页面中单击修改服务配置。
- 3. 在弹出的页面中勾选待添加的新类型的引擎,完成后单击下一步。
- 4. 根据界面提示配置引擎详细信息,配置参数与新建工作空间的配置一致,详情可参见创建工作空间。

修改工作空间已绑定引擎的数量或引擎详情

如果您想增加已绑定引擎的数量或修改已绑定引擎的详细配置信息,可参考以下步骤操作。如果您需要新增一 种类型的引擎,您可参考修改工作空间的服务配置(新增引擎类型)进行操作。

- 1. 进入项目空间管理页面。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 在对应工作空间的操作列单击工作空间配置。
 - iv. 在右侧配置页面中单击更多设置。
- 2. 在工作空间配置页面的计算引擎信息模块,选择引擎类型的页签。
 - 单击增加实例, 可增加已绑定引擎的数量。

⑦ 说明 不同版本的DataWorks中,单个工作空间可绑定的各类引擎的数量有限且不一致,详情可参见各版本功能支持详情,如果您绑定的引擎的数量超过限制,则无法继绑定。

。 修改已绑定引擎的参数。页面中无法修改的参数即不支持修改。

stgreSQL AnalyticDB for MySQL CDH + Titleracite(
+ 1000000	
开发环境 MaxCompute项目名称: doc_test_prod_dev	~
MaxCompute访问者导份: 任务执行者	
	开发环境 MaxCompute项目名称: doc_test_prod_dev MaxCompute访问者身份: 任勞执行者

修改工作空间模式

工作空间的模式包括简单模式(不区分开发环境和生产环境)和标准模式(开发环境和生产环境隔离), 主账 号可将简单模式的工作空间升级为标准模式, 标准模式的工作空间不支持修改为简单模式。

- 1. 登录DataWorks控制台,单击相应工作空间后的工作空间配置。
- 2. 单击工作空间配置对话框中的更多设置,即可进入工作空间配置页面。

您也可以进入数据开发页面,单击右上角的工作空间管理,进入工作空间配置页面。

\$	💸 DataStudio	
Ø	数据开发	数16开发 2 歳 12 0 0 位
*	组件管理	Q 文件名称/创建人 C
Q	临时查询	> 解決方案 品
G	运行历史	 ▼ 型防滞産 ▼ 里
4	手动业务流程	> 🛁 数据集成
亩	回收站	> NaxCompute
		> 🐏 EMR > 😂 数据服务
		▶ 🏪 算法
		> 🔕 通用
		> 器 自定义

3. 在工作空间配置页面,单击模式后的升级为标准模式。

\$	DataWorks	v	
	=		
۵	工作空间配置	配置	
28	成员管理	基本属性	
0	权限列表	工作空间ID:76834	创建日期 : 2018-07-23 11:06:12
~	MaxCompute高级配置	工作空间名称	模式:简单模式 升级为标准模式
*	数据源管理	状态正常	负责人:
		显示名	描述: 🖉

4. 在升级为标准模式对话框中,输入开发环境下的MaxCompute项目名称,并勾选确认要升级此工作空间,单击确定。

注意 此处您需要新建一个简单模式工作空间作为开发环境下的MaxCompute项目,并可以 在工作空间配置 > 计算引擎信息页面,选择使用个人账号或工作空间所有者操作生产环境的数据。

5. 单击确认升级提示框中的确认。

修改工作空间基本配置

如果您想修改工作空间的显示名称等基本配置,可参考以下步骤进行操作。

- 1. 进入工作列表页。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
- 2. 在右侧配置页面中单击工作空间配置。
- 3. 在弹出的页面中修改显示名称、描述等基本信息。
- 4. 更多其他工作空间的信息修改,可单击更多设置,根据界面提示进行修改。

2.3. 权限管控

2.3.1. DataWorks权限管控概述

DataWorks在**产品管控、功能使用**等方面提供了完善的权限管控体系,其中产品管控通过阿里云RAM Policy 实现,管控对象是控制台操作,如创建工作空间等;功能使用权限通过RBAC实现,通过定义成员角色并授权 来管控功能使用,并根据产品功能范围分为**全局级**和空间级。本文为您详细介绍DataWorks的权限体系。

权限体系框架

根据DataWorks权限管控的对象进行分类,DataWorks的权限体系框架可分为以下几种。

管控对象	管理方式	权限管控的界面范围
产品管控-控制台	RAM Policy ⑦ 说明 RAM Policy 是阿里云提供的用语法 结构描述的一组权限的 集合,可以精确地描述 被授权的资源集、操作 集以及授权条件。 权限管控详情请参见管控权限 详情:细分RAM Policy。	Office Office <thoffice< th=""> <thoffice< th=""> <thoffice< t<="" td=""></thoffice<></thoffice<></thoffice<>
功能使用-空间级	内置RBAC (Role-based access control) ⑦ 说明 空间级的功 能使用权限使用空间级 的用户角色来管控。	● DataWorks DataStudio ① tot 2 ● ● ● DataWorks DataStudio ② tot 2 ● ● ● DataWorks DataStudio ② tot 2 ● ● ● DataWorks(③ DataStudio ① tot 2 ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●
功能使用-全局级	内置RBAC (Role-based access control) ⑦ 说明 全局级的功 能使用权限使用全局级 的用户角色来管控。	● Control Extent 10 THE CONTROL CONT

2.3.2. 管控权限详情: 细分RAM Policy

DataWorks在**产品管控、功能使用**等方面提供了完善的权限管控体系,其中产品管控通过阿里云RAM Policy 实现,管控对象是控制台操作,如创建工作空间等操作,本文为您详细介绍DataWorks控制台的权限体系。

权限管控范围

DataWorks的"产品管控"权限指的是与所有在DataWorks管理控制台中执行的操作相关的权限。例如,工作 空间列表页面中创建空间、禁用空间、删除空间等操作;资源组列表页面中的创建独享资源组、配置独享资 源组网络等操作;报警配置页面的配置联系人等操作。

	âI	作台 📁 华东2(上海) 🗸		Q 搜索	费用 工单 ICP 备案	<u>م</u>	业 支持	App		1 P	7 0	简体	0
DataWorks		DataWorks / 工作空间列表								₹ <i>₹</i> ₿	品动态	帮助文档	ă
概范 工作空间列表		当前使用的是 企业版 , 版本到明日为 2022年1月11	Ξ.			版本	《升级 版	本延期	查看版本	详情	购买独享	资源组	
资源组列表		创建工作空间 请输入工作空间/显示名	Q									C	
报警配置	^	工作空间名称/显示名	模式	创建时间	管理员	状态	操作						
报警资源 报警联系人		ndhulft ndhulft	简单模式 单环境	2021年11月10日 14:40:44	santie_doctest@test.aliyunid.com	~	进入首都 进入运给 工作空间	5 进入表 集中心 进 可配置 弓	対据集成 i 注入数据地 □撃配置 j	进入数排 图 进, 更多	居开发 入数据服务	5	
开放平台 New 计算引擎列表	^	antii Antiin	简单模式 单环境	2021年10月21日 15:54:38	santie_doctest@test.aliyunid.com	~	进入首页 进入运输 工作空间	5 进入支 住中心 过 可配置 引	数据集成 注 注入数据地 川撃配置 『	进入数排 图 进; 更多	想开发 入数据服务	5	
MaxCompute Graph Compute		terigu terigu	简单模式 单环境	2021年10月21日 14:10:50	santie_doctest@test.aliyunid.com	~	进入首都进入运输	瓦 进入支 住中心 进 可配置 弓	牧据集成 注 注入数据地 □撃配置 ₪	进入数排 图 进, 更多	居开发 入数据服务	5	
Hologres		Test 0.1 Test 0.1	标准模式 开发跟生产隔离	2021年9月28日 10:59:09	santie_doctest@test.aliyunid.com	~	进入首页 进入运输 工作空间	5 进入委 住中心 送 可配置 弓	対据集成 注 注入数据地 □撃配置 및	进入数1 图 进, 更多	居开发 入数据服务	5	

您可以进入DataWorks管理控制台查看DataWorks控制台支持执行的操作。

针对DataWorks控制台的操作:

- DataWorks为您提供了内置的权限策略便于您直接使用,给RAM用户授权大范围的权限,详情可参见下文的控制台大范围权限管控:系统内置策略。
- DataWorks也为您提供了自定义权限所需的细分RAM Policy组成,您可以根据需要自定义设置权限策略进一步细化权限策略管控的范围,后续授权时选择自定义策略,详情可参见下文的控制台细分权限管控:自定义RAM Policy。

使用限制

通过细分RAM Policy对Dat aWorks的管控权限进行自定义管控的功能,当前仅在部分用户中受限应用,还未全面上线。如果您还没有使用此功能的权限,敬请等待功能的全面上线。

控制台大范围权限管控:系统内置策略

DataWorks为您内置了如下系统策略,您可以直接给RAM用户授权时选择此权限策略,为RAM用户授予对应的 权限。

系统内置权限策略	权限控制
AliyunDataWorksFullAccess	DataWorks工作空间所有操作权限
AliyunBSSOrderAccess	购买相关服务的权限

给RAM用户授权的操作步骤请参见下文的授权操作步骤。

控制台细分权限管控: 自定义RAM Policy

自定义RAM Policy的结构包含:

- 版本号。
- 授权语句列表。每条授权语句包括授权效果(Effect)、操作(Action)、资源(Resource)以及限制条件(Condition,可选项)。



⑦ 说明 更多RAM Policy的语法结构详情请参见权限策略语法和结构。

当前针对DataWorks控制台各细节操作的权限管控, DataWorks细分出以下的自定义策略组成。

⑦ 说明 Resource说明:

- 实际自定义策略时,您需要将下表中Resource中占位符 \$ 表示的内容替换为真实ID值。例如, \$regionid 需要替换为真实的地域ID值、 \$accountid 需要替换为阿里云主账号的UID。
- * 为通配符,实际使用时您可以替换为具体的参数值,用来做进一步权限管控范围的细化。例如,将workspace/* 替换为workspace/workspaceid 则表明权限策略生效的范围为指定的这个工作空间内。

DataWorks							
概览		当前使用的是基础版,版本到期日为	· 永久 。			版本升级 版本延期	查看版本详情 购买独享资源组
工作空间利用		Createwor	kspace				
LIFERING		创建工作空间 doc	Q				6
资源组列表		工作空间名称/显示名	模式	创建时间	管理员	状态 操作	ModifyWorkspace
报 验配置 报警资源	^	No. doc. No. Reydon, Inte	简单模式 单环境	2021年10月14日 09:52:00	●私戸書に下意所示。@test.aliyunid.com	进入首页 进入数 ✓: 工作空间配置 引 进入运维中心 进	据开发进入数据集成 攀配置进入数据服务 入数据地图更多
报警联系人 开放平台 New		No. doc. Sol. Rev. doc. Sol.	简单模式 单环境	2021年9月7日 17:30:49	●松戸航空下生活中。②testaliyunid.com	进入首页 进入数 ✓ : 工作空间配置 引 进入运维中心 进	据开发 进入数据集成 章配置 进入数据服务 DeleteWorkspan 入数据地图 更多
计算引擎列表 MaxCompute	^ <	No. doc. NN. Rey, doc, not	简单模式 单环境	2021年8月23日 15:49:57	●松戸敷在下生にも、@testaliyunid.com	进入首页 进入数 ✓: 工作空间配置 引 进入运维中心 进	勝升友 进入 等配置 进入 外数器地図 人数器地図 人数器地図
Graph Compute Hologres		No., doc. Int. Res., doc. Int.	标准模式 开发跟生产隔离	2020年12月29日 18:23:52	BERTHER Ditestaliyunid.com	进入首页 进入数 ✓: 工作空间配置 引 进入运维中心 进	据开发 进入数据集成 擊配置 进入数据服务 入数据地图 更多
数据湖集成 (湖仓一体)		No. doc. test.	简单模式 单环境	2020年11月10日 11:32:16		Chablevv の: 激活 删除工作空	的KSpace 间
		Bir, doc test. Bir, doc, test	标准模式 开发跟生产隔离	2020年11月2日 16:45:21	●部門版C下意所を、@test.aliyunid.com	进入首页 进入数 ✓ : 工作空间配置 引 进入运维中心 进	据开发 进入数据集成 季配置 进入数据服务 入数据地图 更多
Action		Resourc	e				Action说明
CreateWorkspace acs:dataworks:\$regionid:\$accountid:workspace/* 创建工作空					创建工作空间		

• 功能分类: 工作空间

Action	Resource	Action说明
ModifyWorkspace	acs:dataworks:\$regionid:\$accountid:workspace/\$workspaceN ame	修改工作空间
DeleteWorkspace	acs:dataworks:\$regionid:\$accountid:workspace/\$workspaceN ame	删除工作空间
DisableWorkspace	acs:dataworks:\$regionid:\$accountid:workspace/\$workspaceN ame	禁用工作空间
EnableWorkspace	acs:dataworks:\$regionid:\$accountid:workspace/\$workspaceN ame	启用工作空间

• 功能分类:资源组

	_	ListResourceGroup	CreateResourceGroup					
概范		独享资源组 公共资源组 自定义多	资源组					
工作空间列表		0.50.007/22200 0.50.00.02222000	●日本語2-3-293時日 i市は2人物学会が400日 ○					C
资源组列表			And a second sec					÷
报警配置	~	资源组名称/备注	英型	状态▽	到期时间	数量	使用率	操作 ModifyResourceGroup
报警资源			1.11					查看信息 运输助手 扩容/缩容 续费 网络设置 修改归属工作空间 变更规格
报警联系人		10 M I I I I I I I I I I I I I I I I I I					10 C	升配 刷新 续费 修改归属工作空间
开放平台 New		100.00						查看信息 运缴助手 扩容 缩容 续费 网络设置 修改归属工作空间 变更现格
MaxCompute		- 19 A.	1000			ShowResou	urceGroupDetail	查看信息 扩容 缩容 续费 网络设置 修改归属工作空间 变更现格
Graph Compute		r_shanghai_range ger集群使用资源组	调度资源组	✓ 运行中	2022-02-23	1	0%	查看信息 运弹助手 扩容 缩容 续费 网络设置 修改归属工作空间 支更规格
Hologres 数据胡集成(湖仓一体)		1000						查看信息 扩容 缩容 续费 网络设置 修改归属工作空间 支更规格
								查看信息 运维助手 扩容 缩容 续费 网络设置 修改归属工作空间 变更现格
Action		Resource		Act	ion说明		注意事	项

DataWorks公共云合集·基础操作

Dat aWorks

Action	Resource	Action说明	注意事项
List Resource Grou p	acs:dataworks:\$regionid:\$accountid: exclusive_resource_group/*	展示独享资源组页 签。如果没有此授 权,则在控制台中 看不到独享资源组 的页签。	ListResourceGroup和 ShowResourceGroupDet ail通常组合应用于控制用 户查看独享资源组页签的 权限。
ShowResourceGr oupDet <i>a</i> il	/ResourceGr het ail acs:dataworks:\$regionid:\$accountid: exclusive_resource_group/\$resourceGr oupName		 仅有 List ResourceGroup权 限时:用户可看到独享 资源组页签,单页签中 没有内容,为空白页 面。 同时有 List ResourceGroup、 ShowResourceGroup Detail:用户可看到独 享资源组页签,且页签 中展示 ShowResourceGroup Detail指定的资源组详 情。 ① 说明 授权 ShowResourceGrou
			pDetail前必须先授 权 ListResourceGroup ,仅有 ShowResourceGrou pDetail权限时无法 查看独享资源组详 情。
CreateResourceG roup	acs:dataworks:\$regionid:\$accountid: exclusive_resource_group/*	创建独享资源组。	非购买独享资源组,此处 为授权用户可根据已购买 的订单在DataWorks的控 制台上创建资源组。资源 组的购买、扩容、缩容、 续费、规格变更等资源购 买相关操作需要有 AliyunDataWorksFullAcc ess和 AliyunBSSOrderAccess权 限。
ModifyResourceG roup	acs:dataworks:\$regionid:\$accountid: exclusive_resource_group/\$resourceGr oupName	修改独享资源组。	-

• 功能分类:报警信息

						-
DataWorks	DataWorks / 报警配置一报	醫资源				产品动态 帮助文档
概范	 DataWorks的报警资源 费。了解更多> 	原已于2020-07-01开始商业化。相	我都你使用的DataWorks版本,	我们每月将免费赠送一定额度的报警短信	和报警电话,附赠额度用完之后将按照	短信0.045元/条,电话0.15元/通的标准进行计
工作空间列表	当日报警资源使用]情况				SetUpperLimits
报警配置 ListAlarmReso	urce 2022-01-19					设置报警数量上限 C
报警资源	报警资源	数据集成		智能监控	数据质量	总计
祝整联系人	短信	0		0	0	0
	电话	0		0	0	0
井飯平台 West インジョン インシー インシー インシー インシー インシー インシー インシー インシ	本月报警资源概览	<u>9</u>				
MaxCompute	报警资源	本月总用量	附赠额度	已用附赠额度	剩余附赠额度	按量付费用量
Graph Compute	短信	0	100	0	100	0
Hologres	电话	0	0	0	0	0
DateWorks	DataWorks / 16888-98-168	欧Ⅲ·死 ↓				
DataWorks	DataWorks / 报警配置一报	警联系人	-			产品动态 帮助文档
DataWorks 概览	DataWorks / 报警配置一级 ① 如下为已加入至DataW	醫联系人 Works工作空间的RAM子账号或R	AM角色(Role)成员信息。			产品和此 ⁹ 相助文档
DataWorks 概定 工作空间列表	DataWorks / 报酬配置一级 ① 如下为已加入至DataW ① 手机、邮件将被应用3	警察系人 Works工作空间的RAM子账号或R 面运维中心任务监控、数据质量监	AM角色(Role)成员信息。 12拉、审批流程通知。			7*65555 ⁹ 1685285
DataWorks 概范 工作空间列表 资源组列表	DataWorks / 採園企園一振引 ③ 如下为已加入至DataW ③ 手机、邮件将被应用益 RAM子歌号 RAM子歌号	設現為人 Works工作空间的RAM子账号或R 回三値中心任务监控、数据氏量編 Role	AM角色(Role)成员信息。 注控、审批流程遵知。			P* <u>Ban</u> k [®] Remixes
DataWorks 概范 工作空间列表 波源和列表 报题配置 个	DataWorks / 探察配置一探 ③ 如下为已加入至DataW ④ 手机、邮件将被应用3 RAM子账号 RAM	證联系人 Works工作空间的RAM子账号或R 定运维中心任务监控、数据质量监 Role	AM角色(Role)成员信息。 註述、审批编程通知。			<u>7*63302€ 1680.248</u>
DataWorks 概意 工作空间列表 流游组列表 探察配置	DataWorks / 报题配置一段 ③ 如下为已加入至DataW ④ 手机、邮件将被应用3 RAM子聚号 RAM 等。 (RAM) 取 (RAM) 和 (RAM) A (RA	副歌 (高 広 加 中 の は の の の の 本 子 教 号 成 れ 子 教 号 成 れ 子 教 号 成 れ 子 教 号 成 れ 子 教 号 成 れ 子 教 号 成 れ 子 教 号 成 れ 子 教 号 成 れ 子 教 号 成 れ 子 教 号 成 れ 子 教 号 成 れ 子 教 号 成 れ 子 教 号 成 れ 子 教 号 成 れ 子 教 号 成 れ 子 教 号 成 れ 子 教 号 成 れ 子 教 号 成 れ 子 教 日 の た か 一 か 子 教 日 の た の た の の の の の の の の の の の の の	AM角色 (Role) 成员信息。 1拉、中世派和建筑机。	٩		<u>₩63302</u> 和助文45
DataWorks 概定 工作空间列表 按照配置 ^ 限额资源 ListContact 报警联系人	DataWorks / 探察配置一級 ③ 如下为已加入至DataWorks ④ 手机、邮件将被应用3 RAM子账号 RAM子账号	副歌馬人 Works工作空间的RAM子教号或R 面這個中心任务協注、飲服民業協 Role 高力に 属控素RAM子教号 RA	AM角色 (Role) 成员信息。 由拉、中批流程道知。 3 M图示名	Q ##16	803B	アビステルシージ 作用的文化的 C ModifyContacts 操作
DataWorks 概意 工作空间列表 按照配置 探照改置 派 联系改要 ListContact 报整改要人 开放平台 [20]	DataWorks / 探察配置一段 ③ 如下为已加入至DataW ④ 手机、邮件将被应用3 RAM学账号 RAM RAM学账号 zhaohuifen	副歌 (新人 Works. 江作空间的RAM子報号或RA 電道地中心任务論社、就預5篇編集 Role 高力11 「 成定表RAM子報号 RA	AM编色 (Role) 成员信息。 由拉、申批派程建和。 3 M图示名 sobulien	Q 手机号 无	節箱 无	C ModifyContacts 18/1 Set
DataWorks 概意 工作空间列表 透影和列表 授聖記章 ^ 探聖記章 ListContact 提聖表承人 开放平台 CCC	DataWorks / 探察配置一段 ③ 如下为已加入至DataW ④ 手机、邮件将被应用3 RAM子账号 RAM :S 阿珍切问控制(RAM) 截0 RAM子账号 zhaohuifen Bin_User	副歌 (新人 Works: I作空间的RAM子報号或RA E 該領中-O任务論社、 就預5集編編 Role 高力11	AM编卷 (Role) 成员信息。 a技、中社场和建筑和。 3 M图示名 aschurien User	Q 手机号 无 无	<i>邮箱</i> 无 无	C ModifyContacts 18/1 Sea Sea Sea
DataWorks 概意 工作空间列表 资源和列表 规图配置 ^ 报图配置 / 报图数源 UstContact 报图数系人 开放平台 CCC 计留引厚列表 ^ MacCompute	DataWorks / 探察配置一段 ③ 如下为已加入至DataW ④ 手机、邮件将被应用3 RAM子账号 RAM S 网络印刷控制(RAM)配引 RAM子账号 zhaohuifen Bin_User	副歌馬人 Works 正作空间的RAM子教号或R 高速地中心任务出社、数据医量型 Role 第751 第25天RAM子教号 RA 2h 2h 8h	AM编色 (Role) 成员信息。 a拉、中批论和显然A。 3 M图示名 asobuifen	Q 手机号 无 无	<i>邮</i> 箱 无 无	ア (1) 下一页 >
DataWorks 概范 工作空间列表 透影和列表 授聖記章 个 报聖記章 ListContact 报聖表承 开放平台 CCC Li第3(写列表 个 MacCompute Graph Compute	DataWorks / 探察配置—解 ③ 如下为已加入至DataW ④ 手机、蛇杆将被应用3 RAM子账号 RAM :S 同步机间控制(RAM)截3 RAM子账号 zhaohuifen Bin_User	副歌 (新人 Works 正作空间的RAM子報号)成用 正道地中心任务協注、 数据质量送 Role 第八寸に 構造表RAM子報号 RA 2h 2h 8h	AM角色 (Role) 成员信息。 由拉、中批流程道知。 3 M图示名 sobuifen LUser	Q 手机号 无 无	<i>邮箱</i> 无 无	ア (1) 下一页 >

Action	Resource	Action说明
ListContacts	acs:dataworks:\$regionid:\$accountid:contacts_ram_user/*	列出联系人
ModifyContacts	acs:dataworks:\$regionid:\$accountid:contacts_ram_user/*	修改联系人信息
ListAlarmResource	acs:dataworks:\$regionid:\$accountid:alarm_resource/*	列出报警资源
SetUpperLimits	acs:dataworks:\$regionid:\$accountid:alarm_resource/*	设置报警数上限

进行管控台细分权限的授权时,您需要首先参考上表中的权限细分来自定义权限策略,然后给RAM用户绑定自 定义的权限策略进行授权操作,操作步骤请参见下文的<mark>授权操作步骤</mark>。

授权操作步骤

您可以使用阿里云账号(主账号)或具有管理权限(AdministratorAccess)的RAM用户进行授权操作,详细操作步骤如下。

1. (可选)新建自定义权限策略。

如果您希望通过细分的RAM Policy来做细粒度的权限管控,您需要根据实际需要先完成创新权限策略,如 果您使用内置策略进行大范围的授权则无需进行此步骤。

您可以使用阿里云主账号登录RAM控制台后,在**权限管理 > 权限策略**页面新建自定义权限策略,操作详 情请参见创建自定义权限策略。 当前DataWorks仅支持使用脚本编辑的方式进行创建,创建时,您可以根据需要参考当前已支持的细分自 定义权限策略来配置权限,要点如下。

RAM 访问控制	← 创建权限策略							
概览								
身份管理へ	 您正在使用全新设计的权限策略编辑器,新版致力于认 	上自定义策略更易懂易用。我们非常	常重视您的使用反馈,请告诉我们您的想法。					
用户	可视化编辑 Beta 脚本编辑							
用户组	A DATI CARAGE A HISTOR							
角色	策略文档长度 216 个字符							
设置	1 - {							
町曾 022	2 version : 1, 3 * "Statement": [
2000 官理 へ	4 * { 5 "Effect": "Allow",							
1200日年	6 "Action": "datawork ListResourceGr	pup", hai:191223 ::exclusive	resource group/*					
权阻禁败		<u>^</u>						
Action	Resource	Action说明	注意事项					
1			ListResourceGroup和ShowResourceGroupDetali通常组合の用工作利用自由素を体験で確認力なない環境					
		展示独享资源组页签。如果没有此授权,	○ 仅有ListResourceGroup权限时:用户可看到独享资源 如而驾,单而驾中没有内容,为众日而而					
ListResourceGroup	acs.dataworks.Sregionid Saccountid exclusive_resource_group*	则在控制台中看不到线事资源相的页签。	和风田,中风田子说中了男,为王山风幽。 。同时有LisResourceGroup, ShowResourceGroupDetail:用户可看到线掌资源组 页歪。直至弦中展示ShowResourceGroupDetail指定 的资源指详情。					
ShowResourceGroupDetail	acs.dataworks.Sregionid.Saccountid.exclusive_resource_group/SresourceGroupName	根据资源组名称展示资源组详情。	⑦ 調轉 提校ShowResourceGroupDetal前必须先 接校LisResourceGroup. 仅有 ShowResourceGroupDetal权限时无法查查按 事资源组详确。					
CreateResourceGroup	acs:dataworks:Sregionid Saccountid exclusive_resource_group/* 創識技家資源組。 市場実施家資源組。 新春には加いの体验分支約台上自識資源機、資源組約向 上部業務に、 市高、構造、体験、所格支援等資源研究性共振作業 有以NunBSSCriderAccess切扱。							
ModifyResourceGroup	acs.dataworks.\$regionid:\$accountid.exclusive_resource_group/\$resourceGroupName	修改独享资源组。						

- ⑦ **说明** Resource说明:
 - 实际自定义策略时,您需要将下表中Resource中占位符 \$ 表示的内容替换为真实ID值。例如, \$regionid 需要替换为真实的地域ID值、 \$accountid 需要替换为阿里云主账号的UID。
 - * 为通配符,实际使用时您可以替换为具体的参数值,用来做进一步权限管控范围的细化。例如,将 workspace/* 替换为 workspace/workspaceid 则表明权限策略生效的范围 为指定的这个工作空间内。

典型场景的脚本编辑代码示例请参见下文内容。

- o 应用示例1: 自定义角色, 授予角色修改工作空间权限
- 应用示例2: 自定义角色, 授予角色查看并管理某个独享资源组的权限
- o 应用示例3: 自定义角色, 授予角色查看报警资源、设置报警数量上线、查看联系人权限
- 应用示例4: 自定义角色, 授予角色查看所有上海地域资源组、创建及修改独享资源组的权限
- 2. 给对应的RAM用户授权。

您可以使用阿里云主账号登录RAM控制台后,在**权限管理 > 授权**页面为RAM用户授权,操作详情请参 见<mark>为RAM用户授权</mark>。

应用示例1: 自定义角色, 授予角色修改工作空间权限

脚本编辑的权限策略代码如下。

```
{
    "Statement": [
        {
            "Action": "dataworks:ModifyWorkspace",
            "Effect": "Allow",
            "Resource": "acs:dataworks:$regionid:$accountid:workspace/$workspaceName"
        }
    ],
    "Version": "1"
}
```

应用示例2: 自定义角色, 授予角色查看并管理某个独享资源组的权限

脚本编辑的权限策略代码如下。

```
{
"Statement": [
{
"Action": "dataworks:ListResourceGroup",
"Effect": "Allow",
"Resource": "acs:dataworks:*:1111:exclusive resource group/*"
},
{
"Action": "dataworks:ShowResourceGroupDetail",
"Effect": "Allow",
"Resource": "acs:dataworks:*:11111:exclusive resource group/resourceGroupName2"
},
{
"Action": "dataworks:ModifyResourceGroup",
"Effect": "Allow",
"Resource": "acs:dataworks:*:111:exclusive_resource_group/resourceGroupName2"
}
],
"Version": "1"
}
```

应用示例3: 自定义角色, 授予角色查看报警资源、设置报警数量上线、查看联 系人权限

脚本编辑的权限策略代码如下。

Dat aWorks

```
{
 "Statement": [
   {
     "Action": "dataworks:ListAlarmResource",
     "Effect": "Allow",
     "Resource": "*"
   },
    {
     "Action": "dataworks:SetUpperLimits",
     "Effect": "Allow",
     "Resource": "acs:dataworks:$regionid:$accountid:alarm_resource/*"
   },
   {
     "Action": "dataworks:ListContacts",
     "Effect": "Allow",
     "Resource": "acs:dataworks:$regionid:$accountid:contacts_ram_user/*"
   }
 ],
 "Version": "1"
}
```

应用示例4: 自定义角色, 授予角色查看所有上海地域资源组、创建及修改独享 资源组的权限

⑦ 说明 创建独享资源组权限指选择订单来新建资源组,而非购买独享资源的权限。

脚本编辑的权限策略代码如下。

```
{
  "Statement": [
    {
      "Action": "dataworks:ListResourceGroup",
      "Effect": "Allow",
      "Resource": "acs:dataworks:*:$accountid:exclusive resource group/*"
    },
    {
      "Action": "dataworks:ShowResourceGroupDetail",
      "Effect": "Allow",
      "Resource": "acs:dataworks:cn-shanghai:$accountid:exclusive_resource_group/*"
    },
    {
      "Action": "dataworks:CreateResourceGroup",
      "Effect": "Allow",
      "Resource": "acs:dataworks:cn-shanghai:$accountid:exclusive resource group/*"
    },
    {
      "Action": "dataworks:ModifyResourceGroup",
      "Effect": "Allow",
     "Resource": "acs:dataworks:cn-shanghai:$accountid:exclusive resource group/resourceGrou
pName1"
    },
    {
      "Action": "dataworks:ModifyResourceGroup",
     "Effect": "Allow",
      "Resource": "acs:dataworks:cn-shanghai:$accountid:exclusive resource group/resourceGrou
pName2"
   }
 ],
  "Version": "1"
}
```

2.3.3. 使用权限详情: 用户与角色

2.3.3.1. 用户、角色与权限概述

DataWorks在**产品管控、功能使用**等方面提供了完善的权限管控体系,其中功能使用方面,针对功能的范围 是全局功能还是空间级功能,分别提供全局角色、空间级角色来实现权限管控,本文将为您介绍当前 DataWorks如何通过RABC(Role-Based Access Control)方式来管控功能使用权限。

⑦ 说明 DataWorks的"产品管控"权限指的是与所有在DataWorks管理控制台中执行的操作相关的权限。例如,工作空间列表页面中创建空间、禁用空间、删除空间等操作;资源组列表页面中的创建独享资源组、配置独享资源组网络等操作;报警配置页面的配置联系人等操作。DataWorks针对产品管控方面的权限管控是通过RAM Policy来实现的,详情请参见管控权限详情:细分RAM Policy。

全局功能及空间级功能

登录DataWorks管理控制台并进入DataWorks工作空间后,单击顶部菜单栏左侧的 = 图标,即可显示

DataWorks产品的所有功能模块,如下图所示。

DataWorks Data	Studio			●任务发布 ● 誇项目克糖 ● 运维中心 📫 17 🔍 💚 🔤 🗧
器全部产品 ·>	教報汇整			
Co 数据集成	Co 数据集成			
X DataStudio(截振开发)				
经 运维中心(工作流)	数据开发			
❷ 数据地图	O DATABLAU	X DataStudio(数据开发)		
◎ 数据质量	数 款服务	截案分析		
迁移助手	任务运维			
數葉综合治理	🌦 运维中心(工作流)			
	数据治理			₩C m
	要全中心	◎ 数据质量		280 *** 停止-F9
	秀 数据保护中			保存・Crtris I Crtris I Crtris I Crtris I
	5. million		<u> </u>	1903年 - Linna Z 派説 - Crat-Y I Crat-Y
	Stakoung	SAME ALL ALL A		査询→Craf+F Crad+F
	其他			wase - unit-init - Unit-Shift+K I Cindi-Shift+K
	S. 代码评审	🍿 DataWorks(首页) New		同词28/#-Crit+D1 (Crit+D
	◆ 迁移助手			39123/# - Luti+ / Luni+/ 列始報道 - Shifty Alt
ALCHYMIC:				

单击功能模块即可进入对应功能模块页面并开始使用此功能。进入功能模块页面后:

- 如果界面顶部菜单栏显示DataWorks工作空间名称,则此模块为"空间级别模块",例如DataStudio。
- 如果界面顶部菜单栏不显示DataWorks工作空间名称,则此模块为"全局级别模块",例如数据地图。

⑦ 说明 区分全局级功能和工作空间级功能详情请参见下文的附录:如何区分"空间级别模块"和"全局级别模块"。

全局角色与空间级角色

DataWorks功能使用的权限体系是基于RBAC(Role-based access control)权限模型构建的,对全局功能和 空间级功能可以分别定义全局角色或空间级角色,将一个用户(RAM用户或Role)添加为某个角色后,这个用 户即可拥有此角色包含的DataWorks相关功能模块的使用权限。

您了解的关键概念有:

- 用户:包含RAM用户、RAM Role两类。
- 角色:包含DataWorks空间级角色、DataWorks全局级别角色两类。
- 权限:包含DataWorks空间级别功能模块、DataWorks全局级别功能模块。

DataWorks为您预设了部分全局角色和空间级角色,您可以直接使用这些角色给用户授权,也可以根据需要, 自定义全局角色或空间级角色。用户、角色、权限之间的对应关系,如下图所示。



? 说明

- 在所有角色中, 仅"全局级角色"中的租户管理员角色拥有所有功能模块的使用权限。
- 阿里云主账号下所有RAM用户均为被默认添加为租户成员角色。
- 如果租户管理员自定义了某个全局级别角色,并指定了该角色不具备某些全局级模块的使用权限,则该自定义角色的权限优先级将高于租户成员的权限。

例如:某个主账号下的RAM用户(RAM用户A),默认情况下为**租户成员**角色,可访问**数据地图**功能页面。当租户管理员自定义了某个角色,并指定该角色无**数据地图**访问权限,并将RAM用户A添加为该自定义角色后,RAM用户A则无法访问数据地图功能页面。

权限详情: 全局角色

DataWorks全局角色包括租户管理员、租户成员、租户安全管理员、数据目录管理员、元数据采集管理员、数据治理管理员这几类,各角色的权限详情如下表所示。

角色	权限点	授权人	描述
租户管理员	拥有DataWorks所有产品 功能的权限(不包含 <mark>阿里 云DataWorks管理控制</mark> 台的操作权限)。	租户所有者(阿里云主账 号)、拥 有AliyunDataWorksFullAc cess的RAM用户、拥 有AdministratorAcces s权限的用户、拥有租户管 理员角色的RAM用户可以 将租户管理员角色授权给 其RAM用户。	DataWorks产品最高权限 管理员,可以操作 DataWorks产品内的所有 功能。

Dat aWorks

角色	权限点	授权人	描述
租户成员	 拥有和目前开发角色同样的使用权限。 数据保护伞只读权限 安全中心普通使用权限(不包含全部权限审计) 数据地图普通使用权限(不包含数据目录管理员、元数据采集管理员的权限) 数据分析使用权限 审批中心普通使用权限(不包含审批策略管理) 	无需授权,默认情况下, 当前阿里云主账号下所有 RAM用户均为 租户成员 角 色。	当前阿里云主账号下所有 RAM用户、RAM Role默认 都是 租户成员 。
租户安全管理员	拥有安全中心、审批中 心、数据保护伞的所有权 限。	租户管理员 可以将其他 RAM用户授权为 安全管理 员。	用于管理工作空间的安全 相关配置。
数据治理管理员	拥有 数据治理中心 的普通 使用权限(不包括服务开 通、检查项的创建、启 用、停止) <i>,</i> 及数据治理 中心的管理权限。	租户管理员 可以将其他人 授权为 数据治理管理员 。	用于数据治理中心的相关 管理。
数据目录管理员	拥有 数据地图 普通使用权 限、管理 数据地图的数据 目录功能权限。	租户管理员 可以将其他人 授权为 数据目录管理员 。	可以管理用数据地图的数 据目录功能。
元数据采集管理员	拥有 数据地图 的普通使用 权限和 元数据采集 权限。	租户管理员 可以将其他人 授权为 元数据采集管理 员。	可以使用数据地图的元数 据采集功能。

权限详情:空间级角色

DataWorks为您提供了多种预设空间级角色,也支持您根据权限管控需要自定义空间级角色。

• 预设空间级角色

DataWorks产品提供:项目所有者、空间管理员、开发、运维、部署、访客、安全管理员、模型设计师这几类预设角色。

⑦ 说明 工作空间所有者为阿里云主账号,子账号仅为代主账号创建工作空间。当前不支持给其他 用户授权为项目所有者。各预设角色的对DataWorks各个功能模块的权限详情请参见附录:预设角色 权限列表(空间级)。

• 自定义空间级角色

DataWorks自定义空间级角色可控制自定义的角色是否有用某个空间级功能模块的权限,新建自定义角色的操作请参见step2: (可选)新建及管理自定义空间级角色。

通过空间级角色对空间级功能进行权限管控时,可细分为两类权限管控对象:DataWorks功能操作、引擎功能操作(例如引擎数据库权限,即表、资源等内容的增删改查权限),预设角色和自定义角色对这两类细分的权

限管控对象的权限详情如下表所示。

权限管控对象	预设角色	自定义角色
Dat aWorks功能操 作	Dat <i>a</i> Works以预设好各角色在各空间级功能 上的使用权限,详情请参见 <mark>附录:预设角色权</mark> 限列表(空间级)。	创建自定义角色时,自定义指定此角色是否 有某个空间级功能的操作权限。
引擎功能权限- MaxCompute引擎	 开发环境的MaxCompute引擎权限: 空间级预设角色默认有部分开发环境的 MaxCompute引擎权限,即,空间级预设 角色的用户可直接访问开发环境下 MaxCompute表。 不同空间级预设角色拥有的开发环境的 MaxCompute引擎权限详情可参见下文 的附录:空间级预设角色与MaxCompute 引擎权限的映射关系。 生产环境的MaxCompute引擎权限: 空间级预设角色没有生产环境的 MaxCompute引擎权限,如果要对生产环 境的MaxCompute表行访问等操作,您需 要在安全中心申请权限,详情请参见数据 访问控制。 	创建自定义角色时,如果配置了自定义角色 与MaxCompute引擎的角色进行映射,则自 定义角色拥有映射的MaxCompute引擎角色 的权限。
引擎功能权限-EMR 引擎	通过绑定EMR引擎时,为空间成员配置权限映射 引擎。	获得引擎权限,详情可参考 <mark>准备工作:绑定EMR</mark>
引擎功能权限-CDH	通过为空间下的成员配置空间成员与Linux或Ker <mark>CDH计算引擎</mark> 。	beros账号映射获得引擎权限,详情可参考 <mark>绑定</mark>
引擎功能权限-其他 引擎	各引擎在联合DataWorks使用前,都需要先将引 绑定到DataWorks的工作空间时,需设置在开发 绑定AnalyticDB for PostgreSQL引擎时需指定认 空间级角色的用户(含预设角色和自定义角色) 置的调度身份,即,空间级角色不直接管控非Ma 定调度身份来设置。	擎绑定到DataWorks的工作空间。其他引擎在 环境、生产环境使用该引擎的调度身份(如, 访问数据库的用户名和密码)。 ,在执行其他擎任务时,就是使用绑定引擎设 axCompute引擎权限,此权限由绑定引擎时指

如上文所述:

- DataWorks提供空间级预设角色时,同时将预设角色与工作空间所绑定的开发引擎的角色进行了映射,使得 预设角色也拥有了映射角色所拥有的部分引擎操作权限。
- DataWorks支持自定义空间级角色,创建自定义角色时,您也可以同时设置角色与引擎角色的映射关系,使得自定义角色拥有引擎操作权限。

因此,使用空间级角色为某个用户授权后,此用户即可拥有DataWorks功能的操作权限和引擎的操作权限。以下以MaxCompute引擎为例,为您介绍,使用空间级预设角色和空间级自定义角色授权后,用户实际的操作权限逻辑。

• 场景一: 授权空间级预设角色权限

示例场景:某个RAM用户被一个拥有空间管理员权限的用户添加为工作空间成员,并授权预设的**开发**角色权限。



完成添加成员操作后,在DataWorks层面和MaxCompute引擎层面均有了部分权限,此RAM用户可在 DataWorks层面和MaxCompute引擎层面进行的操作权限详情如下。

- Dat aWorks层面:给RAM用户授予开发角色后,该用户可以在Dat aWorks开发代码并且提交,但是不能将 代码直接发布到生产环境,发布生产的操作需要有运维权限(项目所有者、管理员、运维这几类角色拥 有此权限)。
- MaxCompute引擎层面: 给RAM用户授予开发角色后,在MaxCompute引擎层面同时为这个RAM用户授 予了一个Role_Project_Dev这个角色,这个角色会被赋予部分MaxCompute引擎开发项目的表和项目的 权限。

? 说明

- 当授予阿里云RAM用户预设管理员角色,其拥有的DataWorks功能权限较多,但同样不能直接访问生产表。
- 此处说明的RAM用户特指,未被指定为非调度引擎访问身份(未被指定为生产项目 MaxCompute访问身份)的子账号。
- 场景二: 授权空间级自定义角色权限

场景示例:某个RAM用户被一个拥有空间管理员权限的用户添加为工作空间成员,并授予Dat aoWrks工作空间自定义角色权限。



新建自定义角色时,您可以指定此角色是否与MaxCompute引擎的某个角色做映射,角色新建完成后,可 以在添加成员时关联此自定义角色,完成后,被添加为成员的RAM用户在DataWorks层面和MaxCompute引 擎层面均有了部分权限,此RAM用户可在DataWorks层面和MaxCompute引擎层面进行的操作权限详情如 下。

⑦ 说明 新建自定义角色的操作请参见step2: (可选)新建及管理自定义空间级角色,添加成员并 授权的操作请参见step3:添加及管理工作空间成员。

- DataWorks层面: DataWorks上给一个RAM用户授予DataWorks自定义角色,该用户只能访问自定义角色 所赋予的模块权限。
- MaxCompute引擎层面:
 - 如果该DataWorks自定义角色未设置引擎权限映射,那么RAM用户没有MaxCompute引擎的操作权限,无法通过命令进行引擎查询等相关操作。
 - 如果该DataWorks自定义角色设置了引擎权限映射,那么RAM用户有自定义角色所映射的引擎权限, 所拥有的操作权限与映射的MaxCompute引擎的角色一致。

⑦ 说明 除了RAM用户被指定为调度引擎访问身份(生产项目MaxCompute访问身份)这一场景外,其他场景下,RAM用户被添加为工作空间成员后,默认没有生产环境操作权限,操作及访问生产表需要在安全中心进行权限申请,详情请参见:通过新版安全中心申请表权限。关于MaxCompute访问身份详情说明,可参考文档:绑定MaxCompute计算引擎。

附录:空间级预设角色与MaxCompute引擎权限的映射关系

DataWorks空间级预设角色与MaxCompute开发项目的权限映射与权限详情如下表所示:

Dat aWorks

Dat <i>a</i> Works成员角 色	MaxCompute角色	DataWorks开发环 境/MaxCompute DEV引擎项目数据权 限	DataWorks生产环 境/MaxCompute PROD引擎项目数据 权限	DataWorks平台权 限特征
项目管理员	Role_Project_Admi n	 引擎层面:当前 项目下project/t able/fuction/re source/instance /job/package的 所有权限。 DW层面:可进行 数据开发,并且 发布任务至生产 环境。 	默认无权限、需要在 安全中心走审批流程 申请	指项目空间的管理 者。可以对该项目空 间的基本属性、数据 源、当擎配目空间计 算引擎起行管理,并为 项目成员赋予项目管 理员、开发、运维、 部署、访客角色。
开发	Role_Project_Dev	 引擎层面:当前 项目下<i>project/f</i> <i>uction/resource</i> <i>/instance/job/p</i> <i>ackage/table</i>的 所有权限。 DW层面:可进行 数据开发,但无 法发布任务至生 产环境。 	默认无权限、需要在 安全中心走审批流程 申请	开发角色的用户能够 创建工作流、脚本文 件、资源和UDF以及 新建和删除表,同时 可以创建发布包,但 不能执行发布操作。
运维	Role_Project_Pe	当前项目空间 下 <i>project/fuction/</i> <i>resource/instance/</i> <i>job</i> 的所有权限,拥 有package的Read 权限和table的 Read/Describe权 限。 ⑦ 说明 引 擎层面有权 限,但在 DataWorks 上,运维角色 不能直接在界 面运行节点进 行任务执行操 作。	默认无权限、需要在 安全中心走审批流程 申请	运维角色的用户由项 目管理员分配运维权 限,拥有发布及线上 运维的操作权限,没 有数据开发的操作权 限。
部署	Role_Project_Depl oy	默认无权限。	默认无权限、需要在 安全中心走审批流程 申请	部署角色与运维角色 相似,但是它没有线 上运维的操作权限。

DataWorks公共云合集·基础操作

Dat <i>a</i> Works成员角 色	MaxCompute角色	DataWorks开发环 境/MaxCompute DEV引擎项目数据权 限	DataWorks生产环 境/MaxCompute PROD引擎项目数据 权限	DataWorks平台权 限特征
访客	Role_Project_Guest	默认无权限。	默认无权限、需要在 安全中心走审批流程 申请	访客角色的用户只具 备查看权限,没有权 限进行编辑工作流和 代码等操作。
安全管理员	Role_Project_Secur ity	默认无权限。	默认无权限、需要在 安全中心走审批流程 申请	安全管理员仅在数据 保护伞模块中使用, 用于敏感规则配置、 数据风险审计等。
无	Project Owner	MaxCompute项目 空间的所有者,拥有 该项目空间的所有权 限。	有权限	无
无	Super_Administrat or	MaxCompute项目 空间的超级管理员, 拥有项目空间的管理 类权限以及项目空间 内所有类型资源的全 部权限。	有权限	无
无	Admin	每一个项目在创建 时,何前动,并自创建一个 Admin角色,并自创建一个 Admin角色,前定词动,并已的。 可的可的可用之一个。 对理、的不可见。 和一个。 和一个。 和一个。 和一个。 和一个。 和一个。 和一个。 和一个	有权限	无

附录:如何区分"空间级别模块"和"全局级别模块"

界面顶部菜单栏显示DataWorks工作空间名称的模块为"空间级别模块",例如DataStudio。

≡	仰 DataWorks DataStudio		❷ 任务发布	❷ 跨项目克隆	❷ 运维中心	ф Ø	ల్ష	The second secon
(/)	数据开发 と同日日の	<u>ን</u> ዊ						
	Q、 文件名称/创建人	™						
	▶ 解决方案	88						
Q	> 业务流程							
G			运行 - F8					
			停止 - F9					
÷			保存 - Ctrl+S Cmd+S					
ш			撤消 - Ctrl+Z Cmd+Z					
			重做 - Ctrl+Y Cmd+Y					
			查询 - Ctrl+F Cmd+F					
			替换 - Ctrl+H Cmd+Alt+F					
			删除一行 - Ctrl+Shift+K C	md+Shift+K				
			同词选择 - Ctrl+D Cmd+D					
~			块注释 - Ctrl+/ Cmd+/					
			列编组 - Shift+Alt					
٢								
≡								

界面顶部菜单栏不显示DataWorks工作空间名称的模块为"全局级别模块",例如数据地图。

≡	① DataWorks 数据地图 首页	数据总览 全部数据 我的数据 配置管理 数	据发现 更多 🗸	ር 🗹 🔍 🗢 🕴
		· · · · · · · · · · · · · · · · · · ·	据地图	
	MaxCo	mpute > 请输入表名、描述等关键字		搜索
il	丘期浏览的表	近期读取的表	热门浏览	热门读取
	dataworks_demo_2021.ods_pg_ad ods_pg_admission_score	▶ lineage_demo.table_student_f table test 分区 表的描述	▶ lineage_demo.table_student_f table test 分区 表的描述	V lineage_demo.table_student_f table test 分区表的描述
	✤ lineage_demo.table_student_f table test 分区表的描述	workshop0423.rpt_user_info_d rpt_user_info_d	vc_dpe_e2.xc_rpt_user_info_d 测试表的描述	workshop0423.rpt_user_info_d rpt_user_info_d
	xal_workshop.bank_data bank_data	dataworks_demo_2021.ods_pg_ad ods_pg_admission_score	workshop0423.rpt_user_info_d rpt_user_info_d	dataworks_demo_2021.ods_pg_ad ods_pg_admission_score
	workshop0423.rpt_user_info_d rpt_user_info_d	1 lineage_demo.table_student_c_13 table test 分区	1 lineage_demo.table_student_f_2007 table test 分区	1 ineage_demo.table_student_c_13 table test 分区
	V lineage_demo.table_student_c_10 table test 分区	Ineage_demo.table_student_c_10 table test 分区	Ineage_demo.table_student_c_1 table test 分区	fineage_demo.table_student_c_12 table test 分区

2.3.3.2. 角色及成员管理:空间级

为了更细粒度的管控权限实现最小化授权,DataWorks为您提供不同的空间级预设角色,也支持自定义空间级 角色,针对实际业务制定权限管控粒度。您可以将阿里云账号(用户)添加为工作空间成员,通过关联角色授 予其角色所拥有的权限。本文为您介绍空间级角色及成员管理操作。

背景信息

DataWorks的工作空间为您提供了成员及角色等身份,您可以针对不同用户的工作空间使用需求,授予其相应的功能权限角色。DataWorks提供预设工作空间级别角色,不同角色拥有的DataWorks相关功能权限请参见附录:预设角色权限列表(空间级)。DataWorks预设角色权限与MaxCompute引擎(开发项目)角色与的对应关系请参见权限详情:空间级角色。

当DataWorks的预设角色不能满足您的需求时,您还可以通过**角色管理**自定义针对空间级功能的DataWorks角色。

添加自定义角色		×
* 输入角色名称:	test	
✓ DataWorks权限		
HoloStudio		1
StreamStudio	● 无权限 ○ 只读 ○ 读写	1
发布管理	💿 无权限 💿 只读 🔵 读写	
数据服务	● 无权限 🔵 只读 🔵 读写	
数据集成	● 无权限 ○ 只读 ○ 读写	
数据建模(DataBlau DDM)	无权限 🦳 只读 🔵 读写	
数据开发	● 无权限● 只读● 读写	
数据质量	 无权限 只读 读写 	
运维中心	💽 无权限 🔵 只读 🔵 读写	
资源优化	💿 无权限 🔵 只读 🔵 读写	
迁移助手	 无权限 只读 读写 	
✓ 配置引擎权限映射		
▲ 请通过阿里云主账号或被授予	MaxCompute项目的ODPS admin、Superadmin角色的子账号进行操作	
	* 项目: 请选择 > * 角色: 请选择 >	
添加		
		-
	开始配置	取消

- DataWorks自定义空间级角色可控制某个该角色是否有用某个空间级别模块的权限,如定义一个角色无数据 分析功能的权限,后续关联此角色的用户即无法访问操作数据分析功能模块。
- 如果您当前使用的是MaxCompute引擎,您也通过配置引擎权限映射使该角色拥有引擎资源的相关操作权限。

新增自定义空间级角色的操作详情请参见下文的step2: (可选)新建及管理自定义空间级角色。

使用限制

- 仅DataWorks企业版工作空间才可以添加自定义角色,详情请参见DataWorks各版本详解。您可以参考版本 服务计费说明,升级DataWorks工作空间至企业版。
- 仅空间管理员角色和工作空间所有者,可以添加成员、修改成员角色、删除成员及已创建的自定义角色。
- 仅支持使用阿里云主账号,或被授予MaxCompute项目admin、Superadmin角色的RAM用户,配置自定义的DataWorks新角色与MaxCompute引擎的权限映射关系。

step1: 进入成员管理页面

- 1. 登录DataWorks控制台。
- 2. 在左侧导航栏,单击工作空间列表。
- 3. 进入**工作空间配置**页面。

您可以通过以下两种方式进入工作空间配置:

○ 单击相应工作空间后的工作空间配置。在工作空间配置对话框中,单击更多设置,进入工作空间配置页面。

工作空间配置	
基本信息	
工作空间名称	
显示名	
* 模式	简单模式(单环境)
描述	
高级设置	
* 启动调度周期 💿	π
* 能下载Select结果 📀	π
更多设置	关闭

○ 单击相应工作空间后的进入数据开发。在数据开发页面,单击右上角的工作空间管理
 图标,进入工作空间配置页面。

6	💸 DataStudio		■ ,	•						∂ 任务发布	∂运维	中心 🔍 💎
	临时查询	ዶ [‡ C ⊕	Sq SELECT	01 ×	Sq sele	ct	🆽 (dw_user_info_i	all_d			工作空间管理
(1)	Q 文件名称/创建人	T		3		Þ	•	C 🗱				
*	✔ 临时查询			odps s	5ql *******							*******
Q	Sq select 我锁诉	萣 07-04 15:35		author	r: a time:2	010-0	7-04 16					
©	Sq SELECT01 ∄	我锁定 07-04 16:15			******							********
ê												

4. 在工作空间配置页面左侧导航栏,单击成员管理,进入成员管理页面。

step2: (可选)新建及管理自定义空间级角色

工作空间预设角色权限无法修改,如果预设角色不满足权限管控需求,您可以通过**角色管理**来自定义空间级 角色,指定此角色是否有用某个空间级别模块的权限。

- 1. 参考step1: 进入成员管理页面进入成员管理页面后, 单击角色管理页签。
- 2. 单击角色管理页签右上角的添加自定义角色。
- 3. 在添加自定义角色对话框输入自定义的DataWorks角色名称。例如: test。
- 4. 为自定义角色配置DataWorks不同空间级别模块的权限。

添加自定义角色		×
* 输入角色名称:	test	
➤ DataWorks权限		
HoloStudio	● 无权限 ○ 只读 ○ 读写	1
StreamStudio	● 无权限 ○ 只读 ○ 读写	
发布管理	 无权限 只读 读写 	
数据服务	 无权限 只读 读写 	
数据集成	● 无权限 ○ 只读 ○ 读写	
数据建模(DataBlau DDM)	● 无权限 ● 只读 ● 读写	
数据开发	 无权限 只读 读写 	
数据质量	 无权限 只读 读写 	
运维中心	 无权限 只读 读写 	
资源优化	 无权限 只读 读写 	
迁移助手	 无权限 只读 读写 	

- 无权限:表示该角色无相应模块的查看权限。
- 只读: 表示该角色只能查看相应模块的数据信息,不能修改模块数据。
- 读写:表示该角色可以修改相应模块的数据信息。
- 5. 单击**配置引擎权限映射**下的**添加**,为自定义的DataWorks新角色配置与其他引擎权限的映射关系,使该 角色拥有对应引擎权限。

⑦ 说明
○ 目前仅支持配置与MaxCompute引擎的权限映射。
◎ 当前仅支持使用阿里云主账号,或被授予MaxCompute项目admin、Superadmin角色的RAM 用户,配置自定义角色与MaxCompute引擎的权限映射关系。
◇ 配置引擎权限映射
▲ 请通过阿里云主账号或被授予MaxCompute项目的ODPS admin、Superadmin角色的子账号进行操作
○*引擎类: MaxCompute ∨ *项目: ∨ *角色: admin × ∨
添加

例如,授权自定义角色(test)访问目标MaxCompute项目时,使用MaxCompute的admin角色。各引擎与DataWorks权限的对应关系请参见<mark>附录:空间级预设角色与MaxCompute引擎权限的映射关系</mark>。

6. 单击开始配置。

当界面提示创建成功时,您即成功完成自定义角色的创建,后续进行添加成员时,可将成员关联此角色。

您可以在角色管理页签,查看DataWorks工作空间已配置的预设角色及自定义角色,或编辑、删除自定义角色。预设角色的权限信息请参见附录:预设角色权限列表(空间级)。

			<u>දී</u> බු ද) 前体 💎
三 〇 工作空間配置	成员管理 <u>命任告</u> 理			添加自定义角色
坐 成员管理	1 预设角色		自定义角色	
秋原列表	角色名称	攝作	角色名称	攝作
✓ MaxCompute高级配置	空間管理员	-	88.2	984R 2019
	款据分析(III	-		编辑 删除
O SUBSECT	部署	-	1000	编辑 删除
开設集群管理	开发	-	自定义角色_数据开发无权限_运维中心_只读权限	编辑 删除
◆ 扩展程序配置	模型设计师	-		
	访察	-		
	项目所有者	-		
	运维	-		
	安全管理员	-		

step3: 添加及管理工作空间成员

1. 参考step1: 进入成员管理页面进入成员管理页面后, 在成员管理页签右上角单击添加成员。

2. 在添加成员对话框中的待添加账号模块,勾选需要添加的成员账号。

加成员				
您可以前往 RAM控制	则台 新建子账号,并单击 刷新 同	同步至此页面。		
•选择组织成员:	待添加账号		已添加账号	
	请输入成员名称	Q	请输入成员名称	Q
	moriartychen		BD	
		<		
	_ 1/53 项	3	1项	
* 批量设置角色:	空间管理员 部署 员	开发 🔽 模型设计	师 访客 项目所有者	运维 安全管
			4	取消

- 3. 单击>图标,将需要添加的RAM用户移动至已添加账号中。
- 4. 在批量设置角色处勾选需要关联的角色,关联后此账号即拥有角色对应的权限。
- 5. 单击确定,完成添加工作空间成员。

在成员管理页签,您还可以执行如下成员管理操作:

● 查看成员信息。

您可以查看当前DataWorks项目空间下,所有成员的账号、角色等信息,也可以根据成员名称、云账号、以及某类别的角色进行筛选,查看目标条件的成员信息及相应角色的成员数量。方便您对成员及相应角色进行集中管控。

	100	я "			<u>දී</u> ව එ, ee#	V 80.00.000
- • Itense	成员管	理由色管理			·译加动)	重着角色权限
基 成员管理	请输入的	2员名或云账号 Q 自定义角	· 自:			
⑦ 权限列表	<u>全部</u>	空间管理员 9 数据分析师 0 部署	4 开发 20 機型设计师 4 访客 6 项目所有者 1 运	2推 5 安全管理员 3		
✓ MaxCompute高级配置		成员	云烁号	角色	加入时间	操作
③ 数据源管理		810.00		空印管理员 × 开設 × → → → → → → → → → → → → → → → → → →	2018-07-26 15:14:52	所有者
▶ 开滚集群管理		1010		开发 × 访客 × 模型设计师 × ×	2018-08-30 16:05:38	8519.
◆ 扩展程序配置		1018	- Converting (2019)	────────────────────────────────────	2019-03-15 14:25:43	818
		1.0	And a start of the	开設 ×	2019-03-22 16:20:44	88 19
		stations	description of the log	☐ 开戏 × 运递 × ▼	2019-10-16 22:03:59	819
		100.0	- Charles and a start water	空印管理员 × 开发 × 运续 × ×	2020-11-30 11:04:41	151 0
		101	And a second	空间管理员 × 防害 × × ×	2019-11-12 17:45:22	819
		10.000	and the second second	开波 × 访察 × 機型设计师 × ▼	2018-08-30 11:25:15	80
		34	and a second	────────────────────────────────────	2018-08-27 13:33:02	1819
			00000,000000	开发 × 部署 × // / / / / / / / / / / / / / / / /	2018-08-29 16:41:22	819
	批量移	76			1 2 3 下一页 > 每页显示	ж: 10 🗸

● 删除成员。

您可以在**成员管理**页面,单击目标成员**操作**列的**删除**,删除单个成员。如果DataWorks工作空间中存在多 个待删除的成员,则您可以勾选多个成员,批量删除。

DataWorks				介 団 �� m ##	•
= ② 工作空间配置	成员管理 角色管理			添加成	查 着角色权限
基 成员管理	请输入成员名或云账号 Q 自	·建义角色: 读选择 ✓			
役 权限列表	全部 空间管理员 9 数据分析师 0	部署 4 开发 20 模型设计师 4 访客 6 项目所有者 1	运程 5 安全管理员 3		
✓ MaxCompute高级配置	成员	云烁号	角色	加入时间	操作
② 数据源管理	International	100 CONTRACTOR	空印管理员 × 开发 × V	2018-07-26 15:14:52	所有者
▶ 开源集群管理	100 1000	and the second sec	开发 × 防審 × 模型设计师 × ✓	2018-08-30 16:05:38	1819
③ 扩展程序配置	1988	- (ac-c-4, (b-c))(c-4)	开发× ~	2019-03-15 14:25:43	移除
		- American State (Control of State	开发 ×	2019-03-22 16:20:44	819
	· ·····	description (see a	────────────────────────────────────	2019-10-16 22:03:59	89
		- No. of Concession, Name	空印管理页 × 开发 × 运续 × V	2020-11-30 11:04:41	819
		And the second second	空间管理员 × 尚容 × ×	2019-11-12 17:45:22	819
	Internet	and a second second second	开没 × 防審 × 機型设计师 ×	2018-08-30 11:25:15	819
	IN	100000 (000000000)	开发 × 安全管理员 × ✓	2018-08-27 13:33:02	8 19:
		here (4,0+1)(4)	开波 × 部署 × ✓	2018-08-29 16:41:22	移除
	投量移除			く 上一页 1 2 3 下一页 > 毎页显示	R: 10 V

附录:查看用户权限

您可以在MaxCompute_SQL任务中,执行如下语句,查询自己的权限信息。

```
show grants --查看当前用户自己的访问权限。
show grants for <username> --查看指定用户的访问权限,仅由空间管理员才有执行权限。
```

更多查看权限的命令请参见通过MaxCompute SQL查询权限信息。

2.3.3.3. 角色及成员管理: 全局级

DataWorks支持以租户全局为管控粒度实现全局功能的权限管控,为您提供为精细化的全局级权限管控。例如 控制某用户是否有全局模块数据地图的类目管理权限管控。DataWorks为您提供部分全局预设角色权限,同时 支持您自定义全局角色。本文为您介绍全局级角色及成员管理操作。

背景信息

进入DataWorks功能模块界面后,当顶部菜单栏不显示DataWorks工作空间名称时,此模块为"全局级别模块",例如**数据地图**。针对此类全部功能模板,DataWorks为您提供了全局级成员及角色等身份,您可以实际全局功能的使用需求,授予不同用户其相应的功能权限角色。

DataWorks为您提供了部分全局预设角色权限,同时支持您自定义全局角色,预设角色及其权限详情如下所示。

预设全局角色	作用全局功能	权限详情
租户管理员	拥有DataWorks所有产品功能的权限 (不包含阿里云DataWorks管理控制 台的管控操作权限,管控权限详情请 参见管控权限详情:细分RAM Policy)。	DataWorks产品最高权限管理员,可以操作DataWorks产 品内的所有功能。
租户成员	所有全局模块	当RAM用户账号被加入某个工作空间后,将默认为租户成 员角色,可访问所有全局模块。

预设全局角色	作用全局功能	权限详情
安全管理员	安全中心、审批中心、数据保护伞	安全中心所有权限审批中心的自定义审批策略权限数据保护伞所有权限

当DataWorks的预设角色不能满足您的需求时,您还可以自定义全局角色,控制某个全局角色是否有用某个全局级模块的权限。当前支持通过自定义全局角色来管控的全局功能模块如下所示。

可管控全局功能	可管控权限详情
数据保护伞	 无权限 可使用:全部只读、全部可操作
数据地图	 无权限 可使用:普通使用权限
数据综合治理	无权限可使用:普通使用权限、数据治理管理
数据分析	 无权限 可使用:普通使用权限
审批中心	 无权限 可使用:普通使用权限、管理审批流程
安全中心	 无权限 可使用:普通使用权限

使用限制

仅DataWorks企业版工作空间才可以添加自定义角色,详情请参见DataWorks各版本详解。您可以参考版本服务计费说明,升级DataWorks工作空间至企业版。

step1: 进入全局成员管理页面

- 1. 登录DataWorks控制台。
- 2. 在左侧导航栏,单击工作空间列表,单击进入数据开发,先进入数据开发页面。
- 3. 在菜单栏左上角单击的 三图标,选择全局成员管理,进入全局管理页面。

step2: (可选)新建及管理自定义全局级角色

工作空间预设角色权限无法修改,如果预设角色不满足权限管控需求,您可以通过**角色管理**来自定义全局级 角色,指定此角色是否有用某个全局级功能的权限。

- 1. 参考step1: 进入全局成员管理页面进入全局成员管理页面后, 单击角色管理页签。
- 2. 单击角色管理页签右上角的添加自定义角色。
- 3. 在添加自定义角色对话框输入自定义的DataWorks角色名称。例如: test。
- 4. 为自定义角色配置不同全局级功能的权限。

* 输入角色名称:	test		
➤ DataWorks权限			
▲ 请通过租户管理员	或被授予AliyunDataWorksFullAccess、Administrate	or策略的 了账号 进行操作!	
数据保护伞		○ 无权限 🤇	可使用
		全部只读 全	部可操作
数据地图		○ 无权限 《	可使用
		普通	使用权限
数据综合治理		○ 无权限 《	可使用
		普通使用权限数据	站理管理
数据分析		○ 无权限 《	可使用
		普通	使用权限
审批中心		○ 无权限 (可使用
		普通使用权限	目 审批流程
安全中心		○ 无权限 《	可使用
		普通	使用权限

5. 单击开始配置。

当界面提示创建成功时,您即成功完成自定义角色的创建,后续进行添加成员时,可将成员关联此角色。

step3:为用户授予全局角色权限

- 参考step1:进入全局成员管理页面进入全局成员管理页面,进入成员管理页签。
 成员管理页签为您展示当前阿里云主账号下所有RAM用户,这些RAM用户即为全局成员。
- 2. 在对应全局成员的角色列添加或删除全局角色,完成对成员的全局角色授权。

			 wsadwsadwsad 				
☰	全局成员管理					🖞 থ্ 💎 🚥 🖬	doctent (hertuniyarikkoon
成 员管理 角色管理 请输入成员名或者云账号进行 全部 租户所有者 1 租	翌茶 Q 自定义角色: 戸管理员 1 租户成员 9	请选择 >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	型贝 0 元数据采集管理员 0 数据	治理管理员 0			i
成员	ī	云账号	所属工作空间	角色		加入时间	操作
santie_doctest@test.aliyunid.	com	uartie_doches/@dext.alignmid.com	文档测试空间_上海,doc_ st_3,test10,Max_F1,dfe,q	test_2,doc_te 租户管理员 ×	~	2020-10-21 16:41:59	租户所有者
Alice	A	Alice	文档测试空间_上海	租户成员 ×	~	2021-11-11 14:36:17	-
ram_test	ra	ram_test		租户成员 ×	~	2022-01-04 10:41:57	-
doctest	s	santie		租户成员 ×	~	2021-12-01 11:57:36	-

⑦ 说明 当RAM用户账号被加入某个工作空间后,将默认为租户成员角色,可访问所有全局模块。

2.3.3.4. 附录: 预设角色权限列表(空间级)

DataWorks提供项目所有者、空间管理员、数据分析师、开发、运维、部署、访客、安全管理员、模型设计师 这几种角色,本文将为您介绍具体角色的权限说明。

DataWorks提供预设工作空间级别角色,各预设角色建议的操作人员描述如下所示。

角色	描述
项目所有者	工作空间所有者拥有工作空间的所有权限,正常为阿里云主账号。例如,可以根据 需求给RAM用户授予相应角色、删除本工作空间非项目所有者的成员等。
空间管理员	除拥有开发角色和运维角色的全部权限外,还可以进行添加/移出工作空间成员并 授予角色,创建自定义资源组等操作。
数据分析师	仅具有数据分析模块的操作权限。数据分析模块详情,请参见数据分析。
开发	负责数据开发页面的设计和维护工作。
运维	负责在运维中心页面管理全部任务的运行情况并进行相应处理。
部署	仅在多工作空间模式时审核任务代码并决定是否提交运维。
访客	仅有只读权限,可以查看数据开发页面的业务流程设计和代码内容。
安全管理员	仅有数据保护伞模块的操作权限,无其它模块权限。数据保护伞模块详情,请参见数据保护伞。
模型设计师	可以查看模型,进行数仓规划、数据标准、维度建模和数据指标等内容的编辑,但 是不能发布模型。

不同预设角色对DataWorks各空间级功能的操作权限详情请参见下文的:

- 数据管理
- 发布管理
- 按钮控制
- 代码开发
- 函数开发
- 节点类型控制
- 资源管理
- 工作流开发
- 数据集成
- 数据建模
- 数据分析

此外,空间级预设角色对MaxCompute类引擎的开发环境数据也具有一定的操作权限,详情请参见权限详情:空间级角色。

数据管理

权限点	项目所 有者	空间管 理员	数据分 析师	开 发	运 维	部 署	访 客	安全管 理员	模型设 计师
自己创建的表删除	有	有	无	有	无	无	无	无	无
自己创建的表类目设置	有	有	无	有	无	无	无	无	无
自己收藏的表查看	有	有	无	有	无	无	无	无	无
新建表	有	有	无	有	无	无	无	无	无
自己创建的表取消隐藏	有	有	无	有	无	无	无	无	无
自己创建的表结构变更	有	有	无	有	无	无	无	无	无
自己创建的表查看	有	有	无	有	无	无	无	无	无
自己申请的权限内容查 看	有	有	无	有	无	无	无	无	无
自己创建的表隐藏	有	有	无	有	无	无	无	无	无
自己创建的表生命周期 设置	有	有	无	有	无	无	无	无	无
非自己创建的表数据权 限申请	有	有	无	有	无	无	无	无	无
更新开发环境的表	有	有	无	有	有	有	无	无	无
删除开发环境的表	有	有	无	有	无	无	无	无	无
数据预览	有	有	无	有	有	有	有	有	无

发布管理

权限点	项目所 有者	空间管 理员	数据分 析师	开 发	运 维	部 署	访 客	安全管 理员	模型设 计师
创建发布包	有	有	无	有	有	无	无	无	无
查看发布包列表	有	有	无	有	有	有	有	无	无
删除发布包	有	有	无	有	有	无	无	无	无
执行发布	有	有	无	无	有	有	无	无	无
查看发布包内容	有	有	无	有	有	有	有	无	无

按钮控制

DataWorks公共云合集·基础操作

权限点	项目所 有者	空间管 理员	数据分 析师	开 发	运 维	部 署	访 客	安全管 理员	模型设 计师
按钮 停止	有	有	无	有	无	无	无	无	无
按钮 格式化	有	有	无	有	无	无	无	无	无
按钮 编辑	有	有	无	有	无	无	无	无	无
按钮 运行	有	有	无	有	无	无	无	无	无
按钮 放大	有	有	无	有	无	无	无	无	无
按钮 保存	有	有	无	有	无	无	无	无	无
按钮 展开/收起	有	有	无	有	无	无	无	无	无
按钮 删除	有	有	无	有	无	无	无	无	无

代码开发

权限点	项目所 有者	空间管 理员	数据分 析师	开 发	运 维	部 署	访 客	安全管 理员	模型设 计师
保存提交代码	有	有	无	有	无	无	无	无	无
查看代码内容	有	有	无	有	有	有	有	无	无
创建代码	有	有	无	有	无	无	无	无	无
删除代码	有	有	无	有	无	无	无	无	无
查看代码列表	有	有	无	有	有	有	有	无	无
运行代码	有	有	无	有	无	无	无	无	无
修改代码	有	有	无	有	无	无	无	无	无
文件下载	有	有	无	无	无	无	无	无	无
文件上传	有	有	无	有	无	无	无	无	无

函数开发

权限点	项目所 有者	空间管 理员	数据分 析师	开 发	运 维	部 署	访 客	安全管 理员	模型设 计师
查看函数详情	有	有	无	有	有	有	有	无	无
创建函数	有	有	无	有	无	无	无	无	无
查询函数	有	有	无	有	有	有	有	无	无

Dat aWorks

权限点	项目所 有者	空间管 理员	数据分 析师	开 发	运 维	部 署	访 客	安全管 理员	模型设 计师		
删除函数	有	有	无	有	无	无	无	无	无		
节点类型控制											
权限点	项目所 有者	空间管 理员	数据分 析师	开 发	运 维	部 署	访 客	安全管 理员	模型设 计师		
节点 PAI	有	有	无	有	无	无	无	无	无		
节点 MR	有	有	无	有	无	无	无	无	无		
节点 CDP	有	有	无	有	无	无	无	无	无		
节点 SQL	有	有	无	有	无	无	无	无	无		
节点 XLIB	有	有	无	有	有	有	有	无	无		
节点 Shell	有	有	无	有	无	无	无	无	无		
节点 虚拟节点	有	有	无	有	有	有	有	无	无		
节点 script_seahawks	有	有	无	有	无	无	无	无	无		
节点 dtboost_analytic	有	有	无	有	无	无	无	无	无		
节点 dtboost_recommand	有	有	无	有	无	无	无	无	无		
节点 pyodps	有	有	无	有	无	无	无	无	无		

资源管理

权限点	项目所 有者	空间管 理员	数据分 析师	开 发	运 维	部 署	访 客	安全管 理员	模型设 计师
查看资源列表	有	有	无	有	有	有	有	无	无
删除资源	有	有	无	有	无	无	无	无	无
创建资源	有	有	无	有	无	无	无	无	无
上传jar文件	有	有	无	有	无	无	无	无	无
上传taxt文件	有	有	无	有	无	无	无	无	无
上传archive文件	有	有	无	有	无	无	无	无	无

工作流开发

DataWorks公共云合集·基础操作

权限点	项目所 有者	空间管 理员	数据分 析师	开 发	运 维	部 署	访 客	安全管 理员	模型设 计师
运行/停止工作流	有	有	无	有	无	无	无	无	无
保存工作流	有	有	无	有	无	无	无	无	无
查看工作流内容	有	有	无	有	有	有	有	无	无
提交节点代码	有	有	无	有	无	无	无	无	无
修改工作流	有	有	无	有	无	无	无	无	无
查看工作流列表	有	有	无	有	有	有	有	无	无
修改owner属性	有	有	无	无	无	无	无	无	无
打开节点代码	有	有	无	有	无	无	无	无	无
删除工作流	有	有	无	有	无	无	无	无	无
创建工作流	有	有	无	有	无	无	无	无	无
创建文件夹	有	有	无	有	无	无	无	无	无
删除文件夹	有	有	无	有	无	无	无	无	无
修改文件夹	有	有	无	有	无	无	无	无	无

数据集成

权限点	项目所 有者	空间管 理员	数据分 析师	开 发	运 维	部 署	访 客	安全管 理员	模型设 计师
数据集成 - 节点编辑	有	有	无	有	无	无	无	无	无
数据集成 - 节点查看	有	有	无	有	无	无	无	无	无
数据集成 - 节点删除	有	有	无	有	无	无	无	无	无
项目同步资源管理菜单	有	有	无	有	有	有	无	无	无
项目同步资源组列表	有	有	无	有	有	有	有	无	无
项目同步资源组创建	有	有	无	有	有	有	无	无	无
项目同步资源组管理机 器列表	有	有	无	有	有	有	无	无	无
项目同步资源组添加机 器	有	有	无	有	有	有	无	无	无
项目同步资源组删除机 器	有	有	无	有	有	有	无	无	无

Dat aWorks

权限点	项目所 有者	空间管 理员	数据分 析师	开 发	运 维	部 署	访 客	安全管 理员	模型设 计师
项目同步资源组修改机 器	有	有	无	有	有	有	无	无	无
项目同步资源组获取资 源组ak	有	有	无	有	有	有	无	无	无
项目同步资源组删除	有	有	无	有	有	有	无	无	无
项目资源消耗监控	有	有	无	无	无	无	无	无	无
运维中心任务修改资源 组	有	有	无	有	有	有	无	无	无
同步任务列表菜单	有	有	无	有	有	有	无	无	无
任务转脚本	有	有	无	有	有	有	无	无	无
获取项目成员列表	有	有	无	有	有	有	无	无	无
新建代码接口	有	有	无	有	有	有	无	无	无
保存/更新代码接口	有	有	无	有	有	有	无	无	无
根据fileld获取代码接 口	有	有	无	有	有	有	有	无	无
获取数据集成节点列表	有	有	无	有	有	有	无	无	无
搜表接口	有	有	无	有	有	有	无	无	无
搜字段接口	有	有	无	有	有	有	无	无	无
查询数据源列表接口	有	有	无	有	有	有	有	无	无
新建数据源接口	有	有	无	无	有	无	无	无	无
查询数据源详情接口	有	有	无	有	有	有	无	无	无
更新数据源接口	有	有	无	无	有	无	无	无	无
删除数据源接口	有	有	无	无	有	无	无	无	无
测试连通性	有	有	无	有	有	有	无	无	无
数据预览	有	有	无	有	有	有	无	无	无
检查是否开通OTS Stream	有	有	无	有	有	有	无	无	无
开通OTS	有	有	无	有	有	有	无	无	无
查询ODPS建表语句	有	有	无	有	有	有	无	无	无

DataWorks公共云合集·基础操作

权限点	项目所 有者	空间管 理员	数据分 析师	开 发	运 维	部 署	访 客	安全管 理员	模型设 计师
新建ODPS表	有	有	无	有	有	有	无	无	无
查询ODPS建表状态	有	有	无	有	有	有	无	无	无
迁移数据库表	有	有	无	无	无	无	无	无	无

数据建模

权限点	项目所 有者	空间管 理员	数据分 析师	开 发	运 维	部 署	访 客	安全管 理员	模型设 计师
查看模型	有	有	有	有	有	有	有	有	有
编辑模型	有	有	无	有	有	无	无	无	有
发布模型	有	有	无	无	有	无	无	无	无

数据分析

权限点	项目所 有者	空间管 理员	数据分 析师	开 发	运 维	部 署	访 客	安全管 理员	模型设 计师
查看数据分析产品页面	有	有	有	有	有	有	有	有	有
使用数据分析产品	有	有	有	有	有	有	无	有	有

⑦ 说明 自定义角色默认无数据分析师角色权限,该类角色如果需要使用数据分析产品,请联系空间 管理员授权数据分析师角色权限,授权详情请参见角色及成员管理:空间级,自定义角色,详情请参见权限详情:空间级角色。

2.3.4. 规划与配置角色权限

本文为您介绍如何为DataWorks工作空间添加并管理成员和角色,以及不同角色的权限功能差异。

角色规划

DataWorks产品内部包含"空间级角色"和"全局(Region)级别角色"两类。

- 在DataWorks的空间级别角色中,您可以根据不同的业务场景为不同的RAM用户设置不同的工作空间权限,DataWorks目前有预设角色,同时也支持您自定义角色。您将RAM用户添加到DataWorks工作空间时,可以选择添加为当前工作空间的预设角色或者自定义角色。详情请参见角色及成员管理:空间级。
- DataWorks的全局级别角色同样包含预设角色、自定义角色两类。管理员可以自由定义拥有某些全局级别功能模块使用权限的全局(Region)级别角色。

? 说明

如果**租户管理员**自定义了某个全局(Region)级别角色,并显示地指定了该角色不具备某些全局(Region)级模块的使用权限(例如:无**数据地图**访问权限),则该自定义角色的权限优先级将高于**租户成员**的权限。

授予空间级别角色

您可以将目标RAM用户或RAM Role添加至指定DataWorks工作空间,该RAM用户或RAM Role添加至工作空间后,即拥有"空间级别角色"。

仅具有空间管理员权限的用户可以授予其他用户"空间级别角色"。授权操作,详情请参见添加工作空间成员和 角色。

⑦ 说明 阿里云主账号以及拥有 AliyunDataWorksFullAccess 权限的RAM用户具有空间管理员权限。

授予全局级别角色

授予某个RAM用户或RAM Role "全局级别角色"则可以通过"全局级别角色管理"来实现。

? 说明

- 仅租户管理员角色和拥有AliyunDataWorksFullAccess的云账号可以授权他人全局级别角色。
- 租户管理员可以将其他人授权为租户管理员。

3.运维中心 3.1.运维中心概述

运维中心包括运维大屏、实时任务运维、周期任务运维、手动任务运维、智能监控、智能诊断、资源运维和引 擎运维功能模块,为您提供任务操作与状态、引擎、资源等多方位的运维能力。

运维中心功能模块

当您在DataStudio中完成节点开发,并提交和发布至生产环境后,即可在运维中心对任务进行运维操作,包 括周期调度任务的自动调度与手动运行、任务运行详情查看、任务运行状态监控、任务运行使用的资源监控与 自动运维;实时任务的执行管控、运行详情查看与监控报警配置;调度任务运维大屏以及数据集成离线同步与 实时同步任务运维专页,针对任务运维关键指标的查看。



下表为运维中心各模块功能使用的简单说明:

模块	描述	环境支持情况
查看运维大屏	运维大屏以报表的形式为您展示任务的运行的重要指标。包括 调度任 务的关键运维指标统计与数据集成离线同步与实时同步任务运维专 页。	开发环境运维中心无 此模块
实时同步任务运行与 管理	在实时任务运维界面,您可以针对实时任务进行启动、停止、下线与 监控报警相关操作。	_
查看并管理周期任务	周期任务运维为您展示任务提交到调度系统后的调度任务列表与调度 环境产生的周期实例。周期任务面板展示当前所有的调度节点,您可 以在周期任务面板修改周期调度任务使用的资源组、责任人等信息。 DataWorks每晚会根据周期任务自动生成第二天待运行的周期实例, 同时支持您手动触发周期任务,分别生成 补数据实例 和测试实例,您 可以通过生成的实例查看任务的运行详情。	开发环境运维中心无 法自动调度生成周期 实例。

模块	描述	环境支持情况
手动任务运维	手动任务运维为您展示任务提交到调度系统后的手动任务列表与经过 手动触发运行后生成的 手动实例。手动任务 面板展示当前所有的手动 触发式任务,您可以在手动任务面板修改任务执行时使用的调度资源 组,责任人等信息。 在手动任务面板中,选择类型为手动业务流程,支持您手动触发任务 运行并且生成手动实例,您可以查看生成的手动实例的运行详情。	-
智能监控	 监控对象:周期任务、实时计算任务、资源水位。 监控支持为自动调度的周期任务设置监控报警,监控在自动调度场景下的周期实例运行情况、支持独享资源组水位监控及实时计算任务运行状态监控。 监控方式:自定义规则、智能基线 支持针对指定对象(节点、业务流程、工作空间、基线、实时计算任务、独享调度资源组、独享数据集成资源组)的监控报警,同时也支持针对整个"业务线"的基线预警,并通过您配置的报警方式来发送相应的报警信息。 报警方式:支持短信、邮件、电话、钉钉群和webhook,已支持设置值班表,将对应时间的报警消息报警给对应的值班人员。 	开发环境运维中心无 此模块
资源运维	资源运维主要对 独享调度资源组 及 独享数据集成资源组 的使用情况 (使用率、运行中任务列表、排队任务列表)及进行监控和自动运 维,支持查看资源组使用趋势。	-
引擎运维	目前支持对EMR引擎的运维。在DataWorks实例任务中,E- MapReduce计算引擎任务会被分为若干作业有序执行,您可以使用 DataWorks的引擎运维功能,查看各个E-MapReduce作业的详细信 息,及时查找并清理运行有误的作业,避免该类作业阻塞下游任务, 影响实例任务的正常运行。	开发环境运维中心无 此模块
智能诊断	DataWorks运维中心为您提供运行诊断功能,帮助您掌握任务运行的 全链路信息,以便迅速定位问题。	开发环境运维中心无 此模块

同时, 需关注以下注意事项:

- 在标准模式工作空间下,您可以通过界面左上角切换进入到开发运维中心和生产运维中心。
- 仅生产环境运维中心支持自动调度。即只有在生产环境运维中心,您才能在周期实例面板中看到自动调度 执行的周期实例。

生产任务运行逻辑

生产任务包括周期任务、手动任务和实时任务。



• 周期任务运维

当节点被提交发布至运维中心后,您可以在运维中心的**周期任务**中看到该节点(周期任务列表中的任务为 所有周期性调度的任务),DataWorks每晚会生成第二天待运行的周期实例,同时支持您手动触发周期任 务,分别生成**补数据实例**和测试实例。

实例类型	应用场景	与周期任务的关系 (对应实例如何生成)	实例触发方式 (对应实例如何触发 执行)	任务运行条件	
		平台每晚根据某一时间 点周期任务快照信息自 动生成第二天待运行的 周期实例。		以周期实例上的任务 为例,任务开始运行 前需满足以下条件: • 依赖的所有父节点	
周期实例	周期性自动化进行ETL 操作。	⑦ 说明 开发 环境运维中心不 会自动调度生成 周期实例。	平台自动触发周期实 例执行。	 实例状态都为成功 状态。 ○ 已到达任务节点设 置的任务定时运行 时间。 ○ 四克次还在日 	
	需要对当前节点及下游		人为针对对周期任务	 。 兩度贡源元定。 • 本任务不是冻结状 态。 	
补数据实 例	间段或未来某段时间的 数据的操作,即需要对 历史或未来时间段的数 据进行ETL操作。	人为针对当前周期任务 手动触发补数据操作, 并生成补数据实例。	进行补数据操作后, 会生成补数据实例, 并同时触发补数据实 例执行。		
	需要对当前周期任务进 行测试,观察执行结果 是否符合预期。		人为针对周期任务的	⑦ 说明 运维 中心中,不同的 实例颜色代表实	
执行测试 并查看测 试实例	⑦ 说明 测试 实例的运行其实 际上也在真实执 行代码逻辑的。	人为针对当前周期任务 手动触发测试操作,并 生成测试实例。	进行测试操作后,会 生成测试实例,并且 同时触发测试实例执 行。	例处于不同的状 态中,实例运行 状态示意详情可 参考:附录:实 例状态。	

您可以在**周期任务运维**下的**周期实例**面板、**补数据实例**面板、**测试实例**面板中查看各类实例的运行情况,根据实例的状态了解数据产出是否异常,不同实例状态的详情请参见<mark>附录:实例状态</mark>。

- ? 说明
 - 周期实例根据周期任务每晚(23:30~24:00)定时自动生成,测试实例、补数据实例根据补数据和测试操作时间点的周期任务手动触发,因此补数据和测试会以最新的任务配置情况生成对应的实例。
 - 新建节点发布生产环境后,当晚会全量生成第二天待运行的周期实例,但新建节点发布时,您可以通过节点调度配置,选择实例生成方式来控制是否立即生成当天的周期实例,详情请参考时间属性配置说明。
- 手动任务运维

当手动业务流程被提交发布至运维中心后,您可以在运维中心 > 手动任务运维 > 手动任务页面中看到该 节点,当您需要运行该手动业务流程时,可以在手动任务面板对该任务执行运行操作,手动触发任务运行 后,会产生手动实例,您可以在手动实例页面查看该实例执行详情。

• 实时任务运维

当实时任务被提交发布至运维中心后,您可以在运维中心 > 实时任务运维页面对实时任务进行启动、下 线及报警配置等相关操作,实时任务运维分为实时计算任务和实时同步任务,您可以分别在实时任务运维 下的实时计算任务、实时同步任务页面对其进行管理。详情请参见:实时计算任务运行与管理、实时同步任务运 行与管理。

任务运行状态监控:智能监控

您可以使用**智能监控**功能监控周期调度任务状态、周期任务自动调度产生的**周期实例**运行状态、实时计算任 务运行状态,配置相应监控规则(自定义规则、智能基线)后,平台一旦检测到规则被触发,将会根据您设置 的自定义规则及智能基线的报警方式,通过邮件、短信、电话或钉钉群消息等形式通知您,方便您及时发现并 处理异常问题。

周期任务状态监控

DataWorks每晚根据周期任务生成第二天待自动调度运行的周期实例,所以为保障周期任务可以正常产生 周期实例并且自动调度运行,DataWorks内置了报警规则来对周期任务进行定期监控扫描,如有异常便会 自动报警。

系统内置全局报警规则(非空间级别报警规则)对周期任务的状态进行监控。报警包括**孤立节点**与**节点成** 环。

? 说明

- DataWorks每天定时9点、12点、16点对周期任务状态进行扫描,如有异常将会自动发送报警。
 但扫描时间点前10分钟内产生异常不会纳入本次扫描,该异常将会被纳入下一个周期的任务状态扫描中进行扫描。
- 全局规则为系统内置规则,您无须手动新建,默认以短信、邮件的方式报警给节点责任人。但 您可以在规则管理页面中针对全局规则修改报警接收人。

节点孤立报警

孤立节点指节点在周期任务或周期实例中展开父节点时,没有依赖任何父节点,这类型节点在自动调度 时不会被调度调起,所以孤立节点不会自动调度运行,如果该孤立节点下游依赖较多,则会造成严重的 后果。孤立节点产生后会自动报警,如果收到孤立节点报警请及时处理。

⑦ 说明 在DataWorks上,除了工作空间根节点外,自定义的每个周期调度的节点都需要有父节点 依赖才可以被正常调度运行。

节点成环报警

节点成环指的是某节点为上游节点但又同时依赖了自己的下游节点。导致依赖关系成环,这类型节点在自动调度时不会被调度调起,节点成环后会自动报警,如果收到节点成环报警请及时处理。

周期实例运行状态监控

在DataWorks上,周期任务周期性调度时将产生周期实例,DataWorks支持通过周期任务自定义规则监控 配置,实现对周期实例运行状态的监控,包括指定对象的自定义规则监控报警和重要任务提前预警的基线 预警功能。 自定义规则报警

通过新建自定义规则实现对指定对象的报警。

您可以对指定对象(任务节点、基线、工作空间、业务流程)上节点的状态(完成、未完成、出错、周期未完成、超时、自动重跑后仍出错等及实时计算任务的出错)设置监控报警。支持短信、邮件、电话、钉钉机器人和webhook的方式发送报警给责任人(节点责任人、指定责任人或值班表中的排班等),详情请参考自定义规则。

基线预警

通过设置基线优先级来保障基线上的任务按时执行,数据顺利产出。

当有任务需要重保并且上游任务依赖较为复杂时,可以使用基线功能将重要任务移到该基线上。当影响 基线在预期时间内数据产出的情况发生时,基线将为您快速定位影响当前节点数据产出的关键路径,找 到阻塞任务(即影响保障数据产出的关键实例)并第一时间发出报警,保障重要数据在预期时间内顺利 完成。当影响基线的任务(基线上的任务及其上游任务)出错或者变慢,您将会收到相应的报警。其报 警分为基线预警与事件报警。支持短信、邮件、电话、钉钉机器人和webhook的方式发送报警给责任人 (节点责任人、指定责任人或值班表中的排班等),关于基线使用详情请参见:基线管理。

■ 基线预警

当基线根据基线的承诺时间、预警余量等预测无法保障基线上的任务在预警时间内完成时,便会触发 基线提前预警功能。

■ 事件报警

当影响基线上数据产出的任务(基线上的任务及其上游任务)出错或者变慢时,将会收到相应的事件 报警。

○ 实时计算任务运行状态监控

您可以通过自定义规则,对实时计算任务的运行状态进行监控,当任务节点出错时通过配置的短信、邮件、电话、钉钉机器人和webhook等方式发送报警给责任人(节点责任人、指定责任人和值班表中的排班等)。详情请参考自定义规则

任务运行资源监控与运维:自动运维

您可以使用**智能监控 > 规则管理**中的自定义规则,来对独享资源组使用率及独享资源组上等待的实例数设置 监控报警,并且支持结合自动运维功能通过短信、邮件、电话、钉钉机器人等方式发送报警给责任人(节点 责任人、指定责任人)。关于自定义规则使用详情请参考自定义规则。

自动运维功能支持通过自定义规则中的独享资源组监控规则,并在限定条件下对独享资源组进行运维操作来 保障重要任务的顺利运行,比如终止运行中的实例。详情请参考自动运维。

任务运行问题定位:运行诊断

● 调度任务运行的必要条件

节点发布至运维中心后,任务开始运行前需满足以下条件。

⑦ 说明 当任务未运行时您可以使用智能诊断功能快速进行对问题进行定位与排查,详情可参考文档: 智能诊断。

i. 依赖的所有父节点实例状态都为成功状态

配置节点依赖关系可以保障节点在运行时能取到正确的数据(读取上游表数据时,上游表的最新数据 已产生),避免当前节点读取数据时,上游表数据还未产出,导致当前节点读取数据出现问题,所 以,当节点存在上游节点依赖时,当前节点的执行,需要等待所依赖的父节点都执行成功。详情请参 见:同周期调度依赖逻辑说明。 ? 说明

- 父节点非成功状态包括父节点未运行、父节点状态失败、父节点被冻结(暂停)、父节点 处于运行中。
- 冻结状态的节点将阻塞下游节点执行。
- ii. 已到达任务节点设置的定时运行时间

数据开发中定义节点的调度属性时,需要为节点设置调度环境下节点运行的"理想时间"。

- 当当前节点依赖的所有父节点都在当前节点设置的定时运行时间点前执行成功,那么当父节点全都 执行成功后,当前节点将进入等待状态,待到达任务节点的定时时间后启动运行。
- 当当前节点依赖的所有父节点在当前节点设置的定时运行时间点后执行成功,那么当父节点全都执行成功且有充足的调度资源时,当前节点将立即运行起来。
- iii. 调度资源充足

数据开发中定义节点的调度属性时,您可以设置调度环境下节点运行所使用的独享调度资源组。同时,您也可以在运维中心的周期任务页面进行修改。

当当前节点所有的父节点运行成功,并且已经到达节点设置的定时运行时间时,如果节点指定的调度 资源组无可用资源时,节点进入等待资源的状态,直到占用该调度资源的任务执行成功,将资源释放 后启动运行。

iv. 本任务不是冻结状态

如果当前节点处于冻结状态,那么节点满足上述三个条件后仍不会运行,并导致依赖该周期实例的下 游实例无法运行。如果您需要解冻该节点,详情请参见:任务冻结与解冻。

• 任务运行失败原因诊断

您可以使用智能诊断功能和DAG的上游分析功能对未运行任务进行快速诊断。

其中,智能诊断功能支持对未开始运行的任务进行任务运行的必要条件检查,支持对运行失败的任务进行 原因分析并给出诊断建议等。

智能诊断能力概览:

- 任务运行的全链路诊断。包括父节点运行情况检查、当前节点设置的定时时间检查、调度资源详情(任务使用的调度资源组使用趋势与任务排队数,当前任务等待资源时哪些任务占用资源)、任务执行详情 (任务出错时提供原因分析与诊断建议)。
- 当前实例的基本信息与实例运行的关键时间点(上游节点运行完成时间、当前节点设置的定时时间、开始等待调度资源时间、开始运行时间、运行成功时间)。
- 当前节点影响的基线及其当天基线实例运行情况。
- 当前节点平均运行时间、开始执行时间、等待资源消耗时间趋势图及当前节点运行历史的实例列表。

关于智能诊断功能使用详情请参见:智能诊断。

任务产出的表数据监控:数据质量

您可以使用数据质量功能对周期实例、补数据实例、测试实例产出的表数据进行监控。

周期任务在生产环境通过自动调度或手动触发来运行,分别产生周期实例,补数据实例和测试实例,您可以 对**周期任务**设置数据质量规则来监控周期任务生成的实例(**周期实例、补数据实例、测试实例**)运行后所 产出的表数据是否符合预期。 数据质量通过表配置的分区表达式来匹配节点产出的表分区,数据质量规则关联产出该表数据的调度节点,当 节点运行后(运维中心中通过测试实例,补数据实例,周期实例的运行及重跑操作产出表数据),会触发该数 据质量规则校验,您可以设置规则的强弱来控制节点是否失败退出,从而避免脏数据影响扩大,同时,您也可 以通过订阅规则来第一时间接收节点产出表数据异常的报警。详情请参考文档数据质量。

附录: 实例状态

运维中心按照不同的颜色与标识来辨识任务正处于运行流程中的哪个阶段,不同的实例颜色与标识代表实例处于不同的状态下。不同的实例颜色与标识所对应的任务状态如下所示。关于任务运行必要条件详情请参见:任 务运行问题定位:运行诊断。

序号	状态类型	状态标识	运行流程图
1	运行成功状态	\oslash	
2	未运行状态	Θ	रिग्रेन
3	运行失败状态	\otimes	执行成功 依赖的父节点都 已到当前设置的节点 节点设置的调度资源 已执行成功 定时调度时间
4	正在运行状态	•	
5	等待状态	0	失敗
6	暂停/冻结状态	۲	

3.2. 查看运维大屏

运维大屏展示当前您需要重点关注的运维指标、任务整体运行情况、调度资源变化趋势等运维概览,以及离线 同步、实时同步任务的运行状态分布、数据同步进度等数据集成信息,帮助您提升任务运维效率。

进入运维大屏

- 1. 登录DataWorks控制台。
- 2. 在左侧导航栏,单击工作空间列表。
- 3. 选择工作空间所在地域后,单击相应工作空间后的进入数据开发。
- 4. 单击左上方的
 图标,选择全部产品 > 任务运维 > 运维中心(工作流),默认进入运维大屏 > 运维

 概览页面。

查看运维概览

目前运维概览页面的统计范围是周期任务或周期实例,不包括其它类型的任务和实例。您可以在运维大屏页 面查看以下内容:

• 重点关注区域展示以下内容:

- 您当前需要重点关注的周期实例,包括失败实例、运行慢实例、等资源实例。同时满足以下条件的实例
 即为运行慢实例:
 - 实例处于运行中的状态。
 - 实例运行超过30分钟。
 - 实例的运行时长与过去10天的平均运行时长相比,已经超过15分钟及以上。
- 您当前需要重点关注的**孤立节点、暂停节点和过期节点**:
 - 孤立节点此处统计的为周期任务中没有父节点的节点个数。当节点没有上游依赖时,该节点会变为孤立节点,将无法再自动调度运行。例如,修改父节点输出名称后,造成依赖丢失。关于孤立节点产生原因、影响及解决方案详情,您可以参考文档:什么是孤立节点,如何处理孤立节点?。
 - 暂停节点此处统计的为暂停(冻结)状态的周期任务。周期任务暂停后,该周期任务产生的实例也将 是冻结状态,冻结状态的实例将不会运行,并且会阻塞其下游节点执行。关于冻结操作详情介绍与影 响说明可参考文档:任务冻结与解冻。
 - 过期节点此处统计的为节点调度生效日期已过的节点。节点在调度生效日期内将自动产生周期实例并
 运行,调度生效日期范围外将无法产生周期实例并且自动调度。



