# 阿里云
# 应用高可用服务

## 故障演练

文档版本：20190625

# 法律声明

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

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的”现状“、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含”阿里云”、Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

# 通用约定

| 格式 | 说明 | 样例 |
| --- | --- | --- |
|  | 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。 |  **禁止:**<br>重置操作将丢失用户配置数据。 |
|  | 该类警示信息可能导致系统重大变更甚至故障，或者导致人身伤害等结果。 |  **警告:**<br>重启操作将导致业务中断，恢复业务所需时间约10分钟。 |
|  | 用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。 |  **说明:**<br>您也可以通过按Ctrl + A选中全部文件。 |
| > | 多级菜单递进。 | 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。 | 单击 确定。 |
| `courier` 字体 | 命令。 | 执行 `cd /d C:/windows` 命令，进入Windows系统文件夹。 |
| *## * | 表示参数、变量。 | `bae log list --instanceid` *`Instance_ID`* |
| []或者[a\|b] | 表示可选项，至多选择一个。 | `ipconfig` *`[-all\|-t]`* |
| {}或者{a\|b} | 表示必选项，至多选择一个。 | `swich` *`{stand \| slave}`* |

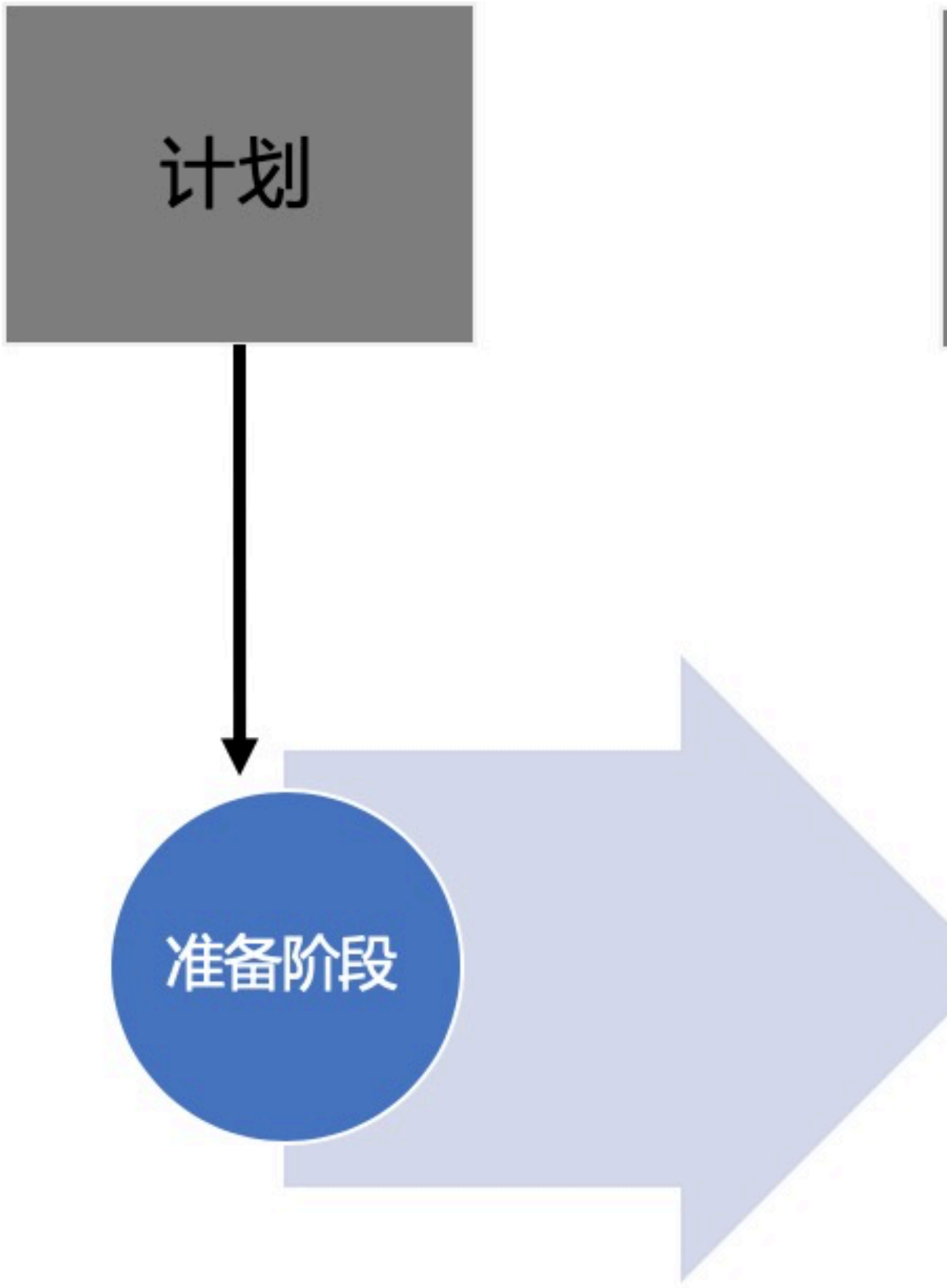
# 目录

# 1 故障演练概述

故障演练是一款遵循混沌工程实验原理并融合了阿里巴巴内部实践的产品，提供丰富故障场景实现，能够帮助分布式系统提升容错性和可恢复性。

## 流程

故障演练建立了一套标准的演练流程，包含准备阶段、执行阶段、检查阶段和恢复阶段。通过四阶段的流程，覆盖用户从计划到还原的完整演练过程，并通过可视化的方式清晰的呈现给用户。


计划
准备阶段

## 适用场景

故障演练可适用于以下典型场景：

- 衡量微服务的容错能力

  通过模拟调用延迟、服务不可用、机器资源满载等，查看发生故障的节点或实例是否被自动隔离、下线，流量调度是否正确，预案是否有效，同时观察系统整体的 QPS 或 RT 是否受影响。在此基础上可以缓慢增加故障节点范围，验证上游服务限流降级、熔断等是否有效。最终故障节点增加到请求服务超时，估算系统容错红线，衡量系统容错能力。

- 验证容器编排配置是否合理

  通过模拟杀服务 Pod、杀节点、增大 Pod 资源负载，观察系统服务可用性，验证副本配置、资源限制配置以及 Pod 下部署的容器是否合理。

- 测试 PaaS 层是否健壮

  通过模拟上层资源负载，验证调度系统的有效性；模拟依赖的分布式存储不可用，验证系统的容错能力；模拟调度节点不可用，测试调度任务是否自动迁移到可用节点；模拟主备节点故障，测试主备切换是否正常。

- 验证监控告警的时效性

  通过对系统注入故障，验证监控指标是否准确，监控维度是否完善，告警阈值是否合理，告警是否快速，告警接收人是否正确，通知渠道是否可用等，提升监控告警的准确和时效性。

- 定位与解决问题的应急能力

  通过故障突袭，随机对系统注入故障，考察相关人员对问题的应急能力，以及问题上报、处理流程是否合理，达到以战养战，锻炼人定位与解决问题的能力。

## 故障演练与 AHAS 服务体系

故障演练作为 AHAS 的一部分，与 AHAS 其他功能组成了一套完善的高可用保障服务，可以帮助用户实现包括架构、业务、人员的全面高可用提升。故障演练在其中承担着问题发现、问题验证、高可用经验沉淀的作用。