重点关注根据您的需求可以选择统计昨天、前天、近7天或指定时间段内(最多支持7天)的数据。您可以 单击某个指标进入相应的节点或实例列表,及时关注相关指标并进行治理,以避免影响业务。

 运行状态分布区域展示当前所有业务日期为昨天、前天、近7天或指定时间段内(最多支持7天)正常调度 实例的运行状态分布,统计的是页面请求发生时的数据。您可以单击饼图中的某个色块,进入相应状态的 实例列表。



• 任务完成情况区域展示页面请求当天00:00~23:00内,当前工作空间中周期实例的完成情况。您可以查看 昨日、今日和历史平均的运行成功或未运行的任务,并可以指定需要查看的任务类型。



对比正常周期性调度昨日、今日和历史平均的任务完成情况,统计出折线图。如果三条折线偏移过多,表 示在某个时间段内有异常情况,您需要进一步检查与分析。

⑦ 说明 由于运维中心统计方法的变更(仅统计生产环境的实例),今日任务完成情况折线会产生明显波动。

调度资源分配区域默认展示指定时间段内(最多支持7天)调度资源组使用率与各时间点运行在该资源组上的实例数量的变化趋势。您可以单击右上方的资源组选择下拉列表切换资源组分别查看当前工作空间下的各调度资源组,其资源组使用率与运行的实例数变化趋势。



实例数量为对应时间段运行在该资源组上实例数量,资源组使用率为对应时间段运行在该资源组上的实例 所占用的资源的百分比。

? 说明

- 。 资源组使用率不记录数据集成任务占用的资源。
- 资源组使用率与资源组上运行的实例数量为资源组级别,例如:如果您使用的独享调度资源组为多个工作空间共享,那么此处统计的为多个工作空间下,该资源组总的资源使用率和实例数量变化趋势。
- 运行时长排行区域展示运行时长、等待资源时长和运行变慢时长的排行。该区域每日更新数据,对运行在前一天,且运行成功的实例的运行时长进行排行。

┃运行时	长排行 ~				上次更新: 2021-10-08
✓ 运行中 等待到	3年4月7 资源时长排行)	节点名称	负责人	执行时长	A
变慢热	¥亍 7000048746	测试小时任务是否并行	20	2分	
2	7000048746	测试小时任务是否并行	100	1分25秒	
3	7000048746	测试小时任务是否并行	1	1分8秒	
4	7000048834	测试告答	5	1分4秒	
5	7000048377	oss_odps_di	20	59秒	
					*

 近一个月出错排行区域每日更新数据,统计最近一个月的任务出错排行,为您展示任务出错次数最多的前 10个节点。您可以查看节点ID、节点名称和出错次数。

┃近──个	月出错排行			上次更新: 2020-06-18
	节点ID	节点名称	负责人	出错次数
1		mi	10.0	256
2		end and street spins		87
3				46
4		410.010000		43
5		endpoint College	1.48	18
6		energie in the second		17
7		And the second second		16
8		8.0	1000	16
9		en.i.m		16
10		Contraction of the local distance	1810	16

调度任务数量趋势区域每日更新数据,展示在一段时间内,生产环境的周期任务、周期实例的数量变化趋势。您最多可以查看近一年的调度任务数量变化。



任务类型分布区域展示页面请求发生时,各种类型的任务数量。饼图最多支持展示8种任务类型,如果统计的任务超过8种类型,会进行合并展示。



查看离线同步运维

在运维大屏页面,单击数据集成,默认展示某个时间范围内离线同步的信息。您可以在右上方设置统计数据的时间范围。

启动		×
是否重置位点:	重置位点	
启动时间点位:	请选择日期与时间	曲
时区:	(GMT+08:00) China Time - Shanghai	~
Failover:	10 分钟内, Failover次数超过 3 任务自动结束	
* 脏数据策略:	💿 零容忍,不允许 🔵 不限制 🔵 有限控制 🕜	
		确定 取消

您可以查看离线同步任务的以下内容:

运行状态分布区域展示所选业务日期内调度实例的运行状态分布,统计的是页面请求发生时的数据。您可以单击饼图中的某个色块,进入相应状态的实例列表。



● 数据同步进度区域展示所选业务日期内离线同步的数据量,包括总数据量、公网总流量和记录总量。



• 同步数据量统计区域根据数据源类型展示所选业务日期内,同步数据量的拉取和写入曲线。



- 最新榜单Top10区域分别展示10个最新失败实例和最新成功实例,以便您全局查看最新的同步任务状态。
- 同步任务执行详情区域支持通过筛选提交时间、任务状态和任务名称等条件,快速搜素任务实例,并可以单击实例ID,查看该实例的运行详情。



查看实时同步运维

在运维大屏 > 数据集成页面,单击实时同步,查看实时同步任务的以下内容:

 运行状态分布区域展示当前实时同步实例的运行状态分布,统计的是页面请求发生时的数据。您可以单击 饼图中的某个色块,进入相应状态的实例列表。



• 概览区域展示当前工作空间下,所有实时任务的数据速度和记录速度的总和。

概览	
*	
数据速度(BPS) O B	记录速度 (RPS) 0

• 任务延迟Top10区域展示延迟最高的10个任务,以便您迅速定位高延迟任务。

【任务延迟 Top10			
	任务名称	业务延迟	数据集成独享资源组
1	OB_2_DATAHUB	23 毫秒	1000.0100000.000
2	mysql_to_datahub	24 毫秒	control di condition, casa

• 任务报警信息区域展示最近的报警信息,以便您快速查看异常情况。

1 任务报警信息								
发生时间	任务ID	任务名称	接收人	报警级别	指标	设定阈值	报警时阈值	通知方式
2020-09-15 14:57:47	287	100101.1.500	B0000001.0	WARNING	心跳告警			邮件
2020-09-15 14:57:16	408	conclusi, Joint, Children and Staffs	- MEMORY 1.1	WARNING	lkkk	0		短信
2020-09-15 14:52:47	287	address 1 http://www.com/address.com/addres	80005.0	CRITICAL	心跳告警			紅紅
2020-09-15 14:52:47	287	100010-1.000		CRITICAL	心跳告警			邮件
2020-09-15 14:52:47	287	100000.2.000	distantio of	WARNING	心跳告警			紅紅
2020-09-15 14:52:47	287	adjuster, J. No.	8000%.d	WARNING	心跳告警			邮件

• Failover信息区域展示指定时间内的Failover消息,以便总览任务Failover的状态。

Failover信息			
在最近 30 分钟	内共Failover 0 次		
时间	实例ID	任务名称	Failover事件

3.3. 实时任务运维

3.3.1. 实时计算任务运行与管理

DataWorks的实时计算任务页面为您展示了任务实例的基本信息及运行详情,方便您查看目标任务的运行情况。同时,您还可以为目标任务设置监控报警,在任务运行出错时及时发现并处理异常。

使用限制

DataWorks当前仅支持对EMR Spark Streaming及EMR Streaming SQL类型的节点提交的任务进行运维。

进入实时计算任务页面

- 1. 登录DataWorks控制台。
- 2. 在左侧导航栏,单击工作空间列表。
- 3. 选择工作空间所在地域后,单击相应工作空间后的进入数据开发。
- 4. 单击左上方的**三**图标,选择**全部产品 > 任务运维 > 运维中心(工作流)**。
- 5. 在左侧导航栏, 单击实时任务运维 > 实时计算任务, 进入实时计算任务页面。

查看实时计算任务列表

运维中心的实时计算任务页面为您展示了生产环境的实时计算任务列表,您可以在该页面查看任务的基本信息,执行启动、终止、下线等运维操作,并设置任务出错监控。

≡	ntaWorks 🏟	运帅	中心	· ·									& DataStudio & 任务》	柿 🗘	n 4	•	
e	运维大屏		搜索:	输入任务ID/名称 Q 状态:	请选择	∨ 任务供	型: 油泡焊	✓ 引撃炭型: 消泡料		引撃突例: 街道部	i v	责任人: 谢远探测	HEA V BAN	0			
\$\$\$	实时任务运维	^		任务名称	任祭旧	状态	开始运行」	结束运行」)	运行时长 小	任务委员	리왕육편	可能な例	发布时间	摄作			1 ^
	实时计算任务									EMB Soark							+
	实时同步任务			spark_streaming_002		已失败	-		0s	Streaming	EMR	test_emr_033003	2021-05-21 16:47:25	扁动	终止 下!	ま 设置出社	
n	周期任务运维	*		spark_streaming_001	Contraction of the	已失败		-	0s	EMR Spark Streaming	EMR	test_emr_033003	2021-05-21 15:04:26	启动	终止 下編	8 1 12日前日	1
ଜ	手动任务运维	*	•									1					F
*	智能监控	*															
	资源运维																
۲	引擎运维	*															
Ŷ,	智能诊断																
																	w
			4														F.
≡			启动	授 <u>此</u> 下线 设置监控	•								每贝显示: 20 1	×	《上一页	下一页	

区域	描述
	在该区域,您可以通过 任务ID、任务名称、状态、任务类型、引擎类型、引擎实例 及 责任人 等筛选条件 过滤出需要查询的实时计算任务。
1	 ⑦ 说明 任务ID或任务名称搜索的结果会受到其它筛选条件的影响。只有同时满足所有筛选条件时,才会显示相应的查询结果。 任务名称搜索支持模糊匹配,输入关键词后,即可显示包含关键词的所有实时计算任务。
	在该区域,您可以查看目标任务的基本信息及运行情况。 • 任务名称:为您展示任务的名称。单击任务名称,即可进入目标任务的详情页面。 • 任务ID:为您展示任务的ID。 • 运行状态:任务的运行状态,包括未运行、生成中、等待资源、启动中、运行中、已停止、已失败、重启中以及下线中等状态。 • 开始运行:任务的开始运行时间。 • 结束运行:任务的结束运行时间。 • 运行时长:任务的运行时长,单位为秒。 • 任务类型:任务的类型。
2	② 说明 DataWorks当前仅支持对EMR Spark Streaming及EMR Streaming SQL类型的节点提交的任务进行 运维。
	 引擎类型:任务所使用的引擎类型。DataWorks当前仅支持对E-MapReduce引擎的实时计算任务进行运维。 引擎实例:任务所绑定的E-MapReduce引擎名称。 发布时间:任务的发布时间,格式为yyyy-MM-dd HH:mm:ss。 责任人:任务所在工作空间的责任人。
	⑦ 说明 您可以根据任务的开始运行、结束运行以及运行时长,按照时间的升序或降序对目标任务进行排序。
3	在该区域,您可以对目标任务进行如下操作。 • 启动:启动未运行的实时计算任务。 • 终止:终止运行中的实时计算任务。 • 下线:下线已提交发布的实时计算任务。 • 设置监控:用于监控目标实时计算任务的运行情况。当任务运行出错时,会根据您设置的短信、邮 件、电话、钉钉群机器人或WebHook等方式发送报警至相应的报警接收人。 配置报警接收人所使用的邮件、电话信息,详情请参见设置报警接收人的手机或邮箱信息,配置钉钉群 机器人及获取Webhook地址,详情请参见场景实践:发送报警消息至钉钉群。
4	您可以勾选多个目标任务后,在该区域批量 启动、终止、下线 目标任务,或为目标任务 设置监控 报警。

查看实时计算任务详情

在任务列表单击目标任务实例名称,即可进入任务的详情页面。您可以在任务的详情页面查看任务的**详细信息、运行日志、操作日志**以及任务的运行**代码**,方便您快速定位对目标任务执行的相关操作记录及操作内容,及时获取任务运行出错的报错详情并跟进处理。

详细信息 运行日志 操作日志 代码	
基本信息	
名称:	任务ID:
任务类型: EMR_SPARK_STREAMING	责任人:
所属工作空间:	最新版本: 1
发布时间: 2021-05-21 16:47:25	调度资源组:
Region: 上海	引擎类型: EMR
引擎实例:	
运行信息	
实例ID:	状态:已失败
执行版本: 1	开始运行: -
开始等待资源: -	等待资源耗时: -
结束时间: -	运行时长: Os

3.3.2. 实时同步任务运行与管理

实时同步任务页面为您展示实时同步任务提交至调度系统后,经过调度系统运行后产生的实时同步任务实例。

进入实时任务运维页面

- 1. 登录DataWorks控制台。
- 2. 在左侧导航栏,单击工作空间列表。
- 3. 选择工作空间所在地域后,单击相应工作空间后的进入数据开发。
- 4. 单击左上方的**三**图标,选择**全部产品 > 任务运维 > 运维中心(工作流)**。
- 5. 在左侧列表,单击**实时任务运维**后的**一**图标。
- 6. 在下拉列表下单击实时同步任务,即可查看所有实时同步任务实例信息。

查看实时任务列表

实时任务实例页面以列表形式对被调度的实时同步任务进行运维和管理,包括查看实时同步任务基本信息、检 查运行信息和基本属性配置以及启动和下线实例等功能。

n DataWorks	运维中心 <u>生产11</u> 。 ***	翠 华东2标准111 📕 🗸 :_DPE_E2											& DataStudio Ω	ହ ୧୪ 🖓 ୪	dataworks_d
云维大屏															
到时任务运机	任务节点名 东源数据源出	尔: B· 全部		~	任务状态: 全i 来源数据源名称:	B ²			~	责任人: 最近操作人:	诸选择			~	
2时同步任务	去向数据源则	L: 全部		~	去向數据源名称:				20	編集成资源组:	全部			· · ·	
周期任务运维 主动任务运维	×	17 10 10 10	d frain		July 2410		B) C 4B (kosk)/B	BY IS ALL	來源数据	未源数据	去向数据源类	去向数据	-	17.00	
198622	~	11904	177 E	四天	10.99%EX2		alcontract in a clini	RECEIRTAN	源典型	27.	핖	源	***A 2	28(1)	3
習語沙斯	5016608	0 sync_date	₩67中 ()		6.9 10	2021-06-01 17:02:59 China Time- Shanghai	2021-08-13 18:0017	detexorks.d.	mpaghani ng	no, repelj, drend	detahub	na jininina Inganis	dataworks_d	(启动 停止 下創	1 授幣设置
	5008500	u	未运行				2021-08-12 18:32:40	detexp/kb.d.	mysegibini 14	shburiyee User	000	odps_first	dataworks_d	启动 停止 下线	. 1989 PM
	5016/ BH	n	异常 😣		9.6 秒	2021-04-06 11-04-07	2021-04-28 16:50:58	deserts_d.	mpol	an, myanji		onlyn_first	dataworks_d	启动 停止 下线	1 12502
	501\$81 K	14 kelka	未运行				2821-04411 14:31:02	denseria _n d.	mpod	scuryed. deno2	latio	Kalla	dataworks_d	启动 停止 下线	1 19903
	5015840	8 90.30	异常 😣				2821-08-28 14/35/28	dataworks.d.	datahubat Naari	scubitaña buzob	050	odpa_feet	dataworks_d	启动 停止 下线	6 按数设置
	5016406	H NUMBER NUMBER	停止 Θ		3.6 秒	2022-03-09 10-21-04 Drive Terrer - Dranghai	2821-05-89 30-40-58	deinsarta _y d.	mand	scudsuol uter	bala	an Jania ya	dataworks_d	(启动停止下部	1 投资设置
	5011602	0.0000000000000000000000000000000000000	日本 😣		7.4 秒	2023-04-29 1518-12	2021-00-08 11:35-08	detexorks.d.	republici rep	no, report, techny	604	odps.feat	dataworks_d	启动 停止 下线	1 报告设置
	5014088	and programs and	停止 Θ		1.7 秒	2023-40-25 10-48-42 GMT-4	2821402-25 10:45:40	deterrorks.d.	report	xuryiqi	hdio	sc,hologr et	dataworks_d	启动 停止 下线	1 报警设置
	501487	17 Not	停止 Θ				2021-02-08 17 12-01	deterrorks.d.	princh	na policel M	8029	odps.first	dataworks_d	启动 停止 下线	t ferige
	5014040	N desited	未运行				2021-02-64 22:29:09	determin_d.	mpal	scuryed. denic2	detechnik	scuderante but	dataworks_d	启动 停止 下线	1 125025
	-		^		0.0.54	2021-03-24	2021-03-09			xc_mysqL_					
	启动 停止	下线 新增貨幣 操作货行	3									每页显示	t: 20 ¥ 共3	0条记录 <	1 2
I	在 (K) 您 可 以 向 数 据 ?	在该区域,您可以通过筛选条件过滤出需要查询的实例。 您可以根据 任务节点名称、任务状态、责任人、来源数据源类型、来源数据源名称、最近操作人、去向数据源类型、去向数据源名称、数据集成资源组 等条件进行精确筛选。													
	展示	0													
	在该区	域,您可以查 <u></u>	看任务家	实例信息 5	息, 包括:										
	• ID:	刈 窓 展 示 仕 务	头彻的	D°											
	● 任务	名称 :为您展	展示任务	实例名	称。单击	任务名称	R, 即可	J进入信	E务实	《例详	情页面	0			
	• 运行状态:包括运行中、未运行、等待资源、异常、停止五种状态。														
	● 描述	: 对实例的简	「单描述	0											
	 业务延迟:当前时间距离同步位点的时间间隔。 														
	● 当前	「同步位点: 🗎	当前消费	到增量	数据的时	间。									
	• 最近	操作时间:	最后操作	实例的	」时间点。										
	● 														
	- 取人	1床1F八 ・取加	5 灯头例	操作的	I用户。										

- 来源数据源:数据同步的来源数据源。
- 去向数据源类型:去向数据源的类型。
- 去向数据源:数据同步的去向数据源。
- 责任人:实时任务所在的工作空间责任人。
- 当前起始位点:目前开始消费增量数据的时间点。
- 数据读取速度(BPS):数据同步读取数据的速度。
- 记录速度: 写入日志的数据同步读取数据速度。
- 数据集成资源组:任务实例运行使用的数据集成资源组。

区域	描述	

	在该区域,您可以对所选择的任务实例进行如下操作。
	● 启动 :启动未运行的同步任务。
	● 停止 :停止运行中的同步任务。
3	 下线:下线未运行、异常及停止的同步任务。
	● 新增告警:勾选对应同步任务后,单击新增告警,即可进行同步任务设置告警规则。
	 操作告警:勾选对应同步任务后,单击操作告警,即可对同步任务告警进行删除、启用和停止操作,同时支持通过选择告警指标类型来批量修改所有对应的告警规则。

实时任务详情页面

Dat aWorks

在任务列表单击实例名称,即可进入任务详情页面。在页面上方可以切换至不同子页面,任务详情页面可以查 看任务的运行信息、日志、基本属性、任务配置、Failover记录和DDL记录。

运行信息 日志 著	基本属性 任务配置	Failover记录 DD	L记录					任务状态:运行中 (``)	
数据读取								С 刷新	
步骤类型	线程号	总行数	总字节数	总等待时间	速度 (条记录/ 秒)	速度 (MB/秒)	业务延时	脏数据条数	
MySQL1	0	1	8B	17.24 day	0	OB	0.26 s	0	
数据写入									
步骤类型	线程号	总行数	总字节数	总等待时间	速度 (条记录/ 秒)	速度 (MB/秒)	业务延时	脏数据条数	
DataHub1	0	1	8B	3 ms	3 ms 0		0.26 s	0	
数据库事件状态 总宽:插入:1次更新 事件关联表 来源表名称:	: 0 次, 删除: 0 次, DDL: 0	次							
来源表名称		插入		更新		8	除		
xc_demo.wpw_date		1		0					

同步任务相关操作

• 启动同步任务

在操作区域单击启动,即可进入任务的启动页面。设置好相关参数,单击确定,即可启动运行该任务。

启动		×
是否重置位点:	重置位点	
启动时间点位:	请选择日期与时间	**
时区:	(GMT+08:00) China Time - Shanghai	~
Failover:	10 分钟内, Failover次数超过 3 任务自动结束	
* 脏数据策略:	● 零容忍,不允许 🔵 不限制 🔵 有限控制 🕐	
		确定取消

当勾选重置位点选项时,启动时间点位和时区为必填参数。

● 停止同步任务

单击停止, 在弹出的对话框中选择单击停止, 即可停止运行中的任务。

• 下线同步任务

单击下线,在弹出的对话框中选择单击下线,即可下线当前任务。

- 报警设置
 - 在报警事件页面,设置发生时间范围、报警级别和规则查看报警事件。
 - 在报警规则页面,显示当前同步任务实例已创建的报警规则。您可以在报警规则列表查看已创建规则的指标和状态等信息,也可以在操作区域对规则进行编辑、暂停、删除和模拟测试等操作。如果需要新创

建规则,请单击右侧的**新建规则**。

新建规则		×
* 名称: 描述: * 指标:	 任务状态	
* 词值:	WARNING 在 5 分钟内,无心跳 CRITICAL 在 5 分钟内,无心跳	
*报警间隔:	5 分钟内只发一次报警	
WARNING:	邮件 🗌 短信 📄 电话 🔄 钉钉	
CRITICAL:	邮件 🗌 短信 📄 电话 🔄 钉钉	
* 接收人 (非钉钉):	请选择	
		确定取消

参数	描述	是否必选
名称	填写新建的规则名称。	是
描述	对新建的规则进行简单描述,便于快速排查定位。	否
指标	设置描述规则的维度指标类型,指标类型和阈值需配合 设置。指标类型包括:任务状态、业务延迟、 Failover、脏数据和DDL不支持。	是
阈值	 触发报警的阈值需要和指标类型配合设置。 当指标类型为任务状态时,设置触发报警规则的最短时间间隔。 当指标类型为业务延迟时,设置业务延迟大于等于特定时间后的报警持续时间。 当指标类型为脏数据时,设置特定时间段内脏数据超过多少条才能触发报警。 当指标类型为DDL不支持时,仅需选择报警级别。 对于非分布式运行的任务,设置特定的时间段内Failover超过多少次触发报警;对于分布式运行的任务,设置指定的时间段内单个子任务Failover超过多少次触发报警。 	是
报警间隔	设置两次报警之间最小时间间隔,默认为5分钟,最短 时间间隔可设置1分钟。	是

参数	描述	是否必选
	设置WARNING级别报警通知方式,包括邮件、短信、 电话、钉钉。	
WARNING	 ② 说明 ● 以邮件方式接收告警:若子账号接收告警 需要由主账号在用户信息中为子账号完善 个人邮箱信息 ● 以短信方式接收告警:若子账号接收告警 需要由主账号在用户信息中为子账号完善 个人手机号信息。 ■ 以电话方式接收告警:若子账号接收告警 需要由主账号在用户信息中为子账号完善 个人手机号信息。 ■ 以自定义钉钉机器人接收告警:要求是普 通群,且自定义机器人需要添加关键字" DataWorks"。 	否
CRITICAL	设置CRITICAL级别报警通知方式,包括 邮件、短信、 电话、钉钉。	否
接收人(非钉钉)	设置发送报警的联系人。	是

3.4. 周期任务运维

3.4.1. 周期任务运维概述

任务提交和发布至生产环境后,即可在周期任务列表中对任务进行运维操作。包括自动调度和手动运行周期任务、查看任务运行详情、暂停任务、下线任务等。本文为您介绍周期任务的运维操作详情。

文档概述

周期任务是指在调度系统按照调度配置周期性自动调度的任务。您可以在运维中心 > 周期任务列表查看指定 工作空间下的周期任务。本文为您介绍周期任务的运维操作详情,包括:

- 创建并发布周期任务
- 查看周期任务
- 运行周期任务
- 暂停调度运行
- 下线周期任务
- 还原已下线的任务
- 管理周期任务
- 查看周期任务操作记录

创建并发布周期任务

任务需要先通过发布流程发布至生产调度系统后,您才可以在运维中心的周期任务列表中看到该任务。操作详 情请参见:数据开发概述。

1、新建任务	2、提交任务至开发环境	3、发布任务至生产环境
1、创建业务流程	提交单个任务 Immai	Contraction (Contraction) Contraction (Contraction) Contraction Contraction
	提交整个业务流程下的任务	已提交的节点都将在任务发布界面产生一条待发布记 录,发布节点后,该节点中代码才会在生产环境生效。 说明:工作空间开启发布检查后,检查通过的节点可 以进行发布操作。若发布失败,可进入发布包列表查 委士地增低
业へ业务流程画板 2、在业务流程面板新建任务		
▲ Terrent town. • Ο ≡ II Ο ○ II 0 ■ II Ο ○ II 0 ■ II Ο ○ II 0 ■ II 0 ○ ○ II 0 ■ II 0 ○ II	E manue e e	4、确认生产调度环境下的任务配置
	Contract Distance Distance	
Status		Name Composition Composition Composition Composition Composition Composition Composition
业务流程面板可通过拖拽开发组件实现流程化 的业务逻辑展现与代码开发。	提交发布的操作。提交后任务将在开发环境生效。 说明:简单模式工作空间只有一个生产环境,提交后,您无需 发布,代码即可进入调度系统周期性执行。	发布成功后,你可以进入周期任务而板确认节点代码,调
		度依赖、调度参数、节点调度时所用的调度资源组等配置 是否符合预期。

? 说明

- 对生产环境的任务、资源、函数进行新增、修改、下线等操作均可以通过提交和发布流程来实现。
- 任务的变更均需执行发布操作,发布成功后该变更才会在生产环境中生效。发布时可能存在发布 失败、发布流程被阻塞或发布版本不符合预期等情况,因此,建议您在任务发布后,进入运维中 心 > 周期任务界面进行排查:
 - 通过DAG图查看任务的上下游依赖关系是否正确。
 - 通过<mark>节点详情</mark>查看生产环境中该任务的参数配置等是否符合预期。

查看周期任务

周期任务列表中为您展示生产环境下调度系统按照调度配置自动定时运行的任务,您可以检查任务代码、调度参数、调度依赖、节点血缘等是否符合预期。同时,支持您通过列表及DAG图的方式,快速查看任务的详细信息并执行相关操作。详情请参见:查看并管理周期任务。

? 说明

- 仅发布至生产环境的任务才可以在运维中心的周期任务列表中看到。若任务配置不符合预期,请参考创建并发布周期任务进行检查。
- 当任务展开父节点时没有依赖任何父节点,或者某任务为上游任务但又同时依赖了自己的下游任务,导致依赖关系成环。这两种类型的任务在自动调度时将不会被调度调起。

运行周期任务

• 周期任务运行方式

DataWorks每晚会根据周期任务自动生成第二天待运行的周期实例,同时支持您手动触发周期任务,分别生成补数据实例和测试实例。



实例类型	应用场景	与周期任务的关系 (对应实例如何生成)	实例触发方式 (对应实例如何触发 执行)
	周期性自动化进行ETL操 作。	平台每晚根据某一时间点周期任 务快照信息自动生成第二天待运 行的周期实例。	平台自动触发周期实 例执行。
周期实例		⑦ 说明 开发环境运维中 心不会自动调度生成周期实 例。	
补数据实例	需要对当前任务及下游任务 执行补过去某一时间段或未 来某段时间的数据的操作, 即需要对历史或未来时间段 的数据进行ETL操作。	人为针对当前周期任务手动触发 补数据操作,并生成补数据实 例。	生成补数据实例的同 时触发补数据实例执 行。
执行测试并查看测试 实例	需要对当前周期任务进行测 试,观察执行结果是否符合 预期。		
	⑦ 说明 测试实例 的运行也在真实执行代 码逻辑。	人为针对当前周期任务手动触友 测试操作,并生成测试实例。	生成测试实例的同时 触发测试实例执行。

? 说明

• 不同实例生成方式对自动调度的周期实例的影响:

新建任务时,您可以通过设置实例生成方式来控制任务发布至生产环境后生成实例的时间,即 任务发布至生产环境后什么时候可以在生产运维中心周期实例面板中看到该任务生成的实例, 详情请参见: <u>实例生成方式</u>。

- T+1次生成:
 - 23:30之前提交发布的任务, 第二天周期实例生效, 并自动调度运行。
 - 23:30之后提交发布的任务, 第三天周期实例生效, 并自动调度运行。
- 发布后即时生成:
 - 23:30之前提交发布的任务,当天生成周期实例并自动调度。
 - 任务定时运行时间在提交发布时间点10分钟后时,DataWorks会正常生成 实例,该实例会真实跑数据。
 - 任务定时运行时间距离提交发布时间不足10分钟时,DataWorks会生成一个已完成的实例,实例状态为实时生成的过期实例,过期实例不会真实 跑数据。
 - 23:30之后提交发布的任务,实时转实例功能不生效。需要等到提交发布后的第 三天实例才生效。
- 已有的实例从T+1次日生成变更为发布后即时生成,对当天实例的影响:已运行的实例会保留,未运行的实例将会被替换为实时转出的实例。
- 任务运行条件和问题定位



调度任务运行需要满足上游任务运行成功、当前任务定时时间已到、调度资源充足、当前任务未冻结等条件,详情请参见:调度任务运行的必要条件。

如果任务为未运行状态,建议您先通过DAG面板中的上游分析功能,快速定位阻塞当前任务运行的关键上游 任务,再通过运行诊断功能,诊断关键实例未运行的原因或存在的问题,当任务依赖复杂时,您可以通过该 功能快速定位问题,提升运维效率。



暂停调度运行

↓ 注意 请勿随意操作project name_root节点,该节点为工作空间根节点。周期任务的实例均依赖于该节点,如果冻结该节点,周期实例将无法运行。

暂停(冻结)周期任务:如果某任务及其下游任务在未来一段时间内不需要自动调度运行,您可以选择暂停(冻结)该任务。

◦ 冻结操作:

 ➡ 运维大屏 ※ 实时任务运维 	搜索: 市点答称/市点□ Q 市点类型: 請选择节点类型 > 責任人: 请选择责任人 > 導度資源组: 请选择 現的节点 今日傍衣的节点 <
□ 周期任务运维 ^ 周期任务	名称 田< 〇 日 市点名称/市点D Q C 回 Q E 11
內如头初 补数据实例 测试实例	● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●
 	offline_odps_once_full 展开文节点 PILEの PILE
 资源运维 ④ 引擎运维 ~ 	
と 智能诊断	<u> 市成市</u> <u> 市成市</u> <u> 市成市</u> 重査実例 査査主 重査 重査 重査 重査
(余 操作历史	
=	

- 冻结后现象:冻结周期任务后,生成的周期实例会处于冻结状态,并导致依赖该周期实例的下游实例无法运行。
- 暂停(冻结)实例:如果周期任务生成的某个实例不需要运行,您可以选择冻结该实例,冻结操作仅在该 实例生效。不影响当天其他实例与周期任务后期生成的实例的运行。如果您要恢复当前实例的运行,可以 选择解冻并重跑该实例,重跑成功后,下游未运行的实例将会自动运行起来。

•	运维大屏		
\$ \$\$	实时任务运维	~	
t	周期任务运维	^	12100万点 12100元元成下点 重風200万点 空慢で点 で刷新 展开搜索
	周期任务		
	周期实例		
	补数据实例		▲上約分析 早下游分析 正行珍断
	测试实例		
ନ	手动任务运维	~	编辑节点 xc dpe e2
w	智能监控	~	
	这 海运维		0 = (dur 0s) 查看更多详情
			C
	引擎运维	~	# # 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
S.	智能诊断		0 and a construction of the construction of th
~			④ 第二章
(K	操作历史		#
			0 (dur 0a) 来想经作 >
Γ_			「 xu_par_verno 毎 (法法) 本 (法法) 本 (法法)
_			里多 < 1/66 ▶ 当前展示2个节点,2层

周期任务置为空跑:如果某任务一段时间内不需要自动调度运行,但不阻塞下游任务运行时,您可以将任务的调度类型设置为空跑调度,空跑状态的周期任务将生成空跑的实例(运行日志为空,无执行时长,不会实际跑数据)。
S CALLER & CALLER AND AND AND A	\odot	=
	E D 24 4. 267 28	7
MaxCompute引擎实例: xc_DPE_E2_engine 华东2(上海)	X 调度配置	2
1 odps sql	时间属性 ^	201 201
3 author:pre_test_xieyi	* 案例生成方式 @: 🚺 T+1次日生成 🔿 发布后期时生成	
4 create time:2021-11-25 15:36:32 5 **********************************		血 缘 关
6 select		С. Ж
		llx
		*
	● vr 交換任务 于测度	
	08 F at (1277-011)	
2	◎ xc_空跑任务的下游任务	
■ 歴性 上下文 运行日志 操作日志	代码	* *
	TE-00, 7	
DATORES. OUPS SQL		-
(現長史里: 呈現社方) パタはす。 温に出た。	· · · · · · · · · · · · · · · · · · ·	
		-
et i montegitti ,	41年9年1488 无日志,不台川资源 · 内	

 ⑦ 说明 在数据开发界面将任务的调度类型设置为暂停调度与生成运维中心对周期任务进行暂停操 作效果一致。但在数据开发界面的修改仅在开发环境生效,若要在生产环境生效,需要进行发布操作。
 详情请参见: 创建并发布周期任务。

⑦ 说明 冻结解冻周期任务和周期实例的影响请参见:任务冻结与解冻。

下线周期任务

当某任务或业务流程不需要运行时,您可以选择下线该任务或批量下线该业务流程中的任务。操作流程如下:

- 1. 在数据开发界面删除任务(此操作将会在开发环境下线目标任务)。
- 2. 在任务发布界面将任务删除操作发布至生产环境(此操作将会在生产环境下线目标任务)。
- 3. 发布后,在运维中心周期任务界面确认是否已无法搜索到该任务。



? 说明

- 任务执行下线操作时,DataWorks将校验任务在开发环境和生产环境是否存在下游依赖,Q当任务 无下游依赖时才可以被下线,下线操作常见问题请参见:下线节点。
- 任务下线后,您在周期任务界面将无法找到该任务,但已生成的周期实例不会删除,实例到期 (30天)后会自动删除,不支持人为删除。
- 批量下线业务流程中的任务,操作详情请参见: 批量操作。

还原已下线的任务

下线后任务将进入回收站,您可以在回收站还原任务。如果任务需要重新发布,您可以通过上线流程重新上线 该任务,详情请参见:创建并发布周期任务。

⑦ 说明 还原后的任务配置与原配置一致,但节点ID会重新生成。

管理周期任务

您可以在周期任务列表对任务进行管理操作:

- 支持修改任务运行时的相关属性:例如,批量修改调度资源组、修改数据集成任务执行资源组、批量冻结 解冻任务等。详情请参见:查看并管理周期任务。
- 支持针对周期任务设置周期实例调度运行时的状态监控。详情请参见:创建并管理自定义规则。
- 支持针对周期任务设置周期实例、补数据实例、测试实例运行时产出的表数据质量监控。详情请参见:数据质量概述。

⑦ 说明 如果您需要重新定义周期任务的相关属性,可以进入数据开发界面,找到该任务并打开配置窗口进行配置。详情请参见:配置基础属性。如果要批量对任务进行修改,可以进入批量操作界面进行操作。

Dat aWorks

¢	运维大屏		搜索·		米刑 · 法法证书:	「美田」、「「春年人」	清洗探書により、	调度资源归,清洗择	▶ 我的带点
\$ \$\$	实时任务运维	~	132.8%		×主·		W2件页II// *	和皮贝加加口。 用皮引手	* * * * * * * * * * * * * * * * * * *
tī	周期任务运维	^	今E	3修改的节点 🔄 暂停 (冻结) 节点	孤立节点	过期节点			C 副新 展开搜索
	周期任务		_	Ar 44a	** 510	1037TT90 1	行友学型	=/*	
	周期实例		-	白你	UILAT	18以口册 1	忙另突至	贡忙入	阀 使TF
	补数据实例			dsg_demo1	10000020002	2010/07/07 10:00:00	1011.03	terroris, terrori	日 DAG图 测试 补数据 更多
	测试实例			1	100001-000		10110	inserts, land	暂停 (冰结) 日 恢复 (解冻) 多
ନ	手动任务运维	~		test	-	10000000000000000	101112	(Accession, Service)	查看实例 分 法加报整 多
᠕	智能监控	~							修改责任人
⊡	资源运维			movie_rate_tag	1000	10.110.000	1011.05	ansaraha, Janual	日 添加到基线
۲	引擎运维	~		ODPS_SQL_1		101-011-011-011-01	101113	(0.000 (0.000))	8 6 6 6 6 6 6 6 6 6 6 6 6 6
ራ	智能诊断			test_tb12	1000000000	ARC 14 1 1 1 1	8507	Bachs, Stor	日 查看血缘 多 查看节点详情
ଜ	操作历史			product	-	101-011-011-011-01	101113	tion of a device	日 DAG图 测试 补数据 更多
				start	10000	10.00	#17.5	inserts, inst	日 DAG图 测试 补数据 更多
			•						- nxnen i 300+ i 314640 i 35⊘
=			修改	责任人 修改调度资源组 修改	2数据集成资源组	暂停 (冻结) 恢复	(解冻) 更多		: 1/4 > 毎页显示: 40 >

查看周期任务操作记录

支持您通过任务或实例的操作日志,或者单击左侧导航中的操作历史查看操作记录。

¢	运维大屏		a) an ar						
\$\$\$	实时任务运维	R: D型台称/D型D Q D加快定: 前选择D型失量 ◇ 黄柱木: 前选择责任人 ◇ 吨度黄黄理: 前选择 ◇ 其我的D加 今日除60的D加 前将 (An	1) TD MR						
n	周期任务运维	孤立节点							
	周期任务								
	周期实例		E (#) 22						
	补数据实例	₽ 下路分析	请谨慎操作						
	测试实例								
ଜ	手动任务运维								
᠕	智能监控								
	资源运维								
۲	引擎运维								
ų,	智能诊断								
		devide your							
		属性 上下文 豐作日志 代码	* *						
		时间 操作人 操作							
ନ	操作历史	2019-09-04 14:16:31 新聞 / 2000/2004/05							

还支持您在数据开发界面通过**版本 > 对比**功能,获取版本变更的详细情况。详情请参见:发布任务。

	c x										⊚ ≡
•										総布ス	田住さ
MaxComp	and 医关闭:		X								5
	@exclude_input=xc_dw_user_info_all_d odps sql	E:							《上一页 🚺		
		BIOW-		B E本	据交时间	AIB42	156	翻注	語作		
	create time:2019-09-03 18:05:45			V18(77)	已提交开发环境,还未发布至生	产环境 🔨	B#X		主者 对比		御 夫 展
	从合并后的数据表xc_dw_user_info_all_d 很忍有效敢强骂入xc_rpt_user_info_d INSERT OVERWRITE TABLE > Control (dt-'\${bdp.system.bizdate}')			V17	已提交还未发布状态下,再次打	交新版	中间版本		主任 Sitt		
	SELECT uid , MAX(region)			V16	2008年1月24日16340301	Si C	中间版本		重晋 Ritt		
	, MAX(device) , COUNT(0) AS pv , MAX(eender)		•	V15(生)	已发布至生产环境	1923	EX#		主동 (对比		102
	, MAX(age_range) , MAX(zodiac)			V14	2022年1月24日 11:13:04	156	中间数本	Mat	唐晉 i 25년		
	FROM me an aver average and where dt = '\${bizdate}'			V13	2021年11月15日 16:02:02	1922	日波布		· · · · · · · · · · · · · · · · · · ·		
	GROUP BY uid;			V12	2021年11月11日 10:12:36	192 <u>2</u>	已发布		書話:対比		
					2021年8月12日 11:55:35	198 2	已发布		●若 i 对比		
				V10	2020年11月30日 10:38:58	198 X	已发布		主語 対比		
				V 9	2020年11月30日 10:27:00	851 8	已发布		●● 1 対比		

3.4.2. 查看并管理周期任务

周期任务是指调度系统按照调度配置自动定时运行的任务。您可以在周期任务页面通过列表及DAG图的方式, 快速查看任务的详细信息并执行相关操作。

背景信息

- 周期任务的创建、发布、运行、暂停调度、下线等任务运维操作说明详情请参见: 周期任务运维概述。
- DataWorks每晚会根据周期任务自动生成第二天待运行的周期实例,同时支持您手动触发周期任务,分别生成补数据实例和测试实例,您可以通过生成的实例查看任务的运行详情。

使用限制

仅Dat aWorks专业版及以上版本用户才可以使用DAG图的节点聚合、上游分析、下游分析功能。版本升级详情 请参见版本服务计费说明。

注意事项

- 请勿操作project name_root 节点,该节点为工作空间根节点。周期任务的实例均依赖于该节点,如果冻 结该节点,周期任务实例将无法运行。
- 任务的变更均需执行发布操作,发布成功后该变更才会在生产环境中生效。发布时可能存在发布失败、发 布流程被阻塞或发布版本不符合预期等情况,因此,建议您在任务发布后,进入运维中心>周期任务界面 进行排查:
 - 通过DAG图查看任务的上下游依赖关系是否正确。
 - 通过节点详情查看生产环境中该任务的参数配置等是否符合预期。

进入周期任务页面

- 1. 登录DataWorks控制台。
- 2. 在左侧导航栏,单击工作空间列表。
- 3. 选择工作空间所在地域后,单击相应工作空间后操作列的 :- 图标,选择运维中心。
- 进入运维中心后,在左侧导航栏,单击周期任务运维>周期任务。进入周期任务页面,查看周期任务 列表或DAG图。

查看周期任务列表

三 🏟 DataWorks 运维中心	e Davaštudo 🖞 🗍 🖏 🖏 💎
 ・ 运進大屏 ・ ・ ・	5点名称/市合の Q、編末方集: 湖西府解決方案 マ 並男前程: 空気前程 マ 市会美型: 湖西府市会美型 マ 要任人: dataworks_demo2 マ Region: 湖西府で向の マ 構変変現程: 湖西府 マ 🚺
	C 8# 028#
周期実例	한 15.800 (해외전체) 선명원회 표준시 대표전체 대전전권원 (가 Region 인명원원 정보) (1994년 인명원원 전
81826559	日義文 公共成支援滞留 日義文 会社成支援滞留 日本
潮流交例	(151) (152)
N #885545	180.000 W
	· 传受理任人 [法让于随时
	倉田山 五世市の神道
し 「「「「」」 「「」」 「「」」 「」」 「」」 「」」 「」」 「」」 「	
G MADE	
操作	描述
	在上图的模块①区域,您可以通过筛选条件过滤需要查询的任务。
	您可以根据 <mark>节点名称、节点ID、解决方案、业务流程、节点类型、责任人、Region、调度资源</mark> 组、引擎类型、引擎实例、基线、我的节点、今日修改的节点、暂停(冻结)节点、孤立节 点和过期节点等条件进行精确筛选。
筛选	 ⑦ 说明 ● 暂停(冻结)节点:冻结状态的任务不会自动调度,并且会阻塞其下游任务运行。即暂停(冻结)的周期任务将产生冻结状态的实例,冻结状态的实例将不会调度运行,并且
	 会阻塞具下游任务运行。 孤立节点:指当前任务没有依赖任何父节点,这类型任务不会自动调度运行,如果该孤立节点下游依赖较多,则会影响下游任务的运行。详情请参见:孤立节点。 过期节点:不在有效期内的节点将不会生成实例,且不会运行。
任务列表	在上图的模块②区域,以列表方式为您展示满足过滤条件的周期任务。

周期任务列表为您展示任务提交到调度系统后的调度任务列表,您可以对任务进行相关的运维操作。

操作	描述
単个任务操作	 描述 在上图的模块③区域,您可以对周期任务进行如下操作: DAG图: DAG图为您展示周期任务的上下游依赖关系,您可以在DAG图中对任务进行相关操作。详情请参见:查看周期任务的运行是否符合预期,执行测试操作后,将产生测试实例,您可以通过测试实例页面查看此次测试任务的运行情况,详情请参见执行测试并查看测试实例。 ⑦ 说明 测试会真实跑数据,请谨慎操作。 补数据:用于对周期任务执行补历史数据或未来时间段数据的操作,执行补数据操作后将产生补数据实例,您可以通过补数据实例页面查看此次补数据任务的运行情况,详情请参见执行补数据并查看补数据实例。 更多: 暂停(冻结):当任务及其下游任务一段时间内不需要运行时,您可以选择暂停(冻结)任务。 ⑦ 说明 请勿操作projectname_root节点,该节点为工作空间根节点。周期任务的实例均依赖于该节点,如果冻结该节点,周期任务实例将无法运行。 恢复(解冻):用于恢复冻结状态的周期任务。 查看实例:单击后将跳转至当前任务生成的周期实例页面。 添加报警:用于自定义报警规则,帮助您监控周期任务每日的运行状态。配置完成后,您可以在智能监控 > 规则管理页面查看此次添加的监控规则。详情请参见:自定义规则。 修改责任人:用于修改当前任务物责任人。 添加到基线:用于将当前任务添加至基线上,基线使用详情请参见;智能基线概述。
	• 添加到基线 .用于符当即证券添加主基线上,基线使用详慎填多见. 智能基线截处。 • 修改调度答源组 :用于修改调度任务运行所使用的调度答源组。
	 配置质量监控:用于为当前任务配置质量监控规则,当任务运行时可以对数据进行校验。详情 请参见:数据质量概述。
	○ 查看血缘:用于查看当前任务的血缘信息。
	◎ 单击查看节点详情,用于查看任务的基本属性,相关参数介绍详情请参见:查看节点详情。
批量操作	在上图的模块④区域,您可以批量选择任务,进行 修改责任人、修改调度资源组、修改数据集成资 源组、暂停(冻结)、恢复(解冻)、添加报警、添加到基线和下线节点 等操作。
每页条数展 示	在上图的模块⑤区域,您可以修改每页展示的条数,最大支持每页展示200条。

查看周期任务DAG图

在列表模式下单击周期任务名称或操作栏中的**DAG图**,即可打开该任务的DAG图,DAG图默认展示当前任务的一层父节点和一层子节点,您可以通过DAG图面板的菜单栏查看更多的上下游依赖,同时支持您右键单击任务进行相关操作。

₩ Q & P 节点名称/节点D Q	0	1 ୯ ୭ ୧ ୧ ଅ ଅ
		生产环境, 请谨慎操作
当前展示4个节点、3层	rstambe.dqc ods_log_info_d_dqc rstams oors sot dw_user_info_d 展开文节点 gate 展开文节点 gate 電音中点端電 a音中点端電 富有小湖 gate 一 oper sot pr_user_info_d gate gate <th>下点D: 名称: 注意 定氏 注意 現現現解:日頃度 度任人: 満述: 原刊詳述</th>	下点D: 名称: 注意 定氏 注意 現現現解:日頃度 度任人: 満述: 原刊詳述
功能	描述	
DAG面板功能	在上图的模块①区域,支持您对DAG图进行节点聚合样式的操作。详情请参见:DAG图功能介绍。 • 节点聚合:支持您按任务状态、所属工作空间、贵聚合。 • 下游分析:当周期任务的数量或层级过多时,您可前任务影响的下游任务数。 • 调整DAG图显示样式:您可以根据实际需要,单词图的显示样式。	★、下游分析、调整DAG图显示 每任人、优先级等维度进行任务 可以进行下游分析功能,统计当 击右上角的功能图标,调整DAG
单个任务操作	在上图的模块②区域,您可以对目标任务执行相关操 • 展开父节点:支持按照层级展开当前任务的上游任 通过该功能知道哪些任务将影响当前任务的数据产 • 展开子节点:支持按照层级展开当前任务的下游任 通过该功能知道当前任务将影响哪些任务的数据产 • 查看代码:用于确认生产环境当前任务的代码详惯 任务的最新版本是否已经发布成功。生产任务上线 • 编辑节点:单击后即可跳转至数据开发页面打开当 ⑦ 说明 其他操作请参考查看周期任务列表。	作。 任务,最多支持展开6级。您可以 记。 任务,最多支持展开6级。您可以 记。 请。若不符合预期,请确认当前 请参见: <mark>发布任务</mark> 。 á前任务。

功能	描述
单个任务详情	在上图的模块5区域。您可以查看任务属性,并可以单击 展开详情 查看当前任务的详 细属性、上下文参数、操作日志、代码等信息,详情请参见: <mark>查看节点详情</mark> 。

查看节点详情

支持您查看生产环境中周期任务的属性(调度参数、所使用的资源组等)及代码是否符合预期。

	■ ② & □ 节点名称/节点ID	Q		$\mathbb{C} \mathbb{B} \vdots \vdots \vdots \vdots \vdots \vdots \vdots \vdots \vdots $	EI (II) 23
	▼ 下游分析			生产环境,	请谨慎操作
		rds_数据同步_dqc			
		离线同步	ODPS SQL	J	
			γ		
		uin_uun			
		0	DPS SQL		
»					
	属性 上下文 操作日	志代码			≈ ≈
	名称:	节点ID:		点类型: ODPS SQL	^
	责任人	所属工作空间:	·····································	置监控:未完成 🖸	
	调度类型: 正常任务	调度周期: 日调度	调度	度资源组: 公共调度资源组 🗹	
	Region: 华东2	引擎类型: MAXCON	IPUTE 31	擎实例:	
	执行时间: 00 00 03 * * ?	出错是否重试: 否	超田	村定义:系统默认 2	
	执行参数:	调度生效周期: 1970	0-01-01 00:00:00 - 9999-01-0 优务	先级: 8	
	₩/₽, / <i>ν.+</i> -o+Ξν₩ J-4J-fk L	-1	石 2	+.	-
					• • •

功能	描述			
	您可以查看任务在生产环境的调度属性。界面参数详情请参见: <mark>调度配置</mark> 。			
属性	⑦ 说明 当数据开发界面与生成运维中心周期任务界面的参数配置不一致时,建议您优先确认任务变更后,是否发布成功。			
上下文	为您展示当前任务的输入参数和输出参数,上下文功能相关介绍详情请参见:配置节 点上下文。			
操作日志	用于查看任务的操作记录,包括:操作时间、操作人、具体操作等。			
代码	用于查看生产环境当前任务的最新代码。当代码不符合预期时,请确认任务的最新版本是否已经发布成功。详情请参见:发布任务。			

3.4.3. 查看周期实例

周期实例是周期任务达到启用调度所配置的周期性运行时间时,被自动调度的实例快照。您可以通过列表及 DAG图的方式,快速查看实例的详细信息并执行相关操作。

使用限制

- 仅DataWorks专业版及以上版本用户才可以使用运行诊断功能,目前支持您免费体验,但建议升级至专业版以获取更多产品能力。详情请参见智能诊断。
- 仅DataWorks专业版及以上版本用户才可以使用DAG图的节点聚合、上游分析、下游分析功能。详情请参见版本服务计费说明。
- 部分功能仅支持有运维权限的用户进行操作,如果功能入口置灰或者没有功能入口,请单击右上角工作空间管理 > 成员管理,进入成员管理页面检查目标用户是否有运维权限。详情请参见:角色及成员管理:空间级。
- 不支持手动删除周期实例,实例在过期后(30天左右)平台将自动删除。如果任务不需要再运行,可以选择冻结实例。
- 公共调度资源组实例保留一个月(30天)、日志保留一周(7天)。
- 独享调度资源组任务实例、日志保留一个月(30天)。
- 完成状态的实例日志大于3M时,每天会定时进行清理。

注意事项

- 周期任务定时生成周期实例,实例会按最新的代码运行任务。如果您的任务在实例生成后修改了代码并重新提交发布,则未运行的实例会拉取最新的代码运行任务。
- 如果您需要监控任务的执行情况,需要先对该任务设置相应的监控规则,详情请参见:智能监控概述。对 于配置了监控报警的任务,如果运行失败但您未收到报警信息,请检查是否已在报警联系人页面配置了您 的手机号码与邮箱地址,详情请参见:报警信息。
- 首次生成周期实例的时间与您选择的实例生成方式有关,实例生成方式包括T+1次日生成和发布后即时生成,详情请参见: 实例生成方式。
 - ⑦ 说明 手动重跑任务不会触发自定义规则报警。

任务运行与问题定位

调度任务运行需要满足上游任务运行成功、当前任务定时时间已到、调度资源充足、当前任务未冻结等条件, 详情请参见:任务运行条件。

如果任务为未运行状态,建议您先通过DAG面板中的上游分析功能,快速定位阻塞当前任务运行的关键上游任务,再通过运行诊断功能,诊断关键实例未运行的原因或存在的问题,当任务依赖复杂时,您可以通过该功能快速定位问题,提升运维效率。

查看周期实例列表

Dat aWorks

DataWorks公共云合集·运维中心

_			25-32-41										
=		A38440	± 1										, parazrono 亡 [] ≪ MOM ▲ 6
•	运维大屏	节)	の現実:	节点名称/节点ID Q	业务日期	B 前天 前天 全部	2022-04-06 2022	2-04-06 🟥 业务	対意: 00:00 (§) 23:59	③ 节点类型:	请选择节点类型 🗸	· 运行日期: 请选择运行日期	前 解决方案:请选择解决方案 ~ ①
>>> +7	>>約仕労运程 ∨	<u>4</u> 25	時流程:	业务流程 ~	Region:	请选择region	> 词度资源组 3	北海邦 マ 引	· 通道沒有引意供型	✓ 引撃突例	: 请选择	✓ 基线: 遗选择基线	 麦任人: 通送探责任人
1	REFE	(B)	<u>変換型</u> :	全部 >	运行状态	5: 选择运行状态	✓ 我的节点		1 我的未完成节点	重跑过的节点 🗌 🗄	变模节点 重置	淸空	
	周期实例												
	补数据实例		a	5本信息	1	分典型	患住人	009508.11	7E6363161 11	亚语日轴 1	运行时长 /	11 11111111	幾作 2
Ð	潮试实例 手动任务运维 🗸		i	05		拟节点		1	2022-04-07 00:00:00	2022-04-06	0s	2022-04-07 00:01:32	DAG图 去诊新 重跑 费多 重调 下語
^ ∧ ⊡	智能监控 ~		1		æ	拟节点		1	2022-04-07 00:03:00	2022-04-06	0s	2022-04-07 00:03:46	111成功 终止125行 111時(万論) 111時(万論)
୍ ୧୨	引擎运维 🗸	C		03)	线同步	-	1	2022-04-07 00:21:00	2022-04-06	0s	2022-04-07 00:21:01	100-c (H(K)) 査審由条 皇著节点洋橋 重着進行日志
				38	廃	线同 步		1	2022-04-07 00:04:00	2022-04-06	35	2022-04-07 00:06:08	DAG图!去诊断!重陶!要多
		C			00	DPS SQL		1	2022-04-07 00:00:00	2022-04-06	0s		DAG觀(去诊新(主張)更多
			1	09	00	OPS SQL	100111-0010	1	2022-04-07 00:07:00	2022-04-06	0s	2022-04-07 00:07:01	DAG图(去诊断)重陶(更多
ଜ	 操作历史	C		15	00	PPS SQL		1	2022-04-07 00:05:00	2022-04-06	18	2022-04-07 00:59:23	DAG器(去诊新(重氮)更多
		C			来	线周步	-	1	2022-04-07 00:10:00	2022-04-06	85	2022-04-07 01:28:48	DAG图(去诊断)重陶(更多
				00000000	账	值节点	-	1	2022-04-07 00:10:00	2022-04-06	Os	2022-04-07 00:10:02	DAG器(去沙新)並用(更多
			终止运	テ 重調 置成功	暫停 (冻	結) 恢复 (解冻)	a						< 1/1 > ⑤项显示: 40 ∨
	操作			描述									

操作	描述
筛选	在上图的模块1区域,您可以通过筛选条件过滤出需要查询的实例。
	 ⑦ 说明 ● 默认筛选的业务日期是当前时间的前一天。 ● 支持通过实例ID进行过滤,节点ID与实例ID的关系如下: ● 查看某小时或分钟任务当天产生的小时或分钟调度的实例,您可基于节点ID进行过滤。 ● 查看某小时或分钟任务当天指定的小时或分钟实例,您可通过实例ID进行精准过
	 对于配置了监控报警的实例,您可以勾选过去24h产生告警的节点,过滤当前时间点24 小时内产生了报警的实例列表。过去24小时内产生过报警的实例,可通过单击该实例 DAG图中的告警图标查看报警详情。DAG图查看实例报警详情请参见:智能监控概述。
任务列表	在上图的模块2区域,以列表方式展示满足过滤条件的周期实例。

操作	描述
操作	 描述 在上图的模块3区域,您可以对任务进行相关操作: AAG图:DAG图为您展示周期实例的上下游依赖关系,您可以在DAG图中进行相关操作。详情请参见:DAG图为您展示周期实例的上下游依赖关系,您可以在DAG图中进行相关操作。详情请参见:DAG图为您展示周期实例的上下游依赖关系,您可以在DAG图中进行相关操作。详情请参见:EAGE的物件,这不可以使用这功能定任何题。详情请参见智能诊断。 重整:用于对运行成功或运行失败的任务进行重题操作,当任务执行成功后可以触发下游未运行状态任务的调度。常用于对运行成功或运行失败的任务进行重题下游中点的操作,您可以触发下游未运行状态任务的调度。常用于处理出错节点和漏跑节点。 ② 说明 仅支持重题成功和失败状态的任务。 更多: 重题下游:用于对运行成功或运行失败的任务,过算之时状态任务的调度。常用于处理数据修复。 ⑦ 说明 仅支持重题成功和失败状态的任务,如果勾选了其它状态的任务,页面会提示已选节点中包含不符合运行条件的节点,并禁止提交运行。 置成功:用于将运行失败的任务设置为成功状态。当任务执行失败,但希望该任务不阻塞下游住务的运行时,可以使用此功能。常用于处理出错节点。 重成力:用于将运行失败的任务设置为成功状态,当任务执行失败,但希望该任务不阻塞下游任务的运行时,可以使用此功能。常用于处理出错节点。 重成了:用于终止不需要运行的任务,任务终止运行后,将失败退出。 ① 说明 仅支持对状态为等待时间、等待资源和运行中的实例进行终止运行操作。 暂停(基金):当前实例及其下游实例都不需要执行的时候,可选择使用此功能。周期实例中的冻结只针对当前实例,且正在运行中的实例,法运行。 使用 你没持对法态为实例解示。 如果该实例还未运行,则上游任务运行完单后,会自动运行。 如果该实例还未运行,则上游任务运行完单后,会自动运行。 如果该实例还未运行,则上游任务运行完单后,会自动运行。 如果该实例还未运行,则上游任务运行完单后,会自动运行。 ① 说明 你法操作仅作用于该实例,如果您周期任务仍是冻结状态,那么靠二天生成的实例仍会是冻结状态。 更有如任务:更有优性有关。 ① 说明 你法操作仅作用于该实例,如果您周期任务仍是冻结状态,那么靠二天生成的 ① 说明 你法操作仅作用于该实例,如果您周期任务仍是冻结状态,那么靠二天生成的实例仍会是冻结状态。
	 • 查看 口 从 FT 间 · 亘 目 当 則 关 ற 的 举 平 后 忌 • 查看 运 行 日 志 : 查看 当 前 实 例 的 运 行 日 志 。
批量操作	在上图的模块4区域,您可以批量选择任务,进行 终止运行、重跑、置成功、暂停(冻结) 和 恢复 (解冻) 等操作。
每页条数展 示	在上图的模块5区域,您可以修改每页展示的条数,最大支持每页展示200条。

查看周期实例DAG



功能	描述
	 您可以右键单击目标节点,执行相关操作。 展开父节点:用于查看当前节点的上游任务,可通过该功能知道哪些节点将影响当前节点的数据产出。支持按照层级展开父节点,最多支持一次展开6级。 展开子节点:用于查看当前节点的下游任务,可通过该功能知道哪当前节点将哪影响哪些节点的数据产出。支持按照层级展开子节点,最多支持一次展开6级。 查看代码:用于确认生产环境当前节点的代码。若不符合预期,请确认节点最新版本是否已经发布成功。 编辑节点:单击后即可跳转至数据开发页面打开当前节点。 续跑:用于续跑失败状态的任务,续跑将从任务执行失败的位置继续运行,即当任务中存在多段sql的情况下,续跑可从具体失败的sql处重新运行该任务。
	 ⑦ 说明 • 仅MaxCompute的SQL任务支持此操作。 • 2021年1月之后购买的独享调度资源组请提工单联系我们升级。
单个实例操作	 紧急操作:紧急操作只对当前节点本次有效。 去除依赖:用于对指定任务进行紧急去依赖的操作,可通过该功能解除当前节 点的依赖关系。常用于上游失败并与此实例没有数据关系时,紧急去除任务的 上游依赖,使当前任务运行起来。
	⑦ 说明 需要根据任务代码、任务血缘确认该操作是否会对数据造成影响。
	 强制重跑:可用于强制重跑当前节点,支持对运行成功、运行失败、未运行的周期实例进行此操作,常用语数据修复。 强制重跑下游:可用于重刷业务日期昨天和前天的数据,仅支持对运行成功、运行失败的周期实例进行此操作,常用于数据修复。
	⑦ 说明 仅工作空间管理员、租户管理员、阿里云主账号可以发起强制 重跑下游的操作。
	⑦ 说明 其他操作请参考查看周期实例列表。
单个实例详情	在上图的模块5区域。您可以进行如下操作:

查看实例详情

20-			🔮 хс_	oss_数据同步 离线同步		
属性	上下文	运行日志	操作日志	代码		* *
名称: "	no_otort		节点10:		实例ID:	*
节点类型: 虚	拟节点		责任人:		所属工作空间:	
调度类型: 正	常任务		调度周期: 日	调度	调度资源组: 公共调度资源组	
任务状态: 运	行成功		定时时间: 2	022-04-07 00:11:00	开始等待资源: 2022-04-07 16:12:13	
等待资源耗时	-		运行时长: 0	s	开始运行时间: 2022-04-07 16:12:13	
结束时间: 20)22-04-07 16:12:13		实例状态: 卤	訓業型实例	Region:	
引擎类型:			引擎实例:		出错是否重试: 否	-
· 結束时间: 20 引擎类型: ∢	JZZ-04-07 16:12:13			制尖型実例	кеgion: 出错是否重试: 否	•

功能	描述			
	您可以在此页签查看任务在生产环境的调度属性。界面参数详情请参见: <mark>调度配置</mark> 。			
属性	 ⑦ 说明 • 伟点D和实例D的关系: 动作动动动动动动动动动动动动动动动动动动动动动动动动动动动动动动动动动动动			
上下文	为您展示当前任务的所有输入参数和输出参数,详情请参见:配置节点上下文。			
运行日志	田干杏看任条详细执行过程			
	用于旦有证为许细汎们过任。			
操作日志	用于查看任务或实例的操作记录(时间、操作人、具体操作)。			
代码	用于查看生产环境当前任务的最新代码。若不符合预期,请确认任务最新版本是否已 经发布成功。详情请参见: <mark>发布任务</mark> 。			

常见问题

- 实例异常排查
 - 运维中心周期任务没有找到该任务
 - 有周期任务,但是周期实例一个都没有

- o 有周期任务,有其他周期实例,但是这个任务没有产生周期实例
- o 有周期任务, 有周期实例但未运行
- o 任务运行必须满足的条件有哪些?
- 任务未运行排查
 - 任务运行条件
 - 任务到定时时间,为什么还没运行?

3.4.4. 执行补数据并查看补数据实例

DataWorks的补数据功能用于在指定日期范围内运行周期任务,支持补当前节点及其下游节点,您可以在补数 据实例页面查看实例的运行状态,以及对补数据实例进行终止、重跑和解冻等操作。本文为您介绍如何对周期 任务进行补数据,并管理补数据实例。

背景信息

周期任务开发完成并提交发布后,任务会按照调度配置定时运行。如果您希望在指定时间段运行周期任务,可 以使用补数据功能。DataWorks周期任务的补数据操作,支持补当前节点及其下游节点,详情请参见补数据, 其支持的补数据模式如下:

- 当前节点:用于对当前节点进行补数据操作。
- 当前节点及下游节点:用于对当前节点及下游节点进行批量补数据操作,通常在当前节点的下游节点数量
 较少时使用该功能,支持指定部分下游节点。
- 海量节点模式:用于对当前节点及下游节点进行批量补数据操作,通常在当前节点的下游节点数量较多时 使用该功能,支持按照项目筛选下游节点。同时,支持节点白名单和黑名单,选中或过滤目标节点。
- 高级模式:用于灵活选择一批节点进行批量补数据操作,节点之间可以不存在依赖关系。您可以在DAG图 上使用可视化方式添加节点,也可以在任务列表中添加节点。
 - 您可以在当前周期任务的DAG图中,结合DAG的聚合功能,按照工作空间、所属责任人或优先级等维度将 节点划分为组,直接勾选该节点组,快速添加目标类别的节点为补数据节点。实例的DAG图,详情请参 见DAG图功能介绍。
 - 您可以在周期任务界面,通过目标条件进行筛选过滤,在左侧任务列表中,勾选符合条件的节点为需要 补数据的节点。

使用限制

- 仅华南1(深圳)、中东东部1(迪拜)地域支持周期任务使用高级模式进行补数据。
- 不支持手动删除补数据实例,实例在过期后(30天左右)平台将自动删除。如果任务不需要再运行,可以 选择冻结实例。
- 公共调度资源组实例保留一个月(30天)、日志保留一周(7天)。
- 独享调度资源组任务实例、日志保留一个月(30天)。
- 完成状态的实例日志大于3M时,每天会定时进行清理。

注意事项

- 当补一个区间的数据时,在启动补数据的第一天存在一个运行失败的实例,则当天的补数据实例会被置为 失败状态,第二天的实例也不会启动运行(只有当天的全部任务实例都成功,第二天的任务实例才会开始 运行)。
- 小时或分钟任务补某一天数据时,当天所有实例是否并发执行与任务是否设置自依赖有关,自依赖的任务 执行补数据时,如果补数据任务中第一个实例前一天的周期实例没有运行,则该补数据任务也无法触发运 行。如果补数据任务的第一个实例前一天没有周期实例,则补数据实例会直接触发运行。

- 如果周期实例和补数据实例均启动运行,为了保证周期实例的正常运行,您需要终止补数据实例的运行。
- 补数据实例过多或并行数过高可能导致周期调度的资源紧张,请您根据需求合理配置。
- 为避免补数据实例占用过多的资源而影响周期实例的运行,平台对补数据实例制定以下规则:
 如果补数据选择业务日期为昨天(T-1),即补当天数据时,补数据任务优先级由任务所在基线优先级决定。
 - 如果补数据选择业务日期为历史业务日期(T-2)时,补数据任务将按照以下规则对任务进行降级处理:
 - 7级、8级任务优先级降成3级。
 - 5级、3级任务优先级降为2级。
 - 1级任务优先级保持不变。

补数据

- 1. 进入**数据开发**页面。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 选择工作空间所在地域后,单击相应工作空间后的进入数据开发。
- 2.
- 3. 在左侧导航栏,单击周期任务运维 > 周期任务,进入周期任务页面。
- 4. 目标节点补数据。
 - i. 单击周期任务列表中的目标任务, 打开该任务的DAG图。

您也可以单击 图标展开周期任务列表,单击目标任务操作列的DAG图,打开该任务的DAG图。

ii. 右键单击目标节点,鼠标悬停至**补数据**,选择相应的补数据模式,配置补数据的相关参数。

⑦ 说明 您也可以在周期任务页面,单击。图标展开周期任务列表,单击目标任务操作列的补数
 据,选择相应的补数据模式,对该周期任务进行补数据。

≡	🍿 DataWorks 遠	运维中心 生产1/	taStudio 🕜 任务发布	ር ፲ ዲ 💎 🗕 🗧
e	运维大屏		8. 速洪悠 🗸	
\$\$\$	实时任务运维 🖌 🖌 🖌		1. NHA20+ *	
t1	周期任务运维 个			
	周期任务	名称 田 日 市点名称/市点0 Q		
	周期实例			
	补数据实例	□ ▼下游分析		生产坏境, 请建俱操作
	测试实例	Complete and address	展开父节点	
ର	手动任务运维		展开子节点	
м	智能监控 >	o oneclickHolo	查看节点详情 音看代码	
	资源运维	虚拟节点	编辑节点	
۲	引擎运维 🗸 🖌 🗸		查看实例	
Ų,	智能诊断		 	
		holo_di_	秋数据 >	当前节点
		高线同步	暂停 (冻结)	当前节点及下游节点
		And a strength or production of	恢复 (解冻)	海量节点模式

补数据的模式及参数配置如下:

○ 使用**当前节点**模式补数据。

	补数据		×
	▲ 补数据实例	过多或并行数过高可能导致周期调度的资源紧张。	
	补数据名称:	P_oneclickHolo	
	当前任务:	oneclickHolo,	
	业务日期:	2021-08-23 世 删除	
		添加	
	是否并行:		
	顺序:	● 业务日期正序 ○ 业务日期倒序	
•			•
		确定取	硝

参数配置如下。

参数	描述
补数据名称	系统会自动生成补数据名称,您可以根据需要进行修改。
当前任务	当前选中的任务。
业务日期	 选择执行补数据任务的业务日期,业务日期精确到天。 如果您需要对该任务的多个不连续的时间段进行补数据,则可以单击添加,配置多个时间段。 当配置的业务日期晚于当前日期,则您可以选择立即运行定时时间大于此刻的补数据实例,后续当实际时间大于配置的业务日期时间,系统会立即运行该补数据实例。 例如,当前日期为 2021-08-24 ,业务日期为 2021-09-17 ,则等实际日期为 2021-09-18 (立即大于业务日期)时,该补数据实例会立即启动运行。 ⑦ 说明 建议补数据的时间不要太长,以免出现任务需要等待资源的情况。
是否并行	 您可以通过选择是否并行,控制同时生成多少个补数据实例来进行补数据,具体如下: 不并行:按照业务日期顺序串行执行,前一个补数据实例运行结束后下个补数据实例才会运行。 并行:根据您配置的并行数,按照业务日期将补数据实例拆分成若干组并行执行。同一时刻有多个不同业务日期的补数据实例运行。

参数	描述
	用于配置当前补数据任务生成的补数据实例个数,当进行补数据时,所有补数据实例并行执 行。
	⑦ 说明 当是否并行参数选择并行时,需要配置当前参数。
	并行数取值最小为2组,最大为10组。多个实例并行执行的情况如下:
并行数	 如果业务日期的跨度时间少于配置的并行组数,则并行执行。例如业务日期是1月11日~1月 13日,并行数配置为4组,则只会生成三个补数据实例(每个补数据实例对应一个业务日 期),三个实例同时并发执行。
	如果业务日期的跨度大于选择的并行组数,则系统会结合配置的业务日期顺序兼有串行和并行。例如业务日期是1月11日~1月13日,并行数配置为2组,则会生成两个补数据实例(其中一个补数据实例会有两个业务日期,这两个业务日期对应的任务串行执行),两个补数据实例并行执行。
顺序	您可以选择按照业务日期正序或业务日期倒序进行补数据。

○ 使用当前节点及下游节点模式补数据。

补数据 ×								
▲ 补数据实例过多或并行数过高可能导致周期调度的资源紧张。								
补数据名称:	补数据名称: P_oneclickHolo							
业务日期:	2021-08-23 法加	2021-08-23	曲 删除					
是否并行:								
顺序:	● 业务日期正序 🔵 业务日	期倒序						
选择需要补数	据的节点	0		=	2477			
搜索: 按名称进行	J搜索	Q		Jz	宝驭: 请选择	~		
任务名	称		任务类型 🎧	责任人 🎧	工作空间 🎧			
			虚拟节点	10000	1.000	•		
- 11	1212.9.57***		离线同步	10000	100			
			离线同步	grant at	1100	Ŧ		
4						ł		
					确定即	硝		

参数配置如下。

参数	描述					
补数据名称	系统会自动生成补数据名称,您可以根据需要进行修改。					
业务日期	 选择执行补数据任务的业务日期,业务日期精确到天。 如果您需要对该任务的多个不连续的时间段进行补数据,则可以单击添加,配置多个时间段。 当配置的业务日期晚于当前日期,则您可以选择立即运行定时时间大于此刻的补数据实例,后续当实际时间大于配置的业务日期时间,系统会立即运行该补数据实例。 例如,当前日期为 2021-08-24,业务日期为 2021-09-17,则等实际日期为 2021-09-18 (立即大于业务日期)时,该补数据实例会立即启动运行。 ③ 说明 建议补数据的时间不要太长,以免出现任务需要等待资源的情况。 					
是否并行	 您可以通过选择是否并行,控制同时生成多少个补数据实例来进行补数据,具体如下: 不并行:按照业务日期顺序串行执行,前一个补数据实例运行结束后下个补数据实例才会运行。 并行:根据您配置的并行数,按照业务日期将补数据实例拆分成若干组并行执行。同一时刻有多个不同业务日期的补数据实例运行。 					
并行数	 并行数取值最小为2组,最大为10组。多个实例并行执行的情况如下: 如果业务日期的跨度时间少于配置的并行组数,则并行执行。例如业务日期是1月11日~1月13日,并行数配置为4组,则只会生成三个补数据实例(每个补数据实例对应一个业务日期),三个实例同时并发执行。 如果业务日期的跨度大于选择的并行组数,则系统会结合配置的业务日期顺序兼有串行和并行。例如业务日期是1月11日~1月13日,并行数配置为2组,则会生成两个补数据实例(其中一个补数据实例会有两个业务日期,这两个业务日期对应的任务串行执行),两个补数据实例并行执行。 					
顺序	您可以选择按照业务日期正序或业务日期倒序进行补数据。					
选择需要补 数据的节点	 您可以根据任务名称、层级等条件进行筛选过滤,选择需要补数据的目标节点。 ③ 说明 通过任务名称进行搜索时支持模糊匹配,即输入关键字,即可显示所有名称中包含关键字的节点。 选择需要补数据的节点区域为您展示的是当前节点及其包含的所有下游节点(包含下游节点的子节点),您可以选择当前节点及其下游节点的部分或全部节点任务(包含当前节点)进行补数据。 					

○ 使用海量节点模式补数据。

补数据				×			
▲ 补数据实例过多或并行数过高可能导致周期调度的资源紧张。							
补数据名称:	P_oneclickHolo		No. 10180-101				
业务日期:	2021-08-23 2021-0	8-23	□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□				
	添加						
顺序: 【 坛工作 杰语	 业务日期正序 业务日期创序 3.3.3.3.3.3.3.3.3.3.3.3.3.3.3.3.3.3.3.						
 按工1+至内 ✓ 包含当前节点 							
您的所有项目			补数据的项目				
项目筛选	Q		项目筛选	Q			
Date		•					
dem		>					
flink	and the second sec	<	暂无数据				
Flink							
shar		•					
195项			0项				
节点白名单 🕐 :	节点搜索			~			
节点黑名单 ⑦:	节点搜索			~			
•							
				确定取消			

参数配置如下。

参数 描述	
顺序 您可以选择按照业务日期正序或业务日期倒序进行补数据。	
您可以根据业务需求,在 您的所有项目 区域,选择目标工作空间,将其添加至 补数 目区域,为目标工作空间中所选节点进行补数据。	居的项
 按工作空间选择需要补数据的节点 第选DataWorks工作空间支持模糊匹配,即输入关键字,即可显示所有名含关键字的工作空间。 您只能选择当前所在地域中的DataWorks工作空间进行补数据操作。 您可以为该补数据任务配置节点白名单或黑名单,选中或过滤无需执行补节点。 您可以选择该补数据任务是否包含当前节点,取值如下: 如果勾选包含当前节点,则补的是当前节点及下游节点的数据。 如果未勾选包含当前节点,则当前节点会空跑,补的是当前节点点的数据。 	称中包 数据的 下游节
除选中的DataWorks工作空间中包含的节点外,仍需要进行补数据的节点。	
⑦ 说明 目前仪文持通过节点D股家目标节点。	
选中的DataWorks工作空间中不需要进行补数据的节点。	
节点黑名单 ⑦ 说明 目前仅支持通过节点ID搜索目标节点。	

○ 使用**高级模式**补数据。

在高级模式中,您可以使用DAG图的聚合功能及节点的类型、责任人等过滤条件,批量为节点之间不存 在依赖关系的多个节点进行补数据。

≡	ntaWorks 🏟	运輸	中心 🕮		● DataStudio ● 任务发布	₵ ₴ ₰ ♥ ■₽₽₽₽₽₽
e 555	运输大屏 实时任务运维	*	搜索:	市点名称/市点D Q 市場調	nin andronen v Refu andrest v Arene and v house of the Cash ta aire and the Cash ta aire and the Cash ta aire a	C ser ETHE
n	周期任务运维	^		名称		C 5 0 0 5 3 -
	周期任务		-			4+
	周期实例			-		王广环境,调建俱保作
	1180ESCRI			Approximation of the local distance of the l		
	RECEN		-			
ର	手动任务运维	~		A DECEMBER OF A		
*	智能监控	~	-	-		
	239554					
۲	引擎运行	•		same to		
Ŷ,	智能沙 斯			1000 000 000 000 000 000 000 000 000 00		
				2222022		
				10000-000 *		
					384 2.432 holo holo state	
				#		
				1.00.000		
				Sherrow Cont		
				1272702.027***		
					1489-147-0 10	
=				× 1/10 >	Lengengever (Kous) vogs (

高级模式补数据步骤如下:

- a. 选择需要补数据的节点。
 - 在当前周期任务的DAG图中,您可以结合DAG图的聚合功能(区域1),按照不聚合、按所属工作空间聚合、按责任人聚合或按优先级聚合等维度将节点划分为组,直接勾选该节点组(区域2),快速添加目标类别的节点为补数据节点。DAG图的聚合功能,详情请参见DAG图功能介绍。
 - 您也可以在周期任务界面,通过节点名称、节点类型、责任人、调度资源组等筛选条件进行 过滤(区域3),在左侧任务列表中,勾选符合条件的周期任务为需要补数据的节点(区域 4),单击下方的添加,将目标周期任务添加至补数据任务中。

⑦ 说明 该方式是将整个目标周期任务添加至补数据任务中,因此系统会对该周期任务中的所有节点进行补数据。如果您希望对周期任务中的部分节点进行补数据,则可以单击目标周期任务的名称,进入DAG图界面,选择需要进行补数据的节点。

b. 查看补数据节点。

补数据节点添加完成后,您可以在区域5的**补数据**面板,查看已选择的节点。同时,您还可以执行 如下管理操作:

- 单击目标节点后的 图标,跳转至该节点的DAG图界面,您可以重新勾选该DAG图中需要补数据 的节点。
- 单击目标节点后的 图标, 在当前补数据任务中删除该节点。
- c. 在区域5的**补数据**面板,单击**去配置**,配置补数据任务的相关参数。

补数据		×
▲ 补数据实例	別过多或并行数过高可能导致周期调度的资源紧张。	
补数据名称:	P_oneclickHolo_	
已选任务数:	4个 修改	
业务日期:	2021-08-23 - 2021-08-23	
	添加	
是否并行:		
顺序:	● 业务日期正序 ─ 业务日期倒序	
4		Þ
	确定	取消

参数配置如下。

参数	描述
补数据名称	系统会自动生成补数据名称,您可以根据需要进行修改。
已选任务数	当前补数据任务中包含的节点个数。您可以单击 修改 ,修改当前补数据任务包含的节点。
业务日期	 选择执行补数据任务的业务日期,业务日期精确到天。 如果您需要对该任务的多个不连续的时间段进行补数据,则可以单击添加,配置多个时间段。 当配置的业务日期晚于当前日期,则您可以选择立即运行定时时间大于此刻的补数据 案例,后续当实际时间大于配置的业务日期时间,系统会立即运行该补数据实例。 例如,当前日期为 2021-08-24 ,业务日期为 2021-09-17 ,则等实际日期为 2021-09-18 (立即大于业务日期)时,该补数据实例会立即启动运行。 ⑦ 说明 建议补数据的时间不要太长,以免出现任务需要等待资源的情况。
是否并行	 您可以通过选择是否并行,控制同时生成多少个补数据实例来进行补数据,具体如下: 不并行:按照业务日期顺序串行执行,前一个补数据实例运行结束后下个补数据实例才会运行。 并行:根据您配置的并行数,按照业务日期将补数据实例拆分成若干组并行执行。同一时刻有多个不同业务日期的补数据实例运行。

参数	描述
并行数	 并行数取值最小为2组,最大为10组。多个实例并行执行的情况如下: 如果业务日期的跨度时间少于配置的并行组数,则并行执行。例如业务日期是1月11日~1月13日,并行数配置为4组,则只会生成三个补数据实例(每个补数据实例对应一个业务日期),三个实例同时并发执行。 如果业务日期的跨度大于选择的并行组数,则系统会结合配置的业务日期顺序兼有串行和并行。例如业务日期是1月11日~1月13日,并行数配置为2组,则会生成两个补数据实例(其中一个补数据实例会有两个业务日期,这两个业务日期对应的任务串行执行),两个补数据实例并行执行。
顺序	您可以选择按照 业务日期正序 或 业务日期倒序 进行补数据。

5. 单击**确定**,启动补数据。

管理补数据实例

补数据任务启动后,会生成相应的补数据实例,您可以在运维中心页面,单击左侧导航栏的周期任务运维 > **补数据实例**,查看补数据实例的基本信息及运行详情,并进行终止运行、重跑等管理操作。进入运维中心, 详情请参见进入运维中心。

☰ 🛊 DataWorks 运	维中心 生产化												🔗 DataStudi	。 の 任务发布 🇘 🗊 🔍	*
・ 运维大屏	没意: 1	节点名称/节点ID	 补数据名标 	F:	请选择补数据	洺称		~	创建人:	请选择责任	X	~	创建日期:	2021-08-24	
🐝 实时任务运维 🗸 🗸	运行状态:	选择运行状态	✓ 业务日期:		选择业务日期	3		Ē	节点类型:	请选择节点	类型	~	责任人:	请选择责任人	~
13 周期任务运维 ^	地域:	请选择region	· 國際资源:	B: [请远绎			\mathbf{v}	引擎实例:	请选择		~	基线:	请选择基线	~
周期任务	2 我的节点 2	我发起的 重重 刷新													◎ 数据表へ
周期实例	任务名称		检查状态		运行状态	5	创建人		创建日期		节点数	业务	目期	操作	*
补数据实例	+ 0	2							2021.08.24						
測試实例			-						2021-00-24					RAILSELL RAILING SCA	
♀ 手助任务运师 ×	- P.	AD. 201001-0011	1						2021-08-24		1	1		批量终止「批量重調」复用	
▲ 智能监控 ~	- 2021-01-	01 00:00:00													
		名称	责任人			定时时间		я	始运行时间计		结束时间计		运行时长小	操作	
			-	-		2021-01-02 00:25:0	0	20	121-08-24 10:43:32		2021-08-24 10:44:37		1m4s	终止运行「重张」更	3
		0	-			2021-01-02 00:29:0	0	20	21-08-24 10:44:44		2021-08-24 10:45:53		1m8s	终止运行「重视」更	5
		0	-			2021-01-02.00:18:0		20	21-08-24 10:46:19		2021-08-24 10:46:36		16s	终止运行「重施」更	8
		0	100			2021-01-02.00:10:0)	20	21-08-24 10:45:56		2021-08-24 10:46:15		196	终止运行「重難」更	5
		Ø				2021-01-02.00:16:0	0	20	21-08-24 10:43:35		2021-08-24 10:44:27		52s	终止运行「重魏」更	8
		0	-			2021-01-02.00:11:0)	20	21-08-24 10:43:29		2021-08-24 10:43:29		0s	终止运行「重発」更近	8
=	後止進行	<u> 28</u>											每页量	研: 29 マ く上一	र हन्त 🕫 🛛
区域	描述														

区域 描述 1 在 底区域、您可以通道博选条件过滤出需要查询的实例、		
2 在这区域,您可以通过筛选条件边流出需要查询的实例。 第□以根据节点名称,节点IO,补数据名称,创建人、创建日期,运行状态、业务日期,我的节点,我 数据的等条件进行精确筛选。 1 ⑦ 识明 • 在该区域,您需要单击展开搜索,才会显示节点类型调度资源组引学实例等更多的筛选条 作。 • 节点名称支持模糊匹配,即输入关键字,即可显示所有名称中包含关键字的节点。 * 方态名称支持模糊匹配,即输入关键字,即可显示所有名称中包含关键字的节点。 * 任务名称:为您展示补数据实例的相关信息,包括: • 任务名称:方您展示补数据实例的相关信息,包括: • 任务名称:方您展示补数据实例的相关信息,包括: • 任务名称:方您限示补数据实例的相关信息,包括: • 任务名称:约定有压。(即区域3)。 • 检查状态: · 任务名称:方应的运行信息(即区域3)。 • 检查状态: · 任务名称:约相关的资金的方式的运行信息(即区域3)。 · 他查求法:约定时,未运行,等待资源,异常,你此五种状态。 · 包括全线,创业要的自需。 · 法后代,创业资本有关数 · 市点个款:分据实例自由通道、 · 中点有量: 创建补数据实例的目题。 · 中点有量:公计数据实例自用。 · 在家区域,您还可以对补数据实例自用。 · 在地名主,论可以选择批量等业补数据实例中运行状态为等待运行和运行中的实例,执行业操作后,对 应该例能被置为失败状态。 ⑦ 说明 • 不支持手动删除补数据实例。 • 不是理事: 批量事新运行升数据实例。 • 小型理事: 就是有力或法方未适行,运行成功或运行失败的实例认识数据实例,实例在过期后(30天左右)平台称自动删除,如果任务不需要 	区域	描述
3 在该区域,您可以查看补数据实例的相关信息,包括: 4 先务名称:为您展示补数据实例的名称。单击实例名称前的+图标,即可显示该实例的运行日期,实例 包含的节点及节点的运行信息(即区域3)。 4 法检末 : 兰前补数据实例的检查状态。 6 法在状态: 兰帕补数据实例的自要点账号。 6 法位计, 创建补数据实例的同里云账号。 6 物量计数据实例的同里云账号。 7 M健 日期, 创建补数据实例的问题。 7 就用 7 NU 8 不支持手动删除补数据实例, 实例在过期后 (30天左右) 平台将自动删除。如果任务不需要 满运行,可以选择冻结实例。 9 NI DataWork当前不支持对状态为未运行、运行成功或运行失败的实例执行终止运行操作。 9 NI DataWork当前代文持批量重邀状态为运行失败的补数据实例。 19 NI DataWork当前代文持批量重邀状太为运行失败的补数据实例和数据实例。 19 NI DataWork当前代文持批量重邀状态为运行失败的补数据实例和数据。 19 NI DataWork当前代文计数据集件的节点集,方便您快速达定要到。 19 NI 使用舞量方式就量式补数据生成的实例,不支持复用。	1	在该区域,您可以通过筛选条件过滤出需要查询的实例。 您可以根据节点名称、节点ID、补数据名称、创建人、创建日期、运行状态、业务日期、我的节点、我 发起的等条件进行精确筛选。 ⑦ 说明 • 在该区域,您需要单击展开搜索,才会显示节点类型调度资源组引擎实例等更多的筛选条 件。 • 节点名称支持模糊匹配,即输入关键字,即可显示所有名称中包含关键字的节点。
 再运行,可以选择冻结实例。 DataWorks当前不支持对状态为未运行、运行成功或运行失败的实例执行终止运行操作。 批量重跑:批量重新运行补数据实例。 ① 说明 DataWorks当前仅支持批量重跑状态为运行失败的补数据实例。 复用:您可以选择复用某次补数据操作的节点集,方便您快速选定需要进行补数据的节点。 ⑦ 说明 使用海量节点模式补数据生成的实例,不支持复用。 	2	在该区域,您可以查看补数据实例的相关信息,包括: 任务名称:为您展示补数据实例的名称。单击实例名称前的+图标,即可显示该实例的运行日期,实例包含的节点及节点的运行信息(即区域3)。 检查状态:当前补数据实例的检查状态。 运行状态:包括运行中、未运行、等待资源、异常、停止五种状态。 创建人:创建补数据实例的阿里云账号。 创建日期:创建补数据实例的日期。 节点个数:补数据实例包含的节点数量。 业务日期:运行补数据实例的日期。 在该区域,您还可以对补数据实例执行如下操作: 批量终止:您可以选择批量终止补数据实例中运行状态为等待运行和运行中的实例,执行此操作后,对应实例将被置为失败状态。 ⑦ 说明 不支持手动删除补数据实例,实例在过期后(30天左右)平台将自动删除。如果任务不需要
 复用:您可以选择复用某次补数据操作的节点集,方便您快速选定需要进行补数据的节点。 ⑦ 说明 使用海量节点模式补数据生成的实例,不支持复用。 		再运行,可以选择冻结实例。 DataWorks当前不支持对状态为未运行、运行成功或运行失败的实例执行终止运行操作。 批量重跑:批量重新运行补数据实例。 ⑦ 说明 DataWorks当前仅支持批量重跑状态为运行失败的补数据实例。
⑦ 说明 使用海量节点模式补数据生成的实例,不支持复用。		 复用:您可以选择复用某次补数据操作的节点集,方便您快速选定需要进行补数据的节点。
		⑦ 说明 使用海量节点模式补数据生成的实例,不支持复用。

DataWorks公共云合集·运维中心

区域	描述
3	在该区域, 您可以查看补数据实例所包含节点的相关信息,包括: 名称:为您展示补数据实例所包含节点的名称。单击节点名称,即可进入节点的详情页面。 责任人:节点任务所在的工作空间责任人。 定时时闻:节点任务设置的定时运行时间。 开始运行时间:节点任务设置的定时运行时间。 结束时间:节点任务运行的结束时间。 运行时长:节点任务运行的结束时间。 运行时长:节点任务运行的结束时间。 运行时长:节点任务运行的结束时间。 逐行时长:常点任务运行的结束时间。 逐行时长:常点任务运行的结束时间。 逐行时长:常点任务运行的结束时间。 * 运行时长:常点任务动行称: 终止运行:终止运行状态为等待运行和运行中的节点,执行此操作后,该节点将被置为失败状态。 ⑦ 说明 DataWorks当前不支持对状态为未运行、运行成功或运行失败的节点执行终止运行操作。 重跑:重新运行目标节点任务。 ⑦ 说明 仅支持重跑状态为运行成功和运行失败的节点。 更多 > 重跑下游:重新运行目标节点任务的状态置为成功。 更多 > 暂停(冻结):将暂停(冻结)时点做复调度。 更多 > 查看血缘:查看节点的血缘关系图。
4	您可以在区域3勾选多个目标节点任务,在本区域(区域4)单击 终止运行 或 重跑 ,即可批量终止运行或重 新运行所选的节点任务。

实例状态说明

序号	状态类型	状态标识
1	运行成功状态	\odot
2	未运行状态	Θ
3	运行失败状态	8
4	正在运行状态	۲
5	等待状态	0
6	冻结状态	۲

3.4.5. 执行测试并查看测试实例

测试实例是对周期任务进行测试时产生的实例,您可以在测试实例界面,通过列表及DAG图的方式快速查看实例的详细信息。

背景信息

您可以通过对周期任务执行测试操作,以验证当前周期任务在生产环境的调度运行是否符合预期。测试操作将 产生测试实例,您可以在测试实例界面查看实例运行情况。

- 对周期任务执行测试操作详情请参见:查看并管理周期任务。
- 通过列表模式查看测试实例执行详情请参见:查看测试实例列表。
- 通过DAG图查看测试实例执行详情请参见:测试实例DAG图。

⑦ 说明 测试实例的运行会真实执行代码逻辑,请谨慎操作。

使用限制

- 功能使用限制:
 - 仅DataWorks专业版及以上版本用户才可以使用运行诊断功能,目前支持您免费体验,但建议升级至专 业版以获取更多产品能力。详情请参见智能诊断。
 - 仅DataWorks专业版及以上版本用户才可以使用DAG图的节点聚合、上游分析、下游分析功能。详情请参 见版本服务计费说明。
 - 部分功能仅支持有运维权限的用户进行操作,如果功能入口置灰或者没有功能入口,请单击右上角工作
 空间管理 > 成员管理,进入成员管理页面检查目标用户是否有运维权限。详情请参见:角色及成员管理:空间级。
- 日志与实例使用限制:
 - 不支持手动删除测试实例,实例在过期后(30天左右)平台将自动删除。如果任务不需要再运行,可以 选择冻结实例。
 - 公共调度资源组实例保留一个月(30天)、日志保留一周(7天)。
 - 独享调度资源组任务实例、日志保留一个月(30天)。
 - 。完成状态的实例日志大于3M时,每天会定时进行清理。

执行测试操作并生成测试实例

- 1. 进入运维中心。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏,单击工作空间列表。
 - iii. 选择工作空间所在地域后,单击相应工作空间后操作列的 :图标,选择运维中心。
- 2. 对周期任务执行测试操作。

进入**运维中心**后,在左侧导航栏,单击**周期任务运维 > 周期任务**。进入**周期任务**页面。您可以通过以 下两种方式对周期任务执行测试操作。

- 方法一:在周期任务列表模式下,单击相应任务操作列的测试。如果当前页面非列表模式,您可单击 页面中间的 图标返回列表模式。
- 方法二:在DAG模式下,右键单击相应实例的DAG图,单击菜单中的测试。如果当前页面非DAG图模

- 式,您可单击相应实例操作列的DAG图,打开该任务的DAG图。
- 3. 配置冒烟测试相关参数。

冒烟测试	×								
如果业务日期选择昨天之前,则立即执行任务。 如果业务日期选择昨天,则需等到定时时间才能执行任务。									
* 冒烟测试名称: * 选择业务日期:									
参数	说明								
冒烟测试名称	您可以输入冒烟测试的名称。								
选择业务日期	选择需要进行测试的业务日期。 ⑦ 说明 业务日期选择昨天之前,则立即执行任务。 如果业务日期选择昨天,则需等到任务定时时间才会执行任务。 								

查看测试实例列表

测试实例列表为您展示周期任务执行测试操作后生成的测试实例,您可以查看测试实例的执行情况,并对实例 执行相关操作。

Dat aWorks

= @ C	DataWorks)ź	5维中心 🖁	71	•						в	DataStudio 🔮 🗹 🖏 1844 💎
🕒 运维大屏		沒來	节点名称/节点ID Q 节点类	1: HBFTARD ~	责任人 : 请选择责任人	✓ 运行	日期: 请选择运行日期	曲 业务日期: 清	选择业务日期 曲	运行状态: 选择运行状态 🗸	Region: 请选择region ~ 1
 第二日 第二日	运输 🗸	调度多	I課題 Please Select V 引擎类型:	请选择引擎类型 >	引擎实例: Please Select	→ 基线:	请选择基线	✓ 我的节点	我今天测试的节点	暫停(冻結)节点 重量	清空
周期任务			warana awala	In Am 201 TE	wer / C	(2.4×25 I)	rheadershill 18	July DW IN	Cold In	T146-0000 18	○期新校起搬表
周期实例			幸 本信息	TRAC	RILA	N/2/DWC41	지도마가마가에 시 다	20199 CL AD 41	地行时代 化	AL GUERRALL	se garte 🗸
indered Millered	1			1000.00		1	2022-04-07 00:01:00	2022-04-06	115	2022-04-07 17:35:47	20 DAG图 去诊断 重點 更多 🔇
 ・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	运维 • · •			10010		8	2022-03-28 00:29:00	2022-03-27	56s	2022-03-28 14:09:52	2(DAG图 去诊新 重加 更多
 日報総合戦 日報 日報				100101	10111,012	8	2022-03-28 00:10:00	2022-03-27	4s	2022-03-28 11:08:38	2(DAG器 去诊断 重跑 更多
				100000	1000-010,000.	8	2022-03-28 00:29:00	2022-03-27	52e	2022-03-28 11:07:48	2(DAG图 去诊断 靈胸 更多
	:	4	1287 (1995) (1995) (1995	(204) (1027 (1628)	0						< 1/1) ・ 1/1) ・ 1/1) ・ 40 マ
操作	Ē		描述								
筛选	Ē		在上图的模 您可以根据 测试的节点 ⑦ 说明	快1区域,ᡬ 节点名称 、 〔和暂停()〕 默认筛〕	您可以通过 节点ID、 东结)等条 选的业务日	筛选条 节点类 €件进行 期是当	€件过滤出 €型、运行 行精确筛选 首前时间的	需要查询 「日期、 」 ¹ 。 前一天。	的实例。 ⊻务日期 ∃	调度资源组、	我的节点、我今天
任务	列表		在上图的模	块2区域,↓	以列表方式	展示满	捉过滤条	件的测试	实例。		

操作	描述
	 在上图的模块3区域,您可以对测试实例进行相关操作: DAG图: DAG图为您展示测试实例的上下游依赖关系,您可以在DAG图中进行相关操作。详情请参见:DAG图功能介绍。 去诊断(运行诊断):用于对任务进行全链路分析,当任务运行不符合预期时,您可以使用该功能定位问题。详情请参见智能诊断。 重跑:用于对运行成功或运行失败的任务进行重跑操作,当任务执行成功后可以触发下游未运行状态任务的调度。常用于处理出错节点和漏跑节点。
	⑦ 说明 仅支持重跑成功和失败状态的任务。
	 更多: 重跑下游:用于对运行成功或运行失败的任务进行重跑下游节点的操作,您可以自行选择需要 重跑的下游任务,任务执行成功后可以触发下游未运行状态任务的调度。常用于处理数据修复。
	⑦ 说明 仅支持重跑成功和失败状态的任务,如果勾选了其它状态的任务,页面会提示已选节点中包含不符合运行条件的节点,并禁止提交运行。
	 置成功:用于将运行失败的任务设置为成功状态。当任务执行失败,但希望该任务不阻塞下游任务的运行时,可以使用此功能。常用于处理出错节点。 终止运行:用于终止不需要运行的任务,任务终止运行后,将失败退出。
操作	⑦ 说明 仅支持对状态为等待时间、等待资源和运行中的实例进行终止运行操作。
	 暂停(冻结):当任务及其下游任务一段时间内不需要运行时,可选择使用此功能。详情请参见:暂停调度运行。
	⑦ 说明 请勿操作projectname_root节点,该节点为工作空间根节点。周期任务的实例均依赖于该节点,如果冻结该节点,周期任务实例将无法运行。
	 • 恢复(解冻):用于将冻结状态的实例解冻。 ■ 如果该实例还未运行,则上游任务运行完毕后,会自动运行。 ■ 如果上游任务都运行完毕,则该任务会直接被置为失败,需要手动重跑后,实例才会正常运行。
	⑦ 说明 解冻操作仅作用于该实例,如果您周期任务仍是冻结状态,那么第二天生成的 实例仍会是冻结状态。
	 查看血缘:查看当前实例的血缘信息。 查看节点详情:查看当前实例的基本信息。 查看运行日志:查看当前实例的运行日志。
批量操作	在上图的模块4区域,您可以批量选择任务,进行 终止运行、重跑、置成功、暂停(冻结) 和 恢复 (解冻) 等操作。
每页条数展 示	在上图的模块5区域,您可以修改每页展示的条数,最大支持每页展示200条。

测试实例DAG图

在列表模式下单击测试实例名称或操作栏中的**DAG图**,即可打开该实例的DAG图。您可以在DAG图中,右键 单击实例进行相关操作。



查看实例详情

在DAG图模式下,单击对应实例,在右下角弹出的弹出中单击**展开详情**,将为您弹出实例详情页面,您可以 查看当前实例的属性、上下文、运行日志、操作日志、代码等详细信息。

	✓ xc_oss_数据同步	
-	离线同步	
属性 上下文 运行日志	操作日志 代码	* *
名称: .ko_demo_dem	节点D:	实例ID:
节点类型: 虚拟节点	责任人:	所属工作空间:
调度类型: 正常任务	调度周期: 日调度	调度资源组: 公共调度资源组
任务状态: 运行成功	定时时间: 2022-04-07 00:11:00	开始等待资源: 2022-04-07 16:12:13
等待资源耗时: -	运行时长: Os	开始运行时间: 2022-04-07 16:12:13
结束时间: 2022-04-07 16:12:13	实例状态: 虚拟类型实例	Region :
引擎类型:	引擎实例:	出错是否重试: 否 ▼
4		• • •

功能	描述
	您可以在此页签查看任务在生产环境的调度属性。界面参数介绍详情请参见: <mark>配置资</mark> <mark>源属性</mark> 。
属性	 ⑦ 说明 ● 任务状态解读:任务状态与任务运行有关,若任务处于未运行、等待时间、等待资源、冻结等状态时,您可以参考任务运行条件进行排查。 ● 等待资源耗时:当任务出现长时间等待资源的情况,您可以使用智能诊断功能定位当前任务运行时,哪些任务在占用资源,快速找到异常任务进行排查。 ● 运行时长过长:当任务运行时间与历史运行情况相比明显变长,您可参考以下场景进行处理。 ● 非同步任务:可答询对应引擎负责人。 ● 离线同步任务:可能任务某个阶段执行较慢或处于长时间等待资源状态。详情请参见:如何排查离线同步任务运行时间长的问题?。

功能	描述
上下文	为您展示当前任务的所有输入参数和输出参数,上下文的功能介绍详情请参见:配置 节点上下文。
运行日志	用于查看任务详细执行过程。
操作日志	用于查看任务或实例的操作记录(时间、操作人、具体操作)。
代码	用于查看生产环境当前任务的最新代码。若不符合预期,请确认任务最新版本是否已 经发布成功。详情请参见: <mark>发布任务</mark> 。

任务运行问题定位

如果任务运行不符合预期,您可以通过运行诊断功能,诊断任务未运行的原因或存在的问题。

3.4.6. DAG图功能介绍

运维中心DAG图支持您查看节点的上下游依赖关系,此外,还支持节点聚合、上下游分析功能,帮助您提高运 维效率。本文为您介绍DAG图的相关功能。

使用限制

仅Dat aWorks专业版及以上版本用户才可以使用DAG图的节点聚合、上游分析、下游分析功能。版本升级详情 请参见版本服务计费说明。

DAG图功能列表说明

单击相应节点后的**DAG图**,打开该节点的DAG图。您可以在DAG图中进行如下操作(本文以周期实例的DAG图 为例):

节点提表: 节点名称/节点D Q 业务日	3期: 昨天 前天 全部 2022-04-25 2022-04-25 曲	节点类型: 清选择节点类型 > 调	線鉄源組 滴透择 > 我的节点 我的出错节点	我的未完成节点 型跑过的节点 变得节点 受得节点 受得节点 展开搜索
基本信息	■ ⓒ ✿ & 戸 1048/75点D Q			3 C H Q Q 🗉 🛠 😫
C 246 (dur 0a)		guri 9548 root satha		生产环境,请谨慎操作

• 节点聚合

当周期实例的节点数量或层级过多时,您可以按照节点状态、工作空间、责任人、优先级等维度进行节点 聚合,查看指定目标维度的节点数量。方便您了解各个维度节点的数量,合理分配并执行节点。以下示例 为您展示,节点**不聚合及按优先级聚合**的显示情况。

⑦ 说明 仅实例(包括自动调度的周期实例、手动触发生成的补数据实例和测试实例)DAG图可按照 节点状态维度进行节点聚合。 • 节点不聚合时,显示如下图所示。

名称		9 % (Q © X
1000.000.000.000	零 下游分析	
100000000000000000000000000000000000000		
and the second second		
1000.000		
	i i i i i i i i i i i i i i i i i i i	
provide lines		•
	aql spark spark sql hive DJAR How DJAR Spark DMR Spark DMR Spark SQL DMR How	EMR Hive

节点按照优先级聚合时,显示如下图所示。您可以快速了解到当前节点的子节点共有6个优先级为7的节点。

名称	■ ② 点 日 市点高物/市点の Q	୯ଅ ୧େର
and Concerning	早下踏分析	
100 C		
2792-279-2011		
-		
-	共1个节点	
	(F 1 (1)	
	共合个市点	
	口 1 (6)	

● 上游分析

周期实例通常存在上下游依赖关系,如果目标节点长时间未运行,则您可以对其进行上游节点分析,在DAG 图中查看阻塞当前节点运行的上游节点,快速定位并及时处理相关运行报错,提升节点的执行效率。

? 说明

- 仅实例(包括自动调度的周期实例、手动触发生成的补数据实例和测试实例)DAG图支持进行 上游分析。
- 仅支持对未运行的节点执行**上游分析**操作。最多展示6层。如果您希望展示更多层级,请单击左 上角的**继续分析**。
- 上游分析可帮助您快速找到当前节点所依赖的父节点中仍处于未运行成功状态的节点,这些节点可能处于运行中、运行失败、等待时间等、待资源或冻结(暂停)等状态,父节点处于这些状态时将阻塞下游节点运行。

示例2_节点长时间未运行,则您可以选中该节点,单击左上角的上游分析。

节点搜索: 1628-46736 0	业务日期: 昨天 前天 全部 2021-03-08 2021-03-08 🖽	节点类型: 请选择节点类型	✓ 调度资源组 请选择 ✓ _ 我的节	点 🗌 我的出错节点
我的未完成节点	↓ 变慢节点			C 刷新 展开报
基本信息	Ⅲ ⑥ Ø & □ 节点名称/节点ID Q			C B () Q II
 ○ 2.与业务引擎交互 ■ 	▲上游分析 〒下游分析			
		0_3		
		ODPS SQL		
	\circ	2_ ODPS SQL		
	ODPS SQL	ODPS SQL	PyODPS 3	
更多 < 1/1 >				

分析结果得出,导致阻塞当前节点运行的上游节点为**数据表同步**及指标统计,如下图所示。通过该分析结果,您可以快速定位阻塞节点,并及时处理相关报错。

- ? 说明 后续操作:
 - 您可以在定位到关键实例后,再基于聚合项进行过滤,查看影响当前节点数据产出的关键实例 所在工作空间,或节点责任人等,帮助您提高运维效率。
 - 定位到阻塞当前实例运行的关键实例后,您还可以右键该实例,选择运行诊断功能快速分析当前实例未运行成功的原因。详情请参见:智能诊断。

爺 DataWorks │ 运维中心	antinenten v		ロロシャン ロシン 日本
节点搜索: •••••••••••	Q 业务日期: 昨天 前天 全部 2021-03-08 2021-	3-08 註 节点类型: 请选择节点类型 > 调度资源组 请选择	▶ 我的节点 我的出错节点
我的未完成节点重跑	拉的节点 🔄 变慢节点		C 刷新 展开搜索
基本信息	田 🕑 🕸 🗳 市 市点名称/节点ID	Q	C 9 ()
 ○ 2_与业务引擎交互 	× 取消上游分析		
	◎ min min min min min min		
	0_2	• 0_1	
	ODPS SQL	ODPS SOL	
	e >		
		ODPS SQL 附项目 ODPS S	跨项目 ODPS S
		0 2]
		ODPS SQL	
更多 < 1/1	当前展示10个节点 5层		

• 下游分析

展开DAG图时,默认展示三层,即包括当前节点在内,及当前节点的一层父节点与一层子节点。无法直观查 看节点影响范围。当周期实例的节点数量或层级过多时,您可以使用下游分析功能,通过节点状态、工作 空间、责任人、优先级等维度进行筛选,统计不同层级目标维度的节点数量,或统计所有层级目标维度的 节点总数量。 ? 说明

- 聚合状态下进行合并或分层展示:单击下游分析后,默认按照责任人聚合,统计所有层级该维度的节点总数量。
- 不聚合、成组展示:不聚合场景下,若下游节点超过10个将默认成组展示。该方式下,您可以 比较直观的看到节点的业务影响。
- 不成组展示:不聚合场景下,单击拆分组,可对成组的节点进行拆分查看,该方式下,您可以 直观的了解到所有节点的上下游关系。
- 下游分析使用分层展示下游节点分析结果时,最多展示6层。如果您希望展示更多层级,请单击左上角的继续分析。

示例对tag节点进行下游分析,结果如下:

合并:按照当前聚合维度对当前已分析的下游进行合并。非聚合状态下,将对下游节点进行成组展示。
 示例使用合并方式,并按照所属工作空间聚合,将所有下游节点合并至同一层,展示所属不同工作空间的节点数量。

L ALRO P M		C 副新 展开搜测
₩ <mark>@</mark> & ¤	节点各称/节点0 Q	୯୫ ଜେଇଅ
× 取消下游分析	(合并) 分照	
	O tag	
	COPP SQL	
	共7个节点	
	(a) (b) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c	1)
	(c)	
分层:按照当前聚合维度对当前已分析的下游进行分层展示。示例使用分层方式,并按照所属工作空间
 聚合,在不同层级展示不同工作空间的节点数量。

■ 「「「P DataWorks」 运進中心 ● 援索: 市点名称/市点D ※※ □ 过期市点	Q 节点类型: 请选择节点类型	✓ 责任人: 请	选择责任人 💙	调度资源组 请选择	 <i>d</i>² DataStudio <i>d d</i> <i>v</i> <i>1</i> 我的节点 	任务发布 Q. 输入搜索内容 今日修改的节点 1	♀ 2 重停 (冻结)节点	 3 Q
9								C 刷新 展开搜索
▲ ● 2 ■ 节点名称/*	D Q						CB	④ ⊝ Ⅲ 单 👯
 * 取消下部分析 会井 ひ 	3) 現代予約者 (2)	Щ С О С О С О	2/194 2/194 2/194 2/194 0 4/195 0 4/195 0 0 4/195 0 0 4/195 0 0 4/195 0 0 4/195 0 0 4/195 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0		(1)	© (1)		

● 选择DAG图的显示样式。

您可以根据实际需要,单击DAG图所在界面右上角的功能图标,调整DAG图的显示样式。例如,全屏展示、适配屏幕展示等。

示例对0_2节点的下游节点,展示拆分组及成组的DAG图。

• 拆分组的DAG图如下所示。通过该方式,您可以直观的了解到所有节点的上下游关系。

搜索: 节点名称/节点ID Q	节点类型;	请洗择节点类型	✓ 责任人:	: 请选择责任人	∨ 调度资源	组请洗择	▼ □ 我	的节点 今日	修改的节点	暂停(冻结)节点	1 孤立节点	点
	i mota	Control 1 - Marcala	, starte	- maarroomer c	110000					1117 (NOPE) 111		
											C 刷新 展3	开搜索
● A P 节点名称/节点	id Q									CB	ା ଚ ଚ ୮	1
× 取消下游分析 继续分析 0					1.000.00							
		-			COPS SQL							
				0.254.0								
				COPS SQL								
			*			*						
			0_3			L MALVER MAN	ан ал ан ал ан					
			0_31				3.					
		MRI I OPS SO.	0_31 ####################################	2	Real Loops So.		Haman	NUE I COPE SQ.				
		NRE LOOPS SOL	0_31		MRIE COPS SQL		аналананананананананананананананананана	NVER LOOPS SQL				
		Part Balling	0_31 00P5 50.		NOTE COPE SQ.		8 80.	NGE LOOPS SQ.				
		REEL COPE SO.	0.3*	2.11	JRABI (GOPS SQ.	1.1 COP4 BC	а. 2. На полодина в 602.	Note Loope so.				
		MILE COPIE SOL	0.31		Heat (con so.		а. А. В 602.	Pate I core so.				
			0.3	2.1144141	Head Constant		4 1	NAIS (COM NO.				
		Para and a second secon	0.3				4 8 9 602	Rea (604 60.				
			0.3 0.0 5 50.				а. Аналана 19 бод.	Rea (004 60.	9			
		Hatel core so.	2.09 50. .009 50. 					Nata I cove los.	9			

○ **成组**的DAG图如下所示。通过该方式,以每行呈现5个节点,有序的为您展示下游节点,方便您快速获取 下游节点的总数量。

搜索: 节点名和 过期节点	你/节点ID Q 节点类型: 请选	择节点类型 > 责任人: 请选	择责任人 > 调度资源组	请选择 > _ 我的节点	今日修改的节点 暂停(冻结	志) 节点 🗌 孤立节点
						€ 刷新 展开批
III 🍄 😃	中 市点名称/节点ID Q					୯ 🖪 ୧୧ ଅ
× 取消下游分	分析 继续分析 🛛 合井 分层					
			LIBBAR			
			ODPS SQL			
			0_2			
Ø			OUP'S SOL			
	共13个节点					
	stars TREES or a project a	Hard have always provide	Red also bee also area.	Real alors into alorse rare	1.00.20000000000	1
	跨项目 ODPS SQL	ODPS SQL	ODPS SQL	ODPS Script	ODPS SQL	
	ELECTRONIC CONT.	2	Real discontraction stars	Best star into stars over	toot	
	跨项目 ODPS SQL	ODPS SQL	ODPS SQL	PyODPS 3	ODPS SQL	
	· · · · · · · · · · · · · · · · · · ·	1. DOTED AND	B.S. S. Street B.S. Science Street			
	跨项目 ODPS SQL	ODPS SQL	跨项目 ODPS SQL			
	· · · · · · · · · · · · · · · · · · ·					

 DAG图为您展示节点上下游依赖关系,其中实线为同周期依赖,虚线为跨周期依赖。还支持您右键单击目标 节点,执行相关操作。右键单击目标节点,执行相关操作。下图以周期实例为例。

≡	🍿 DataWorks 运维	心	• •		
e	运维大屏				
\$\$\$	实时任务运维 🖌 🖌	节点搜索: 节点名称/节	5点ID Q 业务日期:	昨天 前天 全部 2020	-11-25 - 2020-11-25 🏥
t	周期任务运维				
	周期任务	基本信息		A	
	周期实例	0		*	
	补数据实例	#22		*	▼ 跨项目
	测试实例	\odot		(⊙ 13	Start
ର	手动任务运维	#32			展开子节点
᠕	智能监控	#32			查看运行日志
ራ	智能诊断	#32			运行诊断 New
		#32			全 君代码 编辑节点
					查看血缘
		#32	- 000100		查看更多详情
					终止运行
		#32		« »	重跑下游
		Θ			置成功
		#32	- 00/80		· 采跑
		⊘ #7¢			暂停(冻结)
		#/(s)		-	恢复 (解冻)

操作	描述
展开父节点/子节 点	当一个业务流程有3个及以上节点时,运维中心展示节点时会自动隐藏节点。您可以通过展开 父子层级查看全部节点的内容。
查看运行日志	查看当前实例正在运行、成功、失败等状态的运行日志。 ⑦ 说明 。 公共调度资源组实例保留一个月,日志保留一周。 。 独享调度资源组实例、日志保留一个月。
运行诊断	DataWorks运维中心为您提供运行诊断功能,帮助您掌握节点运行的全链路信息,以迅速定 位问题。详情请参见 <mark>智能诊断</mark> 。
查看代码	查看当前实例的代码。
编辑节点	单击后,即可跳转至数据开发页面,对当前节点的内容进行修改。
查看血缘	查看当前实例的血缘关系。
查看更多详情	单击后,即可查看当前节点的 属性、上下文、运行日志、操作日志 和 代码 等信息。
终止运行	仅等待运行、运行中状态的实例可以进行终止运行的操作。进行此操作后,该实例将为失败 状态。
重跑	可以重跑某节点,节点执行成功后可以触发下游未运行状态节点的调度。常用于处理出错节 点和漏跑节点。 ⑦ 说明 仅支持重跑未运行、成功和失败状态的节点。
重跑下游	可以重跑某节点及其下游节点,需要您自定义勾选,勾选的节点将被重跑,节点执行成功后 可以触发下游未运行状态节点的调度。常用于处理数据修复。 ⑦ 说明 仅支持勾选未运行、完成和失败状态的节点,如果勾选了其它状态的节 点,页面会提示已选节点中包含不符合运行条件的节点,并禁止提交运行。
置成功	将当前实例的状态改为成功,并运行下游未运行状态的节点。常用于处理出错节点。 ⑦ 说明 仅支持将状态为失败的节点置成功,不能置成功业务流程。
续跑	节点执行失败后,可以续跑此节点。仅MaxCompute类型节点支持该功能。 ⑦ 说明 2021年1月之前购买的独享调度资源组请提工单联系我们升级。

操作	描述						
紧急操作	当前实例在非常紧急的情况下的操作,紧急操作只对当前节点本次有效。 选择 去除依赖 ,即可解除当前节点的依赖关系。常用于上游失败并与此实例没有数据关系 时,						
暂停(冻结)	周期实例中的冻结只针对当前实例,且正在运行中的实例。冻结的周期实例不自动调度运行 (不会真实跑数据),并且会阻塞它的下游节点执行(下游节点都不会自动调度)。 场景示例:如果您小时节点每小时调度一次,当天生成24个小时实例,如果其中一个小时实 例不需要执行,您可以选择冻结该实例,冻结其中一个小时实例不会影响其他实例执行。						
恢复(解冻)	可以将冻结状态的实例解冻。 • 如果该实例还未运行,则上游节点运行完毕后,会自动运行。 • 如果上游节点都运行完毕,则该节点会直接被置为失败。需手动重跑后,方会正常运行。 ⑦ 说明 解冻操作仅作用于该实例,如果您周期任务仍是冻结状态,那么第二天的 实例仍会是冻结状态。						

3.4.7. 强制重跑下游

DataWorks支持强制重跑下游功能,用于强制重新调度指定的节点及其下游节点,并支持通过工作空间过滤要 重跑的下游节点。本文为您介绍强制重跑下游功能的注意事项和操作步骤。

使用限制

- 仅DataWorks专业版及更高版本可以使用强制重跑下游功能。低版本用户请先升级再使用该功能,详情请参见版本服务计费说明。
- 仅工作空间管理员、租户管理员、阿里云主账号可以发起强制重跑下游的操作。
- 通过工作空间过滤要重跑的下游节点时, 仅支持过滤选择当前用户为管理员的工作空间。
- 仅业务日期为昨天和前天,并且状态为运行成功、运行失败的周期实例支持强制重跑下游的操作。
 实例不满足上述条件时,强制重跑下游功能入口会被置灰。

• 每个租户在同一个地域(Region)、同一时间下, 仅支持执行一次强制重跑下游的操作。

注意事项

- 强制重跑下游会将选中的所有下游节点全都置为未运行状态,如果下游节点在运行中,将会被强制停止。
- 当节点的重跑属性配置为运行成功或失败后不可重跑,执行强制重跑下游操作会导致重跑运行失败。如果您需要修改重跑属性,详情请参见:重跑说明。

操作步骤

- 选择目标实例,进入强制重跑下游页面。
 右键单击目标实例(业务日期为昨天和前天,并且状态为运行成功或运行失败的周期实例),选择紧急操作>强制重跑下游。弹出强制重跑下游配置页面。
- 2. 配置强制重跑下游。
 - i. 基本设置。

强制重跑下游 ×							
▲ 该操作会将所有下游节点置为未完成并重新开始调度。请务必明确影响并知会相关人员。							
基本	基本设置选择工作空间 确认影响						
* 业务日期:	2022-01-04 2022-01-03						
* 起始节点:	序号	节点名称		节点ID	责任人	工作空间	操作
	1	X(这 _一	发			고려 고려 고려 주易	
	请输入节	点ID,如果有多个	,可以用;分隔				添加
* 重跑起点:	○ 起始寺	5点 💿 起始节	点的一级子节点				
* 重跑跨天实例:	○是(• 否					
						下一步	取消
参数		描	述				
业务日期 选择需要强制重跑的周期实例所在的业务日期。仅支持对业务日期为昨 天和前天的周期实例发起强制重跑下游的操作。						日期为昨	
起始节点		选	择当前工作空	² 间下需要发;	起强制重	跑下游操作起始节点。	
重跑起点 选择从起始节点开始执行还是从起始节点的下游节点(一级子节点)开 始执行。					5点)开		



ii. 单击**下一步**,确认信息无误后,单击**确定**完成基础信息设置。

 ↓ 注意 单击下一步后将无法修改当前页面的信息,请务必确认信息无误以节省计算开销并缩 短您的等待时间。 iii. 选择工作空间,通过工作空间过滤需要重跑的下游节点。

在左侧**下游节点所在的工作空间**列表框中选择需要强制重跑的下游节点所在的工作空间,单击 添加至右侧列表框中。选中的工作空间内的下游节点将会被强制重跑。未选中的将会空跑。

强制重跑下游					×
▲ 该操作会将所有下游节点置为未完成并重新开始调度。请务	必明确影响并	谨慎操作。			
基本设置	选择工作3	间	> ñ	角认影响	
下游节点总数:25 您可以选择强制重跑指定工作空间中的下游节点。					^
下游节点所在的工作空间		要重跑的节点所在的	的工作空间		
项目筛选 Q		项目筛选		Q	
	>				
	<		Not Found		
- 1/2 项		0项			
				下一步	▼

? 说明

通过工作空间过滤要重跑的下游节点时, 仅支持过滤选择当前用户为管理员的工作空间。

iv. 单击**下一步**,确认信息无误后,单击**确定**完成基础信息设置。

 ↓ 注意 单击下一步后将无法修改当前页面的信息,请务必确认信息无误以节省计算开销并缩 短您的等待时间。

v. 确认影响。

强制重跑下游				×
▲ 法撮作 △ 按 댮 右 下)	发生白垩为中学成并重新工作调度	法久心明确影响并谋植竭作		
欧洲正本村加田	日に、日本である。	。调力20吋吨家門开建模床IF。		
基2	本设置	选择工作空间	确认影响	
影响节点数:	1			
影响基线:	基线名称	优先级	责任人	
		没有数据		
* 强制重跑下游原因:				
			确定	取消
参数		描述		
影响的节点数		为您展示要进行	强制重跑的节点数。	
影响基线		列表中为您展示」 出。基线监控详 ⁴	此次操作将影响哪些基线数捷 情请参见: <mark>智能基线概述</mark> 。	的产
强制重跑下游原	因	输入需要进行强	制重跑下游的原因。	

vi. 单击确定,系统将根据您的设置,将选中节点的状态全部置为未运行,并重新开始调度。

场景示例

- 场景描述: 假设对A节点发起强制重跑下游的操作,选择重跑A工作空间,不重跑B工作空间的节点。
- 运行详情: B工作空间下的节点B将不会重跑, A工作空间下的C节点会被强制重跑。

Sq Node_A	<mark>工作空间A</mark>
Sq Node_B	工作空间B
Sq Node_C	<mark>工作空间A</mark>

后续步骤

您可以在操作历史中查看操作详情。

3.5. 手动任务运维

3.5.1. 运行并管理手动任务

手动业务流程发布至生产环境后您可以在该页面手动触发任务运行。支持您通过列表或DAG图方式,快速查看 任务的详细信息并执行相关操作。

? 说明

- 手动任务提交至调度系统后,不会自动运行,只有手动触发才会运行。详情请参见:管理手动业 务流程。
- 目前DataWorks V1.0创建的手动任务显示在**手动任务**选项下, DataWorks V2.0及以上版本创建的 手动任务显示在**手动业务流程**选项下。

进入手动任务页面

- 1. 登录DataWorks控制台。
- 2. 在左侧导航栏,单击工作空间列表。
- 3. 单击相应工作空间后的进入运维中心。
- 4. 在左侧导航栏,单击**手动任务运维 > 手动任务**,查看手动业务流程、DAG图。

手动业务流程列表

手动业务流程列表为您展示已发布的手动任务流程。

≡	🏟 DataWorks	运維	中心 生产1		•				ć	P DataStudio 🖌 任务发布 🛛 🔮 😨 💐	. 💎 📰 i
¢	运维大屏		JILder 2004		na man	0	MR/T 1 . 2002/32/06/11 1		4		
\$\$\$	实时任务运维	~	11950001	3235 DIGE 12 10		ų	HIGHERICA				で周囲
а	周期任务运维	~		名称		节点D	修改日期 1	任务类型	责任人	操作	
ନ	手动任务运维	^				-	2022-01-10 15:08:44	手动业务流程		DAG图 运行 查看实例 3	
	手动任务		-								
	于705-09 2989/51/5										
~~ □	(1985)王 梁源云 像	Ť									
	引擎运维	~									
Ų,	智能诊断										
8	慶仲历史		修改责	化三人 一 傳改 職業 國際相	下线手动业务》	az 3				< 1/1 > 報页显示	: 40 🗸
ł	喿作		ł	描述							
ĵ	帝选		F t	在上图的模块 您可以根据 业 ⑦ 说明 方可展示。	1区域 <i>,</i> 务流程 任务行	您可以: 名称 、 ⁼ 名搜索的:	通过筛选条件 节点ID、责任 结果会受到身	+过滤出需要查询的任约 E人、我的节点和今日 其它筛选条件的影响,	务。 修改的节点 只有同时满足	等条件进行精确筛〕 所有筛选条件的结	先。
0	DAG图		1	单击操作栏中 血缘等信息。	的 DAG 详情请	图 ,即回 参见: 手	可打开当前业 □ <mark>□动业务流程</mark> Ⅰ	务流程的DAG图。您回 DAG图。	「以在DAG图中	中查看业务流程的代	词、

操作	描述							
运行	 单击操作栏中的运行,在弹出的对话框中配置参数后,单击运行即可按照运行范围运行手动业务流程,产生手动实例。 运行 ▲ ************************************							
查看实例	单击操作栏中的 查看实例 ,即可跳转至 手动实例 页面,查看手动业务流程的运行结果。详情请参 见:查看并管理手动实例。							
更多	单击操作栏中的 更多 ,可以进行 修改责任人、修改调度资源组、 下线手动业务流程 等更多操作。							
批量操作	在上图的模块3区域,您可以批量选择任务,进行 修改责任人、修改调度资源组、下线手动业务流 程的操作。							

手动业务流程DAG图

单击业务流程名称或操作栏中的**DAG图**,即可打开当前业务流程的DAG图。您可以在DAG图中,右键单击业务流程进行相关操作。

and the Children of the Childr	
(※) 実时任务运命 、 ユニラのWills コニラのWillsコミット く Differ Difference	
□ 周期任务运業 → ○ 各称 节点	(19)1月11日 (19)1月11日 (19)1日 (19)11日 (19)1日 (19)1日 (19)1日 (19)11日 (19)1日 (19)11日 (19)1101 (19)11001 (19)100000000000000000000000000000000000
	生产环境,请谨慎操作
手动任务	
F-2724-09	
の「西部学生	
 < , <li< th=""><th>运行 重雪时以马 确定节点 当雪血承 快改调变词原目</th></li<>	运行 重雪时以马 确定节点 当雪血承 快改调变词原目
	病1个节点,1层
操作	描述
运行	单击运行,在弹出的对话框中配置参数后,单击运行即可按照运行范围运行 求动业务流程,产生手动实例。 运行 (************************************

操作	描述
查看代码	查看当前节点的代码。
编辑节点	单击后可跳转至 数据开发 页面,对业务流程内容进行修改。
查看血缘	查看当前业务流程的血缘关系图。
修改调度资源组	修改当前业务流程所在的资源组。

3.5.2. 查看并管理手动实例

手动实例是手动任务运行后产生的实例。您可以通过列表及DAG图的方式,快速查看实例的详细信息并执行相关操作。

< ↓ 注意

- 目前DataWorks V1.0创建的手动任务运行生成的实例显示在**手动实例**选项下, DataWorks V2.0及 以上版本创建的手动任务运行生成的实例显示在**手动业务流程**选项下。
- 目前仅支持对周期实例运行状态设置监控报警,手动实例、补数据实例和测试实例均不支持设置 任务运行状态的监控报警。
- 您需要购买DataWorks专业版及以上版本,手动实例即可支持运行诊断功能,详情请参见智能诊断。

进入手动实例页面

- 1. 登录DataWorks控制台。
- 2. 在左侧导航栏,单击工作空间列表。
- 3. 单击相应工作空间后的进入运维中心。
- 4. 在左侧导航栏,单击手动任务运维 > 手动实例,查看手动实例的列表、DAG图详情。

手动实例列表

≡	ntaWork	s I 运	集中心 <mark>4</mark>	2 ~1 1	R.	~				& DataStudio 🛛 任务发布 🛛 🗍	ग <i>थ</i> , 💎 💴 :
¢	运维大屏		类型:	手动业务流程 🗸	搜索	Q	责任人: 请选择责任人	✓ 业务日期: 2022·01·09	(1) 运行日期: 请选择运行日期		
\$ \$\$	实时任务运维	~									C周新
а	周期任务运维	~		基本信息			责任人	业务日期 11	结束时间↓	操作	
ନ	手动任务运维	^		1.1				2022-01-00-00-00-00			
	手动任务			10000.0000				2022-01-09-00.00.00		DAGELI HURSSIL	
	手动实例										
৵	智能监控	~									
	资款运维										
۲	引擎运维	~									
ų,	智能诊断										
ð	操作历史										
_			Ħ							< 1/1 > 4	岐島示: 40 ~

操作	说明
筛选	在上图的模块1区域,您可以通过筛选条件过滤需要查询的任务。 您可以根据 类型、业务流程名称、责任人、业务日期 和 运行日期 等条件进行精确筛 选。
DAG图	在上图的模块2区域,单击操作栏中的 DAG图 ,即可打开当前业务流程的DAG图。查看实 例运行的结果。详情请参见: <mark>手动实例DAG图</mark> 。
批量终止	如果实例在运行中,可以单击 批量终止 ,停止任务运行。

手动实例DAG图

单击业务流程名称或操作栏中的**DAG图**,即可打开当前业务流程的DAG图。您可以在DAG图中,右键单击业 务流程进行相关操作。

类型:	手动业务流程 > 搜索:	业务流程名称	F Q 责任人:	请选择责任人 🗸 🖌	业务日期:	2022-01-09	Ħ	运行日期: 请选择运行日期	CR997
	基本信息	1	节点名称/节点ID	Q					
	***								生产环境,请谨慎操作
		•				¢		 ・豊香進行日志 ・通行砂断 (100) ・豊香(103) ・価値) ・価値 ・価値	
	ع د ۱/۱) 計	前展示1个节点,1层						

⑦ 说明 在手动业务流程列表中,点击某个手动业务流程后,右侧DAG图为您展示该业务流程的DAG。 可能包含多个任务。

操作	说明			
	查看当前实例正在运行、成功、失败等状态的运行日志。 ⑦ 说明			
查看运行日志	 不支持手动删除实例,实例在过期后(30天左右)平台将自动删除。 运行在公共调度资源组上的任务,其实例保留一个月,日志保留一周。当运行完成的实例日志大于3M时,每天会定时进行清理。 运行在独享调度资源组上的任务,其实例和日志会保留一个月。 			

操作	说明			
	Dat <i>a</i> Works运维中心为您提供运行诊断功能,帮助您掌握任务运行的全链路信息,以迅速定位 问题。详情请参见 <mark>智能诊断</mark> 。			
运行诊断	€〕注意			
杏看代码	杏看当前节占的代码			
编辑节点	单击后,即可跳转至 数据开发 页面,对当前业务流程的内容进行修改。			
查看血缘	查看当前实例的血缘关系。			
终止运行	终止任务的运行,仅对当前实例有效。			
	可以重跑指定任务,常用于处理出错节点和漏跑节点。			
重跑	⑦ 说明 仅支持重跑成功和失败状态的任务。			

3.6. 智能监控

3.6.1. 智能监控概述

智能监控功能支持您通过配置自定义规则,实现对任务运行状态及资源使用情况的监控;通过配置智能基线, 保障复杂依赖场景下重要数据在预期时间内正常产出。同时,您还可以根据业务需求自定义资源组运维规则, 实现资源组的自动化运维。

功能介绍

各模块功能介绍如下:

功能	描述
智能基线	智能基线能够及时捕捉导致基线上任务无法按时完成的异常情况并提前预警,保障 复杂依赖场景下重要数据能在预期时间内顺利产出。 • 创建及管理基线,详情请参见:基线管理。 • 查看基线实例运行情况,详情请参见:基线实例。 • 查看事件报警信息,详情请参见:事件管理。

功能	描述
规则管理	 您可以在规则管理界面管理全局规则,同时,还支持您根据业务需求自定义监控规则。 全局规则:系统内置全局报警规则(非空间级别报警规则)对周期任务的状态进行监控,避免周期任务异常导致周期实例无法正常调度执行,详情请参见:周期任务状态监控。 自定义监控规则: 周期任务周期性调度时将产生周期实例,DataWorks支持通过周期任务配置自定义监控规则,实现对周期实例运行状态的监控。详情请参见:周期实例状态监控。 支持您创建自定义规则并指定规则对象为实时计算任务,实现对实时计算任务运行状态的监控,详情请参见:实时计算任务运行状态监控。 支持您创建自定义规则并制定对象类型为独享数据集成资源组或独享调度资源组,实现对独享资源组使用情况的监控,详情请参见:独享资源使用监控。
报警信息	您可以在 报警信息 界面查看监控报警信息,包括自定义规则报警信息、全局规则 报警信息、智能基线报警信息等。
值班表	您可以在值班表界面自定义值班表信息,以便配置监控规则报警方式时可以选择发送报警信息给值班表对应的值班人。 ⑦ 说明 值班表支持设置主值班人和备值班人,默认报警给主值班人员, 但是当报警两次后,第三次及之后的报警信息将同时报警给主、备值班人员。
自动运维	您可以根据业务需求,对独享资源组创建运维规则并关联已创建的监控规则,当关 联的监控规则被触发时,系统将对运行在目标独享资源组上且满足 过滤条件 的任 务自动执行运维动作,以达到自动运维的目的。 ⑦ 说明 目前只支持关联调度资源的监控规则。

任务运行状态监控

监控对象

您可以通过配置相应的监控规则(自定义规则、智能基线),对周期调度任务状态、周期任务自动调度产生的 周期实例运行状态、实时计算任务运行状态、资源使用情况进行监控。

• 周期任务状态监控

DataWorks每晚根据**周期任务**生成第二天待自动调度运行的**周期实例**,所以为保障周期任务可以正常产生 周期实例并且自动调度运行,DataWorks内置了全局报警规则(非空间级别报警规则)来对**周期任务**的状 态进行监控,如有异常便会自动报警。报警包括**孤立节点与节点成环**。

规则类型	监控对象	触发条件	报警说明
------	------	------	------

Dat aWorks公共云合集·运维中心

讨象	触发条件	报警说明
	孤立节点产生后会自动报警,如 果收到孤立节点报警请及时处 理。	◎ DataWorks每天定时9点、12
ち点:指任务展开 気时,没有依赖任 ち点。	⑦ 说明 在DataWorks 上,除了工作空间根节点 外,自定义的每个周期调度 的任务都需要有父节点依赖 才可以被正常调度运行。所 以孤立节点不会自动调度运 行,如果该孤立节点下游依 赖较多,则会造成严重的后 果。	点、16点对周期任务状态进行 扫描,如有工作空间内存在孤 立节点或节点成环等异常现 象,系统将会自动发送报警。 但扫描时间点前10分钟内产生 异常不会纳入本次扫描,该异 常将会被纳入下一个周期的任 务状态扫描中进行扫描。 • 全局规则为系统内置规则,您 无须手动新建,默认以短信、 邮件的方式报警给节点责任
むな・指す任を为	节点成环后会自动报警,如果收 到节点成环报警请及时处理。	人。但您可以在 规则管理 页面 中针对全局规则修改报警接收 人。
下点成坏:指呆任务力 上游任务但又同时依赖 了自己的下游任务,导 致依赖关系成环。	⑦ 说明 节点成环后在 自动调度时不会被调度调 起。	 ◆ 支持您在规则管理页面关闭全 局报警规则。
	京点:指任务展开 话时,没有依赖任 点。 斯王:目时依赖 务为赖 务下派成环。	密触发条件孤立节点产生后会自动报警,如 果收到孤立节点报警请及时处理。流点:指任务展开 (时, 沒有依赖任? 说明 在DataWorks 上,除了工作空间根节点 外,自定义的每个周期调度 的任务都需要有父节点依赖 才可以被正常调度运行。所 以孤立节点不会自动调度运行,如果该孤立节点下游依 赖较多,则会造成严重的后 果。近: 許点成环后会自动报警,如果收 到节点成环报警请及时处理。近 说明 节点成环后在 自动调度时不会被调度调 起。

• 周期实例运行状态监控

在DataWorks上,周期任务周期性调度时将产生周期实例,DataWorks支持通过**周期任务**配置自定义监控规则,实现对**周期实例**运行状态的监控,包括指定对象的自定义规则监控报警和重要任务提前预警的基线预警功能。

规则类型	监控对象	触发条件
自定义规则	指定 任务节点、基线、工作空 间、业务流程上的任务。	 当任务运行完成、未完成、出错、周期未完成、超时、自动重跑后仍出错时,将触发报警。 当对象类型为工作空间时,除上述触发条件外,还包括任务转实例完成、实例数量波动。

规则类型	监控对象	触发条件				
	基线任务及影响基线上任务数据产 出的上游任务会被纳入基线监控范 围。	○ 基线报警:				
智能基线	 ⑦ 说明 • 您可以通过设置基线优先级来保障基线上的任务按时执行,数据顺利产出。 • 当有任务需要重保并且上游任务依赖较为复杂时,可以使用基线功能将重要任务移到该基线上。 	 当预测到基线上任务无法在承诺时间内完成时,系统将会根据定义的报警方式发送基线报警信息。详情请参见核心逻辑介绍:基线报警。 事件报警: 基线任务及其上游任务出错,或关键路径上的任务变慢时,会产生事件,并发送事件报警信息,详情请参见:事件管理。 				

- 实时计算任务运行状态监控
 - 规则类型: 自定义规则。
 - 监控对象:实时计算任务。
 - 触发条件:当任务出错时触发报警。
- 资源使用情况监控
 - 规则类型:自定义规则。
 - 监控对象: 独享调度资源组、独享数据集成资源组。
 - ∘ 触发条件:
 - 资源组使用率大于某个数值并持续指定时长时触发报警。
 - 资源组上等待资源的实例数大于某个数值并持续指定时长时触发报警。

报警方式

配置监控规则后,平台一旦检测到规则被触发,将会根据您设置的报警方式,通过邮件、短信、电话或钉钉群 消息等形式通知您,方便您及时发现并处理异常问题。

规则类型	报警方式	疲劳度控制
------	------	-------

规则类型	报警方式	疲劳度控制
	 报警给指定人员:支持邮件、短信、电 话报警给节点责任人、指定责任人或值班 表中的排班等。 	支持最大报警次数、最小报警间隔与免打 扰时间控制。 ⑦ 说明
	⑦ 说明 您需要购买DataWorks 专业版及以上版本,才可以使用电话 告警功能。	 最大报警次数:报警的最大次数,超过设置的次数后,不再产生报警。 最小报警间隔:两次报警之间的最小时间间隔。报警间隔期间,
自定义规则、全局规 则、智能基线	 ● 报警给群组:支持钉钉群机器 人、WebHook。 	若任务恢复正常,剩余报警次数将不再触发报警。免打扰时间:设置了免打扰时间
则、目配至此	 ⑦ 说明 • 仅DataWorks企业版及旗舰 版支持使用WebHook功能。 • WebHook功能仅支持推送报 警信息至企业微信或飞书。 • 如果您需要使用其他的 WebHook,请工单联系我 们。 	后,则在该时间段内系统将不会 发送报警警。 例如,当设置了任务状态为超 时、出错、未完成时触发报 警,且该任务设置的免打扰时间 为 00:00 到 08:00 , 则该时间段内将不会发出报警信 息,如果到达8点,任务仍处于 上述异常状态,将会发出报警信 息。

资源自助运维

DataWorks的自动运维功能可以实现根据已设置好的监控规则和运维规则来自动执行任务的运维操作。

• 触发条件:关联的监控规则被触发。

? 说明

- 目前支持对资源组利用率和资源组上等待资源的实例数进行的监控。
- 。 目前仅支持对独享调度资源组进行自助运维。
- 定位目标实例:对满足实例类型、实例调度周期、实例优先级、实例状态、实例所在工作空间等过滤条件的实例进行自助运维操作。
- 运维动作:终止运行实例。

? 说明 支持单次最大终止2000个实例。

任务报警信息诊断

您可以通过任务DAG图或运行诊断页面,查看指定任务产生的报警信息。

● 通过任务DAG图查看

对于配置了监控报警且在当前时间点24小时内产生了报警的实例,您可以在周期实例页面,打开指定实例的DAG图,单击实例右上角红色告警标识(图中区域②),在弹出的监控详情窗口为您展示将当前任务纳入监控的规则或基线列表,以及各个规则或基线的触发情况,您可以单击右上角的查看报警信息跳转至报警信息页面查看报警详情,或单击规则/基线名称跳转至该规则的配置页面,查看规则配置详情。

≡ 🏟 DataWorks	运维中心 和MARK	Д II & max 🗣 d 📃 👘
 ➡ 运维大屏 ∭ 实时任务运维 → □ 周期任务运维 → 	17点開発 17点目的/17点の Q 並身目期 約天 10万 全部 2022-05-21 2022-05-21 [] 17点発起 前品目で点外型 > 保護出界機 前品目 、 別の未知点で点 単現は切り示 全球であ はあ2000年まで知りてあ。	
周期任务 周期实例		∅ С ┺ Q Q 回 ③ ∷ 生产环境, 请谨慎操作
+ 政策(広例) 務議(広例) 〇 手助任务运業 ~		● 背留
	● #7C	×
 ① 引撃运集 ~ 〇 指能诊断 	2 00% SQL 规则/基结名称 2 00% SQL 规则	截发情况 未触发
♀ 操作历史	Mark I	已触发
=	夏多 < 1/1 → 当期展示2个节点,2展	田子

⑦ 说明 您可以勾选过去24h产生告警的节点(图中区域①)对周期实例进行过滤。

• 通过任务运行诊断页面查看

对于配置了监控报警的任务,您还可以在该任务运行诊断页面的提示信息区域,单击查看详情,在弹出 的监控详情窗口为您展示将当前任务纳入监控的规则或基线列表,以及各个规则或基线的触发情况,您可 以单击右上角的查看报警信息跳转至报警信息页面查看报警详情,或单击规则/基线名称跳转至该规则的 配置页面,查看规则配置详情。

≡	DataWorks	一运维	中心 P DataStudio						口 団 む max 🧌	7
e %	运维大屏 实时任务运输 回期任务运输	~	這行失敗)	5233			1	免费体验 运行诊断为专业版功能,您目前可以免测验,建议升级到专业版。	_{要体} 去升级
ti ≪	周期1155运输 手动任务运输 智能监控	ž	全经指诊状 基本信息 影响基线 	历史实例 2条,查看详情		_				
	资源运输		上游依赖		O 1	定时检查	0	周度资源	0 E\$	執行
۲	引擎运维	~	2022-05-23 14:44:33~2022-05-	initiae Ba	MaxCompute DQC	代码详惯				
Υ.	新版沙斯		23 144433 (dur 0s) Gateway: 道行失敗	日志诊断 目前无法给出课 计算资源诊断	监控详情			×		•
				(or notice of)	规则/基线名称 规则		触发情况 未触发 日本19			
ຶ	 最作历史						Bax	知 英 政		
=										

• 通过报警信息页面查看

您也可以通过报警信息页面查看智能监控模块产生的所有报警信息,并通过报警详情查看报警的完整触发 流程,包括触发该报警的监控规则、报警触发条件、您收到报警的原因等。详情请参见:报警信息。

3.6.2. 智能基线

3.6.2.1. 智能基线概述

智能基线能够及时捕捉导致基线上的任务无法按时完成的异常情况并提前预警,保障复杂依赖场景下重要数据 能在预期时间内顺利产出,帮助您降低配置成本、避免无效报警、自动监控所有重要任务。

应用场景

• 管理任务优先级

在任务数量越来越多,而资源有限的情况下,当发生资源抢占时,您可以将重要任务添加至基线上,并为 基线设置较高的优先级,以保证重要的任务可以优先分配到资源。

• 计算任务预计完成时间

任务的运行受到资源和上游任务运行情况的影响,您可以将任务添加至基线上,DataWorks将计算出该任务 每天或每小时的预计完成时间,便于您查看任务的预计完成时间。

• 保障任务在承诺时间前运行完成

您可以将任务添加至基线上,并设置基线承诺时间,当系统预测到基线上任务无法在承诺时间前完成,或 上游任务有出错或变慢的情况,将发送报警信息,您可以根据报警信息及时处理问题,保障任务在承诺时 间前运行完成。

相关概念

- 基线:用户将重要任务加到基线上并设置承诺时间后,系统将根据任务运行情况计算基线任务的预计完成 时间。当系统判断基线任务可能无法在承诺时间前完成,系统将发出告警。
- 承诺时间:任务运行成功的最晚时间点。即面向数据应用,任务承诺在该时间点前完成。如果希望为运维人员预留一定的时间处理异常,您可以为基线设置预警余量,则系统会将承诺时间-预警余量 作为触发报警的预警时间,判断任务能否在预警时间前运行成功。
- 预警时间:即 承诺时间-预警余量 。
- 基线任务: 被添加到基线上的任务。
- 基线实例:系统使用基线实例计算任务每次运行的预计完成时间。基线实例的状态包括:安全、预警、破线。
 - 安全: 预计完成时间<预警时间 。
 - 预警: 预警时间<预计完成时间<承诺时间 。
 - 破线: 预计完成时间>承诺时间。
- 关键路径:影响基线任务的多条路径中,运行耗时最长的路径。
- 事件:基线任务及其上游任务出错,或关键路径上的任务变慢时,会产生事件。事件会影响基线任务的按时完成。

功能介绍

将重要任务添加到基线上后,DataWorks将根据基线的优先级保障基线任务的资源,并根据基线任务的上下游 依赖关系确定监控范围,根据该监控范围内任务的运行情况触发基线报警或事件报警。



如上图所示,智能基线的主要功能如下。

1. 新建并管理基线

您可以在基线管理页面新建和管理基线:

- 将需要重保的任务移动至基线,设置基线承诺时间等基本信息,并设置基线的报警策略(报警发送的 方式、接收人等),后续会依据基线的设置对需要监控的任务进行监控预警。
- 您也可以指定基线的优先级,基线优先级可以决定任务运行的优先程度,基线优先级越高,基线上任务的优先级也会越高。优先级高的任务,在调度资源紧张的情况下,将优先获得调度资源。

基线新建及管理的操作详情请参见基线管理。

2. 确定监控范围

DataWorks会结合基线上任务的依赖关系确定需要监控的范围,将可能会影响基线数据产出的所有任务都 进行监控。详情请参见下文的核心逻辑介绍:监控范围。

- 3. 触发报警并发送报警信息
 - 基线报警
 - a. 推算时间

DataWorks会根据基线设置的承诺时间,结合基线上任务一段时间内的平均运行时长,推算出基线 上任务的最晚开始时间和最晚完成时间,并逐步推算出监控范围内所有任务的相关时间。详情请 参见核心逻辑介绍:基线报警。

进行时间推算所需的计算资源由基线实例提供,开启状态的基线每天会生成基线实例,用于推算 监控范围内所有任务每次运行的相关时间。您可以在基线实例页面查看基线实例列表和运行情况,详情请参见:基线实例。

b. 触发报警

DataWorks会根据基线配置的报警策略、推算出的时间,结合实际监控范围内任务的运行情况,自动触发报警并发送给报警接收人。当预测到基线上任务无法在承诺时间内完成时,系统将会根据 定义的报警方式发送基线报警信息,详情请参见核心逻辑介绍:基线报警。

○ 事件报警

监控范围确定后,当基线任务及其上游任务出错,或关键路径上的任务变慢时,将会产生相应的事件,并发送事件报警信息,您可以在DataWorks的事件管理页面查看当前已存在的事件列表,详情请参见事件管理。

费用说明

- 基线实例个数:所有开启状态的基线都会产生基线实例,DataWorks根据当天23:59产生的基线实例数量收 取费用。详情请参见:基线实例计费说明。
- 报警短信和报警电话个数:基线报警将产生短信和电话费用,详情请参见:报警短信与报警电话计费。

使用限制

仅DataWorks标准版及以上版本支持使用智能基线功能。低版本用户请先升级再使用该功能,详情请参见: DataWorks各版本详解。

核心逻辑介绍:监控范围

基线创建完成后,某个任务被移动至基线后,智能基线不会监控基线上任务的所有上下游任务,对该任务相关的上下游监控范围如下:

- 上游任务:影响基线上任务数据产出的上游任务会被纳入基线监控范围。
- 下游任务:下游任务不会被纳入基线监控范围,即基线任务的下游任务及上游任务的其他分支下游任务运行出错时不会触发报警。



如上图所示,假设整个DataWorks有6个任务节点,任务D和任务E是基线上的任务,则影响任务D和任务E数据 产出的上游任务,即上图中的任务A、任务B、任务D和任务E,均会被纳入监控范围,监控范围内是我任务出 现异常(出错或变慢),均会被智能基线察觉,而任务C和任务F不受智能基线所监控。

核心逻辑介绍:基线报警

您可以将重要任务添加到基线上,设置基线承诺时间和预警余量后:

- DataWorks会将 承诺时间-预警余量 作为预警时间,结合监控范围内任务的历史平均运行时长,基线实 例依次推算得出监控范围内各任务的最晚完成时间和最晚开始运行时间。
- 任务实际运行时,当监控范围内的任务的运行状态可能会导致基线任务无法在预警时间前完成时,DataWorks会触发基线报警。

时间推算

1. 基线的预计完成时间推算

基线的**预计完成时间**是根据添加到基线的任务在历史一段时间内的平均完成时间估算得出,当基线任务 有多个时,基线的**预计完成时间**以各基线任务平均完成时间中最晚的时间为准。



推算结果:

- 基线A的预计完成时间 = 任务D历史一段时间内的平均完成时间 = 5:00
- 基线B的预计完成时间 = 任务E历史一段时间内的平均完成时间 = 6:00

⑦ 说明 基线B上有两个基线任务,其中任务E平均完成时间晚于任务F的平均完成时间,则任务 E是影响基线B最晚完成时间的任务,所以基线B的预计完成时间由任务E的平均完成时间决定。

2. 监控范围内任务的预计最晚完成时间和预计最晚开始时间推算



推算场景 示例	推算公式	推算结果示例
基线任务 最晚开始 时间推算	基线任务最晚开始时间=基线 预计完成时间 - 基线 任务平均运行时长	 任务D的最晚开始时间=基线A预计完成时间(5:00)-任务D平均运行时长(20min)=4:40 任务E的最晚开始时间=基线B预计完成时间(6:00)-任务E平均运行时长(2h)=4:00 任务F的最晚开始时间=基线B预计完成时间(6:00)-任务F平均运行时长(1.5h)=4:30
基线上游 任务时间 推算	 基线上游任务最晚完成时间=下游任务最晚开始时间 基线上游任务最晚开始时间=基线上游任务最晚完成时间-基线上游任务平均运行时长 	 任务C的最晚完成时间=任务D的最晚开始时间 任务C的最晚开始时间=任务C最晚完成时间 (5:00)-任务C平均运行时长(20min)=4:40
被纳入多 条垫范围内 的共任务用 间 指 算	多条基线任务共同的上游任务的最晚完成时间=下 游基线任务中开始运行时间最早的时间	4:00早于4:10,即任务B的最晚完成时间为4:00 时,方可同时满足基线A和基线B。进而可以推算得 出:任务B的最晚开始时间=任务B最晚完成时间 (4:00)-任务B平均运行时长(2h)=2:00。

报警机制

基线报警是对开启状态且打开告警开关的基线提供的一个报警提醒,您可以基于基线的**预计完成时间**配置基 线的**预警余量和承诺时间**。DataWorks将根据监控范围内任务历史运行情况推算出的任务预计最晚完成时 间,并结合基线上任务的实际运行情况进行监控,当预测到基线上任务无法在基线预警时间(基线承诺时间-预警余量)内完成时,系统将根据该基线定义的基线报警接收人发送基线报警信息。

⑦ 说明 预警余量和基线承诺时间设置不合理可能导致报警不符合预期,详情请参见:合理配置基线承 诺时间和预警余量。



• 任务运行前,基线报警策略:

⑦ 说明 基线在每日任务运行前,通过计算当日基线监控范围内任务历史一段时间内的平均完成时间,将无法在基线预警时间内完成的异常信息,第一时间发送给基线报警接收人员。在基线上的任务依赖关系错综复杂,并且依赖变更频繁场景下,基线可为您提前发现问题并预警。

- 当基线根据基线任务历史一段时间内的平均完成时间,推算出基线任务预计完成时间晚于基线预警时间时,将会触发基线预警。您可以在基线管理界面查看计算出来的基线任务预计完成时间,详情请参见:创建基线。
- 当基线根据基线任务的上游任务历史一段时间内的平均完成时间,推算出上游任务预计完成时间晚于基 线预警时间时,将会触发基线预警。
- 任务运行过程中,基线报警策略:

当基线上任务实际完成时间晚于基线预警时间时,将会触发基线预警。

核心逻辑介绍:事件报警

监控范围确定后,当监控范围内的任务出现异常时,智能监控会生成事件,并基于对该事件的分析触发报警。 任务的异常包括:

- 出错:任务运行失败。
- 变慢:任务本次运行时间相比过去一段时间内的平均运行时间明显变长。

⑦ 说明 如果一个任务先变慢、再出错,会生成两个事件。

您可以进入事件管理页面查看产生的事件详情。

核心逻辑介绍:关键路径与关键实例

基线上需要保障的任务,其依赖关系可能错综复杂,DataWorks提供甘特图功能帮助您快速定位阻塞基线上数据产出的关键路径与关键实例,其中影响基线任务产出的多条路径中,耗时最长的路径为基线关键路径。



04:00 04:30 05:00 05:30 06:00 06:30 07:00 07:30 08:00 08:30 09:00 09:30 #700002341636 #700005040134 [shell] xc_甘特图_L1_A #700005044162 [shell] xc_甘特图_L2_B #700005040253 [shell] xc_甘特图_L3_F #700005040264 -[project:xc_DPE_E2]变慢 [shell] xc_甘特图_L4_I #700005040269 捕获时间:07:47:54 [shell] xc_甘特图_L5_K #700005040270

甘特图用于反映目标任务的关键执行路径。以上述案例为例,关键路径展示及异常捕获时间如下所示:

3.6.2.2. 基线管理

如果您需要保障重要任务能按时完成,您可以使用基线管理功能将重要任务添加至基线上并设置承诺时间,系 统将根据任务运行情况计算基线任务的预计完成时间。当系统判断基线任务可能无法在承诺时间前完成时,将 发出告警。本文为您介绍如何创建和管理基线。

背景信息

智能基线能够及时捕捉导致任务无法按时完成的异常情况并提前预警,保障复杂依赖场景下重要数据能在预期 时间内顺利产出。详情请参见:智能基线概述。

使用限制

● 版本要求:

仅DataWorks标准版及更高版本可以使用基线管理功能。低版本用户请先升级再使用该功能,详情请参见: DataWorks各版本详解。

- 权限管控:
 - 新建基线: 仅空间管理员、阿里云主账号和租户管理员可以创建基线。角色添加详情请参见:添加工作 空间成员和角色。
 - 删除、修改、开启或关闭基线: 仅租户管理员和基线责任人可以删除、修改、开启或关闭基线。
- 报警方式:事件报警和基线报警支持**邮件、短信、电话钉钉群机器人**和WebHook报警方式,使用限制如下:
 - 电话: 仅DataWorks专业版及以上版本,才可以使用电话告警功能。
 - WebHook:
 - 可用版本: 仅DataWorks企业版及旗舰版支持使用WebHook功能。
 - 可用地域:仅华东2(上海)、西南1(成都)、华北3(张家口)、华北2(北京)、华东1(杭州)、 华南1(深圳)、中国(香港)、欧洲中部1(法兰克福)、亚太东南1(新加坡)地域支持使用 WebHook功能报警。
 - 可用报警方式: 仅支持推送报警信息至企业微信或飞书。

? 说明

- DataWorks仅支持使用钉钉群、企业微信和飞书的WebHook地址,如果您需要使用其他的 WebHook,请工单联系我们。
- RAM用户如果希望通过短信、电话等方式接收告警信息,则需要在报警联系人页面,将目标 RAM用户添加为报警联系人。当任务运行报错时,DataWorks会将相应的报警信息发送至对应 联系人。详情请参见:查看和设置报警联系人。

创建基线

- 1. 进入运维中心。详情请参见: 进入运维大屏。
- 2. 单击左侧导航栏中的智能监控 > 智能基线。
- 3. 在基线管理页签中,单击+新建基线,并配置基线的基本信息和报警行为,即可创建基线。

≡	🕅 DataWorks										ļ	の 简体	~	
¢	运维大屏	基线实例 事件管理	基线管理	0										
\$ \$\$	实时任务运维 🗸 🗸													
t	周期任务运维	责任人: 请输入责任人名字/ID		工作空间: 请输入工作	空间	基线名称								
ର	手动任务运维 🗸 🗸	类型: ✔ 天基线 ✔ 小时基线	优先级:	🗸 1 🗹 3 🗹 5 🗸	7 🔽 8	仅显示开启	捜索]						
᠕	智能监控 ^	+新建基线 3												
	智能基线	基线名称	责任人	工作空间	优先级↓↑	报警接收人	开启	承诺时间	预警时间	操作				
	规则管理	10,7681,111		1011108		10.00	否	每天 09:00	每天 08:50	详情 编辑 删除	开启	变更记录		
	报警信息	-												
	值班表 自动远维	10.2525.111	pport	101111281		10000	否	每天 09:00	每天 08:30	详情 编辑 删除	开启	变更记录		

4. 配置基线的基本属性。

新建基线								×
基本信息								
基线名称:								
所属工作空间:	请选择						~	
责任人:	请输入表	责任人名字/ID					~	
基线类型:	• 天基	线 🔵 小时基线						
保障任务:	序号	节点名称	责任人	工作空间	定时时间	预计完 成时间	操作	
				没有数据				
优先级:	请选择	~						
预计完成时间:	(历史数据不足, 暂无法预估)							
承诺时间:	每天 请选择时间							
预警余量:	0	分钟						

参数	说明
基线名称	输入基线的名称。
所属工作空间	选择当前需要保障的任务所属的工作空间。
责任人	基线负责人。
基线类型	定义基线监测周期,包括天监测和小时监测。 • 天基线:按照天维度对任务进行监测,天调度任务可选择此监测方式。 • 小时基线:按照小时维度对任务进行监测,小时调度任务可选择此监测方式。
保障任务	 选择需要添加到当前基线上进行数据产出保障的任务。 任务节点:输入任务节点名称或ID后,单击右侧的添加按钮进行添加,可以添加多个任务节点至当前基线。 业务流程:输入业务流程名称或ID后,单击右侧的添加按钮进行添加,默认将当前业务流程下的所有任务添加至当前基线。 说明 选择业务流程后,建议仅添加业务流程最下游的节点任务,添加后,影响该节点数据产出的上游节点都将被纳入基线监控范围,不推荐将业务流程上所有的任务都添加到基线上。
优先级	定义基线的优先级,数值越大基线优先级越高,基线上任务的优先级也会越高。优先级高的 任务,在调度资源紧张的情况下,将优先获得调度资源。 ⑦ 说明 该优先级将映射为MaxCompute计算任务的优先级。
预计完成时间	系统将根据基线任务的历史运行数据计算基线的预计完成时间。如果历史数据不足,则系统 会提示 历史数据不足,暂无法预估 。
承诺时间	 基线上的任务承诺最晚完成时间(需要保障的数据最晚产出时间),基线将根据该时间来推算预警时间,承诺时间需要基于预计完成时间来配置。保证 承诺时间 - 预警余量时间 晚于基线任务的预计完成时间。 ② 说明 基线预警时间-承诺时间-预警余量,如果实际完成时间晚于 承诺时间 - 预警余量时间,则会触发报警。例如,设置承诺时间为3:30,预警余量为10分钟,如果系统预测任务无法在3:20分完成,则会发出基线报警。 小时基线需要指定保障数据产出的小时实例,并设置该实例(小时任务某周期)最晚完成时间。
预警余量	用于定义基线的预警时间。建议预警余量基于基线上任务的运行时长来配置。详情请参见:合理配置基线承诺时间和预警余量。

5. 配置基线的报警行为。

您可以通过报警开关控制基线是否开启报警,并定义基线上数据无法按时产出时的基线报警策略,及影响 基线数据产出的任务及其上游任务出错或变慢时的事件报警策略。

9 如木收到古警1	言思,请参考又相	到进行排亘。		
許关:				
线报警 事件报警				
方式:	🔽 短信 🗸	邮件 🔽 电话 🔽 钉钉群机器人 🧧	VebHook	
]群机器人:	@所有人	Webhook地址	操作	
			发送测试消息删除	Ð
Hook:	Webhookit	b <u>ti-</u>	操作	
			发送测试消息删除	Ð
	 基线责任 	人 检验联系方式		
	○ 值班表	请选择值班表 💙		
	◯ 其他	请输入接收人名字/ID ¥		
理警次数:	3	次		
警间隔:	30	分钟		
ì时间:	00:00	() - 00:00	\bigcirc	

参数	说明
报警开关	 单击报警开关按钮开启或关闭事件报警和基线报警。 开启报警开关 当预测到基线上任务无法在承诺时间内完成时,系统将根据定义的报警方式发送基线报警信息。详情请参见:核心逻辑介绍:基线报警。 当基线任务及其上游任务出错,或关键路径上的任务变慢时,系统将根据定义的报警方式发送事件报警信息。您可以在DataWorks的事件管理页面查看当前已存在的事件列表。详情请参见:事件管理。
	⑦ 说明 报警开关打开后,您可以选择进行基线报警还是事件报警,建议对重要 任务同时配置基线报警和事件报警。
	 关闭报警开关:基线不会产生任何报警。如果基线任务为开启状态,则基线实例会正常 生成、基线优先级仍然生效。

参数	说明
	 报警方式:支持邮件、短信、电话的方式报警给基线责任人、值班表中的当日值班人员或者指定人员。同时支持钉钉群机器人报警或通过添加WebHook地址的方式报警给其他应用(目前支持钉钉、企业微信和飞书)。 报警方式为钉钉群机器人、webhook时,您可单击操作列的发送测试消息,测试该报警是否可以正常发送。配置钉钉机器人详情请参见:场景实践:发送报警消息至钉钉群。 报警方式为短信、邮件、电话时,您可以单击校验联系方式,校验手机号、邮箱是否填写正确。
基线报警	 ⑦ 说明 ◎ 您需要购买DataWorks专业版及以上版本,才可以使用电话报警功能。 ■ 如果您选择通过电话方式进行报警,为了避免短时间内产生大量报警电话,DataWorks会对报警电话进行过滤,同一个用户在20分钟内最多只会接收到一通报警电话,其余报警电话将被降级为短信,请知悉。
	 报警疲劳度控制 最大报警次数:报警的最大次数,超过设置的次数后,不再产生报警。 最小报警间隔:两次报警之间的最小时间间隔。 免打扰时间:设置了免打扰时间后,则在该时间段内系统将不会发送告警。 例如,当目标任务设置的免打扰时间为 00:00 到 08:00 时,则该时间段内基 线与事件的报警将不会触发。如果到达8点,事件仍处于异常状态,则会发出报警信 息。

参数	说明
	 报警类型:包括出错和变慢报警。 出错:基线监控范围内的任务运行失败。 变慢:基线监控范围内的任务本次运行时间和过去一段时间内的平均运行时间相比,明显变长。
	 报警方式:支持邮件、短信、电话的方式报警给基线责任人、值班表中的当日值班人员 或者指定人员。同时支持钉钉群机器人报警或通过添加WebHook地址的方式报警给其 他应用(目前支持钉钉、企业微信和飞书)。
	 报警方式为钉钉群机器人、webhook时,您可单击操作列的发送测试消息,测试该 报警是否可以正常发送。配置钉钉机器人详情请参见:场景实践:发送报警消息至钉 钉群。
	 报警方式为短信、邮件、电话时,您可以单击校验联系方式,校验手机号、邮箱是否 填写正确。
事件报警	⑦ 说明
	■ 您需要购买DataWorks专业版及以上版本,才可以使用电话报警功能。
	如果您选择通过电话方式进行报警,为了避免短时间内产生大量报警电话,DataWorks会对报警电话进行过滤,同一个用户在20分钟内最多只会接收到一通报警电话,其余报警电话将被降级为短信,请知悉。
	◎ 报警疲劳度控制
	 最大报警次数:报警的最大次数,超过设置的次数后,不再产生报警。
	 最小报警间隔:两次报警之间的最小时间间隔。
	 免打扰时间:设置了免打扰时间后,则在该时间段内系统将不会发送告警。
	例如,当目标任务设置的免打扰时间为 00:00 到 08:00 时,则该时间段内基 线与事件的报警将不会触发。如果到达8点,事件仍处于异常状态,则会发出报警信 息。

6. 单击**确定**,完成基线的创建。

基线管理

≡	仰 DataW	orks													Ç	む 前体	~ 10.000
¢	运维大屏			基线实例	事件管理	基线管理											
\$\$\$	实时任务运维		Г														
tā	周明任务运维			责任人: 请输入	责任人名字/ID		工作空间: 请输入工作	空间	基线名称								
ନ	手动任务运维			类型: 🔽 天基线	: 🔽 小时基线	优先级:	💙 1 🔽 3 🗹 5 🗹	7 🔽 8	仅显示开启	搜索							
᠕	智能监控		ļ	+ 新建基线													
	智能基线			基线名称		责任人	工作空间	优先级 』	报警接收人	开启	承诺时间	预警时间	操作				0
	规则管理			10.0081-0	141	-	10717-008	1	standarda.	是	每天 09:00	每天 08:50	详情	编辑	关闭	变更记录	Ŭ.
	报警信息																
	值班表			10.000		10,100	t	1	Antonio de la composition de l	否	每天 09:00	每天 08:30	详情 删除	编辑	开启	变更记录	<u>ا</u>
	自动运维					-											
	资源运维			1818		_3111_1	C	1	40000010.000.0	否	每天 02:00	每天 02:00	详情 删除	编辑	开启	变更记录	E.I.

操作	说明
过滤	在上图的区域①,您可以通过筛选条件过滤目标基线。 您可以根据 责任人、工作空间、基线名称、基线类型、优先级、基线开启状态 等条件进行精确筛 选。
基线操作	在上图的区域②,您可以单击相应基线后的详情、编辑、开启/关闭、变更记录或删除进行相关操作。 • 详情:单击详情,即可查看基线任务的基本情况。 • 编辑:单击编辑,即可直接修改基线。 • 变更记录:单击变更记录,即可查看基线的历史变更记录。 • 开启或关闭:控制基线任务的状态,开启后方可生成周期实例。保障基线每日生成基线实例,您可以在基线实例面板查看每日基线详情。 • 删除:单击删除,即可直接删除基线任务。

添加任务

任务只可以添加在一条基线上,例如,当任务A已添加至基线A上,此时新建基线B并添加任务A时,任务A将 从基线A转移至基线B。

 ⑦ 说明 当开启状态的基线上无任务时,该基线将变为空基线并产生空基线实例,关于空基线的说明详 情请参见:为什么在基线实例页面会显示基线状态为空基线?。

您可以通过以下2种方式添加任务至基线上:

- 进入基线管理页面,单击右上角的新建基线进行添加。
- 进入周期任务页面,选择相应任务后的更多 > 添加到基线。
 - 单个任务添加至基线

G	之 运维中心		:::	~										& DataStudio	₽ 任务发布	್ನ	·
e	运维大屏		搜索:	节点名称/节点ID (Q 节点类型	请选择节点	类型 🗸 共	任人: 请选	择责任人	~ (我的节点 🗌 今	日修改的节点	_ 暫停()	5结)节点			
\$\$\$	实时任务运维	~															〇刷新「展开搜索
а	周期任务运维	^		名称	Ť	5点ID	修改日期↓		任务类型		责任人	调度类	型资	源组 🍸	操作		
	周期任务			1000	7	00002680048	2019-09-12 13:54	4:34	Athena Node		100.00	日调度	R	认资源组	DAG图	测试	补数据 ▼ 更多 ▼
	周期实例			11.0	7	00002680047	2019-09-12 13:54	4:26	EMR_HIVE		-	日调度	ž	认资源组	DAG图	测	暂停(冻结) 恢复(解冻)
	补数据实例			1000.0	7	00002680046	2019-09-12 13:54	4:10	虚节点		-	日调度	R	认资源组	DAG	测	查看实例 🚽 👻
0	#104.9499				7	00002679618	2019-09-12 11:13	3:59	testNode		10.00 A	日调度	R	认资源组	DAG图	观	添加报警 修改责任人
~	宇宙に方にな	Ĵ			7	00002672064	2019-09-11 20:13	3:58	ODPS_SQL		-	日调度	R	以资源组	DAG图	»	添加到基线
					7	00002677489	2019-09-11 19:51	1:28	ODPS_SQL		100 C	日调度	ž	认资源组	DAG图	2	1%は、気が成明 査習血縁 マ
					7	00002677488	2019-09-11 19:51	1:24	Athena Node		-	日调度	R	认资源组	DAG	N	上下游

批量添加任务至基线

	ix逾虑花1~3分钟填写DataWorks满意意问卷,您的支持将帮助我们越来越好 ·································											
≡	🍿 DataWe	orks I j	云维中心 生产	1	~		₽ DataStudi	io 🕂 🛛 🔍 🖄 🤻				
¢	运维大屏		编奏。	- - - - - - - - - - - - - -	刑 , 注注权志,		注进权主/「1					
\$ \$\$	实时任务运维	~	(2余) 调度资源		里· 開週早口7 目 △日修改			~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~				
u	周期任务运维	^	9403C34C8/				7/m	C #	制新 展开搜索			
	周期任务			名称	节点ID	修改日期 11	任务类型	操作				
	周期实例											
	补数据实例			Contraction of the second line second line of the s	1000000000	2022-03-21 10:44:41	IntoNode	DAG图 测试 补数据 更	18			
	测试实例											
ନ	手动任务运维	~		A REAL PROPERTY AND A REAL	Contractor of	2022-03-21 10:44:31	talnitCheckDon	DAG图 测试 补数据 要	3			
᠕	智能监控	~		In the second second second	Terrar Disease of	2022-03-21 10:44:21	http://www.h	DAG图 测试 补数据 要	8			
	资源运维	ļ		a set of a second second								
۲	引擎运维			and the second second second		2022-03-21 10:43:59	SQL	DAG图 测试 补数据 更	18			
R	操作历史		-									
					The statements	2022-03-21 10:43:47	「「 加 加 把 智	g 图 测试 补数据 更	18			
			•				添加到基					
			修改	告任人 修改调度资源组 修改素	7根年成资源组	暂停(冻结) 恢复	(解冻) 百多					

后续

- 开启状态的基线每天会生成基线实例,您可以在基线实例页面查看基线运行详情,详情请参见:基线实例。
- 您可以进入操作历史界面查看基线的操作记录。

3.6.2.3. 基线实例

系统使用基线实例计算任务每次运行的预计完成时间。本文为您介绍如何查看基线实例详情。

背景信息

智能基线能够及时捕捉导致任务无法按时完成的异常情况并提前预警,保障复杂依赖场景下重要数据能在预期 时间内顺利产出。详情请参见:智能基线概述。

使用限制

- 仅DataWorks标准版及更高版本可以使用基线实例功能。低版本用户请先升级再使用该功能,详情请参见:DataWorks各版本详解。
- 仅开启状态的基线才会生成基线实例。详情请参见:基线管理。

查看基线实例列表

- 1. 进入运维中心。详情请参见: 进入运维大屏。
- 2. 单击左侧导航栏中的智能监控 > 智能基线。
- 3. 单击基线实例进入基线实例页面。

	දී ප. m# マ====									
➡ 运维大屏 基线实例 事件管理 書	鉄管理									
第 实时任务运维 →										
□ 周期任务运维 ∨ 业务日期: 2022年3月20日	费任人: 词给入责任人名字/ID 相关事件ID: 请给入事件ID 1									
♀ 手动任务运维 ∨ 工作空间: 请输入工作空间	基线名称: 请输入基线名称 类型: ✔ 天基线 ✔ 小时基线									
▲ 智能監控 ▲ 优先級: 2 1 2 3 2 5 2 7	2 8 基紙状态: 💟 安全 💟 預警 💟 破焼 💟 其他 完成: 💟 未完成 💟 已完成 🦉 🧝									
智能基线	*									
	代 ※ 接着 探察 余⑦ 颈1最终实例 当前关键实例⑦ 操作 び が同 人 量の 預1最终实例 1 前子链实例⑦ 操作									
值班表	······································									
自动运维										
■ 资源运维	the second secon									
 引擎运维 v 100000000 	计值 处理 目前图 相关事件									
し 智能诊断										
♀ 操作历史										
功能	说明									
	在上图的区域①,您可以通过筛选条件过滤日标基线实例。									
过滤项	您可以通过业务日期、责任人、相关事件ID、工作空间和基线名称、基									
	纥奀型、伔旡级 杣 基线状态 等进行精细筛选。									
	在上图的区域②,为您展示基线名称、责任人、工作空间、优先级、基 线状态、完成状态、基线时间、报警接收人、余量、预计最晚实例、当 前关键实例等基线详细信息,同时您可以在操作列对基线进行查看详									
	情、 处理 基线实例和查看 甘特图、相关事件 等操作。									
	◎ 基线状态:									
	■ 安全:任务在预警时间之前完成。									
	■ 预警:任务在预警时间之后未完成,但还未到达承诺时间。									
	■ 破线: 任务在承诺时间之后仍未完成。									
	 ■ 其他·其线所有任冬外干堑信状态或其线沿右关联任冬 									
基线实例列表	 预计 最									
	 当前关键实例:基线关键路径上首个未运行成功的实例。未运行成功包括:任务未运行、等待资源、等待定时时间、正在运行、运行失败等状态。 									
	○ 余量:任务异常的警戒程度。									
	- 当 圣线 [[为] [] [] [] [] [] [] [] [] [] [] [] [] [
	■ 当基线任务未完成时, 余重 = 基线预警时间 - 基线预计完成时间									
	•									
	如果余量为负数,则基线外干预警戓破线状态。									

查看基线实例详情

单击目标基线右侧操作列的详情,即可打开基线实例详情页面。

基金法外)→ 基金法外第第 金融合同時、2022年4-33 東朝: 1 V FREIがStole, REINSTRIE #14.1										
基平復巻 回 運動費調 2022-03-110-030 預計量額(素明) 実内感を、2022の8月11日(00:25.55 当前大規範実所) 実列数称: -	O 単地想要 2022年5日1030 単本A: 東国A:		ಂ ಹೇಕೆರುಡ ಚಿತ್ರಸ. ಕಗರ: ಡಿಲಿಗವರ		19 17					
关键路径					主名曰於因					
任务实例名称	优先级	责任人	預t完成	余量						
contract of contracts		NUMBER OF CONTRACTOR OF	2022-03-11 00:02	12分钟						
-			2022-03-11 00:20	-3:944						
		P	2022-03-11 00:25	4分钟						
历史完成曲线										
18:00 -										
06:00 -										
00:00					#: 00:20					
相关事件										
事件iD 事件描述		事件责任人	触发时间11	犹恭召						
		12: 明明2: 福田								

- 基本信息:若为小时基线,您可以在右上方切换周期。
 - 基线承诺时间:基线上任务的最晚完成时间(需要保障的数据最晚产出时间)。
 - 基线预警时间:即 基线承诺时间-预警余量 。

⑦ 说明 如果您希望为运维人员预留一定的时间处理异常,您可以为基线设置预警余量,则系统
 会将 承诺时间-预警余量 作为触发报警的预警时间,判断任务能否在预警时间前运行成功。详情请参见:核心逻辑介绍:基线报警。

- 预计最晚实例:为您展示基线任务中,预计最晚运行结束的任务。
- 当前关键实例:为您展示基线关键路径上首个未运行成功的实例。未运行成功包括:任务未运行、等待资源、等待定时时间、正在运行、运行失败等状态。
- 处理人:处理当前基线预警的人员。
- 关键路径:为您展示影响当前基线数据产出的关键路径。支持通过列表及甘特图的形式进行查看,关键路径的确定详情请参见:核心逻辑介绍:关键路径与关键实例。

⑦ 说明 甘特图除了为您展示关键路径上任务的运行情况外,还会为您展示任务的历史运行情况。

- 历史完成曲线:为您展示基线历史运行情况。
- 相关事件:为您展示造成任务无法按时完成的异常事件,包括任务变慢和出错事件。事件相关说明详情请参见:事件报警策略。

常见问题

为什么在基线实例页面会显示基线状态为空基线?
3.6.2.4. 事件管理

基线任务及其上游任务出错,或关键路径上的任务变慢时,将会产生相应的事件,您可以在事件管理页面查看 事件详情。

背景信息

智能基线能够及时捕捉导致任务无法按时完成的异常情况并提前预警,保障复杂依赖场景下重要数据能在预期 时间内顺利产出。详情请参见:智能基线概述。

使用限制

仅DataWorks标准版及更高版本可以使用事件管理功能。低版本用户请先升级再使用该功能,详情请参见: DataWorks各版本详解。

相关概念

- 出错:任务运行失败。
- 变慢:任务本次运行时间相比过去一段时间内的平均运行时间明显变长。

⑦ 说明 如果一个任务先变慢、再出错,会生成两个事件。

查看事件列表

Ξ	ntataWorks 🅼	4	? 🕄 简体 🕯	-
¢	运维大屏	基线实例 事件管理 基线管理		
\$\$\$	实时任务运维 🗸 🗸			
t1	周期任务运维 🗸 🗸	责任人: 请输入责任人名字//D 发现时间: 2022年3月20日 17:56:49 2022年3月21日 17:56:4 団		
ଜ	手动任务运维 🖌 🖌	事件状态: ✔ 新发现 ✔ 处理中 📄 已恢复 📄 已忽略 🖤件类型: ✔ 出措 ✔ 变慢 优先级: ✔ 1 ✔ 3 ✔ 5 ✔ 7 ✔ 8		
*	智能监控へ	任务节点 ♥: 请输入任务节点名称//D 搜索		
	智能基线			
	规则管理	専件ID 状态 工作空间 任务实例 优先 类型 时间 最差基线 抽	操作	0
	报警信息			
	自班表	T B BE BELLERTER I BELLERTER	岸筒 处埋 忽略	
	2015年	The second	with at TD	
	以歐运年	The second secon	平南 <u>火</u> 理 忽略	
. .	Japagat			
IJ	力能	说明		
		在上图的区域①,您可以通过筛选条件过滤目标事件。		
过滤项 您可以通过责任人、事件发现时间、事件状态、事件类型、优先级、任务 节点或任务实例进行精确筛选。对于搜索出的结果,每一行代表一个事件 关联到一个异常的任务)。				务 (即

功能	说明
	在上图的区域②,为您展示事件ID、事件状态、工作空间、任务实例、优先 级、事件类型、时间、最差基线等事件详细信息,同时您可以在操作列对基 线进行查看详情、处理事件和忽略事件等操作。
事件实例列表	 详情: 毕士详情后, 您可以查看毕个事件的详细信息。 处理: 单击处理后, 您可以设置处理事件需要的时长, 设置后在该时间段内 将暂停事件报警, 事件的处理操作记录会被记录。
	 忽略:单击忽略后,目标事件永久停止报警,该条事件的忽略记录会被记录。

查看事件详情

单击目标事件操作栏中的**详情**,即可查看事件发生时间、告警时间、恢复时间、任务的过往运行记录及详细的任务日志。

事件管理 → 事件详情 ← 返回	1210							
1.其太保白								
实例名称:				後型: 出槽				
事件责任人:				指政治: [2]				
工作空间:				代先级: 1				
最差基线								
基线名称:				余量: -1328分钟				
时间信息								
0	0		0	0		0		
发生时间	发现时间		首次报警时间	最后处	理时间	恢复时间		
2022-03-21 13:00:08	2022-03-21 13:0	1:36	2022-03-21 13:02:06					
基线影响							查看基线的	299
基线名称	责任人	工作空间	优先级	业务日期	余量	状态	操作	
100000000000000000000000000000000000000	100.000	programming and	1	2022-03-20	-316分钟	破线	洋清	
医中语导动性								
DJELGARAIG								
任务日志							宣誓日志汝	梢

- 基本信息:为您展示事件产生的实例名称、事件类型、事件责任人等基本信息。
- 最差基线:如果当前事件对应的任务有多个下游任务处于开启状态的基线上时,基线余量时间最少的基线 将被视为最差基线。
- 时间信息:事件发送时间轴,包括事件发生时间,事件发现时间、首次报警时间、最后处理时间和时间恢复时间。单击查看报警信息即可跳转至事件对应的报警详情页面。
- 基线影响:当前事件所影响的基线数据产出的基线列表。
- 历史记录对比:当前事件对应的任务每日数据产出时间变化图。
- 任务日志:当前事件对应的任务当日周期实例运行日志。

3.6.2.5. 附件

3.6.2.5.1. 合理配置基线承诺时间和预警余量

本文为您介绍如何合理的配置基线承诺时间和预警余量。

背景信息

- 智能基线能够及时捕捉导致任务无法按时完成的异常情况并提前预警,保障复杂依赖场景下重要数据能在 预期时间内顺利产出。详情请参见:智能基线概述。
- 基线承诺时间是任务运行成功的最晚时间点。即面向数据应用,任务承诺在该时间点前完成。如果希望为运维人员预留一定的时间处理异常,您可以为基线设置预警余量,则系统会将承诺时间-预警余量作为触发报警的预警时间,判断任务能否在预警时间前运行成功。
- 设置基线承诺时间和预警余量的详情请参见: 创建基线。

使用限制

仅DataWorks标准版及更高版本可以使用基线管理功能。低版本用户请先升级再使用该功能,详情请参见: DataWorks各版本详解。

如何合理的配置基线承诺时间和预警余量

您需要根据基线上任务历史一段时间内的最晚完成时间,将基线承诺时间设置在基线上任务最晚完成时间之后,并预留一定的预警余量。以保证您在收到任务异常的预警信息后能够在承诺时间内完成异常处理。



未合理配置基线承诺时间和预警余量的场景示例

如果基线承诺时间和预警余量设置不合理,可能会造成基线破线,进而导致基线预警不符合预期。

场景一:当基线承诺时间与基线上任务最晚完成时间相同,且基线未设置预警余量时:



任务出现异常时,因为没有预留异常处理时间,容易造成基线破线,基线报警频繁。

场景二:当预警余量设置不合理,即基线预警时间(基线承诺时间-余量)在基线上任务最晚完成时间段内时:



任务出现异常时,因为预留的异常处理时间不够,容易造成基线破线,基线报警频繁。

• 场景三: 当基线承诺时间设置不合理, 即基线承诺时间设置在基线上任务最晚完成时间之前时:



任务将无法在基线承诺时间内完成,基线预警功能会在当天零点进行提前预警。

3.6.3. 规则管理

3.6.3.1. 自定义规则

自定义监控规则帮助您根据实际业务需求,自定义监控规则,监控任务运行状态或资源使用情况,以便及时发现并处理异常。本文为您介绍如何在规则管理页面自定义报警规则,以及添加钉钉群机器人获取Webhook地址。

使用限制

- 自定义监控规则仅对周期实例生效。
- 自定义监控规则支持邮件、短信、电话、钉钉群机器人和WebHook报警方式,使用限制如下:
 - 电话: 仅DataWorks专业版及以上版本,才可以使用电话告警功能。
 - WebHook:
 - 可用版本: 仅DataWorks企业版及旗舰版支持使用WebHook功能。
 - 可用地域:仅华东2(上海)、西南1(成都)、华北3(张家口)、华北2(北京)、华东1(杭州)、 华南1(深圳)、中国(香港)、欧洲中部1(法兰克福)、亚太东南1(新加坡)地域支持使用 WebHook功能报警。
 - 可用报警方式: 仅支持推送报警信息至企业微信或飞书。

⑦ 说明 DataWorks仅支持使用钉钉群、企业微信和飞书的WebHook地址,如果您需要使用其他的 WebHook,请工单联系我们。

创建自定义规则

- 1. 进入运维中心。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 选择工作空间所在地域后,单击相应工作空间后操作列的…图标,选择运维中心。
- 2. 在左侧导航栏, 单击智能监控 > 规则管理。
- 3. 在规则管理页面,单击新建自定义规则。
- 4. 在新建自定义规则对话框中, 配置各项参数。

新建自定义规则							×
基本信息							^
规则名称:	请输入规则名称	<i>ъ</i>					
对象类型:						~	
规则对象:	请输入基线名称						
	序号	基线名称	责任人	工作空间	操作		
	1	2	dataworke de	-	删除		I
任务白名单 1:	请输入节点名称	弥/ID,如果有多个,可以	用,分隔				1
	节点名称	节点ID	责任人	工作空间	操作		
没有数据					l		
触发方式 ———							
触发条件:	选择触发条件	~ ()					
报警行为 ———							1
● 如未收到告警信息,请参考文档进行排查。							
报警方式:	报警方式: 短信 🔄 邮件 🔄 电活 🔄 钉钉群机器人 🔄 WebHook						
疲劳度控制							
最大报警次数:	3 次						
最小报警间隔:	30	分钟					-
						确认取消	ă

i. 配置基本信息。

参数	描述
规则名称	输入新建自定义规则的名称。
	指定监控对象的类型,定义监控的粒度。包括 任务节点、基线、工作空间、业 务流程、独享调度资源组和独享数据集成资源组。
对象类型	 ⑦ 说明 对象类型为基线时,仅对基线上的任务运行状态进行监控,不会监控基线任务的上游任务,如果您需要同时监控基线任务的上游任务,请参见:智能基线概述。
规则对象	指定监控对象。 如果 对象类型 选择 任务节点、基线、工作空间 和 业务流程 ,则需要填写 规则对 象。输入监控对象的名称或者ID后,在列表中选择需要添加的对象,单击 添加 。
任务白名单	指定上述监控对象中,哪些任务不受当前规则监控。 当 对象类型为基线、工作空间、业务流程 时,支持您输入 节点名称/ID ,单 击 添加 按钮将任务至白名单列表中。白名单中的任务将不受监控。
资源组名称	指定需要监控的独享资源组名称。 当 对象类型为独享调度资源组和独享数据集成资源组 时,则需要在下拉列表中 选择需要监控的资 源组名称 。

ii. 配置触发方式。

⑦ 说明 自定义报警逻辑中,任务冻结状态表示任务已完成。

对象类型	触发条件	描述
	完成	表示从任务运行的起始时间点开始监控,在任务运行成功时系统 发送报警。 当对象类型为基线或业务流程时,仅当基线、业务流程上的 所有任务都运行成功时,才会触发报警。 当对象类型为任务节点,且添加了多个节点时,仅当所有节点 都执行完成时,才会触发报警。 当对象类型为工作空间时,触发条件不支持设置为完成。 ③ 说明 对于小时任务,只有全部周期均运行成功后, 该任务才会被判定为完成。

对象类型	触发条件	描述
	未完成	 表示从任务运行的起始时间点开始监控,到指定的目标时间点任务仍未结束运行,则系统发送报警。 ⑦ 说明 此报警与智能基线报警策略不同,智能基线能够及时捕捉导致基线上的任务无法按时完成的异常情况并提前预警,详情请参见:智能基线概述。 场景一:任务的定时调度时间为1点,设置的未完成时间为2点,则1点时该任务开始运行,在2点时任务仍未结束运行,则发送报警。 场景二:设置任务每天0:00~23:59每小时调度一次,当该任务未完成时间设置为12点时,则该规则每天都会触发告警。 场景三:设置某基线完成时间为10点,则只要基线上有一个任务没有在10点完成,便会发送告警。 ⑦ 说明 对于小时或分钟任务,会校验当天所有周期在
		指定时间点是否未运行完成。
	出错	 家统发送报警。 ⑦ 说明 实例任务运行出错即在运维中心 > 周期任务运维 > 周期 实例的基本信息列,目标实例显示_⊗状态。
		 当对象类型为基线、工作空间或业务流程时,只要基线、工作空间或业务流程内的某个任务运行出错,便会触发告警。 任务每次运行出错都会触发报警,如果某个任务出错报警次数为2,当该任务重跑2次均出错时,将会触发4次报警。 如果您需要在任务自动重跑后仍出错时再触发报警,则可以将触发条件配置为:自动重跑后仍出错。
任务节点、基线、工 作空间、业务流程		

对象类型	触发条件	描述
	周期未完成	表示在指定的周期内,任务仍未结束运行,则系统发送报警。通 常用于监控以小时为周期单位的实例任务。 当业务流程设置了 周期未完成 监控后,系统会根据您设置的周期 N,对业务流程中的任务(例如,天任务、小时任务、分钟任务 等),进行第N个周期任务的监控。如果实例数少于N时,则会忽 略该任务的告警。 例如,设置的周期为3,业务流程中有如下两个任务,则告警监 控情况如下: • 小时任务A:每2小时调度一次,运行一次耗时25min。运行起 始时间为每日0点0分,则该任务一天(24小时)共有12个任 务周期,0点为第一个周期,则第3个周期为4点。该任务正常 运行时,第3个周期任务会在4点25分执行完毕。此时设置触发 条件为 周期未完成 ,且指定该任务第3个周期在4点30未完成 时候发送报警信息,则如果任务某天运行到4点30还未完成 时,将会发送报警。
	超时	表示从任务运行的起始时间点开始监控,到指定的运行时长后,该任务仍未结束运行,则系统发送报警。通常用于监控任务的运行时长。 ⑦ 说明 如果监控的任务执行失败,到指定的运行时长 后,该任务仍处于未成功状态,将触发超时告警。
	自动重跑后 仍出错	表示从任务运行的起始时间点开始监控,如果任务运行出错且自动重跑后仍出错,则系统发送报警, ⑦ 说明 如果您需要在任务每次运行出错时就触发报 警,则可以将触发条件配置为:出错。
	任务转实 例 完成	仅当对象类型为工 作空间 时,可以选择该触发条件。
	实例数量波 动	仅当对象类型为工 作空间 时,可以选择该触发条件。DataWorks 每天在24:00前生成次日需要运行的周期实例,当工作空间的实 例数量与历史均值相比发生一定幅度的波动时,您将收到该报 警。

对象类型	触发条件	描述
独享调度资源组、独	资源组利用 率	当 利用率 大于某个数值并持续指定时长时,触发该报警。 示例:利用率大于50%并持续15分钟时,触发报警。
享数据集成资源组	资源组上等 待资源的实 例数	当 等资源实例数 大于某个数值并持续指定时长时,触发该报警。 示例: 等资源实例数 大于10并持续15分钟,触发报警。

iii. 配置报警行为。

报警方式	报警接收人	描述
邮件、短信、电话	支持报警给 任务责任 人、值班表和其他人 员。	 您可以单击校验联系方式,校验手机号或邮箱 是否填写正确。 您需要购买DataWorks专业版及以上版本,才 可以使用电话告警功能。 如果您选择报警方式为电话,则需要选中为了 避免短时间内产生大量报警电 话,DataWorks会对报警电话进行过滤, 同一个用户在20分钟内最多接受到一通报警 电话,其余报警电话将被降级为短信,请知 悉。 您需要先配置值班表,再在此处选择报警给值 班表,值班表配置详情请参见:值班表。
钉钉群机器 人WebHook、	群内人员	 您可单击操作列的发送测试消息,测试该报警是否可以正常发送。如未收到告警信息,请参考智能监控进行排查。 DataWorks仅支持使用钉钉群、企业微信和飞书的Webhook地址,如果您需要使用其他的WebHook,请工单联系我们。 钉钉群机器人安全配置仅支持配置关键字,并且关键字需要包含DataWorks。

iv. 配置疲劳度控制。

参数	描述		
最大报警次数	报警的最大次数,超过设置的次数后,不再产生报警。		
最小报警间隔	两次报警之间的最小时间间隔。		
免打扰时间	设置了免打扰时间后,则在该时间段内系统将不会发送告警。 例如,当设置了任务状态为 超时、出错、未完成 时触发报警,且该任务设置的免 打扰时间为 00:00 到 08:00 ,则该时间段内将不会发出报警信息,如果 到达8点,任务仍处于上述异常状态,将会发出报警信息。		

5. 单击**确定**,即可生成规则。

您可以在规则管理页面单击相应规则后的详情、开启/关闭或删除进行相关操作。

- 详情: 单击详情, 即可查看规则的基本信息。
- 开启或关闭:控制规则的状态,开启后方可监控任务的执行情况。您可以在报警信息面板查看报警详情。
- 删除: 单击删除, 即可直接删除规则。

场景实践:发送报警消息至钉钉群

- 1. 打开需要告警的目标钉钉群,单击右上角的群设置图标。
- 2. 单击智能群助手。
- 3. 在智能群助手页面,单击添加机器人。
- 4. 在**群机器人**页面,单击◎图标。
- 5. 在选择要添加的机器人页面,单击自定义。
- 6. 在自定义对话框中,单击添加。
- 7. 在添加机器人对话框中, 配置各项参数。

参数	描述	
机器人名字	自定义机器人的名称。	
添加到群组	添加机器人的群组,不可以修改。	
	设定后,只有包含关键词的消息内容才会被正常发送。此处必须添 加DataWorks为关键词,注意区分大小写。	
自定义关键词	 ⑦ 说明 最多可以设置10个关键词,消息中至少包含其中1个关 键词才可以发送成功。 	

8. 选中我已阅读并同意《自定义机器人服务及免责条款》,单击完成。

9. 完成安全设置后,复制机器人的Webhook,单击完成。

↓ 注意 请妥善保管Webhook地址,一旦泄露会有安全风险。

10. 进入规则管理页面,单击新建自定义规则,选择报警方式为钉钉群机器人,复制webhook地址至钉钉 群机器人的输入框中。

新建自定义规则								>	<
┃ 触发方式 触发条件: ┃ 报警行为	选择触发条件	~							•
⑦ 如未收到告警信息	息,请参考 <mark>文档</mark> 进行排查。								
报警方式: 钉钉群机器人:	短信邮件@所有人	电话 Webhoo	✔ 钉钉群机器人 ok地址		操作 发送测试消息 删除	(Ð		
疲劳度控制 ——									
最大报警次数: 最小报警间隔:	3		次 分钟						
免打扰时间:	00:00 至 00:00		Q						
						确	i认 [」]	取消	

DataWorks智能监控推送企业微信或飞书的报警信息样例如下。

webhook 🚥	企业微信样例	€书样例 由定义机器人 机器人 🖬 通过webhook将自定义服…
DataWorks Rem	ind	DataWorks自定义报警
Remind Name: ddd	bb	
Remind Unit: Task		规则名称: webhook_test
Occur Time:: 2021	-07-29	触发条件:任务
01:24:46		触发时间: 2021-07-29 11:29:21
Remind Type: com	pleted	规则类型:未完成
Task List: (show siz	ze 5)	任务列表: (共1个,完整列表请前往
Some Nodelds: 10	0000	运维中心)
Remind Founder:		部分节点ID: 7000-4105005
dataworks_3h1_1		规则创建人: dataworks_3h1_1
More information:		查看更多信息:
http://t.tb.cn/1Wi6 qOIOTAQ	tbmT4AXdSYf	http://t.tb.cn/4tEL1n0JVPpcJaT adkZT4D

3.6.3.2. 值班表

您可以使用DataWorks值班表功能设置实例运维报警的值班机制。当自定义规则的报警接收人设置为值班表后,DataWorks将按照排班将报警信息发送给对应的值班人,以便值班人员及时发现并处理问题。本文为您介 绍如何新建并管理值班表。

使用限制

仅DataWorks专业版及以上版本,才可以使用值班表功能。如果您使用的是低版本的DataWorks,则需要升级 至专业版及以上版本,才可以使用该功能。升级DataWorks版本,详情请参见版本服务计费说明。

注意事项

- 当前DataWorks版本到期后,您将无法访问值班表页面,不能查看、新增、修改或删除值班表及对应的排班。
- 如果当前专业版或以上版本的DataWorks配置了值班表,当版本到期后未及时续费或降级了当前版本,则 DataWorks会为您保留原有的值班表数据。当重新升级到专业版或更高版本后,原有的值班表功能可以正常 使用。
- 当您创建的自定义规则中,报警接收人配置了值班表后,在产生报警时值班表对应的值班人员才会收到报警信息。配置自定义规则,详情请参见自定义规则。

新建值班表并配置值班安排

- 1. 进入运维中心。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 选择工作空间所在地域后,单击相应工作空间后的进入运维中心。
- 2. 在运维中心页面的左侧导航栏,单击智能监控 > 值班表。
- 3. 在值班表区域,单击+图标,配置新建值班表的各项参数。

新建值班表				:
名称:	test			
管理员:		×		~
值班提醒:	前1天 🗸 🗸	9 🕓		
提醒方式:	短信 邮件	电话 ✔ 钉钉群机	器人	
	@所有人	Webhook地址		操作
	新増	没有数	E	
				确认取消
参数			描述	
名称			输入值班表的	的名称。
管理员			默认为当前到 DataWorks	登录阿里云的账号, 工作空间的主账号:

参数	描述
值班提醒	开启值班提醒后,将提前通知值班人下次值班的时间, 默认配置为 不提醒 。您可以根据需求选择前1天、前2 天或前一周提醒。 当选择前1天、前2天或前一周时,则可以配置具体的 提醒时间以及 提醒方式 。
提醒方式	DataWorks支持使用 短信 邮件、 电话 及 钉钉群机器 人(Webhook地址)等方式进行值班提醒。添加钉钉 群机器人获取Webhook地址详情请参见场景实践:发 送报警消息至钉钉群。 仅当值班提醒配置为前1天、前2天或前一周时,需要 配置该参数。

4. 单击确认。

5. 在新建的值班表页面,单击右上角的新建值班,创建值班信息。

您也可以直接单击指定日期的白色块,新建值班信息。

← 2021-05	曲 → □	到本月 全部值到	觃人 ~			十新建值班			
名称: test 值班提醒: 前1天14	k:00	管理员: 韩琳 提醒方式: 短	木 豆信	当前值 使用帽	当前值班人: 韩琳 使用情况: 详情				
周日 9	周— 10	周二 11	周三 12	周四 13	周五 14	周六 15			
16	17	18	19	20 ± 00:00 ± 00:00	21 备 00:00 截止23:59	22			
23	24	25	26	27	28	29			
30	31	1	2	3	4	5			

6. 配置新建值班对话框的各项参数。

参数	描述
值班人	当前值班表中配置的值班人员。您可以选择当前登录 DataWorks工作空间的主账号或RAM用户。
类型	○ 主:配置值班人为主值班人。○ 备:配置值班人为备用值班人。

参数	描述
全天	 是:开始时间及结束时间,均为全天24小时值班。 否:可以自定义开始时间或结束时间的值班时间点。
开始时间	开始值班的时间。当 全天 配置为否时,可以自定义值班 时间点。
结束时间	结束值班的时间。当 全天 配置为否时,可以自定义值班 时间点。
重复	默认为不重复。您也可以选择重复的周期为每日、每周、每周一到周五、每月、每年以及自定义。 当配置了重复周期后,当前新增的值班信息会添加在所 选的周期时间内。
时间间隔	时间间隔 可以配置为当前值班时间区域前后的任意天 数。 仅当 重复 配置为 自定义 时,需要配置该参数。
结束重复	用于指定重复周期的结束时间。默认为永不结束。 当 重复 配置为 每日、每周、每周一到周五、每月、每 年以及 自定义 时,需要配置该参数。

示例新建的值班信息如下图。

≡	ntaWorks 🅼	一运	維中心									Ġ	DataStud	io L I L	V	
¢	运维大屏		值班表	ŧ	+		← 2021-05	曲 →	到本月	全部值	班人 ~				十新建值期	
\$\$\$	实时任务运维	*	搜索:	输入值班表名称	Q		名称: test			管理员:			当前值	証 氏: 韩琳		
t1	周期任务运维	*	管理员:	请选择	~		值班提醒: 前1天14:	:00		提醒方式: 知	逗信		使用情	青况: 详情		_
Q	手动任务运维	~					周日	周一		周二	周三	周	Щ	周五	周六	_
*	智能监控	^	test			^)		1	Î
	基线实例		1000													
	基线管理															-11
	事件管理						2	3		4	5	6		7	8	
	规则管理															
	报警信息															
	值班表						9	10		11	12	13	3	14	15	-11
ራ	智能诊断															1
				< 1/	1 >	-	16	17		18	19		00:00	21 含 00:00 截止23:59	22	Ţ

? 说明

- 新建的值班信息中, 蓝色代表主值班人, 黄色代表备用值班人。
- 新建值班时,您只能选择为当前日期及之后日期创建值班信息,当前日期之前的日期则无法 创建值班信息。
- DataWorks支持在相同的值班时间配置多个主或备值班人。

管理值班表

- 在值班表区域,您可以执行如下操作:
 - 通过值班表名称或管理员, 快速搜索符合条件的值班表。
 - 单击目标值班表后的 ··· 图标, 可以修改或删除已创建的值班表。

⑦ 说明 仅值班表管理员可以编辑或删除值班表或已创建的值班安排。

○ 双击目标值班表,则可进入值班表的详情页面,查看值班表的详细信息。

 搜索: 輸入值班表名称 Q 管理员: 请选择 ✓ test ··· 则试 ··· 	值班表		+
 管理员: 请选择 test ・・・ 则试 ・・・ 	搜索:	输入值班表名称	Q
test ····	管理员:	请选择	~
则试	test		••••
	则试		

- 在目标值班表详情页面,您可以执行如下操作:
 - 在①区域,您可以查看值班表的基本信息以及当前的值班人。当您配置的自定义监控规则使用了目标值 班表,则该值班表会按照排班情况发送相关报警信息,您可以单击使用情况后的详情查看具体使用信 息。管理自定义规则,详情请参见自定义规则。
 - 在②区域,您可以通过值班人类型,筛选目标值班人的值班安排,也可以查看指定月份的值班安排。默 认显示当前月份所有值班人的值班安排。

如果您选择了查看非当前月份的值班,在查看完毕后可以单击回到本月,快速返回至当前月份的值班界面。

 • 在值班安排界面,您可以查看已创建值班的值班详情,包括值班人、值班日期、值班类型(属于主值班 或备值班)等。单击已创建的值班,则可选择修改或删除当前值班。

← 2021-05		到本月 全部值球	旺人 ~ 2			十新建值班
名称: test		管理员:		当前值		
值班提醒: 前1天14	k:00	提醒方式: 短	2信	使用情	祝:详情	
周日	周—	周二	周三	周四	周五	周六
25	26	27	28	29		1 Î
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20 (1) (1) (1) (1) (1) (1) (1) (1)	21 翁 00:00 截止23:59	22

3.6.4. 自动运维

您可以根据业务需求对独享资源组自定义运维规则,实现独享资源组的智能监控和自动化运维。本文为您介绍如何通过DataWorks的运维中心配置自动运维规则。

背景信息

DataWorks的自动运维功能可以实现根据已设置好的监控规则和运维规则来自动执行任务的运维操作。您可以 根据具体业务需求,对运行在目标独享资源组上的实例任务定制监控指标,再结合自身业务逻辑自定义运维规 则,并关联目标实例任务的监控规则。当实例任务满足过滤条件时,将自动触发并执行运维动作,以达到自动 运维的目的。

使用限制

- 仅支持项目管理员设置、修改和删除自动运维规则。
- 仅支持对独享资源组上运行的实例任务配置自动运维规则。
- 目前仅支持查看30天内的自动运维规则执行记录。
- 目前仅支持当调度资源组使用率水位过高时终止运行指定任务的自动运维操作。
- 一条自动运维规则仅支持关联一条监控规则,但多条自动运维规则支持关联同一条监控规则。

进入自动运维

- 1. 进入数据开发页面。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 选择工作空间所在地域后,单击相应工作空间后的进入数据开发。

2.