控制台
有效衡量目标
架构高可用
防御式编程
可观察架构
业务高可用
SLA评估
资损预防
人员高可用
监控覆盖度

# 2 名词解释

使用故障演练前，您需要先了解相关名词。

| 名词 | 名词解释 |
| --- | --- |
| 演练 | 一个完整的、可执行的流程配置 |
| 草稿 | 未配置完成的演练或修改后未保存的演练。 |
| 演练任务 | 一次演练的执行记录，每执行一次演练就产生一个任务。 |
| 演练报告 | 一个演练任务的结果，一个演练任务对应一份报告。 |
| 阶段 | 活动的集合，一个演练共分4个阶段，每个阶段包含多个活动，演练会按照阶段顺序执行。 |
| 活动 | 演练流程中的一个可运行的节点，一个演练由多个活动组成。 |
| 演练参数 | 一个活动运行时所需要的参数，参数决定了活动执行的结果。 |
| 演练对象 | 实施演练的目标主机或设备。目前以活动为最小颗粒度，不同活动可以对不同的主机或设备执行。 |
| 场景 | 活动的分类，例如磁盘满载和磁盘IO高都属于磁盘这个场景。 |
| 标签 | 用户对演练自定义的含义。一个演练可以有多个标签，便于用户对演练进行分类和管理。 |

# 3 快速入门

本文提供一个快速上手的故障演练示例，让您快速了解如何进行故障演练。

## 前提条件

您已经购买了 ECS 服务器。

## 操作步骤

### 安装探针

1. 进入 AHAS 产品主页，开通 AHAS 服务。具体步骤参见开通 AHAS。
2. 登录 AHAS 控制台，在探针管理页面安装探针。具体步骤参见阿里云 ECS 主机接入。

### 新建演练

3. 在 AHAS 控制台左侧导航栏中选择故障演练 > 演练列表。
4. 在故障演练页面右上角单击新建演练。
5. 在新建演练页面，填写演练名称。

### 执行演练

6. 在 Step2：执行阶段，单击 +新建活动，选择CPU满载，单击添加。

7. 在右侧框中，选择演练对象，即安装了探针的 ECS 机器。单击确定。


8. 在演练详情页右上角，单击保存演练。

9. 单击开始演练，演练将开始进行，可看到演练任务页面。

当 CPU 满载执行完毕之后，可看到如下页面。


此时，不需要立刻执行恢复活动。可以登录演练机器，查看 CPU 信息。当观察到 CPU 满载生效之后，再开始执行恢复阶段活动。

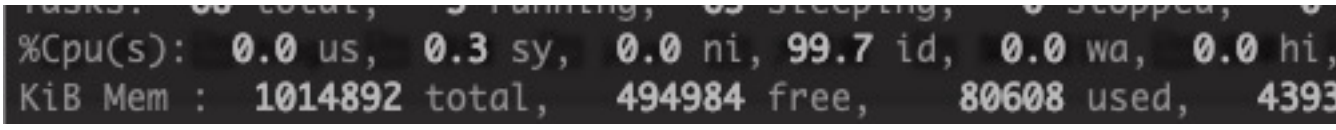

**说明:**

演练任务页面上直接查看系统信息的功能即将上线。

恢复演练

**10.点击当前卡片，选择继续执行，进入到恢复阶段。**


**恢复 CPU 满载运行成功之后，登录到演练机器上面查看 CPU 信息，可看到指标已经恢复正常。**



# 4 创建演练

基于阿里巴巴多年业务的真实线上故障库的积累，AHAS 故障演练模块为您预定义了丰富的测试任务，检验应用的高可用能力。目前，仅支持对部署在阿里云 ECS 实例上的应用进行故障演练。

## 前提条件

安装 AHAS 探针，参见阿里云 ECS 主机接入。

## 背景信息

一次完整的故障演练包括以下四个阶段：

- 准备阶段：故障注入前的准备工作，可根据执行阶段所选的活动，自动添加准备阶段所依赖的演练活动，您无需手动选择。

  例如，当故障注入影响 Java 进程时，则需要安装特定的 Agent，对故障注入所需的类进行字节码的增强；当故障注入仅影响 I/O、CPU、Disk 等，则不需要安装 Agent。
- 执行阶段：执行故障注入。您可以同时选择多个故障类型。参见演练活动参数说明。
- 检查阶段：检查故障注入的效果是否符合预期。当故障注入成功之后，通过检查阶段的活动，可以自动检查故障注入是否生效。
- 恢复阶段：清除故障。当故障演练自动结束、您主动终止或者演练中的任何环节出现异常后，系统都会进入恢复阶段，自动清除相应的故障，使故障演练对象恢复演练前的状态。