3. 在左侧导航栏,选择智能监控 > 自动运维,进入自动运维页面。

运维规则管理

自动运维页面可以为您展示已创建的运维规则列表及规则执行记录信息,您可以根据需求,执行如下操作:

 ⑦ 说明 仅支持项目管理员权限设置和修改规则,如需添加项目管理员权限,请参见添加工作空间成员 和角色。

规则管理 执行记录					
search	Q				添加规则
名称	关联监控规则	生效工作空间	最后修改时间	修改人	操作
xm_res_test4	xm_res_test2	springboot_标准模式2	2021-06-01 15:55:37	dataworks_3h1_1	查看 删除
wzp_test	xm_res_test2	epringboot_标准模式2	2021-06-01 15:58:47	dataworks_3h1_1	查看 删除
test001	xm_res_test2	springboot_标准模式2	2021-06-03 10:42:59	dataworks_3h1_1	查看 删除

• 添加规则

规则管理页面,单击右上方的添加规则,按照要求填写如下信息:

新建规则						×						
						-						
③ 当关联的监控规则	则被触发,系统将对满足过滤条	合件的实例自动执行运维动	作。目前只支持法	关联调度资源监控规	见则。							
* 名称:	请输入规则名称					- 1						
* 关联监控规则:	H01.047.01		~	去创建监控规则								
* 运维动作:	请选择				~							
过滤条件						- 1						
* 资源组:	dag/and/01/23/08/30/H					- 1						
* 工作空间:	请选择				~							
* 实例类型:	请选择				~							
* 优先级:	请选择				~							
* 状态:	请选择				~							
白名单						- 1						
序号	任务名称	责任人	工作空间		操作							
没有数据												
						- 1						
请选择			~	添加								
约束规则						- 1						
* 生效时段:	请选择时间	① 至 请选择时	前	C								
* 最大生效次数:	次											
* 最小生效间隔:	分钟					+						
					确定	取消						

类别	参数	说明
	名称	输入新建的运维规则名称。
其木信自	关联监控规则	关联需要使用的监控规则。如需新建监控规则,请参见自定义规则。 ? 说明 目前仅支持关联调度资源监控规则。
штт (ЦЛХ	运维动作	选择达到监控规则条件触发监控规则时,实例需要执行 的运维操作。 ⑦ 说明 目前运维动作仅支持选择终止运行实 例。
	资源组	创建监控规则时设置的资源组名称。 ⑦ 说明 此处监控对象仅支持选择独享数据集成 资源组和独享调度资源组。
	工作空间	选择该运维规则生效的工作空间。
	实例类型	选择使用该运维规则的实例类型。包括 周期实例、补数 据实例、测试实例、手动任务和手动业务流程
过滤条件	调度周期	调度周期包括分钟、小时、天、周和月。
	优先级	选择运维规则执行优先级,数值越大优先级越高。取值 如下: 。 1 。 3 。 5 。 7 。 8
	状态	选择实例运行状态。包括 等待资源 和运行中。
白名单	白名单	展示该规则可使用任务白名单,如需添加任务白名单, 请在下拉列表选择需要使用的白名单,单击 添加 。 ⑦ 说明 自动运维规则对白名单任务不生效。
	生效时段	设置规则生效的起始时间和结束时间。
约束规则	最大生效次数	设置规则触发的最大次数。
	最小生效间隔	设置规则触发的最小时间间隔。

● 搜索规则

在规则管理页面左上方的搜索栏,可以通过输入规则名称进行模糊匹配来快速查找目标规则。

- 查看、修改和删除规则
 - 如需查看规则信息,请在规则管理页面,单击目标规则所在行右侧的查看。
 - 如需修改规则定义,请单击查看规则对话框右下方的修改。
 - 如需删除目标规则,请单击目标所在行右侧的**删除**,在弹出的对话框单击确认。
- 查看规则执行记录

在执行记录页面,为您展示运维规则的执行情况,包括执行时间、规则责任人及涉及任务数等,如需查看 详细信息,请单击目标执行记录右侧的**查看详情**,即可查看规则自生效时刻起24小时内的资源使用率水位 曲线信息。

记录详情						×		
等资源实例数	如利用率							
100.00%								
80.00%						<u> </u>		
60.00%	\							
40.00%								
20.00%						— L,		
0.00%	2021-05-27 01:26:40		2021-05-27 15: ● 资源使用率	:20:00 <u>3</u>	2021-05-28 05:13:20			
┃ 被终止运行的	的实例 (共{total}个)							
节点名称	业务日期	实例状态	实例类型	责任人	工作空间	环境		
test	2021-04-09 00:00:00	等待资源	补数据实例	dataworka_25/1_1	springboar, 10 mtBrC1	生产		
test	2021-04-09 00:00:00	等待资源	补数据实例	detereorks_3h1_1	epringbour, BORNET2	生产		
test	2021-04-09 00:00:00	等待资源	补数据实例	determine (Brit)	springhant_SSREET.	生产		
test	2021-04-09 00:00:00	等待资源	补数据实例	dataworka_21/1_1	opringbourt, 10 MtBat21.	生产		
test	2021-04-09 00:00:00	等待资源	补数据实例	determorks_3h1_1	opringbour, NVRNET2	生产		
test	2021-04-09 00:00:00	等待资源	补数据实例	determine (Brit)	springhout, SOURCE	生产		
test	2021-04-09 00:00:00	等待资源	补数据实例	deteworks_SV1_1	springboor, 10 MtBaCL	生产		
					く 上一页 <mark>1</mark>	下—页 >		

 等资源实例数/利用率:以水位曲线形式为您展示等资源实例数和利用率情况,您还可以将鼠标悬停在 水位线显示区域内的任意位置,即可显示此刻资源实例数和利用率情况。

• 被终止运行的实例:为您展示被终止运行的任务列表。

其他

现在您已经学习了如何配置运维规则,配置好运维规则后,系统会自动根据运维规则对资源组使用情况进行监控,资源组监控详情请参见资源运维。

3.6.5. 报警信息

您可以使用报警信息功能查看智能监控模块产生的所有报警信息。包括智能基线产生的基线预警信息和事件报 警信息、自定义规则报警及全局规则报警产生的报警信息。本文为您介绍如何查看报警详情。

背景信息

- 您可以在报警信息页面查看接收到的报警详情,包括触发该报警的监控规则、报警触发条件、您收到报警的原因等。如果您对相关报警存在疑问,您可以根据报警信息类型,查看对应类型的规则触发报警的机制,详情请参见:智能监控概述。
- 如果您不希望收到某类报警,请将报警详情页面的URL发送给规则创建人,相关责任人会在规则管理页面或 智能基线页面将您从规则的报警接收人中移除。规则创建人请参见基线管理、自定义规则界面。

进入报警信息页面

- 1. 进入运维中心。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 选择工作空间所在地域后,单击相应工作空间后操作列的: 图标,选择运维中心。
- 2. 进入运维中心后,在左侧导航栏,单击智能监控 > 报警信息。进入报警信息页面。

查看报警信息

您可以在报警信息页面查看智能监控模块产生的所有报警信息。

= 🕼 DataWorks		运维中心						l 🗗 🕫 🔍	简体	~
 ● 送準大屏 ※ 实时任务运维 1. 周期任务运维 	• •	共型: ✔ 规则 ✔ 基线 其它 报警方式: ✔ 短信 ✔ 邮件	: 規則D/名称: 清給入例 4747 电话 4747群机器	100 / 名称 マ 実例D : (清給入実例D) は WebHook	接收人: 请选择接	юл v	报警时间: 2022	₩5月24日 17:00:10 · 2022₩5月25日 17:00:1 団		0
♀ 手动任务运维	*	报警时间 11	规则/基线	对象	接收人	报警方式	发送状态	内容预览	操作	0
✓ 智能监控 智能基线 規則管理	^	2022年5月25日 16:50:35		任祭: 憲任人 工作会		自印年	发送成功		详情	
报警信息	1	2022年5月25日 16:35:55		任务: 事任人		邮件	发送成功	売成	详情	
自动运维 资源运维		2022年5月25日 16:30:25	The Party of	任勞: 責任人 工作空	(*************************************	部件	发送成功	Jo3)	详情	
① 引撃返後〇 引撃返後	ř	2022年5月25日 16:30:25		任务: 责任人 工作会	(******,0°,	短信	发送成功	J03])	详情	
		2022年5月25日 16:20:24		任务: 妻任人 失败 工作空		₿5/‡	发送成功		详情	
		2022年5月25日 16:16:39	节点孤立报警	节点: 浅 工作空 (目	100	的体	发送成功	6-7-7-8-8	详情	
		2022年5月25日 16:16:39	节点孤立接替	节点: 後 工作空 11	-	坦信	发送成功	1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	详情	
		2022年5月25日 16:16:39	节点孤立报酬	节点: ang 工作空		邮件	发送成功	e-contraction of the second seco	详情	
		2022年5月25日 16:16:39	节点现立接管	节点:ang 工作空	1000	坦信	发送成功	Contract Stratt Labor.	详情	
☞ 操作历史		2022年5月25日 16:16:39	节点孤立报警	多个对象,点由展开 🖌		自5(牛	发送成功	0.000	详情	

区域	描述
筛选	在上图的模块①区域,您可以通过筛选条件过滤需要查询的报警信息。 您可以根据 规则类型、规则ID/名称、实例ID接收人、报警时间、报警方式 等条件进行 精确筛选。
报警信息列表	在上图的模块②区域,以列表方式为您展示满足过滤条件的报警信息。您可以单击操作列 的 详情 进入报警详情页查看详细报警过程。

? 说明

- 当查看事件报警时,规则/基线列将展示当前事件所影响的基线列表。
- 当查看规则报警时,对象列将展示该规则监控对象中,触发报警的对象。

查看报警详情

您可以在**报警详情**页面查看报警的完整触发流程,包括触发该报警的监控规则、报警触发条件、您收到报警 的原因等。

报警详情		×
•	报警对象 任务: 业务日期: 责任人:d 工作空间:	● 不想收到报警
2022-05-25 16:50:35	 报警情况 未完成 	
•	触发规则/基线 规则/基线:	
•	接收对象	
•	特殊处理	
2022-05-25 16:50:35	发出告答 报答方式:邮件 报答状态:发送成功 报答内容: Landon and and and and and and and and and an	
		天闭

- •报警对象:哪个异常任务触发了该报警,即规则监控对象中触发报警的对象。
- 报警情况:该任务触发了那种异常报警,即规则的报警触发条件。如果您对相关报警触发条件存在疑问,您可参考文档进行确认。详情请参见:智能监控概述。
- 接收对象: 谁接收该报警, 一般为规则定义的报警接收人。但如果规则定义的报警接收人未合理配置个人 手机或邮箱, 将会先向主账号发送报警信息, 当主账号也未设置手机或邮箱时, 系统将会向基本接收管 理中产品的欠费、停服、即将释放等信息的消息接收人发送报警信息, 为避免出现上述情况, 您可 在Dat aWorks控制台配置报警联系人完整的手机号、邮箱等信息。详情请参见: 查看和设置报警联系人。

⑦ 说明 如果您不希望收到某类报警,请将报警详情页面的URL发送给规则创建人,相关责任人会在规则管理页面或智能基线页面将您从规则的报警接收人中移除。

3.6.6. 常见问题

本文为您介绍在运维中心设置了智能监控后,任务运行失败的常见报错及解决方案。

没有收到任务运行失败的告警短信/邮件

- 问题现象: 在运维中心设置了智能监控后, 任务运行失败了却没有收到告警的短信或邮件。
- 可能原因:

出现上述问题通常是因为设置的报警接收人的手机号码和邮箱信息不正确。

当DataWorks任务运行失败时,系统会向DataWorks智能监控中设置的报警接收人的手机号码和邮箱发送报 警信息,如果手机号码和邮箱信息不正确,则系统会将报警信息发送到阿里云账号基本接收管理中设置的 产品欠费、停服、即将释放等相关信息的消息接收人。此种情况下DataWorks的报警接收人将收不到报警信 息。

● 解决方案:

您需要确认并修改DataWorks智能监控中设置的报警接收人的手机号码和邮箱信息,确保系统可以向其发送 报警消息。

您可以单击右上角的账号名称,在用户信息中查看或修改邮箱和手机号码信息。

⑦ 说明 RAM账号的邮箱和手机号信息需要对应的主账号进行修改。



3.7. 智能诊断

您可以使用智能诊断功能对任务进行全链路分析,当任务运行不符合预期时,您可以使用该功能快速定位问题。

背景信息

• 任务运行条件说明:

调度任务运行需要满足的条件包括:上游任务运行成功、当前任务定时时间已到、调度资源充足、当前任 务未冻结等。关于任务运行条件说明,详情请参见:任务运行条件。

• 任务未运行问题定位:

如果上游任务为未运行状态且上游依赖层级较深时,建议您先通过DAG面板中的上游分析功能,快速定位阻 塞当前任务运行的关键上游任务,再通过**智能诊断**功能,诊断该关键任务未运行的原因,提升运维效率。

功能概述

智能诊断功能可以基于以下维度对任务进行诊断:

- 检查上游任务运行情况。
- 检查是否已到达任务设置的定时运行时间。
- 检查调度资源使用情况,可在此处查看当前任务等待资源时,哪些任务在占用资源。
- 检查任务执行情况。

关于任务运行全链路诊断详情请参见: 全链路分析。

除上述任务运行诊断外,智能诊断还为您提供提下信息帮助您了解任务的运行情况:

- 支持您查看任务运行的关键时间点。详情请参见基本信息。
- 支持您查看任务影响的基线及对应基线运行情况。详情请参见影响的基线。
- 支持您查看任务的历史运行情况。详情请参见历史实例运行情况。

使用限制

仅Dat aWorks专业版及以上版本,支持使用运行诊断功能。您目前可以免费体验,但建议您升级到专业版以获 取更多产品能力。版本升级详情请参见:Dat aWorks各版本详解。

进入智能诊断

- 1. 进入运维中心。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 选择工作空间所在地域后,单击相应工作空间操作列的 :-- 图标,选择运维中心。
- 2. 进入智能诊断。
 - 方法一:在左侧导航栏,单击周期任务运维或手动任务运维,选择周期实例、补数据实例、测试实 例或手动实例,通过以下方式进入相应实例的智能诊断页面。
 - 单击相应实例前的状态图标,进入实例的**智能诊断**页面。

节点搜索: 节点名称/节点ID Q 业务	马日期: 昨天 前天 全部	3 2022-04-26 2022-0	04-26 🟥 节点	[美型: 请选择节点类型 ↓	
调度资源组 Please Select 🗸 _ 我的节	点 我的出错节点	我的未完成节点 重調	包过的节点	变慢节点	C 刷新 展开搜索
基本信息	任务类型	责任人	优先级 🔰	操作	
 <i>#</i>7 04- 	Test.	-	1	DAG图「去诊断「重跑「更多	<u>.</u>

- 在实例列表模式下,单击相应实例操作列的去诊断。如果当前页面非列表模式,您可单击页面中间的。图标返回列表模式。
- 在DAG模式下,右键单击相应实例的DAG图,选择菜单中的运行诊断。如果当前页面非DAG图模式,您可单击相应实例操作列的DAG图,打开实例的DAG图。
- 在DAG模式下,单击相应实例的DAG图,在右下角的弹框中,单击任务状态右侧的去诊断。
- 方法二:在左侧导航栏,单击智能诊断进入智能诊断页面。

⑦ 说明 智能诊断只支持通过实例ID定位具体实例,您可通过实例详情页获取实例ID。

全链路诊断

DataWorks根据任务运行的必要条件,将依次检查上游任务的运行状态、当前任务设置的定时时间、调度资源的使用情况、当前任务的执行情况:

● 上游依赖

智能诊断的上游依赖页面,为您展示当前任务的上游任务运行情况,如果上游任务未运行成功,将阻塞当前任务运行,您可以单击上游任务操作列的运行诊断,查看上游任务未运行成功的原因。

⑦ 说明 如果上游任务为未运行状态且上游依赖层级较深时,建议您先通过DAG面板中的上游分析功

能,快速定位阻塞当前任务运行的关键上游任务,再通过**智能诊断**功能,诊断该关键任务未运行的原因,提升运维效率。

历史实例									
	💙 定	时检查			🕑 调度资源			2 任务执行	
运行状态	调度关型	责任人	实例类型	工作空 间	业务日期	定时时间	周期序号	运行次数	操作
运行成功	日调度	r	周期实例		2022-01-19	2022-01-20 00:21:00	1	0	运行诊断
运行成功	日调度	r	周期实例		2022-01-19	2022-01-20 00:21:00	1	0	运行诊断
	5史实例 	5史实例	正行成功 現度実型 更任人 遠行成功 日現度 二	正行状态 IP 调度类型 责任人 实例类型 运行成功 日调度 周期实例 运行成功 日调度 周期实例	正行状态 冒 调度类型 责任人 实例类型 工作空 同 运行状态 冒 调度类型 责任人 实例类型 工作空 同 运行成功 日调度 周期实例	正行状态 IP 環度実型 責任人 天例共型 工作空 迎身日期 运行成功 目現度 第 原期末例 2022-01-19 运行成功 日現度 第 周期末例 2022-01-19	正行状态 冗 调度笑型 責任人 文例类型 工作空 业券日期 定时时间 运行状态 冗 调度类型 責任人 玄例类型 工作空 业券日期 定时时间 运行成功 日順度 原用实例 2022-01-19 2022-01-20 00:21:00 运行成功 日順度 原期实例 2022-01-19 2022-01-20 00:21:00	正行状态 TP 環度类型 責任人 文明典型 工作空 迎号日期 短时时间 周期序号 运行状态 TP 環度类型 責任人 玄明樂型 工作空 业务日期 左时时间 周期序号 运行成功 目頃度 勇任 周期实例 2022-01-19 2022-01-20 00-21:00 1 运行成功 日頃度 ● 周期实例 2022-01-19 2022-01-20 00-21:00 1	正行状态 17 環境業型 責任人 変例樂型 工作空 迎身日期 運射时间 周期序号 運行次数 旅行成功 日明度 一 周期実例 こ 2022-01-19 2022-01-20 00:21:00 1 0 旅行成功 日明度 一 周期実例 こ 2022-01-19 2022-01-20 00:21:00 1 0

● 定时检查

定时检查将为您检查当前任务的定时时间是否已到。仅当上游依赖检查通过时,才会触发该检查。

日 日本 日本 伝見 影响基线	历史实例				×
🕑 上游依赖		📀 定时检查	《 调度资源	♥ 任务执行	
2022-04-14 00-20:15 ~ 2022-04- 14 00:20:19 (dur 4s) Gatewaycn- shanghal.ig1ge518isq55dic99jxuhé 运行规功	宠时检查通过				

⑦ 说明 在数据开发模块中定义任务的调度属性时,需要为设置任务在调度环境中的预期运行时

间,但任务实际运行时间可能由于上游任务运行失败等问题导致任务实际执行时间晚于任务定时时间。

● 调度资源

智能诊断的**调度资源**页面,为您展示资源使用趋势。如果当前任务该检查未通过,那么表示当前任务运行时所使用的调度资源不足,任务将进入等待资源状态,直到占用该调度资源的任务执行结束,释放资源后,才会运行当前任务。您可以根据资源使用趋势,合理安排任务的定时时间,以错峰运行。



功能	描述
	为您展示前任务所使用的调度资源组名称、正在该资源组上运行的任务数、 该资源组上正在等待运行的任务数。
调度资源信息	⑦ 说明 公共调度资源组:每天0点~9点是DataWorks的任务高峰 期,公共调度资源紧张,可能出现任务等待资源的情况。您可以错峰运 行调度任务或者前往DataWorks控制台添加独享资源组或自定义资源组 来缓解资源紧张的情况。
资源使用趋势	为您展示当前调度资源组各时间段内的资源使用率,以及当前任务等待资源的耗时情况。
占用资源的任务	为您展示当前任务等待资源的时间段内,占用该调度资源组的任务列表。

● 任务执行

任务执行为您展示当前任务的执行日志、已关联的数据质量规则执行详情及代码详情等信息,对于运行失败的任务,智能诊断功能会根据日志信息,为您提供智能诊断建议,帮助您快速获取任务的出错原因。

全链路诊断 基本信息 影响基线	<mark>夫败</mark> 历史实例				\times
✓ 上游依赖		오 定时检查	☑ 调度资源	8 任务执行	
2022-04-13 21:00:01 ~ 2022-04- 13 21:00:01 (dur 0s) Gateway: 运行失败	 診断信息 日志 Ma 日志診断 「错误信息」: 「错误而因」: 「诊断違议」: 计算资源诊断 正在诊断计算资源 () 	xCompute DQC 代码详情			

~	
〔♪ 注意	仅MaxCompute任务支持查看诊断信息和MaxCompute作业列表。

功能	描述
诊断信息	 日志诊断:获取关键报错信息、报错原因并给出诊断建议。 计算资源诊断:若任务长时间处于等待引擎执行资源的状态下,将在此处给出提示。
日志	为您展示任务的详细执行过程。
MaxCompute	 为您展示MaxCompute作业列表及计算资源详情。 ⑦ 说明 一个DataWorks实例会分成若干个MaxCompute Instance依次执行。 当任务满足运行条件后,DataWorks会根据任务类型将任务下发到对应的计算引擎上执行,当计算资源紧张时,可能会出现等待计算资源的情况,导致任务运行变慢。
DQC	如果任务关联了数据质量规则,在任务运行时,将触发数据质量规则运行,您可在 此处查看数据质量规则详细执行情况。
代码详情	为您展示当前任务的代码详情。

基本信息

您可以在基本信息页面查看当前任务运行的关键时间点和基本信息,相关属性的详细介绍请参见:调度配置。

全链路诊断 基本信息 影响基线	历史实例			×
0	0	0	0	0
上游节点运行完成	到定时时间	等调度资源	开始运行	运行成功
2022-04-13 00:00:10	2022-04-13 03:10:00	2022-04-13 03:10:01	2022-04-13 03:10:04	2022-04-13 03:10:08
名称:	节点ID:		实例ID:	
节点类型: ODPS SQL	责任人:	100 A 100	所属工作空间:	10 p 1 pp (A)
调度类型: -	调度周期: 分	钟调度	调度资源组: 公共调度资源	原组
任务状态: 运行成功	运行时长: -		实例状态: 实例运行成功	
Region: -	引擎类型: M.	AXCOMPUTE	引擎实例:	
出错是否重试: 否	超时定义: 系	统默认 🕜	执行参数: -	
优先级: 8	备注: -			

影响基线

您可以在**影响基线**页面查看将当前任务纳入监控范围的基线列表及基线运行情况。智能基线详情请参见:智能 基线概述。

全链路诊断 基本信息] 影响基线 历史实例						×
工作空间	责任人	基线名称	优先级	基线状态	完成	基线时间	余量
$\sum_{k=1}^{n-1} x ^2+ x ^2+ x ^2$		n, 102,004 (Alasenia, Minal (Sealine		破线	预计:2022-04-14 07:01:43	预警:2022-04-13 03:04:00 承诺:2022-04-13 03:04:00	-3小时57分 钟43秒钟

历史实例

您可以在**历史实例**页面查看如下信息:

- 当前任务各维度指标的变化趋势:通过可视化图表方式为您展示当前任务近15天内运行耗时、开始运行时间、等待调度资源耗时的变化趋势。
- 当前任务的历史实例运行情况:通过列表方式为您展示当前任务的历史实例运行详情,包括实例开始运行时间、结束时间、运行时长、等待资源耗时等。支持您单击操作列的运行诊断跳转至对应实例的诊断详情页面。



3.8. 资源运维

资源运维功能用于监控DataWorks中任务运行所用资源组的使用情况,本文为您介绍如何使用资源运维功能。

背景信息

您可以根据具体业务需求对资源组上运行的实例任务定制监控指标和自动运维规则,资源运维页面会以可视化 方式为您展示资源组使用情况及实例任务执行情况,实现资源组及实例任务的智能监控和自动化运维,减少繁 杂的人工操作,提高运维管理效率。

使用限制

- 当前华东2(上海)、华东1(杭州)地域不支持使用DataWorks的资源运维功能。
- DataWorks的资源运维功能仅支持对独享调度资源组和独享数据集成资源组进行运维。

进入资源运维

1. 登录DataWorks控制台。

- 2. 在左侧导航栏,单击工作空间列表。
- 3. 选择工作空间所在地域后,单击相应工作空间后的进入数据开发。
- 4.
- 5. 在左侧导航栏,选择资源运维,默认显示当前DataWorks使用的独享资源组列表。

查看资源组列表

资源运维页面为您展示当前DataWorks任务运行正在使用的资源组(即运行状态的资源组),此处仅支持展示 独享调度资源组和独享数据集成资源组。您可以根据需求,执行如下操作:

≡	= DataWorks E88-0-0 / U & Castudo / 15826 () U & Castudo / 15826 (
e «	运维大屏 宽时任告运维 >	独享资源组 董君全章				C search		Q
17	mus/csizie v	资源组名称	後型	状态	到期时间	规格	数量	使用率1
6		HZ_ZYJ_DLDUXIANG	款摄集成资源组	运行中	2021-07-10 00:00:00	24c48g	1	0.00%
Ge		HZ_ZYJ_DUXIANG	调度资源组	运行中	2021-07-10 00:00:00	24c48g	1	0.00%
Ŵ	智能监控	ox_schedul	调度资源组	运行中	2021-07-09 00:00:00	8c16g	1	0.00%
	资源运维	zmr_di_060701	调度资源组	运行中	2021-07-08 00:00:00	16c32g	1	0.00%
۲	引動活体 🖌	dd_zyz_01	调度资源组	运行中	2021-06-19 00:00:00	8c16g	2	0.00%
Ŷ,	智能诊断	cx_resource_new	数据集成资源组	运行中	2021-07-08 00:00:00	16c32g	1	0.00%
		dogfood0604002	调度资源组	运行中	2021-07-05 00:00:00	8c16g	2	0.00%
		dogfood0602001	调度资源组	运行中	2021-07-03 00:00:00	16c32g	1	0.00%
		test_wjh	调度资源组	运行中	2021-06-28 00:00:00	12c24g	1	0.00%
		dogfood0525008SCH	调度资源组	运行中	2021-06-26 00:00:00	16c32g	1	0.00%
		dogfood0525005	调度资源组	运行中	2021-06-26 00:00:00	8c16g	1	0.00%
		dogfood0525004	调度资源组	运行中	2021-06-26 00:00:00	16c32g	1	0.00%
		DIIDE_0b7f36353cde4f3996df85cf84f040b4	数据集成资源组	运行中	2021-06-20 00:00:00	4c8g	1	0.00%
		es_test	款提集成 资源组	运行中	2021-06-19 00:00:00	4c8g	2	0.00%
		DIIDE_86c4bbf71c6d4cafaf7b87944295f66e	数据集成资源组	运行中	2021-06-19 00:00:00	8c16g	2	0.00%
		dogfood05140002	调度资源组	运行中	2021-06-15 00:00:00	12c24g	1	0.00%
		dogfood05140001	调度资源组	运行中	2021-06-15 00:00:00	4c8g	1	0.00%
		xm_pre_di_test	款提集成资源组	运行中	2021-06-13 00:00:00	4c8g	1	0.00%
		xm_pre_gateway2	调度资源组	运行中	2021-06-12.00:00:00	4c8g	1	0.00%
		xm_pre_gateway	调度资源组	运行中	2021-06-12 00:00:00	4c8g	1	0.00%
		DIIDE_c1e1a24bfac442c3a48d344357215c44	数据集成资源组	运行中	2021-07-09 00:00:00	4c8g	1	0.00%
		DIIDE_3a29dd10c0c04778a31e85ea049f963b	款攝集成資源組	运行中	2021-07-06 00:00:00	8c16g	2	0.00%
=		molin_prod_di	数据集成资源组	运行中	2021-09-09 00:00:00	12c24g	2	0.00%

● 筛选目标资源组。

您可以在右上方的搜索栏输入所要查找的目标资源组名称,单击Q,即可进行模糊匹配来快速查找目标资源组信息。

● 查看其他资源组。

您可以单击独享资源组后的查看全部,即可跳转至DataWorks管理控制台页面下的资源组列表。

参数名	说明			
资源组名称	创建资源组时自定义的资源组名称。单击资源组名称,可以跳转至资源组详情页面,具体 介绍请参见 <mark>查看资源组详情</mark> 。			
	创建资源组时选择的资源组类型,此处仅支持 数据集成资源组 和调 度资源组 。			
类型	⑦ 说明 资源运维功能仅支持独享资源组。			

参数名	说明
状态	 当前资源组的运行状态,包括: 运行中 已到期 已释放 启动中 创建失败 更新中 更新失敗 已释放 超时
到期时间	当前资源组可用的截止日期。
规格	创建资源组时定义的资源组规格。
数量	创建资源时定义的资源数量。
使用率	资源组当前使用率,单击 使用率 后的[JF],可以对整个资源组列表按照资源使用率从低到 高或从高到低进行排序。

查看资源组详情

资源组详情页面为您展示目标资源组的具体使用情况,包括资源组的基本信息、资源使用率水位曲线和任务列 表,并支持自定义监控规则,帮助您实现资源的智能监控和自动运维,查看和设置运维规则请参见自动运维。

资源组首页 > diResGroup					Í
▲ 读资源组将于8天后到期,请读费					×
diResGroup 近隣道殿型10 期間 8c16g 関際型量1 服务状态运行中 (2021-0 政際型単定支援集構設研究 東京操作	1 16-18 (9)(9)	● 当前使用車 ● 0.79% 単粒	◎ 還行中 0 _个	☑ ¹⁸¹ 八中 4↑	2
還行中的高线同步实例 排队中的高线同步实例 医行中的实时间步实例 更新时间-2021-0	6-09 16:56:16	3		Ē	1动运维
名称 责任人 工作空间	开始同步时间	同步时长	同步速度	同步行数	
M.2.367	381	有政策		< 上一页 1 下一	- 页 〉
资源使用率 100.00%		4	C	2021-06-08 17:01:15 - 2021-06-09 17:01:15	5 🏥
80.00%					
60.00%					
40.00%					
20.00%					
0.00%	• 33	食源使用率			
区域	参数		说明		

区域	参数	说明
1	该区域为您展示资源组的基本信息。	如果您需要修改资源组信息,例如对 资源组进行扩容、缩容等操作,您可 以单击 更多操作 ,即可跳转至 DataWorks管理控制台的 资源组列 表。
2	该区域为您展示该资源组的使用情况,包括 当前使用率 、当前使用该资源组且处于运行中状态的实例数和等资源且处于排队中状态的实例个数。	如果您需要查看为该资源组配置的监 控规则,可以单击 监控 ,即可跳转 至 运维中心>智能监控>规则管理 。
3	可以为您分类展示当前资源组上运行 的所有实例列表。根据运行状态和同 步方式进行实例分类,具体如下: •运行中的离线同步实例 •排队中的离线同步实例 •运行中的离线同步实例	您可以单击 自动运维 ,即可跳转 至运维中心>智能监控>自动运维, 有关自动运维功能具体介绍请参见 <mark>自</mark> 动运维。
4	以水位曲线形式为您直观展示当前资 源组资源使用率。	你可以通过设置起始时间来筛选需要 查看资源使用率情况的时间段。 您还可以将鼠标悬停至水位曲线上, 查看此刻资源利用率。

3.9. 引擎运维

在DataWorks实例任务中, E-MapReduce计算引擎任务会被分为若干作业有序执行。您可以使用DataWorks的 引擎运维功能,查看各个E-MapReduce作业的详细信息,及时查找并清理运行有误的作业,避免该类作业阻 塞下游任务,影响实例任务的正常运行。

使用限制

- DataWorks目前仅支持对E-MapReduce计算引擎进行运维,并且您需要升级E-MapReduce执行包的版本后 才能正常获取运维数据。如果您当前需要使用引擎运维功能请提交工单升级E-MapReduce执行包的版本。
- 如果您的DataWorks项目空间没有添加E-MapReduce引擎,则运维中心左侧导航栏不会显示引擎运维。
- 目前仅华南1(深圳)地域支持使用DataWorks的引擎运维功能。

进入引擎运维

- 1. 进入数据开发页面。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 选择工作空间所在地域后,单击相应工作空间后的进入数据开发。
- 2.
- 3. 在左侧导航栏,选择**引擎运维 > E-MapReduce**,进入E-MapReduce引擎运维页面。

查看引擎作业

E-MapReduce引擎运维页面,为您展示了当前地域,与所选DataWorks项目空间绑定的E-MapReduce引擎, 在DataWorks中下发的E-MapReduce作业的详细信息,您可以根据需求执行如下操作:

☰ 愉 DataWorks 道	途中心	×									& DataStudio	Ø 任务发布 📫 👔	l 🕄 🗢	
🕒 运维大屏	搜索:	EMR > 編入案例D	Q	状态: 講选择	✓ EMF	R作业类型: 请选择	×	•						
	2	作业ID	AH-69	状态	英型	队列	开始时间↓	结束时间 11	责任人	实例类型	工作空间	DataWorks定例ID	所屬DataWorks实例	^ A
13 周期任気运権 >			-	FINISHED	MAPREDUCE	default	2021-03-11 01:32:57	2021-03-11 01:33:22	Sec.	周期实例	10.000	-		10
★ ###55200			inter-	FINISHED	MAPREDUCE	default	2021-03-11 01:31:56	2021-03-11 01:32:20	-	周期实例			100000000000000000000000000000000000000	-111
 9 引擎运施 				EINIGHED	MADREDUICE	dalarda	2021 02 11 01 21 22	2021 02 11 01 21 42	to be detailed	BIDT N	and the local data	-		- 11
E-MspReduce				rinianeo	MAPREDUCE	Gerauit	2021-03-11-01:31:33	2021-03-11-01.31.943		(1999) (1990) (1990)				-
い · · · · · · · · · · · · · ·				FINISHED	MAPREDUCE	default	2021-03-11 01:30:55	2021-03-11 01:31:19	2.252	周期实例	a da se dad	and a second second		_
		100000000000000000000000000000000000000	inite a	FINISHED	MAPREDUCE	default	2021-03-11 01:29:53	2021-03-11 01:30:18	diam'n.	周期实例	10000	10.000	100000000000000000000000000000000000000	
			ining.	FINISHED	MAPREDUCE	default	2021-03-11 01:30:30	2021-03-11 01:30:40	-	周期实例	******		-	
			10100	FINISHED	MAPREDUCE	default	2021-03-11 01:29:29	2021-03-11 01:29:39	In Special	周期实例	plexity	-	-	
			-	FINISHED	MAPREDUCE	default	2021-03-11 01:28:49	2021-03-11 01:29:16	1415	周期实例	and the second s	-		
			laine.	FINISHED	MAPREDUCE	default	2021-03-11 01:27:45	2021-03-11 01:28:10	1000	周期实例	-	10.00		
		-	Telline I	FINISHED	MAPREDUCE	default	2021-03-11 01:27:48	2021-03-11 01:28:13	in the second	周期实例	and other	0.0000		
		-	100	FINISHED	MAPREDUCE	default	2021-03-11 01:28:24	2021-03-11 01:28:35	1227	周期实例	allowing the	-		
		10000-00010-000-000	Testing .	FINISHED	MAPREDUCE	default	2021-03-11 01:27:20	2021-03-11 01:27:30	1.00	周期实例	10000	10.00	1000	
		-	laine-	FINISHED	MAPREDUCE	default	2021-03-11 01:27:23	2021-03-11 01:27:33	-	周期实例	10.00		-	
			10100	FINISHED	MAPREDUCE	default	2021-03-11 01:26:14	2021-03-11 01:26:24	In Section	周期实例	phone			
			-	FINISHED	MAPREDUCE	default	2021-03-11 01:26:43	2021-03-11 01:27:09	1.00	周期实例	10000	-	and the second second	
	4								in lands	Name and Address of the				
=	终止	医行对症的DataWorka实例 3						每页显示: 40			1234	28 下一页 >	1/28 到篇 页 👌	112 v

● 筛选目标E-MapReduce作业。

您可以在E-MapReduce引擎运维页面的顶部菜单栏,根据不同维度进行筛选,快速查看目标作业。

类别	描述
	您可以通过E-MapReduce的 作业ID 查找目标作业,或通 过DataWorks的 实例ID ,查找该实例任务下的所有EMR 作业。
实例ID	 ⑦ 说明 • 引擎运维页面默认为您展示最近三天的引擎实例数据。 • 通过E-MapReduce的作业ID或DataWorks的实例ID搜索时,仅支持展示最近30天的引擎实例数据。

类别	描述
	 NEW:表示E-MapReduce作业刚创建。 NEW_SAVING:表示正在保存E-MapReduce作业。 SUBMITTED:表示E-MapReduce作业已提交执行申请。 ACCEPTED:表示调度程序已接受E-MapReduce作业的执行申请。 RUNNING:表示E-MapReduce作业正在运行。
状态	⑦ 说明 如果目标作业长期处于运行状态, 您可以手动终止运行目标作业所在的DataWorks 实例任务,防止运行有误的作业长期占用资源, 阻塞下游任务,详细内容请参见下文终止运行Da taWorks实例任务。
	 FINISHED:表示E-MapReduce作业已运行结束。 SUCCESSED:表示E-MapReduce作业运行成功。 FAILED:表示E-MapReduce作业运行失败。如果作业呈现该运行状态,则您需要及时识别并处理相关报错,以免阻塞下游作业,影响任务的正常运行。 KILLED:表示由作业的执行者或管理员终止了E-MapReduce作业的运行。
EMR作业类型	当前仅支持查看MAPREDUCE及SPARK类型的作业。

- 查看目标作业的详细信息。
 - 您可以查看E-MapReduce作业的基本信息,包括作业Ⅳ、作业状态、开始运行时间、结束运行时间、所属DataWorks实例及实例Ⅳ等。

⑦ 说明 不同E-MapReduce作业所属的DataWorks实例名称可能相同,但如果不同E-MapReduce 作业的运行**开始时间**不同时,则被认为属于不同的DataWorks实例任务。您可以通过作业所属的DataWorks实例ID判断任务是否属于同一个DataWorks实例任务。

- 通过作业的运行状态,识别运行出错的作业(例如,状态为FAILED的作业为运行失败的作业),单击作 业ⅠD或所属DataWorks实例名称,进入相应任务详情界面,确认并处理报错。
- 您可以对开始时间或结束时间,进行升序或降序排列,直观的查看作业的运行先后顺序及运行时长,了 解目标作业的运行情况。
- •终止运行DataWorks实例任务。

如果E-MapReduce作业长期处于运行状态,即作业状态显示为RUNNING,则您可以勾选目标作业,单击引 擎运维页面左下角的终止运行对应的DataWorks实例,终止运行目标作业所在的DataWorks实例任务。 该类作业往往是由于运行内部出错,导致作业不能自动终止,防止该作业长时间占用资源,阻塞并影响其 他作业的运行,则需要手动将其停止,及时处理相关报错。

↓ 注意

- 当多个E-MapReduce作业属于同一个DataWorks实例任务时,当终止运行任意E-MapReduce作业时,都会导致DataWorks实例任务的运行状态置为失败。
- 当前仅支持对运行中的DataWorks实例任务进行终止运行操作。
- 执行终止运行操作后,目标实例任务的状态将被置为失败。置为失败状态后,当前实例任务会 阻塞该任务的下游节点运行。请知悉风险并谨慎操作。

3.10. 操作历史

操作历史页面展示用户在运维中心各页面上的操作记录。

使用限制

- 仅记录在当前工作空间下的运维操作。
- 仅支持展示三十天内的操作记录。目前记录的操作如下:
 - 任务:下线、冻结、解冻操作。
 - 实例:冻结、解冻、终止运行、强制重跑下游操作。
 - 基线:新建、开启、关闭、修改、删除操作。
- 仅工作空间的运维人员、空间管理员、租户管理员可以查看此页面。
- 仅记录页面操作,不记录OpenAPI操作。

操作步骤

- 1. 进入操作历史页面。
 - i. 进入运维中心。详情请参见: 进入运维大屏。
 - ii. 单击左侧导航栏中的操作历史。
- 2. 进入操作历史页面后,您可以查看相关操作历史详情。

≡	🅅 DataWorks 生产1	10000 000000 20000 000000000000000000000	• •				🕂 🔍 简体	
¢	运维大屏	操作历史 ●						
\$ \$\$	实时任务运维 🖌 🖌 🖌	- 操作对象: 任务 3	立例 🔽 基线 🛛 拳型:	其线-新建 ⊻ 其线-开启	√ 已洗择 5/5 项 ∨ 操	作人:	×	
t,	周期任务运维 🗸 🗸	操作日期: 2022-03-09	2022-03-22 🟥 😋 🖡	新新			0	
ନ	手动任务运维 🗸 🗸	+8.44-0+1/2	+日//	+8.1/- 1	NETZ 0+A	42-+-	18.00	
᠕	智能监控 🖌 🖌	2022-01-10 17-15-59	操作突空	操作人	IV IX ASIAN	10,354	」 第1日 第1日 第1日 第1日 第1日 第1日 第1日 第1日	
	资源运维	2022/01/10 17:10:00	1000			380 F/4690	三百 日十四	
۲	引擎运维 🗸 🗸							
Ŷ,	智能诊断							
ଜ	操作历史							
							< 上一页 1 下-	-页 >

操作	描述
筛选	在上图的模块1区域,您可以通过筛选条件过滤出需要查询的操作历史。 您可以根据 操作对象、类型、操作人、操作日期 等条件进行精确筛选。目前 支持查看操作历史的操作类型包括: • 任务:下线、冻结、解冻操作。 • 实例:冻结、解冻、终止运行、强制重跑下游操作。 • 基线:新建、开启、关闭、修改、删除操作。
查看详情	在上图的模块2区域,以列表样式为您展示操作记录的操作时间、操作类型、操作人、涉及节点、状态等信息。 您还可以单击操作列的查看详情,查看相关操作的结果,支持您单击任务名称,跳转至节点所在的任务列表、实例列表或数据开发页面。

3.11. 常见问题

3.11.1. 数据未正常产出或任务未运行如何处理

问题现象

在数据地图查看数据时,发现当天数据未正常产出,或者在运维中心的周期实例发现任务节点状态为灰色的未运行状态,如下图所示。

0	workshop_start 虚拟节点	
	Ļ	
۲	OSS数据同步 _{演线同步}	
Θ	ods_log_info_d	

可能原因

未运行节点的上游节点未成功运行,上游节点实例可能被冻结,或还在等待运行。上游节点未成功运行,则下 游节点不会启动运行。

解决方案

您需要点开任务DAG图,选中实例后右键查看父节点,查看上游节点实例状态。

• 上游节点实例状态: 紫色

说明上游节点实例被冻结了,上游任务将不会运行,且阻塞下游任务。此种情况下,您可以在实例详情的 操作记录中查看相关记录,了解上游节点冻结的操作记录。

• 上游节点实例状态: 黄色

说明上游节点在等待运行中:

- 上游节点可能设置了定时运行,但是还未到定时时间,上游节点当前在等待任务定时时间。此种情况下 您可以单击右下角的查看实例详情,看上游节点的定时时间,待到了定时时间后上游节点即开始运行, 上游节点运行完成后下游节点即启动运行。
- 上游节点可能在等待调度资源,您可在运行日志中查看是否正在等待调度资源。若出现此种情况,说明 当前项目下正在运行的任务占用的调度资源已达到可使用的资源上限,您可重新安排任务运行时间将各 任务错峰运行,或购买独享调度资源组,独享调度资源组详情可参见独享调度资源组概述。
- 上游节点实例状态: 灰色

说明上游节点也未运行,您需继续查看上游节点的父节点状态。

3.11.2. 找不到任务实例如何处理

问题现象

提交发布任务后,在运维中心的实例列表页面,找不到任务对应的实例。

可能原因

找不到任务对应的实例则说明实例未成功创建。DataWorks会根据任务节点中的调度配置中**生成实例方** 式和调度依赖的配置结果,使用不同方式为您创建运行节点任务的实例。实例未成功创建的可能原因如下。

节点成为孤立节点时(即节点的调度依赖中,没有配置任何依赖的上游节点),则无法正常创建任务实例,在实例列表中找不到对应的实例。

此类场景通常发生在创建好节点和业务流程后,后期修改了节点的依赖配置,删除了节点的**依赖的上游节** 点配置结果后提交发布,导致节点成为孤立节点,无法生成对应的实例。

- 生成实例方式配置为T+1次日生成时:
 - 如果您提交发布的时间早于23:30,则DataWorks会在当天的23:30为您创建实例,提交发布后的第二 天可以在运维中心的实例列表页面找到对应实例。
 - 如果您提交发布的时间晚于23:30,则DataWorks会在提交发布后的第二天的23:30为您创建实例,第
 三天可以在运维中心的实例列表页面找到对应实例。

如果您在实例创建成功前在运维中心的实例列表中查看实例,则找不到对应的实例。

• 生成实例方式配置为发布后即时生成时:

⑦ 说明 旧版工作流不支持将生成实例方式配置为发布后即时生成。

- 提交发布的时间晚于23:30时,则DataWorks不会创建实例,此种情况下实例无法成功创建,您在实例 列表中找不到对应的实例。
- 提交发布的时间早于23:30时,则DataWorks会即刻为您创建任务实例,您可以在实例列表个找到对应 的实例。
 - 任务定时运行时间在提交发布时间的十分钟之后(例如提交发布时间是18:00,定时运行时间是18: 30),则DataWorks会正常创建实例运行任务,您可以在实例列表中找到对应的实例。
 - 任务定时运行时间距离提交发布时间不足十分钟(例如提交发布时间是18:00,定时运行时间是18:05),则DataWorks会生成一个已完成的实例,节点实例属性为实时生成的过期实例。

此种场景下,如果您提交发布的时间晚于23:30,则找不到对应的实例。但是需要注意找得到实例时,实例是否是过期实例。

解决方案

1. 检查节点的调度依赖配置,确保节点不是孤立节点。

在数据开发页面的**调度依赖**配置中,查看各个节点的**依赖的上游节点**是否有配置内容,确保节点不是孤 立节点。

2. 检查节点的提交发布时间,在23:30之前提交发布节点。

在**发布中心的发布包列表**页面,查看发布时间,如果发布时间晚于23:30,您可以在第二天的23:30 之前重新提交发布节点。

⑦ 说明 如果节点的生成实例方式配置为发布后即时生成,您还可以再检查一下任务定时运行时间与提交发布的时间间隔是否超过十分钟,如果没有的话,则需要重新提交发布,确保定时运行时间在提交发布时间的十分钟之后,否则生成的实例为过期实例,无法正常运行节点任务。
4.发布中心 4.1.发布中心概述

DataWorks的发布中心为当前任务发布功能的升级版,主要解决多环境间节点、函数、资源、组件等对象的发 布问题。通常,当目标端需要创建与原端相同的对象时,可以借助该功能快速实现对象的发布。本文为您介绍 发布中心的应用场景、逻辑说明、发布流程等内容。

应用场景

基于不同的发布环境,DataWorks的发布中心可分为跨工作空间发布和跨云发布两种场景,具体如下:

• 跨工作空间发布

跨工作空间发布为原跨项目克隆功能,主要用于同地域同一阿里云账号下,不同工作空间的节点、函数、 资源、组件等对象的发布。

由于简单模式工作空间只有一个环境,不区分开发环境和生产环境,不能直接隔离开发和生产。如果您需要实现开发和生产隔离,则可以使用**跨工作空间发布**功能,将用于开发测试的简单模式工作空间中的任务 发布至用于生产调度的简单模式工作空间,实现简单模式工作空间开发环境和生产环境的隔离。跨工作空间发布常用的场景如下:

- 将目标对象从一个简单模式的工作空间发布至另一个简单模式的工作空间。
- 将目标对象从一个简单模式的工作空间发布至另一个标准模式的工作空间。
- 跨云发布

跨云发布主要用于跨账号、跨地域或跨云平台(例如,金融云、政务云等)节点、函数、资源、组件等对 象的发布。其本质也是将节点由原工作空间迁移发布至目标工作空间,只是两个工作空间所归属的地域、 账号、云平台不同。

逻辑说明

发布中心的发布操作是以节点为单位进行的,文件夹、业务流程、以及节点的依赖关系会同步发布至目标工作 空间。如果节点存在依赖关系,则必须在上游节点成功发布至目标工作空间后,下游节点才能执行发布操作。 发布节点的变更情况如下:

- 节点发布时,系统会更改所有前缀为原工作空间名称的节点输入输出,将其前缀由原工作空间名称更换为目标工作空间名称。同时,根据节点是否存在跨工作空间依赖,您可以基于业务需求配置发布环境的依赖 映射参数,不同配置情况,节点发布后的上下游依赖及输入输出名称不同,详情请参见下文的节点依赖关系及输入输出变更。
- 发布MaxCompute引擎类任务,系统会修改任务代码中涉及的原工作空间名称为目标工作空间名称,详情 请参见下文的MaxCompute引擎类任务的代码变更。

? 说明

- 配置依赖映射关系,详情请参见配置发布环境,配置调度依赖,详情请参见配置调度依赖。
- 本文均以 工作空间名称,节点名称 格式的输出名称示例,具体输出名称请以实际使用为准。

节点依赖关系及输入输出变更

不同配置场景的节点发布后的变更情况如下。

• 不存在跨工作空间依赖, 节点发布前后输入输出及依赖关系变更



- o task_A的输入名: 由 project1_root 变更为 project2_root 。
- task_A的输出名: 由 project1.task_A_out 变更为 project2.task_A_out 。
- 存在跨工作空间依赖,但未设置跨工作空间依赖映射,节点发布前后输入输出及依赖关系变更



发布后, 节点的变更如下:

- 节点的输入输出: 所有的节点输入输出名称中, 涉及的 project1 前缀变更为 project2 。
- 节点的跨工作空间依赖:原 project1.task_A 跨工作空间依赖的 project3.task_A ,发布后, pro
 ject2.task_A 依然跨工作空间依赖 project3.task_A 。
- 存在跨工作空间依赖,设置了跨工作空间依赖映射,节点发布前后输入输出及依赖关系变更



发布后, 节点的变更如下:

- o 节点的输入输出:所有的节点输入输出名称中,涉及的 project1 前缀变更为 project2 。
- • 节点的跨工作空间依赖:原 project1.task_A 跨工作空间依赖的 project3.task_A ,发布后, project2.task_A 跨工作空间依赖变更为依赖 project4.task_A 。

MaxCompute引擎类任务的代码变更

MaxCompute引擎类任务(例如, ODPS SQL、ODPS Spark等)发布到目标工作空间后,执行任务代码时,系 统会将任务代码中涉及的原工作空间名称更换为目标工作空间名称。



使用限制

跨工作空间发布的限制如下:

• 环境限制

DataWorks仅支持对简单模式的工作空间进行跨工作空间发布,将其发布至其他简单或标准模式的工作空间。

- 权限限制
 - 创建发布环境: 仅具有空间管理员角色权限的成员支持创建及配置发布环境。
 - 创建发布包:具有**开发**角色权限的成员可以创建发布包。
 - 发布发布包:具有运维、部署、空间管理员角色权限的成员可以发布发布包。

工作空间添加成员及授权,详情请参见角色及成员管理:空间级。

- 其他限制
 - **跨工作空间发布**用于同地域、同一阿里云账号下,不同工作空间的发布,该场景不支持跨地域发布。如果您需要跨地域发布,请使用**跨云发布**。
 - DataWorks目前不支持发布旧版业务流程至目标工作空间。请将源工作空间中旧版业务流程的任务迁移
 至业务流程中的某个目录后,再发布该业务流程至目标工作空间。

跨云发布的限制:执行跨云发布的账号需要同时为源工作空间和目标工作空间的成员,并且在源工作空间中 具有**项目所有者**或空间管理员角色权限。工作空间添加成员及授权,详情请参见角色及成员管理:空间级。

发布流程

• 跨工作空间发布

跨工作空间发布的所有操作均在原工作空间执行,发布流程如下图。



i. 创建并配置发布环境, 详情请参见创建并配置发布环境。

由于在目标工作空间发布原工作空间的节点时,目标端是一个全新的发布环境,因此,您需要将原端 的引擎实例、资源组及依赖关系等信息映射至目标端。映射后,当发布发布包时,无需在目标工作空 间进行额外的配置,系统便会将发布节点使用的环境替换为映射的环境信息。

 ii. 基于已创建的发布环境,将待发布对象添加至发布包统一发布。创建发并发布发布包,详情请参见创 建发布包。

⑦ 说明 跨工作空间发布场景,发布包创建完成后系统便直接发布,无需您手动执行发布操作。

iii. 查看发布结果,详情请参见<mark>查看发布包列表</mark>。

• 跨云发布

跨云发布的发布流程如下图。



i. 创建并配置发布环境, 详情请参见创建并配置发布环境。

由于在目标工作空间发布原工作空间的节点时,目标端是一个全新的发布环境,因此,您需要将原端 的引擎实例、资源组及依赖关系等信息映射至目标端。映射后,当发布发布包时,无需在目标工作空 间进行额外的配置,系统便会将发布节点使用的环境替换为映射的环境信息。

ii. 在原工作空间中基于已创建的发布环境,将待发布对象添加至发布包并下载发布包至本地。随后将下载的发布包导入至目标工作空间提交发布。创建并发布发布包,详情请参见创建及发布发布包:跨云。

⑦ 说明 跨云发布场景,发布包创建完成后,您需要手动将发布包下载至本地,再导入目标工作空间进行发布。

iii. 在目标工作空间查看发布结果,详情请参见查看发布包列表。

4.2. 创建并配置发布环境

发布中心用于发布同地域同一阿里云账号下,不同项目(例如,简单模式生产、开发环境)间,以及跨账号、 跨地域或跨云平台(例如,金融云、政务云等)节点、函数、资源、组件等对象。创建发布包前,您需要先创 建相应的发布环境。本文为您介绍如何创建并配置发布环境。

使用限制

DataWorks仅支持具有空间管理员角色权限的成员创建及配置发布环境。如果您的账号需要创建发布环境,则需要授权目标账号空间管理员角色权限,详情请参见添加及管理工作空间成员。

进入发布中心

- 1. 登录DataWorks控制台。
- 2. 在左侧导航栏,单击工作空间列表,选择工作空间所在地域后,单击相应工作空间后的数据开发。
- 3. 单击顶部菜单栏的**三**图标,选择**全部产品 > 数据开发与运维 > 发布中心**,进入发布中心。

⑦ 说明 标准模式工作空间中的节点,在数据开发界面提交后将自动进入任务发布界面。目前标准模式的任务发布,仍需使用任务发布功能,发布中心暂不支持标准模式工作空间开发环境到生产环境的发布。

配置发布环境

在发布中心界面,您可以根据业务需求选择配置跨工作空间发布及跨云发布的发布环境,具体配置如下。

1. 进入发布环境管理。

在发布中心页面,单击左侧导航栏的发布环境配置,进入发布环境管理。

2. 创建发布环境。

在**发布环境管理**页面,单击新建发布环境,创建发布环境。

3. 配置环境信息。

在新建发布环境对话框,配置环境信息。

i. 配置环境基本信息。

新建发布环境	
* 环境名称: * 发布类型: * 目标工作空间: * 发布对象所有者 1 ;	 请 輸入 ● 跨工作空间发布 ● ● 跨云发布 请选择 ◆ 默认 ◆
参数	描述
环境名称	新创建的发布环境名称。
发布类型	 发布环境的类型,取值如下: 跨工作空间发布:原跨项目克隆功能。用于同地域同一阿里云账号下,不同项目间 节点发布的场景。 例如,简单模式工作空间的开发环境和生产环境是分隔的,您可以直接使用跨工作 空间发布,无需逐个创建即可将开发环境的节点批量发布至生产环境中进行发布调度,详情请参见发布任务。 跨云发布:用于跨账号、跨地域或跨云平台(例如,金融云、政务云等)节点克隆 迁移的场景。

送择需要创建发布环境的工作空间。 当发布类型为跨工作空间发布:选择当前登录账号已加入的工作空间(即当前账号为目标工作空间的成员)。 ⑦ 说明 如果您需要添加用户为目标工作空间的成员,详情请参见添加及管理工作空间成员。 ● 当发布类型选择跨云发布:您需要手动输入目标工作空间的成员,详情请参见器置工作空间。 ● 当发布类型选择跨工方空间管理页面,即可查看工作空间的成员。如果您需要添加用户为目标工作空间的成员。如果您需要添加用户为目标工作空间的成员,详情请参见添加及管理工作空间。 ● 跨云发布时,您需要同时为原工作空间及目标工作空间的成员。如果您需要添加用户为目标工作空间的成员,详情请参见添加及管理工作空间成员。 / ① 说明 ● 送入工作空间管理页面,即可查看工作空间的成员。如果您需要添加用户为目标工作空间的成员,详情请参见添加及管理工作空间成员。 / ② 说明 ● 送发布类型选择跨工作空间发布时,需要配置该参数。 未完发布包跨工作空间发布时,需要配置该参数。 表示发布包跨工作空间发布时,需要配置该参数。 表示发布包跨工作空间发布时,需要配置该参数。 表示发布包跨工作空间发布时,需要配置该参数。 表示发布包跨工作空间发布时,需要配置该参数。 表示发布包跨工作空间发布时,需要配置该参数。 表示发布包跨工作空间发布时,需要配置该参数。 表示发布包的方上在空间发布的,在目标工作空间为成员,并有达参包的方子。 / ① 以下: ● 試入: DataWorks将创建发布包的用户置为节点的责任人。 ● 发布包划建人: DataWorks将物建发布包的用户置为节点的责任人。 ● 表示包入方上: DataWorks将执行发布操作(即发布包为用户定为节点的责任人)	参数	描述
		选择需要创建发布环境的工作空间。 ■ 当 发布类型为跨工作空间发布 :选择当前登录账号已加入的工作空间(即当前账号 为目标工作空间的成员)。
目标工作空间 • 当发布类型选择跨云发布:您需要手动输入目标工作空间的名称。 ? 说明		⑦ 说明 如果您需要添加用户为目标工作空间的成员,详情请参见添加及管理工作空间成员。
 「⑦ 说明 进入工作空间管理页面,即可查看工作空间名称,详情请参见配置工作空间。 跨云发布时,您需要同时为原工作空间及目标工作空间的成员。如果您需要添加用户为目标工作空间的成员,详情请参见添加及管理工作空间成员。 	日持工作穴词	■ 当 发布类型 选择 跨云发布 :您需要手动输入目标工作空间的名称。
		 ⑦ 说明 进入工作空间管理页面,即可查看工作空间名称,详情请参见配置工作空间。 跨云发布时,您需要同时为原工作空间及目标工作空间的成员。如果您需要添加用户为目标工作空间的成员,详情请参见添加及管理工作空间成员。
 发布对象所有者 发布包含工作空间发布成功后,在目标工作空间发布包中各节点的责任人。取值如下: 默认:DataWorks优先将节点在原工作空间中的责任人置为第一责任人。如果该责任人非目标工作空间中的成员,则系统会将发布包执行人(即发布该发布包的用户)置为节点的责任人。 发布包创建人:DataWorks将创建发布包的用户置为节点的责任人。 发布包执行人:DataWorks将执行发布操作(即发布目标发布包)的用户置为节点的责任人。 		仅当 发东米刑 进择 跨工作空间发东 时,季更配罟该会数
 默认: DataWorks优先将节点在原工作空间中的责任人置为第一责任人。如果该责任人非目标工作空间中的成员,则系统会将发布包执行人(即发布该发布包的用户)置为节点的责任人。 发布包创建人: DataWorks将创建发布包的用户置为节点的责任人。 发布包执行人: DataWorks将执行发布操作(即发布目标发布包)的用户置为节点的责任人。 		表示发布包跨工作空间发布成功后,在目标工作空间发布包中各节点的责任人。取值如下:
 发布包创建人: DataWorks将创建发布包的用户置为节点的责任人。 发布包执行人: DataWorks将执行发布操作(即发布目标发布包)的用户置为节点 的责任人。 	发布对象所有者	默认: DataWorks优先将节点在原工作空间中的责任人置为第一责任人。如果该责任人非目标工作空间中的成员,则系统会将发布包执行人(即发布该发布包的用户)置为节点的责任人。
		 发布包创建人: DataWorks将创建发布包的用户置为节点的责任人。 发布包执行人: DataWorks将执行发布操作(即发布目标发布包)的用户置为节点 的责任人

ii. 配置引擎实例映射。

用于为原工作空间引擎实例配置映射的目标工作空间引擎实例。配置后,在发布任务时,系统会将原 引擎实例中的节点、资源、函数、组件等相关内容绑定至目标引擎实例中,并且目标引擎实例的配置 信息与原引擎实例一致。

? 说明

- **引擎实例映射**区域至少需要保留一个原引擎实例,并为其配置到目标端的映射关系。
- 请根据您的实际业务判断是否需要在目标工作空间中映射原引擎实例。

引擎实例映射				^				
原工作空间引擎实例		引擎类型	目标工作空间引擎实例	忽略映射 🜖				
No. 1		ODPS	请输入					
参数	描述							
原工作空间引擎 实例	显示原工作至	显示原工作空间绑定的所有引擎实例。						
引擎类型	引擎类型。	引擎类型。						
目标工作空间引 擎实例	选择原工作3 您可以进入目 空间环境信	空间引擎实例需要映频 目标工作空间的 发布 息,查看引擎实例信	射的目标端引擎实例。 中心 > 发布环境配置 页面, 息。进入 发布环境配置 ,详 ⁴	单击右上角的 查看工作 情请参见 <mark>发布环境配置</mark> 。				
忽略映射	开启后,目标 源、函数、约 当原工作空间 擎实例开启 2	示工作空间不会映射; 且件等后续将不会被打 目中引擎实例下的节; 8略映射 。	对应原工作空间的引擎实例, 打包发布。 点、资源、函数、组件等不需	原引擎实例中的节点、资 要发布时,您可以为该引				

iii. 配置资源组映射。

用于为原工作空间资源组配置映射的目标工作空间资源组。配置后,在发布任务时,系统会使用映射的目标资源组代替原资源组运行任务。

? 说明

- DataWorks仅支持配置同类型的资源组相互映射。
- 在映射数据集成资源组时,您需要保证目标数据集成资源组的白名单、网络连通性等配置与原数据集成资源组一致,否则可能会影响数据集成任务的正常运行。数据集成资源组的网络连通配置,详情请参见配置资源组与网络连通。

资源组映射			^						
原工作空间资源组	资源组类型	目标工作空间资源组	忽略映射 🚺						
默认资源组	数据集成资源组	请选择 >							
公共调度资源组	调度资源组	请选择 >							
参数	描述								
原工作空间资源 组	显示原工作空间绑定的所有资源组	显示原工作空间绑定的所有资源组。							
资源组类型	资源组类型。								
目标工作空间资 源组	选择原工作空间资源组需要映射的 您可以进入目标工作空间的 发布 口 空间环境信息,查看资源组信息	^{太游电关主。} 							
	开启后,目标工作空间不会映射家 作空间的资源组。	寸应原工作空间的资源组 <i>,</i> 任务发布	5时,仍然使用原工						
忽略映射	⑦ 说明 默认资源组和公共	资源组不支持使用 忽略映射 功能。							
	当原工作空间中的资源组数量比目标工作空间的资源数量多时,您可以为多出的资源组 开启 忽略映射 。开启后,任务发布时,将继续使用原资源组运行任务。								

iv. 配置依赖映射。

该区域显示原工作空间下存在的所有跨项目依赖的上游项目名称,您可以为该项目配置与其映射的目标项目。配置后,任务发布时,系统在目标工作空间中会根据该映射关系,将目标项目配置为发布节点的上游依赖,生成新的上下游依赖关系。节点克隆时的依赖关系说明,详情请参见逻辑说明。

? 说明

- 依赖映射区域默认为您展示 T+1 天自动更新的依赖,如果您需要查看实时更新的依赖,请单击刷新依赖。
- 如果您配置了跨项目依赖映射至新的项目, DataWorks将根据映射关系修改节点的依赖关系, 请确保映射的工作空间可正常添加依赖。常见的无法正常添加依赖的情况, 请参见提交节点报错:当前节点依赖的父节点输出名不存在。
- MaxCompute引擎类任务(例如, ODPS SQL、ODPS Spark等)配置依赖映射后,系统会同步修改该任务的任务代码。详情请参见逻辑说明。

(依赖映射 🙃 ^								
Ⅰ 以下信息T+1自动更新, 対	四需立即刷新,请点击刷新依赖							
原项目名	项目属性	目标项目名称	忽略映射 🕕					
	MaxCompute	请选择 >						
Sec. Sec.	MaxCompute	请选择						
	MaxCompute	请选择 >						
参数	苗述							
原项目名称	显示原工作空间下所有跨项目	还 .示原工作空间下所有跨项目依赖的项目名称。						
项目属性	项目类别。							
目标项目名称	MaxCompute 講師 〇 MaxCompute 講師 〇 MaxCompute ご 〇 MaxCompute ご MaxCompute ご							
忽略映射	开启后,任务发布时,系统将	忽略该项目的依赖映射,继续使用。	原项目作为跨项目依赖					

的项目,发布节点的上下游依赖关系保持不变。

v. 配置数据源映射。

用于为原工作空间的数据源配置映射的目标工作空间数据源。配置后,任务发布时,发布任务中使用的数据源将变更为映射的目标数据源。

⑦ 说明 仅当发布类型选择跨工作空间发布时,需要配置数据源映射。

数据源映射			^						
原数据源名称	数据源类型	目标数据源名称	忽略映射 🚺						
dige (inc.	MaxCompute	请选择							
参数	描述	描述							
原数据源名称	显示原工作空间下所有数据源	显示原工作空间下所有数据源。							
数据源类型	数据源类型。								
目标数据源名称	选择原数据源需要映射的目标	示数据源。							
忽略映射	开启后,任务发布时,系统将	\$忽略该数据源,不会将其映射至	目标工作空间。						

vi. 在备注区域, 配置发布环境的描述信息。

4. 单击新建,成功创建发布环境。

管理发布环境

发布环境管理页面为您展示了已创建的发布环境列表,您可以查看环境名称、状态、创建人等信息。

≡	ntaWorks	发布中心								🔮 🔍 mak 💎	-
6 M		发布环境管理							五治	工作空间环境信息 新建发布	朝和帝
日发	和包列表	环境名称		状	<u>تە</u>	创建人	更新时间	目标空间名称	香注	操作	
ଶ ≋	5环境配置	发布环境2_跨云_xc_dpe_e2		启	用	-	2022年3月1日 14:44:16	xc_dpe_e2	時云发布测试	查看 修改 売隆 禁用 	###
		testhan_跨云		R	170	1.000	2022年3月1日 14:39:20	xc_simple_3h_backup		查看 修改 克隆 蔡用	250
		发布环境1_跨空间_xc_simple_3h_backup	q	启	明	100000.00.000 100	2022年2月28日 18:04:38	xc_simple_3h_backup	测试	查看 修改 克隆 禁用	#19:

同时,可以对已有环境执行如下操作:

- 查看:查看发布环境的配置信息。
- 修改:修改发布环境的配置信息。
- 克隆: 快速生成一个与当前环境配置相同的发布环境。
- 禁用:如果当前环境近期无需使用,您可以将其禁用。禁用后,开发人员在创建发布包界面将无法再基于 该发布环境打包发布节点。
- **启用**:重新启用禁用状态的发布环境。
- 删除:删除废弃使用的发布环境。

⑦ 说明 环境删除后将无法恢复,请您谨慎操作。

后续步骤

发布环境创建完成后,您可以进入**创建发布包**页面创建发布包,详情请参见<mark>创建发布包</mark>及创建及发布发布 包:跨云。

4.3. 创建发布包

本文为您介绍如何在源工作空间创建发布包。

创建发布包

- 1. 创建发布包。
 - i. 在发布中心页面,单击左侧导航栏的**创建发布包**,进入创建发布包页面。

此页面展示的是当前工作空间下已提交的节点,已提交的节点均可在该页面被打包发布。

ii. 添加目标节点至待发布列表。

DataWorks		×									🖞 🔍 💎 🖬 🖓 🗄
三 公開設務知	创建发	就在包 剑试发布现	5394 🗐								每入发布包 特发布列数 ¹⁰
Ⅲ 发布包列表 【 发布环境配置	解决方案 对象类型	5: 谢这师 1: 请选择		 ✓ 业务流程: 消 ✓ 交更共型: 消 	选择 选择	× ×	描交人: 请选择 据交时间: 却始日期	- 始來日期	> 节点D:	清能入	2
		ID	8称	节点版本号	提交人	対象與型	交更关型	节点状态	播交时间	开发环境测试	獲作 3
		12014246	sd_shell_node	2	dataworks_3h1_1	Shell	更新	已提交,侍发布	2021年10月13日 11:28:0 7		查看 发布 从待发布包修除
		12014212	sd_node1	3	dataworks_3h1_1	ODP5 SQL	更新	已爆交,待发布	2021年10月13日 11:26:1 6		董書 发布 从待发布包移除
		12013755	sd_v1	1	dataworks_3h1_1	虚拟节点	更新	已趨交,待发布	2021年10月13日 10:39:0 7		查看:发布:从待发布包移除
		12013749	sd_bccagg.jar	1	dataworks_3h1_1	JAR	更新	已提交,待发布	2021年10月13日 10:38:3 3		查看 发布 从得发布包移除
		12011779	n_test_sql2	1	dataworks_3h1_1	ODPS SQL	更新	已爆交,待发布	2021年10月12日 17:23:2 8		查看 发布 从侍发布包修除
		12010955	n_test_sql	1	dataworks_3h1_1	ODP5 SQL	更新	已爆交,特发布	2021年10月12日 16:43:0 1		查看 发布 从 行发布包移除
		12010767	n_test_shells	1	dataworks_3h1_1	Shell	更新	已提交,待发布	2021年10月12日 16:41:3 4		查看 发布 从侍发布包移除
		11988474	sql_093003	1	dataworks_3h1_1	ODPS SQL	新增	已建交、待发布	2021年9月30日 16:22:50		查看 发布 从侍发布包修除
		11988473	sql_093002	1	dataworks_3h1_1	ODPS SQL	814 818	已盛交,特发布	2021年9月30日 16:22:46		重着 发布 从将发布包移除 查看 发布 从将发布包移除
	化化	系加到发布包 · 待发布列表	2 批量发布						< 上一页 1 2	3 4 … 6 下—页	> 1/6 到篇 页 确定
区域	ł	苗述									
1	7	生该区域,	选择已创	建的发布	环境。						
2	7	生该区域,	您可以通	过 解决方	案、业务	·流程、对	象类型等	条件进行	搜索,筛	选指定条件	牛的节点。
	7	生该区域,	勾选需要	发布的节	点, 单击	批量添加	到发布包	, 添加目相	示节点至待	寺发布列表	
3		⑦ 说明	月								

- iii. 创建发布包。
 - a. 单击待发布列表。
 - b. 在待发布列表对话框单击创建发布包。
 - c. 配置发布包名称,确认发布环境及发布列表。

会在第二天生效:在 钟后的实例会正常转;	23:30后发着 出或更新;5	ic变更慢作,将会在第三 定时时间在发布时间十分	天生效。 特之内的或者在23:30之后发布的实例不		
				*	
引擎类型		目标工作空间引擎实例	۶J		
ODPS		deploy_test_prod			
EMR		已忽略			
				L	
	目标	工作空间资源组唯一标识	只符		
原组	默认	资源组			
原组	已忽	略			
	公共	调度资源组			
	已忽	略			
				L	
	修改范围		目标项目名		
	DATAWOR	RKS	deploy_test_prod		
	DATAWOR	RKS	已忽略		
E	MAXCOM	PUTE	已忽略		
	引撃共型 ODPS EMR 類组 現組 E	引撃类型 ODPS EMR 目标 開始 開始 開始 開始 開始 日本は、 日本は	引擎装型 ODPS ODPS EMR	引撃装型 目标工作空间引擎尖例 ODPS deploy_test_prod EMR 已忽略 ITFC中空间资源组唯一标尺符 開紙 目标工作空间资源组唯一标尺符 開紙 日初略 日初略 公共调度资源组 日初略 日初略 日初略 日初報 日初報 日初期 日初期 日初期	引撃类型 目标工作空间引撃实例 ODPS deploy_test_prod EMR 已忽略 第社 第社资源组 第社 日标工作空间资源组唯一标识符 開組 第社资源组 日本 日本 日本 日本 日本 日本 開組 日本 日本 日本 <t< td=""></t<>

d. 单击创建发布包,完成发布包的创建。

4.4. 创建及发布发布包: 跨云

DataWorks支持将需要发布的对象(节点、函数、资源、组件等)批量添加至发布包统一发布。跨云(即跨账 号、跨地域或跨金融云、政务云等云平台)发布场景,发布包创建完成后,要先在原工作空间下载,再将下载 的发布包导入目标工作空间并发布。本文为您介绍该场景如何创建及发布发布包。

前提条件

已创建并配置发布发布包时所需的发布环境,详情请参见创建并配置发布环境。

使用限制

操作账号需满足以下权限要求:

- 操作账号需要有原工作空间的空间管理员、运维或部署角色权限,才能在原工作空间中创建、发布发布包。
- 操作账号需要有目标工作空间的空间管理员、运维或部署角色权限,发布包才能在目标工作空间发布成功。

如果您需要为用户添加相应的角色权限,详情请参见添加及管理工作空间成员。

发布包创建及发布流程

跨云发布的逻辑说明,详情请参见逻辑说明。

跨云发布场景,创建及发布发布包的流程如下:

- 1. 进入原工作空间创建发布包,详情请参见原工作空间:创建发布包。
- 2. 进入原工作空间下载已创建的发布包至本地,详情请参见原工作空间:下载发布包。
- 3. 进入目标工作空间导入下载的发布包并提交发布,详情请参见目标工作空间:导入发布包。

原工作空间:创建发布包

1. 进入创建发布包页面。

在发布中心页面,单击左侧导航栏的创建发布包,进入创建发布包页面。

? 说明

- 此页面展示的是当前工作空间下已提交的节点。已提交的节点均可在该页面被打包发布。
- 仅当节点所在的发布包发布状态为发布成功或发布失败时,该节点才能被再次打包。
- 2. 创建发布包。

在创建发布包页面,将需要发布的对象(节点、函数、资源等)添加至发布包,统一发布。

	发布中心	and the second	• •							ú	9 4, ma 💎 💴
E	创建发	布包 跨云发布环	5 11 ~ 1								导入发布包 特发布列表
8章发布包列表 121发布环境配置	解决方宫 对象类型	: 请选择 : 请选择		 > 业务流程: > 交更典型: 	書选择 書选择	× ×	爆交人: 请选择 爆交时间: 起始日期	- 结束日;	> 市前0:	调输入	2
		ID	名称	节点版本号	描交人	对象类型	交更关型	节点状态	揭交时间	开发环境测试	^{接作} 3
		10000036273	RDS_数据同步	1	0.000	高线同步	新聞	检查通过	2022年3月16日 11:03:22	未测试	查看 发布 从持发布包移除
		10000015021	ip2region.jar	2	0.000	JAR	更新	检查通过	2022年3月16日 11:03:22		查看 发布 从持发布包移除
		10000036272	workshop_start	1	0.000	虚拟节点	新増	已謂交,侍发布	2022年3月16日 11:03:22	未测试	查看 发布 从待发布包移除
		10000036274	zy_demo_跨项目克隆	1	0.000	ODPS SQL	新增	检查通过	2022年3月16日 11:03:22	未測试	查看 发布 从师发布包移除
		10000015022	getregion	2	0.000	函数	更新	检查通过	2022年3月16日 11:03:22		查看 发布 从侍发布包移除
		10000015028	rpt_user_info_d	2	0.000	ODPS SQL	更新	检查通过	2021年12月22日 21:39:46	未測试	查看 发布 从侍发布包移除
		10000015027	dw_user_info_all_d	2	0.000	ODPS SQL	更新	检查通过	2021年12月22日 21:39:42	未则试	查看 没布 添加到待没布包
		10000015026	ods_log_info_d	2	0.000	ODPS SQL	更新	检查通过	2021年12月22日 21:39:37	未阅试	查看 发布 添加到待发布包
		10000015025	rds_数据同步	2	0.000	嘉线同步	更新	检查通过	2021年12月22日 21:39:33	未测试	查看 发布 清加到待发布包
		10000015024	oss_数据同步	2	0.000	离线同步	更新	检查通过	2021年12月22日 21:39:29	未到试	查看 没布 添加到待没布包
	北副石	1982年48日 - 特没布列表									上一页 1 2 下一页 >

待发布	列表					
待发布	6项					创建发布包
ID	名称	提交人	对象类型	变更类型	节点状态	操作
4310	4 RDS_数据同步	distante (secolo	高线同步	新増	检查通过	查看 移除
4310	ip2region.jar	100000-0000	JAR	更新	检查通过	查看 移除
4310	2 workshop_start	100000-000	虚拟节点	新增	已提交,待发布	查看 移除
4310	5 zy_demo_跨项目亮隆	description (second	ODPS SQL	新増	检查通过	查看 移除
4310	8 getregion	And a second second	函数	更新	检查通过	查看 移除
3288	3 rpt user info d	States and states	ODPS SOL	更新	检查通过	音吾 移除

i. 选择发布环境。

在区域1,选择待创建的发布包需要使用的跨云发布环境。

ii. 筛选指定条件的发布对象。

在区域2,您可以通过**解决方案、业务流程、对象类型**等条件进行搜索,筛选指定条件的发布对象。

iii. 添加目标对象至发布包。

在区域3,勾选需要发布的对象,单击**批量添加到发布包**,添加目标对象至待发布列表。添加后, 在**待发布列表**按钮上会显示待发布对象的数量。

 ⑦ 说明 如果您需要发布的对象个数较少,也可以直接单击目标对象操作列的发布,进入发 布确认页面执行后续发布流程。

iv. 确认发布信息并创建发布包。

- a. 查看待发布列表。
 - 查看列表概况:单击待发布列表,即可查看发布列表中待发布对象的ID、对象类型、变更类型、节点状态等基本信息。
 - 查看对象详情或移除对象:单击目标对象操作列的查看或移除,即可查看对象的详细信息或 从发布包中删除该对象。
- b. 确认发布信息。

单击创建发布包,在发布确认对话框,输入发布包名称,检查发布包的发布环境及发布列 表等信息。

布确认						×
 1.T+1次日生成突例的节点,在23:30前 2.发布即时生成实例的节点,定时时间还 会生效,将会在第三天生效。 	发布变更操作,会在第二天 E发布时间十分钟后的实例:	生效;在23:30原 会正常转出或更剩	后发布变更操作 新;定时时间在	,将会在第三 发布时间十分	天生效。 分钟之内的或者在23:30之后发布的实例不	
发布包名称: test	8					
发布环境 跨云发布环境 发布列表 引擎实例映射						
原工作空间引擎实例		引撃类型		目标工作空间	间引擎实例	
(1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0		ODPS		in.e.		
资源组映射					^	
原工作空间资源组	资源组类型		目标工作空间]资源组唯一标	示识符	
默认资源组 (八共调度资源)	数据集成资源组					
	49051X (4915				<u>^</u>	
原项目名	项目属性			E	标项目名称	
and produce a pro-	MaxCom	pute		B	忽略	
10.000	MaxCom	ipute		B	忽略	
					创建发布包	关闭

c. 确认信息无误后, 单击创建发布包, 系统将创建该发布包。

发布包创建完成后,您需要在原工作空间的**发布包列表**页面下载发布包,用于后续导入目标工 作空间进行提交发布。

⑦ 说明 发布包创建后,仅支持保留30天,请您及时下载保存。超期后该发布包将无法 下载,您需要重新创建新的发布包。

? 说明

- T+1 次日生成实例的节点,发布生效的时间如下:
 - 如果在23:30前执行发布操作,则发布内容会在第二天生效。
 - 如果在23:30后执行发布操作,则发布内容会在第三天生效。
- 发布即时生成实例的节点,发布生效的时间如下:
 - 定时时间在发布时间十分钟后的实例会正常转出或更新发布操作。
 - 定时时间在发布时间十分钟之内,或者在23:30之后发布的实例,发布内容将 会在第三天生效。

原工作空间:下载发布包

1. 进入发布包列表。

在发布中心左侧导航栏单击发布包列表,进入发布包列表页面。

2. 下载发布包。

在发布包列表页面,您需要将目标发布包下载至本地,用于后续导入目标工作空间。

≡	n DataWorks	发布中心	A DESCRIPTION OF TAXABLE A									ć	e) max 💎 🛛	and a second
	=	发布句及	連											
6 66	1226包	X10 637	XFG											
語 发	的制度	没布人:	请选择	~	发布时间:	请选择日期	#	发布状态:	请选择	~	发布环境:	请选择		~
13 201	东境在置	申请人:	请选择	Ŷ	申请时间:	请远华日期		发布包提索:	请输入没布包名称或D				(1 28
		10	Water An Pr Da	NAME IT 10	Address and state	(h)# (db.2884100		Water I	Marken and Mill		Mahadend Jander	18.0-	
		10	20560	22/03/198	224034538	中調入	100 THE R. (10)		201	254043143		22194033	SRIP	•
		4	workshop再次发布	发布中心-生产	時云发布	daman in planet	2021年12月22日 20:00		- The second second second	2021年12月22日 20	0:06:13	发布成功	豊富 下航	2

i. 筛选目标发布包。

在发布包列表页面,您可以查看发布包的名称、环境、类型等基本信息,并根据发布人、发布时间、发布状态、发布环境等条件筛选目标发布包。

ii. 单击目标发布包操作列的下载。

⑦ 说明 发布包创建后, 仅支持保留30天, 请您及时下载保存。超期后该发布包将无法下载, 您需要重新创建新的发布包。

iii. 在**下载发布包**对话框确认下载信息无误后,单击**确认**,下载发布包。

发布包下载完成后,您需要将下载的发布包导入至目标工作空间进行发布。

目标工作空间:导入发布包

1. 进入创建发布包页面。

进入目标工作空间创建发布包页面的步骤与原工作空间类似,详细步骤请参考创建发布包。

2. 导入发布包。

- i. 单击创建发布包页面右上角的导入发布包。
- ii. 输入**发布包名称**,并选择上传方式。

您可以根据业务需求,选择本地文件上传或OSS上传。

导入发布包	上传方式:本地文件	×	导入发布包	上传方式: OSS上传 ×
▲ 任务首次发布时,会	夏蓋发布工作空间里同名目录下的同名任务,请确认发布。		▲ 任务首次发布时,会	要盖发布工作空间里同名目录下的同名任务,请确认发布。
* 发布包名称:			* 发布包名称:	
上传方式:	 ● 本地文件 ○ OSS上传 		上传方式:	○ 本地文件
选择文件:	上传文件		oss地址:	清縮入 校验
备注:	导入发布包时,支持上传一个发布文件。		备注:	导入发布包时,支持上传一个发布文件。

 本地文件:直接上传已下载的发布包。当待上传的发布包小于或等于30MB时,您可以使用该方式 上传。

? 说明 本地文件最多支持上传30MB大小的文件。

 OSS上传: 您需要先将已下载的发布包上传至OSS进行存储,获取OSS的存储地址。当待上传的发 布包大于30MB时,您只能使用该方式上传。

上传发布包至OSS,详情请参见上传文件,获取OSS存储地址,详情请参见分享文件。

iii. 预览并发布发布包。

发布包上传成功后,您可以单击**发布包预览**,检查发布包的发布环境、发布列表信息。确认无误 后,单击**发布**,发布发布包。

⑦ 说明 如果目标工作空间中存在与创建的发布包同名的发布任务,则创建的发布包首次发布时,会覆盖该同名任务,请您确认无误后再执行发布操作。

后续步骤

跨云发布的发布包发布后,您可以进入当前工作空间的**发布包列表**页面,查看发布结果。详情请参见<u>查看发</u> 布包列表。

4.5. 查看发布包列表

发布包列表页面,为您展示所有已创建发布包的名称、发布类型、发布时间、发布状态等基本信息,您可以根 据发布状态判断发布包的发布情况。或根据业务需求,查看发布包的环境配置、发布对象等详情信息,以及删 除、回滚发布包。本文为您介绍发布包列表的相关内容。

前提条件

已创建并发布发布包, 详情请参见创建发布包及创建及发布发布包: 跨云。

进入发布包列表

- 1. 进入发布中心。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏,单击工作空间列表,选择工作空间所在地域后,单击相应工作空间后的数据开发。
 - ⅲ. 单击顶部菜单栏的**⊟**图标,选择**全部产品 > 数据开发与运维 > 发布中心**,进入发布中心。

2. 进入发布列表。

在发布中心页面,单击左侧导航栏的发布包列表,进入发布包列表页面。

查看发布包信息

在发布包列表,您可以查看所有已创建发布包的基本信息,或根据业务需求执行发布包的删除、回滚等操作。

	发布中心	· ·								ß	4 max ♥
	发布包列表	R.									
	发布人:	请选择	~	发布时间:	请选择日期		送布状态: 请选择		✓ 没布环境:	请访择	~
2 发布环境配置	申请人:	请选择	~	申请时间:	请选择日期		一 发布包搜索: 清编入发布包	1名称歌D			1 ***
	ID 发行	布包名称	发布环境		发布英型	申请人	申请时间	发布人	发布时间	发布状态	强作
	55 tes	it.	发布演示·跨工作	空间发布	間工作空间发布	investories.	2022年3月22日 17:27:05	1001010-0010	2022年3月22日 17:27:05	发布成功	11 BR
	52 tes	it322	默认环境		詞云发布	description in	2022年3月22日 16:49:27	and the second	2022年3月22日 16:54:01	面涼中	±* (2)
	45 wo	irkshop应务等工作空间发布演示	发布演示-跨工作	空间发布	跨工作空间发布	and the second second	2022年3月16日 11:03:16	- Second system of	2022年3月16日 11:03:17	囲滾成功	22
	5 wo	irkshop再次发布	默认环境		跨云发布	and the second second	2022年1月26日 15:02:07	description i	2022年1月26日 15:06:28	发布天奴 ③	
											< 上一页 1 下一页 >
区域	描述	2									
1	1包申••发•••• 。请日发在准名发发9 导 • •	人及发布人的区目, 人及发布人的区目, 有人:指执行之后。 后面的发布状态包括 方面的发布的方面的表示。发布包括 方面的发布的方面的表示。 方面的发布的方面的表示。 方面的发布的方面的表示。 方面的发布的方面。 方面的发布的方面。 方面的发布的方面。 方面的发布的方面。 方面的发布的方面。 方面的发布的方面。 方面的发布的方面。 方面的方面。 方面的方面。 方面的方面。 方面的方面。 方面的方面。 子面的方面。 子面的方面。 子面的方面。 子面的方面。 子面的方面。 子面的方面。 子面面的方面。 子面面的方面。 子面面的方面。 子面面的方面。 子面面的方面。 子面面的方面。 子面面的方面。 子面面的方面。 子面面的方面。 子面面的方面。 子面面的方面。 子面面的方面。 子面面面的方面。 子面面的方面。 子面面面面的方面。 子面面面面面面面面面面。 子面面面面面面面面面面面面面面面面面。 子面面面面面面面面	【别发发儿子己子臣包包失 布 跨 正包】 别布布括 创生在发发败 场 云在成如句操: 建成发布布原 景 发 被功	、 的作 , , 布成失因 下 布 回回 下 用的 正等。功败。 / 场 奚 奚 户月 石 彾 功败。 / 为 奚 奚 3	- 。 - 。 - · · · · · · · · · · · · · · · · · · ·	下包中所有: 下包中有一 E被导入至 包支持该状 之布版本。 之布版本。	对象均发布成 个对象发布失 目标工作空间 态。	;功,该发布 :败,该发布 。	5包才会显示; 5包便会显示;	成功状态 支败状态	。您可以

Dat aWorks

区域	描述
	 在该区域,您可以查看目标发布包的名称、发布环境、发布状态等信息,或对目标发布包执行如下操作: 查看发布包详情:用于查看发布包的发布环境配置信息及发布对象详情。 您可以查看发布包中所有对象的发布状态,当发布包为发布失败状态时,可根据发布包详情定位发布失败的具体对象,快速处理报错。 回滚发布包:用于回滚已发布的发布包至前一个发布版本。 删除发布包:如果当前发布包后续无需使用,则可删除该发布包。 ⑦ 说明 发布包删除后将无法恢复,请您谨慎操作。
2	 下载发布包: 跨云发布场景下, 创建完发布包后, 需要下载当前发布包至本地, 用于导入目标工作空间 进行发布。
	⑦ 说明 仅支持跨云发布环境的发布包执行下载操作。
	 更新发布状态:跨云发布场景下,已发布的发布包需要被再次发布时,如果当前发布包不是发布成 功或发布失败状态,则您需要手动更新发布状态为发布成功或发布失败,否则,该发布包无法被再次发 布。
	⑦ 说明 仅支持跨云发布环境的发布包更新发布状态。

5.开放平台 5.1. 开放平台概览

DataWorks开放平台是DataWorks对外提供数据和能力的开放通道。DataWorks开放平台提供开放 API(OpenAPI)、开放事件(OpenEvent)、扩展程序(Extensions)的能力,可以帮助您快速实现各类应 用系统对接DataWorks、方便快捷的进行数据流程管控、数据治理和运维,及时响应应用系统对接DataWorks 的业务状态变化。

进入开放平台

- 1. 以阿里云账号登录DataWorks控制台。
- 2. 在左侧导航栏,单击开放平台,即查看开放平台下所有功能模块。

DataWorks		NataWorks / 开放平台					
概览	① DataWorks开放平台由OpenAPI (开放API)、OpenEvent (开放满思) 组成,可适用于各类产品集成与被集成的场景。其中,OpenAPI已经商业化,开放满思正在火热公测中,升级至DataWorks企则。						
工作空间列表		OpenADI OpenEuert Extensions					
资源组列表		Openaven Openaven Extensions					
报警配置	▲ ③ 您的 DataWorks企业版 还有 17 天到期,2022年2月11日 起将会停止消息订阅,请及时续费。续费DataWorks企业版 查看开放房意文档						
报警资源		OpenAPI 免费额度 数据日期: 2022-01-23 OpenAPI 调用次数 数据日期: 2022-01-23					
报警联系人		0% 金融版時代 1,000,000 次/日 末日新61,000,000 次 ○次					
开放平台 New		705(80)827 (1/00/00/ ///); 44/3837 (1/00/00/ //					
计算引擎列表	^	OpenAPI 调用明细 数据日期: 2022-01-23					
MaxCompute		日期 2022-01-16 - 2022-01-23 値 API送別 请选择 🗸 API吸別 请选择					
Graph Compute		日期 API名称 API模别 API模别 基于调用次数 本月累计调用次数					
Hologres							
数据湖集成 (湖仓一体)	沒有意词则符合条件的记录						
		く 上一页 1 下-	页 >				

功能简介

● 开发API (OpenAPI)

DataWorks开放平台的OpenAPI功能为您提供开放API能力,通过开放API实现本地服务和DataWorks服务的 交互,提升企业大数据处理效率,减少人工操作和运维工作,降低数据风险和企业成本。您还可以在 OpenAPI区域查看OpenAPI免费额度使用情况、调用次数和调用明细等信息。

关于OpenAPI功能,具体请参见开放API(OpenAPI)。

● 开放事件(OpenEvent)

DataWorks开放平台的开放事件功能为您提供消息订阅服务,通过订阅DataWorks事件状态,应用系统对接 DataWorks,实时获取相关内容的状态变化,帮助您及时响应,满足您的个性化决策需求。关于开放事件功 能,具体请参见。

关于开放事件功能,具体请参见开放事件。

● 扩展程序(Extensions)

DataWorks通过OpenEvent能力为您提供消息推送订阅能力,同时您可以将服务程序注册为DataWorks的扩展程序,通过扩展程序来卡点并响应订阅的事件消息,以此实现通过扩展程序对特定事件进行消息通知与流程管控。

关于扩展程序功能,具体请参见扩展程序。

5.2. 开放API (OpenAPI)

Dat aWorks开放平台提供OpenAPI能力,通过调用Dat aWorks的OpenAPI使用Dat aWorks的各项功能,实现应用和Dat aWorks的集成和交互。

背景信息

DataWorks的OpenAPI提供100多个API功能,通过调用API,可以快速进行批量操作和系统集成对接等工作, 提高数据开发效率,满足企业定制化需求。您还可以通过开放平台,轻松获取OpenAPI的使用情况,包括免费 额度使用情况、调用次数和调度明细等信息。关于OpenAPI详细信息,请参见DataWorks OpenAPI概述。

使用限制

- 可用版本: DataWorks企业版和旗舰版, 详情请参见DataWorks各版本详解。
- 可用地域:华东1(杭州)、华东2(上海)、华南1(深圳)、华北2(北京)、华北3(张家口)、北京政务云、华东2(上海)金融云、西南1(成都)、中国(香港)、亚太东南1(新加坡)、印度尼西亚(雅加达)、美国(硅谷)、印度(孟买)、美国(弗吉尼亚)和德国(法兰克福)、英国(伦敦)、日本(东京)、澳大利亚(悉尼)、马来西亚(吉隆坡)。

查看OpenAPI使用情况

- 1. 以阿里云账号登录DataWorks控制台。
- 2. 在左侧导航栏选择开放平台。
- 3. 在右侧浏览区域单击OpenAPI,即可查看OpenAPI的使用情况。

DataWorks	taWorks / 开放平台					
概范	● DataWorks开放平台由OpenAPI (开放API)、OpenEvent (开放消息) 组成,可适用于各类产品集成与被集成的场景。其中,OpenAPI已经商业化,开放消息正在火热公测中,开级至DataWorks企业版即可参与公测。					
工作空间列表	OpenAPI OpenEvent Extensions					
资源组列表						
报警配置 へ	⑥ 您的 DataWorks企业版 还有 17 天到期, 2022年2月11日 起将会停止涡息订阅, 请及时续费。续费DataWorks企业版 查看开放涡息文档					
报警资源	OpenAPI 免费额度 對限日期: 2022-01-23 OpenAPI 週用次数 對限日期: 2022-01-23					
报警联系人	0%					
开放平台 Rev	免费额度共1,000,000 次月,本月剩余1,000,000 次					
计算引擎列表 ^	OpenAPI 调用明细 数据日期: 2022-01-23					
MaxCompute	日期 2022-01-16 - 2022-01-23 自 API送知 请选择 メ API级别 请选择 メ					
Graph Compute	日期 API名称 API模别 API模别 API模别 当天调用次数 本月累计调用次数					
Hologres						
数据湖集成 (湖仓一体)	沒有查询则符合条件的记录					
	く 上一页 1 下一页 >					

OpenAPI功能区简介

OpenAPI页面可以为您展示OpenAPI的使用情况,具体如下:

- OpenAPI免费额度:可以查看每个自然月截止当前日期API的免费额度使用情况。目前,DataWork企业版每 个自然月会为每个阿里云账号在每个地域提供1,000,000次/月的免费额度,免费额度使用完后,结合版本按 量计费,具体请参考DataWorks OpenAPI概述。
- OpenAPI调用次数:可以查看每个自然月截止当前日期所有OpenAPI的总调用次数,按自然月累计,每月1日清零重新累计。
- OpenAPI调用明细:可以查看特定时间段内API的调用信息,包括调用日期、API类别、API级别、当天调用 次数和本月累计调用次数。您可以通过API类型和API级别筛选所要查找的API信息。关于API名称、API类别 和API级别的详细介绍,请参考DataWorks OpenAPI概述。

5.3. 开放事件(OpenEvent)

5.3.1. OpenEvent概述

DataWorks的OpenEvent旨在将DataWorks的各类状态变更以消息的方式触达到用户,便于用户订阅并作出个性化的响应,例如,您可以通过OpenEvent订阅表变更,从而实现核心表的实时监控;通过OpenEvent订阅任务变更,以定制化开发实时任务大屏。本文为您介绍OpenEvent的实现原理与配置要点。

注意事项

- 目前OpenEvent面向DataWorks企业版和旗舰版用户公测,华北2(北京)、华东1(杭州)、华南1(深圳)、华东2(上海)、华北3(张家口)、华南1(深圳)、西南1(成都)、美国(硅谷)、美国(弗吉尼亚)、德国(法兰克福)、日本(东京)、中国(香港)、新加坡地域支持该功能。
- OpenEvent通过Event Bridge实现消息的订阅与消费,所以您需要先开通Event Bridge产品服务,关于 Event Bridge产品介绍详情请参见: 什么是事件总线Event Bridge。
- 仅支持阿里云主账号或者有AliyunDataWorksFullAccess权限的RAM用户权限的用户可以直接进行界面相关 配置,关于DataWorks权限说明,详情请参见控制台细分权限管控:自定义RAM Policy。

支持订阅的事件类型

DataWorks支持订阅的事件从事件类型上可以分为普通事件和扩展点事件。

- 普通事件: 支持用户订阅此类事件的消息。
- **扩展点事件**:用户订阅此类事件的消息后,还可联合DataWorks的扩展程序功能,在收到事件消息后实现 自定义响应。

DataWorks当前支持订阅的事件覆盖了运维中心、DataStudio和数据治理三个功能模块中的事件,详情如下 表所示。

所属模块	事件细分	事件描述与类型	事件示例
	实例状态变 更事件	 描述:任务在运维中心运行时,实例状态发生变更而触发的事件。 类型:实例状态变更事件为普通事件,您可以订阅此类事件消息,获知事件状态的实时变化情况。 	例如,运维中心的周期实例,从未运行变为运行 中、从运行中变为运行失败。

所属模块	事件细分	事件描述与类型	事件示例
运维中心	节点变更事 件	在运维中心(生产环境)的节 点变更事件,包括普通事 件和扩展点事件。	 普通事件:您可以在本地开发程序用于接收该类型的事件消息,以便您基于自身业务进行个性化响应。 新增节点事件 更新节点事件 则除节点事件 删除节点事件 扩展点事件:兼容普通事件特性,此外,该类型事件还支持将本地程序注册为DataWorks扩展程序,实现扩展程序对扩展点事件的自定义逻辑处理与流程管控。 下线节点前置事件 冻结节点前置事件 解冻节点前置事件 解冻节点前置事件 新结节点前置事件 新结节点前置事件
	补数据操作 事件	 描述:在运维中心(生产环境)的节点进行补数据操作触发的事件。 类型:属于扩展点事件。 	 事件示例如下。 对周期任务进行补当前节点的操作。 对周期任务进行补当前节点及下游节点的操作。 通过海量节点模式进行补数据。 通过高级模式进行补数据。
监控告警	监控告警事 件	 描述:在运维中心中,触发 DataWorks的告警机制的事件。 类型:属于普通事件 	 事件示例如下。 基线告警示例:某个基线上任务预计完成时间太晚,触发基线告警。 事件告警式例:某个基线上任务运行失败,触发事件告警。 规则告警示例:创建了一个针对孤立节点的报警规则后,当工作流中出现了孤立节点,触发了此规则告警。
DataStudio	文件变更事 件	 描述:在DataStudio中,文件(节点、资源或函数)出现状态变更时触发的事件。 类型:属于扩展点事件,您可以订阅此类事件消息,并联合扩展程序的功能,订阅此类事件后实现自定义响应。 	 事件示例如下。 文件提交扩展点事件: 例如提交某个节点、资源或函数。 文件发布扩展点事件: 例如发布某个节点、资源或函数。 节点运行扩展点事件: 例如运行节点代码。 文件删除扩展点事件: 例如删除某个节点、资源、函数。

所属模块	事件细分	事件描述与类型	事件示例		
	表变更事件	 描述:在DataStudio的表管理中,表状态变更时触发的事件。 类型:属于扩展点事件,您可以订阅此类事件消息并 	 事件示例如下。 表提交扩展点事件:例如在表管理中提交某个表。 表发布扩展点事件:例如在表管理中发布某个表。 		
		联合扩展程序的功能,订阅 此类事件后实现自定义响 应。	⑦ 说明 仅标准模式工作空间支持在 表管理中发布表,因此表发布扩展点事件 仅存在于标准模式下。		
数据治理	审批中心事 件	 描述:在数据治理模块中, 创建申请单和完成申请单触 发的事件。 类型:属于普通事件,您可 以订阅此类事件消息,获知 事件状态的实时变化情况。 	 事件示例如下。 创建申请单事件:例如在数据地图中,在权限 管理页面单击申请函数和资源权限,进行申 请单的创建。 完成申请单事件:例如在数据地图中,当访问 账号为管理员时,在权限管理页面单击待我 审批选项卡,审批所有项目下的表、资源和函 数的权限。 		

⑦ 说明 由于网络时延问题, OpenEvent的消息发送可能存在一定延迟。

配置流程

了解上述原理后,您可根据实际情况规划并应用DataWorks的OpenEvent功能,实现关键事件的消息订阅,应用流程与配置要点如下。

- 1. 在事件总线 Event Bridge 控制台开通事件总线 Event Bridge, 创建自定义总线。
- 2. 在DataWorks端开启消息订阅,添加事件分发通道。
- 3. 在事件总线 Event Bridge 控制台配置事件总线。
- 4. 在事件总线 Event Bridge 控制台发布事件,详情请参见发布事件。

应用优势

DataWorks的OpenEvent具备以下优势:

• 节省运维成本:

通过开启DataWorks开放平台的消息订阅功能,您可以很方便的订阅事件消息。

• 提升开发效能:

通过在DataWorks界面上创建相关配置,您无需再进行发送消息相关配置,仅需做好相关SDK开发的准备配置工作及订阅消息配置,节省繁杂的手工操作。

• 实现个性化需求:

Dat aWorks提供丰富的事件类型,您可以通过订阅不同事件类型的消息,快速感知业务状态变化信息,及时 做好个性化响应和决策。

5.3.2. 开启消息订阅

DataWorks的OpenEvent通过EventBridge为您提供事件消息订阅与发送服务。在EventBridge中创建自定义总 线后,您需要在DataWorks中开启消息订阅并配置消息事件分发通道,以便于在EventBridge中创建规则并完 成事件的发布。本文为您介绍如何开启消息订阅服务与配置自定义总线。

注意事项

- 目前OpenEvent面向DataWorks企业版和旗舰版用户公测,华北2(北京)、华东1(杭州)、华南1(深圳)、华东2(上海)、华北3(张家口)、华南1(深圳)、西南1(成都)、美国(硅谷)、美国(弗吉尼亚)、德国(法兰克福)、日本(东京)、中国(香港)、新加坡地域支持该功能。
- OpenEvent通过EventBridge实现消息的订阅与消费,所以您需要先开通EventBridge产品服务,关于 EventBridge产品介绍详情请参见: 什么是事件总线EventBridge。
- 仅支持阿里云主账号或者有AliyunDataWorksFullAccess权限的RAM用户权限的用户可以直接进行界面相关 配置,关于DataWorks权限说明,详情请参见控制台细分权限管控:自定义RAM Policy。

开通EventBridge

- 1. 开通Event Bridge服务,请参考开通事件总线Event Bridge并授权。
- 2. 进入EvenBridge控制台,在左侧导航栏单击事件总线。
- 3. 在页面右上角单击自定义事件总线后的快速创建按钮,按照提示创建自定义总线。

⑦ 说明 您在创建用于DataWorks消息订阅的自定义总线时,先跳过添加事件源和创建事件规则的步骤。

开启事件订阅服务

在进行事件订阅之前,您需要先开启DataWorks上的开放事件服务。

- 1. 进入DataWorks控制台,单击左侧导航栏开放平台。
- 2. 在开放平台界面,单击OpenEvent,选择事件订阅页签,进入事件订阅页面。
- 3. 在事件订阅页签中,单击启用消息订阅,开启DataWorks事件消息订阅服务。

添加事件分发通道

1. 授权DataWorks访问EventBridge资源。

如果您是首次使用,会出现点击授权按钮,并且添加事件分发渠道按钮将置灰,您需要单击点击授权按钮,授权DataWorks访问EventBridge资源,授权后将自动在RAM访问控制的身份管理 > 角色中增加自动创建一个服务关联角色AliyunServiceRoleForDataWorksOpenPlatform。关于该RAM角色的相关说明,详情请参见AliyunServiceRoleForDataWorksOpenPlatform权限说明。

⑦ 说明 主账号及有AliyunDataWorksFullAccess权限的RAM账号可直接单击授权。无上述权限的 子账号将无法进行授权操作,您需要为该子账号授予Aliyundataworksfullaccess权限,操作详情 请参见:主账号如何给子账号添加Aliyundataworksfullaccess权限?。

2. 添加事件分发通道。

支持指定将DataWorks某个工作空间中的事件分发到Event Bridge的某个自定义总线中,之后您可以通过 Event Bridge来订阅和消费DataWorks事件消息。添加事件分发通道步骤如下:

- i. 单击**添加事件分发通道**,进入添加事件消息分发弹框。
- ii. 单击要分发事件的工作空间,在下拉列表中展示的是您当前已经加入的工作空间,您可以单击选择 需要发送消息的工作空间。

iii. 在弹框中单击**指定分发到Event Bridge中的自定义总线**,在下拉列表中选择上述空间中的事件需要 发送到的自定义总线。您可以单击**创建自定义总线**,跳转到Event Bridge控制台界面。

配置自定义事件总线

您可以根据需要在<mark>事件总线Event Bridge控制台</mark>配置事件规则,详情请参见事件规则概述。Dat aWorks发送到 Event Bridge的消息格式可以参见附录:发送给Event Bridge的消息格式。

发布事件

配置完事件总线后,您可以在<mark>事件总线Event Bridge控制台</mark>发布事件。您可以通过发布事件测试自定义事件总 线配置的事件规则,具体请参见发布事件。

管理事件分发通道

DataWorks提供事件分发通道的管理操作,您可对事件分发通道进行启用、禁用、删除等操作。

- 启用: 启用该事件分发通道后,该工作空间的事件消息将会推送至Event Bridge中对应的事件总线。
- 禁用: 禁用该事件分发通道后,该工作空间的事件消息将停止向Event Bridge中对应的事件总线推送,您将 无法通过Event Bridge接收到相应的事件消息。
- 删除:删除该事件分发通道后,该工作空间的事件消息将停止向Event Bridge中对应的事件总线推送,您将 无法通过Event Bridge接收到相应的事件消息。

后续步骤

OpenEvent基于EventBridge实现DataWorks事件消息的分发,您可以在EventBridge订阅事件消息。关于 EventBridge相关说明,详情请参见:<u>什么是事件总线EventBridge</u>。

5.3.3. 附录: DataWorks服务关联角色

本文为您介绍在DataoWorks上授权使用Event Bridge所产生的服务关联角色。

背景信息

在使用EventBridge实现DataWorks上的事件消息订阅与消费时,您需要先授权DataWorks访问您的 EventBridge资源才可基于EventBridge实现消息的订阅与消费,授权成功后会自动在RAM控制台增加一个服务 关联角色AliyunServiceRoleForDataWorksOpenPlatform,此角色用于DataWorks开放平台的服务关联角 色,DataWorks开放平台将使用此角色来访问您在其他云产品中的资源。关于RAM服务关联角色相关说明,详 情请参见服务关联角色。

AliyunServiceRoleForDataWorksOpenPlatform权限说明

- 1. 登录RAM控制台, 在左侧导航栏选择 身份管理>角色。
- 2. 在角色页面的搜索框中,输入*AliyunServiceRoleForDataWorksOpenPlatform*,自动搜索到名称 为AliyunServiceRoleForDataWorksOpenPlatform的RAM角色。
- 3. 点击AliyunServiceRoleForDataWorksOpenPlatform角色名称查看角色基本信息。
- 4. 在角色基本信息页面,点击**权限管理**,在权限管理界面查看当前角色被授予 的AliyunServiceRolePolicyForDataWorksOpenPlatform的权限策略详情,如下。

```
{
  "Version": "1",
   "Statement":[
     {
       "Effect": "Allow",
       "Action": [
        "eventbridge:GetEventBus",
        "eventbridge:GetRule",
         "eventbridge:CreateEventBus",
         "eventbridge:DeleteEventBus",
         "eventbridge:PutEvents",
         "eventbridge:ListEventBuses"
      ],
       "Resource": "*"
     },
     {
        "Action":"ram:DeleteServiceLinkedRole",
        "Resource":"*",
        "Effect":"Allow",
        "Condition":{
         "StringEquals": {
            "ram:ServiceName": "openplatform.dataworks.aliyuncs.com"
          }
        }
     }
  ]
}
```

删除AliyunServiceRoleForDataWorksOpenPlatform角色

- 1. 登录RAM控制台, 在左侧导航栏选择身份管理 > 角色。
- 2. 在角色页面的搜索框中,输入*AliyunServiceRoleForDataWorksOpenPlatform*,自动搜索到名称 为AliyunServiceRoleForDataWorksOpenPlatform的RAM角色。
- 3. 在右侧操作列,单击删除。

常见问题

问题描述:为什么我的RAM用户无法自动创建DataWorks服务关联角 色AliyunServiceRoleForDataWorksOpenPlatform?

解决方案: 您需要拥有指定的权限,才能自动创建或删 除AliyunServiceRoleForDataWorksOpenPlatformAliyunServiceRoleForDataWorksOpenPlatform。 因此,在RAM用户无法自动创建AliyunServiceRoleForDataWorksOpenPlatform时,您需为其添加以下 权限策略。

DataWorks公共云合集·开放平台

```
{
   "Statement": [
      {
       "Action": [
               "ram:CreateServiceLinkedRole"
           ],
         "Resource": "acs:ram:*:主账号ID:role/*",
         "Effect": "Allow",
          "Condition": {
              "StringEquals": {
                   "ram:ServiceName": [
                       "openplatform.dataworks.aliyuncs.com"
                  1
               }
          }
    }
 ],
 "Version": "1"
}
```

⑦ 说明 请把主账号ID替换为您实际的阿里云账号(主账号)ID。

相关文档

• 服务关联角色

5.3.4. 附录:发送给EventBridge的消息格式

本文介绍DataWorks开放消息的消息结构、不同事件类型事件的消息格式及各字段含义,帮助您快速获取和感知事件消息的状态变更信息。

消息总体格式

一个完整的事件消息由以下几个部分组成,示例如下。

```
{
"datacontenttype": "application/json;charset=utf-8",
 "data": {
   "beginWaitTimeTime": 165270057****,
   "dagId": 44655****,
   "dagType": 0,
   "taskType": 0,
   "modifyTime": 165270057****,
   "createTime": 165254323****,
   "appId": 3****2,
   "tenantId": 235454***432001,
   "opCode": 31,
   "flowId": 1,
   "nodeId": 100269****,
    "beginWaitResTime": 165270057****,
   "taskId": 453268****,
   "status": 3
 },
 "id": "539fd8f4-4ea1-4625-aa8b-6c906674***",
 "source": "acs.dataworks",
 "specversion": "1.0",
 "subject": "",
 "time": "2020-11-19T21:04:41+08:00",
 "type": "dataworks:InstanceStatusChanges:InstanceStatusChanges",
 "aliyunaccountid": "123456789098****",
 "aliyunpublishtime": "2020-11-19T21:04:42.179PRC",
 "aliyuneventbusname": "default",
 "aliyunregionid": "cn-hangzhou",
 "aliyunpublishaddr": "172.25.XX.XX"
}
```

字段名称	字段类型	说明
dat a	object	 消息实体格式,具体请参见下文各事件类型的消息实体格式详情: 消息实体格式: 实例状态变更事件 消息实体格式: 文件变更事件(文件提交、发布、运行、删除) 消息实体格式: 表变更事件(表提交到开发、生产环境) 消息实体格式: 节点变更事件(新增、修改、删除节点等) 消息实体格式: 补数据操作事件 消息实体格式: 审批中心(创建、完成审批单) 消息实体格式: 运维中心告警
id	String	用来定位事件消息的唯一ID。
type String		事件类型,取值包括commit-file(提交文件)、deploy-file(发布文件) 等,全量取值及其对应的事件请参见下文的 <mark>附录:事件详情(eventType)</mark> 。

部分字段详细说明请参见下表:

⑦ 说明 其余字段的详细说明,您可以参考事件概述。

消息实体格式: 实例状态变更事件

示例如下。

```
{
 "datacontenttype": "application/json;charset=utf-8",
 "data": {
   "beginWaitTimeTime": 165270057****,
   "dagId": 44655****,
   "dagType": 0,
   "taskType": 0,
   "modifyTime": 165270057****,
   "createTime": 165254323****,
   "appId": 3****2,
   "tenantId": 235454***432001,
   "opCode": 31,
   "flowId": 1,
   "nodeId": 100269****,
   "beginWaitResTime": 165270057****,
   "taskId": 453268****,
   "status": 3
  },
  "id": "539fd8f4-4ea1-4625-aa8b-6c906674***",
  "source": "acs.dataworks",
  "specversion": "1.0",
 "subject": "",
 "time": "2020-11-19T21:04:41+08:00",
 "type": "dataworks:InstanceStatusChanges:InstanceStatusChanges",
 "aliyunaccountid": "123456789098****",
 "aliyunpublishtime": "2020-11-19T21:04:42.179PRC",
 "aliyuneventbusname": "default",
 "aliyunregionid": "cn-hangzhou",
 "aliyunpublishaddr": "172.25.XX.XX"
```

```
}
```

字段名称	字段类型	说明
finishTime	Long	调度任务实例运行完成的具体时间
beginWaitTime Time	Long	调度任务实例开始等时间的具体时间
beginRunningTi me	Long	调度任务实例开始运行的具体时间
dagId	Long	DagId

DataWorks公共云合集·开放平台

Dat aWorks

字段名称	字段类型	说明
dagType	Integer	Dag的类型,取值如下: • 0:周期调度任务 • 1:手动任务 • 2:冒烟测试 • 3:补数据 • 4:手动业务流程 • 5:临时业务流程
taskType	Integer	任务实例的调度类型,取值如下: NORMAL(0):正常调度任务。该任务被日常调度。 MANUAL(1):手动任务。该任务不会被日常调度。 PAUSE(2):冻结任务。该任务被日常调度,但启动调度时直接被置为失败状态。 SKIP(3):空跑任务。该任务被日常调度,但启动调度时直接被置为成功状态。 SKIP_UNCHOOSE(4):临时工作流中未选择的任务,仅存在于临时工作流中,启动调度时直接被置为成功状态。 SKIP_CYCLE(5):未到运行周期的周或月任务。该任务被日常调度,但启动调度时直接被置为成功状态。 CONDITION_UNCHOOSE(6):上游实例中有分支(IF)节点,但是该下游节点未被分支节点选中,直接置为空跑任务。 REALTIME_DEPRECATED(7):实时生成的已经过期的周期实例,该类型的任务直接被置为成功状态。
modifyTime	Long	任务实例的修改时间
createTime	Long	任务实例的创建时间
appld	Long	工作空间的ID。您可以调用 <mark>ListProjects</mark> 查看空间ID信息。
tenantld	Long	调度任务实例所在工作空间的租户ID
opCode	Integer	调度任务实例的操作码:该字段可忽略
flowId	Long	业务流程的ID,周期调度任务实例的业务流程默认为1,手动业务流程和内部工 作流调度任务实例为实际的业务流程ID
nodeld	Long	调度任务实例对应的节点ID
beginWaitResT ime	Long	调度任务实例开始等资源的具体时间
taskld	Long	调度任务实例ID

字段名称	字段类型	说明
status	Integer	任务的状态,取值如下: • 0(未运行) • 2(等待定时时间dueTime或cycleTime到来) • 3(等待资源) • 4(运行中) • 7(下发给数据质量进行数据校检) • 8(正在进行分支条件校检) • 5(执行失败) • 6(执行成功)

消息实体格式: 文件变更事件(文件提交、发布、运行、删除)

• 文件提交、发布事件

```
{
 "datacontenttype": "application/json;charset=utf-8",
 "data": {
   "operator": "**********", //本次操作人
   "projectId": 12****56,
   "tenantId": 12****56,
   "nodeId": 12****, //节点ID
   "fileType": 12****, //文件类型
   "fileName": "******",
   "changeType": "UPDATE", //NEW UPDATE DELETE
   "fileCreateTime": "2021-01-15 14:03:02",
   "fileId": 12****,
   "fileVersion": 3,
   "dataSourceName":"odps first"
 },
 "id": "539fd8f4-4ea1-4625-aa8b-6c906674***",
 "source": "acs.dataworks",
 "specversion": "1.0",
 "subject": "",
 "time": "2020-11-19T21:04:41+08:00",
 "type": "dataworks:FileChange:CommitFile",
 "aliyunaccountid": "123456789098****",
 "aliyunpublishtime": "2020-11-19T21:04:42.179PRC",
 "aliyuneventbusname": "default",
 "aliyunregionid": "cn-hangzhou",
 "aliyunpublishaddr": "172.25.XX.XX"
```

```
}
```

字段名称	字段类型	说明
operator	String	发布人用户UID
projectId	Long	项目空间ID

字段名称	字段类型	说明
tenantId	Long	租户ID
nodeld	Long	调度节点ID
fileT ype	Long	 文件的代码类型,常用的代码类型包括: 6 (Shell) 10 (ODPS SQL)、 11 (ODPS MR) 23 (数据集成) 24 (ODPS Script) 99 (虚拟节点) 221 (PyODPS 2) 225 (ODPS Spark) 227 (EMR Hive) 228 (EMR Spark) 229 (EMR Spark SQL) 230 (EMR MR)

• 文件删除、运行事件

```
{
 "datacontenttype": "application/json;charset=utf-8",
 "data": {
   "operator": "************", //本次操作人
   "projectId": 12****,
   "tenantId": 12****,
   "nodeId": 12****, //节点ID
   "fileType": 12****, //文件类型
   "fileName": "******",
   "fileCreateTime": "2021-01-15 14:03:02",
   "fileId": 12****
 },
 "id": "539fd8f4-4ea1-4625-aa8b-6c906674****",
 "source": "acs.dataworks",
 "specversion": "1.0",
 "subject": "",
 "time": "2020-11-19T21:04:41+08:00",
 "type": "dataworks:FileChange:DeleteFile",
 "aliyunaccountid": "123456789098****",
 "aliyunpublishtime": "2020-11-19T21:04:42.179PRC",
 "aliyuneventbusname": "default",
 "aliyunregionid": "cn-hangzhou",
 "aliyunpublishaddr": "172.25.XX.XX"
}
```

字段名称	字段类型	说明
operator	String	发布人用户UID
projectId	Long	项目空间ID
tenantld	Long	租户ID
nodeld	Long	调度节点ID
fileT ype	Long	 文件的代码类型,常用的代码类型包括: 6 (Shell) 10 (ODPS SQL) 11 (ODPS MR) 23 (数据集成) 24 (ODPS Script) 99 (虚拟节点) 221 (PyODPS 2) 225 (ODPS Spark) 227 (EMR Hive) 228 (EMR Spark) 229 (EMR Spark SQL) 230 (EMR MR) 239 (OSS对象检查) 257 (EMR Shell)

消息实体格式:表变更事件(表提交到开发、生产环境)

{	
	"datacontenttype": "application/json;charset=utf-8",
	"data": {
	"operator": "**********",
	"projectId": 12****56,
	"tenantId": 12*****56,
	"extensionBizId": "123456",
	"tableName":"table1",
	"tableType":"ODPS", // ODPS、EMR
	"maxComputeProject":"project1"
	},
	"id": "539fd8f4-4ea1-4625-aa8b-6c906674c113",
	"source": "acs.dataworks",
	"specversion": "1.0",
	"subject": "",
	"time": "2020-11-19T21:04:41+08:00",
	"type": "dataworks:TableChange:CommitTable",
	"aliyunaccountid": "123456789098****",
	"aliyunpublishtime": "2020-11-19T21:04:42.179PRC",
	"aliyuneventbusname": "default",
	"aliyunregionid": "cn-hangzhou",
	"aliyunpublishaddr": "172.25.XX.XX"

```
}
```

字段名称	字段类型	说明
operator	String	发布人用户UID
projectId	Long	项目空间ID
tenantId	Long	租户ID
extensionBizId	String	扩展程序卡点流程ID
tableName	String	表名
tableType	String	表类型: • ODPS • EMR
maxComputeP roject	String	对应maxCompute的project名称

消息实体格式: 节点变更事件 (新增、修改、删除节点等)

• 节点新增、修改、删除事件

```
{
 "datacontenttype": "application/json;charset=utf-8",
 "data": {
   "nodeName": "",
   "programType": "ODPS SQL",
   "cronExpress": "00 20 00 * * ?",
   "schedulerType": "NORMAL",
   "ownerId": "191223248874***",
   "priority": 1,
   "baselineId": 11780****,
   "repeatability": true,
   "modifyTime": 164636454****
   "createTime": 164636454****,
   "datasource": "odps_first",
   "tenantId": 28378****10656,
   "nodeId": 100***150,
   "projectId": 30**95,
   "operator": "********** //本次操作人
  },
 "id": "539fd8f4-4eal-4625-aa8b-6c906674***",
 "source": "acs.dataworks",
 "specversion": "1.0",
 "subject": "",
  "time": "2020-11-19T21:04:41+08:00",
 "type": "dataworks:NodeChange:NodeChangeDeleted",
 "aliyunaccountid": "123456789098****",
 "aliyunpublishtime": "2020-11-19T21:04:42.179PRC",
 "aliyuneventbusname": "default",
 "aliyunregionid": "cn-hangzhou",
 "aliyunpublishaddr": "172.25.XX.XX"
```

- 13

字段名称	字段类型	说明
nodeName	String	节点名称
Dat aWorks

字段名称	字段类型	说明
programType	String	常用的代码及其对应的文件类型包括: • 6 (Shell) • 10 (ODPS SQL) • 11 (ODPS MR) • 24 (ODPS Script) • 99 (虚拟节点) • 221 (PyODPS 2) • 225 (ODPS Spark) • 227 (EMR Hive) • 228 (EMR Spark) • 229 (EMR Spark SQL) • 230 (EMR MR) • 239 (OSS对象检查) • 255 (EMR Spark Shell) • 259 (EMR Presto) • 260 (EMR Impala) • 900 (实时同步) • 1091 (Hologres开发) • 1093 (Hologres SQL) • 1100 (赋值节点) • 1221 (PyODPS 3)
cronExpress	String	 周期调度的cron表达式,该参数位于DataWorks控制台中,数据开发任务的调度配置 > 时间属性 > cron表达式对应。配置完调度周期及定时调度时间后,DataWorks会自动生成相应cron表达式。示例如下 每天凌晨5点30分定时调度:003005**? 每个小时的第15分钟定时调度:0015***? 每你小时的第15分钟定时调度:0015***? 每隔十分钟调度一次:0000/10***? 每天8点到17点,每个十分钟调度一次:0000-59/108-23***? 每月的1日0点20分自动调度:002001*? 从1月1日0点10分开始,每过3个月调度一次:00100011-12/3? 每周二、周五的0点5分自动调度:000500**2,5 由于DataWorks调度系统的规则,cron表达式有以下限制: 最短调度间隔时间为5分钟。 每天最早调度时间为0点5分。

• 节点冻结、解冻、下线事件

DataWorks公共云合集·开放平台

```
{
 "datacontenttype": "application/json;charset=utf-8",
 "data": {
   "operator": "************",
   "projectId": 12****56,
   "tenantId": 12*****56,
   "extensionBizId": "12****",
   "tableName":"table1",
   "tableType":"ODPS", // ODPS、EMR
   "maxComputeProject":"project1"
 },
 "id": "539fd8f4-4ea1-4625-aa8b-6c906674****",
 "source": "acs.dataworks",
 "specversion": "1.0",
 "subject": "",
 "time": "2020-11-19T21:04:41+08:00",
 "type": "dataworks:TableChange:CommitTable",
 "aliyunaccountid": "123456789098****",
 "aliyunpublishtime": "2020-11-19T21:04:42.179PRC",
 "aliyuneventbusname": "default",
 "aliyunregionid": "cn-hangzhou",
 "aliyunpublishaddr": "172.25.XX.XX"
```

```
}
```

字段名称	字段类型	说明
operator	String	发布人用户UID
projectId	Long	项目空间ID
tenantld	Long	租户ID
nodelds	Array	被操作节点ID列表
ext ensionBiz Id	String	扩展程序卡点流程ID

```
消息实体格式:补数据操作事件
```

```
{
 "datacontenttype": "application/json;charset=utf-8",
 "data": {
   "excludeNodeIds":[],
   "rootNodeId": 1000083****,
    "startFutureInstanceImmediately": false,
   "useMultipleTimePeriods": true,
   "operator": "132087671033****",
   "multipleTimePeriods": "[{\"bizBeginTime\":\"2022-04-17\",\"bizEndTime\":\"2022-04-17\"}]
",
   "parallelGroup": 1,
   "rootNodeProjectId": 12****8,
   "isParallel": false,
   "name": "P_fff_20220418_21*****",
    "tenantId": 16935****3377,
    "includeNodeIds":
   [
    1000083****
   ],
   "projectId": 1****,
   "order": "asc",
   "extensionBizId": "12****"
  },
  "id": "539fd8f4-4ea1-4625-aa8b-6c906674****",
 "source": "acs.dataworks",
 "specversion": "1.0",
 "subject": "",
 "time": "2020-11-19T21:04:41+08:00",
 "type": "dataworks:BackfillDataOperate:BackfillData",
 "aliyunaccountid": "123456789098****",
 "aliyunpublishtime": "2020-11-19T21:04:42.179PRC",
 "aliyuneventbusname": "default",
 "aliyunregionid": "cn-hangzhou",
 "aliyunpublishaddr": "172.25.XX.XX"
```

字段名称	字段类型	说明
name	String	补数据工作流名称
rootNodeld	Long	补数据工作流根节点ID
rootNodeProje ctId	Long	补数据工作流根节点所在的项目空间ID
includeNodeld s	Array	补数据的节点ID列表
excludeNodeId s	Array	无需补数据的节点ID列表,在这个列表中的节点会生成空跑实例,空跑实例被 调度后会直接运行成功,不会执行脚本内容。
biz BeginT ime	String	任务的开始时间。仅小时调度任务需要设置该参数,格式为HH:mm:ss,取值 范围为 00:00:00~23:59:59。

字段名称	字段类型	说明
bizEndTime	String	任务的结束时间。仅小时调度任务需要设置该参数,格式为HH:mm:ss,取值 范围为 00:00:00~23:59:59。
isParallel	Boolean	是否可以并行运行
parallelGroup	Integer	并行运行的分组数,1表示不分组
startFutureInst anceImmediat ely	Boolean	是否跳过调度时间,立刻调起未来实例
order	String	业务日期运行的顺序, "acs"表示按业务日期正序运行

消息实体格式: 审批中心(创建、完成审批单)

示例如下。

{

• 创建审批单事件。

```
"datacontenttype": "application/json;charset=utf-8",
"data": {
 "appId":19****,
"assignee":"28609853964174****",
"assigneeName":"yupeng.sunyp",
"createTime":165209436****,
"eventType":"approval",
"process":{
  "applicant":"28609853964174***",
  "applicantName":"yupeng.sunyp",
  "approvalContent":{
    "applyPeriod":"299796480****",
    "applyReason":"测试",
    "arrayData":[
      {
        "ownerBaseId":"182293110403****",
        "columns":[
         {
            "columnType":"string",
            "columnGuid":"odps.b_mc1.loghub_07****.id",
            "columnComment":"",
            "columnName":"id"
          },
          {
            "columnType":"string",
            "columnGuid":"odps.b mc1.loghub 07****.name",
            "columnComment":"",
            "columnName":"name"
          },
          {
            "columnType":"string",
            "columnGuid":"odps.b_mc1.loghub_07****.times",
            "columnComment":"",
            "columnName"."times"
```

```
corunname . crines
     }
    ],
    "columnMetaList":[
     {
        "columnType":"string",
        "columnGuid":"odps.b_mc1.loghub_07****.id",
        "columnComment":"",
        "columnName":"id"
      },
      {
        "columnType":"string",
        "columnGuid":"odps.b mc1.loghub 07****.name",
        "columnComment":"",
        "columnName":"name"
      },
      {
        "columnType":"string",
        "columnGuid":"odps.b mc1.loghub 07****.times",
        "columnComment":"",
        "columnName":"times"
      }
    ],
    "objectType":"TABLE",
    "odpsTable":"loghub 07****",
    "envType":1,
    "projectGuid":"odps.b mc1",
    "objectGuid":"odps.b_mc1.loghub_07****",
    "tenantId":0,
    "objectName":"loghub 07****",
    "ownerAccountName": "ALIYUN$dataworks 3h1 1(dataworks 3h1 1)",
    "odpsProject":"B MC1",
    "projectName":"B MC1",
    "actions":[
     "Select",
     "Describe"
    ],
    "projectId":19****,
    "workspaceId":"19****"
 }
],
"contentType":"application/json",
"granteeAccounts":[
 {
    "granteeId":"28609853964174***",
    "granteeTypeSub":103,
    "granteeType":1,
    "granteeName":"RAM$dataworks 3h1_1:yupeng.sunyp"
  },
    "granteeId":"23785763111910****",
    "granteeTypeSub":105,
    "granteeType":1,
    "granteeName":"RAM$dataworks 3h1 1:dev"
  }
```

```
1,
     "odpsProjectName":"B MC1",
     "projectEnv":"1",
     "resourceSummary":"loghub 07****",
     "tenantId":28074952195****,
     "workspaceId":194209
   },
   "assignmentCategory":"MaxCompute",
    "createTime":1652094363000,
   "processDefinitionId":"definition-3dcc9ce7-d29d-435d-a908-60d4355f****",
   "processId":"52853586998****",
   "status":"Pending",
    "title":"MaxComputeTable",
    "updateTime":165209436****
 },
 "processId":"52853586998****",
 "status":"Submit",
 "taskId":"52853587001****",
 "tenantId":28074952195****,
 "updateTime":165209436****
 },
 "id": "539fd8f4-4ea1-4625-aa8b-6c906674****",
 "source": "acs.dataworks",
 "specversion": "1.0",
 "subject": "",
 "time": "2020-11-19T21:04:41+08:00",
 "type": "dataworks:ApprovalChange:ApprovalChangeCreated",
 "aliyunaccountid": "123456789098****",
 "aliyunpublishtime": "2020-11-19T21:04:42.179PRC",
 "aliyuneventbusname": "default",
 "aliyunregionid": "cn-hangzhou",
 "aliyunpublishaddr": "172.25.XX.XX"
}
```

• 完成审批单事件。

```
[
 {
   "ownerBaseId":"182293110403****",
    "columns":[
     {
        "columnType":"decimal(38,18)",
       "columnGuid":"odps.px_wokebench_1.oracle_dss_a1214_1136_1_invalid_column.i_d",
       "columnComment":"",
       "columnName":"i d"
     }
   ],
    "columnMetaList":[
     {
       "columnType":"decimal(38,18)",
        "columnGuid":"odps.px wokebench 1.oracle dss a1214 1136 1 invalid column.i d",
       "columnComment":"",
       "columnName":"i d"
     }
   ],
    "objectType":"TABLE",
   "odpsTable":"oracle_dss_a1214_1136_1_invalid_column",
   "envType":1,
    "projectGuid":"odps.px_wokebench_1",
    "objectGuid": "odps.px wokebench 1.oracle dss a1214 1136 1 invalid column",
   "tenantId":0,
   "objectName":"oracle dss a1214 1136 1 invalid column",
   "ownerAccountName":"ALIYUN$dataworks 3h1 1(dataworks 3h1 1)",
    "odpsProject":"dataworks_pref_test****",
    "projectName":"dataworks pref test***",
   "actions":[
     "Select",
     "Describe"
   ],
    "projectId":10****,
   "workspaceId":"10****"
 }
]
```

字段名称	字段类型	说明
appld	Long	工作空间ID
assignee	String	处理人Baseld
assigneeName	String	处理人name
comments	String	备注
createTime	Long	申请单创建时间戳
processId	String	审批单ID
status	String	审批单状态

字段名称	字段类型	说明
taskld	String	审批任务ID
tenantId	String	租户ID
updateTime	String	申请单更新时间戳
eventType	String	事件类型
process	Object	审批任务对象
applicant	String	申请baseld
 applicant Name	String	申请人名称
 assignment Cat egory	String	申请内容类型
createTime	String	申请单创建时间戳"desc"表示按业务日期倒序运行
 processDefiniti onId	String	工作流定义ID
processId	String	工作流ID
status	String	工作流状态
title	String	工作流标题
updateTime	Long	申请单更新时间戳
 approvalConte nt	Object	审批内容对象
 applyPeriod	String	申请时长
 applyReason	String	申请原因
 contentType	String	内容类型
 odpsProjectNa me	String	项目名称
 resourceSumm ary	String	资源描述

字段名称	字段类型	说明
tenantld	Long	租户ID
 workspaceld	Long	工作空间ID
projectEnv	String	所属环境
 granteeAccoun ts	Array	授权列表
 granteeld	String	授权主体id
 granteeType	String	授权类型
 granteeTypeSu b	String	 授权主体子类下,代表账号类型,具体如下: 生产云账号(生产调度使用的账号):ACCOUNT_PRD(101) 应用云账号:ACCOUNT_APP(102) 个人云账号:ACCOUNT_USER(103) 部门云账号:ACCOUNT_DEPT(104) MOCK账号:ACCOUNT_MOCK(106) 他人云账号:ACCOUNT_OTHER_USER(105)
 granteeName	String	授权主体名称
larravData	Array	参见授权内容列表

消息实体格式:运维中心告警

	Long	
 workspaceld	Long	工作空间ID
projectEnv	String	所属环境
 granteeAccoun ts	Array	授权列表
 granteeld	String	授权主体id
 granteeType	String	授权类型
 granteeTypeSu b	String	授权主体子类下,代表账号类型,具体如下: • 生产云账号(生产调度使用的账号):ACCOUNT_PRD(101) • 应用云账号:ACCOUNT_APP(102) • 个人云账号:ACCOUNT_USER(103) • 部门云账号:ACCOUNT_DEPT(104) • MOCK账号:ACCOUNT_MOCK(106) • 他人云账号:ACCOUNT_OTHER_USER(105)
 granteeName	String	授权主体名称
arrayData	Array	参见授权内容列表

● 基线告警 { "alarmType": "SLA_ALERT", "baselineId": 13700****,

```
"baselineName": "ods层任务专用",
"baselineStatus": 3,
"bizDate": 165444480****,
"inGroupId": 14,
"nodeId": 1000039****,
"projectId": ****,
"taskId": 307289****,
"tenantId": 26095350312****
```

}

字段名称	字段类型	说明
alarmType	String	告警类型: • SLA_ALERT • REMIND_ALERT • TOPIC_ALERT
baselineId	Long	基线ID
baselineName	String	基线名称
baselineStatus	Integer	基线的状态,包括: o -1(异常) o 1(安全) o 2(预警) o 3(破线)
bizDate	Long	业务日期时间戳。
inGroupId	Integer	基线实例的周期号。天基线为1,小时基线的取值范围为[1,24]。
nodeld	Long	导致基线异常的节点ID
projectId	Long	基线所属项目空间ID
taskld	Long	导致基线异常的实例ID
tenantld	Long	租户ID

● 事件告警

```
{
    "alarmType": "TOPIC_ALERT",
    "nodeId": 1000031****,
    "projectId": ****,
    "taskId": 307275****,
    "taskStatus": 5,
    "tenantId": 28074952195****,
    "topicId": 108****
```

```
}
```

字段名称	字段类型	说明
alarmType	String	告警类型: • SLA_ALERT • REMIND_ALERT • TOPIC_ALERT
topicId	Long	事件的ID

字段名称	字段类型	说明
taskStatus	String	触发事件的节点的实例状态
nodeld	Long	触发事件的节点ID
projectId	Long	触发事件的节点所属项目空间ID
taskld	Long	触发事件的节点的实例ID
tenantld	Long	租户ID

• 规则告警

```
。 当规则对象为任务节点、基线、工作空间、业务流程时
```

```
{
    "alarmType": "REMIND_ALERT",
    "nodeIds": "1000040****,1000040****,1000040****",
    "projectId": ****,
    "remindId": ****,
    "remindName": "出错报警",
    "remindType": "ERROR",
    "remindUnit": "NODE",
    "taskIds": "307290****,307290****,307290****",
    "tenantId": 25902581602****
}
```

字段名称	字段类型	说明
alarmType	String	告警类型: SLA_ALERT REMIND_ALERT TOPIC_ALERT
nodelds	String	触发规则告警的节点列表
remindId	Long	规则ID
remindT ype	Stirng	触发条件,包括: FINISHED(完成) UNFINISHED(未完成) ERROR(出错) CYCLE_UNFINISHED(周期未完成) TIMEOUT(运行超时)
projectId	Long	节点所属的项目空间

字段名称	字段类型	说明
remindUnit	String	对象的类型,包括: ■ NODE(任务节点) ■ GATEWAY_RES(调度独享资源组) ■ DI_RES(数据集成资源组)
tenantld	Long	租户ID
taskids	String	触发规则告警的实例列表
remindName	String	规则名称

• 当规则对象为资源组时:

}

```
{
   "alarmType": "REMIND_ALERT",
   "projectId": ****,
   "remindId": 20000****,
   "remindName": "独享资源组告警",
   "remindType": "RES_GROUP_THRESHOLD",
   "remindUnit": "GATEWAY_RES",
   "resourceGroupIdentifier": "S_res_group_19582071655****_165096585****",
   "resourceGroupName": "emr_exclusive_scheduld",
   "resourceGroupType": "GATEWAY",
   "tenantId": 19582071655****
```

字段名称	字段类型	说明
alarmType	String	告警类型: SLA_ALERT REMIND_ALERT TOPIC_ALERT
remindId	Long	规则ID
remindT ype	Stirng	触发条件,包括 FINISHED(完成) UNFINISHED(未完成) ERROR(出错) CYCLE_UNFINISHED(周期未完成) TIMEOUT(运行超时) RES_GROUP_THRESHOL(资源组利用率) RES_GROUP_WAIT_AMOUNT(资源组等资源实例数)
projectId	Long	节点所属的项目空间

字段名称	字段类型	说明
remindUnit	String	对象的类型,包括: ■ NODE(任务节点) ■ GATEWAY_RES(调度独享资源组) ■ DI_RES(数据集成资源组)
tenantld	Long	租户ID
remindName	String	规则名称
resourceGrou pldent ifier	String	资源组的唯一标识符
resourceGrou pName	String	资源组名称
resourceGrou pType	String	资源组类型: ■ GATEWAY(调度资源组) ■ DI(数据集成资源组)

附录:事件类目详情 (eventCategoryType)

事件类型	事件码
实例状态变更事件	instance-status-change
文件变更事件	file-change
表变更事件	table-change
节点变更事件	node-change
补数据变更事件	backfill-data-operate
审批中心事件	approval-change
监控告警事件	monitor-alert

附录:事件详情 (eventType)

系统模块	事件名称	事件编码	事件分类
	新增节点	node-change-created	普通事件
	更新节点	node-change-updated	普通事件
	删除节点	node-change-deleted	普通事件
	冻结节点	freeze-node	扩展点事件
运维中心			

系统模块	事件名称	事件编码	事件分类				
	解冻节点	unfreeze-node	扩展点事件				
	下线节点	undeploy-node	扩展点事件				
	补数据	backfill-data	扩展点事件				
	提交文件	commit-file	扩展点事件				
	发布文件	deploy-file	扩展点事件				
*** 년 고 부	运行文件	run-file	扩展点事件				
蚁 施开及	删除文件	delete-file	扩展点事件				
	提交表	commit-t <i>a</i> ble	扩展点事件				
	发布表	deploy-table	扩展点事件				
中世中心	创建申请单	approval-change- created	普通事件				
申加十心	完成申请单	approval-change- finished	普通事件				
监控告警	运维中心告警	workbench-monitor- alert	普通事件				

授权内容列表

• assignmentCategory类型为MaxCompute时 arrayData的数据内容。

```
[
 {
   "ownerBaseId":"182293110403****",
    "columns":[
     {
        "columnType":"decimal(38,18)",
       "columnGuid":"odps.px_wokebench_1.oracle_dss_a1214_1136_1_invalid_column.i_d",
       "columnComment":"",
       "columnName":"i d"
     }
   ],
    "columnMetaList":[
     {
       "columnType":"decimal(38,18)",
        "columnGuid":"odps.px wokebench 1.oracle dss a1214 1136 1 invalid column.i d",
       "columnComment":"",
       "columnName":"i d"
     }
   ],
    "objectType":"TABLE",
   "odpsTable":"oracle_dss_a1214_1136_1_invalid_column",
    "envType":1,
    "projectGuid":"odps.px_wokebench_1",
    "objectGuid": "odps.px wokebench 1.oracle dss a1214 1136 1 invalid column",
   "tenantId":0,
   "objectName":"oracle dss a1214 1136 1 invalid column",
   "ownerAccountName":"ALIYUN$dataworks_3h1_1(dataworks_3h1_1)",
    "odpsProject":"dataworks_pref_test****",
    "projectName":"dataworks pref test***",
   "actions":[
     "Select",
     "Describe"
   ],
    "projectId":10****,
   "workspaceId":"10****"
 }
]
```

字段名称	字段类型	说明
ownerBaseld	String	表owner的baseld
ownerAccount Name	String	表owner的name
objectType	String	对象类型
odpsTable	String	表名称
envType	String	环境
projectGuid	String	项目Guid
objectGuid	String	对象Guid

字段名称	字段类型	说明
objectName	String	对象名称
odpsProject	String	odps项目名称
projectName	String	项目code
projectId	Long	工作空间ID
workspaceld	String	工作空间ID
actions	Array	操作类型列表
columns	Array	列列表
columnType	String	列类型
columnGuid	String	列Guid
 columnComm ent	String	列备注
 columnName	String	列名
columnMetaLi st	Array	列元数据列表
columnType	String	列类型
columnGuid	String	列Guid
 columnComm ent	String	列备注
 columnName	String	列名

• assignment Category类型为Dat aService时 arrayDat a的数据内容。

```
[
 {
   "resourceId":"DsApiDeploy/1822931104031845/workspaceId/227859/dsDeployId/1|41738195594
7846|1|280749521950784|22****",
   "ownerName":"ysf0111****",
   "resourceVersion":1,
   "name":"api_api",
   "dsDeployId":"1|417381955947846|1|280749521950784|22****",
   "workspaceName":"da_simple_20****",
   "id":"41738195594****",
   "type":1,
   "ownerId":"25888213955403****",
   "url":"https://ds-cn-shanghai.data.aliyun.com/?projectId=227859&type=api&id=4173819559
47846&version=1&defaultProjectId=22****",
    "workspaceId":"22****"
 }
```

```
]
```

字段名称	字段类型	说明
resourceld	String	资源ID
ownerName	String	资源ownerName
ownerld	String	资源责任人Baseld
resourceVersio n	Long	资源版本
name	String	资源名称
dsDeployId	String	数据服务发布id
workspaceNa me	String	工作空间名称
workspaceld	String	工作空ID
id	String	资源唯一ID
type	String	资源类型: 1(AP1) 2(函数) 3(服务编排)
url	String	数据服务链接地址

• assignment Category类型为DataStudio时 arrayData的数据内容。

[
	{	
		"sourceDatasourceType":"mysql",
		"sourceTableName":"test_datahub_type_copy",
		"targetDatasourceType":"odps",
		"sourceDatasourceName":"test_Approval",
		"targetTableName":"test_approval_test_datahub_type_copy",
		"targetDatasourceName":"odps_first",
		"workspaceId":"24****"
	},	
	{	
		"sourceDatasourceType":"mysql",
		"sourceTableName":"test_datahub_type_copy",
		"targetDatasourceType":"odps",
		"sourceDatasourceName":"test_Approval",
		"targetTableName":"test_approval_test_datahub_type_copy",
		"targetDatasourceName":"odps_first",
		"workspaceId":"24****"
	}	
]		

字段名称 字段类型 说明 sourceDatasourc String 源端数据类型 eType sourceDatasourc String 源端表类型 eName sourceTableNa String 源端表类型 me targetDatasourc String 目的端数据类型 еТуре targetDatasourc String 目的端表类型 eName targetTableNam String 目的端表类型 е

5.4. 扩展程序(Extensions)

背景信息

1.

5.4.1. 扩展程序概述

DataWorks通过OpenEvent能力为您提供消息推送订阅能力,同时您可以将服务程序注册为DataWorks的扩展 程序,通过扩展程序来卡点并响应订阅的事件消息,以此实现通过扩展程序对特定事件进行消息通知与流程管 控。本文为您介绍通过OpenEvent和扩展程序实现消息通知与流程管控的功能概述。

使用限制

- 目前仅面向DataWorks企业版和旗舰版用户公测,华北2(北京)、华东1(杭州)、华东2(上海)、华北3(张家口)、华南1(深圳)、西南1(成都)、美国(硅谷)、美国(弗吉尼亚)、德国(法兰克福)、日本(东京)、中国(香港)、新加坡地域支持该功能。
- 仅支持阿里云主账号或者有AliyunDataWorksFullAccess权限的RAM用户权限的用户可以直接进行界面相关 配置,关于DataWorks权限说明,详情请参见管控权限详情: 细分RAM Policy。

产品能力

DataWorks作为阿里云一站式大数据开发与治理平台,在各项能力支持上预设置了很多拓展点,例如,在标准 模式下,简略来看,节点创建后完成代码开发后,需先提交到开发环境中进行验证,完成后再正式发布到生产 环境中,后续每天根据调度配置定时运行,如果节点不再需要,可以删除节点。



在这个流程中DataWorks预置了几个扩展点,您可以基于扩展点增加自定义管控流程,例如下列场景:

• 增加代码评审流程。

当用户单击提交节点时,流程进入自定义代码评审流程中,不直接提交到开发环境中进行验证。自定义评 审流程通过后再提交至开发环境。

• 增加发布前审核流程。

当用户单击发布节点时,流程进入自定义发布前审核流程中,不直接发布到生产环境。自定义审核流程通 过后,再发布至生产环境。

上述场景中的流程管控,是通过DataWorks提交文件扩展点和发布文件扩展点实现流程卡点,并通过您注册在 DataWorks上的扩展程序实现自定义流程管控。您可以将自己开发的应用注册到DataWorks上成为扩展程序, 并通过OpenEvent和Open API进行扩展程序与DataWorks的通信,进而实现自定义流程控制。处理流程大致如 下:

1. 通过OpenEvent实现扩展消息订阅。

OpenEvent通过公共Kafka为您提供部分事件的消息订阅,详情请参见OpenEvent概述。完成订阅后,您可 以获取到对应事件的消息详情,例如,提交节点为文件提交类事件,这类事件的消息详情会包含节点ID、 操作人UID、项目空间ID等信息。各类事件的消息格式详情请参见附录:消息格式。

2. 通过扩展程序实现扩展点事件处理。

DataWorks支持将用户的服务程序注册为扩展程序,订阅接收到事件消息后,通过扩展程序处理事件并将 处理结果通过OpenAPl返回给DataWorks。其中,支持通过扩展程序处理事件的扩展点列表详情请参见下 文的支持的扩展点事件列表。

3. 通过OpenAPI实现扩展点事件处理结果回调。

扩展点事件触发等待扩展程序处理过程中,当前的流程会变为"检查中"状态,OpenAPI会将扩展程序的处理结果返回给DataWorks,DataWorks根据反馈结果状态继续流程。

- 处理结果为通过:事件符合扩展程序校验逻辑,当前流程通过后继续,状态从"检查中"变为"通过"。
- 处理结果为不通过:事件不符合扩展程序校验逻辑,需阻塞此次流程,状态从"检查中"变为"不通过"。

② 说明 您可通过Updat elDEEvent Result 这个AP将Dat aSt udio中的扩展点事件的处理结果返回给 Dat aWorks。

支持的扩展点事件列表

当前支持通过扩展程序处理的扩展点事件为DataStudio中的部分操作,分类及事件详情如下表。

支持的扩展点	扩展点事件	回调校验结果的OpenAPI
文件提交扩展点	文件提交前置事件 :例如节点、资源、函数的提交操 作。	
	文件发布前置事件 :例如节点、资源、函数的发布操 作。	
文件发布扩展点	⑦ 说明 仅标准模式工作空间支持该扩展点。	
文件运行扩展点	文件运行前置事件 :例如运行节点代码。	您可通 过UpdateIDEEventResult
文件删除扩展点	文件删除前置事件 :例如在删除某个节点、资源、函 数。	这个API将事件处理结果回 调至DataWorks。
表提交扩展点	表提交前置事件 :例如表管理提交表。	
	表发布前置事件 :例如表管理发布表。	
表发布扩展点	⑦ 说明 仅标准模式工作空间支持该扩展点。	

以节点提交为例,流程如下图所示。

Dat aWorks

S	操作检查		-	C	Sq 触发	发文件提交前词	置事件检	查 x												
G	类型: 全部					E1 61	لم		\$	\odot	Þ		С			∋	28	ಹೆ		
	状态:全部											提交	文件,	操作格	金查中 :					
	名称/操作时间	类型	状态	<u>^</u>		1od 2**	ps sq] ******				****									
4	触发文件提交前置事件检 查 2021-10-22 11:57:00	提交文件		_		3au 4cr	- <u>e</u> ,	检查中						¥ (2				×		
亩 fx	触发代码运行前置事件检 查 2021-10-20 17:17:12	运行	检查中 >			6 sele		检查通	过后将	自动拼	行操作	₽,-请存	E [操	作检查	」中查	看结果				
2	xc_table_prod 2021-10-20 16:36:00	表提交到开发																		
∷ ₽	扩展程序_pyodps_代码运 行前置事件 2021-10-20 16:25:34	运行	检查中 >																	
操作检	Ť 👩	*				×									_ // >		~			
67 Pr - A									点	击提交! 的扩展;	触发扩展 程序校引	医柱/序柱 金)	金查(将触发	工作空	间目前	所开启	的所有了	(件提交)	前置爭
检查								6) ग	以进入1	创建发行	节包界 面	面查看这	文件检查	查状态					
提交的	检查 通訊程序					XII			可,	点击状; 详情	态查看相	交验详情	青、可行	生此查考	舌 当前技	暴作触发	发的扩展	展程序校	验列表及	及校
投资	型 始查 测试程序	状态 松吉中	操作 查看详情 查看文档		服料检查	关闭			可	以进入1 点击状; 详情	创建发7 态查看相	交验详情	面查看了 青、可?	文件检查 在此查	查状态 看当前书	景作触发	发的扩展	展程	序校	序校验列表及

更多扩展点的触发应用示例可参见。应用示例:数据开发过程触发事件检查

操作流程

如上文所述,通过OpenEvent、扩展程序、OpenAPI实现自定义管控流程的完整的操作包括以下几个步骤。

1. 准备工作:

开启消息订阅,并注册扩展程序,获取开发扩展程序所需的必要信息。详情请参见准备工作。

2. 开发部署扩展程序。

根据已获取的信息,开发扩展程序,并将扩展程序部署为一个应用服务。详情请参见开发部署扩展程序。

3. 测试应用扩展程序。

扩展程序开发部署完成后,您即可在DataWorks的测试工作空间中测试扩展程序是否符合预期,完成测试 后上线扩展程序,由管理员在使用的工作空间中启用扩展程序。详情请参见测试应用扩展程序。

5.4.2. 准备工作

本文为您介绍开发使用扩展程序前的准备工作。

使用限制

• 仅DataWorks企业版及以上版本支持扩展程序功能。

⑦ 说明 如果使用的企业版DataWorks版本到期后,所有扩展程序均会失效,无法再触发事件检查。 已触发且未到终态的检查会自动通过。

- 扩展程序注册并开启后,多个扩展程序可关联同一个触发事件,即同一个事件支持触发多个扩展程序。
- 有内部节点的组合类节点(机器学习(PAI)节点、do-while节点、for-each节点)触发检查时,需要内部 节点都检查通过才可继续进行下一步操作。

准备工作1:开启消息订阅

扩展程序功能通常需要联合OpenEvent功能一起使用,通过OpenEvent来订阅扩展点事件,再通过扩展程序对 事件进行处理,因此在开发使用扩展程序前,您需先开启消息订阅,获取消息订阅的KafKa信息、订阅事件的 消息体格式,用于后续扩展程序的代码开发。 ● 开启消息订阅并获取Kafka相关信息的操作详情请参见开启消息订阅: DataWorks端(基于Kafka)。

OpenAPI	OpenEvent Exten	sions				
OpenEvent可将D	ataWorks中的事件消息通过	Kafka提供消息订阅服务。O	penEvent默认提供了Kafka Topic、用	I户、Group,启用消息订阅后,即	可以使用Kafka SDK来订阅DataWorks事件消息。	
启用满息订阅						
	-					
Kafka 接入点信息						
类型		网络	协议	接入	<u>_</u>	
SASL		VPC	SASL PLAINT	TEXT 192.	168.90.6:9094,192.168.90.8:9094,192.168.90.7:9094 🗎	单击复制
SSL		公网	SASL_SSL	106.	14.15.145:9093,139.224.117.213:9093,106.14.209.162:90	093 🖻
Kafka 用户信息						
用户名		密告 査者	<u>「用户信息</u>	操作	可删除后重新创建	
doctest		******** 复制 🗍 显示	重置	删除		
Kafka Topic 信息						
Торіс		订阅内容		操作	查看订阅内容详情	
topic_default	428543003249281	已订阅 1 个工作空间,	2 个事件类型	查看	订阅内容	
Kafka Group 信!	10 2-				公测期间最多只能创建 5 个Grou	p 🕂 新建Group
Group ID		描述		操作		
doctest				删除		

• 了解订阅事件的消息体格式的详情请参见附录: 消息格式。

准备工作2: 注册扩展程序

开发使用扩展程序前,您需要先在DataWorks上注册好一个扩展程序获取对应的Extension Code,用于后续的 扩展程序代码开发,操作步骤如下。

- 1. 登录DataWorks控制台并进入扩展程序页面。
 - i. 以阿里云主账号或拥有AliyunDataWorksFullAccess权限的RAM用户登录DataWorks控制台。
 - ii. 在左侧导航栏选择开放平台,在开放平台页面选择Extensions页签,进入扩展程序页面。

DataWorks		DataWorks / 开放平台
概范		● DataWorks开放平台由OpenAPI(开放API)、OpenEvent(开放涡思)组成,可适用于各类产品集成与被集成加场量。其中、OpenAPI已经商业化,开放涡息正在火热公园中,开吸至DataWorks企业宽即可参与公司。 责。
工作空间列表资源组列表		OpenAPI OpenEvent Extensions
报警配置	^	🛞 您的 DataWorks企业版 还有 15 天到期,2022年2月11日 起将会停止所思订阅,请及时续费。续费DataWorks企业版 查看开始消息文档
报警资源		+ 回 扫 序 2 篇
报警联系人		
开放平台 🔤 1		江南非 梁柏序 建 栗柏序 医理
计算引擎列表	^	④ 使用扩展程序必读: 1. 当心业饭过或,扩展程序被下线或茶用时,扩展程序构合大效,扩展程序控制的相关流程检查与磁塞体不起作用,相关流程系统将直接通过。
MaxCompute		2.扩展程序处于"测试中、"已上线"状态时,并不会置接生效,需要更到宏启用的工作空间中的"工作空间管理"页面中启用才能正式生效。 3.扩展程序处于"测试中状态,则只有"测试用工作空间可启用;处于"已上线"状态时,则是"适用工作空间范围"内的工作空间可启用,
Graph Compute		
Hologres		扩展程序名称 扩展点 适用模块 负责人 注册时间 Extension Code 操作
数据湖集成 (湖仓一体)		没有意识到符合条件的记录

2. 单击注册扩展程序,在弹窗中配置填的注册信息。

⑦ 说明 准备工作阶段,您仅需配置必填的注册信息,待后续扩展程序开发调试完成,部署后可再 进一步完善扩展程序的其他信息。

Dat aWorks

注册扩展程序 (Exte	ension)		×	* 适用模块 数据开发
* 扩展程序名称 1 doctest				* 负责人 @ 3 doctest
*处理的扩展点 💡				◎测试用工作空间 @
待选择		已选择		4 doc_test_2
请输入	Q	请输入	Q	
 代码运行前置事件 文件删除前置事件 		文件提交前置事件		17 展相学年前地址
表提交到开发前置 表提交到生产前置	事件 < 事件			扩展程序文档地址 🔮
文件发布前置事件				
5项		1项		9 展転予告気配置 ♥ (法論入内容、参約格式为 key=value、一行一个参約)
* 所属事件 文件变更事件				HETELY COLLET, prophylocoller of senses (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,

其中,准备工作阶段必填的参数有扩展程序名称、处理的扩展点、负责人和测试用工作空间,其他参数可 在扩展程序开发部署完成后进一步完善。详细的参数解释如下。

参数	如何配置
扩展程序名称	自定义扩展程序的名称,用于标识扩展程序。
	选择该扩展程序用于处理哪几类扩展事件触发的消息。当前支持的扩展点及其详情请参见支持的扩展点事件列表,您可根据实际情况选择。
处理的扩展点	 ⑦ 说明 选择完成后,配置界面会自动根据选择结果匹配所属事件和适用模块,无需手动配置。
负责人	选择扩展程序的负责人,方便扩展程序使用者遇到问题时能及时联系到负责人。
测试用工作空间	指定用于扩展程序测试用的工作空间。 扩展程序无需进行上线操作,即可在测试用工作空间中生效。 即开发人员可以在上线前,在测试空间中进行完整链路的测试验证,通过触发事件, 来测试DataWorks通过Kafka发送消息、扩展程序接收消息并进行消息审核与回调。
扩展程序详情地址	填写介绍扩展程序详情的地址,用于帮助扩展程序的用户更好的理解和使用此扩展程序。 您可以在开发部署扩展程序时,同时开发一个扩展程序的详情展示页面,完成后把页 面地址配置在此处,以便使用者在触发扩展程序校验时,可通过此链接查看扩展程序 完整的校验过程,例如,此次扩展程序检查链路和阻塞的原因。
扩展程序文档地址	填写扩展程序的帮助文档地址,供扩展程序的使用者阅读。 您可以在开发部署扩展程序时,同时开发一个扩展程序的帮助文档页面,完成后把页 面地址配置在此处,以便使用者学习了解扩展程序的校验逻辑与属性。

参数	如何配置					
扩展程序参数配置	DataWorks支持在扩展程序开发过程中使用参数来提高扩展程序开发和应用效率,并 为您提供了典型应用场景的内置参数供您直接使用,您也可以自定义参数,在扩展程 序代码开发时,将需要应用的参数添加到此处即可。 支持添加多个参数,一行一个参数,参数格式为 key=value 格式。更多参数的使 用详情和要求请参见参考:扩展程序参数配置。					
扩展程序选项配置	扩展程序开发者可通过该选项定义可供扩展程序使用者使用的功能配置项,扩展程序 选项配置可实现该扩展程序选项,在不同工作空间进行个性化管控。扩展程序开发者 需要在此界面通过JSON字符串定义选项。例如:您可以通过选项配置让扩展程序使用 者自行管控sql长度。JSON格式可参考文档:参考:扩展程序选项配置。					

3. 单击确定,完成扩展程序注册。

此时您可以在扩展程序页面查看到扩展程序的Extension Code。

enAPI OpenEv	ent Extensions									
您的 DataWorks企业版 还有 15 天到期,2022年2月11日 起将会停止消息订阅,请及时续费。续费DataWorks企业版 查看开放消息文档										
展程序设置注册扩展程序	扩展程序管理									
 使用扩展程序必读: 生企业版过期、扩展程序被下线或禁用时,扩展程序被会失效,扩展程序控制的相关流程检查与超塞将不起作用,相关流程系统将直接通过。 扩展程序处于"测试中"、"它上线"状态时,并不会直接生效,需要到欧启用的工作空间中的"工作空间管理"页面中启用才能正式生效。 扩展程序处于"测试中"状态,则只有"测试用工作空间"句启用;处于"已上线"状态时,则是"适用工作空间范围"内的工作空间同启用。 										
扩展程序名称	扩展点	注册时间	Extension Code	状态	操作					

后续步骤

完成准备工作后,您即已获取开发扩展程序所需的Kafka信息、事件消息体格式、Extension Code等必要信息,后续可开展扩展程序代码开发及部署工作,详情请参见开发部署扩展程序。

5.4.3. 开发部署扩展程序

DataWorks支持您使用自定义的扩展程序来处理订阅的扩展点事件,并将处理结果通过OpenAPl返回给 DataWorks。本文为您介绍开发自定义扩展程序的注意事项和流程。

前提条件

已完成准备工作,即已开启消息订阅、注册扩展程序的操作,并获取开发扩展程序所需的信息,详情请参见<mark>准</mark> 备工作。

开发扩展程序流程

以订阅处理DataStudio中的扩展点事件为例,扩展程序用来处理订阅的扩展点事件时,实际实现的完整底层 消息时序图如下所示。



根据上图可见,进行扩展程序相关代码开发有以下几个注意事项。

- 1. 订阅并解析扩展点事件消息。
 - 订阅扩展点事件消息的代码开发示例请参见开发消息订阅:用户端。
 - 解析扩展点事件消息时,可参见<mark>附录:消息格式</mark>了解不同事件消息的格式详情。
- 2. 获取扩展点事件详情并处理。
 - 您可通过调用API, 指定API中的 messageId 参数来获取详细的事件详情。

? 说明

- DataStudio中的扩展点事件可使用GetIDEEventDetail这个API。
- messageId 的参数取值可通过解析订阅获取的事件消息获取。
- 在扩展程序代码开发过程中,您可以使用参数来提高开发和应用效率,例如,您可以使用内置扩展程 序参数 extension.project.disabled ,使扩展程序对某个工作空间不生效。更多扩展程序参数的使

用配置请参见参考:扩展程序参数配置。

- 扩展程序对各类扩展点事件的处理逻辑示例可参见应用示例:消息处理代码DEMO。
- 3. 返回处理结果至DataWorks。

您可通过API将扩展程序对扩展点事件的处理结果返回给DataWorks。

⑦ 说明 DataStudio中的扩展点事件可使用UpdateIDEEvent Result这个API。

扩展程序示例

本示例使用Java语言介绍用户侧的拓展程序开发流程,您可以参考示例根据实际情况进行开发。

1. 添加Pom依赖、Kafka依赖。

添加KAFKA依赖
<dependency></dependency>
<proupid>org.apache.kafka</proupid>
<artifactid>kafka-clients</artifactid>
<version>2.2.0</version>
<exclusions></exclusions>
<exclusion></exclusion>
<groupid>org.slf4j</groupid>
<artifactid>slf4j-api</artifactid>
添加API依赖
<dependency></dependency>
<proupid>com.aliyun</proupid>
<artifactid>aliyun-java-sdk-core</artifactid>
<version>4.5.1</version>
<dependency></dependency>
<proupid>com.aliyun</proupid>
<pre><artifactid>aliyun-java-sdk-dataworks-public</artifactid></pre>
<version>3.4.14</version>

2. 增加处理并回调逻辑代码示例

```
package com.aliyun.openservices.kafka.ons;
import java.util.ArayList;
import java.util.List;
import java.util.Properties;
import com.alibaba.fastjson.JSONObject;
import com.aliyuncs.DefaultAcsClient;
import com.aliyuncs.IAcsClient;
import com.aliyuncs.dataworks_public.model.v20200518.UpdateWorkbenchEventResultRequest;
import com.aliyuncs.dataworks_public.model.v20200518.UpdateWorkbenchEventResultResponse;
import com.aliyuncs.exceptions.ClientException;
import com.aliyuncs.profile.DefaultProfile;
import com.aliyuncs.profile.DefaultProfile;
import org.apache.kafka.clients.CommonClientConfigs;
import org.apache.kafka.clients.consumer.ConsumerConfig;
```

import org.apache.kafka.clients.consumer.ConsumerRecord; import org.apache.kafka.clients.consumer.ConsumerRecords; import org.apache.kafka.clients.consumer.KafkaConsumer; import org.apache.kafka.clients.producer.ProducerConfig; import org.apache.kafka.common.config.SaslConfigs; import org.apache.kafka.common.config.SslConfigs; public class KafkaConsumerDemo { public static void main(String args[]) { //设置sasl文件的路径 JavaKafkaConfigurer.configureSasl(); //加载kafka.properties Properties kafkaProperties = JavaKafkaConfigurer.getKafkaProperties(); Properties props = new Properties(); //设置接入点,请通过控制台获取对应Topic的接入点 props.put(ProducerConfig.BOOTSTRAP SERVERS CONFIG, kafkaProperties.getProperty("b ootstrap.servers")); //设置SSL根证书的路径,请记得将XXX修改为自己的路径 //与sasl路径类似,该文件也不能被打包到jar中 props.put(SslConfigs.SSL TRUSTSTORE LOCATION CONFIG, kafkaProperties.getProperty("ssl.truststore.location")); //根证书store的密码,保持不变 props.put(SslConfigs.SSL TRUSTSTORE_PASSWORD_CONFIG, "KafkaOnsClient"); //接入协议,目前支持使用SASL SSL协议接入 props.put(CommonClientConfigs.SECURITY PROTOCOL CONFIG, "SASL SSL"); //SASL鉴权方式,保持不变 props.put(SaslConfigs.SASL MECHANISM, "PLAIN"); //两次poll之间的最大允许间隔 //可根据实际拉取数据和客户的版本等设置此值,默认30s props.put (ConsumerConfig.SESSION TIMEOUT MS CONFIG, 30000); //设置单次拉取的量,走公网访问时,该参数会有较大影响 props.put (ConsumerConfig.MAX PARTITION FETCH BYTES CONFIG, 32000); props.put(ConsumerConfig.FETCH MAX BYTES CONFIG, 32000); //每次poll的最大数量 //注意该值不要改得太大,如果poll太多数据,而不能在下次poll之前消费完,则会触发一次负载均衡 ,产生卡顿 props.put (ConsumerConfig.MAX POLL RECORDS CONFIG, 30); //消息的反序列化方式 props.put(ConsumerConfig.KEY DESERIALIZER CLASS CONFIG, "org.apache.kafka.common. serialization.StringDeserializer"); props.put (ConsumerConfig.VALUE DESERIALIZER CLASS CONFIG, "org.apache.kafka.commo n.serialization.StringDeserializer"); //当前消费实例所属的消费组,请在控制台申请之后填写 //属于同一个组的消费实例,会负载消费消息 props.put(ConsumerConfig.GROUP ID CONFIG, kafkaProperties.getProperty("group.id")); //hostname校验改成空 props.put(SslConfigs.SSL ENDPOINT IDENTIFICATION ALGORITHM CONFIG, ""); //构造消息对象,也即生成一个消费实例 KafkaConsumer<String, String> consumer = new org.apache.kafka.clients.consumer.Ka fkaConsumer<String, String>(props); //设置消费组订阅的Topic,可以订阅多个 //如果GROUP ID CONFIG是一样,则订阅的Topic也建议设置成一样 List<String> subscribedTopics = new ArrayList<String>(); //如果需要订阅多个Topic,则在这里add进去即可 //每个Topic需要先在控制台进行创建

```
subscribedTopics.add(kafkaProperties.getProperty("topic"));
       consumer.subscribe(subscribedTopics);
       //循环消费消息
       while (true) {
           try {
               ConsumerRecords<String, String> records = consumer.poll(1000);
               //必须在下次poll之前消费完这些数据,且总耗时不得超过SESSION TIMEOUT MS CONFIG
               //建议开一个单独的线程池来消费消息,然后异步返回结果
               for (ConsumerRecord<String, String> record : records) {
                   System.out.println(String.format("Consume partition:%d offset:%d", re
cord.partition(), record.offset()));
                   System.out.println("Consume message:" + record.key() + "," + record.v
alue());
                   //1、解析结果 消息结构参考https://help.aliyun.com/document_detail/21536
7.htmlhttps://www.alibabacloud.com/help/dataworks/latest/appendix-message-formats
                   JSONObject messageContent = JSONObject.parseObject(record.value());
                   String messageId = messageContent.getString("messageId");
                   //2、自定义逻辑处理
                   //doSomeThing();
                   //3、回调API,此处仅做代码示意,请按最佳实践原则来编码
                   DefaultProfile profile = DefaultProfile.getProfile("cn-shanghai", "<y</pre>
our-access-key-id>", "<your-access-key-secret>");
                   /** use STS Token
                    DefaultProfile profile = DefaultProfile.getProfile(
                    "<your-region-id>",
                                            // The region ID
                    "<your-access-key-id>",
                                                 // The AccessKey ID of the RAM account
                    "<your-access-key-secret>", // The AccessKey Secret of the RAM acc
ount
                    "<your-sts-token>");
                                                // STS Token
                    **/
                   IAcsClient client = new DefaultAcsClient(profile);
                   String extensioncode = "xxxx";
                   UpdateWorkbenchEventResultRequest request = new UpdateWorkbenchEventR
esultRequest();
                   request.setMessageId(messageId);
                   request.setExtensionCode(extensioncode);
                   //允许流程通过设置OK,不允许通过设置为FAIL
                   request.setCheckResult("OK");
                   try {
                       UpdateWorkbenchEventResultResponse response = client.getAcsRespon
se(request);
                       System.out.println(new Gson().toJson(response));
                   } catch (ServerException e) {
                       e.printStackTrace();
                   } catch (ClientException e) {
                       System.out.println("ErrCode:" + e.getErrCode());
                       System.out.println("ErrMsg:" + e.getErrMsg());
                       System.out.println("RequestId:" + e.getRequestId());
                   }
           } catch (Exception e) {
               trv {
                   Thread.sleep(1000);
               } catch (Throwable ignore) {
```



部署扩展程序

扩展程序代码开发调试完成后,您可以将其部署在本地或阿里云ECS上,部署为一个应用服务,用于后续使用。同时您也可以开发一个扩展程序的在线信息查询的页面和帮助文档页面,用于查看每一个事件的处理过程和状态、学习了解扩展程序的逻辑。后续您可以将页面地址完善到扩展程序的注册信息中,并将扩展程序中使用了的参数添加到扩展程序参数配置中。

☰ (-) 阿里云 ☆	工作台 💿 将东2(上海) 🗸			Q 搜索		费用 工单	ICP 备案 企业	业 支持	App 📴	Δ.	© #
DataWorks	DataWorks / 开放平台		注册扩展程序 (Extension)		×				7	*品动态	帮助文档
概范页	开放平台		* 测试用工作空间 🕖		^						
工作空间列表	① DataWorks开放平台由C	DpenAPI (开放API) 、OpenEv	请选择	~	1	测中,升级至DataWo	ks企业版即可参与	还测.			
资源组列表	OpenAPI OpenSver	t Extensions	扩展程序详情地址 🕢								
报警配置へ	Openari										
报警资源	扩展程序设置		扩展程序文档地址 🕢								
报醫联系人	注册扩展程序 扩	"展程序管理									
开放平台 🔤	使用扩展程序必读: 1 当众业的时期 1		扩展程序参数配置 🕢								
计算引擎列表 ^	 2.扩展程序处于"测 3.扩展程序处于"测 	试中"、"已上线"状态时,并不 试中"状态,则只有"测试用Tf	请输入内容,参数指式为 key=value,一行一个!	鬱数	-						
MaxCompute	4. 扩展程序在上线机	前需要提交至平台审核,平台省									
Graph Compute	扩展程序名称	扩展点	扩展程序选项配置 🕢				状态		操作		
Hologres		文件发布前置事件	通过JSON字符串描述		3		审批通	ist i	羊情 Sepa		上线
数据湖集成(湖仓一体)		文件发布前置事件			7		刺滅中	1	羊情 编辑	1 1 <u>校</u>	上线(:
		文件提交前置事件	描述				测试中	1	羊情 编辑	1 1 1 1 1 1	上线丨:
		代码运行前置事件	遺驗入		-		测试中	1	羊肉 编辑	1 1 1 1 1 1	上线 :
		代码运行前置事件		确认 取消			审批未		羊肉 網絡	1 1 1 1 1 1	上线 :
		文件提交前置事件					审批未		羊情 编辑	i 1这	上线
		文件提交前置事件	数据开发 dataworks_demo2	2022年4月20日 22:02:42 93ef5db	0c83743		审批来	it 🔺 i	新聞 潮報	i 11交	上线

扩展程序注册信息各参数的详情如下。

参数	如何配置
扩展程序名称	自定义扩展程序的名称,用于标识扩展程序。
	选择该扩展程序用于处理哪几类扩展事件触发的消息。当前支持的扩展点及其详情请参 见 <mark>支持的扩展点事件列表</mark> ,您可根据实际情况选择。
处理的扩展点	⑦ 说明 选择完成后,配置界面会自动根据选择结果匹配所属事件和适用模块,无需手动配置。
负责人	选择扩展程序的负责人,方便扩展程序使用者遇到问题时能及时联系到负责人。
测试用工作空间	指定用于扩展程序测试用的工作空间。 扩展程序无需进行上线操作,即可在测试用工作空间中生效。 即开发人员可以在上线前,在测试空间中进行完整链路的测试验证,通过触发事件,来测 试DataWorks通过Kafka发送消息、扩展程序接收消息并进行消息审核与回调。

参数	如何配置
扩展程序详情地址	填写介绍扩展程序详情的地址,用于帮助扩展程序的用户更好的理解和使用此扩展程序。 您可以在开发部署扩展程序时,同时开发一个扩展程序的详情展示页面,完成后把页面地 址配置在此处,以便使用者在触发扩展程序校验时,可通过此链接查看扩展程序完整的校 验过程,例如,此次扩展程序检查链路和阻塞的原因。
扩展程序文档地址	填写扩展程序的帮助文档地址,供扩展程序的使用者阅读。 您可以在开发部署扩展程序时,同时开发一个扩展程序的帮助文档页面,完成后把页面地 址配置在此处,以便使用者学习了解扩展程序的校验逻辑与属性。
扩展程序参数配置	DataWorks支持在扩展程序开发过程中使用参数来提高扩展程序开发和应用效率,并为您 提供了典型应用场景的内置参数供您直接使用,您也可以自定义参数,在扩展程序代码开 发时,将需要应用的参数添加到此处即可。 支持添加多个参数,一行一个参数,参数格式为 key=value 格式。更多参数的使用详 情和要求请参见参考:扩展程序参数配置。
扩展程序选项配置	扩展程序开发者可通过该选项定义可供扩展程序使用者使用的功能配置项,扩展程序选项 配置可实现该扩展程序选项,在不同工作空间进行个性化管控。扩展程序开发者需要在此 界面通过JSON字符串定义选项。例如:您可以通过选项配置让扩展程序使用者自行管控 sql长度。JSON格式可参考文档:参考:扩展程序选项配置。

5.4.4. 测试应用扩展程序

完成扩展程序的开发、部署后,您可以在DataWorks的扩展程序页面对扩展程序进行测试验证,验证完成后上线扩展程序,上线后管理员即可启用。本文为您介绍如何测试并上线扩展程序。

前提条件

- 已完成准备工作,即已开启消息订阅、注册扩展程序的操作,并获取开发扩展程序所需的信息,详情请参见准备工作。
- 已完成开发部署扩展程序,操作详情请参见开发部署扩展程序。

背景信息

在注册扩展程序时,您已指定进行扩展程序测试的工作空间,在完成上述前提条件后,您可直接在测试工作空间内启用扩展程序并进行测试验证,验证完成后将扩展程序上线,完成上线后,管理员即可对所有工作空间选择是否应用此扩展程序。

- 1. 测试操作请参见下文的测试扩展程序。
- 2. 上线操作请参见下文的上线与管理扩展程序。
- 3. 启用操作请参见下文的启用扩展程序。

测试扩展程序

1. 查看测试工作空间。

登录DataWorks控制台,单击**开放平台 > Extensions**进入扩展程序页面,在对应扩展程序的**详情**中查看 确认测试工作空间是哪个工作空间。

Dat aWorks

☰ €Э阿里云	合 工作台	◎ 华东2 (上海	ij) ✔ Q 搜索	费用 工单 ICP 备案	《 企业 支持 》	App 🗈 🛕	日 ⑦ 简体 🌔
DataWorks	Da	ataWorks / 开放平台	注册扩展程序 (Extension)	×			产品动态
戰克 工作空间列表 资源组列表		 DataWorks开放参与公测。 OpenAPI Op 	 ※前近用工作空间 	×	nAPI已经商业化,	开放消息正在火热	松测中,升极至DataWorks企
报警配置 报警资源 报警联系人	^	 您的 DataWork 扩展程序设置 注册扩展程序 	扩展程序文档地址 🕢		消息文档		
开放平台 @@@ 计算引擎列表 MaxCompute	^	• 使用扩展程 1. 当企业版 2. 扩展程序 3. 扩展程序	扩展程序参数配置 ● 请输入内容,参数指式为 key=value,一行一个参数		系统将直接通过。 i式生效。 可启用。		
Hologres		扩展程序名称 doctest	描述		0f3bf224d4ae	状态测试中	操作 编辑上线 努 🗈
数照朔集成(湖仓一体)		•		RC28			< 1-70 [F]

- 2. 进入测试工作空间的空间配置页面。
 - i. 登录DataWorks控制台,单击工作空间列表,进入工作空间列表页。
 - ii. 在测试工作空间后单击工作空间配置 > 更多配置,进入工作空间配置页。
- 3. 启用扩展程序。

单击左侧导航栏的**扩展程序配置**后,打开启用扩展程序开关,在弹出的页面中统一授权,即可对测试工 作空间启用此扩展程序。

📃 🏟 Data	Works	doc_test_2 J doc_test_2	۱.,					d 🗗 🕄	🗢 ana jama pad	ejund son 🗄
三 〇 工作空间配置		注册扩展程序								
基 成员管理		名称	标识符		扩展点	来源	负责人			启用
农限列表			0.000.0040-004040	CERCORAN	文件提交前置事件	自定义	santie	未启用	2022-01-26 16:33:49	2
MaxCompute测	· 保配置									
③ 数据源管理	_									
, 开源集群管理	_				an den		v			
扩展程序配置				扩展程序数据	受权		^			
				扩展程序【docte • 文件变更事件 (st】请求授权本工作空间中的以下事件满思和Op OpenAPE UpdateIDEEventResult)	enAPI:				
	_			若启用该扩展程序 能根据事件满息和	《系统将会把本工作空间中的以上事件满息和Ope 10penAPI对DataWorks相关流程进行逻辑处理和	nAPI授权给该扩展程序,该扩加 响应。	集程] 库可			
			8	找己知晓并想	象受以上数据授权所可能产生的网络					
						4 同意授权	INCOME.			

4. 测试扩展程序。

根据扩展程序代码中订阅处理的扩展点事件类型,在测试工作空间中触发此类事件,验证扩展程序是否生效、处理结果是否符合预期。

例如,扩展程序是用于管控节点提交这一扩展点事件时,您即可进入测试工作空间的DataStudio页面 后,操作节点提交,测试是否触发扩展程序进行流程卡点及处理。

(7)	操作检查			C	≤ 触发文件	提交前置事	件检查 ×												
G	类型:全部 、 状态:全部 、				•	জি ।	5]			Þ		C		8	D %	\$	പ്പ		
					提交文件,操作检查中>														
~	名称/操作时间	类型	状态	Â						*****									
4	触发文件提交前置事件检 查 2021-10-22 11:57:00	提交文件	检查中 >	F		au	2 检查	≠ <u></u> 2						×					
亩 fx	触发代码运行前置事件检 查 2021-10-20 17:17:12	运行				sele	检查〕	过后来	<u>切后将自动执行操作,请在[操作检查]</u> 中查看结果。										
	xc_table_prod 2021-10-20 16:36:00	表提交到开发																	
•	扩展程序_pyodps_代码运 行前置事件 2021-10-20 16:25:34	运行	检查中 >																
操作检	ē 3				>	¢		点	击提交	触发扩展	展程序检	渣(将触发	工作空	间目前所开	启的	所有文件	提交前置事	
名称: 触发文件提交前置事件检查							件	的扩展	程序校验	佥)									
12:021	地型							2 可	以进入	创建发布	市包界面	重查看文	2件检查	E 状态					
提交	前位查 测试程序	检查中	查看详情 查看文档		重新检查 关闭	1		3 可验	点击状 详情	态查看相	交验详情	青, 可右	E此查看	雪当前 掛	操作触发的	扩展,	程序校验到	列表及校	

更多触发示例请参见应用示例:数据开发过程触发事件检查。

上线与管理扩展程序

扩展程序测试完成后,您即可进行上线操作,完成上线后管理员可以在适用的工作空间中启用对应的扩展程序。

您可在开放平台的扩展程序页签中,对已测试完成并且经平台审批通过的扩展程序,在该扩展程序的操 作列,单击上线将扩展程序进行上线。

OpenAPI OpenEve	nt Extensions										
您的 DataWorks企业版还有 15 天到期,2022年2月11日起将会停止满思订阅,请及时续费。续费DataWorks企业版 查看开放满意文档											
扩展程序设置 注册扩展程序 扩展程序管理											
 使用扩展程序必该: 当企业版过期,扩展程序被下纸或菜用时,扩展程序就会失效,扩展程序控制的相关流程检查与阻塞将不起作用,相关流程系统将直接通过。 扩展程序处于"测试中"、"已上纸"状态时,并不会直接生效,需要到欲启用的工作空间中的"工作空间管理"页面中启用才能正式生效。 扩展程序处于"测试中"状态,则只有"测试用工作空间"可启用;处于"已上线"状态时,则是"适用工作空间范围"内的工作空间可启用。 											
扩展程序名称	扩展点	适用模块	负责人	注册时间	Extension Code	操作					
doctest	文件提交前置事件	数据开发	doctest	2022年1月26日 16:33:49	0afb5c3ab49247b49e(详情编辑 上线 禁用 删除	•				

同时您也可以查看当前扩展程序的状态、查看扩展程序的详情,并对扩展程序进行编辑、禁用、删除的操作。

扩展程序的状态包括:

- 测试中:新注册的扩展程序默认为测试中,已上线的扩展程序下线后也会变为测试中的状态,测试中的扩展程序仅测试空间可使用。
- 审批中:测试中的扩展程序点击提交后,将处于审批中的状态,经平台审批通过的扩展程序状态将变更为审批通过,审批通过后扩展程序才可以进行上线操作。若平台审批未通过,该扩展程序的状态将变更为审批未通过。
- 已上线:在扩展程序的操作列,单击上线后,扩展程序状态变为已上线,上线后,所有扩展程序适用的工作空间都可以使用扩展程序(空间管理员可在工作空间配置界面查看并启用该扩展程序)。
- 已禁用:在扩展程序的操作列,单击禁用后,扩展程序状态变为已禁用,此时该扩展程序失效,启用此扩展程序的工作空间的触发事件流程直接通过,不会触发检查或阻断。

 已过期:当DataWorks企业版版本过期后,扩展程序状态会变更为已过期状态,此状态的扩展程序失效, 与已禁用状态类似。

启用扩展程序

扩展程序上线后,管理员即可通过查看扩展程序详情、与扩展程序负责人了解扩展程序详情,在适用的工作空间中启用扩展程序,启用后,对应工作空间中的扩展事件即会被管控。启用操作如下。

- 1. 进入待应用的工作空间的空间配置页面。
 - i. 登录DataWorks控制台,单击工作空间列表,进入工作空间列表页。
 - ii. 在工作空间后单击工作空间配置 > 更多配置,进入工作空间配置页。
- 2. 启用扩展程序。

单击左侧导航栏的**扩展程序配置**后,打开启用扩展程序开关,在弹出的页面中统一授权,即可对工作空间启用此扩展程序。



⑦ 说明 若您的扩展程序存在选项配置,您可在启用扩展程序前点击设置按钮先对进行选项配置。

5.4.5. 附录

背景信息

1.