> 说明：
>
> - 通常，执行阶段和恢复阶段至少有一个演练活动，其他阶段根据需要配置。
> - 在任何阶段选择一个演练活动后，AHAS 将自动识别并添加其他阶段的依赖活动。
>
>   例如，在执行阶段选择 CPU 满载，则准备阶段会自动添加检查探针版本，恢复阶段会自动添加恢复 CPU 满载。

主要的故障类型和演练活动：

| 故障类型 | 故障演练活动 |
| --- | --- |
| CPU | CPU 满载 |
| Disk | 填充磁盘、磁盘 I/O 高 |
| Docker | 删除 Docker 容器 |
| Network | 域名访问不通、网络延迟、网络丢包 |
| Process | 杀进程 |

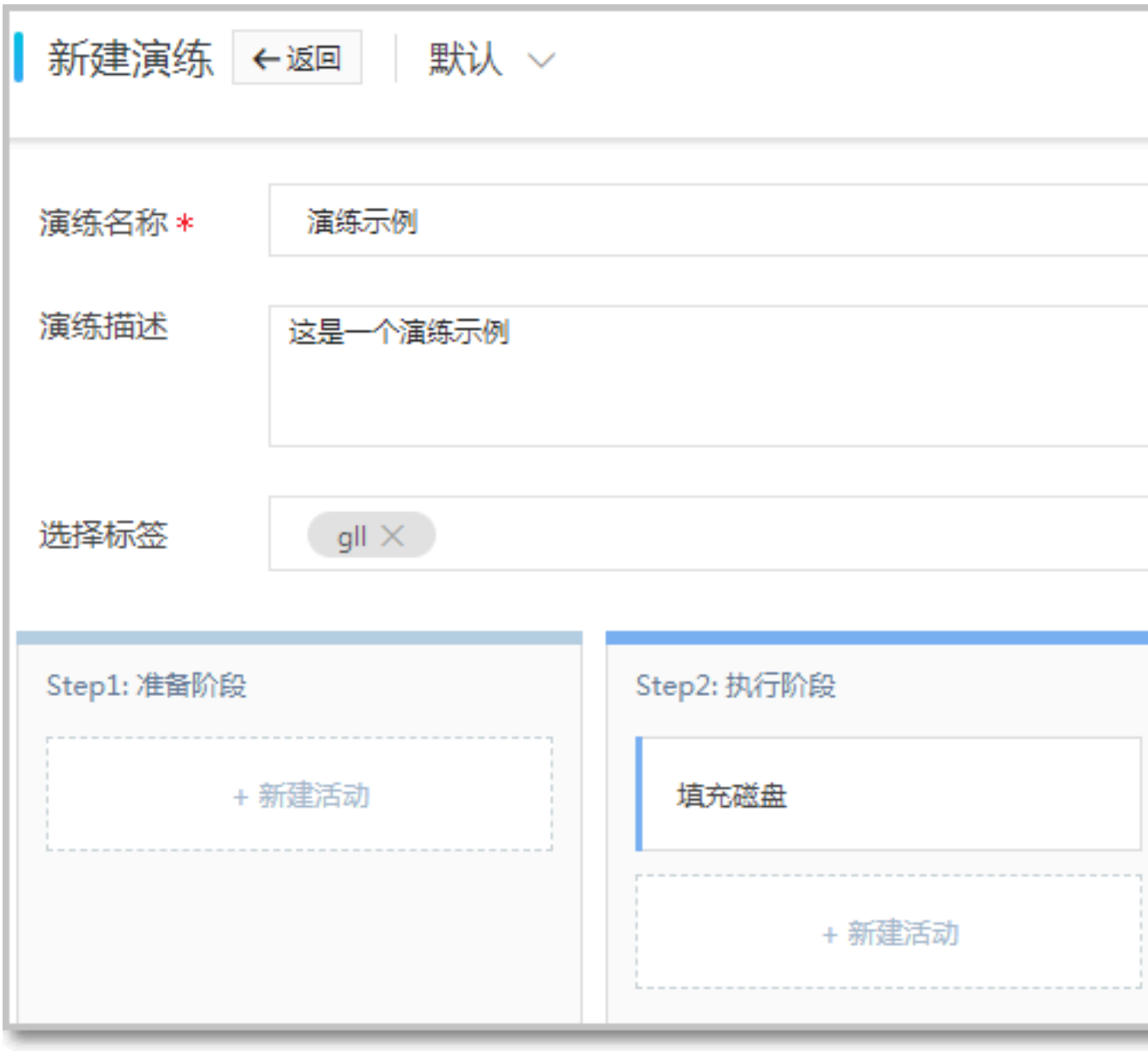
**操作步骤**

1. 登录 *AHAS* 控制台，在左侧导航栏中选择故障演练 > 演练列表。
2. 在故障演练页面右上角，单击新建演练。
3. 在新建演练页面，配置该演练场景的基本信息：

| 配置项 | 配置说明 |
| --- | --- |
| 演练名称 | 填写演练名称。 |
| 演练描述 | 为该演练添加描述。 |
| 添加标签 | 自定义演练标签，便于演练的查询和统计。<br>在添加标签框中，输入标签内容，按回车键完成输入。一个演练场景中，最多可输入五个标签。<br>标签将自动保存，下次可通过搜索或直接从下拉列表中选择已有的标签。 |

4. 在执行阶段下，单击 +新建活动，从列表中选择目标活动，例如填充磁盘，单击添加。

   系统将自动添加依赖活动，例如恢复阶段的恢复磁盘填充。




   在准备、执行、检查或恢复阶段下，重复以上步骤，可添加多个演练活动。

5. 单击已添加的演练活动，例如填充磁盘，在右侧配置面板中配置相关参数。

   参数因演练活动而异。具体参数配置说明参见演练活动参数说明。