5.4.5.1. 参考: 扩展程序参数配置

开发扩展程序代码时, DataWorks为您提供了内置参数, 同时也支持您自定义参数, 以提高扩展程序代码开发 效率和效果, 例如, 结合参数实现扩展程序仅对指定工作空间生效。本文为您介绍如何使用扩展程序的参数。

背景信息

使用扩展程序参数进行代码开发时,您需要在扩展程序的注册信息中定义好扩展程序参数,将后续可能需要在 扩展程序代码开发过程中使用的内置参数或自定义参数添加到扩展程序的注册信息中。扩展程序开发者可通过 调用APIGetExtension获得该扩展程序的参数配置。

三 (一) 阿里云 命 工作台 💿	华东2(上海) ~	Q 搜索 费用 工单	ICP 备案 企业 支持 App 🖸 🎝 🗟 🗑 🔞 简体 🧐
DataWorks DataWorks	/ 开放平台 注册扩展程序 (Extension)	×	产品动态 帮助文档
概题 Data	Works开放平台曲OpenAPI(开放API)	1 正在火烧公司	中,升极至DataWorks企业版的可参与公
工作空间列表	● 负责人 😧		
资源组列表	PI OpenEvent Extension: doctest (santie)	~	
1997.2 · · · · · · · · · · · · · · · · · · ·	DataWorks企业版 还有 15 天到期, 2 * 测试用工作空间 ●		
报警资源	doc_test_2		
步 展程 报警联系人	扩展程序详情地址 •		
开放平台 [2000]	87.00442195 B.200442195182.00		
计算引擎列表	使用扩展程序必读: 1.当企业版过期,扩展程序被下结成基		
MaxCompute	2.扩展程序处于"测试中"、"已上线"状2 3.扩展程序处于"测试中"状态,则只有"		
Graph Compute	扩展程序参数配置 😡		
扩展	程序名称 扩展点 请输入内容,参数格式为 key=value,一行一个	·参数 sion Code	18/1
数据湖集成(湖仓一体)	est 又件提交所置4	c3ab49247b4	Jee (FO) SALA LEE AND MIRE
	描述		く 上一页 1 下一页 >
	词输入		
			(attraction of the second s
		1650 人命	
			6

- 支持添加多个参数,一个参数一行。
- 添加内置参数时:

内置参数格式为 key=value 的格式,当前支持的内置参数有:

- 内置参数1: extension.project.disabled, 工作空间黑名单配置
- 内置参数2: extension.project.disabled.append,工作空间黑名单追加配置
- 内置参数3: extension.project.{eventCode}.disabled, 扩展点黑名单配置
- 内置参数4: extension.project.{eventCode}.disabled.append, 扩展点黑名单追加配置
- 添加自定义参数时:

```
建议也使用 key=value 的格式。
```

内置参数1: extension.project.disabled , 工作空间黑名单配置

• 应用场景:

通过此参数,使扩展程序对指定工作空间不生效。

参数说明:

extension.project.disabled=projectId1,projectId2,projectId3

其中, *project Id*为不生效的工作空间ID, 您可通过List Project s获取, 或在Dat aWorks控制台的工作空间配置 页面查看获取。

- ⑦ 说明 支持一个参数应用代码里指定多个工作空间ID,使用逗号(,)分隔。
- 代码示例:

extension.project.disable=projectId1

• 注意事项:

不能重复配置(即出现key相同的情况),如果重复配置,以最后一个配置结果为准。

内置参数2: extension.project.disabled.append , 工作空间黑名单追加配置

● 应用场景:

当 extension.project.disabled 参数没有多个value值时(即参数取值没有配置为逗号分隔的多个工作 空间ID),您可以通过此参数追加工作空间,使扩展程序对追加的指定工作空间不生效。

参数说明:

extension.project.disabled.append=projectId1,projectId2,projectId3

其中, *project Id*为不生效的工作空间ID, 您可通过List Project s获取, 或在Dat aWorks控制台的工作空间配置 页面查看获取。

⑦ 说明 支持一个参数应用代码里指定多个工作空间ID,使用逗号(,)分隔。

代码示例:

extension.project.disable.append=projectId1

内置参数3: extension.project.{eventCode}.disabled , 扩展点黑名单配置

• 应用场景:

通过此参数,使扩展程序对指定工作空间的扩展点不生效。

参数说明:

extension.project.{eventCode}.disabled=projectId1,projectId2,projectId3

其中

- event Code: 为不生效的扩展点事件类型, 取值请参见附录:事件详情(event Type)。
- *project Id*:为不生效的工作空间ID,您可通过List Project s获取,或在Dat aWorks控制台的工作空间配置页面查看获取。

⑦ 说明 支持一个参数应用代码里指定多个工作空间ID,使用逗号(,)分隔。

代码示例:

extension.project.commit-file.disable=projectId1
extension.project.delete-file.disable=projectId2,projectId3

● 注意事项:

不能重复配置(即出现key相同的情况),如果重复配置,以最后一个配置结果为准。

内置参数4: extension.project.{eventCode}.disabled.append , 扩展点黑名单追加配置

• 应用场景:

当 extension.project.{*eventCode*}.disabled.append 参数没有多个value值时(即参数取值没有配置为 逗号分隔的多个工作空间ID),您可以通过此参数追加工作空间,使扩展程序对追加的指定工作空间的扩展 点不生效。 ● 参数说明:

extension.project.{eventCode}.disabled.append=projectId1,projectId2,projectId3

其中

- event Code:为不生效的扩展点事件类型,取值请参见附录:事件详情(event Type)。
- *project Id*:为不生效的工作空间ID,您可通过List Project s获取,或在Dat aWorks控制台的工作空间配置页面查看获取。

⑦ 说明 支持一个参数应用代码里指定多个工作空间ID,使用逗号(,)分隔。

• 代码示例:

```
extension.project.commit-file.disable.append=projectId1
extension.project.delete-file.disable.append=projectId2,projectId3
```

5.4.5.2. 参考:扩展程序选项配置

扩展程序选项是由扩展程序开发者定义,供扩展程序使用者使用的功能配置项,可实现该扩展程序选项在不同 工作空间进行个性化管控。例如,扩展程序开发者可以通过选项配置让扩展程序使用者自行管控SQL长度,实 现扩展程序在不同工作空间限制的SQL长度不同。

背景信息

- 扩展程序选项由扩展程序开发者在注册扩展程序时定义的该选项,供扩展程序使用者在具体工作空间下对 该扩展程序选项进行设置。
 - 扩展程序开发者注册扩展程序,详情请参见:准备工作2:注册扩展程序。
 - 扩展程序使用者在空间内使用扩展程序时,可在启用扩展程序前在下图位置配置选项。

≡	🕼 DataWorks							🖞 🛛 থ্	简体 💎	dataworks_demo2
	=									
۵	工作空间配置	注册扩展程序								
25	成员管理	名称	标识符	扩展点	来源	负责人	状态	注册时间	启用	操作
0	权限列表	100 Carlos	A	文件发布前置事件,代码运行前置事件	自定 义	dataworks_de mo2	未启 用	2022-04-26 15:51:02		12mm
~	MaxCompute高级配置	-	in all second second	文件发布前置事件,文件提交前置事件,代 码运行前置事件	自定 义	dataworks_de mo2	未启 用	2022-04-26 15:48:48		12m
© ▶	奴斯該官理 开源集群管理	100 C	20.00	文件提交前置事件,代码运行前置事件	自定 义	xiangcui	未启 用	2022-04-26 11:44:50		12m
۲	扩展程序配置	Sec. 2	10 million (1999)	代码运行前置事件,文件提交前置事件	自定义	xiangcui	未启用	2022-04-26 10:58:40		12 111
					dauda		ata ata			

● 扩展程序开发者可通过APIGet OptionValueForProject获取扩展程序使用者(空间管理员)在具体工作空间 内该选项的配置。

注意事项

扩展程序选项配置信息及相关文案由扩展程序开发者提供,平台不为内容风险负责。

支持的组件
目前选项配置支持的组件有七种:文本、单行输入框、多行输入框、单选选择下拉、多选选择下拉、复选和单选组件。如下图所示:

文本 (?):	文本内容	
单行输入框 (?):		
多行輸入框 ⑦:		
单选选择下拉 ②:	请选择	~
多选选择下拉 ②:	请选择	~
复选 ②:	opt1 opt2 opt3	
单选 ②:	O opt1 O opt2 O opt3	

选项配置

,

扩展程序开发者需要在注册扩展程序时,通过JSON字符串方式在选项配置中定义选项。

t
"type": "object",
"properties": {
"组件名称(不可重复)":{
"type": "string",
"title ": "填写见下表",
"x-decorator": "填写见《参数解析》表格内容",
"x-component": "填写见《参数解析》表格内容",
"x-decorator-props": {
"tooltip": "描述"
},
"x-component-props": {
"dataSource": "填写见《参数解析》表格内容",
"mode": "multiple"
}
}
}
}

您可以参考以下内容进行配置:

- 组件相关参数解析,详情请参见:参数解析。
- 组件配置示例,详情请参见:配置示例:设置SQL任务消耗的CU资源阈值、配置示例:使用7类组件。
- 注册与管理扩展程序,详情请参见: 准备工作2: 注册扩展程序。

参数解析

字段	组件类型	值
	文本	文本
	单行输入框	单行输入框

字段	组件类型	值		
title	多行输入框	多行输入框		
	单选选择下拉	单选选择下拉		
	多选选择下拉	多选选择下拉		
	复选	复选		
	单选	单选		
	文本	Formltem		
	单行输入框	Formltem		
	多行输入框	Formltem		
x-decorator	单选选择下拉	Formltem		
	多选选择下拉	Formitem		
	复选	Formitem		
	单选	Formltem		
	文本	PreviewText.Input		
	单行输入框	Input		
	多行输入框	Input.TextArea		
x-component	单选选择下拉	Select		
	多选选择下拉	Select		
	复选	Checkbox.Group		
	单选	Radio.Group		
	文本	可不填		
	单行输入框	可不填		
	多行输入框	可不填		

Dat aWorks

字段	组件类型	值
dataSource	单选选择下拉	如有需要可填写: (其中每一对大括号都是 唯一的, value和label不可重复。label对应 的值为展示到界面的文案, value的值是文案 对应的标记, 自行定义即可。以下为示例代 码: [{ "value": "10001", "label": "opt1" }, { "value": "10001", "label": "opt1" }]
	多选选择下拉	如有需要可填写: (其中每一对大括号都是 唯一的, value和label不可重复。label的值 是展示到界面的文案, value的值是文案对应 的标记, 自行定义即可。以下为示例代码: [{ "value": "10001", "label": "opt1" }, { "value": "10001", "label": "opt1" }]
	复选	如有需要可填写: (其中每一对大括号都是 唯一的, value和label不可重复。label的值 是展示到界面的文案, value的值是文案对应 的标记, 自行定义即可。以下为示例代码: [{ "value": "10001", "label": "opt1" }, { "value": "10001", "label": "opt1" }]

字段	组件类型	值
		如有需要可填写:(其中每一对大括号都是 唯一的,value和label不可重复。label的值 是展示到界面的文案,value的值是文案对应 的标记,自行定义即可。以下为示例代码:
	单洗	[{ "value": "10001",
	+22	"label": "optl" }, { "value": "10001", "label": "optl" }]

配置示例:设置SQL任务消耗的CU资源阈值

下图为设置SQL任务消耗的CU资源阈值配置过程与最终效果示例。

注册扩展程序 (Extension)	×								
扩展程序文档地址 🕢									*
		标识符	扩展点	来源	负责人	状态	注册时间	扁用	操作
扩展程序参数配置 🕢			文件提交前置事件,代码运行前置事件	目定义	xiangcul	末启用	2022-04-26 11:44:50	۲	-
语检入由实 参数终于为 key-value行个参数		が 展程序选项 eseb			~	末慮用	2022-04-26 10 58:40		设置
19982 (F31F) (#96.113.025 No)=9000() 13 1 #960		取 经 XX 未 a12a 选项			帮助文档 02	未启用	2022-04-19 20:35:50	•	10 M
选项配置	11	2635 0 * EMAQUER REX	2字信息由本扩展程序开发者目立义提供,平台	不为内容风险负	5. 02	未启用	2022-03-07 14:55:44	•	10 2
扩展程序选项配置 🕢		Cu数调值 ③:				未腐用	2022-02-21 19:16:26	•	Q.M.
("type`robject","properties";["cuNumber";["type`rstring","title";"Cu取调值","x- decorator";"Formitem","x-component";"Input","x-decorator-props";["tooltip";"请输入SQL任务 的词称(cu我调查")]);	\square	Syste 数据授权 • 文件变更事件 (OpenAPI: Update	DEEventResult)			末启用		•	10 2
		1617			02	未启用	2022-02-14 16:15:00		QШ.
描述		Syste				末启用	2022-02-10 16:54:46		
请输入		System_SimpleProcessingChecker	文件提交前置事件,文件发布前置事件	系统	SYSTEM	已启用	2021-09-18 14:16:38		设置
政	湖								

上述选项配置中的JSON配置如下。

{
"type": "object",
"properties": {
"cuNumber": {
"type": "string",
"title": "Cu 数阈值",
"x-decorator": "FormItem",
"x-component": "Input",
"x-decorator-props": {
"tooltip": "请输入SQL 任务的消耗 CU 数阈值 "
}
}
}
}

配置示例:使用7类组件

下图为使用7类组件的配置过程与效果示例。

注册扩展程序 (Extension)	×	ור	扩展程序选项	×
			最终效果	
扩展程序文档地址 🜒			选项	動文档
			1 本区域选项信息及相关文字信息由本扩展程序开发者自定义提供,平台不为内容风险负责。	
扩展程序参数配置 🕢			文本 ①: 文本内容	
请输入内容,参数格式为 key=value,一行一个参数			单行输入框 ②:	
选项配置			▶ 多行输入框 ②:	
扩展程序选项配置 🕢				
["type":"object"."properties":["text":["type":"string"."title":"这本"."x- decorator":"FormItem","x-component":"PreviewText.Input","x-decorator-props":			単选选择下拉 ③: 请选择 🗸	
('tooltp:?:抽还文件').'default:?:文本內容').'input':['type':string'.'title:?:車行報入框'.'x- decorator':FormHen'.'x-component':Input':Adecorator-props':['tooltp:?'抽述文 件').'x-component-props':(]).'textarea':['type':'string'.'title':'多行输入框'.'x-	•		多选选择下拉 ③: 请选择 🗸 🗸	
描述	- 1		复选 ②: opt1 opt2 opt3	
语检 λ			单选 ②: O opt1 O opt2 O opt3	
1998-1927 N	. 1			
	- 1		数据授权	
		r	● 文件变更事件 (OpenAPI: UpdateIDEEventResult)	
H الم	取消			
			取消	确定

上述选项配置中的JSON配置如下。

⑦ 说明 如果您需要一次性使用七个组件,可复制以下JSON并修改部分数据即可(如果是文本组件修改default,其余的组件可以修改dataSource或组件名称)。

```
{
   "type":"object",
   "properties":{
      "text":{
          "type":"string",
          "title":"文本",
          "x-decorator":"FormItem",
          "x-component":"PreviewText.Input",
          "x-decorator-props":{
             "tooltip":"描述文件"
          },
          "default":"yuayuan"//仅文本的默认值可以这样填写
       },
       "input":{
          "type":"string",
          "title":"单行输入框",
          "x-decorator":"FormItem",
          "x-component":"Input",
          "x-decorator-props":{
             "tooltip":"描述文件"
          },
          "x-component-props":{
          }
       },
       "textarea":{
          "type":"string",
          "title":"多行输入框",
          "x-decorator":"FormItem",
          "x-component":"Input.TextArea",
```

```
"x-decorator-props":{
       "tooltip":"描述文件"
   },
   "x-component-props":{
   }
},
"select":{
   "type":"string",
   "title":"单选选择下拉",
   "x-decorator":"FormItem",
   "x-component":"Select",
   "x-decorator-props":{
       "tooltip":"描述文件"
   },
   "x-component-props":{
       "dataSource":[
           {
               "value":"10001",
               "label":"opt1"
           },
            {
               "value":10002,
                "label":"opt2"
            },
            {
                "value":10003,
               "label":"opt3",
               "disabled":true
           }
       ]
   }
},
"selectmore":{
   "type":"string",
   "title":"多选选择下拉",
   "x-decorator":"FormItem",
   "x-component":"Select",
   "x-decorator-props":{
       "tooltip":"描述文件"
   },
   "x-component-props":{
       "dataSource":[
           {
               "value":"10001",
               "label":"opt1"
           },
            {
               "value":10002,
               "label":"opt2"
            },
            {
               "value":10003,
               "label":"opt3",
               "disabled":true
            }
```

```
],
        "mode":"multiple"
   }
},
"checkbox":{
   "type":"array",
   "title":"复选",
   "x-decorator":"FormItem",
   "x-component":"Checkbox.Group",
   "x-decorator-props":{
       "tooltip":"描述文件"
    },
    "x-component-props":{
        "dataSource":[
           {
                "value":"10001",
                "label":"opt1"
            },
            {
                "value":10002,
                "label":"opt2"
            },
            {
                "value":10003,
                "label":"opt3",
                "disabled":true
           }
       ],
       "mode":"multiple"
   }
},
"radio":{
   "type":"number",
   "title":"单选",
   "x-decorator":"FormItem",
    "x-component":"Radio.Group",
   "x-decorator-props":{
       "tooltip":"描述文件"
    },
    "x-component-props":{
        "dataSource":[
           {
                "value":"10001",
                "label":"opt1"
            },
            {
                "value":10002,
                "label":"opt2"
            },
            {
                "value":10003,
                "label":"opt3",
                "disabled":true
           }
       ],
```

```
"mode":"multiple"
}
}
}
```

5.4.5.3. 应用示例: 消息处理代码DEMO

完成开放消息的配置后,您可以通过本地开发并在DataWorks注册扩展程序的方式,来对指定工作空间下某些页面功能(即进行扩展点事件的相关操作)进行管控。扩展程序订阅Kafka后,DataWorks会将操作产生的事件消息通过Kafka发送至您的扩展程序,并根据扩展程序返回的结果来决定该操作是否生效。

前提条件

本文以具体场景示例扩展程序接收并处理扩展点事件消息,但您需要先了解以下几方面内容:

- 已经完成Kafka相关配置,确保DataWorks事件消息可以正常推送,详情可参考文档:开启消息订阅: DataWorks端(基于Kafka)。
- 已经根据开发消息订阅:用户端中提供的代码样例,配置好扩展程序订阅消息,确保扩展程序 (Consumer)可以正常接收DataWorks推送的消息。
- 已经了解扩展程序接收、处理事件消息,并将处理结果通过回调API返回DataWorks的完整链路,详情可参考文档:开发部署扩展程序。

背景信息

在Dat aWorks上,扩展点事件可通过扩展程序功能实现流程卡点与事件消息响应,避免因开发人员数据开发专业技能欠缺、数据质量保障意识薄弱,导致代码修改和开发后未进行严格测试便发布生产,或者进行其他高危操作(误删关键表、资源函数等)而出现故障的问题。将人为检视转变为程序审批(开发扩展程序并注册在Dat aWorks),避免这类问题的发生。



目前支持的扩展点,及扩展程序处理扩展点事件消息的示例如下:

支持的扩展点	扩展程序处理事件消息示例
文件提交扩展点	立
文件发布扩展点	又什旋文、及仰爭什府志处珪小例
表提交扩展点	
表发布扩展点	专担六 华左束供逃自从两二例
	衣旋父、 及 仰争针消忌处理亦例

支持的扩展点	扩展程序处理事件消息示例
文件删除扩展点	文件删除事件消息处理示例
代码运行扩展点	代码运行事件消息处理示例

使用限制

调用API的账号需要有AliyunDataWorksFullAccess权限。关于阿里云主账号授予RAM子账号 AliyunDataWorksFullAccess权限的操作详情可参考文档:DataWorks权限管控概述。

文件提交、发布事件消息处理示例

示例场景:当进行节点提交、发布操作时,如果节点代码中包含CREATE TABLE关键字,则抛出警告但不阻塞 流程;当节点代码中包含DROP TABLE关键字时,则阻塞节点提交、发布操作。

```
String messageId = ""; // messageId从消息中解析
String extensionCode = ""; // 扩展程序标识符从DataWorks页面查询
Long projectId = 0L; // 项目空间ID从消息中解析
IAcsClient client = null; // 初始化POP客户端
GetIDEEventDetailRequest request = new GetIDEEventDetailRequest();
request.setMessageId(messageId);
request.setProjectId(projectId);
GetIDEEventDetailResponse response = client.getAcsResponse(request);
GetIDEEventDetailResponse.EventDetail.CommittedFile dataSnapshot = response.getEventDetail().
getCommittedFile();
String checkResult = "OK";
String checkResultTip = "Success";
if (dataSnapshot.getContent().contains("CREATE TABLE")) {
   checkResult = "WARN";
    checkResultTip = "最好不要包含建表语句";
}
if (dataSnapshot.getContent().contains("DROP TABLE")) {
   checkResult = "FAIL";
    checkResultTip = "不能包含删表语句";
}
UpdateIDEEventResultRequest request = new UpdateIDEEventResultRequest();
request.setCheckResult(checkResult);
request.setExtensionCode (extensionCode);
request.setCheckResultTip(checkResultTip);
request.setMessageId(messageId);
client.getAcsResponse(request);
```

表提交、发布事件消息处理示例

示例场景:表提交或者发布时必须指定生命周期,否则不允许提交或发布。

```
String messageId = ""; // messageId从消息中解析
String extensionCode = ""; // 扩展程序标识符从DataWorks页面查询
Long projectId = 0L; // 项目空间ID从消息中解析
IAcsClient client = null; // 初始化POP客户端
GetIDEEventDetailRequest request = new GetIDEEventDetailRequest();
request.setMessageId(messageId);
request.setProjectId(projectId);
GetIDEEventDetailResponse response = client.getAcsResponse(request);
GetIDEEventDetailResponse.EventDetail.TableModel dataSnapshot = response.getEventDetail().get
TableModel();
String checkResult = "OK";
String checkResultTip = "Success";
if (dataSnapshot.getLifeCycle() == null || dataSnapshot.getLifeCycle() == 0) {
    checkResult = "FAIL";
    checkResultTip = "表必须指定生命周期";
}
UpdateIDEEventResultRequest request = new UpdateIDEEventResultRequest();
request.setCheckResult(checkResult);
request.setExtensionCode(extensionCode);
request.setCheckResultTip(checkResultTip);
request.setMessageId(messageId);
client.getAcsResponse(request);
```

文件删除事件消息处理示例

示例场景:工作空间下不允许任何文件(资源、函数、节点)被删除,当指定工作空间下进行文件删除操作时 将会被阻塞。

```
String messageId = ""; // messageId从消息中解析
String extensionCode = ""; // 扩展程序标识符从DataWorks页面查询
Long projectId = 0L; // 项目空间ID从消息中解析
IAcsClient client = null; // 初始化POP客户端
String checkResult = "FAIL";
String checkResultTip = "不允许删除文件";
UpdateIDEEventResultRequest request = new UpdateIDEEventResultRequest();
request.setCheckResult(checkResult);
request.setExtensionCode(extensionCode);
request.setCheckResultTip(checkResultTip);
request.setMessageId(messageId);
client.getAcsResponse(request);
```

代码运行事件消息处理示例

场景示例:当节点代码中包含CREATE TABLE关键字时,节点运行前将抛出警告;当节点代码中包含DROP TABLE关键字时,将阻塞节点运行操作。

```
String messageId = ""; // messageId从消息中解析
String extensionCode = ""; // 扩展程序标识符从DataWorks页面查询
Long projectId = 0L; // 项目空间ID从消息中解析
IAcsClient client = null; // 初始化POP客户端
GetIDEEventDetailRequest request = new GetIDEEventDetailRequest();
request.setMessageId(messageId);
request.setProjectId(projectId);
GetIDEEventDetailResponse response = client.getAcsResponse(request);
GetIDEEventDetailResponse.EventDetail.FileExecutionCommand dataSnapshot = response.getEventDe
tail().getFileExecutionCommand();
String checkResult = "OK";
String checkResultTip = "Success";
if (dataSnapshot.getContent().contains("CREATE TABLE")) {
   checkResult = "WARN";
   checkResultTip = "最好不要包含建表语句";
}
if (dataSnapshot.getContent().contains("DROP TABLE")) {
   checkResult = "FAIL";
    checkResultTip = "不能包含删表语句";
}
UpdateIDEEventResultRequest request = new UpdateIDEEventResultRequest();
request.setCheckResult(checkResult);
request.setExtensionCode(extensionCode);
request.setCheckResultTip(checkResultTip);
request.setMessageId(messageId);
client.getAcsResponse(request);
```

5.4.5.4. 应用示例:数据开发过程触发事件检查

已完成扩展程序的开发,并在DataWorks管控台中完成扩展程序的注册和上线后,当您在扩展程序生效的工作 空间中使用某些页面功能(即进行扩展点事件的相关操作)时,DataWorks会将事件消息发送至您的扩展程 序,并根据扩展程序返回的结果来决定该操作是否生效。

前提条件

已完成开放消息的配置、扩展程序的配置。

背景信息

DataWorks支持对扩展点事件进行消息通知,通过扩展程序来自动化校验并响应,其中扩展点事件包括文件发 布前置事件、文件提交前置事件、表提交前置事件等。扩展点事件的概念和详细列表请参见支持的扩展点事件列 表。 其中扩展点事件的自动校验及响应需通过扩展程序实现。DataWorks已为您提供了系统内置扩展程序,包括禁止简单加工、不允许createtable等,您可以根据需要选择是否开启。除此之外,您可以自行开发扩展程序,并将其注册到DataWorks。您可以在工作空间页面查看扩展程序,并控制是否需要在当前工作空间开启该扩展程序校验。

=	🎁 DataWorks	RT123 rt123 2 ~					查看扩展程序来	源:	ntie_doctest@test	
		选择扩展程序	序生效的工作空间		• 来源为"系统"	的是内置扩展程序	i i i			
Ø Ins		注册扩展程序					• 米源为"目定"	("的是注册的扩展	程序	
基 成员管		名称	标识符		适用环境	**	联系人	检查类型	修改时间	启用
(2) (2) (2) (2) (2) (2) (2) (2) (2) (2)		禁止简单加工	System_SimpleProcessingChecker		数据开发 3	系统	SYSTEM	文件发布前置事件	2021-10-26 10:17:27	
MaxC		不允许create table	System CreateTableChecker		教授开发	医位	SYSTEM	文件发布前置事件	2021-10-26 10:17:27	•
(a) (b) (b) (b) (b) (b) (b) (b) (b) (b) (b		禁止使用select*	检查器详情 Sys				× stem	文件发布前置事件	2021-10-26 10:17:26	•
→ #23 (① ITEL	istere Istalia	SQL长度不超过阈值	•扩展程序名称: Sys	分区表查询必须带分区			STEM	文件发布前置事件	2021-10-26 10:17:27	•
		禁止使用max_pt函数的校验	* 1) 闽田Topic: * 处理的扩展点/事件:	文件发布前置事件			STEM	文件发布前置事件	2021-10-26 10:17:27	•
		禁止使用Insert Into+重改厚性	 适用模块: Sys 扩展程序类型: 	 数据开发 (1) 流程控制 (1) 信息提示 			✓ STEM	文件发布前置事件	2021-10-26 10:17:27	•
		大表不能带distinct	* 适用工作空间范围: Svs	请选择			✓ STEM	文件发布前置事件		
			 扩展程序生效方式: 	 透用工作空间管理员自主选择 	道用工作空间强制生女	ġ.				
		4 分区表查询必须带分区	Sys 方度不: 扩展程序文档地址:	https://dgc-cn-shanghai.data.aliyur	n.com/#/knowledge-base/ch	ecker-eve	STEM	文件发布前置事件	2021-10-26 10:17:26	•
		点击查看扩 展程序详情	描述:	分区表直询必须带分区检查器注册						
							_			

使用限制

系统内置扩展程序需要开启数据治理中心后,才可以在工作空间配置页面展示。数据治理中心模块开启与模块 相关说明可参考文档:数据治理中心概述。

扩展程序启用后,实际在数据开发过程中,扩展事件被触发后的消息通知与响应有以下限制。

• 系统内置扩展程序启用后,暂时仅支持对ODPS SQL类的节点生效。

即启用后,仅ODPS SQL类节点中相应扩展点事件被触发时,流程会被阻断,通过内置扩展程序对对应的逻辑进行校验并返回校验结果。非ODPS SQL节点对应的事件被触发后,流程会被阻断,但是内置扩展程序不会校验逻辑是不是符合预期,直接返回校验成功。

- 扩展点事件中,表发布前置事件和表提交前置事件暂时仅支持MaxCompute表提交与发布时触发并进行检测。
- 有内部节点的组合类节点(机器学习(PAI)节点、do-while节点、for-each节点)触发检查时,需要内部 节点都检查通过才可继续进行下一步操作。

能力概述

数据开发模块支持您本地服务对以下扩展点事件进行消息接收,此外,还支持将本地程序注册为DataWorks扩展程序,通过扩展程序接收扩展点事件消息,实现扩展程序对扩展点事件的自定义逻辑处理,并通过回 调UpdateIDEEventResultAP将处理结果返回至平台,实现DataWorks上的流程管控。支持的扩展点如下:

- 文件操作:运行文件、提交文件、发布文件、删除文件
- 表操作:提交表、发布表

支持的扩展点

DataWorks目前在数据开发过程中,支持对以下的扩展点事件进行消息通知,通过扩展程序实现校验并响应。

• DataStudio-运行文件

Ø	操作检查			c		行前置事件检查 ×						
0	类型: 全部				•	ন 🖪 🗠	§ 🕑 🖻	0.C 🗹		s &		
	状态: 全部									5, 操作检查中 >		
	名称/操作时间					odps sql						
4	触发代码运行前置事件检查 2021-10-20 16:07:57	1 运行		<u>校前</u> 中 > ◀		-author:datawor						×
	起之中1023 1037-37 触发文件发布前置事件检查		5-17-84			Create Lime:20	21-10-20	检查通过后均白油			憲法里	Â
Π	2021-10-20 15:58:07	the second se	EX14					12EDBUD647EP	ABM JERT FA		1899,84	
f×	10221-10-2015:55:39		r)	检查中 >								
			¥									
最作检查	i <mark>(</mark> 3)				×							
名称: 102	於代码运行偏置申件检查						6	触发扩展和	程序检查			
100.25	e Genoed kei	852	5 8 00	en e				NH 5 48 /L-1	いホテナデルコ	5.0.3.		
运行前	检查测试程序	121	100 T	ANTIN STICK			9	世八保作村	亚直贝宣有检查	主心來		
							(操作触发:	工作空间目前所 在此查看当前打	所开启的代码运 量作触发的扩展	公前置事件类型的指示。	扩展程序 验详特
				1				1又112,1111	다니트 셈 ㅋ 베 1	*TF/04/04/01/07/08	(1主/] 12 12 71 42 24 13	世旧
<u>ع احر</u>	studio_坦;	ホサ州										
atc		X X IT										
77	操作检查				C 50 钟发文	供提交前置事件检	查 x					
())	米田・ 수部											
9						া বি	(\$)		」C' ⊑	, e d	X	
*	状态:全部								是交文件,操作			
0	名称/操作时间	ġ	类型	状态		odps sql		*****				
4	触发文件提交前置事	¥件检	日大竹体		3		***					
	2021-10-22 11:57:00) #	ECAH	NEETH >	4		12日中			2		×
亩	触发代码运行前置事				6	sele	检查通过后将自	自动执行操作,	请在「操作检	查」中查看结身	果。	
fx	查 2021-10-20 17:17:12	2	运行									
	ve table prod											
<u>=0</u>	2021-10-20 16:36:00	ז ל	度提交到开发									
⊞	扩展程序_pyodps_f	调运										
Ŷ	行前置事件 2021-10-20 16:25:34	1 1	运行	检查中 >								
			÷.									
操作检查	I 🚯					×	6 57	- 提交触发扩展彩	呈序检查 (将會	中发工作空间日言	前所开启的所有文件提	·
an	10 V 20 10 V 20 W 20 10 W						一件的	」扩展程序校验)				A 11 4 7
白标: NC	次又升提又明直事件指查 型		火态				 चb 	(讲入创建发布6]界面杳若文件:	检查状态		
提交前	检查·测试程序			查看详细 查看文档			- 10					
							可点 验询	(击状态查看校验 ≰情	à详情,可在此3	查看当前操作触	发的扩展程序校验列	表及校
					重新检查关键	8						
- + د	Studio_告	ᆂᅕᄴ	-									
Jaco		шХп										
Sa 19 2	发文件发布前置事件检查	Ξ×										\odot
						AA 0						
		· (\$)	•			ኤ					世 发布 7	7 运维
	1odps sql											
		ataworks_										
		ime:2021-										
	6 select '1'											
_		. 春花园	12.00									
=	DataWorks 任务发行	f5 YYW_2	6 74 1	Ē.					8	DataStudio 🕜 运维中	⇔ů ⊈ 17 €, ♥da	lataworks_3h1
	= 创建	发布包										待发布列表
SA BIRK	发布包 解决方	家: 3	新选择		程: 请选择			全部			请输入节点ID/名称	
0= 99269	包列表 交更类	型: N	寿选择		(ii)<=: YYYY-M	IM-DD	田 投索					
			名称	节点版本号	提交人	对象类型	交更类型	节点状态	提交时间	开发环境测试	握作	



• DataStudio-删除文件

Dat aWorks



• DataStudio-提交表



• DataStudio-发布表

S	操作检查			C III xc.table.prod x	
©	类型: 全部				
	状态: 全部				
	名称/操作时间			*提父到土产,操作检查中> *提父到开及,操作体	
α ≞	xc_table_prod 2021-10-20 16:13:39	表提交到生产	检查中 > ◀		
亩	xc_table_prod 2021-10-20 16:12:19	表提交到开发		MaxCompute引擎实例 检查通过后指自动执行操作,请在【操作检查】中查看结果。	
f×	102代码运行前置即件检查 2021-10-20 16:07:57	运行		2	
₽	触发文件发布前置事件检查 2021-10-20 15:58:07	提交文件			
•	触发文件发布前置事件检查 2021-10-20 15:55:39			中文名 M202支25余納西1941社会 版主题 M2045 ◇ 二版主题 M2045 ◇ 1981主要 C	
ଜ	触发文件发布前置事件检查 2021-10-20 15:54:05	提交文件	检查中 >		
Ø	<u> </u>				
操作	检查 3	Ť		× 1 触发扩展程序检查	
名称	: xc_table_prod				
ŧ				2 进入操作检查页查看检查记录	
	没有崩检查		查看详细 查查文档		
				Entrone XM 3 操作触发工作空间目前所开启的表发布前置事件类型的扩展程序构可在此查看当前操作触发的扩展程序校验列表及校验详情	2验.

5.4.5.5. 应用示例:运维中心触发事件检查

完成扩展程序的开发,并在DataWorks管控台中完成扩展程序的注册和上线后,当在扩展程序生效的工作空间 中使用页面部分功能(操作扩展点)时,DataWorks会将事件消息发送至您的扩展程序,并根据返回结果来决 定该操作是否生效。本文为您介绍运维中心操作扩展点触发扩展程序校验的页面过程。

前提条件

已完成开放消息的配置、扩展程序的配置。

背景信息

DataWorks支持您对扩展点事件进行事件消息通知,此外,还支持您基于DataWorks支持的扩展点,通过开启 对应扩展点事件的扩展程序来实现流程卡点,并通过实现扩展程序对扩展点事件的自定义逻辑处理。若要实现 上述能力,您需要先在DataWorks管理控制台注册扩展程序,并在工作空间配置处启用该扩展程序。

≡	🕅 DataWorks	× ×								
	=									*
\$	工作空间配置	注册扩展程序								
*	成员管理	名称	标识符	扩展点	来源	负责人	状态	注册时间	启用	操作
9	权限列表		8	补数据前置事件,节点下线前置事件,节点 解冻前置事件,节点冻结前置事件	自定 义	dataworks_de mo2	已启 用	2022-04-19 20:35:50		设置
*	MaxCompute高级配置		٤	文件提交前置事件,文件发布前置事件	自定 义	xiangcui	未启 用	2022-03-08 15:40:27		设置
•	数据源管理 开源集群管理	100	8	文件提交前置事件,文件发布前置事件	自定 义	dataworks_de mo2	未启 用	2022-03-07 14:55:44		设置
€	CDH集群配置		٤	文件提交前置事件,文件发布前置事件	系统	SYSTEM	未启 用	2022-02-21 19:16:26		设置
€	EMR集群配置		٤	文件提交前置事件,文件发布前置事件	系统	SYSTEM	未启 用	2022-02-16 16:03:12		设置
٢	扩展程序配置			文件提交前置事件	自定 义	dataworks_de mo2	未启 用	2022-02-14 16:15:00		设置
			¢	文件提交前置事件,文件发布前置事件	系统	SYSTEM	未启 用	2022-02-10 16:54:46		设置
			s	文件提交前置事件,文件发布前置事件	系统	SYSTEM	未启 用	2021-09-18 14:16:38		设置
			5	文件提交前置事件,文件发布前置事件	系统	SYSTEM	未启 用	2021-09-18 14:15:48		设 加
			Custom Colort Asterials Chaples	· ·	To lot	OVOTEM	未启	0001 00 10 1415 00		V0.99

? 说明

- 仅有空间管理员权限的用户可在工作空间配置启用扩展程序。
- 扩展程序相关说明,详情请参见:扩展程序概述。
- 用户注册的扩展程序来源为自定义,上图中来源为系统的扩展程序,是DataWorks数据治理中心模块提供的检查项,检查逻辑由平台定义,且目前仅数据开发模块适用,若您想要进一步了解,详情可参见:数据治理中心概述。

能力概述

运维中心模块支持您本地服务对以下扩展点事件消息接收,此外,还支持将本地程序注册为DataWorks扩展程序,通过扩展程序接收扩展点事件消息,实现扩展程序对扩展点事件的自定义逻辑处理,并通过回调 APIUpdateWorkbenchEventResult将处理结果返回至平台,实现DataWorks上的流程管控。运维中心支持的扩展 点如下:

- 节点冻结
- 节点解冻
- 补数据
 - 对周期任务进行补当前节点的操作。
 - 对周期任务进行补当前节点及下游节点的操作。
 - 通过海量节点模式进行补数据。
 - 通过高级模式进行补数据。
- 节点下线

⑦ 说明 此处节点特指生产运维中心周期任务列表中的任务,即生产运维中心周期任务界面对任务进行上述操作,可触发扩展程序校验,开发环境运维中心对周期任务进行上述操作,不会触发扩展程序校验,实例界面进行实例的冻结与解冻操作,不会触发扩展程序校验。

支持的扩展点

• 运维中心-节点冻结、节点解冻

下图以操作节点冻结扩展点为例进行说明。



• 运维中心-补数据



• 运维中心-下线节点

			_							
	M DataWorks	s 运维	e中心 生产11						& DataStudio 🖺	「 2、 简体 💛 (
¢	运维大屏		搜索: 700003	 O 节点类型: 请洗探T 	·····································	请洗探崇任	→ 運度資源(1):	请洗择 ✔ 日報的节点 □ 4	日條改的节点	暂停 (冻结) 节点 🗌 孤立节点
\$\$\$	实时任务运维	~	2+100 II							
n	周期任务运维	^								○ 刷新 展开搜索
	周期任务		✓ 名称	下线操作检查中 200	修改日期.1	任务类型·		潮度周期 潮度密源组 7		操作
	周期实例	-	-					3		
	計数据实例		test_select	· 70	2022-03-23 10:11:54	ODPS SQL	晶作检查详情	日调度 公共调度资源	(组	DAG版: 2515. I *NS2图 / 更多
	モンロタ市体									
4	19 00 00 00 00	, i					名称: test_select			
	20307-00						操作类型: 下线			
	218024						检查点	状态	操作	
	2010/2016	Ť						提供检查由	本无兴趣	本要分批
V	H 105 52 601							採作检查中	旦目奸谓	巨 相又怕
							扩展程序			
ଜ	操作历史									重新检查 关闭
							扩展点			
							添加报警 添加到基线			
			4				下线节点			• • • • • • • • • • • • • • • • • • •
=			修改责任人修改调度	資源但 修改数据集成资源组	暫停 (冻结) 恢复	(解冻)	更多		< 1/	1 > 每页显示: 40 ~
						œ \#	᠈ᄰᄻᆎᅕᄽᄻᆂᆂ	* - 3 可	在此查看当	前操作触发的扩展程序
1	仕 周期仕务 行下线节占:	が面, 握作)	探1作节点卜线打/ 触发工作空间		二 二、二、二、二、二、二、二、二、二、二、二、二、二、二、二、二、二、二、二	直, 进.	(深作位宣许情贝)	^{宣有} 校	验列表及材	验详情
	启的节点下	线前置	了。 <u>那夜</u> 工日至回 置事件检查的扩展和	程序校验	四三元不					

6.数据建模

6.1. DataWorks智能数据建模

6.1.1. 概述

随着业务的快速发展,企业数据成几何倍增长,数据量庞大、复杂、各类数据间标准不一致,往往会出现数据 难以管理的现象。DataWorks智能数据建模服务,将无序、杂乱、繁琐、庞大且难以管理的数据,进行结构化 有序的管理。使企业中的数据产生更多的价值,将数据价值最大化。

前提条件

您需要开通DataWorks智能数据建模后,才可以使用该产品功能,详情请参见智能数据建模计费标准。

使用限制

DataWorks工作空间中,不同角色使用智能数据建模的限制如下:

- 浏览模型详情:访客、空间管理员、模型设计师、项目所有者等所有DataWorks工作空间中的角色均可浏览数据模型的相关信息。更多DataWorks工作空间角色,详情请参见角色及成员管理:空间级。
- 编辑模型信息: 仅空间管理员、开发、运维、模型设计师角色支持编辑模型信息。如果您需要执行该操作,则可授权目标用户拥有对应的角色权限。授权详情请参见角色及成员管理: 空间级。
- 发布数据模型: 仅空间管理员、运维角色支持发布已创建的数据模型。如果您需要执行该操作,则可授权目标用户拥有对应的角色权限。授权详情请参见角色及成员管理: 空间级。

概述

DataWorks数据建模支持数仓规划设计、制定并沉淀企业数据标准、维度建模、数据指标定义,通过使用 DataWorks数据建模,您可以将建模设计产出的维度表、明细表和汇总表物化到计算引擎中并进一步应用。



● 数仓规划

使用DataWorks进行数据建模时,您可以在数仓规划页面进行数据分层、业务分类、主题域和业务过程设 计。

○ 数据分层

您可以结合业务场景、数据场景综合考虑设计数仓的数据分层,DataWorks为您默认创建业界通用的五层数仓分层:

- 数据引入层 ODS (Operational Data Store)
- 明细数据层 DWD (Data Warehouse Detail)
- 汇总数据层 DWS (Data Warehouse Summary)
- 应用数据层 ADS (Application Data Service)
- 公共维度层 DIM (Dimension)

您也可以根据业务需求创建其他分层数据层,创建数据分层的操作请参见创建数仓分层。

○ 业务分类

当企业业务比较复杂,不同类型业务彼此间需要共享数据域,但是又希望能在模型设计和应用过程中快速定位本业务的数据时,您可结合真实业务情况,规划不同的业务分类,在后续建模的维度表和明细表中,将其关联到对应的业务分类中。创建业务分类的操作可参见业务分类。

○ 数据域

数据域是一个较高层次的数据归类标准,是对企业业务过程进行抽象、提炼、组合的集合,是企业业务 人员在使用数据时第一个分组入口,可以帮助企业业务人员快速的从海量的数据中快速圈定到自己的业 务数据。

数据域面向业务分析,一个数据域对应一个宏观分析领域,比如采购域、供应链域、HR域、电商业务域 等。数据域的设定建议由统一组织或者人员(如数据架构师或者模型小组成员)进行管理和设定,数据 域设计人员需要能对企业有深刻的业务理解,更多的表达对业务的解释和抽象。使用DataWorks进行数据 域的规划建设的操作,请参见数据域。

○ 业务过程

业务过程是对业务活动流程的描述,例如在电商领域,加购、下单、支付等都可以是一个业务过程。进行业务效果分析时,业务过程有非常典型的应用,例如常用漏斗分析,即将购买商品的业务活动分解为 浏览商品、加入购物车、下订单、付款、确认收货等业务过程,统计每个业务过程的"订单数",将可以 针对"订单数"这一指标做漏斗分析。使用DataWorks创建业务过程的操作请参见业务过程。

● 数据标准

DataWorks数据建模支持在建模前规划制定数据标准,或在建模使用过程中根据业务情况沉淀企业业务的数据标准。通过规范约束标准代码、度量单位、数据标准、命名词典,来保障后续建模与应用过程中数据处理的一致性。

例如,现有注册表和登录表两张表,注册表中存储了会员ID,字段名为user_id,登录表中也存储了会员 ID,字段名为userid,此时针对会员ID这个数据创建统一的数据标准,例如指定数据处理的标准代码、指定 字段的属性要求(例如字段的数据类型、长度、默认值等)、指定数据的度量单位。创建好数据标准后, 后续在建模过程中涉及到会员ID这个字段的设置时,即可直接关联此标准,以此来保障所有会员ID字段的标 准统一。

DataWorks中数据标准的创建操作请参见字段标准。

● 维度建模

DataWorks的数据建模理念遵循维度建模思想,使用DataWorks的维度建模功能进行数据仓库建模设计时:

○ 维度表

结合业务的数据域的规划,提取出各业务数据域中进行数据分析时可能存在的维度,并将维度及其属性 通过维度表的方式存储下来。例如,在进行电商业务数据分析时,可用的维度及其属性有:订单维度 (属性包括订单ID、订单创建时间、买家ID、卖家ID等)、用户维度(性别、出生日期等)、商品维度 (包括商品ID、商品名称、商品上架时间)等,此时您就可以将这些维度和属性创建为订单维度表、用户 维度表、商品维度表等,将维度属性记录作为维度表的字段。后续您可将这些维度表部署到数仓中,通 过ETL将实际维度数据按照维度表定义的方式进行存储,方便业务人员在后续的数据分析时进行取用。

○ 明细表

结合业务过程的规划,梳理分析各业务过程中可能产生的实际数据,将这些实际数据字段通过明细表的 方式存储下来。例如下订单这一业务过程中,您可以创建下订单这一明细表,用于记录下单过程可能产 生实际数据字段,例如订单ID、订单创建时间、商品ID、数量、金额等。后续您可将这些明细表部署到数 仓中,通过ETL将真实的数据按照明细表的定义方式进行汇总存储,便于业务分析时取用。

○ 汇总表

您可以结合业务数据分析和数仓分层,将一些明细的事实数据和维度数据先进行汇总分析,创建汇总 表,后续数据分析时直接取用汇总表中的数据即可,无需再取用明细表和维度表中的数据。

○ 逆向建模

逆向建模主要用于将其他建模工具生成的模型反向建模至DataWorks的维度建模中。例如,当您已通过其他建模工具生成模型,此时,想更换为DataWorks的智能建模进行后续建模工作,则可以使用逆向建模功能。该功能无需您再次执行建模操作,即可帮助您快速将已有模型反向建模至DataWorks的维度建模中, 节省了大量的时间成本。

维度表、明细表、汇总表的创建操作请参见创建维度表、创建明细表、创建汇总表。逆向建模操作,详情 请参见<mark>逆向建模</mark>。

● 数据指标

DataWorks的数据建模提供数据指标功能,为您提供统一的指标体系建立能力。

指标体系由原子指标、修饰词、时间周期和派生指标构成。

- **原子指标**:是基于某一业务过程下的度量,如"支付订单"业务过程中的"支付金额"。
- 修饰词:是对指标统计业务范围的限定,如限定"支付金额"的统计范围为"母婴类产品"。
- 时间周期:用于明确指标统计的时间范用或者时间点,如制定统计"支付金额"的时间周期为"最近7 天"。
- 派生指标:由原子指标、修饰词、时间周期组合定义。如,统计"最近7天" "母婴类产品"的"支付金额"。

指标体系的创建操作请参见数据指标概述。

数据建模的必要性

• 海量数据的标准化管理

企业业务越庞大数据结构就越复杂,企业数据量会随着企业业务的快速发展而迅速增长,如何结构化有序 地管理和存储数据是每个企业都将面临的一个挑战。

• 业务数据互联互通,打破信息壁垒

公司内部各业务、各部门之间数据独立自主形成了数据孤岛,导致决策层无法清晰、快速地了解公司各类 数据情况。如何打破部门或业务领域之间的信息孤岛是企业数据管理的严峻考验。

• 数据标准整合,统一灵活对接

同一数据不同描述,企业数据管理难、内容重复、结果不准确。如何制定统一的数据标准又不打破原有的 系统架构,实现灵活对接上下游业务,是标准化管理的核心重点之一。

• 数据价值最大化,企业利润最大化

在最大程度上用好企业各类数据,使企业数据价值最大化,为企业提供更高效的数据服务。

6.1.2. 数仓规划

6.1.2.1. 数仓规划概述

使用DataWorks进行数据建模时,数仓架构师或者模型小组成员可以在数仓规划页面对数据分层、业务分类、数据域、业务过程、数据集市、主题域进行设计。完成设计后,模型设计师在建模过程中可以依赖数仓规划中的数据分层、业务分类、数据域、业务过程等对所建模型进行分层化域管理。

数仓分层

您可以结合业务场景、数据场景综合考虑设计数仓的数据分层,DataWorks为您默认创建业界通用的五层数仓 分层:

- 数据引入层 ODS (Operational Data Store)
- 明细数据层 DWD (Data Warehouse Detail)
- 汇总数据层 DWS (Data Warehouse Summary)
- 应用数据层 ADS (Application Data Service)
- 公共维度层 DIM (Dimension)

您也可以根据业务需求创建其他分层数据层,创建数据分层的操作请参见创建数仓分层。

分层归属

DataWorks为您提供了**贴源层、公共层、应用层**等分层归属,您可以将不同功能作用的分层挂载至相应归属,通过层级化结构呈现输出。下图以默认的五层数据分层为例,数据分层规划完成后,后续的表数据存储可 根据规划分别存储至不同的数据分层中。



- 贴源层:用于数据库、日志、消息等基础源数据的引入。源数据经过一系列ETL操作进入贴源层,该层级只能挂载贴源表。
- 公共层:用于加工、整合贴源层输入的公共数据,建立统一的指标维度,构建可复用面向分析和统计的明 细事实数据和汇总数据。该层级可挂载明细表、维度表、汇总表。
- 应用层:基于实际应用需求,获取公共层加工整合后的数据,面向具体应用场景或指定产品进行的个性化数据统计。该层级可以挂载应用表、维度表。

业务分类

当企业业务比较复杂,不同类型业务彼此间需要共享数据域,但是又希望能在模型设计和应用过程中快速定位 本业务的数据时,您可结合真实业务情况,规划不同的业务分类,在后续建模的维度表和明细表中,将其关联 到对应的业务分类中。创建业务分类的操作可参见业务分类。

数据域与业务过程

数据域与业务过程通常位于公共层,根据业务类别、数据来源、数据用途等多个维度,对公共层数据进行区域 及区域下业务活动进行划分。

● 数据域

数据域是一个较高层次的数据归类标准,是对企业业务过程进行抽象、提炼、组合的集合,是企业业务人员在使用数据时第一个分组入口,可以帮助企业业务人员快速的从海量的数据中快速圈定到自己的业务数据。

详情请参见数据域。

• 业务过程

业务过程是对业务活动流程的描述,例如在电商领域,加购、下单、支付等都可以是一个业务过程。进行 业务效果分析时,业务过程有非常典型的应用,例如常用漏斗分析,即将购买商品的业务活动分解为浏览 商品、加入购物车、下订单、付款、确认收货等业务过程,统计每个业务过程的"订单数",将可以针 对"订单数"这一指标做漏斗分析。使用DataWorks创建业务过程的操作请参见业务过程。

数据集市与主题域

数据集市与主题域通常位于应用层,面向用户的实际业务需求,对指定的某类业务进行场景或产品的细粒度划分,制定不同分析视角,基于公共层的整合数据进行业务个性化的数据统计。

• 数据集市

数据集市是对某个业务分类制定细化的业务主题,并通过主题域,基于不同分析视角对目标集市中的数据 进行主题划分,最终面向业务应用统计分析数据。例如,运营平台数据集市。

详情请参见数据集市。

主题域

主题域用于将数据集市按照分析视角进行划分,通常是联系较为紧密的数据主题的集合。您可以根据业务的关注点,将这些数据主题划分至不同的主题域。例如,电商行业通常分为交易域、会员域、商品域等。 详情请参见主题域。

维度管理

维度管理用于标准化定义维度,帮助您对数据建模过程中产生的维度进行统一建设和管理,保证维度唯一,避 免重复建设。详情请参见<mark>维度管理</mark>。

建模空间

当您所需要管理多个DataWorks工作空间且需要复用一套数仓规划时,面对跨多个工作空间的复杂数据体系,可以通过设计空间来共享一套数据建模工具,针对整个数据体系进行统一地数仓规划、维度建模及指标定义等工作。

DataWorks智能数据建模在进行跨多个工作空间数据的统一建模情况下,为您提供了**模型设计工作空间**与数据研发工作空间的概念,通过建模空间功能实现模型跨工作空间应用。



建模空间的详细介绍请参见建模空间。

6.1.2.2. 业务分类

如您所在的集团或公司业务复杂且较为庞大时,可创建业务分类,对数据进行业务区分,便于对业务数据进行 更加方便的管理。本文为您介绍如何创建业务分类并关联数据域及管理数据集市。

新建业务分类

- 1. 进入数据建模。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 选择工作空间所在地域后,单击相应工作空间后的数据开发。
 - iv. 单击左上方的**三**图标,选择**全部产品 > 数据建模 > 数仓规划**,默认进入数仓分层页面。
- 2. 在数仓规划页面的左侧导航栏,单击业务分类,进入业务分类页面。

3. 创建业务分类。

i. 在业务分类页面, 鼠标悬停至 ⊞图标, 单击新建一级业务分类。

ii. 配置新建一级业务分类对话框的各项参数。

新建—级业务分类	×
英文缩写: * 业务名称: * 负责人: * 备注:	请输入英文缩写 请输入中文名称
	0/2048 商认 取消
参数	描述
英文缩写	业务分类的英文缩写。 英文缩写由小写英文字母、数字、下划线(_)组成,并且以小写英文字母开头。最大 长度不可超过128个字符。
业务名称	业务分类的中文名称。 中文名称由中文、英文字母、数字、下划线(_)、and(&)、英文括号组成,并且以 中文、英文字母或数字开头。最大长度不可超过128个字符。
负责人	创建业务分类的责任人。默认为当前登录工作空间的用户。
备注	业务分类的相关描述,便于后期查看业务内容。

4. 单击确认,完成创建业务分类。

如果已创建的业务分类已变更或已不需要,您可在左侧目录树右击业务名称,选择**编辑**或删除进行操 作。

⑦ 说明 删除业务分类前,您需要先删除业务分类的所有绑定关系。

设置关联数据域

创建完成业务分类后,您可以在该业务分类页面的**关联数据域**区域,设置当前业务分类需要关联哪些数据 域。设置完成后,后续该业务分类下的数据建模即可使用此处关联的所有数据域。

业务分类	Ð C	器 销售业务							
请输入业务分类	Q								
∨ ☷ 天猫精灵		备注:	-						
✓ B≣ sales									
> 2011 销售业务		客服 (c)	rm)						
		□ 财务(f)	财务 (financial) 供应维 (supply)		创建时间	备注	操作		
			ales)		2021-06-28 10:46:56		编辑 删除		
		□ 市场 (n	nk)		2021-06-28 10:46:48		编辑 删除		
			드 (ui) ser)						
		🗌 设备 (d	evice)						
		□ 谭程管 确定	理 (cm)						
		请选择		~ -	一键加载全部数据域				
		已选数据域:							
		中文名	英文名	英文缩写	备注		操作		
		默认	default data domain	default	默认数据域。如果业务过程没有指定所属数	据域, 会被默认放到该数据域中。	删除		

管理数据集市

创建完成业务分类后,您可以在该业务分类页面的**数据集市管理**区域,查看当前业务分类挂载的数据集市列 表,并根据业务需求对目标数据集市进行编辑或删除等管理操作。

(?) 说明	删除 指定数据集市,	不仅会删除该数据集市与业务分类的绑定关系,	同时,	会将该数据集市
一并删除,	请您谨慎操作。			

业务分类	∄ C	🗄 销售业	务					
请输入业务分类	Q		~~~~	,	LUKER J PJ		10701	
> 믎 天猫精灵		产品销售	chanpin		2021-06-28 10:46:56		编辑	删除
> 晋 销售业务		课程销售	kecheng	ţ	2021-06-28 10:46:48		编辑	删除
		关联数据 可选数据线: 请选择 已选数据域: 中文名	英 文名	×	加數全部数据域 备注 			操作
		数据集市管 集市名称 金融业务 支付业务	管理 英文编写 finance pay		 (1)(3)(3)(3)(3)(3)(3)(3)(3)(3)(3)(3)(3)(3)	HTURNUULSINGHUM-U-T。 留注	操作编辑编辑	新 新 新 家

新建子业务分类

在一级业务分类下,如有其他子业务分类,可继续新建子业务分类。创建方法与一级业务分类相同。

业务分类 请输入业务分类	⊕ C Q	🏭 销售业务					
 > Ⅲ 天猫楠灵 > Ⅲ sales > Ⅲ 销售业务 	新建子业务分类 编辑 删除	基本信息 英文缩写: 测试-级业 创建人: 修改人: 备注:	务分类	业综名称: 创建时间: 修改时间:	销售业务 2021-06-28 10:46:25 2021-06-28 10:46:25		编辑
		新建子业务分类 业务名称 产品销售 课程销售	英文编写 chanpin kecheng	创建时间 2021-06-28 10:46:56 2021-06-28 10:46:48		备注	操作 編輯 删除 編輯 删除

后续步骤

业务分类创建完成后,您可在创建维度表、明细表、汇总表、应用表时关联对应的业务分类,也可在**维度建** 模页面,通过业务数据视角查看表分类详情。关联表详情请参见创建维度表、创建明细表、创建汇总表、创建应用 表。

6.1.2.3. 数仓分层

6.1.2.3.1. 创建数仓分层

数仓分层是结合对业务场景、实际数据、使用系统的综合分析,对数据模型进行的整体架构设计及物理层的详 细划分。用于将不同用途的数据,归类划分至不同的分层,便于您更好地组织、管理、维护数据。本文为您介 绍如何创建并管理数仓分层。

背景信息

数据仓库是所有数据的集合,包括日志信息、数据库数据、文本数据、外部数据等都集成在数据仓库中。数仓 分层、数据域、业务过程、数据集市、主题域等要素,共同确定了您此次建模的逻辑数仓架构。其中,数据域 及业务过程位于公共层,用于公共层数据模型的构建。数据集市及主题域位于应用层,用于面向具体业务应用 的模型构建。

数仓分层保障了数据在进入数据仓库之前都经过清洗和过滤,使原始数据不再杂乱无章,优化了查询过程,有 效的提高了数据获取、统计和分析的效率。同时,数仓分层实现了各种不同维度数据的关联,使多维分析更加 方便,为从多角度、多层次地数据分析和决策制定提供便捷。

规划数仓分层

数仓分层需要结合业务场景、数据场景、系统场景进行综合考虑设计,您可以根据实际业务需求,规划您数据 模型的分层。

DataWorks的数仓分层功能,默认为您创建了数据引入层ODS(Operational Data Store)、公共维度层 DIM(Dimension)、明细数据层DWD(Data Warehouse Detail)、汇总数据层DWS(Data Warehouse Summary)及应用数据层ADS(Application Data Service)等五层,各个分层的功能介绍 如下:

• 数据引入层 ODS (Operational Data Store)

ODS层用于接收并处理需要存储至数据仓库系统的原始数据,其数据表的结构与原始数据所在的数据系统中的表结构一致,是数据仓库的数据准备区。ODS层对原始数据的操作具体如下:

• 将原始的结构化数据增量或全量同步至数据仓库中。

- 将原始的非结构化数据(例如,日志信息)进行结构化处理,并存储至MaxCompute。
- 根据实际业务需求,记录原始数据的历史变化或对原始数据进行简单的清洗。

ODS层的数据表,命名必须以 ods 开头,并且生命周期为366天。

• 明细数据层 DWD (Data Warehouse Detail)

DWD层通过企业的业务活动事件构建数据模型。基于具体业务事件的特点,构建最细粒度的明细数据表。 您可以结合企业的数据使用特点,将明细数据表的某些重要维度属性字段适当冗余,即宽表化处理。同时,也可以减少明细数据表及维度表的关联,提高明细表的易用性。

• 汇总数据层 DWS (Data Warehouse Summary)

DWS层通过分析的主题对象构建数据模型。基于上层的应用和产品的指标需求,构建公共粒度的汇总指标 事实表。

例如,从ODS层中对用户的行为做一个初步的归类汇总,抽象出来一些通用的维度,假设维度为*时间、IP、ID*,并根据这些维度统计出相关数据,比如用户每个时间段在不同登录IP购买的商品数。则在DWS层可以进一步添加一层轻度的汇总,可以让计算更加的高效。例如在此基础上计算仅7天、30天、90天的行为会节省很多时间。

• 应用数据层 ADS (Application Data Service)

ADS层用于存放数据产品个性化的统计指标数据,输出各种报表。例如,某电商企业,在6月9日至6月19 日,杭州地域出售的各大球类商品的数量及排行情况。

• 公共维度层 DIM (Dimension)

DIM层使用维度构建数据模型。基于实际业务,通过定义维度,确定维度主键,添加维度属性,关联不同维度等操作,构建整个企业的一致性数据分析维表,帮助您降低数据计算口径和算法不统一的风险。



对于创建的分层,您可以选择平铺展示或层级结构展示,如下图所示。

展示方式	
平铺展示 不区分层级结构,所有分层统一平铺放置。	
DataWorks为您提供了贴源层、公共层、应用层、其他等分层归属,您可以将创建的数据层挂载至不同的分层归属。 ● 贴源层:主要用于数据库、日志、消息等基础数据的引入。该层级只能挂载贴源表。 ● 公共层:主要完成公共数据的加工与整合,建立统一的维度,构建可复用面向分析和统证明细事实表,以及汇总公共粒度的指标。该层级可挂载明细表、维度表、汇总表。 ● 应用层:基于实际应用需求,获取公共层加工整合后的数据,进行持续的个性化数据建改造。该层级可以挂载应用表及维度表。 ● 使用层:基于实际应用需求,获取公共层加工整合后的数据,进行持续的个性化数据建立改造。该层级可以挂载应用表及维度表。 ● 其他:用于存放存量创建的系统默认层级及用户自定义的层级。 ● 系统默认层级:您需要提交技术支持人员,帮助您修改该类层级的分层归属。 ● 用户自定义的层级:您需要通过编辑操作,修改层级的分层归属。 ● 用户自定义的层级:您需要通过编辑操作,修改层级的分层归属。	分十的 安及

⑦ 说明 分层归属一旦设置(包括新建时配置及编辑修改配置)则无法变更,请您合理规划层级归属。

创建数仓分层

系统已默认为您创建了五个分层(ODS层、DIM层、DWD层、DWS层、ADS层)可以满足绝大部分的场景和需求,如您存在部分需要个性化处理的场景,可参见如下步骤创建新的数仓分层。

个性化处理的场景示例:抽象出TMP(临时表)层,在创建表的时候严格的将表归属到对应的分层中,对每个 分层设定一些标准和校验规则,譬如命名规范、生命周期设置等。

1. 进入数据建模。

- i. 登录DataWorks控制台。
- ii. 在左侧导航栏, 单击工作空间列表。
- iii. 选择工作空间所在地域后,单击相应工作空间后的数据开发。

Ⅳ. 单击左上方的国图标,选择全部产品 > 数据建模 > 数仓规划,默认进入数仓分层页面。

2. 创建分层。

i. 单击新建分层, 在新建数仓分层对话框配置分层的基本信息。

新建数仓分层		×
英文缩写: *	请输入英文缩写	
英文名: *	请输入英文名称	
中文名: *	请输入中文名称	
负责人: * 分层归属: *		
模型类型: *	· · · · · · · · · · · · · · · · · · ·	
备注:	0/2058	
	取消 确	і ,

参数	描述
英文缩写	数仓分层的英文缩写,唯一标识数仓分层。 英文缩写由英文字母、数字、下划线(_)组成,并且以英文字母开头,最 大长度为128字符。
英文名	数仓分层的英文名称。 英文名称由英文字母、数字、下划线(_)、and(&)组成,并且以英文字 母或数字开头,最大长度为2048字符。
中文名	数仓分层的中文名称。 中文名称由中文、英文字母、数字、下划线(_)、and(&)、英文括号组 成,并且由中文、英文字母或数字开头,最大长度为2048字符。
负责人	创建数仓分层的负责人。默认为当前登录账号。

参数	描述	
分层归属	 「協社 「該参数与模型类型配合使用。用于将创建的分层划分至不同的层级,并关联对应的模型类型。 「點源层:主要用于数据库、日志、消息等基础数据的引入。该层级只能挂载贴源表。 公共层:主要完成公共数据的加工与整合,建立统一的维度,构建可复用面向分析和统计的明细事实表,以及汇总公共粒度的指标。该层级可挂载明细表、维度表、汇总表。 应用层:基于实际应用需求,获取公共层加工整合后的数据,进行持续的个性化数据建设及改造。该层级可以挂载应用表及维度表。 其他:用于存放存量创建的系统默认层级及用户自定义的层级。 系统默认层级:您需要提交技术支持人员,帮助您修改该类层级的分层归属。 用户自定义的层级:您需要通过编辑操作,修改层级的分层归属。 ① 说明 如果位于其他层级的数据分层未修改层级归属,则后续创建模型表时无法挂载相应表至该层级。 	
模型类型	该参数与 分层归属 配合使用。用于将对应的模型类型挂载至不同的分层归 属。 ■ 贴源表:只能挂载至 贴源层 。 ■ 明细表:只能挂载至公共层。 ■ 维度表:可以挂载至公共层及应用层。 ■ 汇总表:只能挂载至公共层。 ■ 应用表:只能挂载至应用层。	
备注	数仓分层的备注信息。您可以输入目标数仓分层的功能描述,方便您快速了 解每个分层的功能作用,将对应业务数据存放至合适的分层。 最大长度为2058字符。	

3. 单击确认,完成创建。

后续步骤

数仓分层创建完成后,您需要创建数仓分层检查器,用于规范数仓分层中表的命名规范,详情请参见<mark>配置数仓</mark> 分层检查器。

6.1.2.3.2. 配置数仓分层检查器

您可通过数仓分层页面,配置数仓分层检查器以规范分层中表名的命名要求,便于后期管理。本文将为您介绍 如何通过数据建模控制台查看数仓分层详情,以及创建数仓分层检查器的操作方法。

前提条件

已创建待配置的数仓分层,数仓分层创建方法请参见创建数仓分层。

背景信息

数仓分层检查器用于规范目标分层中表的命名,将同一分层中表名称的命名格式统一,便于您通过表名称,即 可了解到该表所属的业务类型、作用功能、数据粒度等信息。同时,可以帮助您减少后期的运维成本。

系统默认创建的数仓分层和您自定义新建的数仓分层均可以配置数仓分层检查器。

⑦ 说明 您可重复执行如下步骤为不同的数仓分层配置检查器。

- 1. 在数仓分层页面,单击已创建的分层,进入该数仓分层的详情页。
- 2. 配置数仓分层的检查器。

在数仓分层详情页,通过如下步骤配置检查器。

? 说明

数仓分层检查器用于规范目标分层中表的命名,将同一分层中表名称的命名格式统一,便于您通过表 名称,即可了解到该表所属的业务类型、作用功能、数据粒度等信息。同时,可以帮助您减少后期的 运维成本。

i. 在检查器管理区域,单击添加。

ii. 配置检查器的参数信息。

检查器管理 80 数2/28 五型结果				
新婚检查器 秋以、① 星岳开白 ① 星岳突的性音 ① 日本 日				
英文擁有: * 语输入检查器规文缩写	名称: * 通知人检查器名称 英型: * 第名检查器 V			
親則定义: * 请输入表前缀 <u>业务分类英文增</u> 写 ∨ ④ ○ 描述:	07048			
参数	描述			
英文缩写	检查器的英文缩写。 英文缩写由小写英文字母、数字、下划线(_)组成,并且以小写英文字母 开头,最大长度为128字符。			
名称	检查器的名称。 检查器的名称由中文、英文字母、数字、下划线(_)、and(&)、英文括 号组成,并且由中文、英文字母或数字开头,最大长度为2048字符。			
类型	目前仅支持配置为 表名检查器 ,用于检查挂载至目标分层中的表名称是否 符合该数仓分层的命名规范。			
规则定义	用于检查表名称是否符合数仓分层命名规范的规则。您需要输入表的前缀信息(例如,目标数仓分层要求表名称以 dim 开头),通过表前缀进行检测。同时,也可以单击④图标,使用业务分类、数据域、业务过程等多种类别的英文缩写来定义规则。			
默认	表示是否将该检查器配置为默认检查器。			
是否开启	表示配置的检查器是否生效。配置的检查器开关默认为打开状态,即默认检 查器生效。			
是否实时检查	 是否开启实时检测,默认关闭。 实时检测开启时:在创建或修改表时,会对表名称自动实时检测,确保表名称符合命名规范。 实时检测关闭时:在创建或修改表时,不会对表名称实时检测,您需要在创建或修改完成后,手动点击分层详情页中的触发检测,对表名称进行检测。 			
描述	检查器的描述信息。			

iii. 单击**保存**,保存后,检查器的配置才会生效。

- iv. 启动检查器并查看检测结果。
 - a. 检查器配置完成后,您可以单击**触发检测**,则系统默认启动该数仓分层中的所有检查器,根据 检查器的配置内容,检测分层中所有表的命名规范。
 - b. 检查结束后, 您可以单击查看结果, 查看本次检测结果。
 - 示例一:某公司希望目标数仓分层中的表以 dim_ 开头,则检测器中配置的规则定义为 ^dim_.
 * 。当数仓分层中包含命名为 dim_sku、 dim_store、 dimension_warehouse、 fact_order的表时,触发检测后,检测结果如下表所示。

表名称	是否符合数仓分尾的命名抑范
dim_sku	匹配
dim_store	匹配
dimension_warehouse	不匹配
fact_order	不匹配

示例二:某公司希望DWD数仓分层中的表使用数仓分层的英文缩写_表所在的数据域_自定义表命名标签缩写_df/di 格式命名。当该数仓分层中包含命名为*dwd_sale_order_df、dwd_sale_order_di、dwd_sale_order_df*、*dwd_sale_order_df*、*dwd_sale_order_df*、*dwd_sale_order_df*、*dwd_sale_order_df*

表名称	是否符合数仓分层的命名规范		
dwd_sale_order_df	匹配		
dwd_sale_order_di	匹配		
dwd_sale_order	不匹配		
dws_sale_order_df	不匹配		

后续步骤

数仓分层创建并配置完成后,您可继续进行数仓规划中的数据域、业务过程与业务分类的规划和创建。详情请参见数据域、业务过程和业务分类。

6.1.2.3.3. 查看并管理数据分层

本文将为您介绍如何查看并管理已创建成功的数据分层及相关信息。

您可以在数据分层页面,对已创建的数据分层进行如下管理操作:

≡	🍿 DataWorks 💈	合规划	∂ 数仓规划	❷ 数据标准	❷ 维度建模	∂ 数据指标	👌 থ	V
8≣ : ♦	业务分类 数仓分层			Dimension	八十進帝日			
8	公共层 🗸	叙信が海倫被告哲型努物域、叙維物域、系統物域提出行時指考測的行行、一般分以ト江通: UDS: Operational Usta Store, 知識是, UNN: Dimension, 公共推測量, DWD: Data Warehouse Detail, 胡嘎数据层, DWS: Data Warehouse Summary, 汇总数据层, ADS: Application Data Service, 应用数据层,						
		新建分局 						
÷	建模空间	始勝県 公共宏 应用是 其他						
		金用数据层 ADS 由于特致据层 由了考致就度产 最合性化的液 最后更新2021 变素人: dataw ⑦ 正总数据层 DWS の 間朝数据层 DWS の 間期数据层 DWS の 間期数据层 DWS の 間期数据层 通信更新2021 ①素人: dataw の 世後本にはaw の 世後太 の 世後太 ・ ・	维度层 DIM 基于维度建模 念思想,建立。 最后更新2021 负责人: data	₽ 	 数据5 ODS 存放朱 的原始 最后更 负责人 区 修改 	は 大 展 数 振 至 … 新 2021 : dataw… 面 删除		

• 查看数据分层概览。

您可以在**数据分层**页面,查看所有已创建数据分层的基本信息,包括数据分层的层级归属、分层名称、功 能描述、最新更新时间及对应负责人等。快速了解数据模型的分层规划,后续将对应数据表挂载至合理的 分层中。

• 查看数据分层详情。

单击对应的数据分层进入该分层的详情页。在基本信息区域,即可查看目标数据分层的英文缩写、英文 名、中文名、负责人、分层归属、模型类型创建人、创建时间,以及最近一次修改该数据分层的修改 人及修改时间等信息。您也可以根据业务需求,单击右上角的编辑,修改目标分层的相关信息。

? 说明

- 系统创建的数据分层创建人默认为system, 负责人为当前工作空间的责任人。
- DataWorks支持修改或删除系统默认创建的数据分层。

基本信息	编辑
英文缩写: ADS	英文名: application data store
中文名: 应用数据层	负责人:
分层归属: 其他	模型类型:
创建人: system	创建时间: 2021-05-31 18:02:19
修改人:	修改时间: 2021-07-22 22:34:16
备注: 用于存放数据产品个性化的统计指标数据。	

• 修改数据分层信息。

单击目标分层的 🛙 📾 图标, 即可修改该数据分层的基本信息。

• 删除数据分层。

单击目标分层的 💩 🛲 图标, 即可删除该数据分层。

⑦ 说明 如果数据分层中已经挂载了数据表,则您需要将分层中的所有表删除后,才能删除该数据 分层。

6.1.2.4. 公共层

6.1.2.4.1. 数据域

数据域是联系较为紧密的数据主题的集合,通常是根据业务类别、数据来源、数据用途等多个维度,对企业的 业务数据进行的区域划分,将同类型数据存放在一起,便于您快速查找需要的内容。不同使用目的数据,分类 标准不同。例如,电商行业通常分为交易域、会员域、商品域等。

背景信息

数据域是一个较高层次的数据归类标准,是对企业业务过程进行抽象、提炼、组合的集合,是企业业务人员在 使用数据时第一个分组入口,可以帮助企业业务人员快速的从海量的数据中快速圈定到自己的业务数据。



由上图所示,数据域规划创建好之后:

- 您可以规划当前数据域下业务的业务过程,后续创建好业务过程后挂载至数据域中,便于后续全数据域视 图下查看所有涉及的业务过程。
- 在后续的数据指标及维度建模的过程中,可以关联对应指标及维度表到数据域中,便于后续全数据域的视 图下查看指标及维度表。

本文为您介绍如何创建数据域,并为您介绍创建数据域后,规划创建业务过程的操作入口,详情请参见<mark>创建数</mark> 据域、查看数据域详情并添加业务过程。

创建数据域

DataWorks的数据建模默认为您创建了名为默认的数据域,如果业务过程没有指定所属数据域,则创建的**业务过程**会被默认放至该数据域中。您可以根据需要建模的业务类型,创建符合企业实际应用的数据域。创建数据域操作如下。

- 1. 进入数据建模。
 - i. 登录DataWorks控制台。
- ii. 在左侧导航栏, 单击工作空间列表。
- iii. 选择工作空间所在地域后,单击相应工作空间后的数据开发。
- iv. 单击左上方的**三**图标,选择**全部产品 > 数据建模 > 数仓规划**,默认进入数仓分层页面。
- 2. 在数仓规划页面的左侧导航栏,单击公共层>数据域,进入数据域页面。
- 3. 创建数据域。
 - i. 在数据域页面, 单击+新建域。
 - ii. 在新建数据域对话框, 配置各项参数。

新建数据域	或	×
英文缩写 * 英文名 * 中文名 * 负责人 * 备注	清选择 清縮入 「清縮入 「清縮入 「清縮入	
参数	描述	
英文缩 写	数据域的英文缩写。用于数据分层检查器配置 规则定义 时使用,检查目标数据数据分层的命名规范,详情请参见配置数仓分层检查器。 英文缩写由小写英文字母、数字、下划线(_)组成,并且以小写英文字母开等 字符。	居表是否符合所挂载 头,最大长度为128
英文名	数据域的英文名称。 英文名称由英文字母、数字、下划线(_)、and(&)组成,并且以英文字母 长度为2048字符。	}或数字开头,最大
中文名	数据域的中文名称。 中文名称由中文、英文字母、数字、下划线(_)、and(&)、英文括号组成 文字母或数字开头 <i>,</i> 最大长度为2048字符。	<i>₹,</i> 并且由中文、英
负责人	创建数据域的负责人。系统创建的数据域负责人是system。	
备注	数据域的备注信息。 最大长度为2048字符。	

4. 单击确认,完成创建。

查看数据域详情并添加业务过程

成功创建数据域后,您可以查看目标数据域的详细信息,并基于该数据域创建需要分析的业务数据活动(即业务过程)。

- 1. 在数据域页面,单击已创建的数据域,进入该数据域的详情页。
- 2. 查看数据域的详细信息。

在基本信息区域,即可查看目标数据域的英文缩写、中文名、英文名、负责人、创建人、创建时间, 以及最近一次修改该数据域的修改人及修改时间等信息。您也可以根据业务需求,单击右上角的编辑, 修改目标数据域的相关信息。

⑦ 说明 DataWorks不支持删除系统默认创建的数据域。

基本信息			编辑
英文缩写: default	英文名: default data domain	中文名:默认	负责人: system
备注: 默认数据域。如果业务过程没有指定所属数	数据域,会被默认放到该数据域中。		
创建人: system	创建时间: 2021-06-22 17:38:46		
修改人: system	修改时间: 2021-06-22 17:38:46		

3. 为目标数据域添加业务过程。

创建数据域后,系统默认为当前数据域添加命名后缀为 __default 的业务过程。您可前往**业务过程**页, 单击对应业务过程**操作**列的**编辑**,修改**数据域**为当前数据域。

⑦ 说明 如您需要创建新的业务过程,请参见业务过程。

管理数据域

您可以在数据域页面,对已创建的数据域进行如下管理操作:

≡	🍿 DataWorks	数仓规划	• •		● 数仓规划 ● 数据标准	∂ 维度建模 ∂ 数据指标	👌 🍳 🦁 🗾 🗄
8≣	业务分类	******					
۲	数仓分层	安义括理			读 日本域 门床裙 教学授荣		
♣	公共层	ACIE/4/2018 1 20/27 1 20:37 (21:2:0/	erente, okaoliiiikov	(1), 7) 医亚方沃底师起致16。 知义勿2	ak, Linovak, Tjirovak, ext-kakter		
	数据域						
	业务过程		客服	财务	(供应链	新售	市场
⊟	应用层、	×	crm	financial	supply	sales	mk
191	维度管理	+	最后更新2021 负责人: dataw	最后更新2021 负责人: dataw	最后更新2021 负责人: dataw	最后更新2021 负责人: dataw	最后更新2021 负责人: dataw
Ŷ	建模空间	新建城					
			区 修改	乙 惨改	区修改 面 删除	区修改 自制除	ビ 修改

● 查看数据域概览。

您可以在**数据域**页面,查看所有已创建数据域的基本信息,包括数据域的名称、功能描述、最新更新时间 及对应负责人等。方便您快速筛选业务数据。

修改数据域。

单击目标数据域的 近棘 图标,即可修改该数据域的基本信息。

• 删除数据域。

单击目标数据域的 📾 📾 图标, 即可删除该数据域。

? 说明

- DataWorks不支持删除系统创建的数据域。
- 如果您的数据域中已经添加了业务过程,则您需要将数据域中的所有业务过程删除后,才能删除该数据域。

6.1.2.4.2. 业务过程

业务过程是企业在指定的数据域中所执行的业务活动,是数据建模所需要分析的逻辑主体。例如,交易域中可 以有加入购物车、下单、支付等业务过程。本文为您介绍如何创建并使用业务过程。

前提条件

已根据规划准备好对应数据域,系统已默认提供数据域,如需创建新的数据域请参见<mark>数据域</mark>。

背景信息

业务过程是对业务活动流程的描述,例如在电商领域,加购、下单、支付等都可以是一个业务过程。进行业务 效果分析时,业务过程有非常典型的应用,例如常用漏斗分析,即将购买商品的业务活动分解为浏览商品、加 入购物车、下订单、付款、确认收货等业务过程,统计每个业务过程的"订单数",即可以针对"订单数"这 一指标做漏斗分析。



如上图所示,您可以对数据域下的业务规划具体的业务过程,创建业务过程后:

- 后续为每个业务过程规划创建用于衡量业务过程的原子指标和派生指标。
- 后续业务过程中,实际业务的真实数据表(明细表)也可关联具体的业务过程。

创建业务过程

DataWorks的数据建模已为您创建好一个默认的业务过程(英文名为**default**,中文名称为**默认**),如果您的 业务比较简单,无需明细多个业务过程,那您无需再创建新的业务过程,在后续建模配置过程中,直接使用此 默认业务过程即可。

如果您的业务比较复杂,则您可以根据实际业务需求,创建符合企业实际应用的业务过程。创建业务过程操作 如下。

- 1. 进入数据建模。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 选择工作空间所在地域后,单击相应工作空间后的数据开发。
 - iv. 单击左上方的**三**图标,选择**全部产品 > 数据建模 > 数仓规划**,默认进入数仓分层页面。
- 2. 在数仓规划页面的左侧导航栏,单击公共层 > 业务过程,进入业务过程页面。
- 3. 创建业务过程。
 - i. 在业务过程页面, 单击新建业务过程。

ii. 在**新建业务过程**对话框,配置各项参数。

新建业务过	程。	×
英文缩写: *	请输入英文缩写	
英文名: *	请输入英文名称	
中文名: *	请输入中文名称	
数据域: *	默认 く	•
父级业务过程	请选择父级业务过程	•
负责人: *		•
备注:		
	0/2048	
	确认	取消
参数	描述	
	业务过程的英文缩写。	
英文缩写	英文缩写由小写英文字母、数字、下划线(_)组成,并且以小写英文字 128字符。	母开头,最大长度为
	业务过程的英文名称。	
英文名	英文名称由英文字母、数字、下划线(_)、and(&)组成,并且以英文 长度为2048字符。	ζ字母或数字开头 <i>,</i> 最大
	业务过程的中文名称。	
中文名	中文名称由中文、英文字母、数字、下划线(_)、and(&)、英文括号 文字母或数字开头,最大长度为2048字符。	号组成 <i>,</i> 并且由中文、英
数据域	选择业务过程所属的数据域。	
父级业务 过程	可选参数。您可以选择配置新建的业务过程归属于某个父级业务过程。如 建的业务过程即为父级业务过程。	□果不配置该参数,则 新
负责人	创建业务过程的负责人。系统创建的业务过程负责人为system。	
备注	业务过程的备注信息。 最大长度为2048字符。	

4. 单击**确认**,完成创建。

后续步骤

完成业务过程创建后,您可以:

- 后续为每个业务过程规划创建用于衡量业务过程的原子指标和派生指标。
- 后续业务过程中,实际业务的真实数据表(明细表)也可关联具体的业务过程。

6.1.2.5. 应用层

6.1.2.5.1. 数据集市

数据集市是基于业务分类,面向特定应用场景或者产品的数据组织。通常位于数据应用层,依赖于公共层的整 合数据。本文为您介绍如何创建数据集市并管理主题域。

前提条件

已创建业务分类,用于对指定业务进行特定场景化的细分。详情请参见业务分类。

背景信息

数据集市是对某个业务分类制定细化的业务主题,并通过主题域,基于不同分析视角对目标集市中的数据进行 主题划分,最终面向业务应用统计分析数据。例如,运营平台数据集市。

创建数据集市

- 1. 进入数据建模。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 选择工作空间所在地域后,单击相应工作空间后的数据开发。
 - iv. 单击左上方的**三**图标,选择**全部产品 > 数据建模 > 数仓规划**,默认进入数仓分层页面。
- 2. 在数仓规划页面的左侧导航栏,单击应用层 > 数据集市,进入数据集市页面。

3. 创建数据集市。

i. 在数据集市页面, 鼠标悬停至 图标, 单击新建一级数据集市。

ii. 配置新建一级数据集市对话框的各项参数。

新建—级数据集市		×
英文缩写: *	请输入英文缩写	
集市名称: *	请输入中文名称	
集市类型:	请选择	
所属业务分类:	请选择	
负责人: *	jainala jamil	
备注:		
	0/2048	
	确认目	则消
参数	描述	
	数据集市的英文缩写。	
英文缩写	英文缩写由小写英文字母、数字、下划线(开头,最大长度为128字符。	_)组
	数据集市的名称。	
集市名称	集市名称由中文、英文字母、数字、下划线 成,并且以中文、英文字母或数字开头,最	〕(_) 大长度

数据集市的类型。	
----------	--

最大长度为2048字符。

集市类型	 业务集市:面向业务需求的集市。 数据应用集市:面向数据产品需求的集市。 公共集市:如果您希望创建各集市需要通用的应用层模型,则可以选择该类型。
所属业务分类	数据集市所属的业务分类。详情请参见 业务分类 。
负责人	创建数据集市的负责人。默认为当前登录账号。
备注	数据集市的备注信息。

4. 单击确认,完成创建。

如果已创建的数据集市已变更或已不需要,您可在左侧目录树右击数据集市名称,选择**编辑**或删除进行 操作。

⑦ 说明 删除数据集市前,您需要先删除数据集市的所有绑定关系。

管理主题域

主题域是将数据集市按照分析视角,进行数据主题的划分。您可以在目标数据集市页面的**主题域管理**区域, 查看当前数据集市挂载的主题域列表,并根据业务需求对目标主题域进行编辑或删除等管理操作。

⑦ 说明 删除指定主题域,不仅会删除该主题域与数据集市的绑定关系,同时,会将该主题域一并删除,请您谨慎操作。

乞付业务 顶向某个应用场景或者产品的数据组织,一般会依赖数据公共层。								
	⊕ C	创建人: dataworks_demo2		创建时间:	2022-01-17 20:09:57			
请输入数据集市	Q	修改人: dataworks_demo2 备注:		修改时间:	2022-01-17 23:52:54			
> 🎝 金融业务								
> 🎄 支付业务		新建子数据集市						
		集市名称	英文缩写	创建时间		备注	操作	
				没有数据				
	Ì	主题词管理						
		主题域名称	英文缩写	创建时间		备注	操作	
		收银台	cashier	2022-01-17 20:13:59			编辑	删除
		単 利 単 利 単 利 単 利 単 利 単 利 単 利 単 利 単 利 単 利	bill	2022-01-17 20:12:13			编辑	删除
								88

创建子数据集市

在一级数据集市下,如有其他子数据集市,可继续新建子数据集市。创建方法与一级数据集市相同。

金融业务 面向某个应用场暴音	成者产品的数据组织	,一般会依赖	^烫 数据公共层。						
		⊕ C	其木仁自						编辑
请输入数据集市		Q	英文缩写: finance		集市名称:	金融业务			
> & 金融业务			所属业务分类: 销售业务		集市类型:	₩ 务 集 市			
→ 品 支付业务	新建子数据集市		创建人:		创建时间:	2022-01-17 20:11:31			
- XX X(153255	编辑		修改人:		修改时间:	2022-01-17 20:11:31			
	删除		备注:						
			新建子数据集市						
			集市名称	英文缩写	创建时间		备注	操作	
			银行业务	bank	2022-01-17 20:23:52			编辑	删除
			主题域管理						
			主题域名称	英文缩写	创建时间]	备注	操作	-
									85

后续步骤

完成数据集市的创建后,后续面向具体业务分析创建应用表时,可以关联该数据集市。创建应用表,详情请参见创建应用表。

6.1.2.5.2. 主题域

主题域用于将数据集市按照分析视角进行划分,通常是联系较为紧密的数据主题的集合。您可以根据业务的关注点,将这些数据主题划分至不同的主题域。例如,电商行业通常分为交易域、会员域、商品域等。

前提条件

已创建数据集市,用于确定主题域分析的业务类别主题。详情请参见数据集市。

创建主题域

1. 进入数据建模。

- i. 登录DataWorks控制台。
- ii. 在左侧导航栏, 单击工作空间列表。
- iii. 选择工作空间所在地域后,单击相应工作空间后的数据开发。
- iv. 单击左上方的**三**图标,选择**全部产品 > 数据建模 > 数仓规划**,默认进入**数仓分层**页面。
- 2. 在数仓规划页面的左侧导航栏,单击应用层 > 主题域,进入主题域页面。
- 3. 创建主题域。
 - i. 在主题域页面, 鼠标悬停至 图标, 单击新建一级主题域。
 - ii. 配置新建一级主题域对话框的各项参数。

新建—级主题域		×
英文缩写: *	请输入英文缩写	
主题域名称: *	请输入中文名称	
负责人: *	and the second	~
所属数据集市: *	请选择	~
备注:		
	0/204	48
	确认	取消

参数	描述
英文缩写	主题域的英文缩写。 英文缩写由小写英文字母、数字、下划线(_)组成,并且以小写英文字母 开头,最大长度为128字符。
主题域名称	主题域的名称。 主题域名称由中文、英文字母、数字、下划线(_)、and(&)、英文括号 组成,并且以中文、英文字母或数字开头,最大长度为2048字符。
负责人	创建主题域的负责人。默认为当前登录账号。
所属数据集市	主题域所属的数据集市。详情请参见 <mark>数据集市</mark> 。
备注	主题域的备注信息。 最大长度为2048字符。

4. 单击确认,完成创建。

如果已创建的主题域已变更或已不需要,您可在左侧目录树右击主题域名称,选择**编辑**或删除进行操作。

创建子主题域

在一级主题域下,如有其他子主题域,可继续新建子主题域。创建方法与一级主题域相同。

收银台 将数据集市按照分析视角进行切:	分,比如在电商行业	, 通常分为会员、交易、商品等。				
请输入主题域 > ◎ 收报台 > ◎ 账单	• C • C • Q • 新建子主题域 編辑 · 翻除 · · ·	基本信息 英文第号: cashier 所屬数据集市: 支付业务 创建人: 修改人: 酱注:	基本信息 英文缩写: cashier 主题域名称: 收银台 所屋数据集市: 支付业务 创建人: 创建时间: 2022-01-17 20:13:59 修改人: 修改时间: 2022-01-17 20:13:59 銜注:			
		新建了主题线 主题域名称 收银台编码	英文缩写 number	创建时间 2022-01-17 20:14:42	皆注	撮作 编辑 删除

6.1.2.6. 维度管理

维度管理用于标准化定义维度,帮助您对数据建模过程中产生的维度进行统一创建和管理,保证维度唯一,避免重复创建。本文为您介绍如何创建并管理维度。

背景信息

维度管理是在数仓规划阶段,提前规划出创建模型时可能需要使用到的维度及相关属性,并集中管理。后续在 创建汇总表或应用表时,您可以关联该维度进行统计分析。

使用限制

维度管理中不支持创建相同的维度,即维度的英文缩写必须唯一。

创建维度

- 1. 进入数据建模。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 选择工作空间所在地域后,单击相应工作空间后的数据开发。
 - iv. 单击左上方的**≣**图标,选择**全部产品 > 数据建模 > 数仓规划**,默认进入数仓分层页面。
- 2. 在数仓规划页面的左侧导航栏,单击维度管理,进入维度管理页面。
- 3. 创建维度。
 - i. 在维度管理页面, 单击新建维度。

ii. 配置新建维度属性对话框的各项参数。

新建维度属性						×
维度中文名: * 维度英文名: * 维度英文缩写: * 维度属性:	请输入中文名称 请输入英文缩写 维度属性中文名 *	维度属性英文名 *	维度属性英文缩写 *	主键 □ □	操作 删除 删除	
					确认	取消

参数	描述
维度中文名	维度的中文名称。 中文名称由中文、英文字母、数字、下划线(_)、and(&)、英文括号组 成,并且以中文、英文字母或数字开头,最大长度为2048字符。
维度英文名	维度的英文名称。 英文名称由英文字母、数字、下划线(_)组成,并且以英文字母或数字开 头,最大长度为128字符。
维度英文缩写	维度的英文缩写。唯一标识该维度。 英文缩写由小写英文字母、数字、下划线(_)组成,并且以小写英文字母 开头,最大长度为128字符。
维度属性	为当前创建的维度添加维度属性。 输入指定行数,单击添加即可添加相应个数的维度属性,编辑维度属性的中 文名称、英文名称、英文缩写,并指定当前维度的主键。

4. 单击确认,完成创建。

查看并管理维度

您可以通过**维度中文名**或**维度属性中文名**进行搜索(序号1),快速查找指定维度,查看维度的属性信息 (序号2)。同时,也可以根据业务需求,对指定维度执行**编辑**或删除等管理操作。 ⑦ 说明 通过维度或维度属性中文名称搜索,使用的是模糊匹配,即输入关键词,即可查找出所有包含 关键词的维度。

维度管 通过维度	建 管理进行标准化的维度定。	义,使用户能够对数仓建模过程中	产生的维度进行统	;一的建设和管理,保证维度一致,防止重复	夏建设。			
维度中区	文名:		维度属性中分	2名:		<u>搜索</u> 重置		
新建	維度							
	维度中文名	名		维度英文名		文缩写 2	操作	
+	美日	c		category		ry	编辑	删除
+	美日	category		ory catego		category		删除
-	商品	commodity		ty comm		commodity		删除
	序号	维度属性中文名		维度属性英文名		维度属性英文缩写	主键	
	001	商品ID		ID		ID		
	002	商品类型		TYPE		ТҮРЕ	否	

后续步骤

维度创建完成后,后续在创建模型表时可以使用该维度作为数据分析的粒度,详情请参见创建应用表、创建汇 总表。

6.1.2.7. 建模空间

当您所需要管理多个DataWorks工作空间且需要复用一套数仓规划时,面对跨多个工作空间的复杂数据体系,可以通过设计空间来共享一套数据建模工具,针对整个数据体系进行统一的数仓规划、维度建模及指标定义等工作。本文为您介绍DataWorks建模空间的应用场景及配置步骤。

背景信息

DataWorks智能数据建模在进行跨多个工作空间数据的统一建模情况下,为您提供了数据模型设计空间与数据研发工作空间的概念,通过建模空间功能实现模型跨工作空间应用。

其中

- 数据模型设计空间为模型设计师视角, 主要用于统筹管理所有关联研发空间设计的模型数量。
- 数据研发工作空间为数据开发工程师视角,主要用于模型发布以及基于模型设计结果进行ETL数据开发。



模型设计师设置好**数据模型设计空间与数据研发工作空间**的关系后,模型设计师在进行模型(汇总表、维度 表、明细表、应用表)发布时,可以选择将模型物化到被关联的某一个研发空间,后续数据开发工程师即可在 对应的工作空间进行数据开发操作。

需要注意的是:

- 以上图为例,模型设计师在工作空间Project_1中进行模型设计(模型1),这套模型通过模型设计空间物化 部署至Project_2、Project_3后,如果在应用过程中发现模型1需要完善,则仅能在Project_1中进行修 改,Project_2、Project_3中无法直接查看到具体的模型设计内容也无法直接修改。
- Project_1被作为数据模型设计空间后,可以选择是否将模型物化发布到本空间(即Project_1可以将本空间作为物化空间); Project_2、Project_3被作为物化空间后,不能再作为模型设计工作空间。

注意事项

• 通过模型设计空间建立工作空间之间的模型设计研发关系时:

支持选择当前账号拥有空间管理员权限的工作空间,但不能选择已被添加为数据研发工作空间的工作空间。

• 删除工作空间间的模型设计物化关系时:

仅可删除当前账号拥有管理员权限的工作空间。

操作步骤

以下为您介绍如何通过模型设计空间建立DataWorks工作空间之间的模型设计物化关系,便于后续模型发布时的选择。

1. 进入数据建模。

- i. 登录DataWorks控制台。
- ii. 在左侧导航栏, 单击工作空间列表。
- iii. 选择工作空间所在地域后,单击相应工作空间后的数据开发。
- iv. 单击左上方的**三**图标,选择**全部产品 > 数据建模 > 数仓规划**,默认进入数仓分层页面。

- 2. 在数仓规划页面的左侧导航栏,单击建模空间,进入建模空间页面。
- 3. 在建模空间页面的模型空间配置区域,可查看当前的数据模型设计空间和数据研发工作空间映射关系,单击更改配置,可修改配置数据研发工作空间。

⑦ 说明 进入更改配置操作页面后,您仅可以操作(新增或删除)您有空间管理员权限的工作空间。

建模空间面对第多个工作空间的	复杂数据体系,可以通过设计空间来共享一督数据建模工	具,针对整个数据体系进 行统一地数 合规划、维度建模及指标定义等工作。		
模型空间配置 数据模型设计空间: 数据研发工作空间:	阿里云內容服务 (aliyun_content) 请选择	v		
已选择:	✔ DW智能数据建模演示空间 (aidm_aitong_1)	工作空间名称 (中文)	操作	
	确定	DW智能数据建模演示空间	1 删除	
0	<u>銀交</u> 取消			

- 您可以单击数据研发工作空间后的下拉框,选择需要物化发布的工作空间,单击确定,再单击提交, 即可将勾选的工作空间添加为数据研发工作空间。
- 您也可以单击当前展示的已选择的工作空间后的删除按钮,完成后单击提交,即可删除当前建模空间 与所选研发工作空间的映射关系。
- 完成更改配置回到建模空间页面后,您可以在**变更记录**模块看到各个空间管理员对建模空间的变更操作 配置记录。

后续步骤

完成上述步骤后,模型设计完成后,模型发布时,需要选择模型所发布的研发工作空间,操作详情请参见<mark>模型</mark> 发布。

6.1.3. 数据标准

6.1.3.1. 数据标准概述

DataWorks数据建模支持在建模前规划制定数据标准,或在建模使用过程中根据业务情况沉淀企业业务的数据标准。通过规范约束标准代码、度量单位、字段标准、命名词典,来保障后续建模与应用过程中数据处理的一致性,从源头上保障数据的标准化生产,节约后续数据应用和处理的成本。

应用场景

DataWorks的数据标准包含字段标准、标准代码、度量单位、命名词典。



例如,现有注册表和登录表两张表,注册表中存储了会员ID,字段名为user_id,登录表中也存储了会员ID, 字段名为userid,此时针对会员ID这个数据创建统一的数据标准,例如指定数据处理的标准代码、指定字段的 属性要求(例如字段的数据类型、长度、默认值等)、指定数据的度量单位。创建好数据标准后,后续在建模 过程中涉及到会员ID这个字段的设置时,即可直接关联此标准,以此来保障所有会员ID字段的标准统一。

DataWorks中字段标准的创建操作请参见字段标准。

6.1.3.2. 字段标准

字段标准是对含义相同但字段名称不同的数据进行统一规范管理的数据准则,字段标准可定义字段的取值范 围、度量单位等内容。当字段标准发生变化时可快速定位或变更对应的表,极大地提升了应用效率和准确率。 本文将为您介绍字段标准的创建方法。

背景信息

字段标准又称为数据字典,可理解为全局字段管理。可将多个表中含义相同但字段名不同的内容进行关联,并 对该字段制定相关的取值范围、度量单位、标准代码等内容。后期字段标准发生变化时,可快速定位和变更关 联的表,极大程度提升构建效率、应用效率、准确性和后期治理效果。

相关示例如下:

现有注册表和登录表, 注册表中存储了会员ID, 字段名为user_id,登录表中也存储了会员ID, 字段名为userid。此时两个表中包含的会员ID均是同一内容, 但是使用了两种不同的字段名进行描述, 字段出现了歧义便需要人工介入理解进行确认。

解决方法如下:

我们可通过DataWorks数据建模创建新的字段标准,预先在模型中定义会员ID标准,然后在创建注册表和登录 表时,将对应会员ID的字段引用该标准。在后期分析中可自动推荐两张表的关联关系和关联字段,极大地提升 应用效率和准确性。



注意层级关系

- 字段标准在创建时需要放在根目录、目录或标准集下方,相关说明如下。
 - 根目录: 目录中最顶级的目录, 所有的目录、标准集、标准只能放置在根目录下。
 - 目录:用于存放标准和标准集,类似操作系统的文件夹目录。
 - 标准集: 和目录类似, 但标准集内只能存放标准。
- 字段标准之间可存在继承关系,如买家ID标准和卖家ID标准均可继承于会员ID标准。

创建目录与标准集

1. 进入数据标准。

- i. 登录DataWorks控制台。
- ii. 在左侧导航栏, 单击工作空间列表。
- iii. 选择工作空间所在地域后,单击相应工作空间后的数据开发。
- iv. 单击左上方的**三**图标,选择**全部产品 > 数据建模 > 数据标准**,进入数据标准页面。
- 2. 在数据标准页面左侧导航栏,单击字段标准,进入字段标准页面。

3. 新建根目录。

- i. 在字段标准页面, 单击 庄 图标, 选择新建根目录。
- ii. 在弹出的新建根目录对话框中,填写名称,单击确认。

新建根目录		×
* 名称:		
	确认	取消

名称只能由小写英文字母、数字、下划线(_)组成,并且以小写英文字母开头。最大长度不可超过 128个字符。

根目录创建成功后,可在左侧字段标准区域下方的目录树中查看已创建的根目录。

- 4. (可选)新建子目录。
 - i. 在左侧目录树, 选择已创建完成的根目录。
 - ii. 右击根目录名称,选择**新建子目录**。

iii. 在弹出的新建子目录对话框中,填写名称,单击确认。

新建子目录		×
* 名称:		
	确认	取消

名称只能由小写英文字母、数字、下划线(_)组成,并且以小写英文字母开头。最大长度不可超过 128个字符。

- 5. (可选)新建标准集。
 - i. 在左侧目录树,选择已创建完成的根目录或子目录。
 - ii. 右击目录名称,选择新建标准集。
 - iii. 在弹出的新建标准集对话框中,填写名称,单击确认。

新建标准集		×
* 名称:		
	确认	取消

名称只能由小写英文字母、数字、下划线(_)组成,并且以小写英文字母开头。最大长度不可超过 128个字符。

创建标准

- 1. 在左侧目录树下,选择对应目录或标准集。
- 2. 右击目录或标准集名称,选择新建标准。
- 3. 在弹出的新建标准对话框中, 配置如下内容。

新建标准		×	
* 标准编号:	请填写标准编码		
* 英文缩写:	请埴写英文缩写		
* 英文名称:	请埴写英文名称		
* 中文名称:	请填写中文名称		
* 数据类型:	BIGINT		
长度:	请填写长度		
精度:	请填写精度		
非空:			
默认值:	请填写默认值		
业务定义:	请填写业务定义		
所属路径:	请选择所属路径 🗸		
父级标准:	请选择父级标准		
引用代码:	请选择引用代码		
	備认	取消	
参数项	参数描述		
	自定义编码。		
标准编号	标准编码由英文字母、数字、下划线(_)组成,并且以英文	字母开头。	最大长度为64个字
	符。		
	字段标准的英文缩写。		
英文缩写	英文缩写由小写英文字母、数字、下划线(_)组成,并且以	小写英文字	母开头。最大长度
	不可超过128个字符。		
	字段标准的英文名称。		
英文名称	英文名称由英文字母、数字、下划线(_)、and(&)组成,	并且以英文	文字母或数字开头。
	名称最大长度不可超过2048个字符。		
	字段标准的中文名称。		
中文名称	中文名称由中文、英文字母、数字、下划线(_)、and(&)、英文括号	号组成,并且以中
	文、英文字母或数字开头。名称最大长度不可超过2048个字	符。	
	对应字段的数据类型,包括如下:		
数据类型	BIGINT、DOUBLE、DECIMAL、STRING、DATETIME、BOOLE	AN.	

参数项	参数描述
长度	数据类型关联参数。 例如,选择DECIMAL类型,则长度对应DECIMAL(20,4)中的20。
精度	数据类型关联参数。 例如,选择DECIMAL类型,则长度对应DECIMAL(20,4)中的4。
非空	引用该标准的字段值是否可以为空。默认为空。
默认值	引用该标准的字段值未填写时的默认值。最大长度不可超过2048个字符。
业务定义	对业务的定义描述,最大长度不可超过2048个字符。
所属路径	标准所属的目录,您可从已创建的目录、子目录、标准集中选择。
父级标准	可选择已创建的标准作为父级标准,选择后与该标准产生继承关系,帮助您更好地识别字段 的关联关系。 例如,买家ID和卖家ID均属于会员ID,则买家ID标准和卖家ID标准的父级标准就是会员ID标 准。
引用代码	您可从已创建的标准代码中,引用指定标准代码来约束字段的取值范围,标准代码请参见 <mark>标</mark> 准代码。

删除

● 删除标准

⑦ 说明 字段标准如已被引用,则无法删除,需在对应部分删除引用。

i. 单击标准所在的标准集或目录名称。

ii. 在标准列表单击对应标准操作列的删除。

● 删除标准集、子目录、根目录

在左侧目录树直接右击标准集、子目录或根目录的名称,选择删除。

? 说明 不支持跨层级删除目录。

后续步骤

创建完成后,您可在数据建模配置事实表与维度表字段时进行关联,关联方法请参见字段管理:添加表字段。

6.1.3.3. 标准代码

标准代码是数据标准的取值范围,在标准代码中可设置某一数据标准可选择的数据的内容以及范围。例如性别 数据标准的标准代码内容应该为男或女。本文将为您介绍标准代码的创建方法。

背景信息

标准代码与枚举维度比较相似,但是在数据治理的作用要强于枚举维度,标准代码可以与数据质量打通,某个 字段被设定为标准代码以后,该字段的取值范围必须在标准代码以内,如取值范围不在标准代码以内,需要马 上治理。

例如,性别字段,设置标准代码的取值范围为男和女。则数据仅能存储男或女,存在其他数据时就需要马上治理。

登录系统并创建目录

- 1. 进入数据建模。
- 2. 进入标准代码。
 - i. 在数据建模页面的顶部菜单栏,单击数据标准,进入数据标准页面。
 - ii. 在数据标准页面的左侧导航栏,单击标准代码,进入标准代码页面。
- 3. 新建目录。
 - i. 在标准代码页面, 单击 图标, 选择新建目录。

ii. 在弹出的新建目录对话框中,填写名称,单击确认。

新建目录	×
* 名称:	
确认	取消

名称由小写英文字母、数字、下划线(_)组成,并且以小写英文字母开头。

iii. 目录创建成功后,可在左侧标准代码区域下方的目录树中查看已创建的目录。

创建标准代码

1. 选择目录,右击目录名称,单击新建标准代码。

2. 在弹出的新建标准代码对话框, 配置如下参数并添加枚举值。

新建标准	代码								×
代码编号: 代码名称: 英文名称: 描述:	* *	请填写代码: 请填写代码: 请填写英文: 请填写描述	编号 名称 名称				0/2048		
编码取值			编码名称	英英	(名称	编码含义		操作	
					没有数据				
添加	清空							< 1/1	>
								确认	取消

○ 参数配置

参数项	参数描述
代码编号	自定义编号。 代码编号由小写英文字母、数字、下划线(_)组成,并且以小写英文字母开头。最大长 度为128个字符。
代码名称	对应标准代码的名称。 代码名称由中文、英文字母、数字、下划线(_)、and(&)、英文括号组成,并且以中 文、英文字母或数字开头。最大长度不可超过2048个字符。
英文名称	对应标准代码的英文名称。 英文名称由英文字母、数字、下划线(_)、and(&)组成,并且以英文字母或数字开 头。最大长度不可超过2048个字符。
描述	对标准代码的自定义描述,最大长度不可超过2048个字符。

。 添加枚举值

单击新建标准代码对话框中的添加,配置如下参数。

? 说明

- 如需配置多个枚举值,可多次单击添加。
- 如需删除某一条枚举值,可单击对应枚举值操作列的删除。
- 如需清空或重新填写枚举值, 可单击**清空**。

编码取值	编码名称	英文名称	编码含义	操作			
				保存删除			
添加清空				< 1/1 >			
参数项	参数描述						
编码取值	枚举值的编码取值。	枚举值的编码取值。					
英文名称	对应枚举值的英文名称	7 \					
编码名称	对应枚举值的名称。						
编码含义	对枚举值的含义描述,最大长度不可超过2048个字符。						

删除标准代码

您可直接右击左侧目录树中的编码名称,选择删除。

⑦ 说明 如标准代码已被引用则不允许删除,需先删除对应的引用关系。

后续步骤

创建完成后,您可在数据建模配置事实表与维度表字段时进行关联,关联方法请参见字段管理:添加表字段。

6.1.3.4. 度量单位

字段参数的数量单位(如个、厘米等),系统已为您提供了预置的度量单位,您可根据自定义需求创建新的度 量单位。本文将为您介绍度量单位的创建方法。

操作步骤

- 1. 进入数据建模。
- 2. 进入度量单位。
 - i. 在数据建模页面的顶部菜单栏,单击数据标准,进入数据标准页面。
 - ii. 在数据标准页面的左侧导航栏,单击度量单位,进入度量单位页面。

3. 新建度量单位。

i. 在度量单位页面,右键单击目标度量类别,选择新建度量单位。

系统为您提供了货币单位、对象量词比例单位排名单位、时间单位等度量类别。

ii. 在弹出的新建度量单位对话框中, 配置如下参数。

新建度量单位	×
* 英文缩写: * 英文名称: * 中文名称: * 分类:	请填写英文缩写 请填写英文名称 请填写中文名称 货币单位 ✔ 确认 取消
参数项	参数描述
英文缩写	度量单位的编码。 英文缩写由小写英文字母、数字、下划线(_)组成,并且以小写英文字母开头。最大 长度不可超过128个字符。
英文名称	度量单位的英文名称。 英文名称由英文字母、数字、下划线(_)、and(&)组成,并且以英文字母或数字开 头。最大长度不可超过2048个字符。
中文名称	度量单位的中文名称。 中文名称由中文、英文字母、数字、下划线(_)、and(&)、英文括号组成,并且以 中文、英文字母或数字开头。最大长度不可超过2048个字符。
分类	目前仅能选择系统为您提供的分类,包括 货币单位、对象量词、比例单位、排名单 位、时间单位。

ⅲ. 单击确认。

度量单位创建成功后,可在左侧度量单位区域下方的目录树中查看已创建的度量单位。

后续步骤

度量单位创建完成后,您可在创建事实表与维度表的字段时与创建原子指标时进行关联使用,详情请参见发布 与管理表和原子指标。

6.1.3.5. 命名词典

命名词典为您提供管理业务名词、物理表、字段的词根和词素及其标准化翻译功能,您可以在DataWorks新建命名词典,或将已有词典导出,本文为您介绍详细操作步骤。

新建命名词典

如果您仅需添加单个命名词典,您可以直接进行单个命名词典的新建。

- 1. 进入数据建模。
- 2. 进入命名词典。

i. 在数据建模页面的顶部菜单栏,单击数据标准,进入数据标准页面。

- ii. 在数据标准页面的左侧导航栏,单击命名词典,进入命名词典页面。
- 3. 新建命名词典。
 - i. 在**命名词典**页面单击新建。
 - ii. 在弹出的页面中配置命名词典的中文名称、英文名称和英文缩写参数。

新建命名词典		×
中文名称: *	講論入中文名称	
央义名称: * 英文缩写: *	请输入央义名称 请输入英文缩写	
		初认 取消

其中各参数的配置格式要求如下。

参数	格式要求
中文名称	需以汉字、字母或数字开头,支持使用下划线(_)、and符号(&)、括号。
英文名称	需以字母或数字开头,支持使用下划线(_)、and符号(&)。
英文缩写	需填写小写字母或数字。

iii. 完成参数配置后单击确认。
 在命名词典页面即可看到新建的命名词典内容。

导入命名词典

如果您有大量命名词典待创建,您可以选择批量导入命名词典,DataWorks为您提供了导入模板,您可按照模板表格填写好待导入的名称词典后批量导入。

- 1. 参考上述步骤进入命名词典页面后, 单击导入。
- 2. 在导入弹窗中选择导入模式,下载导入模板并批量上传。



i. 在命名词典的导入弹窗中,选择导入模式为跳过已存在的同名对象或覆盖已存在的同名对象。

ii. 单击下载模板下载批量导入命名词典的模板表格。

iii. 在模板表格中,填写并保存待导入的各命名词典的中文名称、英文名称、英文缩写。各参数的格式要求如下。

参数	格式要求
中文名称	需以汉字、字母或数字开头,支持使用下划线(_)、and符号(&)、括号。
英文名称	需以字母或数字开头,支持使用下划线(_)、and符号(&)。
英文缩写	需填写小写字母或数字。

⑦ 说明 下载模板表格后,如果一开始无法编辑表格,可单击表格顶部的启用编辑按钮。

iv. 单击选择文件,选择上述保存的命名词典模板后单击确定。

后续您可以在导入结果的提示弹窗中查看导入是否成功与导入细节。

导入				×
		\checkmark		
		导入完成		
	成功:1条	总计1条	失败: 0 <u>条</u>	
		直有详结		
			重新导入	完成

导出命名词典

您也可以在命名词典的页面选择待导出的命名词典,单击底部的**批量导出**,将选择的命名词典批量导出,或 者直接单击顶部的**批量导出**,将所有的命名词典批量导出。

命名词典	Ę.				
管理业务名词	、物理表、字段的词根和词素及其	其标准化翻译			
	批	量导出所有命名词典			
新建	导入 全部导出 请输	入中文名、英文名、英文缩写	Q		
	中文名	英文名	英文缩写	操作	
	测试1	test1	test1	编辑	制除
	测试	test	test	编辑	删除
					く _{上一} 页 1 下一页 >
	批量导出	所选命名词典			
批量删除	批量导出 已选择2				

后续步骤

创建完成命名词典后,您可以在数仓规划的各层的检查器配置中,选择使用,检查器的配置详情请参见配置数

仓分层检查器。

6.1.4. 维度建模

6.1.4.1. 维度建模概述

DataWorks的数据建模产品遵循Kimball维度建模理论,使用DataWorks的维度建模功能进行数据仓库模型设计时,您可根据业务情况设计并创建维度表、明细表、汇总表、应用表,并可将模型快速发布到相应的研发引擎。同时,还可以使用逆向建模,将已有的MaxCompute物理表逆向生成模型。

建模视角说明

维度建模支持从**公共层、应用层**对创建的模型表进行归类,您可以根据实际需求,选择合适的层级进行建模,不同层级支持的管理视角不同。

≡	🍿 DataWorks 🕯	推度建模		≡	n DataWorks	维度建模	~
0	维度建模	维度建模	E C	Ø	维度建模	维度建模	∃ ₹ G
₩	数据落标 1	公共层 应用层	未分层	T	数据落标		土公司
3	逆向建模	□ ▲ 请选择业务分类	~	в	 逆向建模		ホカ法
	2	⊕ 数据域	Q			♥ ^	-
		✓ 間 业务分类 → 計 (26)			2	✔ ⊕ 数据域	Q
						, 晋业务分类 ₄₎	
		· · · · · · · · · · · · · · · · · · ·				→ 創活教提资产 (7)	
		> ==			<u>o</u>		
	3	∨ 🗄 Data_modeling (19)				> 🗂 维度表 (6)	
		> 🏭 业务过名称1 (6)				> 凸 汇总表 (1)	
		∽⊕ 默认 (8)				> ⊕	
		> 🗂 维度表 (7)				> 🌐	
		→ 🗂 明细表 (1)					
		> 🕀 数据域1 (1)				>	
		~ 🕀 d (4)		Γ_		> ()	
		> 🗂 维度表 (2)				> 🌐 (2)
		→ □ 明细表 (2)				> 🕀	

- 公共层:用于加工、整合贴源层输入的公共数据,建立统一的指标维度,构建可复用面向分析和统计的明 细事实数据和汇总数据。该层级支持使用数据域或业务分类视角管理模型表。
- 应用层:基于实际应用需求,获取公共层加工整合后的数据,面向具体应用场景或指定产品进行的个性化数据统计。该层级仅支持使用业务分类视角管理模型表。
- 未分层:如果已创建的模型表未划分层级归属(即公共层或应用层),则该表会被放置于未分层中。

选择对应层级后,您只能在相应层级下创建目标表。并且可以在目录树切换该层级支持的管理视角,查看并管理层级下的模型表。

维度表

结合业务的数据域的规划,提取出各业务数据域中进行数据分析时可能存在的维度,并将维度及其属性通过维度表的方式存储下来。例如,在进行电商业务数据分析时,可用的维度及其属性有:订单维度(属性包括订单ID、订单创建时间、买家ID、卖家ID等)、用户维度(性别、出生日期等)、商品维度(包括商品ID、商品名称、商品上架时间)等,此时您就可以将这些维度和属性创建为订单维度表、用户维度表、商品维度表等,将维度属性记录作为维度表的字段。后续您可将这些维度表部署到数仓中,通过ETL将实际维度数据按照维度表定义的方式进行存储,方便业务人员在后续的数据分析时进行取用。

更多维度表的设计创建详情可参见创建维度表。

明细表

结合业务过程的规划,梳理分析各业务过程中可能产生的实际数据,将这些实际数据字段通过明细表的方式存储下来。例如下订单这一业务过程中,您可以创建下订单这一明细表,用于记录下单过程可能产生实际数据字段,例如订单ID、订单创建时间、商品ID、数量、金额等。后续您可将这些明细表部署到数仓中,通过ETL将 真实的数据按照明细表的定义方式进行汇总存储,便于业务分析时取用。

更多事实表的设计创建详情可参见创建明细表。

汇总表

您可以结合业务数据分析和数仓分层,将一些明细的事实数据和维度数据先进行汇总分析,创建汇总表,后续 数据分析时直接取用汇总表中的数据即可,无需再取用明细表和维度表中的数据。

更多汇总表的设计创建详情可参见创建汇总表。

应用表

您可以结合业务数据分析和数仓分层,将一些明细的事实数据和维度数据先进行汇总分析,创建汇总表,后续 数据分析时直接取用汇总表中的数据即可,无需再取用明细表和维度表中的数据。

更多应用表的设计创建详情可参见创建应用表。

逆向建模

逆向建模主要用于将其他建模工具生成的模型反向建模至DataWorks的维度建模中。例如,当您已通过其他建模工具生成模型,此时,想更换为DataWorks的智能建模进行后续建模工作,则可以使用逆向建模功能。该功能无需您再次执行建模操作,即可帮助您快速将已有模型反向建模至DataWorks的维度建模中,节省了大量的时间成本。

更多逆向建模的操作详情可参见逆向建模。

6.1.4.2. 创建维度表

维度表用于确定您观察业务状况的视角,存储不同观察视角的维度描述。例如,分析产品销售情况,您可以选 择按产品类别及时间两个维度进行分析,创建产品维度表及时间维度表。本文将为您介绍维度表的创建方法。

前提条件

- 已创建数仓分层。数仓分层是将相同功能作用的表挂载至统一的数据层,方便您查找使用。维度表通常挂载于公共维度层(DIM层),您也可以根据业务需求,选择将维度表挂载至其他数据层。创建数据分层,详情请参见创建数仓分层。
- 已创建数据域,用于确定维度表需要存储哪类业务类型的数据。创建数据域,详情请参见数据域。

背景信息

结合业务的数据域的规划,提取出各业务数据域中进行数据分析时可能存在的维度,并将维度及其属性通过维度表的方式存储下来。例如,在进行电商业务数据分析时,可用的维度及其属性有:订单维度(属性包括订单ID、订单创建时间、买家ID、卖家ID等)、用户维度(性别、出生日期等)、商品维度(包括商品ID、商品名称、商品上架时间)等,此时您就可以将这些维度和属性创建为订单维度表、用户维度表、商品维度表等,将维度属性记录作为维度表的字段。后续您可将这些维度表部署到数仓中,通过ETL将实际维度数据按照维度表定义的方式进行存储,方便业务人员在后续的数据分析时进行取用。



如上图所示:

- 创建维度表时:
 - 可以关联本维度表是用于分析哪个业务分类、哪个数据域,便于后续直接查看具体某一业务分类或数据 域下,有哪些维度表。
 - 可以指定维度表后续在数据建模分析使用时,维度表的数据存储于数仓中的哪个数据分层,一般情况下 维度表可存储于公共维度层(DIM层)。
- 维度表创建后,您可以将维度的属性添加为维度表的字段,且可以对维度表进行关联和分区的设置,并且 可使用统一的数据标准来进行字段设置,保障全数据域中的维度数据的属性是一致的。
- 维度表创建配置完成后,可发布物化至计算引擎,后续在计算引擎中使用维度表进行数据分析。
- 维度表创建配置后,在进行数据派生指标和汇总表的设计创建时,可直接关联使用维度表中的维度。

创建维度表

- 1. 进入维度建模。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击**工作空间列表**。
 - iii. 选择工作空间所在地域后,单击相应工作空间后的进入数据开发。
 - iv. 单击左上方的**三**图标,选择**全部产品 > 数据建模 > 维度建模**,默认进入**维度建模**页面。
- 2. 创建维度表。
 - i. 在**维度建模**页面,鼠标悬停至 <u></u>图标,单击创建维度表。

ii. 配置维度表的基本信息。

📄 创建维度表 🗙							Ξ
◎ 保存							
基本信息							
* 数仓分层:	公共层公共	共维度模型层	~	业务分类:	请选择业务分类	~	
* 数据域:	默认 (defaul	t)	~	存储策略:	请选择存储策略	~	
表名规则:	请选择检查器	2 A	~				
* 表台:	请输入表名 (法检) 主日						
水中又有, 生命周期,	明湘八水2004	天描述,				0/20/18	
* 表类型:	普通维度表	× 14421				0/2040	▲显示基本
参数项		描述					
数仓分层	收仓分层 选择已创建的数仓分层,仅支持选择公共层的公共维度模型层。后续使用此维度表 时,维度表中的数据即存储在数仓的这一层中。创建数仓分层的操作详情请参见创建数 仓分层。				挂度表 见 <mark>创建数</mark>		
数据域		已创建完成的数据域,认	羊情i	请参见 <mark>数据</mark>	域。		
业务分类		已创建的业务分类,详情请参见 <mark>业务分类</mark> 。					
存储策略		维度表存储数据的策略,可以选择 每日增 量、 每日全 量等多种策略。					
表名规则		选择已配置的检查器。用于检测表名称是否符合该检查器制定的表命名规则。配置检查器,详情请参见 <mark>配置数仓分层检查器</mark> 。				配置检查	
表名		表的内部名称,可由小 ^国 头。	写英:	文字母、下	划线(_)、数字组成,并	<u>;</u> 且以小写英文	文字母开
表中文名		表的中文名称,可由中文、英文字母、下划线(_)、数字、and(&)、英文括号组 成,并且以中文、英文字母或数字开头。				括号组	
生命周期		维度表存在的时间周期。	,最关	大可保存36	000天。		
表类型		 根据实际情况,选择如下类型。 普通维度表:普通的维度表,无其他需求可选择该类型。 枚举维度表:可进行枚举的维度表,例如性别维度表。 层级维度表:存在字段需要设置层级关系,支持指标做上卷和下钻使用。例如国家、省份、市县等。 				列如国	
描述		维度表的描述信息。					

3. 配置完成后,单击保存,保存成功后,可在左侧目录树进行查看。

后续步骤

表创建完成后,您仍需为表配置字段管理、关联关系、分区设置等内容,并将表发布至对应环境,详情请参见发布与管理表。

6.1.4.3. 创建明细表

明细表用于存储大量能够体现业务活动状况的实际数据或详细数值,是数据聚合后依据某个维度生成的结果 表。例如,分析产品销售情况,您可以创建销售明细表,用来存储产品维度(作为外键)、时间维度(作为外 键)以及销售总量。本文为您介绍如何创建明细表。

前提条件

- 已创建数仓分层。数仓分层是将不同功能作用的表挂载至统一的数据层,方便您查找使用。明细表通常挂载于数据明细层(DWD层),您也可以根据业务需求,选择将明细表挂载至其他数据层。创建数据分层, 详情请参见创建数仓分层。
- 已创建业务过程,用来确定明细表需要存储哪类具体业务活动的数据。创建业务过程,详情请参见业务过程。
 程。

背景信息

结合业务过程的规划,梳理分析各业务过程中可能产生的实际数据,将这些实际数据字段通过明细表的方式存储下来。例如下订单这一业务过程中,您可以创建下订单这一明细表,用于记录下单过程可能产生实际数据字段,例如订单ID、订单创建时间、商品ID、数量、金额等。后续您可将这些明细表部署到数仓中,通过ETL将 真实的数据按照明细表的定义方式进行汇总存储,便于业务分析时取用。



如上图所示:

- 创明细表时:
 - 可以关联本明细表是用于分析哪个业务分类和业务过程,便于后续直接查看具体某一业务分类或业务过程下,有哪些明细表。
 - 可以指定明细表后续在数据建模分析使用时,明细表的数据存储于数仓中的哪个数据分层,一般情况下 明细表可存储于数据明细层(DWD层)。
- 明细表创建后,您可以为明细表添加字段,且可以对明细表进行关联和分区的设置,并且可使用统一的数据标准来进行字段设置,保障全数据域中的数据的属性是一致的。
- 明细表创建配置完成后,可发布物化至计算引擎,后续在计算引擎中使用明细表进行数据分析。

操作步骤

1. 进入维度建模。

- i. 登录DataWorks控制台。
- ii. 在左侧导航栏, 单击工作空间列表。
- iii. 选择工作空间所在地域后,单击相应工作空间后的进入数据开发。

iv. 单击左上方的**三**图标,选择**全部产品 > 数据建模 > 维度建模**,默认进入**维度建模**页面。

2. 创建明细表。

i. 在**维度建模**页面,鼠标悬停至 <u></u>图标,单击创建明细表。

ii. 配置明细表的基本信息。

🗎 创建明细表 🗙							Ξ
圖 保存							
基本信息							
* 数仓分层:	公共层 明維	田数据层	小小学	务分类:	请选择业务分类	~	
* 业务过程:	请选择	```	存储	诸策略:	请选择存储策略	~	
表名规则:	请选择检查器	R N	*				
* 表名:	请输入表名						
* 表中文名:	请输入表显示	沼					
生命周期:		天 描述:				0/2048	
* 表类型:	事实事务表	~				1	显示基本
参数项		描述					
数仓分层		选择已创建的数仓分层,仅	(支持)	先择公共	层的 明细数据层 。后续使用此明约	细表时,	明细表
~ 0 <i>///</i> 4		中的数据即存储在数仓的这	四层中	P。详情:	请参见创建数仓分层。		
		己创建立学校业务学校。					
业务过柱		匕 刨建元 成 的业务过程,许	-	≶児业分∶	也在。		
业冬分米		已创建的业条分类 详情语	⇒□√	₩条分类			
±7772					0		
存储策略		明细表存储数据的策略,可	「以选邦	≩每日增	量、 每日全 量等多种策略。		
主力切回		选择已配置的检查器。用于	检测者	表名称是 [:]	否符合该检查器制定的表命名规则	り。配置権	佥查器,
表名规则		详情请参见 <mark>配置数仓分层检</mark>	渣器。	,			
表名		表的内部名称,可由小写英	文字母	母、下划:	线(_)、数字组成,并且以小写	英文字母	开头。
表中文名		表的中文名称,可由中文、	英文与	字母、下:	划线(_)、数字、and(&)、英	主文括号组	组成,并
		且以中文、英文字母或数字	开头。	•			
十个田田		四细末方方的叶词 田坦 三			от		
土叩周期		·明细衣仔仕的的间向期。 軍	之口的	木仔3000	して。		
韦米刑		根据实际情况 选择重变重	[名丰	国期村	1昭表武罗加耒		
仪大王		11以加大时间, 加汗尹大司	- 77 AX	、 /uj // ////	而化为术用化。		
描述		明细表的描述信息。					

3. 配置完成后,单击保存,保存成功后,可在左侧目录树进行查看。

后续步骤

表创建完成后,您仍需为表配置字段管理、关联关系、分区设置等内容,并将表发布至对应环境,详情请参 见<mark>发布与管理表</mark>。

6.1.4.4. 创建汇总表

汇总表用于组织一个数据域下相同时间周期、相同维度的多个派生指标的统计数据,为后续的业务查询,OLAP分析,数据分发等提供基础。本文为您介绍如何创建汇总表。

前提条件

- 已创建数据分层。数据分层是将相同功能作用的表挂载至统一的数据层,方便您查找使用。汇总表通常挂载于汇总数据层,用于将一个统计粒度(维度或维度组合)及该粒度下的多个派生指标汇总输出,为后续的业务查询,数据分发等提供基础。您也可以根据业务需求,选择将汇总表挂载至其他数据层。创建数据分层,详情请参见创建数仓分层。
- 已创建数据域。汇总表是基于数据域创建的,用于确定您需要建模的业务类型,承载业务过程。创建数据域,详情请参见数据域。
- 已创建时间周期,用于确定统计数据的时间范围。详情请参见时间周期

背景信息

汇总表是将一个数据域中的多个派生指标通过时间周期及关联维度整合在一起,关联的维度、时间周期和派生 指标,用于生成汇总表中的统计字段,帮助您进行报表等分析展示。您可以使用汇总表呈现相同时间、相同维 度下,多个派生指标所统计的业务情况。

操作步骤

- 1. 进入维度建模。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 选择工作空间所在地域后,单击相应工作空间后的进入数据开发。
 - iv. 单击左上方的■图标,选择全部产品 > 数据建模 > 维度建模,默认进入维度建模页面。
- 2. 创建汇总表。
 - i. 在**维度建模**页面,鼠标悬停至 🕞 图标,单击创建汇总表。
 - ii. 配置汇总表的基本信息。

基本信息						
* 数仓分层:	公共层汇总数据层	~	业务分类:	请选择业务分类	~	
* 数据域:	默认 (default)	~	粒度:	请选择	~	
* 时间周期:	请选择时间周期	~	修饰词:	请选择修饰词	~	
表名规则:	请选择检查器	~				
* 表名:	请输入表名					
* 表中文名:	请输入表显示名					
生命周期:	天 描述:				0/2048	
* 表类型:	普通汇总表 🗸					

参数	描述	示例值
数仓分层	挂载汇总表的数据层。默认选择 公共层的汇总数据 层(DWS)。您也可以根据业务需求,选择将汇总表挂载至 其他数据层。创建数据分层,详情请参见 <mark>创建数仓分层</mark> 。	汇总数据层(DWS)
业务分类	选择已创建的业务分类,详情请参见 <mark>业务分类</mark> 。	销售类
数据域	选择汇总表所挂载的数据域,用于确定汇总表需要整合呈现的统计数据主题。创建数据域,详情请参见数据域。 ⑦ 说明 每个汇总表只能挂载至一个数据域中。	交易域
时间周期	用于确定汇总表需要整合的统计数值的时间范围。例如,近 一天,近一周。 您需要从已创建的时间周期中选择。如果已创建的时间周期 中没有符合您业务需求的时间,则您可以参考 <mark>时间周期</mark> 新 建。	1w(近7天)
修饰词	用于确定统计数据的业务范围。 您需要从已创建的修饰词中选择。如果已创建的修饰词中没 有符合您业务需求的修饰词,则您可以参考 <mark>修饰词</mark> 新建。	线上门店
粒度	选择汇总数据的粒度。	无
表名规则	选择用于检查表命名规则的检查器。此处可选择您在数仓规 划时,在各个数据分层中已创建的检查器,操作详情请参 见 <mark>配置数仓分层检查器</mark> 。	无
表名	汇总表的表名称。建议命名格式为 dws_{数据域}_{统计维 度}_{自定义命名标签}_{时间周期} , 方便您快速了解汇 总表的统计内容。 表名称由小写英文字母、下划线(_)、数字组成,并且以小 写英文字母开头。	dws_trade_buyer_subpay _1d,其中: 数据域为 <i>trade</i> 。 统计维度为 <i>buyer</i> 。 自定义命名标签为 <i>subpa</i> <i>y</i> 。 时间周期为 <i>1d</i> 。
表中文名	表显示出来的中文名称。 中文名称由中文、英文字母、下划线(_)、数字、 and(&)、英文括号组成,并且以中文、英文字母或数字开 头。最大长度为256字符。	买家交易分阶段付款汇总表
生命周期	表存在的时间周期,单位为天。	90天
描述	表的描述信息。	无

参数	描述	示例值
表类型	 您可以根据业务需求,选择如下表类型。 普通汇总表:只能统计单粒度的数据,不会做多个粒度数据的再次汇总。 例如,创建 dws_demo_2 普通汇总表,语句如下。 create table dws_demo_2 (shop_code_comment '维度:门底code' 	
	<pre>shop_code comment ' 编设: 门店code', pay_amt comment ' 指标: 门店销售金额', pay_cnt comment '指标: 门店下单次数') comment '普通汇总表';</pre>	
	您只需要使用 select 语法查询目标指标字段,即可获 取固定粒度(门店粒度)的指标数据。普通汇总表不支持 使用组合逻辑查询多个统计粒度的数据。	
	轻度汇总表:可以统计多个粒度的数据,并将不同粒度的数据进行再次汇总。	
	例如,创建 dws_demo_1 轻度汇总表,语句如下。	普诵汇总表
	<pre>create table dws_demo_1 (merchant_code comment '维度: 商家code', company_code comment '维度: 子公司 code', shop_code comment '维度: 门店code', sku_code comment '维度: 商品code', pay_type comment '维度: 支付类型', pay_amt comment '指标: 销售金额', pay_cnt comment '指标: 下单次数') comment '轻度汇总表';</pre>	目地儿心水
	您可以通过 group by 语法,组合出多种不同粒度的派 生指标,进行统计汇总。 ■ 使用 group by merchant code 语句,统计商家	
	粒度的指标数据。 ■ 使用 group by merchant_code, company_code, s hop_code 语句,统计门店粒度的指标数据。	

3. 单击左上角的保存按钮,保存配置的汇总表信息。

后续步骤

表创建完成后,您仍需为表配置字段管理、关联关系、分区设置等内容,并将表发布至对应环境,详情请参见发布与管理表。

6.1.4.5. 创建应用表

应用表是面向具体业务场景时,用于组织相同时间周期、相同维度的多个原子指标、派生指标或统计粒度的统计数据,为后续的业务查询,OLAP分析,数据分发等提供基础。本文为您介绍如何创建应用表。

前提条件

- 已创建数据分层。数据分层是将相同功能作用的表挂载至统一的数据层,方便您查找使用。应用表通常挂载于应用数据层,用于将一个统计粒度(维度或维度组合)及该粒度下的多个指标汇总输出,为后续的业务查询,数据分发等提供基础。您也可以根据业务需求,选择将应用表挂载至其他数据层。创建数据分层,详情请参见创建数据分层。
- 已创建数据集市或主题域,用于确定统计数据的业务场景。详情请参见数据集市、主题域。
- 已创建时间周期,用于确定统计数据的时间范围。详情请参见时间周期。

背景信息

应用表是将一个数据集市或主题域中的多个原子指标或派生指标,通过时间周期及关联维度整合在一起,关联 的维度、时间周期、原子指标、派生指标,用于生成应用表中的统计字段,帮助您进行报表等分析展示。您可 以使用应用表呈现相同时间、相同维度下,多个指标所统计的业务情况。

操作步骤

- 1. 进入维度建模。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 选择工作空间所在地域后,单击相应工作空间后的进入数据开发。
 - iv. 单击左上方的**三**图标,选择**全部产品 > 数据建模 > 维度建模**,默认进入**维度建模**页面。
- 2. 创建应用表。
 - i. 在**维度建模**页面,鼠标悬停至 🕞 图标,单击创建应用表。
 - ii. 配置应用表的基本信息。

◎ 创建应用表 ×									
◎ 保存									
基本信息									
* 数仓分层:	应用层 应用	用数据层		~	* 集市/主题:	请选择		~	
粒度:	请选择			~					
* 时间周期:	请选择时间	请选择时间周期		~	修饰词: 请选择修饰词		~		
表名规则:	请选择检查	请选择检查器							
* 表名:	请输入表名								
* 表中文名:	请输入表显示名								
生命周期:			天 描述:					0/2048	
* 表类型:	普通应用表	~							▲显示基本
参数		描述							
数仓分层	数仓分层 挂载应用表的数仓分层。默认选择 应用层 分层归属的 应用数据层 。您也可以根据业务 需求,选择将应用表挂载至其他数据层。创建数据分层,详情请参见 <mark>创建数据分层</mark> 。					据业务 <mark>`层</mark> 。			
集市/主题 选择已创建			创建的数据集市或主题域,详情请参见 <mark>数据集市、主题域</mark> 。						
粒度		选择应用	月表的数据统计	十维度	,详情请参见	し维度管理。			

参数	描述
时间周期	用于确定应用表需要整合的统计数值的时间范围。例如,近一天,近一周。 您需要从已创建的时间周期中选择。如果已创建的时间周期中没有符合您业务需求的时 间,则您可以参考 <mark>时间周期</mark> 新建。
修饰词	用于确定统计数据的业务范围。 您需要从已创建的修饰词中选择。如果已创建的修饰词中没有符合您业务需求的修饰 词,则您可以参考 <mark>修饰词</mark> 新建。
表名规则	选择用于检查表命名规则的检查器。此处可选择您在数仓规划时,在各个数据分层中已 创建的检查器,操作详情请参见 <mark>配置数据分层检查器</mark> 。
表名	应用表的表名称。 表名称由小写英文字母、下划线(_)、数字组成,并且以小写英文字母开头。
表中文名	表显示出来的中文名称。 中文名称由中文、英文字母、下划线(_)、数字、and(&)、英文括号组成,并且以 中文、英文字母或数字开头。最大长度为256字符。
生命周期	表存在的时间周期,单位为天。应用表最多支持保存36000天。
描述	表的描述信息。
表类型	根据实际情况,选择如下类型。 ■ 普通应用表:只能统计单粒度的数据,不会做多个粒度数据的再次汇总。 ■ 轻度应用表: 可以统计多个粒度的数据,并将不同粒度的数据进行再次汇总。

3. 单击左上角的保存按钮,保存配置的应用表信息。

后续步骤

表创建完成后,您仍需为表配置字段管理、关联关系、分区设置等内容,并将表发布至对应环境,详情请参见<mark>配置发布表</mark>。

6.1.4.6. 发布与管理表

模型创建成功后,您需为维度表、明细表、汇总表、应用表配置字段管理、关联关系、分区管理等配置,完成 后将表进行发布。本文将为您介绍如何添加字段、配置字段属性、发布以及管理表。

? 说明

- 您也可以通过FML语句创建字段、关联关系及分区,操作方法请参见使用FML语句配置并管理数据表。
- 本文中字段管理、关联关系、分区管理的配置方法均适用于维度表、明细表、汇总表、应用表, 下文中以表泛指这四类表。

字段管理:添加表字段
您可以通过快捷模式或代码模式为表添加字段,其中快捷模式支持如下导入方式:

- 从表/视图导入:从计算引擎已存在的物理表或视图导入字段,即在查找已有表/视图下拉列表搜索并选择
 已存在的物理表或视图进行字段导入。
 - ⑦ 说明 当前仅支持从MaxCompute引擎的表或视图导入。
- 从指标导入: 从数据指标所有派生指标中选择所需指标作为模型字段。

⑦ 说明 仅汇总表和应用表支持选择从指标导入。

• 快捷模式: 从表/视图导入

字段管理							5 A 2 M
			快捷模式	代码模式			
从指标导	▲ 快速导	λ	统计粒度: 1 个	统计指标: 4 个		关联粒	这度/指标
从表/视图 查找已有题	导入 收起 へ 長/视图: ♪ №	MaxCompute 🗸 请輸	1 入表名/视图名搜索	输入表或视图名称		~	
已引用的题	長/视图: 表	odps.dw_scheduler_pr	r 🕑 🔿 × 🙋				
			<i>•</i>				
odps.dw_s	cheduler_pre.ad	dm_dp_ant_autotest_d	s_cloneur7o-载	入指定字段			×
请输入		Q					
三 序	号 字段名称	字段显示名	类型	描述	主键	外键	分区
1	item_id		string	商品id	否	否	否
2	seller_id		string	卖家id	否	否	否

	-	_, _, _ , _						
请輸入			Q					
	序号	字段名称	字段显示名	类型	描述	主键	外键	分区
	1	item_id		string	商品id	否	否	否
	2	seller_id		string	卖家id	否	否	否
3	3	category_id		string	类目id	否	否	否
	4	brand_id		string	品牌id	Ki	否	К
	5	item_name		string	商品名称	否	否	否
	6	description		string	商品描述	否	否	否
导入	送 (关	1 已选中 2 个字段						

i. 单击快捷模式下,从表/视图导入后的展开。

⑦ 说明 明细表和维度表无需执行该步骤。

ii. 在查找已有表/视图的输入框中输入相应名称,即可选择对应的表或视图,并在已引用的表后选择导入表或视图的全部字段或部分字段。

⑦ 说明 输入名称时支持模糊匹配,即输入关键字便可搜索到所有名称中包含关键字的表或视图。

- ▲图标表示导入全部字段。
- ◎图标表示导入部分字段。
- iii. 当选择导入部分字段时,在弹出的窗口中,会为您展示所选表的字段,您可以选择需要添加至模型的 字段,完成后单击界面下方的导入。
- 快捷模式: 从指标导入(适用于汇总表、应用表模型的创建)

字段	管理								К Я 2 У
				快捷模式代码	關模式				
从指	赫导入 《上史速导》			统计粒度: 1 个 纷	计指标:	4 ↑			关联粒度/指标
从表	影视图导入 收起 ^								
查找	论已有表/视图: ✔ M	axCompute	・ 请输	入表名/视图名搜索					~
EB	用的表/视图: 表	odps.dw_sche	eduler_p	r 🛃 🍥 🗙					
从指标导	杁								
名称:	请输入	Q	业务过程:	请选择		~	原	子指标: 请选择原子指标	2 ~
时间周期:	请选择时间周期	~	修饰词:	请选择修饰词		~			
•	中文名称		英文	名称	指标编码	原子指标	时间周期	修饰词	
	近3天_最低的&中文名称_name_2		atom	1_2_3d_test_0906_low	INDD B22D 243E A5	na me _2	近 3 天	中文名称,最低的	
	近3天_最低的&中文名称_aaaa1		8888	aa1_3d_test_0906_low	INDD 578A 51D0 E9	aa aa 1	近 3 天	中文名称,最低的	
导入	3 1 1 日选中 2 个指标							< 上一]	页 1 2 下页 >

i. 单击快捷模式下,从指标导入后的快速导入。

ii. 在弹出的窗口中,会为您展示当前已创建的派生指标,您可以在派生指标中选择需要添加至汇总表、 应用表的字段。同时,您还可以使用**时间周期、业务过程、修饰词、原子指标**筛选指定派生指标。

iii. 完成后单击界面下方的导入。

● 代码模式

代码模式为您提供代码输入的操作方式,单击代码模式后,弹窗中会根据已配置的模型信息自动生成建模 语言,您可以在此基础上修改模型信息,完成后单击下方的确定。

FML		へ FML(Fast Modeling language),适用于维度建模领域的类SQL语言。 帮助文档		
FML	央捷模式	<pre>c_test ALIAS 'dws_doc_test'</pre>	No. of the second second	
Max0 Hive	Compute DDL DDL	d' BIGINT COMMENT 'userId comment', "Name' STRING COMMENT 'userName comment'		ł
7 8 9 10 11 12	*/) PARTITIONED BY (ds ALIAS '业务日期;);	,yyyymmdd' STRING COMMENT '业务日期,yyyymmdd'		

字段管理:设置表字段属性

完成上述操作后,模型所需字段即已被添加至该模型中,此时您可以根据业务需求设置模型各字段的**关联字** 段、冗余字段、关联粒度/指标。

1. 设置字段属性。

字段属性默认展示**字段名称、类型、字段显示名、描述、主键、非空、度量单位、操作**等基础属性。 您可以在展示已添加字段的右上方单击**字段显示设置**,选择需要展示出来的字段属性,并根据需要进行 修改。

2. 设置字段的关联字段标准、关联标准代码。

用于为已添加的字段选择关联的字段标准和标准代码,规范字段的取值内容及范围。

- 关联字段标准:统一规范管理含义相同但字段名称不同的数据,定义字段的取值范围、度量单位等内容。
- 关联标准代码:设置某一字段标准可选择数据的内容以及范围。

3. 设置字段的冗余字段。

在已添加字段的操作列,单击冗余字段,即可设置字段的关联字段。

⑦ 说明 汇总表、应用表仅支持设置关联字段标准。

冗余字段	₽: a							×
						输入表或		
关联表名/	视图名: 🔥 🛚	MaxCompute 🗸	表 odps.dw_sc	heduler_pre				~
关联字段:	请选	择字段名		▶ 冗余字段名	请输	入关键字		Q
	序号	字段名称	字段显示名	描述	主键	外键	分区	
	01	item_id		商品id	否	否	否	
	02	seller_id		卖家id	否	否	否	
	03	category_id		类目id	否	否	否	
	04	brand_id		品牌id	否	否	否	
促左	取当							

4. 设置字段的关联粒度/指标。

汇总表、应用表的各字段数值的统计类型可通过设置字段的**关联类型**来指定,包括**统计粒度、派生指** 标和原子指标三种。

- 统计粒度:用于关联维度表和维度表中的字段。例如,商品维度、商家维度。
- 派生指标:指定汇总表字段需要整合的统计数值的派生指标。例如,近7天盒马App支付下单的金额数。
- **原子指标**:指定汇总表字段需要整合的统计数值的原子指标。例如,支付下单的金额数。

? 说明

- 仅汇总表、应用表支持字段的统计粒度/指标。
- 从表导入的字段、代码模式添加的字段无默认关联类型,您可以手动设置字段的关联类型为 哪种类型。

如果您想设置或修改字段的关联类型,可以参考以下步骤进行修改。

字段	管理							К.Я. К.У.
				t	共建模式 代码模式			
从指	标导入	仕	X.	纺	计粒度: 1个 统计指标: 4	k	1 关联粒度/	/指标
从表 查找	/视图导入 已有表/视图:	收起 ^ M	axCompute	✓ 请输入	表名/视图名搜索		~	
关联粒	度/指标							×
字段关目	联类型:	全部		~				
序号	字段		关联类型		关联内容			
0	hm_order_i _paid_mem	number_of ibers_3d	统计粒度	~	dim_sku(商品表)	✓ sku_coo	je(商品编码)	~
1	hm_order_s	se_uv_1d	统计指标	~	INDD621328A0D4(近1天_盒马_	搜索UV)		~
2	pc_pay_the ount_3d	_order_am	统计指标	~	INDDE6D133AAC6(近3天_pc端_	支付下单金额)		~
3 确定	关闭	Ð						

单击字段管理右侧的**关联粒度/指标**,在弹出的窗口中,修改字段关联的类型和关联的内容,完成后单 击**确定**。

5. 完成设置后单击左上角的保存。

发布表

维度表、明细表、汇总表、应用表,可直接发布物化至绑定的DataWorks引擎中,用于后续计算引擎进行数据 开发、数据分析等。当前支持将表直接发布物化至MaxCompute、Hologres、E-MapReduce、CDH、MySQL 引擎。

您也可以直接将表内容导出,然后手动导入至引擎中进行使用。手动导出支持MaxCompute、Hologres、Hive 引擎,导入导出方法请参见<mark>导入与导出表配置</mark>。

1. 发布模型。

表创建完成后,您可单击表详情页顶部菜单栏的团质和图标,在弹出的发布对话框中配置如下参数。

发布		×
发布类型 *	 ●物理表 ○ 视图 	
引擎类型 *	MaxCompute V	
引擎实例 * 生效环境 *	abc ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~	
发布模式 * 发布备注 *	● 増量发布 ○ 删除重建 申请发布	
	40/2048	
	发布。	肖

参数	描述
发布类型	可选择发布当前模型为 物理表 或 视图 。
工作空间	 可选择当前模型需要物化发布到哪个工作空间,该参数为可选参数。 已设置建模空间:此处可选择在数仓规划 > 建模空间已添加为数据研发工作空间的工作空间。 未设置建模空间:界面不显示此参数,默认物化发布到当前所在的工作空间。 ⑦ 说明 添加数据研发工作空间的操作详情请参见建模空间。
引擎类型	选择数据存储的引擎,目前可选如下几类。 MaxCompute、E-MapReduce、Hologres、CDH、MySQL。
引擎实例	根据需求将表物化至 引擎类型 参数中相应类型的数据存储引擎。
生效环境	 可选择发布至开发或生产环境。 ⑦ 说明 如果您使用的是简单模式的DataWorks工作空间,则仅支持发布表到生产环境。 如果您使用的是标准模式的DataWorks工作空间,则支持发布表到生产环境或开发环境。
发布模式	 增量发布:选择该模式,发布时仅会将目标模型此次变更的内容发布至对应引擎。 删除重建:选择该模式,发布时会将对应引擎中之前已发布的该模型删除,删除后再重新创建此次发布的模型。 ⑦ 说明 如果您的模型是首次发布,则无论选择哪种模式,系统均会在对应的引擎中新增该模型。
发布备注	描述本次发布的相关内容。

- 2. 配置完成后单击发布,系统会将模型发布至对应的开发或生产环境,并物化至所选引擎。
 - ? 说明
 - 您可在**发布**对话框查看当前发布进度。
 - 您可重复执行发布操作将模型发布至不同的环境中。
- (可选)发布完成后,您可单击表详情页顶部菜单栏的 图 专 图标,查看最近一次发布操作的发布结果以及发布的详细信息。

其中,发布状态包含发布成功、发布失败、发布中三种状态。发布成功的日志如下图所示。

发布日志			×
开发环境 生产环境			
发布状态: ❷发布成功 发布模式: 增量发布	引擎类型: MAXCOMPUTE 工作空间: DataWork:	引擎实例: abc	
ext1":null,"ext2":null,"ext3":null,"schedu ist":null,"nodeTags":null,"nodeTagsStr":nu 2022-06-13 16:49:54 INFO - search 4 times	leIntervalType":null,"triggerConfig ll}],"success":true}	g":null,"scheduleExpress":null,"extraConfig":	null,"depthL
2022-06-13 16:49:54 INFO - execute step Ge 2022-06-13 16:49:54 INFO -	tTaskInfoStep finish,result:{inter.	plan.resultMsg=success, inter.plan.resultCod	le=0}
the constant of the second			
Sarage relief.combre			
2022-06-13 16:49:54 INFO - deploy plan exe	cute success		l
			关闭

? 说明

- 当系统提示 deploy plan execute success 时,也表示发布成功。
- 当**发布状态**为发布失败时,您需要根据具体的报错信息尽快定位原因并处理。
- 当**发布日志**对话框显示**暂无发布日志** / 表示模型表从未被发布过。

克隆表

如果您已创建完成部分表,后续希望创建内容一致的表,可使用克隆表快速创建。您仅需找到待克隆的表,右 键后选择克隆表即可,表参数配置与新建表参数配置一致。

维度建模	ተፈረር
公共层	应用层
〓 ∨ 请选择业务分	类 🗸
请输入	Q
~ 🏭 天猫精灵 (2)	
> 🗄 sales (0)	
~ 🕀 客服 (1)	
~ 凸 维度表 (1)	
	a standard of the
> 🕀 用户交互 (1)	删除
> 🏭 销售业务 (0)	克隆

模型管理

创建完成维度表、明细表、汇总表、应用表后,您可以在左侧的目录树双击模型名称,查看对应模型的详情, 在模型详情也顶部菜单栏,进行模型的编辑、发布、模型开发等操作。

≡ 👘 DataWorks	组成建模	••									♂首页 ♂≸	2011年1月1日 2011年1月11日 2011年1月11月11日 2011年1月11日 2011年1月11日 2011年1月111111111111111111111111111111111	准 の 単度建模	🖉 数据指标 🗳	익, 商休 💎	
♥ 總統建構	维度建模	⊕ £ £ C	📒 dwd	×												=
♥ 数据指标	公共局	应用层	図 編輯 C 別新	♂ 关联关系 (1) 没布	日本 回馬	经重规则 已模型开发	出 导出DDL									
🖥 逆向建模	₩ > 请选择业务分析	11 ×														
	请输入	Q	基本信息													
	~ 22 天猫精灵 (2)		* 数仓分层:	公共层 維度层				~	业务分类:	天猫精灵					~	
	> 🗄 sales (0)		* 数据域:	香服 (crm)				~	存储策略:	请远择存储策略					*	
	~ ⊕ 客服 (1)		*表名:	dwd	- A.											
	~ 凸 维度表 (1)		* 表中文名:	测试												
	📒 dwo	ata ata	生命問期:	7			天 描述:								0/2048	▼显示更多
) 曲 田白衣石 (1)															

编辑模型:单击编辑即可对当前表的基本信息和字段进行修改编辑,编辑完成后单击保存即可完成修改。

⑦ 说明 编辑修改表信息后,需发布物化至引擎,才能在后续开发过程中使用最新修改后的表。

- 刷新模型信息:单击刷新可对当前变内容进行刷新,以获取最新的表信息。
- 配置模型关联关系:单击关联关系即可创建或修改当前模型的关联关系。
- **发布模型**: 单击发布即可将表发布物化至引擎。
- 查看模型发布日志:单击日志即可查看表发布日志详情。
- 生成模型的数据质量规则:模型发布生产环境成功后,单击质量规则,即可自动生成数据质量监控规则, 便于后续在生产环境中通过生成的数据质量规则来约束对应模型的数据质量。
- 生成模型代码框架:模型发布成功后,单击模型开发,即可通过模型开发功能生成代码框架,帮助您提高
 后续数据开发效率。
- **导出生成模型的DDL语句**:单击**导出DDL**即可将模型信息转化为对应引擎的建表DDL语句,便于复制或直接导出DDL语句文件。

(?) 说明 当前仅支持将模型信息转化为MaxCompute、Hologres、Hive引擎的建表DDL语句。

后续步骤

表内容发布至目标引擎后,您可以进入DataWorks的数据开发页面,基于该汇总表进行相关的数据开发操作。 数据开发的功能介绍,详情请参见数据开发界面功能介绍。

6.1.4.7. 数据落标

模型发布生产环境成功后,DataWorks智能数据建模支持自动生成数据质量监控规则,便于后续在生产环境中 通过生成的数据质量规则来约束对应表的数据质量。本文为您介绍如何在表发布成功后自动生成数据质量规 则。

前提条件

已将表发布至生产环境。表发布操作请参见发布与管理表。

注意事项

lim_t	est_02 ×															
編輯	C 刷新	♂ 关联关	系 团为	之布	國日志	€质量	规则	ピ模	型开发	む 导出DDL						
基本	信息															
数仓	分层:	公共	层 维度层	n.							~	业务	6分类:	请选择业务分	送	
数据	域:	默认	(default)								~	存储	諸策略 :	请选择存储第	휷略	
表名	:	dim_	test_02													
表中	文名:	dim_	test_导入录	Ę												
生命	周期:							天	描述:							
字段	设管理										4153164	進手	代码描录			
字段	设管理 戏已有表:	请输入表	名搜索					~			快捷	英式	代码模式			
字段 直接	Q管理 戏已有表: 別用的表:	请输入表	名搜索 us.doc.tes	t pro	od dev.moo	delingtabl	etest	*			快趣	英式	代码模式			
字段 查找 已引	设管理 戊已有表: 引用的表:	请输入表 odp	名搜索 s.doc_tes	t_pro	od_dev.mod	delingtabl	etest	*			快趣	英式	代码模式			
字段 董 已 日	设管理 成已有表: 別用的表: 字段名称	请输入表 odp	名搜索 s.doc_tes 类型	t_pro	od_dev.moo 字段显	delingtabl	etest	~ 苗述	主義	踺 非空	快趣	英式	代码模式来源字段	字段类别	关联字段标准	关联标准代码
字 序 章 引	设管理 北已有表: 13月的表: マテ段名和 id	请输入表 odp	名搜索 s.doc_tes 类型 (STRING)	t_pro	od_dev.moo 字段显 ID	delingtabl	etest	~ 苗述	±ti	键 非空 2 V	快趣 来源: modelingtal	模式 表 bletest	代码模式 来源字段 id	字段类别 属性	关联字段标准	关联标准代码
字 部 子 部 - - - - - - - - - - - - -	设管理 伐已有表: 別用的表: id name	请输入表 odp 尔 字符串 字符串	名搜索 is.doc_tes 类型 (STRING) (STRING)	t_pro	od_dev.mod 字段显 ID 名称	delingtabl	etest	~ 苗述		建 非空 2 了	快趣 来源: modelingtat modelingtat	模式 表 bletest	代码模式 来源字段 id name	字段类別 属性 ▼ 属性 ▼	关联字段标准	关联标准代码
字序 查抄 已引 1 2 3	设管理 找已有表: 引用的表: id name region	请输入表 odf 尔 字符串 字符串 字符串	各搜索 is.doc_tes 类型 i (STRING) i (STRING) i (STRING)	t_pro	od_dev.moo 字段显 ID 名称 地区	delingtabl	etest	~ 苗述		键 非空 2 · · · · 3 · · ·	快調相 来源 modelingtal modelingtal	模式 表 bletest bletest	代码模式 来源字段 Id name region	<u>字段类別</u> 属性 マ 属性 マ 属性 マ	关联字段标准 •	关联标准代码
字序 = = = = = = = = = = = = =	设管理 以已有表: 引用的表: id name region user_id	请输入表 odp 尔 字符串 字符串 字符串 字符串	各搜索 s.doc_tes 类型 (STRING) (STRING) (STRING) (STRING)	t_pro	od_dev.moo 字段显 ID 名称 地区 会员ID	delingtabl	etest #	~ 苗述		建 非空 2 2 3 1 1 1 1 1	快速 快速 modelingtal modelingtal modelingtal	模式 表 表 bioletest bioletest	代码模式 来源字段 id name region user_id	字段类别 属性 ¥ 属性 ¥ 属性 ¥ 属性 ¥ 属性 ¥	关联字段标准 * *	关联标准代码
字 部 已 5	设管理 以已有表: 引用的表: id name region user_id product_id	 请输入表 odp 字符串 字符串 字符串 字符串 字符串 字符串 	名搜索 is.doc_tes 类型 (STRING) (STRING) (STRING) (STRING) (STRING)	t_pro	od_dev.moo 字段显 ID 名称 地区 会员ID 产品ID	delingtabl	etest	~ 苗述		建 非空 2 · · · · · · · · · · · · · · · · · · ·	快速 快速 家源 modelingtal modelingtal modelingtal modelingtal	模式 表 表 Obletest Obletest	代码模式 id name region user_id product_id	字段类别 属性 ¥ 属性 ¥ 属性 ¥ 属性 ¥ 属性 ¥ 属性 ¥	关联字段标准 * * *	关联标准代码 行政区划代码

表成功发布至生产环境后,表的字段管理设置中,以下字段支持自动生成数据质量规则:

- 主键和非空字段
- 关联了标准代码的字段

对于上述支持的字段,您可参考下文操作一键生成质量规则。

操作步骤

1. 双击表名称进入表详情页面后,单击顶部的质量规则,进入数据落标页面。

	est_02 ×												
编辑	C 刷新 。	°关联关系 1	2 发布	國日志	€ 质量规则	じ 横型	开发 击	导出DDL					
基本	信息												
* 数仓:	分层:	公共层维	度层						✓ 4ks	务分类:	请选择业务分	送	
* 数据	域:	默认 (defaul	lt)						▼ 存	诸策略:	请选择存储制	6略	
* 表名:	:	dim test 02	2										
* 表中*	文名・	dim test B	、主										
4-61		uni_test_ts	/.ax				H#342.						
(生命)	刮用:					大	宙还:						
字段	设管理												
字段	2管理 8P55ま・3	また) 主々地去				v			快速模式	代码模式			
字段	2管理 800万表: 1	青输入表名搜索				*			快捷模式	代码模式			
字段 查找 已引	2管理 ^{找已有表:} 11用的表:	青输入表名搜索 odps.doc_	test_pr	od_dev.mode	elingtabletest	~			快捷模式	代码模式			
字段 查找 已5	2管理	青输入表名搜索 odps.doc_ 卷和	test_pr	od_dev.mode 字段启示	elingtabletest	~ ::	主键	非卒	快速模式	代码模式	字段类别	关联之时标准	关联标准代码
字段 查找 已引	2管理 ©CE有表: ; ;用的表: 可 字段名称 id	青输入表名搜索 odps.doc_ 类型 字符串 (STRIN	test_pr	od_dev.mode 字段显示	elingtabletest	~ :: 描述	主键	非空	快速模式 快速模式 来源表 modelingtabletest	代码模式 来源字段 id	字段类別 雇性 ▼	关联字段标准	关联标准代码
字段 査払 已引 1 2	2管理 til用的表: il用的表: id name		test_pr G) ▼ G) ▼	od_dev.mode 字段显元 ID 名称	elingtabletest	~ 描述	主键 [2]	非空	快趣模式 快速模式 来源表 modelingtabletest modelingtabletest	代码模式 来源字段 id name	字段类別 属性 * 属性 *	关联字段标准	关联标准代码
字段 査状 こち 1 2 3	R管理 む日有表: i 川田的表: id name region	輸入表名搜索 odps.doc_ 学符串 (STRIN 字符串 (STRIN 字符串 (STRIN 字符串 (STRIN	test_pr G) ▼ G) ▼ G) ▼	od_dev.mode 字段显示 ID 名称 地区	elingtabletest	~ :: 描述	主键 了 □	李菲 · · · · · · · · · · · · · · · · · · ·	快趣模式 来源表 modelingtabletest modelingtabletest	代码模式 来源字段 id name region	字段类別 属性 ▼ 属性 ▼ 属性 ▼ 属性 ▼	关联字段标准	关联标准代码
 字段 査払 己号 1 2 3 4)管理 に日有表: i に用的表: id name region user_id	輸入表名授索 odps.doc_ 学符車 (STRIN 学符車 (STRIN 学符車 (STRIN 学符車 (STRIN 学符車 (STRIN	test_pr G) ▼ G) ▼ G) ▼ G) ▼	od_dev.mode 字段显示 ID 名称 地区 会员ID	elingtabletest	~ 描述	主键 了 口	非空 又 一 一	快进模式 来源表 modelingtabletest modelingtabletest modelingtabletest	代码模式 来源字段 id name region user_id	字段类別 属性 属性 項性 属性 マ 属性	关联字段标准	关联标准代码
 字段 重批 已引 日 1 2 3 4 5)管理 に日有表: 計 川用的表: 「 す取名称 id name region user_id product_id	 第約入表名授案 odps.doc_ 学符串 (STRIN 学符串 (STRIN 学符書 (STRIN 学符書 (STRIN 字符書 (STRIN 字符書 (STRIN 	test_pr G) ▼ G) ▼ G) ▼ G) ▼ G) ▼	od_dev.mode 字段显示 ID 名称 地区 会员ID 产品ID	elingtabletest 示名	~	主键 ② □ □	非空 又 二 二 二 二	快进模式 来源表 modelingtabletest modelingtabletest modelingtabletest modelingtabletest	代码模式 来源字段 id name region user_id product_id	字段类別 属性 属性 属性 属性 属性 マーム 属性	关联字段标准	关联标准代码

2. 在数据落标页面单击生成质量规则,即可一键自动生成对应的质量规则。

如下图所示,自动生成的质量规则默认为启动状态,自动生成规则等级,您也可以单击操作列的查看详 情进入数据质量的按表配置监控规则页面,进行其他规则信息查看,也可进行试跑、启停等规则相关操 作。

dim_test_02 ×	数据落标:	dim_test_02 ×								
 dim_test_02 表中文名: dim_test, 业务过程: 默认 生命周期: 描述: 	2 普通维度表		数据 ³ 创建 ¹ 业务:	感: 默认 时间: 2022-06-01 分类:	1 16:16:26		数仓分层: 维度 最后修改人: sa	E层 antie_docte	st@test.aliyunid.co	m
1 生成质量規则 引擎: MA 引擎实例: odg 运行环境: PRG	搜索字段关键 XCOMPUTE os_first ODUCT	ja J		Q						C 刷新
质量规则名称 字段规则-非空-(id) 字段规则-表中- (region)	相关字 段 id region	规则等级 ¥ ● 弱 ● 弱	 たる 	则建人 antie_doctest@te l.com antie_doctest@te l.com	创刻 st.aliyuni 202 16:: st.aliyuni 202 16:	2006-01 22:42 12-06-01 12-2:42	最后修改人 santie_doctest@tes d.com santie_doctest@tes d.com	t.aliyuni t.aliyuni	最后修改时间 2022-06-01 16:22:42 2022-06-01 16:22:42	操作 直看详情 直看详情
 今段規则・唯一・(id) ★ 税約万同 ▲ 規則管理 へ 規則管理 へ 規則所表 ** 技規版配置 ** 技規版配置 ** (ご 任务直询 * 未確管理 (1) 	id 监控规则 , din dim_test_ 分区表达式 ● 读给入 ds=<[a-2A-20-9 ID:1552661	●弱 n_test_02 02 + - - - - - - - - - - - - - -	 启动 S d 合詞的規模 责任人: si 手动配置 代表規模 	antie_doctest@te l.com	st.allyuni 202 16: mid.com ~	226-01 22:42	santie_doctest@tes d.com 引擎/数	t.aliyuni 課題: MaxCom 契約规则	2022-06-01 16:22:42	查看详情 doc_test_prod
报告模 经管理			□ : : : : : : : : : : : : : : : : : : :	观则名称 字段规则-唯一-(id) 字段规则-非空-(id) 批量停用	ID 28585988 28585987	 > 启动 ● 启动 ● 启动 	规则字段 id id	3월/操作 • 停止 1 • 停止 1	&为别规则 日志 &为别规则 日志 《上一页 1	下-页 >

6.1.4.8. 使用FML语句配置并管理数据表

数据表创建完成后,您可以使用脚本模式通过FML语句配置表的字段及分区信息。本文为您介绍如何使用FML 语句配置数据表的相关内容。

前提条件

Dat*a*Works目前不支持通过FML语言创建数据表,您需要提前使用可视化界面创建好维度表、明细表、汇总表。

- 创建维度表,详情请参见创建维度表。
- 创建明细表, 详情请参见创建明细表。
- 创建汇总表,详情请参见创建汇总表。

背景信息

FML(Fast Modeling Language)是用于在维度建模领域快速构建数据表的一门类SQL语言,其语法采用了类

SQL的标准语法。Dat aWorks数据建模的设计与实现解耦,在设计过程中,您无需考虑各个大数据引擎的实现 方式,建模引擎会根据FML语言定义的Schema去驱动底层各个数据引擎的执行和操作,仅在实际物化(即把 设计的逻辑表转换为底层引擎的物理表)阶段,建模引擎会根据物化的选择,将FML语言转换为对应数据引擎 可识别的SQL语法,并提交任务节点执行相关语句。

使用限制

- DataWorks目前不支持通过FML语言创建数据表或修改表名称,您只能对已创建的数据表进行相关编辑操作。例如,编辑表字段、配置关联关系、配置表的分区等。
- DataWorks目前仅支持通过FML语言将设计的数据表物化至MaxCompute、Hologres及Hive引擎。
- FML语言将SQL中的关键字作为保留字,如果您的数据表或数据表的列使用了关键字命名,则需要使用 符号进行转义,否则创建表会产生报错。

进入目标表的FML编辑页面

- 1. 在维度建模页面,双击目录树下对应的表名称。
- 2. 在表详情页单击代码模式。

在代码模式页签的FML编辑区域,您可以查看当前表的建表FML语句。同时,您也可以配置或修改当前表的相关字段,详情请参见配置目标数据表。

FML FML(Fast Modeling language), 道用于律度遵循领的类SQL语言, 解前文档 1
<pre>1 不支持修改来名 2 CREATE OWS TABLE test ALIAS '测试' 3 /*(4 user_id ALIAS 'userId' BIGINT COMMENT 'userId comment', 5 user_name ALIAS 'userName' STRING COMMENT 'userName comment' 7 PARTITIONED BY 8 (9 ds ALIAS '业务日期, yyyymmdd' STRING COMMENT '业务日期, yyyymmdd' 9); </pre>

配置目标数据表

为了方便查看,所有配置表的语句均体现在创建数据表语句(CREATE TABLE)的格式中,实际上 DataWorks的数据建模不支持使用FML语句创建表。您在实际使用时,仅参考定义表内容(例如,定义约束、 分区等)的语句即可。配置目标表的FML语句格式如下。 --创建新表 CREATE <table_type> TABLE IF NOT EXISTS --表名 [ALIAS <alias>] --定义列属性 <col_name> [ALIAS <alias>] <datatype> [<category>] [COMMENT <comment>] [WITH (<key>=<va lue>,...)] --定义约束 PRIMARY KEY (<col name>), --维度约束 CONSTRAINT <constraint name> DIM KEY (<col name>) REFERENCES <ref table name> (<ref tab le col name>), --层级约束 CONSTRAINT <constraint name> LEVEL <col name:(<col name>)>, --分组约束 CONSTRAINT <constraint_name> COLUMN_GROUP(<col_name>,...), --定义备注 COMMENT 'comment' --定义分区 PARTITION BY (col DATATYPE COMMENT 'comment' WITH ('key'='value',...), ...) --定义属性 WITH ('key'='value', 'key1'='value1', ...) ; tableType : dimDetailType? DIM | factDetailType? FACT | CODE | DWS ; dimDetailType : NORMAL | LEVEL | ENUM ; factDetailType : TRANSACTION | AGGREGATE | PERIODIC SNAPSHOT | ACCUMULATING SNAPSHOT | CONSOLIDATED ; comment : COMMENT 'comment' ;

参数	描述
tableName	数据表的名称。推荐命名由英文字母、数字及下划线(_)组成,长度为128个字符。

参数	描述
<i>x</i>	当您的目标引擎中已存在同名表,如果不指定 if not exists 时,则创建表会报 错,而指定了 if not exists 后,无论引擎中是否存在同名表,则创建表均会成 功。
if not exists	⑦ 说明 当您的引擎中已经存在同步表时,指定了 if not exists 后,创建的新表会替换原有的同名表。
alias	数据表或数据表中列的别名,通常用于定义数据表的中文名称。该参数为可选参数。
tableType	 创建的表所属的类型。目前仅支持使用FML语句创建如下类型的表: 维度表 普通维度表(NORMAL):创建维度表时指定的默认类型。 层级维度表(LEVEL):用于存放例如省、市、区等具有层级关系的数据。 枚举维度表(ENUM):用于存放常用的可枚举的值。例如男、女性别等。 事实表 事实表 事实表,保存最为原子的数据。 周期快照表(PERIODIC_SNAPSHOT):保存具有规律性、可预见时间的事实记录,它统计的是间隔周期内的度量统计,例如,历史至今、自然年至今等。其采用增量更新等方式更新表数据。 累加表(ACCUMULATING_SNAPSHOT):存储事务数据的快照信息,用于积累不确定周期的数据,例如,订单记录快照事实表可包含付款日期,发货日期和收货日期时间点。 DWS表:用于合并特定指标的逻辑汇总表,表的定义语法与维度表和事实表类似。 码表:又称为标准代码,是一种包含行业特定属性的代码表。例如,水电行业,可能会涉及创建是否有合同,供电合同类型等标准代码。
comment	表的备注信息。推荐备注信息的长度为1024字符。
columnDefinition	数据表列的定义。可以包含如下参数: • col_name:数据表的列名。由英文字母、数字、下划线(_)等字符组成。如果列名 使用了FML中的关键字,则需要使用 、符号进行转义。 • alias:数据表中列的别名,通常用于定义数据表中列的中文名称。该参数为可选参 数。 • dataType:目前FML的数据类型包含了BIGINT、SRRING、VARCHAR、CHAR、 DECIMAL、DATETIME等。 • category:用于对数据表的列进行分类。在维度建模中,列可以分为业务属性 (ATTRIBUTE)、度量(MEASUREMENT)、关联字段(CORRELATION)类型。 • ③ 说明 FML语句支持先设计后物化数据表,因此新创建的表可以不设置列信 息。

参数	描述
constraint	 定义表结构的约束情况,取值如下: 主键约束(PrimaryConstraint):格式为 PRIMARY KEY(col1, col2),其中 col1, col2 字段列表必须是已经定义过的列表。 维度约束(DimConstraint):格式为 DIM KEY(col1, col2) REFERENCES tablename(ref1, ref2), 。 层级约束(LevelConstraint):仅在层级维度中生效。用于定义层级维度中的各个层级。
Partitioned BY	用于创建表的分区。
WITH	您在创建表时可以自定义的信息,采用 key=value 的方式, key 和 value 的 值需要使用单引号引起来,防止自定义的信息与FML的关键字冲突。 WITH 中的扩展属 性,可以由物化FML语句的引擎来进行解析处理。

6.1.4.9. 导入与导出表配置

您可将本地表通过Excel的方式导入至DataWorks数据建模中,或将数据建模中的表导出至本地数据仓库中。 本文将为您介绍导入与导出表的操作方法。

导入表

将本地表以Excel的方式导入至数据建模中,导入方法如下。

- 1. 在维度建模页面,单击 🖌 图标,选择Excel导入。
- 2. 在弹出的导入对话框,选择导入模式并上传需要导入的文件,完成后单击确定。

? 说明	如您不知道提供什么类型的文件,	可单击下载模板查看系统为您提供的模板内容进行创
建。		

导入		×
导入模式:	 跳过已存在的同名对象 覆盖已存在的同名对象 	
文件上传:	▲ 选择文件 下载模版	
		取消 确定

3. 系统导入数据完成后,会提示您成功导入的数量和导入失败的数量。您可单击**完成**或**重新导入**。在左侧 目录树查看导入进来的表。

导入				×
		6) 空成		
		总计1条		
	成功: 1条		失败: 0条	
		查看详情		
			重新导入	完成

导出表配置

将已创建成功的表,导出成SQL文件或命令代码。操作方法如下。



- 1. 在维度建模页面,双击对应目录树中表的名称。
- 2. 在右侧表详情页,单击导出DDL并选择对应引擎类型(MaxCompute、Hologres、Hive)。
- 3. 在弹出来的DDL导出对话框中,将显示创建表的语句代码。您可单击**导出文件**生成SQL文件使用,或单击**复制**,然后在对应引擎中执行命令进行操作。

MaxO	Compute DDL导出	×
1 2 3 4 5 6	CREATE TABLE IF NOT EXISTS dim exam (ceshi STRING COMMENT asd STRING COMMENT) COMMENT LIFECYCLE 365;	
	与出文件	制

6.1.4.10. 附录: 使用画板工具创建表

本视频将为您介绍DataWorks数据建模画板功能的使用方法。

6.1.4.11. 逆向建模

当您已有MaxCompute中的大量物理表,且希望能通过DataWorks智能数据建模产品统一管理所有模型,则可 以使用逆向建模功能。该功能无需您再次执行建模操作,即可帮助您快速将已有物理表反向建模至DataWorks 的维度建模中,节省了大量的时间成本。本文为您介绍如何进行逆向建模。

前提条件

- MaxCompute中已有大量的物理表,同时,MaxCompute引擎已绑定至目标DataWorks工作空间。绑定引擎 至DataWorks工作空间,详情请参见配置工作空间。
- 已创建数据域,用于确定数据建模所属的数据域,详情请参见数据域。
- 已创建业务过程,用于确定数据建模所属的业务过程,详情请参见业务过程。
- 已创建业务分类,用于确定数据建模所属的业务分类,详情请参见业务分类。

使用限制

当前仅支持对MaxCompute引擎中生产环境的表进行逆向建模。

建模流程

逆向建模主要用于将大数据引擎中已经存在的物理表反向建模至DataWorks的维度建模中,其建模流程如下:



1. 配置逆向建模策略。

- 建模范围:创建模型前,您需要根据实际业务需求确定需要逆向建模的表范围,即表所在的工作空间、精确的表名称或表名称包含的关键字。
- 建模匹配规则: 您需要根据实际业务需求确定新创建的模型及表名称的匹配规则。
- 建模执行方式:确定建模时是全量新建目标表,还是仅创建维度建模中不存在的目标表。

⑦ 说明 逆向建模为不可逆操作,创建逆向策略并使用该策略生成模型后,该策略则不可被更改, 请您根据实际业务需求提前做好策略规划。

详情请参见配置逆向建模策略。

2. 匹配模型。

系统会根据您配置的建模策略匹配模型。

3. 确认模型信息。

系统初步生成的表模型可能会存在偏差,您需要根据具体业务需求进行模型基础信息的调整。例如,更改 表所在的**数据域、业务过程**等模型基础信息。详情请参见<mark>确认模型信息</mark>。

4. 生成最终模型。

建模完成后,您可以查看建模成功的表类型及数量,以及建模失败的表的错误信息,快速定位问题并及时 处理。

? 说明

- 建模成功的表已被物化至相应引擎,您无需再进行发布。
- 建模成功的表会被存放至维度建模中,您可以进入维度建模页面查看,并进行后续的管理操作,详情请参见管理表。

操作步骤

- 1. 进入逆向建模。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 选择工作空间所在地域后,单击相应工作空间后的数据开发。
 - ⅳ. 单击左上方的■图标,选择全部产品 > 数据建模 > 维度建模,默认进入维度建模页面。
 - v. 在维度建模页面的左侧导航栏, 单击逆向建模, 进入逆向建模页面。
- 2. 启动建模。
 - 首次使用逆向建模功能,则请单击逆向建模界面中间的快速开始,启动建模。
 - 非首次使用逆向建模功能,则请单击建模列表右上角的开始逆向建模,启动建模。
- 3. 配置逆向策略。

⑦ 说明 逆向建模为不可逆操作,创建逆向策略并使用该策略生成模型后,该策略则不可被更改, 请您根据实际业务需求提前做好策略规划。

i. 配置策略信息。

= 🍿 DataWorks	維度建模 标准	🔂 の首页 の数合规3	別 の数据标准 の维度建模	の数据指标	白 & 简体 👇 🗕 🚽
♥ 維度建模♥ 数据常标	建模列表, 逆向建模 ← 逆向建模				
12 逆向建模					
		1 逆向策略	2 确认模型信息	3 生成模型	
		表所在项目: * 引擎安型 * 引擎实例 * 表名匹配规则: ● 表命名规范: 执行方式:	MaxCompute ④ 慎樂匹配 ○ 時者匹配 斎給入来名关键字 业务过程 > 」 武編城 ④ 全量聚重 ○ 增量更新 予始的法律监	▼_ 业务分类 ▼ 0 ⊙	v v v
参数	描述				
	选择	需要逆向建模的	n表所在的DataW	/orks工作空间。	
表所在项目	0	说明			
引擎类型	当前个	又支持逆向建模	「MaxCompute引	擎中生产项目的]物理表。
引擎实例	选择	需要逆向建模的	的表所在的引擎实	:例。	
表名匹配规	支持f 模 標 精 输入 和 名称,	的匹配方式如下 脚匹配:输入: 个工作 一 一 一 一 一 一 一 一 一 一 一 一 一 一 一 一 一 一 一	关键词,即可匹置的表名称为表的表。 的表名称为表的表。 一进行匹配,具体 同的表时,则会相 建模中。 目同的表时,则这 表名称之间使用。	配到包含该关键说 情准全称。 会称包含的关键 如下: 包据配置的执行方 在向建模最终会失 英文分号(;)分	司的所有表名称。 字。系统会根据您输入的表 可式对该表进行逆向建模,创 变败,不会创建该表。 隔,并且英文分号(;)后
表命名规范	用于 建模 ● 解 ● 表 程 〇	₩ 析 表名匹配 新 表 挂 载 至 奇 新 新 新 最 多 の 可 以 信 、 数 据 城 、 业 、 、 、 、 、 、 、 、 、 、 、 、 、	见则 中输入的表行 含名规范所配置的 则中输入的表名: 回含9个下划线, 务分类、自定义 回含对应数仓层级 别到表名称中不 最为空。您可以在	名称,来规范逆向 对应数仓层级下 称和下划线个数。 每两个下划线之 内容等信息。 及时,则会将该表 存在对应的数据。 确认模型信息B	回建模后的表名称,并将逆向 。解析规则如下: 间可以选择配置为 业务过 挂载至对应的层级下。 域、 业务过程 等数仓层级 时修改相应表所属的分层。

参数	描述
	全量覆盖:系统会将所有匹配到的表均逆向创建至DataWorks的维度建模中。 当您需要全量重新建模目标表时,可以选择该方式。
	⑦ 说明 如果维度建模中已存在所匹配的目标表,选择全量覆盖时会将原有已存在的表删除重建。
	增量更新:系统会对所有匹配到的表进行如下处理:
执行方式	a. 识别DataWorks维度建模中是否已创建所匹配的表,将已创建的表过滤 掉。
	b. 将DataWorks维度建模中未创建的表逆向建模至维度建模中。
	当您已知维度建模中已创建部分目标表,并且该表未改动,不希望重复创建 时,可以选择该方式。
	⑦ 说明 如果表名匹配规则中输入的表在目标工作空间中未匹配到,则 逆向建模会失败。

ii. 单击开始创建模型,通过建模策略进行模型解析。

- 4. 确认模型信息。
 - i. 确认模型信息。

系统根据您配置的建模策略生成初步表模型,您可以结合实际业务需求,调整表的类型,以及所属的数仓分层、业务分类、数据域、业务过程等信息。同时,当目标表无需建模时,您可以选择将其删除。

建模列3 (←))	■ · 逆向建模任务20211214 逆向建模任务20211214										
		ジージョン		2 确认模型信息	3 生成模型		4 元成				
				◎ 已解	忻1/1						
8 2	《情报醒:关闭当前页面不影响模型解析,模型解析信息可在;	兰向建模首页任务卡片 处查看并继续进	5逆向建模操作。								×
解析	列表										
	表名	表显示名	*表类型	*数仓分层	*业务过程	*数据域	业务分类	存储策略	负责人	备注	操作
	aligan (daga daga daga daga daga daga daga dag	原始用户表	事实事务表 >	明细数据層 (DWD) >	通過择 >	-	请选择业务分类 >	请选择存储策略 >			删除

ii. 单击开始生成模型, 生成模型。

5. 查看建模结果。

模型生成后,您可以查看本次建模操作成功创建的**明细表、维度表、汇总表**的数量,以及建模失败的表的详细信息,单击**错误日志**,即可快速定位问题原因并及时处理。

? 说明

- 建模成功的表已被物化至相应引擎,您无需再进行发布。
- 建模成功的表会被存放至维度建模中,您可以进入维度建模页面查看,并进行后续的管理操作,详情请参见管理表。

		 一 逆向策略 	确认模型信息	(~ 生成模型	4 完成		
			模型生	成完成			
		ž	本次模型 構成功: 32个 建模失数: 3个 明细表	主成结果 : 4张 / 维度表: 28张 / 汇总表: 0强	ž		
序号	表名		表显示名		表类型	状态	操作
1	dimingen		1298		维度表	🛽 失敗	错误日志
2	dim_site		854		维度表	😆 失敗	错误日志
3	dim_limil_mp2		499,160,892		维度表	🙁 失败	错误日志
3	aim es apri		911,141,142		11.01.70	● 失敗	

查看建模列表

在逆向建模 > 建模列表页面, 您可以查看已创建任务的详细信息及操作日志。

●数合规划 ●数振标准 ● 塩度速模 ●数振指标 🔮 🔍 💎 🚃 🛁 n Da 徐庆建模 建模列表 🖸 逆向建模 任务ID: 油榆) 操作人: 诸法将 > 操作日期: 波法探日期 1 【tk-a3ea939ca6f84dd6ae4d1e59042dd570】 逆向建模任务20211104 日志査君 逆向规则:精准匹配:tes 2 建模成功: 0个 / 建模失败: 0个 / 明细表: 0张 / 维度表: 0张 / 汇总表: 0张 【tk-16a6dd3e74fc44f5b467efd081ab525d】逆向建横任务20211101 任务查看 兰向规则: 積准匹配:* 021-11-01 12-22-42 884 【tk-38b8487c9cdf404fba5f0d26c130228a】逆向建模任务20211029 任务查看 送向规则: 精准匹配:* 【tk-8b2c21cd58724942a5b4fe96ba107b75】 逆向建模任务20211029 日志宣看 兰向规则: 稿准匹配:* 建模成功:0个 建模失败:0个 明细表:0张 维度表:0张 汇总表:0张 く 上一页 1 下一页 >

区域	描述
1	在该区域,您可以通过 任务ID、操作人、操作日期 进行筛选,查看指定条件的逆向建模任务。
2	在该区域,您可以查看逆向建模任务的 逆向规则 、建模结果等详情。 对于已执行的建模任务,单击日志查看即可查看任务的日志信息。 对于未执行的建模任务,单击任务查看即可返回任务详情页继续执行任务。

后续步骤

逆向建模完成后,您可以进入维度建模页面的模型管理目录树查看创建的模型,详情请参见模型管理。同时, 建模成功的表默认已被发布至目标引擎,您可以进入DataWorks的数据开发页面进行相关的数据开发操作。数 据开发的功能介绍,详情请参见数据开发功能索引。

6.1.4.12. 模型开发

维度表、明细表、汇总表等模型表发布后,DataWorks智能数据建模的模型开发功能支持快速生成代码框架, 帮助您提高后续数据开发效率。本位为您介绍如何通过模型开发功能生成代码框架。

前提条件

模型表已发布,操作详情请参见发布与管理表。

使用限制

当前模型开发功能仅支持对发布至MaxCompute引擎的表生成代码框架,发布至其他引擎的场景暂不支持。

发布		×
生效环境 *	✓ 开发	
引擎类型 *	MaxCompute	~
引擎实例 *	doc_test	~
发布模式 *	● 増量发布 ○ 删除重建	
发布备注 *	dim_test_02(dim_test_导入表)申请发布	
		29/2048
		发布取消

操作步骤

1. 双击表名称进入表详情页面后,单击顶部的**模型开发**,进入数据开发(DataStudio)的创建节点配置页面。

📄 dim_t	est_02 ×								
区编辑	C 刷新	♂ 关联关系	☑ 发布	國日志	🖻 质量规则	ビ 模型开发	击 导出DDL		
基本	信息								
* 数仓	分层:	公共层	维度层			~	业务分类:	请选择业务分类	*
* 数据	域:	默认 (def	ault)			~	存储策略:	请选择存储策略	~
* 表名	:	dim_test	_02						
* 表中	文名:	dim_test	导入表						
生命	周期:			天	描述:				0/2048

2. 根据界面提示,确认并配置节点参数。

新建节点		×
最近使用:	ODPS SQL	
* 引擎实例:	doc_test 华东2(上海)	
*节点类型:	ODPS SQL	
* 路径:	业务流程/Data_Modeling	
	上次: 业务流程/Data_Modeling	
* 名称:	dim_test_02	
		í

⑦ 说明 发布至MaxCompute引擎的表,通过模型开发功能自动生成代码框架时,默认会创建为 ODPS SQL类型的节点,并根据表名称自动生成节点名称。

3. 单击提交,完成节点的创建。

成功创建节点后, DataWorks会根据模型表的信息为您自动生成部分代码框架, 后续您可基于此代码框架 快速开展数据开发的工作。

	dim_test_02 ×									
Sq dim_test_02 ×	防治堤 CBM	₽¥₩¥系 (1%8	5 月日志 (向 所暴招の	【2/検型开	お日田の日				
	基本信息									
2	* 数仓分层:	公共屋 進度层			~	业务分类:	请选择	出务分类		~
3create time:2022-06-01 19:01:37	* 数据域:	默认 (default)			~	存储策略:	调选择	存储策略		~
4***********************************	* 责名:	dim test 02								
5 模型名称: dim_test_02(dim_test_导入表)	* 要由文名·	dim tort B) #								
6数据域:默认; 业务过程:默认;	4.0.000	CIIII (163C) 497(20								
7模型描述:	生动周期:		大	趟还:						0/2048
8										
10 SELECT t1.id AS idIO 11 ,t1.name AS name名称 12 .t1.region AS region地区	┃ 字段管理				铁	建模式 代码	夏式			
13 ,t1.user_id AS user_id会员ID	查找已有表:	请输入表名搜索			~					
14 ,t3.user_id_01 AS user_id_01关联表会员ID										
15 _,t1.product_id AS product_id产品ID	字段名称	类型	字段显示名	描述	主键 非空	来源表	来源字段	字段类别	关联字段标准	关联标准代码
16 ,t1.type AS type产品类型	1 id	字符串 (STRING) 🔻	D		•	modelingtabletest	id	尾住 ・	•	•
17	2 name	字符串 (STRING) ▼	名称			modelingtabletest	name	属住 ・	,	
18 FROM doc_test_prod_dev.modelingtabletest t1	3 region	字符串 (STRING) 🔻	地区			modelingtabletest	region	属性 ・	,	行政区划代码
19 LEFT JOIN (SELECT user_id_01	4 user_id	字符串 (STRING) 『	会员ID			modelingtabletest	user_id	尾性 ・	,	•
20 FROM doc_test_prod.dim_test	5 user_id_01	字符串 (STRING) ▼	关联表会员ID			dim_test	user_id_01	属住 ・	•	
<pre>21) t3 ON t1.user_id = t3.\${self_define}</pre>	6 product_id	字符串 (STRING) ▼	产品ID			modelingtabletest	product_id	尾性 ・	,	,
22	7 type	字符串 (STRING) 『	产品类型			modelingtabletest	type	尾性 *	•	•
23	分区字段管理									
25	4 4	的名称	类型		字段显示名	描述	*	主键	非空	
	1 ds		序符串 (STRING)	▼ 11/8	6日期, yyyymm	56 业务日期, yyy	ymmdd			
									_	

后续步骤

后续您可基于自动生成的代码框架进行代码开发,继续完成数据开发工作。

6.1.5. 数据指标

6.1.5.1. 数据指标概述

DataWorks的数据建模提供数据指标功能,为您提供统一的指标体系建立能力。

指标体系

指标是衡量业务特征的统计数值,用于体现企业某一业务活动的业务状况。指标通常分为原子指标和派生指标,与指标相关的参数还包括修饰词和时间周期。派生指标则是由 原子指标 + 时间周期 + 一个或多个修饰词 组成。

				修饰词1	
原子指标	+	时间周期	+	修饰词2	派生指标
				修饰词3	

- 原子指标用于明确业务的统计口径及计算逻辑。例如,下单总数。
- 修饰词用于限定业务活动的范围。例如,线上生鲜门店、线下生鲜门店。

⑦ 说明 一个派生指标中可以包含一个或多个修饰词。

- 时间周期用于确定需要统计的时间范围。例如,一个自然日。
- 派生指标由原子指标、修饰词、时间周期三大要素构成,用于统计目标指标在具体时间、维度、业务条件下的数值表现,反映企业某一业务活动的业务状况。例如,统计一个自然日中线上生鲜门店的下单总数、

一个自然日中线上和线下生鲜门店的下单总数 。

- 创建原子指标,详情请参见原子指标。
- 创建修饰词,详情请参见修饰词。
- 创建时间周期, 详情请参见时间周期。
- 创建派生指标,详情请参见派生指标。

6.1.5.2. 原子指标

原子指标用于明确业务的统计口径和计算逻辑,是基于用户的业务活动(即业务过程)创建的,用于统计业务 活动中某一业务状况的数值。例如,用户的业务活动为购买,则原子指标就可以指定为支付金额。本文为您介 绍如何创建原子指标。

前提条件

已创建业务过程,用于明确需要统计分析的具体业务活动。创建业务过程,详情请参见业务过程。

背景信息

指标是衡量业务特征的统计数值,用于体现企业某一业务活动的业务状况。指标通常分为原子指标和派生指标,与指标相关的参数还包括修饰词和时间周期。派生指标则是由 原子指标 + 时间周期 + 一个或多个修饰词 组成。



- 原子指标用于明确业务的统计口径及计算逻辑。例如,下单总数。
- 修饰词用于限定业务活动的范围。例如,线上生鲜门店、线下生鲜门店。

⑦ 说明 一个派生指标中可以包含一个或多个修饰词。

- 时间周期用于确定需要统计的时间范围。例如,一个自然日。
- 派生指标由原子指标、修饰词、时间周期三大要素构成,用于统计目标指标在具体时间、维度、业务条件下的数值表现,反映企业某一业务活动的业务状况。例如,统计 一个自然日中线上生鲜门店的下单总数 、 一个自然日中线上和线下生鲜门店的下单总数 。

创建原子指标

- 1. 进入数据建模。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 选择工作空间所在地域后,单击相应工作空间后的进入数据开发。
 - iv. 单击左上方的**三**图标,选择**全部产品 > 数据建模 > 数据指标**,默认进入派生指标页面。
- 2. 在数据指标页面的左侧导航栏,单击原子指标,进入原子指标页面。

3. 创建原子指标。

i. 在**原子指标**页面, 单击 ⊞图标新建原子指标。

4. 配置原子指标信息。

i. 配置原子指标基本信息。

🙏 新建原子指标	示 ×		
圖 保存			
原子指标	减基本信息 -		
业务过程:	*	请选择业务过程 🖌	
英文缩写:	*	请输入英文缩写	
		▲ 英文缩写是指标的唯一性标识,一经保存,无法二次修改。	
英文名称:		请输入英文名称	
中文名称:	*	请输入中文名称	
业务口径:	*		
		输入@可以引用其他原子指标	
描述:			
			0/1024
参数	描述		

参数	描述
业务过程	原子指标所属的业务过程。
英文缩写	原子指标的英文缩写。英文缩写是指标的唯一性标识,创建后无法修改。 英文缩写由小写英文字母、数字、下划线(_)组成,并且以小写英文字母开头。
英文名称	原子指标的英文名称。 建议使用原子指标所统计数值的英文,便于您快速了解该原子指标的统计类型。 英文名称由英文字母、数字、下划线(_)、and(&)组成,并且以英文字母或数字开头。
中文名称	原子指标的中文名称。 建议使用原子指标所统计数值的中文,便于您快速了解该原子指标的统计类型。 中文名称由中文、英文字母、数字、下划线(_)、and(&)、英文括号组成,并且以中文、英 文字母或数字开头。

参数	描述
业务口径	用于统一原子指标所统计的业务活动的统计口径,防止统计时产生歧义。 例如, <i>下单实际支付金额</i> 的业务口径为: 用户下单生成订单后,通过支付渠道支付了订单金额 的总和。该金额为扣除掉所有红包优惠后,用户实际支付的金额总和。
描述	原子指标的描述信息。

ii. 配置原子指标的计算逻辑。

计算逻辑仅用于定义目标原子指标在后续使用时的计算方式及单位。在实际使用过程中,您需要根据 此处定义的计算逻辑,编写计算代码,实现真实的计算操作。

计算逻辑	
计算函数:	请选择计算函数 🖌
小数位数:	0
数据单位	请选择数据单位 🗸
是否去重:	○ 是 ● 否
参数	描述
计算函数	 累加(SUM) 计数(COUNT) 求平均(AVG) 最大值(MAX) 最小值(MIN) 率 排名 其他
小数位数	根据您的度量单位,确定小数位数。例如,统计支付金额时,度量单位为元,则小数位数为2。
数据单位	基于原子指标的统计数据类型,选择合适的度量单位。例如,统计支付金额,则可以选择 货币 单位 。如果已创建的单位类型中没有适合您业务使用的,则您可以进入度量单位创建,详情请 参见 <mark>度量单位</mark> 。
是否去重	您需要根据业务需求进行判断是否在统计时需要去重。例如,原子指标为会员数,则统计时需要 进行去重,不能进行单纯的累加。

5. 单击**保存**,完成创建。

导入原子指标

如果您有大量原子指标待创建,则可以选择批量导入原子指标。DataWorks为您提供了导入模板,您可按照模 板表格填写好待导入的原子指标后批量导入。

1. 进入原子指标页面, 鼠标悬停至 函图标, 单击Excel导入。

进入原子指标页面可参考创建原子指标的相关步骤。

2. 在导入对话框中选择导入模式,下载导入模板并批量上传。

导入				×
导入模式: 文件上传:	 跳过已和 上选择如 下载模版 	存在的同名对象 1		
			取消	确定
业务过程(英文编写)	英文缩写	中文名称	英文名称	计算函数
default default	atom_1 atom_2	name_1 name_2	english name 1 english name 2	SUM 3

- i. 在原子指标的导入对话框中,选择导入模式为跳过已存在的同名对象。
 当前仅支持使用跳过已存在的同名对象模式导入。
- ii. 单击下载模板下载批量导入原子指标的模板表格。
- iii. 在模板表格中,填写并保存待导入的各原子指标的业务过程、英文缩写、中文名称等信息。各参数的格式要求如下。

参数	描述
业务过程	原子指标所属的业务过程。此处需填写业务过程的英文缩写。
英文缩写	原子指标的英文缩写。英文缩写是指标的唯一性标识,创建后无法修改。 英文缩写由小写英文字母、数字、下划线(_)组成,并且以小写英文字母开头。
中文名称	原子指标的中文名称。 建议使用原子指标所统计数值的中文,便于您快速了解该原子指标的统计类型。 中文名称由中文、英文字母、数字、下划线(_)、and(&)、英文括号组成,并且以中文、英 文字母或数字开头。
英文名称	原子指标的英文名称。 建议使用原子指标所统计数值的英文,便于您快速了解该原子指标的统计类型。 英文名称由英文字母、数字、下划线(_)、and(&)组成,并且以英文字母或数字开头。

参数	描述
计算函数	 累加(SUM) 计数(COUNT) 求平均(AVG) 最大值(MAX) 最小值(MIN) 率 排名 其他
小数位数	根据您的度量单位,确定小数位数。例如,统计支付金额时,度量单位为元,则小数位数为2。
数据单位	基于原子指标的统计数据类型,选择合适的度量单位。例如,统计支付金额,则可以选择 货币 单位 。如果已创建的单位类型中没有适合您业务使用的,则您可以进入度量单位创建,详情请 参见 <mark>度量单位</mark> 。
是否去重	您需要根据业务需求进行判断是否在统计时需要去重。例如,原子指标为会员数,则统计时需要 进行去重,不能进行单纯的累加。
业务口径	用于统一原子指标所统计的业务活动的统计口径,防止统计时产生歧义。 例如, <i>下单实际支付金额</i> 的业务口径为: 用户下单生成订单后,通过支付渠道支付了订单金额 的总和。该金额为扣除掉所有红包优惠后,用户实际支付的金额总和。
描述	原子指标的描述信息。
负责人账 号ID	创建原子指标的用户ID。

⑦ 说明 下载模板表格后,如果一开始无法编辑表格,可单击表格顶部的启用编辑按钮。

iv. 单击选择文件,选择上述保存的原子指标模板后单击确定。

后续您可以在导入结果的提示对话框中查看导入是否成功及导入详情。

47/				^
		导入完成 ⁽²⁾ 計系		
	成功:1条	0.251745	失敗: 0条	
		10.000		

后续步骤

创建的原子指标后续可以被派生指标引用,用于统计某一业务活动统计目标指标在具体时间、维度、业务条件

下的数值表现。如果原子指标已经被派生指标引用,则您需要删除所有关联的派生指标后,才可以删除原子指标。派生指标引用原子指标,详情请参见<u>派生指标</u>。

6.1.5.3. 修饰词

修饰词是基于选择的数据域创建的,是对数据域中的数据不同维度的修饰,用来限定统计数据的业务范围。例如,统计上海区域生鲜门店的销售金额,上海区域和生鲜门店就是对业务范围的限定修饰。

前提条件

已创建数据域,用于确定修饰词所属的数据范围。创建数据域,详情请参见数据域。

背景信息

修饰词是一种业务修饰,用来圈定或者聚焦统计数据的业务范围和限定。例如,统计线上渠道的销售金额、生 鲜门店的销售金额, 线上渠道 和 生鲜门店 就是聚焦的业务范围和限定。一个或多个修饰词与原子指标、 时间周期,共同构成一个派生指标,用于体现企业某一业务活动的业务状况。派生指标的相关内容,详情请参见派生指标。

创建修饰词

- 1. 进入数据建模。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 选择工作空间所在地域后,单击相应工作空间后的进入数据开发。
 - iv. 单击左上方的**三**图标,选择**全部产品 > 数据建模 > 数据指标**,默认进入派生指标页面。
- 2. 在数据指标页面的左侧导航栏,单击修饰词,进入修饰词页面。

3. 创建修饰词。

- i. 在修饰词页面, 单击 册图标新建修饰词。
- ii. 配置修饰词基本信息。

🔑 新建修饰词	×	≡
圖 保存		
修饰词基本	· · · · · · · · · · · · · · · · · · ·	
修炼词关刑。		
☆ 御戸 御天里 :		
22/8-37	日期/35×234-3	
苗文夕称:		
中文名称: *	時間のなくにする	
业务□径: *	諸範入业务口径	
	0/1024	
描述:		
	0/1024	
へ隐藏高级配		
数仓分层:		
朱巾/土迦:	请边华 ~	38
参数	描述	
修饰词类	■ 普通业务修饰词	
型	■ 维度枚举修饰词	
英文缩写	英文缩写是修饰词的唯一性标识,创建后无法修改。例如,shop_type_fresh	0
	英文缩写由小写英文字母、数字、下划线(_)组成,并且以小写英文字母开到	F°
	修饰词的革文名称 例如 shon type of fresh	
英文名称		武教守工义
		现数于开关。
	修饰词的中文名称。例如,生鲜门店。	
中文名称	中文名称由中文、英文字母、数字、下划线(_)、and(&)、英文括号组成	,并且由中文、英
	文字母或数字开头。	
业务口径	用于限定对目标数据阶统计的业务沽动范围。例如,生鲜类型等售实门店。	
描述	修饰词的描述信息。	
	选择修饰词所属的数仓分层。	
新るシロ	■ 应用数据层	
剱で方法	■ 汇总数据层	
	您需要单击 高级配置 ,才可配置当前参数。	

参数	描述
集市/主 题	当数仓分层配置为应用数据层时,用于选择修饰词所属的集市或主题。
数据域	当 数仓分层 配置为 汇总数据层 时,用于选择修饰词所属的数据域。

4. 单击保存,完成创建。

导入修饰词

如果您有大量修饰词待创建,则可以选择批量导入修饰词。DataWorks为您提供了导入模板,您可按照模板表 格填写好待导入的修饰词后批量导入。

1. 进入修饰词页面, 鼠标悬停至 3. 图标, 单击Excel导入。

进入修饰词页面可参考创建修饰词的相关步骤。

2. 在导入对话框中选择导入模式,下载导入模板并批量上传。

导入						×
导入模 文件上	式: 传: 下载	跳过已存在的同名对 L选择文件 4 糢版 2	€ 1			
				取消	〕 确定	
<u>,</u> 重在分层	● 奏市/主题(英文編写第号	c まま数(交定都写) default default	; 修饰词英型	天文编写 adj_1 adj_2	desc_1 desc_2	3

- i. 在修饰词的导入对话框中,选择导入模式为跳过已存在的同名对象。
 当前仅支持使用跳过已存在的同名对象模式导入。
- ii. 单击下载模板下载批量导入修饰词的模板表格。

iii. 在模板表格中,填写并保存待导入的各修饰词的英文缩写、中文名称、业务口径等信息。各参数的格式要求如下。

参数	描述
修饰词类 型	 ■ 普通业务修饰词 ■ 维度枚举修饰词
英文缩写	修饰词的英文缩写。 英文缩写是修饰词的唯一性标识,创建后无法修改。例如,shop_type_fresh。 英文缩写由小写英文字母、数字、下划线(_)组成,并且以小写英文字母开头。
英文名称	修饰词的英文名称。例如,shop type of fresh。 英文名称由英文字母、数字、下划线(_)、and(&)组成,并且以英文字母或数字开头。
中文名称	修饰词的中文名称。例如,生鲜门店。 中文名称由中文、英文字母、数字、下划线(_)、and(&)、英文括号组成,并且由中文、英 文字母或数字开头。
业务口径	用于限定对目标数据所统计的业务活动范围。例如,生鲜类型等售卖门店。
描述	修饰词的描述信息。
数仓分层	选择修饰词所属的数仓分层。 ■ 应用数据层 ■ 汇总数据层 您需要单击 高级配置 ,才可配置当前参数。
集市/主 题	当数仓分层配置为应用数据层时,用于选择修饰词所属的集市或主题。
数据域	当 数仓分层 配置为 汇总数据层 时,用于选择修饰词所属的数据域。

⑦ 说明 下载模板表格后,如果一开始无法编辑表格,可单击表格顶部的启用编辑按钮。

iv. 单击选择文件,选择上述保存的修饰词模板后单击确定。

后续您可以在导入结果的提示对话框中查看导入是否成功及导入详情。

	导入完成		
成功: 1条	总计1条 查看详情	失败; 0条	

后续步骤

创建的修饰词与原子指标、时间周期后续可以被派生指标引用,用于统计某一业务活动统计目标指标在具体时间、维度、业务条件下的数值表现。如果修饰词已经被派生指标引用,则您需要删除所有关联的派生指标后, 才可以删除修饰词。派生指标引用修饰词,详情请参见派生指标。

6.1.5.4. 时间周期

时间周期是用来明确数据统计的时间范围或者时间窗口,例如近1天,近1自然周。用于在统计派生指标时, 限定业务统计的时间范围。本文为您介绍如何创建时间周期。

背景信息

时间周期与原子指标、修饰词,共同构成一个派生指标,用于体现企业某一业务活动的业务状况。派生指标的 相关内容,详情请参见<mark>派生指标</mark>。

创建时间周期

1. 进入数据建模。

- i. 登录DataWorks控制台。
- ii. 在左侧导航栏, 单击工作空间列表。
- iii. 选择工作空间所在地域后,单击相应工作空间后的进入数据开发。
- iv. 单击左上方的**三**图标,选择**全部产品 > 数据建模 > 数据指标**,默认进入**派生指标**页面。
- 2. 在数据指标页面的左侧导航栏,单击时间周期,进入时间周期页面。

3. 创建时间周期。

i. 单击 ⊞图标, 新建细粒度的时间周期。

DataWorks默认为您创建了天、自然周、自然月等9个基础周期粒度。您可以根据实际业务需求,在该周期粒度的基础上,创建更细粒度的时间周期,例如,近一天、近三周等。方便后续基于您的业务需求,使用细粒度的时间周期做快速的统计分析。

ii. 配置时间周期的基本信息。

🕔 新建时间周	₹ ×	
圆 保存		
基本信息		
周期粒度 *	天 🗸	
英文缩写 *	请输入英文缩写	
	▲ 英文编写是指标的唯一性标识, 一经保存, 无法二次修改。	
英文名称	请输入英文名称	
中文名称 *	请输入中文名称	
描述		
	0/1024	
会粉	HE12P	
参叙	加心	
周期粒度	选择需要创建的细粒度的时间周期所在的基础周期。	
	时间周期的英文缩写。英文缩写是指标的唯一性标识,创建后无法修改。	
英文缩写	英文缩写由小写英文字母、数字、下划线(_)组成,并且以小写英文字母或数字开头。	
	다) (2) (2) (2) (2) (2) (2) (2) (2) (2) (2	
英文名称		
	英大石桥田英大于诗、奴子、下划线(_)、 and (d) 组成,开且以英大于诗或奴子开关。	
	时间周期的中文名称。	
中文名称	中文名称由中文、英文字母、数字、下划线(_)、and(&)、英文括号组成,并且由中文、	芵
	义子 '	
	时间周期的描述信息。	
描述	最大长度为1024字符。	

4. 单击保存,完成创建。

导入时间周期

如果您有大量时间周期待创建,则可以选择批量导入时间周期。DataWorks为您提供了导入模板,您可按照模 板表格填写好待导入的时间周期后批量导入。

1. 进入时间周期页面,鼠标悬停至 <>> 图标,单击Excel导入。

进入时间周期页面可参考创建时间周期的相关步骤。

2. 在导入对话框中选择导入模式,下载导入模板并批量上传。



- i. 在时间周期的导入对话框中,选择导入模式为跳过已存在的同名对象。
 当前仅支持使用跳过已存在的同名对象模式导入。
- ii. 单击下载模板下载批量导入时间周期的模板表格。
- iii. 在模板表格中,填写并保存待导入的各时间周期的周期粒度、英文缩写、中文名称等信息。各参数的格式要求如下。

参数	描述
周期粒度	选择需要创建的细粒度的时间周期所在的基础周期。
英文缩写	时间周期的英文缩写。英文缩写是指标的唯一性标识,创建后无法修改。 英文缩写由小写英文字母、数字、下划线(_)组成,并且以小写英文字母或数字开头。
英文名称	时间周期的英文名称。 英文名称由英文字母、数字、下划线(_)、and(&)组成,并且以英文字母或数字开头。
中文名称	时间周期的中文名称。 中文名称由中文、英文字母、数字、下划线(_)、and(&)、英文括号组成,并且由中文、英 文字母或数字开头。
描述	时间周期的描述信息。 最大长度为1024字符。
负责人账 号ID	创建时间周期的责任人ID。
? 说明	下载模板表格后,如果一开始无法编辑表格,可单击表格顶部的 启用编辑 按钮。

iv. 单击选择文件,选择上述保存的时间周期模板后单击确定。
后续您可以在导入结果的提示对话框中查看导入是否成功及导入详情。

	\checkmark	
	导入完成	
成功:1条		失败: 0条
	UNCONTRACTOR AND	

后续步骤

创建的修饰词与原子指标、时间周期后续可以被派生指标引用,用于统计某一业务活动下的统计指标在具体时间、维度、业务条件下的数值表现。如果时间周期已经被派生指标引用,则您需要删除所有关联的派生指标 后,才可以删除时间周期。派生指标引用时间周期,详情请参见派生指标。

6.1.5.5. 派生指标

派生指标是由原子指标、时间周期、修饰词构成,用于反映企业某一业务活动在指定时间周期及目标范围中的 业务状况。例如,某企业近一周上海地域的销售金额。本文为您介绍如何创建并使用派生指标。

前提条件

- 已创建原子指标,用于确定业务活动数值统计的业务口径及计算逻辑。创建原子指标,详情请参见原子指标。
- 已创建修饰词,用于确定业务活动数值统计的业务范围。创建修饰词,详情请参见修饰词。
- 已创建时间周期,用于确定业务活动数值统计的时间范围。创建时间周期,详情请参见时间周期。
- 已创建维度表,用于确定业务活动需要呈现的分析维度。创建维度表,详情请参见创建维度表。
- 已创建数据域,用于在公共层确定需要统计分析的数据区域。创建数据域,详情请参见数据域。
- 已创建业务过程,用于在公共层创建派生指标时挂载派生指标,反映目标业务活动的业务状况。创建业务过程,详情请参见业务过程。
- 已创建业务分类,用于在应用层确定需要统计分析的业务类别。创建业务分类,详情请参见业务分类。
- 已创建数据集市或主题域,用于在应用层创建派生指标时挂载派生指标,反映目标业务活动的业务状况。
 详情请参见数据集市主题域。

背景信息

派生指标是企业某一业务活动中业务状况的数值统计指示器,通常由 原子指标 + 时间周期 + 一个或多个修饰 词 组成。一个派生指标唯一归属于一个原子指标,用于反映该原子指标在具体时间、业务条件下的数值表现。



• 原子指标用于明确业务的统计口径及计算逻辑。例如,下单总数。

• 修饰词用于限定业务活动的范围。例如,线上生鲜门店、线下生鲜门店。

⑦ 说明 一个派生指标中可以包含一个或多个修饰词。

- 时间周期用于确定需要统计的时间范围。例如,一个自然日。
- 派生指标由原子指标、修饰词、时间周期三大要素构成,用于统计目标指标在具体时间、维度、业务条件下的数值表现,反映企业某一业务活动的业务状况。例如,统计 一个自然日中线上生鲜门店的下单总数 、 一个自然日中线上和线下生鲜门店的下单总数 。

同时,您还可以选择将派生指标关联至一个或多个维度表的字段,通过多个维度分析具体业务活动的统计数 值。创建维度表,详情请参见<mark>创建维度表</mark>。

进入派生指标页面

- 1. 进入派生指标。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 选择工作空间所在地域后,单击相应工作空间后的进入数据开发。
 - Ⅳ. 单击左上方的冒图标,选择全部产品 > 数据建模 > 数据指标,默认进入派生指标页面。

您需要根据实际需求选择创建派生指标的层级。

≡	🍿 DataWorks 촟	2据指标	•		≡	🍿 DataWorks 💈	牧据指标	× -
ĩũ	派生指标	派生指标	Ð	C	ñĭ	派生指标	派生指标	Đ Đ
4	原子指标	公共层 1 应	用层		4	原子指标	公共层 1	应用层
3D-	修饰词	→ 清选择数据域 →	0	~	3 2-	修饰词	□ ▲ 请选择业务分类	
₿	时间周期	✔ ⊕ 数据域	2	Q	₿	时间周期	✔ 詰 业务分类	a
		~ 白 派生指标 (0)		≔			~ 白 派生指标 (0)	:=
		> 🕀 默认 (0)					> 冒畺 测试 (0)	

- 公共层:该层级是基于数据域构建派生指标,创建的派生指标主要用于后续构建可复用面向分析和统计的汇总数据。您可以选择指定数据域,在该数据域下创建派生指标。
- **应用层**:该层级是基于业务分类构建派生指标,创建的派生指标主要用于面向具体应用场景或指定产品进行的个性化数据统计。您可以选择指定业务分类,在该数据域下创建派生指标。

在派生指标页面,您可以选择创建单个派生指标,或批量创建多个派生指标。通常,企业的业务活动 是复杂多样的,在确定的数据域或业务分类下,我们需要分析多个业务活动在不同时间,不同限定范 围的业务状况,因此,建议您使用批量功能,快速生成某类业务活动的多个派生指标。批量创建派生 指标,详情请参见批量创建派生指标。创建单个派生指标,详情请参见创建单个派生指标。

批量创建派生指标

- 1. 在派生指标页面,鼠标悬停至 图标,选择需要创建的指标类型为批量创建派生指标。
- 2. 创建批量生成派生指标的模型。
| ■ In DataWorks | ●教護指标 | ●数合规划 ●数抵标准 ●堆度建模 ●数抵指标 □ ℃ 💗 | | | | | | |
|---------------------------------------|---|---|--|--|--|--|--|--|
| 葡 派生指标 | 派生指标 主 C | | | | | | | |
| ▲ 原子指标 | 公共层 应用层 | | | | | | | |
| ····································· | ● ~ 默认 ~ | 英语课 · 汉语课 · 已选择2 > | | | | | | |
| | 请输入 Q | | | | | | | |
| | ◇□派生指标(0) := | | | | | | | |
| | → 品 默认 (0) | | | | | | | |
| | | ▲ 算服業
③ femde(b注) ④ femde(b注) ● 並代,如道,將基準
③ femde(b注) ● 並代,如道,將基準 | | | | | | |
| | | (© 140点10) ● 位元,北京公務後,第巻席
(◎ 160点1日) ● | | | | | | |
| | | | | | | | | |
| | | ③ maie(東部)
③ maie(東部)
④ maie(東部)
④ 回(中国) ● 道沢、東京の原素 | | | | | | |
| | | | | | | | | |
| | | 选择修饰词 | | | | | | |
| | | ◎ male(男生) | | | | | | |
| | | ② female(女生) | | | | | | |
| | | () female(安生)Smale(劳生) | | | | | | |
| | | 1//0////////////////////////////////// | | | | | | |
| | | ③ 14(近1天) | | | | | | |
| | | ③ 34(近3天) | | | | | | |
| _ | | | | | | | | |
| 山口 | +#2+ | | | | | | | |
| 序亏 | 抽还 | | | | | | | |
| | | | | | | | | |
| 1 | 在该区域,您可以配 | 置 原子指标、时间周期、修饰词 在渲染树(即区域3)的显示顺序。 | | | | | | |
| | | | | | | | | |
| | 在该区域 选择创建 | 派生指标需要的 原子指标 时间周期 修饰词 选择完成后 即可在区域3 | | | | | | |
| | 14 以区域,应注的建成土钼物而女的成 丁钼物、凹间周期、修仰间 ,应挥元成后,即凹住区域3
药收集创建的派开长生 | | | | | | | |
| | 观见付创建的派生组体。 | | | | | | | |
| | | | | | | | | |
| | ? 说明 | | | | | | | |
| 2 | | | | | | | | |
| 2 | ■ 如果没有谷 | 夺合您业务需求的 原子指标、时间周期、修饰词 ,则可以单击 新建 ,创建 | | | | | | |
| | 合适的原品 | 子指标、时间周期或修饰词。 | | | | | | |
| | | | | | | | | |
| | ■ 每个派生排 | 旨标由一个 原子指标、一个时间周期 、以及一个或多个 修饰词 构成。 | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | 该区域以渲染树的形 | 式,为您展示了批量创建的派生指标。指标状态如下: | | | | | | |
| | | | | | | | | |
| | ■: 表示创建的派 | 牛指标在日标业务过程,数据集市或主题域中不存在 | | | | | | |
| | | | | | | | | |
| 3 | | | | | | | | |
| | ■ 🥑 表示创建的派生 | :指标在目标业务过程、数据集市或主题域中已存在。后续在选择生成派生指 | | | | | | |
| | | | | | | | | |
| | 标时,系统会自动 | 过滤掉已创建的派生指标。 | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| 4 | 在该区域,您可以选 | 择派生指标渲染树的显示样式。例如,放大、缩小、居中、全屏展示等。 | | | | | | |
| | | | | | | | | |

i. 选择构成派生指标的原子指标、修饰词、时间周期,并确定其在渲染树中的显示顺序。

ii. 单击**生成派生指标**。

3. 批量生成派生指标。

i. 配置基础信息。

在基础信息设置页签,配置派生指标的基础信息。

主成派生指标							
基	础信息设置	•	选择派生排	标	创建	电派生指标	
数仓分层: *	公共层汇总数据层					~	
业务过程: *	默认 / 默认					~	
关联维度:	asdfd(ff_copy)	~	维度字段			~	添加
	维度表			维度字段		操作	
	asdfd(ff_copy)			f(a)		删除	

参数	描述
数仓分层	 派生指标所属的数仓分层。 汇总数据层:如果您希望创建的派生指标用于后续构建可复用面向分析和统计的汇总数据,则可选择该层级。 应用数据层:如果您希望创建的派生指标用于面向具体应用场景或指定产品进行的个性化数据统计,则可选择该层级。 上图中是以选择汇总数据层为例。
业务过程	派生指标所属的业务过程,用于确定公共层业务活动类别。仅当 数仓分 层配置为 汇总数据层 时,需要配置该参数。
集市/主题	派生指标所属的数据集市或主题域,用于确定应用层业务活动类别。仅当 数 仓分层配置为应用数据层时,需要配置该参数。
关联维度	选择派生指标需要关联的维度字段,单击 添加 。用于后续在汇总表或应用表 中通过该维度分析业务的活动状况。
	⑦ 说明 选择的维度字段默认关联所有需要创建的派生指标。

ii. 单击下一步。

iii. 选择需要生成的目标派生指标。

a. 您可以根据实际业务从创建的指标模型中,选择需要生成的派生指标。默认勾选指标模型中的所 有派生指标。

⑦ 说明 当创建的派生指标在目标业务中已存在,系统会自动过滤掉已创建的派生指标。

生成派生	指标		×
	基础信息设置	选择派生指标	创建派生指标
指标总额	饮: 12, 已选择数: 12		
	派生指标	业务过程	
	近1天_女生&男生_英语课	默认	
	近7天_女生&男生_英语课	默认	
	近1天_男生_英语课	默认	
	近7天_男生_英语课	默认	
	近1天_女生_英语课	默认	
	近7天_女生_英语课	默认	
	近1天_女生&男生_汉语课	默认	
	近7天_女生&男生_汉语课	默认	
	近1天_男生_汉语课	默认	
	近7天 男生 汉语课	默认	
			上一步创建

b. 单击创建,成功批量创建派生指标。

iv. 查看创建的派生指标。

您可以在创建派生指标页签,查看已创建派生指标的基本信息,包括指标编码、指标状态等。

基础信息设置	选择派生指标	创建派生指标
	+C+=.4577	415-#-
中又石林	但你确的	状念
近1天_女生&男生_英语课	INDD31B5CF5B1D	创建成功
近7天_女生&男生_英语课	INDD694EA08799	创建成功
近1天_男生_英语课	INDD082A425CBF	创建成功
近7天_男生_英语课	INDDFCC43AC623	创建成功
近1天_女生_英语课	INDD2B3763348F	创建成功
近7天_女生_英语课	INDD6B903B1F76	创建成功
近1天_女生&男生_汉语课	INDDF4ECCA24DC	创建成功
近7天_女生&男生_汉语课	INDDD0085DA93A	创建成功
近1天_男生_汉语课	INDD09F1B775E0	创建成功
近7天_男生_汉语课	INDD2A4CEA3EA6	创建成功
近1天_女生_汉语课	INDD355BAFE325	创建成功
近7天_女生_汉语课	INDD994DE2FCC2	创建成功

成功创建派生指标后,派生指标渲染树中目标指标的状态将会由。变为。。



创建单个派生指标

1. 选择指标类型。

在派生指标页面, 鼠标悬停至 B图标, 选择需要创建的指标类型为派生指标。

2. 新建派生指标。

i. 配置业务逻辑	0
-----------	---

业务逻辑							
时间周期: *	请选择						
修饰词:	请选择 イント・シート マンド・シート マンド・シート マント・シート しょうしょう マント・シート しょうしょう くろう しょう しょう しょう ひょう しょう しょう しょう しょう くろう しょう しょう しょう しょう しょう しょう しょう しょう しょう しょ						
原子指标: *	遺选择						
关联维度:	维度表 る加						
	维度表 维度字段 操作						
	没有数据						
参数	描述						
时间周期	业务活动数值统计的时间范围。例如,近一天,近一周。 您需要从已创建的时间周期中选择。创建时间周期,详情请参见 <mark>时间周期</mark> 。						
修饰词	业务活动数值统计的范围限定。例如,线上、线下。 您需要从已创建的修饰词中选择。创建修饰词,详情请参见 <mark>修饰词</mark> 。						
原子指标	用于确定目标业务活动及其统计数值的计算逻辑。例如,下单金额。 您需要从已创建的原子指标中选择。创建原子指标,详情请参见 <mark>原子指标</mark> 。						
	用于确定具体业务活动需要分析的维度。例如,商品维度、商家维度。 您需要从已创建的维度表中选择需要的维度字段。创建维度表,详情请参见创建维度表。 关联的维度和派生指标,后续用于生成汇总表、应用表中的统计字段,帮助您进行报表等分析展 示。每个派生指标关联n个维度字段,理论上可以产生的维度粒度组合为 2的n次方 个。例 如,派生指标 统计近一天线上门店的下单金额 ,关联了商品维度、门店维度,则会生成如下 4种统计粒度:						
关联维度	 统计近一天线上的下单金额 统计近一天线上某商品的下单金额 统计近一天线上某门店的下单金额 统计近一天线上某商品在某门店的下单金额 						
	买际汇忌表、应用表中需要统计哪种维度粒度,取决于其选择的维度。详情请参见创建应用表。 选择目标维度字段后,单击 添加 ,才能成功关联维度字段。						

ii. 配置基本信息。

基本信息	
数仓分层: *	公共屋に总数据屋
业务过程: *	请选择业务过程 *
指标编码: *	INDDD01ADDEABF
中文名称: *	请输入中文名称 智能推荐 2019年3月11日 2019
央文名称:	请输入央文治称
1田江:	
	0/1024
参数	描述
数仓分层	 派生指标所属的数仓分层。 汇总数据层:如果您希望创建的派生指标用于后续构建可复用面向分析和统计的汇总数据,则可选择该层级。 应用数据层:如果您希望创建的派生指标用于面向具体应用场景或指定产品进行的个性化数
	据统计,则可选择该层级。 上图中是以选择 汇总数据层 为例。
业务过程	派生指标所属的业务过程,用于确定公共层业务活动类别。仅当 数仓分层 配置为 汇总数据 层时,需要配置该参数。
集市/主 题	派生指标所属的数据集市或主题域,用于确定应用层业务活动类别。仅当 数仓分层 配置为 应用 数据层 时,需要配置该参数。
指标编码	系统自动生成,当前派生指标的唯一标识。
	派生指标的中文名称。单击右侧的 智能推荐 ,即可生成 <mark>时间周期+修饰词+原子指标</mark> 格式的 名称,便于您快速了解该指标所统计的数值含义。
中文名称	⑦ 说明 仅当业务逻辑中时间周期均配置后,中文名称右侧才会显示智能推荐
	中文名称由中文、英文字母、数字、下划线(_)、and(&)、英文括号组成,并且由中文、英 文字母或数字开头。
	派生指标的英文名称。
英文名称	⑦ 说明 仅当业务逻辑中时间周期配置后,英文名称右侧才会显示智能推荐
	英文名称由英文字母、数字、下划线(_)、and(&)组成,并且以英文字母或数字开头。
描述	派生指标的描述信息。

3. 单击**保存**,完成创建。

管理派生指标

您可以在**派生指标**页面,双击目录树中的派生指标文件夹,或单击 图标,即可查看所有派生指标,并对其进行如下管理操作。

≡	🏟 DataWorks 👔	数据指标	· · · ·					∂数1	き规划 🤗 数期	副标准 🥜 维度	建模 🖉 数	掘指标 🇘 🔍	V	-
iii	派生指标	派生指标	₽C	□ 派生	E指标	×								
Ä	原子指标	公共层	应用层		-75	他走头中亡								
30-	修饰词	₩ す	~	8	加建 🎽	(现象大键子								
世	时间周期	请输入	Q		đ	指标编码	英文名称/中文名称	业务过程	数据域	集市/主题	负责人	最新修改时间	攝作	
		> 🗂 派生指标 (0)	=			NDD994DE2FC	chinese_1w_female 近7天_女生_汉语课	默认	默认		17	2022-01-20 15:05:02	编辑	删除
					2	NDDFCC43AC6 23	english_1w_male 近7天_男生_英语课	默认	默认		27	2022-01-20 15:05:02	编辑	删除
					 3	NDDD0085DA9 3A	chinese_1w_male_female 近7天_女生&男生_汉语课	默认	默认		27	2022-01-20 15:05:02	编辑	删除
					5	NDD355BAFE32	chinese_1d_female 近1天_女生_汉语课	默认	默认		57	2022-01-20 15:05:02	编辑	删除
					I	NDD082A425C BF	english_1d_male 近1天_男生_英语课	默认	默认		27	2022-01-20 15:05:02	编辑	删除
						NDDF4ECCA24 DC	chinese_1d_male_female 近1天_女生&男生_汉语课	默认	默认		57	2022-01-20 15:05:02	编辑	删除

• 编辑派生指标。

单击目标派生指标**操作**列的编辑,即可进入该派生指标的编辑页面。您可以在派生指标的编辑页面修改目 标派生指标的**时间周期、修饰词、业务过程**等信息。

⑦ 说明 如果该派生指标已经关联了汇总表,则不能修改时间周期和已经在汇总表中使用过的维度 字段。您需要先在汇总表中解绑目标维度字段,才可以修改时间周期及维度字段。

• 删除派生指标。

单击目标派生指标操作列的删除,即可删除该派生指标。

⑦ 说明 如果目标派生指标已经关联了汇总表,则您需要先在汇总表中解绑该派生指标,解绑后才可以删除该派生指标。

后续步骤

成功创建派生指标后,您可以在汇总表、应用表中将同一个数据域下相同时间周期、相同维度的多个派生指标 的统计数据组织在一起(每个派生指标及关联的维度将构成汇总表中的一个字段),为后续的业务查 询,OLAP分析,数据分发等提供基础。汇总表、应用表使用派生指标,详情请参见创建应用表。

7.HoloStudio

7.1. 概述

本文为您介绍什么是HoloStudio,以及HoloStudio的核心功能。

HoloStudio是基于交互式分析Hologres构建的一站式OLAP开发平台,深度集成于阿里云智能开发平台 DataWorks。

HoloStudio通过可视化方式,为您提供标准化、无门槛的开发服务和一站式构建实时数据仓库服务,以及高效、便捷的管理服务。

HoloStudio在PostgreSQL的标准管理基础上,为您提供更多Hologres功能。

功能

HoloStudio的核心功能如下:

• 外部表

支持一键创建MaxCompute外部表,同步外部表数据,数据预览和分析达到亚秒级。

● 表

为您提供可视化方式和PostgreSQL语句两种建表模式,满足不同开发层次的需求。

• Hologres开发

基于DataWorks调度框架,构建MaxCompute离线数据到Hologres索引数据的周期。

• SQL Console

基于SQL编辑器,为您提供纯SQL语句的开发以及秒级查询体验。

终端

在Web端为您提供psql客户端服务,满足您在终端开发的需求。

7.2. HoloStudio快速入门

DataWorks-HoloStudio作为基于交互式分析(Hologres)引擎的一站式开发平台,在为您提供SQL界面开发的同时,也支持使用可视化UI界面完成Hologres开发,同时也与DataWorks深度集成,方便您完成数据的周期 性调度。

本小节主要为您介绍使用DataWorks-HoloStudio的操作步骤,通过以下简单步骤的综述,帮助您快速掌握 HoloStudio的基本使用流程。



前提条件

在开始使用DataWorks-HoloStudio之前,请确保您已经成功开通Hologres实例,详情请见购买Hologres。

114.1-#		当前配置
形成	キャック いるのの キャルス (40次) キャスス (上海) ギホー (初期)	地域: 华南1 (深圳)
21.00.35.35	s/#aseoB	计算资源: 64核256GB
计异页源	0482200GB	存储资源: 500 GB
	单价170元/核/月,购买超过512核,请您 提交工单 开通。	实例名称: -
		购买时长: 1个月
存储资源	500 GB	配置费用:
	超出部分将自动转为按量付费。	
		<u>1</u>
实例名称		🏦 ::::::::::::::::::::::::::::::::::::
	输入范围:长度限制为2-64个字符。	
	输入记题:长度限制为2-64个子付。	
实例名称		

操作步骤

使用DataWorks-HoloStudio进行开发的完整操作步骤如下:

1. 新建数据库(推荐)

实例购买成功后,系统会自动生成一个名叫postgres的数据库,该数据库分配的资源少且仅供管理用途,实际业务需要您新建一个DB。为节省SQL操作,Hologres管控台支持直接可视化新建DB。

进入管理控制台的实例详情页,单击左侧导航栏的DB管理 > 新增Database,输入DB名即可新建成功。

新增Database	Х
Database名称 请输入Database名称	
 简单权限模型 ● 开启 ● 关闭 	
Hologres兼容PostgreSQL,使用与PostgreSQL完全一致的权限模型(以下简称专家模式)。基于对业务理解和实践经验,Hologres抽象了一套简单权限模型(Simple Permission Model,SPM),以简化用户权限管理的复杂度。 建议您在DB创建时开启SPM。SPM可以被关闭(恢复到专家模式),但不可以被再次开启。更多信息请参见Hologres简 单权限模型文档。 SPM开启后,SPM会为当前DB自动构建4个用户组: 1. <db>_viewer组对此DB下的所有表、外表、view等对象有查询(SELECT)权限。 2.<db>_writer组除了<db>_viewer权限外,对此DB中的所有对象有INSERT、UPDATE和DELETE权限。 3.<db>_developer组拥有此DB中所有表、外表、view等对象,除了<db>_writer权限外,还可以创建新对象或者删除已有对象。 4.<db>_admin组负责此DB的管理,拥有此DB上的所有权限,并可管理这4个组的成员。 授权是把用户加到对应组中。</db></db></db></db></db></db>	i
确定 取	消

⑦ 说明 Hologres兼容PostgreSQL,使用与PostgreSQL完全一致的权限模型(简称专家模式)。 基于对业务理解和实践经验,Hologres抽象了一套简单权限模型(SPM)来简化授权操作,详情可以 参见简单权限模型概述。新建DB时,您可以选择为新DB开启简单权限模型(推荐操作)并通过管控 台来为新用户可视化授权,以简化授权操作,具体操作请参见DB管理。

2. 创建或绑定工作空间

绑定Hologres实例,分为以下两种情况:

○ 无工作空间

如果您没有使用过DataWorks工作空间,需要在DataWorks管控台创建一个同地域的工作空间并绑定。

在DataWorks管理控制台单击左侧导航栏的工作空间列表,并单击创建工作空间,填写工作空间信息。您可以根据自身需求,选择是否同时开通其它计算引擎服务,详情请参见绑定Hologres实例,并填 写实例配置信息。

⑦ 说明 Hologres在标准DataWorks工作空间分开发环境和生产环境,但由于当前HoloStudio 暂不支持环境切换,导致生产环境的数据在HoloStudio不可查询,建议开发和生产绑定同一个 DB。

基本配置	(>) i	选择引擎	3 引擎许
交互式分析Hologres			
* 实例显示名称			
开发环境		生产环境	
*访问身份 🕜:	● 任务执行者	*访问身份 😮:	 阿里云主账号 · · · · · · · · · · · · · · · · · · ·
* Hologres实例名称:	×	• Holograe实例名称:	
* 数据库名称 😦 :		* 数据库名称 ① :	IN INC.
* 服务器:	bangzhou-	•服务器:	
	internal.hologres.aliyuncs.com		hangzhou-
* 端口:	80		internal.noiogres.allyuncs.com
连通性测试:	测试连通性	"靖山:	80
		连通性测试:	测试连通性

○ 已有工作空间

如果您已创建DataWorks工作空间,可以直接使用该工作空间绑定Hologres实例。具体操作步骤如下:

a. 进入Holostudio

在Hologres管控台选择实例开通的region,单击页面上方的前往DataWorks-HoloStudio开发,即可进入对应的工作空间。

	上海) 🔻		
Hologres	Hologres / 实例列表		
概览	Hologres引擎管理		
实例列表	新增引擎实例 登录Hologres数据库 前	i往DataWorks数仓开发	搜索实(
	实例名称 / 实例ID	运行状态	创建时间
		⊘正常运行	2020年1

b. 绑定实例

进入Holostudio后,单击右上角**小扳手**,进入**工作空间配置**页面,单击**绑定HologresDB**进行实例绑定。更多关于绑定实例的详细操作步骤参见绑定Hologres实例。

DetaWorks	•••		Q. Valapanatati i
O TARANKE	(現象)(5		
8 4.188	治理病意用题: ()		
• кала	80.00 × 80 0	Martin Remercine 3 🧭	
	N/NSR/42846:	数以改结自动重要地的间隔:20% 22	
4. BERRY			
	安全設置		
	RT Movie: HIR .	元作于幾号索更自己的市业责任人:	
	沙瑞古名单(配置中的任务可以达回的产物社组地名):		#20#02#
	Phila	80	8.5
		没有数据	
	1 2001年位年		
	MaxCompute EXITIN Hologen		
		+ BEEHologen08	

绑定完成之后,返回HoloStudio,在PG管理页面单击刷新,既可看到已成功绑定的数据库。



3. 授权子账号(可选)

若您需要给子账号授权,可以按照以下步骤操作:

i. 创建子账号

在Hologres管控台实例详情页单击左侧导航栏用户管理 > 新增用户,勾选需要授权的子账号,并选择用户类型。可以直接授权为superuser(拥有实例所有权限,无需额外授权),也可以授权为普通用户(访问Hologres还需要额外授权)。

Hologres / 3	实例列表 /			
÷	◎ 正常运行	新增用户	х	
实例配置 用户管理 DB管理 监控警告	用户管理 阿里云账号(主、子账号)可以作为Hologres实例的用户 本页面只处理用户账号的管理。对用户的授权可量接移 Hologres兼容PostgreSQL权限管理体系。用户分为Sup 新造用户 成员	选择组织成员		4 9
	Internet and the second se	全选		SuperUser
		用户类型 ②		
			确定取消	

ii. 授权子账号

若您的子账号是普通用户normal,还需要给子账号授权才能访问Hologres实例。您可以选择使用专家权限模式或者简单权限模型(推荐)为子账号授权。

■ 简单权限模型(推荐)

在管控台新建DB并开启简单权限模型之后,可以直接将子账号加入对应的用户组来授权。

testdb 权限管理				
Hologres兼容PostgreSQL,使用与PostgreSQL完全一致的权限模型(比	新增授权	Х	!(Simple Permission Model,SPM),以简化用户权限管理的复杂度。	
建议愿在DB创播时开启SPM、SPM可以减关用(伪复到专家模式),但 SPM开启后、SPM会力当前DB自动构建4个用户组: 1.cdb_xiewer组为社DB下的所有表、外表、view等对意有重应。(2.cdb_ximter组度了cdb_xiewer获良外、对此DB中的所有对象有 3.cdb_sdeveloper组派有出DB中所有表、外系、view等对象有 3.cdb_sdeveloper组派有出DB中所有表、关系、view等对象,并可 7.cdb_sdeveloper组派和EDB中所有表、并不可	被硬权账号 这样很权服号 权限策略 Admin ③ Developer ③ ③ Writer ③ ③ Viewer ④			
按以此此用广加的外型相子。 新聞授权 成员	请定 取	消	海臣	操作

关于简单权限模型的使用可以参见简单权限模型的使用。

■ 专家模式授权

专家权限模式授权采用的是标准的PostgreSQL授权语句,您可以参照以下语句进行授权:

GRANT SELECT, INSERT ON TABLE TABLENAME TO "云账号ID/云邮箱";//若是子账号还需要在云账号 ID前加"p4_",即"p4_id"

GRANT SELECT, INSERT ON ALL TABLES IN SCHEMA public TO "云账号ID/云邮箱";//若是子账号 还需要在云账号ID前加"p4_",即"p4_id"

更多关于专家权限模型的使用可以参见专家权限模型。

■ 使用Holostudio开发

若您的子账号需要使用HoloStudio完成Hologres开发,还需要将子账号添加进DataWorks的项目 工作空间。

在DataWorks工作空间配置页面单击**工作空间管理 > 成员管理 > 添加成员**,并给成员授予一定的项目权限。

6	DataWorks											2		1
۵	□ 工作空间配置	成员管	管理								I	作空间管理	資看角色权限	
*	成员管理	请输入质	成员名或者云账号进行排	東京	搜索							Γ	添加成员	L
0	权限列表	全部	项目所有者 1	项目管理员 1	开发 0	运维 0	部署 0	访客 0	安全管理员 0					ľ
v	MaxCompute高级配置		成员	云账号	1	角色				ற்ப.	入时间		操作	
						项目管理员	i ×		~	20	19-08-23 13:39:32		所有者	
		批量測	198							〈上一页 】	下一页 >	每页显示:	10 ~	

4. HoloStudio开发

进入HoloStudio之后,即可进行Hologres开发:

- SQL Console纯SQL开发,详情请参见SQL Console。
- 表管理可视化UI─键建表、数据预览,详情请参见创建及管理内部表。
- 一键创建MaxCompute外部表、查询数据,详情请参见创建及管理外部表。

○ 与DataWorks深度集成,实现周期性调度,详情请参见Hologres开发:周期性调度。

关于Hologres的更多开发工具请参见概述。

7.3. 绑定交互式分析实例

本文将为您介绍如何将交互式分析(Hologres)实例绑定至DataWorks工作空间。

前提条件

- 1. 使用HoloStudio服务前,请确保已经成功开通交互式分析实例,详情请参见购买Hologres。
- 2. 实例创建成功后会默认生成一个名叫postgres的数据库,仅供管理用途,实际业务需要在Hologres管控 合新建一个业务DB,详情参见DB管理。

新建工作空间绑定交互式分析实例

若您在此之前并未开通DataWorks工作空间,可以按照以下步骤来绑定实例:

1. 进入DataWorks管控台

使用阿里云主账号进入DataWorks管控台,选择与实例同一个region,单击左侧菜单栏工作空间列表进 行新建工作空间。

三 (-) 阿里云 (#南1 (深	(月二) -		Q 搜索文档、控制台、AF	PI、解决方案和资源	佣 工单	备案	企业	支持	官网	>_	Ĵ. <i>È</i>	7 (?) 简	体	0
DataWorks	DataWorks / 工作空间列表											产品	动态 6	帮助了	て档
概览	当前使用的是 基础版 ,版本到期日为	习永久 。					版本升级	版本延	10 <u>2</u> 1	音版本详情	购买独	享资源组	且 购买	资源包	
工作空间列表资源组列表	创建工作空间 请输入工作空间	1/显示名	Q											(
计算引擎列表 ^	工作空间名称/显示名	模式	创建时间	管理员		状态	5		升 操f	乍					
MaxCompute Graph Compute		标准模式 开发跟生产隔离				~	正常		M 进 进	作空间配置 入数据集成	开通服务	务配置 居服务	进入数排 更多	田田安	
交互式分析		标准模式 开发跟生产隔离	-			~	正常		v If 进	作空间配置 入数据集成	开通服的进入数据	時配置 居服务	进入数据 更多	居开发	

2. 新建工作空间

在工作空间列表页面单击**创建工作空间**, 输入工作空间名称, 模式选择可根据项目情况选择(详情请参见简单模式和标准模式的区别)。

= (-) 阿里云 华南1 (深圳) ▼			Q 搜索文档、控制台、API、解决力	7案和资源	费用 エ	单省	解 企业	支持	官网	۶.,	٥.	Ä		
DataWorks DataWorks / 工作空间列表			创建工作空间											
概览工作空间列表		当前使用的是 基础	版 ,版本到期日为	9 永久 。	1 基本配置			- 2	选择引擎					
资源组列表		创建工作空间	请输入工作空间]/显示名	基本信息 工作空间名称 	holo_test								
计算引擎列表 へ MaxCompute		工作空间名称/显示	名	模式	显示名	如果不填, 黑	默认为工作?	空间名称						
Graph Compute				开发跟生产隔离	• 模式	标准模式 (チ	开发跟生产	隔离)					~	
交互式分析				标准模式 开发跟生产隔离	描述									
				标准模式 开发跟生产隔离	高级设置									
					* 能下载Select结果 ❷	π								
					下一步	び消								

3. 选择计算引擎

选择计算引擎为**交互式分析Hologres**,并选择对应的实例付费方式,可根据项目情况选择是否同时开通 其他引擎服务。

基	本配置 2 选择引擎 3	引擎详情
≩Data	Works服务	
V	⑦ 数据集成、数据开发、运维中心、数据质量 您可以进行数据同步集成、工作流编排、周期任务调度和运维、对产出数据质量进行检查等。	
译计算	引擎服务	
	✓ MaxCompute 包年包月 按量付费 开发者版 去购买 开通后,您可在DataWorks里进行MaxCompute SQL、MaxCompute MR任务的开发。 充值 续费 升级 降配	
	€。实时计算 共享模式 去购买 ○ 独享模式 开通后,您可在DataWorks里面进行流式计算任务开发。	
	E-MapReduce 开通后,您可以在DataWorks中使用E-MapReduce进行大数据处理任务的开发。	
~	● 交互式分析Hologres ● 包年包月 ○ 按量付费 开通后,您可以在DataWorks里使用Holostudio进行交互式分析(Hologres)的表管理、外部表管理、SQL任务的开发。	
	● AnalyticDB for PostgreSQL 去购买 开通后,您可以在DataWorks中进行分析型数据库AnalyticDB for PostgreSQL任务的开发。	
择机器	学习服务	
		- (

4. 填写实例信息

填写实例信息,当前标准版的工作空间分开发环境和生产环境,以保证数据的安全性。但是由于当前 HoloStduio暂未上线环境切换功能,建议开发和生产绑定同一个DB,否则HoloStudio将无法查询生产环 境的数据。最后单击**创建工作空间**即可完成创建。

基本配直	·····································	择引擎	3 引擎详
交互式分析Hologres			
* 实例显示名称			
开发环境		生产环境	
*访问身份 🕜:	● 任务执行者	*访问身份 🔵:	● 阿里云主账号
* Hologres实例名称:	×	and the second state of th	○ 阿里云子账号
*数据库名称 😮:		* Hologres实例名称:	
*服务器:	J. I-cn-	*数据库名称 🕜:	
	hangzhou- internal.hologres.aliyuncs.com	•服务器:	hangzhou-
* 端口:	80	· ** •	20
连通性测试:	测试连通性	通过:	测试连通性

|--|

环境	配置项	说明	备注		
	访问身份	任务执行者	无		
	Hologres实例名称	下拉框可选	无		
	数据库名	需要绑定的DB名	实例创建后有一个名为 postgres的默认DB(仅 供管理用途),实际业务 使用请前往 <mark>管控台</mark> 新建 Database。		
开发环境	服务器	Hologres实例的网络地 址	实例选择后自动生成		
	端口	Hologres实例的网络地 址	实例选择后自动生成		
	连通性测试	测试是否连通	无		

DataWorks公共云合集·HoloStudio

环境	配置项	说明	备注
	访问身份	主账号/子账号	DataWorks任务提交调度 后,作为在Holo引擎内 执行代码的身份(账 号)。
	Hologres实例名称	下拉框可选	无
生产环境	数据库名	需要绑定的DB名	当前建议绑定开发环境的 DB
	服务器	Hologres实例的网络地 址	实例选择后自动生成
	端口	Hologres实例的网络地 址	实例选择后自动生成
	连通性测试	测试是否连通	无

5. 进入HoloStudio开发

工作空间创建成功后会跳转至管控台,可看到已配置好的工作空间信息,单击工作空间名或者进入数据 开发,即可进入DataWorks,示例如下:

	(深圳) 🗸		Q 搜索文档、控制台、A	PI、解决方案和资源	费用	工单	备案	企业	支持	官网	2	۵.	Ä	0	简体	0
DataWorks		DataWorks / 工作空间列表												ŕ	品动态	新助	文档
概宽			* **						15+114	115-4-7			Let Di-	- 117 X-b 107 (Mr.)	5.4Q DP		
工作空间列表		当期使用的定 會知識, 版本到别口	八水ス 。						NX 447T30	10,43	Elei E.	自 啟 中 1 年	iri ya	(天儒学页)	KAEL MA	天页深	3
资源组列表		创建工作空间 请输入工作空	间/显示名	Q													С
计算引擎列表	•	工作空间名称/显示名	模式	创建时间	管理员			状态	ş		升 操	作					
MaxCompute		holo_test	标准模式					~	正常		H I	作空间配	置 开;	通服务配置	进入数	数据开发]
Graph Compute		holo_test	开发跟生产隔离								进	入数据集	成 进	入数据服务	更多		

进入DataWorks工作间之后,单击全部产品 > HoloStudio,进入HoloStudio开发界面。

\$	🕅 DataStudio	holo_te	est 🔍 🗸			∂ 任务发布	₽ 运维中心	Ľ	v Loone
=	全部产品		数据汇聚	应用开发					
00	数据集成		Os 数据集成	App Studio					
ж	DataStudio(数据开发)								
*	运维中心(工作流)		数据开发	其他					
4	任务发布		💸 DataStudio(数据开发)	6 需求管理	运行 - F8				
Θ	数据地图(数据管理)		🐥 数据服务		停止 - F9				
			ନ Function Studio New		保存 - Ctrl+S Cmd+S				
					撤消 - Ctri+Z Cmd+Z				
			I Holo Studio		重做 - Ctri+Y Cma+Y				
			畫 数据分析		豐词 - Ctrl++ Cmd++				
					替换 - Ctri+H Cmd+Ait+F				

进入HoloStudio后,在**PG管理**单击**刷新**,即可看到已成功绑定DB,之后您可以开始使用HoloStudio开发。



选择已有工作空间绑定交互式分析实例

若您已有DataWorks项目空间,可将实例绑定至该工作空间。具体操作步骤如下:

⑦ 说明 工作空间和开通的Hologres实例必须在同一region才能成功绑定。

1. 进入工作空间

在DataWorks管控台找到已有的工作空间,单击**工作空间名**或者**进入数据开发**,进入到DataWorks开发 界面。

	上海) 🔻		Q	搜索文档、控制台、API、解决方案和资源	费用	工单	备案	企业	支持	官网	2	۵. ۲	0	简体	(
DataWorks	DataWorks / 工作空间列表												产品动态	6 帮助文	:档
概览	业态/未用から 17046 版本の期日	4- s.A					PE	5- * -311.611	RG-1-25 10			助理者的复数	100.00	100 JT 200 300 400	
工作空间列表	⇒用可12773日372 587386786 (0.44+±1781日)	/9 // / .					ho	K4447T90X	10C44536393	E-1872	(44) (41) (11)	则天11平1	(WKSH :	购失风感已	
资源组列表	创建工作空间 请输入工作空间	间/显示名	Q											C	;
计算引擎列表 ^	工作空间名称/显示名	模式	创建时间	管理员	状态			开通服务		操作					
MaxCompute Graph Compute	holo_demo holo_demo	标准模式 开发跟生产隔离			✔ 正常	20		~		工作的进入数	2间配置 数据集成	开通服务配 进入数据服	置 进入 务 更多	数据开发	

2. 工作空间配置

进入DataWorks开发界面后,单击右上角**小扳手**,进入工作空间配置页面,在**计算引擎信息**选 择Hologres,并单击**绑定HologresDB**进行实例的绑定。

DetaWorks							Q 🔮 antoparada ta
	al anna						
O TODARS	利用用的						
8 4.188	ARRAN C						
 KORPIA 	MAX 网络的网络		~ 6 8 ©	101 (Auto 40) (A	2008/2008: 2 🧭		
	BARRING BRIDE		~ 63 ©	新いた場合	主要违约问问题:2日钟 🙁		
V REAL							
	安全設置						
	RTEselectER:			元件子融号	R里自己的市点责任人:		
	沙塘白各单(配置水峡世界可以达问的	(P他让后地名):					#20802#
	Pitte			III C		8.5	
				没有数据			
	11050048						
	MaxCompute WHITE Ho	logra					
				+ #EHologresCB			

3. 绑定实例

在绑定HologresDB页面填写相关信息,并单击测试连通性,显示测试联通性通过表示能正常绑定,单击确定即可。

⑦ 说明 当前标准版的工作空间分开发环境和生产环境,以生产开发相隔离保证数据的安全性。但 是由于当前HoloStduio未上线环境切换功能,建议开发和生产绑定同一个DB,否则HoloStudio将无 法查询生产环境的数据。

8):						
		端口			操作	
			没有数据			
绑定HologresDB						\times
* 实例显示名称:						
生产环境			开发环境			
基础信息			基础信息			
* 访问身份 😗 :	● 阿里云主账号 ○ 阿里云子账号		* 访问身份 🔵:	 任务执行者 		
引擎信息			引擎信息			
* Hologres实例名称:	(planet)	~	* Hologres实例名称:	ng militari militari	~	
* 数据库名称 🔵:			* 数据库名称 🔵:			
*服务器:	ngzhou-		*服务器:	:n-hangzhou-		
	internal.hologres.aliyuncs.com			internal.hologres.aliyuncs.com		
* 端口:	80		* 端口:	80		
连通性测试			连通性测试			
测试连通性:	测试连通性		测试连通性:	測试连通性		

环境	配置项	说明	备注
标题	实例显示名称	自定义	无
	访问身份	主账号/子账号	DataWorks任务提交调度 后,作为在Holo引擎内 执行代码的身份(账 号)。
	Hologres实例名称	下拉框可选	无
生产环境	数据库名	需要绑定的DB名	当前建议绑定开发环境的 DB
	服务器	Hologres实例的网络地 址	实例选择后自动生成
	端口	Hologres实例的网络地 址	实例选择后自动生成
	连通性测试	测试是否连通	无
	访问身份	任务执行者	无
	Hologres实例名称	下拉框可选	无

开发环境	数据库名	需要绑定的DB名	实例创建后有一个名为 postgres的默认DB(仅 供管理用途),实际业务 使用请前往 <mark>管控台</mark> 新建 Database。
	服务器	Hologres实例的网络地 址	实例选择后自动生成
	端口	Hologres实例的网络地 址	实例选择后自动生成
	连通性测试	测试是否连通	无

4. HoloStudio开发

实例绑定成功后,返回到HoloStudio页面,在**PG管理**单击刷新,即可看到已成功绑定的DB,您即可开始 使用HoloStudio开发。

6	💷 Holo Studio	holo_demo holo_demo	~ ~							Ľ	ф °	👻
Ш	PG管理	G C 🕀										
Q	文件名称/创建人											
≖	> ≣ myholodb											
S												
۶.					Welc	ome	to Hold	\C+1	Idio			
					VVEIC			5011				
				开始一	一个新任务:							
					L.		Ŀ,		Ŀ,			
					「二十		▲ 新建SOL Console		新建数据开发			

7.4. 新建数据库

实例购买功后会默认生成一个数据库postgres,但系统分配给该数据库的资源小仅供管理用途,实际业务需要新建数据。本小节将会为您介绍如何在HoloStudio中创建新的数据库。

操作步骤

1. 新建数据库

您可以使用以下两种方式来新建数据库:

○ Hologres管控台新建数据库

Hologres管控台支持以可视化的方式新建数据库,在Hologres管控台进入对应实例的详情页,选择DB 管理,单击新建Database进行创建。

☰ (-) 阿里云 华东2(上海) 🔻		
Hologres	Hologres / 实例列表 /	212 - 1000 - 2.3 - 1000	
概览	< Natural and an	● ● ● ● ● 正常运行	
实例列表	实例配置 用户管理	DB管理 Hologres兼容PostgreSQL。实例创建后有一个名为 <i>postgres</i> 的默认DB(仅供管 <mark>新增Database</mark>	里用途)。实际业务请新建Database。
	DB管理	Database名称	权限策略 🔮
	监控告警	NAMES AND	简单 切换到专家权限模型

⑦ 说明 当前版本的Hologres将会采用两套权限体系来为用户提供更加方便且完善的授权体
 系:专家模式授权和简单权限模型。在创建新DB的同时,您可以根据业务情况来选择是否开启简
 单权限模型,为方便权限管理建议您开启。

。 SQL方式新建数据库

您也可以以SQL命令语句的方式新建数据库,在HoloStudio执行SQL语句新建数据库,语法如下。

CREATE DATABASE <dbname>; CREATE DATABASE testdb; //示例新建一个名称为testdb的数据库</dbname>
 □ ① ● ○ ○ ● ○ 当前DB:postgres 1 CREATE DATABASE testdb;
运行日志
正在执行中 【执行代码: (1)】 CREATE DATABASE testdb 执行成功,查询结果: [0]行,当前返回: [0]行,耗时: [249]ms.

2. 绑定工作空间

新建数据库成功后,需要将数据库绑定至DataWorks工作空间才能在HoloStudio里使用新的数据库。

单击页面右上角的小扳手(),进入工作空间配置页面。选择计算引擎信息下的Hologres > 绑定 HologresDB。

DetaWorks	••				Q. Philippenints
	1.000				
0 Inderst	WERS				
9 A188	HIMMERIAN (
ORTHR	銀从調査査理細:	~ #8 ©	製品改植自动重進の数:3 😕		
√ ManComputeRidEE	WARRAN REG	~ 68 ©	数以改结自动重用时间间隔:2019 22		
V DERRIG					
	2228				
	RTEselectER:		大作子錄号室里自己的市在责任人:		
	沙積古名单(配置建設任务可以流行的学校社会地名):				#50#02#
	1718842		III C	8.7	
			没有数据		
	110020948				
	MaxCompute XXIII Hologna				
			+ BEETHologens00		

3. 配置数据库信息

填写数据库配置信息,单击**测试连通性**,若显示**测试通过**,表明数据库能成功绑定至工作空间,单击**确** 定即可。

⑦ 说明 当前标准版的工作空间分开发环境和生产环境,以生产开发相隔离保证数据的安全性。但 是由于当前HoloStduio未上线环境切换功能,建议开发和生产绑定同一个DB,否则HoloStudio将无 法查询生产环境的数据。

疲名):		0	测试通过			
		端口			操作	
			没有数据			
绑定HologresDB						\times
* 实例显示名称:						
 生产环境 基础信息 访问身份 ●: 引擎信息 	● 阿里云主账号 〇 阿里云子账号		 計一次环境 基础信息 *访问身份 ●: 引擎信息 	🕐 任务执行者		
* Hologres实例名称:	and the second s	~	* Hologres实例名称:	ngalana natara	~	
* 数据库名称 🔵:	finanza -		* 數据库名称 🔵:			
*服务器:	ngzhou- internal.hologres.aliyuncs.com		*服务器:	internal.hologres.aliyuncs.com		
* 端口:	80		* 端口:	80		
连通性测试			连通性测试			
测试连通性:	测试连通性		测试连通性:	测试连通性		
					确定	取消

标题 实例显示名称 自定义 访问身份 主账号/子账号	备注
访问身份	无
	DataWorks任务提交调度 后,作为在Holo引擎内 执行代码的身份(账 号)。

环境	配置项	说明	备注
	Hologres实例名称	下拉框可选	无
生产环境	数据库名	需要绑定的DB名	当前建议绑定开发环境的 DB
	服务器	Hologres实例的网络地 址	实例选择后自动生成
	端口	Hologres实例的网络地 址	实例选择后自动生成
	连通性测试	测试是否连通	无
	访问身份	任务执行者	无
	Hologres实例名称	下拉框可选	无
开发环境	数据库名	需要绑定的DB名	实例创建后有一个名为 postgres的默认DB(仅 供管理用途),实际业务 使用请前往 <mark>管控台</mark> 新建 Database。
	服务器	Hologres实例的网络地 址	实例选择后自动生成
	端口	Hologres实例的网络地 址	实例选择后自动生成
	连通性测试	测试是否连通	无

4. 使用新的数据库开发

新DB绑定成功后,返回HoloStudio,在PG管理页面单击**刷新**,即可显示新的数据库,您可以使用新的数 据库来开发。



后续步骤

支持阿里云子账号绑定DB,但必须满足以下两个条件:

- 子账号拥有绑定DB的连接权限。
- 子账号是当前DataWorks工作空间的管理员

若是子账号在绑定时出现下图所示的灰色按钮,请检查子账号是否满足以上两个条件。

9	DataWorks			•					ಬೆ	V [1 1 1]
	=		沙箱白名单(配置shell任	E务可以访问的IP地址或域名)						
ø	工作空间配置		IP地址					端口		
æ,	成员管理		192.168.0.1					80		
0	权限列表									
s.	数据源管理		92.168.0.2					80		
		ĥ	计算引擎信息							
			MaxCompute 实	时计算 Hologres						
							十绑定HologresDB			

7.5. 子账号使用DataWorks

本文为您介绍主账号如何创建子账号,并授权子账号使用DataWorks进行数据开发。

背景信息

系统默认设置购买实例的主账号为超级管理员Superuser。Superuser拥有该实例的所有权限。

其他用户必须经过主账号授权才可以访问实例。

子账号包括以下两种权限控制。

• RAM权限

RAM权限为可选权限。子账号被授予相应RAM权限后,可以在Hologres管理控制台管控实例,例如购买或 删除实例、升降配实例资源、修改网络类型以及查看实例信息等。

• 实例开发权限

实例开发权限为必选权限。子账号必须经过主账号授予实例的开发权限后,才能连接实例并进行数据开 发。

Hologres支持使用简单权限模型和专家权限模型为子账号授权。两种权限模型的介绍如下。

• 简单权限模型(推荐)

简单权限模型是Hologres基于PostgreSQL的授权体系,为提升用户体验,制定的一种粗粒度的权限模型, 详情请参见<mark>简单权限模型概述</mark>。

• 专家权限模型

专家权限模型是与标准PostgreSQL语句相同的授权体系,简称专家模式。您可以按照标准的PostgreSQL授 权语句为子账号授权,详情请参见专家权限模型。

创建子账号

如果您已有子账号,请跳过该步骤。

- 1. 使用主账号登录阿里云官网。
- 2. 使用主账号登录RAM管理控制台。
- 3. 在左侧导航栏的人员管理菜单下,单击用户。
- 4. 单击创建用户。

您也可以单击添加用户,一次性创建多个子账号。

- 5. 在用户账号信息区域下, 输入登录名称和显示名称。
- 6. 在访问方式区域下,选择控制台访问。
- 7. 设置子账号的登录密码。

8. 单击确定。

授权子账号

● 授予子账号RAM权限。

主账号授予子账号相应的RAM权限后,子账号可以在Hologres管理控制台执行查看、购买或删除实例等操作。详情请参见概述。

• 授予子账号实例开发权限。

子账号必须经过主账号授予实例的开发权限后,才能在Hologres实例中进行数据开发。详情请参见授予RAM 用户实例的开发权限。

添加子账号至DataWorks工作空间

您需要添加子账号至相应的DataWorks工作空间,才可以使用子账号在DataWorks中进行数据开发。步骤如下:

- 1. 进入工作空间配置页面。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 单击目标工作空间操作列的工作空间配置。
 - iv. 在工作空间配置对话框,单击更多设置,进入工作空间配置页面。

您也可以进入数据开发页面,单击右上角的图图标,进入工作空间配置页面。

- 2. 在左侧导航栏,单击成员管理。
- 3. 在成员管理页面,单击右上角的添加成员。
- 4. 在添加成员对话框,单击刷新,同步当前阿里云账号下的所有子账号至待添加账号列表中。

添加成员						
您可以前往	RAM控制	l台新建子账号,并单 <mark>表</mark> 刷	新日步至此页面。			
* 选择组织	喊员:	待添加账号			已添加账号	
		请输入成员名称	Q		请输入成员名称	Q

5. 在**待添加账号**勾选需要添加的成员账号,单击>,移动需要添加的子账号至已添加账号中。

6. 勾选需要授予的角色,单击确定。

工作空间的创建者默认为管理员角色。具体角色的权限说明请参见<mark>附录:预设角色权限列表(空间</mark>级)。角色描述如下表所示。

角色	描述
项目所有者	项目所有者拥有工作空间的所有权限。
空间管理员	除了拥有开发角色和运维角色的全部权限外,还可以进行添加或移出工作空间 成员并授予角色,创建自定义资源组等操作。
开发	负责数据开发页面的设计和维护工作。

角色	描述
运维	负责在运维中心页面管理全部任务的运行情况并进行相应处理。
部署	仅在多工作空间模式时审核任务代码并决定是否提交运维。
访客	仅有只读权限,可以查看数据开发页面的业务流程设计和代码内容。
安全管理员	仅有数据保护伞模块的操作权限,详情请参见 <mark>数据保护伞</mark> 。

7. 使用子账号登录DataWorks控制台,单击数据开发页面的DataStudio。

8. 您可以根据业务需求进行数据开发,详情请参见概述。

移除DataWorks工作空间的子账号

- 1. 进入工作空间配置页面。
 - i. 登录DataWorks控制台。
 - ii. 在左侧导航栏, 单击工作空间列表。
 - iii. 单击目标工作空间操作列的工作空间配置。
 - iv. 在工作空间配置对话框,单击更多设置,进入工作空间配置页面。

您也可以进入数据开发页面,单击右上角的图图标,进入工作空间配置页面。

- 2. 在左侧导航栏,单击成员管理。
- 3. 在成员管理页面,单击目标成员操作列的删除。
- 4. 单击确认。

⑦ 说明 子账号被工作空间移除后,仍然拥有Hologres实例的开发权限。

7.6. SQL Console

SQL Console使用SQL语句开发Hologres,为您提供秒级交互式查询体验。本文为您介绍SQL Console的功能和用法。

文件夹

文件夹模块可以存放新建的临时查询,方便您管理临时查询。

单击左侧菜单栏SQL Console > 新建 > 文件夹,创建一个文件夹。您可以在该文件夹里新建临时查询,使用标准的SQL语句完成对表的命令操作。同时您也可以选中文件夹中的某张表,右键单击,即可移动、重命 名、删除该表。



SQL Console

SQL Console模块可以生成SQL编辑器,方便您使用标准SQL语句操作。使用之前您需要先创建文件夹,如果已 经创建文件夹,则直接进行如下步骤。

1. 单击左侧菜单栏SQL Console,并填写查询的基本信息。

6	🛄 Holo Studio		~	
Ш	SQL Console	온 C O		
Q	文件名称/创建人	文件夹		
₽	> #	SQL Console		
60	> #			Welcome to HoloStudio
			Ŧ	开始一个新任务:
				1 2 5 5 5 5 5 5 5 5 5 5 5 5 5
扂]性			说明
ŧ	占名称			临时查询的名称。
E	目标文件夹			临时查询存放的文件夹位置。
娄	文据库			临时查询的目标数据库。

2. 编写SQL语句,单击上方运行,执行查询,完成之后,即可在下方显示框查看到执行结果。示例新建一张 表并导入数据,查询表的全过程。

```
CREATE TABLE supplier (
s_suppkey bigint NOT NULL,
s name text NOT NULL,
s address text NOT NULL,
s nationkey bigint NOT NULL,
s phone text NOT NULL,
s_acctbal bigint NOT NULL,
s comment text NOT NULL,
PRIMARY KEY (s_suppkey)
);
INSERT INTO supplier VALUES
(1, 'Supplier#000000001', 'gf0JBoQDd7tgrzrddZ', 17, '27-918-335-1736', 575594, 'each sly
ly above the careful'),
(6, 'Supplier#00000006', 'tQxuVm7s7CnK', 14, '24-696-997-4969', 136579, 'final accounts
. regular dolphins use against the furiously ironic decoys. '),
(10, 'Supplier#000000010', 'Saygah3gYWMp72i PY', 24, '34-852-489-8585', 389191, 'ing wat
ers. regular requests ar'),
(18, 'Supplier#00000001', 'PGGVE5PWAMwKDZw', 16, '26-729-551-1115', 704082, 'accounts sno
oze slyly furiously bold'),
(39, 'Supplier#000000039', 'SYpEPWrlyAFHaC91qjFcijjeU5eH', 8, '18-851-856-5633 611565',
88990, 'le slyly requests. special packages shall are blithely. slyly unusual packages sl
eep'),
(48, 'Supplier#000000048', 'FNPMQDuyuKvTnLXXaLf3Wl6OtONA6mQlWJ', 14, '24-722-551-9498',5
63062, 'xpress instructions affix. fluffily even requests boos');
SELECT * FROM supplier;
```

\$	🗓 Holo Studio		~ ~	50					
	SQL Console	ନ ⊑ C ⊕	2 co 3 4	6 6					
Q	文件名称/创建人			• C					
=0	> 🛃 Lammada		当前DB:"						
B			14 INSERT 0 16 (1, 'Sup (1, 'Sup (10, 'Su (39, 'Su (39, 'Su (38, 'Su 21 21 (48, 'Su (38, 'Su 22 23 SELECT ≠ 2 SELECT ≠ 1 s.suppkey 2 2 1 3 6 4 10 5 18 6 39 7 48	NTO supplier VALUE oplier#000000001', opplier#000000000', opplier#000000000000000000000000000000000000	S 9 (6) 3BoQDd 7 tg rz rdd 1 tQxu Vm7 3 7Cn K' , 1 'Saygah 3g YMP 72 i 1 'GKO'ESPHAHwKD2w 1 'SY pEPWr1 yA FHa C91 1 'FNPMQD uy uKv Tn LXX 9 (0) JBoQDd 7 tg rzddZ 1 T 1 Gxu Vm 3 7Cn K 14 1 Saygah 3g YWM 72 19 Y24 Pac Vez YWM 72 19 Y24 Pac Ve	Z', 17, '27-918-335-17 14, '24-696-997-4969', PY', 24, '34-852-489- ', 16, '26-729-551-111 qjFcijjeU5eH', 8, '18 XaLf3Wl6OtONA6mQlWJ', XaLf3Wl6OtONA6mQlWJ', s_phone 27-918-335-1736 24-696-997-4969 34-852-489-8585 26-729-951-1115 18-851-856-5633 6 24-722-551-9498	36', 575594, 'e 136579, 'final 8585', 389191, ' 5', 704082, 'acc -851-856-5633 61 14, '24-722-55 14, '24-722-55 × s.acctbal s.575594 136579 389191 704082 11588990 563062	Comment counts. regular co ing waters. regular co ing waters. regular counts snooze slyly fi 11565', 88990, 'le sl i1-9498', 563062, 'xpr comment each slyly above the co final waters. regular ing waters. regular required accounts nooze slyly le slyby requests. speci xpress instructions affi	caret dolphi reque furiou Lyly 1 ress j
扂	計性				说明				
S	QL编辑框				编写SQ	L语句。			
俦	存				保存SQ	L编辑框中的全部	语句。		

属性	说明
	运行SQL编辑框内选中的语句或全部语句。单击运行 后,开始执行SQL语句,运行结果展现在编辑框底部。
运行	⑦ 说明 不选中则默认运行编辑框中所有语句。
重新加载	刷新SQL编辑框的内容。SQL编辑框只保留已保存的语 句。
停止	停止正在运行的SQL语句。
运行日志	查看SQL语句的运行情况。例如,SQL语句的运行时长 以及报错提示。
结果	查看SQL语句运行成功后目标表的内容。

⑦ 说明 : 对无结果返回的SQL语句,只有运行日志,没有查询结果,例如 create table 。

HoloStudio对查询出的表数据可以复制、搜索和隐藏列。

视频操作演示

SQL Console的操作演示视频如下。

7.7. PG管理

7.7.1. 创建及管理内部表

HoloStudio的PG管理模块,支持使用可视化方式创建、查看以及删除内部表。本文为您介绍在HoloStudio中基于PG管理的内部表操作。

前提条件

- 阿里云账号注册,详情请参见。
- 实名认证, 详情请参见或。
- 开通交互式分析Hologres并绑定至DataWorks工作空间,详情请参见DataWorks快速入门。

创建内部表

- 1. 登录Hologres管理控制台, 鼠标悬停至顶部菜单栏左侧的、图标, 选择目标地域。
- 2. 在实例列表 > Hologres引擎管理页面,单击前往DataWorks数仓开发。
- 3. 在顶部菜单栏,单击■▼图标,选择目标DataWorks工作空间。
- 4. 在左侧导航栏,单击 ■图标,进入PG管理界面。
- 5. 鼠标悬停至 图标,单击表,创建表。

您也可以单击HoloStudio首页的新建表创建表。

≡	ntaWorks	Holo Studio	-	•			
Ш	PG管理	C 0					
Q	文件名称/创建人	V.					
≞					Welcome	e to Holc	Studio
s				开始	一个新任务:		
Ð.					2 بنائی بنگیر	新建SQL Console	(3) 新建数版开发

6. 配置编辑表页面的各项参数。

⊞	×								
提交表									
基本属性									
* Hologres数据)	车 请选择		* 表名		描述	8			
字段		分区表							
添加字段									
字段名		数据类型		主键		可空	数组	描述	操作
					5 15	行数据			

类别	参数	描述
	Hologres数据库	存放新建表的数据库。
基本属性	表名	新建表的表名称。
	描述	表的相关描述。
	添加字段	单击即可添加字段。
	字段名	表中每一列的标识。
	数据类型	字段的属性。
	主键	确定列的唯一标识。
	可空	字段可以为空值。
字段	数组	有序的元素序列。
	描述	字段的相关描述。
	操作	包括删除、上移和下移。

类别	参数	描述
	存储模式	包括 列存 和行存两种存储模式。 默认为 列存 。
	生命周期	如果数据在指定时间内未被修改, 则系统将自动删除数据。 默认生命周期为 永久 。
	聚簇索引	排序索引。 索引的类型和列的顺序密切相关。 聚簇索引帮助您加速执行索引列的 Range和Filter查询。
属性	字典编码列	Hologres支持为指定列的值构建字 典映射。 字典编码可以将字符串的比较转换 为数字的比较,加速Group By和 Filter查询。 默认设置所有text列至字典编码列 中。
	位图列	Hologres支持在位图列构建比特编 码。 位图列可以根据设置的条件快速过 滤字段内部的数据。 默认设置所有text列至位图列中。
分区表	PARTITION BY LIST	指定分区字段。

7. 单击**提交表**。

您可以在创建成功的表中进行数据开发。

查看表数据

在左侧导航栏单击PG管理,双击需要查看的表,显示表编辑页面。

≡	巾 DataWorks Holo Studio	••••						ር 🔍 🔻	1000
Ш	PG管理 ☐ C ⊕	*							
Q	文件名称/创建人	提立事						生成の原語句	#01275%
⊒	*								34341735
(1)	✓ 模式	基本属性							
ы	✓ public	*Hologres数据库 *表名 *表名	描述						
	× 表								
		添加字段							
		字段名	数据类型	主鍵	可空	数组	描述		操作
	> 外部表								
	> 📰 📖	SQL Script							
		1							
	> 📑 zptest								

您可以在该页面查看建表语句或预览表数据,如下所示:

• 单击表编辑页面的生成DDL语句,获取建表SQL语句。

DDL			×
1	BEGIN;	Second -	
2	CREATE TABLE "public"."dw_log_soc_engage_rt" (
3	"actiondate" timestamptz NOT NULL,		
4	"gender" text,		
5	"user_id" text,		
6	"taxonomy1" text,		
7	"bind_id" int4 NOT NULL,		
8	"device_id" text,		
9	"frequent_country" text,		
10	"frequent_province" text,		
11	"frequent_city" text,		
12	"year_old" text,		
13	"app_version" text		
14);		
15	CALL SET_TABLE_PROPERTY('"public"."dw_log_soc_engage_rt"', 'clustering_key', '"bind_id	d	
16	CALL SET_TABLE_PROPERTY('"public"."dw_log_soc_engage_rt"', 'segment_key', '"actiondate	DataWorks	
47	CALL CET TADLE DRODEDTV//"mublic" "du log coe ongogo nt"! 'hitmon columne' '"gondon'	•	
		确认取	消

• 单击表编辑页面的数据预览, 预览表数据。

数	据预	览								×
		A	В	С	D	E	F	G	Н	
	1	actiondate	gender	user_id	taxonomy1	bind_id	device_id	frequent_cc	frequent_pr	
	2	2020-09-28	female	672-27-266	Oin	0	915362440	Latvia	山西	
	3	2020-09-28	female	251- 9 8-194	Bifur	0	443252571	Zimbabwe	重庆 前	
	4	2020-09-28	female	520-84-787	Bard the Bo	0	537723006	Comoros	香港	
	5	2020-09-28	female	520-84-787	Bard the Bo	0	537723006	Comoros	香港	
	6	2020-09-28	female	520-84-787	Bard the Bo	0	537723006	Comoros	香港	
	7	2020-09-28	female	520-84-787	Bard the Bo	0	537723006	Comoros	香港	
;										
								-	议 取	消

删除内部表

如果您需要删除已创建的内部表,则在PG管理界面,鼠标右键需要删除的表,单击**删除表**即可。



7.7.2. 创建及管理外部表

外部表是指不存储数据于交互式分析Hologres中的表,仅用于查看外部表的数据。HoloStudio的PG管理模块,支持使用可视化方式创建、查看以及删除外部表。本文为您介绍在HoloStudio中基于PG管理的外部表操作。

前提条件

- 阿里云账号注册,详情请参见。
- 实名认证,详情请参见或。
- 开通交互式分析Hologres并绑定至DataWorks工作空间,详情请参见DataWorks快速入门。

背景信息

Hologres与大数据生态无缝打通,可以直接加速查询外部表数据,也可以将外部表的数据导入至Hologres中处理。Hologres当前仅支持对MaxCompute表进行操作。

使用限制

- Hologres支持跨工作空间读取外部表数据,您当前使用的账号需要拥有其他工作空间的MaxCompute表访问权限,才可以使用HoloStudio一键创建外部表直接读取MaxCompute表数据。
- 当MaxCompute的源头表存在时,您才可以新建外部表并读取外部表数据。
- Hologres中的外部表与MaxCompute的源头表字段一一对应,您可以直接搜索MaxCompute中的表名称,HoloStudio会自动生成对应字段的外部表。

创建外部表

- 1. 登录Hologres管理控制台, 鼠标悬停至顶部菜单栏左侧的、图标, 选择目标地域。
- 2. 在实例列表 > Hologres引擎管理页面,单击前往DataWorks数仓开发。
- 3. 在顶部菜单栏,单击 🔩 图标,选择目标DataWorks工作空间。
- 4. 在左侧导航栏,单击 ■图标,进入PG管理界面。
- 5. 鼠标悬停至 图标,单击**外部表**,创建外部表。
- 6. 配置编辑表页面的各项参数。



类别	参数	描述			
甘木屋州	Hologres数据库	存放新建外部表的数据库。			
本 平周日	表名	新建外部表的表名称。			
	类型	外部表的服务类型。目前仅支持 MaxCompute表。			
外部服务	表	需要映射的MaxCompute源头表。			
	字段同步	需要同步的字段。			
	分区同步	需要同步的分区。			

7. 单击提交表。

查看外部表数据

在左侧导航栏单击PG管理,双击需要查看的外部表,显示表编辑页面。

≡	ntataWorks Holo Studio		■		¢ & ▼					
Ш	PG管理 □ C O		×							
Q	文件名称/创建人	坦六主			H ctopu 等合 数据系数					
⊒	✓ ■ hitest	TECCAR								
(1)	✓ 模式	基本属性 —								
ы	✓ public	● Hologres数据	库 hitest * 表名 _public.hitest1_f1							
	> 表	外部服务 —	外部服务							
	◇ 外部表	■ 类型 MaxCo	mpute odps_server	*表 odps.hologresdemo4rh.bank_data_odp	\$					
	public.hitest1_f1	字段同步	分区同步							
	> 2			类型	描述					
	`		808	binint	年龄					
			lop	string						
			marital	string	婚否					
		SQL Script								

您可以在该页面查看建表语句或预览表数据,如下所示:

• 单击表编辑页面的生成DDL语句,获取建表SQL语句。

DDL			×
1	BEGIN;		
2	CREATE FOREIGN TABLE "public"."hitest1_f1" (
З	"age" int8,	and a second sec	
4	"job" text,		
5	"marital" text,		
6	"education" text,		
7	"card" text,		
8	"housing" text,		
9	"loan" text,		
10	"contact" text,		
11	"month" text,		
12	"day_of_week" text,		
13	"duration" text,		
14	"campaign" int8,		
15	"pdays" float8,		
16	"previous" float8,		
4 7	"noutcome" tout		
		間以即	<i>ì</i> 自

● 单击表编辑页面的**数据预览**, 预览表数据。

MN2	数据预	谠												×
			Α		В	С	D	E	F	G	i	н		
	1	age		job		marital	education	card	housing	loan		contac	t I	
											đ	汄	取	肖

删除外部表

如果您需要删除已创建的外部表,则在PG管理界面,鼠标右键需要删除的表,单击**删除表**即可。



7.8. 数据开发

7.8.1. 概述

HoloStudio的数据开发模块与DataWorks无缝连接,提供一站式稳定高效的ETL(Extract-Transform-Load) 服务,帮助您使用可视化方式周期性调度作业、创建外部表、同步外部表数据以及上传本地文件。

数据开发的核心功能如下:

• 文件夹

用于存放并管理数据库中的数据开发节点。详情请参见文件夹。

• Hologres开发

用于周期性调度作业。详情请参见Hologres开发:周期性调度。

● 一键MaxCompute表结构同步
支持使用可视化方式批量创建外部表,加速查询MaxCompute的数据。详情请参见一键同步MaxCompute表结 构。

● 一键MaxCompute数据同步

支持使用可视化方式导入MaxCompute的数据至Hologres。详情请参见一键同步MaxCompute数据。

• 一键本地文件上传

用于上传客户端数据至Hologres。详情请参见一键上传本地文件。

7.8.2. Hologres开发: 周期性调度

HoloStudio与DataWorks无缝连通,您可以通过HoloStudio将MaxCompute数据导入Hologres,并基于 DataWorks的底层能力,前往DataWorks进行定时调度,实现周期性导入数据至Hologres。本文为您介绍如何 将MaxCompute源表数据导入Hologres进行周期性调度。

操作步骤

1. 准备MaxCompute表数据。

准备一张MaxCompute数据源表,您可以参考MaxCompute创建表进行建表,也可以直接从数据地图中选用一张表。示例选用数据地图中已创建的表,其建表DDL语句如下。

```
CREATE TABLE IF NOT EXISTS bank data odps
(
                BIGINT COMMENT '年龄',
age
                STRING COMMENT '工作类型',
job
marital
                STRING COMMENT '婚否',
 education
                STRING COMMENT '教育程度',

    card
    STRING COMMENT '発育程度',

    housing
    STRING COMMENT '是否有信用卡',

    loan
    STRING COMMENT '贷款',

contact STRING COMMENT '联系途径',
month STRING COMMENT '月份',
day_of_week STRING COMMENT '星期几',
duration STRING COMMENT '持续时间',
              BIGINT COMMENT '本次活动联系的次数',
DOUBLE COMMENT '与上一次联系的时间间隔',
campaign
 pdays
                DOUBLE COMMENT '之前与客户联系的次数',

    previous
    DOUBLE COMMENT '之前与客户联系的次数

    poutcome
    STRING COMMENT '之前市场活动的结果',

emp var rate DOUBLE COMMENT '就业变化速率',
cons_price_idx DOUBLE COMMENT '消费者物价指数',
 cons conf idx DOUBLE COMMENT '消费者信心指数',
euribor3m DOUBLE COMMENT '欧元存款利率',
nr employed DOUBLE COMMENT '职工人数',
            BIGINT COMMENT '是否有定期存款'
V
);
```

2. 新建外部表。

进入HoloStudio,在PG管理或SQL Console中新建一张外部表,用于映射MaxCompute源表数据。示例新 建外部表的SQL语句如下。

DataWorks公共云合集·HoloStudio

```
begin;
CREATE FOREIGN TABLE "public"."bank_data_foreign_holo" (
"age" int8,
"job" text,
 "marital" text,
 "education" text,
 "card" text,
"housing" text,
 "loan" text,
 "contact" text,
 "month" text,
 "day of week" text,
 "duration" text,
 "campaign" int8,
 "pdays" float8,
 "previous" float8,
"poutcome" text,
 "emp var rate" float8,
 "cons price_idx" float8,
 "cons conf idx" float8,
"euribor3m" float8,
"nr employed" float8,
"y" int8
)
SERVER odps_server
OPTIONS (project name 'odps 4 holoworkshop dev', table name 'bank data odps');
COMMIT;
```

OPTIONS的连接参数说明如下表所示。

参数	描述
project_name	MaxCompute的项目名称。
table_name	MaxCompute的表名称。

3. 新建存储表。

在HoloStudio中新建一张用于接收并存储数据的真实存储表。字段类型需要与外部表保持一致。示例DDL 语句如下。 begin; create table if not exists bank_data_holo (age int8, job text, marital text, education text, card text, housing text, loan text, contact text, month text, day of week text, duration text, campaign int8, pdays float8, previous float8, poutcome text, emp var rate float8, cons price idx float8, cons conf idx float8, euribor3m float8, nr employed float8, y int8, ds text not null) partition by list(ds); call set_table_property('bank_data_holo', 'orientation', 'column'); call set_table_property('bank_data_holo', 'time_to_live_in_seconds', '700000'); commit;

4. 新建分区表的数据开发。

并单击上方新建Hologres开发,输入作业逻辑,单击保存--前往DataWorks调度,示例SQL如下:

- i. 在HoloStudio中,单击左侧导航栏的圆图标,进入数据开发页面。
- ii. 鼠标悬停至□图标,单击Hologres开发。
- iii. 配置新建节点对话框的节点名称、目标文件夹及数据库参数。

新建节点		×
节点名称	节点名称	
目标文件夹	请选择	
数据库	请选择	
	l	提交取消

iv. 单击提交, 成功创建Hologres开发节点。

v. 在新建的Hologres开发节点的编辑界面, 输入如下创建分区表的数据开发语句。

```
create table if not exists bank_data_holo_1_${bizdate} partition of bank_data_holo
 for values in ('${bizdate}');
insert into bank_data_holo_1_${bizdate}
select
   age as age,
   job as job,
   marital as marital,
   education as education,
   card as card,
    housing as housing,
   loan as loan,
   contact as contact,
   month as month,
   day of week as day of week,
    duration as duration,
   campaign as campaign,
   pdays as pdays,
   previous as previous,
   poutcome as poutcome,
    emp_var_rate as emp_var_rate,
   cons price idx as cons price idx,
   cons conf idx as cons conf idx,
   euribor3m as euribor3m,
   nr_employed as nr_employed,
   y as y,
   '${bizdate}' as ds
from bank_data_foreign_holo;
```

- vi. 单击顶部菜单栏左侧的回,保存Hologres开发。
- vii. 单击顶部菜单栏右侧的前往DataWorks调度,进入调度页面。

Ш	N225开发 온다. C €	×	
۹	文件名称/创建人		lorks ill .S
==	~ A	1100	
5		1	
2		* The server of process schema public_data limit to (dod_product_movie_basic_info) from server odps_server into hmads options(if_table_exist 'update'); 6 - 中所的名称和 Selver(a, UDAH-logres, UDAH-logres, MCLAREAREAREAREAREAREAREAREAREAREAREAREAREA	
		14 "area", 15 "aretors", 16 "type", 17 "movie_longth", 18 "movie_longuege", 28 "indu_url, 21 "gs", 22 "RRM Typull:"resign.dwd_product_movie_basic_info 23 WHER & ds'{bitateij; 24 WHER & ds'{bitateij; 25 WHER & ds'{bitateij; 26 WHER & ds'{bitateij; 27 WHER & ds'{bitateij; 28 WHER & ds'{bitateij; 29 WHER & ds'{bitateij; 21 WHER & ds	

5. 新建分区表的调度作业。

i. 在DataWorks中新建Hologres节点。

单击前往DataWorks调度后,页面自动跳转至DataWorks的新建节点。您需要配置新建节点对话框的节点类型、节点名称及目标文件夹,创建Hologres节点。节点类型选择Hologres开发。

新建节点		×
节点类型:	Hologres开发	
节点名称:	周期性调度	
目标文件夹:	业务流程/test	
		提交取消

如果是重新执行已有作业,则可以不需要重新创建节点。

ii. 在新建的节点页面,单击更新节点版本,将分区表的信息同步至该节点。



6. 配置调度信息。

在新建的节点页面,单击右侧导航栏的**调度配置**,配置调度参数。详情请参见<mark>调度参数概述</mark>。 具体配置如下所示:

i. 配置基础属性。

在基础属性区域,将参数赋值为时间节点。

× 调度配置		
基础属性		
名称:	unitaria di Carta di	<u>h</u>
节点ID :		· · · · · · · · · · · · · · · · · · ·
节点类型:	Hologres开发	
* 责任人 :		版本
描述:		
參数:	\$(bizdate)=\$(yyymmdd) ⑦	

ii. 配置时间属性。

在**时间属性**区域,配置**时间属性为正常调度**。其余参数请根据业务实际情况配置,详情请参见时间 属性配置说明。

时间属性		
* 生成实例方式:	○ T+1次日生成 ⑦ ● 发布后即时生成 ⑦ 注:暂不支持组合类节点(循环、PAI)	
* 时间属性:	● 正常调度 ● 空跑调度	
* 重跑属性:	运行成功后不可重跑,运行失败后可以重跑	3
出错自动重跑:		
* 生效日期:	1970-01-01 🟥	
	注:调度将在有效日期内生效并自动调度,反之,在有效期外的任务将不会	自动调度。
暂停调度:		
* 调度周期:	B ~ (3
* 定时调度时间:	00:09	
	注:默认调度时间,从0点到0点30分随机生成	
cron表达式:	00 09 00 * * ?	
* 超时时间:	● 系统默认 ● 指定时间 小时	
	注:任务运行时间超过超时时间后,任务将被系统终止。可指定超时时间最大	大为 72 小时。
依赖上一周期:		

iii. 配置调度依赖。

在调度依赖区域,您需要配置调度依赖为root节点,操作如下:

- a. 将自动解析配置为是。
- b. 单击使用工作空间根节点,自动解析出root节点。
- c. 再将自动解析配置为否,详情请参见配置同周期调度依赖。

・ 調度依赖 ⑦ 自动解析 ○ 是 ● 否 解析能入範出 依赖的上游节点 諸約人父节点输出名称或输出表名 >> + 使用工作空间很节点 自动推荐											
父节点输出名称	root		名称	root		责任人	来源	操作			
本节点的输出	请输入节点输出名称										
-	171000000.44		- Ø				系统默认添加				
			- Ø				手动添加	删除			

您也可以根据业务逻辑选择已有的父节点。

- 7. 发布调度。
 - i. 配置完调度参数,在新建节点页面的顶部菜单栏左侧,单击3图标,保存调度配置。
 - ii. 在新建节点页面的顶部菜单栏左侧,单击团图标,提交调度信息。
 - iii. 在新建节点页面的顶部菜单栏右侧,单击发布,前往任务发布页面。

- iv. 在任务发布页面,选中当前任务,单击目标任务操作列的发布。发布包提交成功后,单击上方菜单 栏运维中心进行手工补数据。
- v. 在确认发布对话框,单击发布。

确认发布	×
节点新增或调度依赖变更22-00前发布完成,周期节点运维第二天才会生效	
周期性调度_2020-11-04_hologres_admin	
查看发布包详情	
发布	关闭

- vi. 发布包提交成功后,单击任务发布页面顶部菜单栏右侧的运维中心,进入运维中心。
- vii. 在左侧导航栏,单击周期任务运维 > 周期任务,勾选并单击目标任务名称。
- viii. 在右侧编辑界面, 鼠标右键单击目标任务节点, 选择**补数据 > 当前节点**。

G	运维中心(工作流)		× -	F							- 0 ×
←	→ C ☆	🗎 workb	ench2-c	n-hangzhou.data	.aliyun.com/?env	=prod#/cycleTas	k				☆ \varTheta :
	iCMS 📙 环境	. 工作链接	妾 📙 学	习链接 📒 语催	素求链接	官网链接					其他书签
≡	帅 DataWo	orks丨运维	中心	11111	•					& DataStudio 🧕	📍 🛛 🔍 👻 🛶 🚦 ፤
e	运维大屏										
\$\$\$	实时任务运维	~	搜索	节点名称/节点ID	Q 节点类型	: 请选择节点类型	❤ 妻任人:		✓ 调度资源	1 请选择 🗸 🗹 我的	节点 今日修改的节点
ц	周期任务运维	^	8	停 (冻结) 节点	孤立节点 过ぎ	時点					
	周期任务										○周新│展开搜索
	周期实例		-	名称		节点D					
	补数据实例		-	PETRO ALL 2 PETRO		100007			生产外境,	请 俚俱 操 作	
	测试实例			同类们生调度		100067					
ନ	手动任务运维	~				100061		holo workshop	屋井小井占		
᠕	智能监控	~			۲	100054		此间。 虚节点	展开子节点	>	
ራ	智能诊断	۵				10005\$			节点详情		
				-		1000Å5 [»]	C	■期件:圖度	查看代码 编辑表达		
								Hologres开发	唐福 D.杰 查看实例		
									查看血缘		
									测试	当前节点	
									*1903周 智停 (冻结)	> 当前节点及下游节点	
_			更多	▼1	/1 >				恢复 (解冻)	海量节点模式	
						<u> </u>			-		15:50
) 在这里輸	心尔要搜索	的内容			· ·		/ 😢 🖉	<u></u>	~ 🕥 🕾	및 Q× 中 🔁 2020/11/4 😨

ix. 根据业务实际情况, 配置补数据对话框的各项参数。完成调度任务的发布。

补数据		×
* 补数据名称:	P_周期性调度_20201104_155246	
* 选择业务日期:	2020-11-03 - 2020-11-03	
* 当前任务:	周期性调度	
* 是否并行:	不并行 🗸	
4		Þ
		确定 取消

8. 任务执行成功后,返回HoloStudio。单击新建节点的 图标,刷新当前页面。您可以在PG管理模块找到目标分区表,查看调度成功的分区表数据。

111															
Q															
≣															
		基本属性													
33															
2															
				数据预	页览								×		
	public.bank_data_holo														
	m public			1	A age	B	C marital	D education	E	F housing	G	H	1		
				2	26	blue-collar	single	high.scho	olno	no	yes	cellular	1		
				3	40	admin.	single	university.	d no	no	no	telephone			
				4	30	admin.	single	university.	d no	no	no	cellular			
				6	32	services	married	basic.9y	unknown	yes	no	cellular			
				7	53	blue-collar	married	basic.6y	unknown	unknown	unknown	cellular			
				;											
												m认 E	R.Wi		

双击目标表名称,在表编辑页面单击数据预览,即可看到已成功导入的数据。

您也可以使用SQL语句查看父表或分区表的数据,示例如下。

SELECT * FROM bank_data_holo;

7.8.3. 一键同步MaxCompute表结构

本文为您介绍如何使用HoloStudio批量创建外部表。

前提条件

- 阿里云账号注册,详情请参见。
- 实名认证,详情请参见或。
- 开通交互式分析Hologres并绑定至DataWorks工作空间,详情请参见DataWorks快速入门。

背景信息

Hologres与MaxCompute在底层无缝连接,支持使用新建外部表的方式加速查询MaxCompute的数据,详情 请参见通过创建外部表加速查询MaxCompute数据。 HoloStudio支持一键MaxCompute表结构同步功能,您可以使用可视化方式批量创建外部表。

您也可以使用 IMPORT FOREIGN SCHEMA 语句批量创建外部表。也可以将该命令语句在数据开发中执行,并前 往调度,实现MaxComput e表新增后Hologres外表也自动新增,详情见Hologres开发:周期性调度。

⑦ 说明 Hologres只能加速查询MaxCompute的内表,不能加速查询MaxCompute的外表和View。

操作步骤

- 1. 登录阿里云官网。
- 2. 新建一键MaxCompute表结构同步节点。
 - i. 进入Hologres管理控制台。
 - ii. 鼠标悬停至顶部菜单栏左侧的 · 图标,选择目标地域。
 - iii. 在实例列表 > Hologres引擎管理页面,单击前往DataWorks-HoloStudio开发。
 - iv. 在Holo Studio界面的左侧导航栏,单击 🔤 图标。
 - v. 进入数据开发页面, 鼠标悬停至顶部菜单栏的 图标。
 - vi. 单击一键MaxCompute表结构同步。
 - vii. 在新建节点对话框中, 输入节点名称, 并选择目标文件夹和数据库。
 - viii. 单击提交。
- 3. 在编辑节点页面, 配置各项参数。

\$	🛄 Holo Studio		■ ~		થ	V
	数据开发	ନ୍ଟ ⊑ C ⊕	• • ×			
Q	文件名称/创建人		■ 6 ⊙	С		
	▼ ■ ▼ ■ ● ■ ● ■ ● ■		基本信息 目标库 目标运行 通标运行表 边标运行表 远程服务器 ● 1 远程服务器 ● 2 近程服务器 ● 2 远程服务器 ● 3 素名规则 *		postgres 南边塔 > 南边塔 > 南边塔 > 南边塔 > 南边塔 > 南边塔 > 小 ● <td< th=""><th></th></td<>	
			正则预览		正则结果预这	
	◎数				描述	
E	目标库				Hologres的数据库名称。	
E	目标Schema				当前数据库的Schema名称。 如果您没有新建Schema,则只能选择默认创建的 public 。如身 Schema,您也可以选择新建的Schema。	果有新建的

参数	描述
远程服务类型	默认为 odps 。 目前仅支持MaxCompute。
远程服务器	您可以直接调用Hologres底层已创建的名为 odps_server 的远程服务 器。详细原理请参见 <mark>Postgres FDW</mark> 。
远程库	MaxCompute的项目名称。
表名规则	您可以使用正则表达式选择需要同步的表名称,默认同步远程数据库中的 所有表。更多筛选外部表的规则请参见IMPORT FOREIGN SCHEMA。
正则预览	查看正则表达式的运行结果。

同步外部表的规则如下:

- 如果同步外部表时存在名称相同的表,则忽略当前表。
- 如果同步外部表时,存在外部表不支持的数据类型,则系统会报错请在正则表达式中排除当前表再重新导入。
- 4. 在新建节点的编辑页面, 单击 🛽 图标, 保存配置信息。
- 5. 在新建节点的编辑页面,单击回图标,批量创建外部表结构。
- 6. 在PG管理页面,查看批量创建的外部表并查询表数据,详情请参见创建及管理外部表。

7.8.4. 一键同步MaxCompute数据

本文为您介绍如何使用HoloStudio导入MaxCompute数据至Hologres。

前提条件

- 阿里云账号注册,详情请参见。
- 实名认证,详情请参见或。
- 开通交互式分析Hologres并绑定至DataWorks工作空间,详情请参见DataWorks快速入门。

背景信息

HoloStudio支持一键MaxCompute数据同步功能,您可以使用可视化方式导入MaxCompute表数据并进行 查询。该方式比创建外部表直接查询数据的性能更好。

您也可以使用SQL语句导入MaxCompute数据,详情请参见使用SQL导入MaxCompute的数据至Hologres。

操作步骤

- 1. 登录阿里云官网。
- 2. 新建一键MaxCompute数据同步节点。
 - i. 进入Hologres管理控制台。
 - ii. 鼠标悬停至顶部菜单栏左侧的 · 图标,选择目标地域。
 - iii. 在实例列表 > Hologres引擎管理页面,单击前往DataWorks-HoloStudio开发。

- iv. 在Holo Studio页面的左侧导航栏,单击 ፼图标。
- v. 进入数据开发页面, 鼠标悬停至顶部菜单栏的 图标。
- vi. 单击一键MaxCompute数据同步。
- vii. 在新建节点对话框中, 输入节点名称, 并选择目标文件夹和数据库。
- viii. 单击提交。
- 3. 在编辑节点页面,配置各项参数。

6	Holo Studio	,			থ	
	хжяж 2 □ С ⊕	x			·	
Q	文件名称/创建人	C ≙ ⊙ C				
⊒⊙	× ≞ II	MaxCompute源表选择				
(/) [2]	 > ■ • ■ • ■ ■ ■ ■ ■ ■ 	外部表来源 外部表表名字 目标表设置 目标库	 已有外部表 新建外部 请选择 Interi001 	凄		
		 Frouchema 目标表描述 日标表描述 同步设置 回步字段 分区配置 委引 				
		▲ 名称		类型	描述	
		SQL Script				

参数	描述
外部表来源	 已有外部表:表示在Hologres中已经建立 MaxCompute数据映射的外部表。 新建外部表:表示外部表在Hologres中未建立 MaxCompute数据映射。如果您需要导入 MaxCompute表数据,请选中新建外部表。
外部表表名字	已创建的外部表表名称。 Hologres中创建的外部表用于映射MaxCompute数 据,需要与导入数据的MaxCompute表结构一一对应。
目标库	Hologres的数据库名称。
目标Schema	当前数据库的Schema名称。 如果您没有新建Schema,则只能选择默认创建 的 public 。如果有新建的Schema,您也可以选择新建 的Schema。
目标表名	需要导入数据的表名称。
目标表描述	目标表的信息描述。

参数	描述
同步字段	需要导入的MaxCompute表字段。 您可以选择导入部分或全部字段。
分区配置	需要导入的分区字段。 一键MaxCompute数据同步功能仅支持导入一级分 区。如果您需要导入多级分区,请在SQL Console中 使用SQL语句实现,详情请参见SQL Console。
索引配置	您可以为目标表创建索引,详情请参见 <mark>设置表属性</mark> 。
SQL Script	自动解析当前可视化操作对应的SQL语句。

- 4. 在新建节点的编辑页面,单击 图标,保存配置信息。
- 5. 在新建节点的编辑页面,单击 回图标,导入MaxCompute表数据。
- 6. 在PG管理页面,查看导入的外部表数据,详情请参见创建及管理外部表或SQL Console。

7.8.5. 一键上传本地文件

本文为您介绍如何使用HoloStudio导入客户端数据至Hologres。

前提条件

- 阿里云账号注册,详情请参见。
- 实名认证,详情请参见或。
- 开通交互式分析Hologres并绑定至DataWorks工作空间,详情请参见DataWorks快速入门。

背景信息

HoloStudio支持一键本地上传文件功能,您可以使用可视化方式导入客户端数据。 您也可以使用 COPY 命令导入客户端数据,详情请参见使用COPY命令导入或导出本地数据。

操作步骤

- 1. 登录Hologres管理控制台。
- 2. 鼠标悬停至顶部菜单栏左侧的、图标,选择目标地域。
- 3. 在实例列表 > Hologres引擎管理页面,单击前往DataWorks-HoloStudio开发。
- 4. 新建内部表。

使用**SQL Console**新建一张用于接收客户端数据的内部表,详情请参见<mark>SQL Console</mark>。示例SQL语句如下。

BEGIN; CREATE TABLE if not EXISTS holo_bank (age int8, job text, marital text, education text, card text, housing text, loan text, contact text, month text, day of week text, duration text, campaign int8, pdays float8, previous float8, poutcome text, emp_var_rate float8, cons price idx float8, cons conf idx float8, euribor3m float8, nr employed float8, y int8); COMMIT;

您也可以使用PG管理新建内部表,详情请参见创建及管理内部表。

5. 新建一键本地上传文件。

i. 在Holo Studio界面的左侧导航栏,单击 🔤 图标。

- ii. 进入数据开发页面, 鼠标悬停至顶部菜单栏的 图标。
- iii. 单击一键本地上传文件。

iv. 在一键本地文件上传对话框中, 配置各项参数。

一键本地文件上传						×
目标库 *	请选择					
目标Schema *	请选择					
选择要导入的数据表*	请选择					
名称		类型		描述		
		没有	数据			
					下一步	Í
参数			描述			
目标库			Hologres的数排	居库名称。		
目标schema			当前数据库的Sd 如果您没有新建 的 public 。如身 建的Schema。	chema名称。 经Schema,则只 果有新建的Sche	!能选择默认创建 ma,您也可以选择	新
选择要导入的数据表			需要导入数据的	〕表。		

v. 单击下一步。

vi. 在一键本地文件上传对话框中, 配置各项参数。

一键本地文件上传		×
洗择文件 *		浏览
····································	· · · · · · · · · · · · · · · · · · ·	
原始字符集 *	GBK	
首行为标题		
参数		描述
		需要上传的本地文件。
选择文件		仅支持上传TXT、CSV和LOG类型的文件。
		■ 逗号
		■ Tab
		■ 分号
		■ 空格
选择分隔符		- 1
		■ #
		• &
		您也可以自定义分隔符。
		■ GBK
		■ UTF-8
原始字符集		■ CP936
		■ ISO-8859
首行为标题		勾选则设置首行数据为标题。

vii. 单击提交。

6. 在SQL Console或PG管理页面,查看已上传的数据。详情请参见SQL Console或创建及管理内部表。

7.9. 终端

HoloStudio的终端是基于SQL语句的psql客户端,支持实时写入及查询高并发的海量数据,您可以使用SQL语 句进行交互式分析Hologres的开发。本文为您介绍Hologres如何创建及使用终端。

前提条件

- 阿里云账号注册,详情请参见。
- 实名认证,详情请参见或。
- 开通交互式分析Hologres并绑定至DataWorks工作空间,详情请参见DataWorks快速入门。

新建终端

- 1. 登录Hologres管理控制台。
- 2. 鼠标悬停至顶部菜单栏左侧的 · 图标,选择目标地域。
- 3. 在实例列表 > Hologres引擎管理页面,单击前往DataWorks数仓开发。
- 4. 在顶部菜单栏,单击 型图标,选择目标DataWorks工作空间。
- 5. 在左侧导航栏, 单击回图标, 进入终端界面。
- 6. 单击 图标, 配置新建终端对话框的各项参数。

新建终端			
	终 端类型 终端名称	Holo Terminal	✓ #交
参数			描述
终端类型			默认为Holo Terminal。
终端名称			自定义。例如 <i>holo</i> 。

7. 单击提交。

使用终端

- 1. 在终端页面,单击目标终端操作列的◎图标。
- 2. 在启动终端对话框,选择目标数据库。

启动终端					×
	数据库				
				确认	取消

3. 单击确认。

您可以使用SQL语句进行Hologres开发,示例语句如下。

```
BEGIN;
CREATE TABLE holo_bank_data (
age int8,
job text,
marital text,
education text,
card text,
housing text,
loan text,
contact text,
month text,
day of week text,
duration text,
campaign int8,
pdays float8,
previous float8,
poutcome text,
emp var rate float8,
cons price idx float8,
cons conf idx float8,
euribor3m float8,
nr employed float8,
y int8
);
COMMIT;
insert into holo bank data values
(53, 'technician', 'married', 'unknown', 'no', 'no', 'no', 'cellular', 'nov', 'fri',
138, 1, 999, 0, 'nonexistent', 1.22, 93.20, 42.00, 4.02, 5195.80, 0),
(28, 'management', 'single', 'university.degree', 'no', 'yes', 'no', 'cellular', 'jun'
, 'thu', 339, 3, 6, 2, 'success', 1.69, 94.05, 39.79, 0.72, 4991.60, 1),
(39, 'services', 'married', 'high.school', 'no', 'no', 'no', 'cellular', 'apr', 'fri'
, 185, 2, 999, 0, 'nonexistent', 1.80, 93.07, 47.10, 1.40, 5099.10, 0);
select * from holo bank data;
```

≡	🍿 DataWo	orks Holo S	Studio	· · · · · · · · · · · · · · · · · · ·	় ব	s 💎	
Ш	终端	G	PC &	W ×			
Q	请输入终端名称	搜索		'fri', 138, 1, 999, 0, 'nonexistent', 1.22, 93.20, 42.00, 4.02, 5195.8			
⊒	名称	负责人	操作	7, o', (26, 'management', 'single', 'university.degree', 'no', 'yes', 'no', 'cellula r', 'jun', 'thu', 339, 3, 6, 2, 'success', 1.69, 94.05, 39.79, 0.72, 49			
S			⊙ 🖞	91.00, 1), (39, 'services', 'married', 'high.school', 'no', 'no', 'no', 'cellular', 'apr ', 'fri', 185, 2, 999, 0, 'nomexistent', 1.80, 93.07, 47.10, 1.40, 5099 .10. 01:			
Ŀ	M	-	⊙ 🛱	INSERT 0 3 Time: 59.199 ms			
	M		💿 🖻	rhdeno =#			
	M	-	 ⊡ 	select * from holo_bank_data; age job marital education card housing loan contact month day_of_week duration c poutcome emp_war_tact cons_price_idx cons_conf_idx euribor%m nr_employed y	campaign	pdays	previous
	< 1					+	
				53 technician married unknown no no no cellular nov fri 138 nonexistent 1.22 93.2 42 4.02 5195.8 0		999	
				28 management single university.degree no yes no cellular jun thu 339 success 1.69 94.05 39.79 0.72 4991.6 1		6	2
				39 services married high.school no no cellular apr fri 185 nonexistent 1.8 93.07 47.1 1.4 5099.1 0	2	999	
				(3 томя)			
				Time: 22.106 ms			

删除终端

如果您需要删除已创建的终端,则单击目标终端操作列的图图标,在弹出的删除对话框单击确认即可。