6. 在演练对象下拉列表中，选择一个或多个主机，注入故障。勾选使用相同配置，所有活动都会自动添加相同 ECS 实例作为演练对象。

   **说明:**

   已安装 AHAS 探针且探针处于在线状态的主机，将出现在演练对象的下拉列表中。

7. 在通用配置区域，进行配置：

a) 输入该活动执行前、后等待的时长，单位是毫秒。

b) 选择是否手工推进演练。

- 是：即手动方式，当前演练活动执行完毕后，需要您手动触发去执行下一个演练活动。手动推进的方式适合需要持续观察故障注入现象的演练场景，例如与 CPU、磁盘相关的活动，通常需要人工确认故障注入生效之后，才会进行下一步。
- 否：即自动方式，当前演练活动执行完毕后，自动执行下一个演练活动。

c) 单击确定。


8. （可选） 鼠标悬停在演练活动卡片上，单击删除图标，可删除不需要的演练活动。

9. 单击新建演练页面右上角的保存。

**说明:**

如未保存，系统将在浏览器中默认保存未完成的演练内容。当您再次进入新建演练页面时，可以选择继续编辑上次的演练草稿。

返回演练列表中，您可以查看刚刚创建的演练。

系统根据您选择的演练活动，自动添加涉及场景的标签，如 C（CPU）、D（Disk）、N（Network）等，方便您快速预览演练类型。

后续操作

演练创建成功后，您可以：

- 执行演练

  在演练列表中的某个演练的操作列，单击开始，执行演练。具体参见执行演练。

- 查看并编辑演练详情

  在演练列表中的某个演练的操作列，单击查看详情，可查看演练内容。单击页面右上角编辑，可修改演练内容。

- 拷贝演练

  在演练列表中的某个演练的操作列，单击拷贝拷贝一个同名的演练，您可以在此基础上编辑演练内容。

- 删除演练

  在演练列表中的某个演练的操作列，单击删除。

# 5 执行演练

在执行故障演练过程中，您可以实时查看演练进度、每个演练活动的运行状态及结果，同时也能够随时结束演练，进行恢复阶段的活动，清除故障演练影响。

## 前提条件

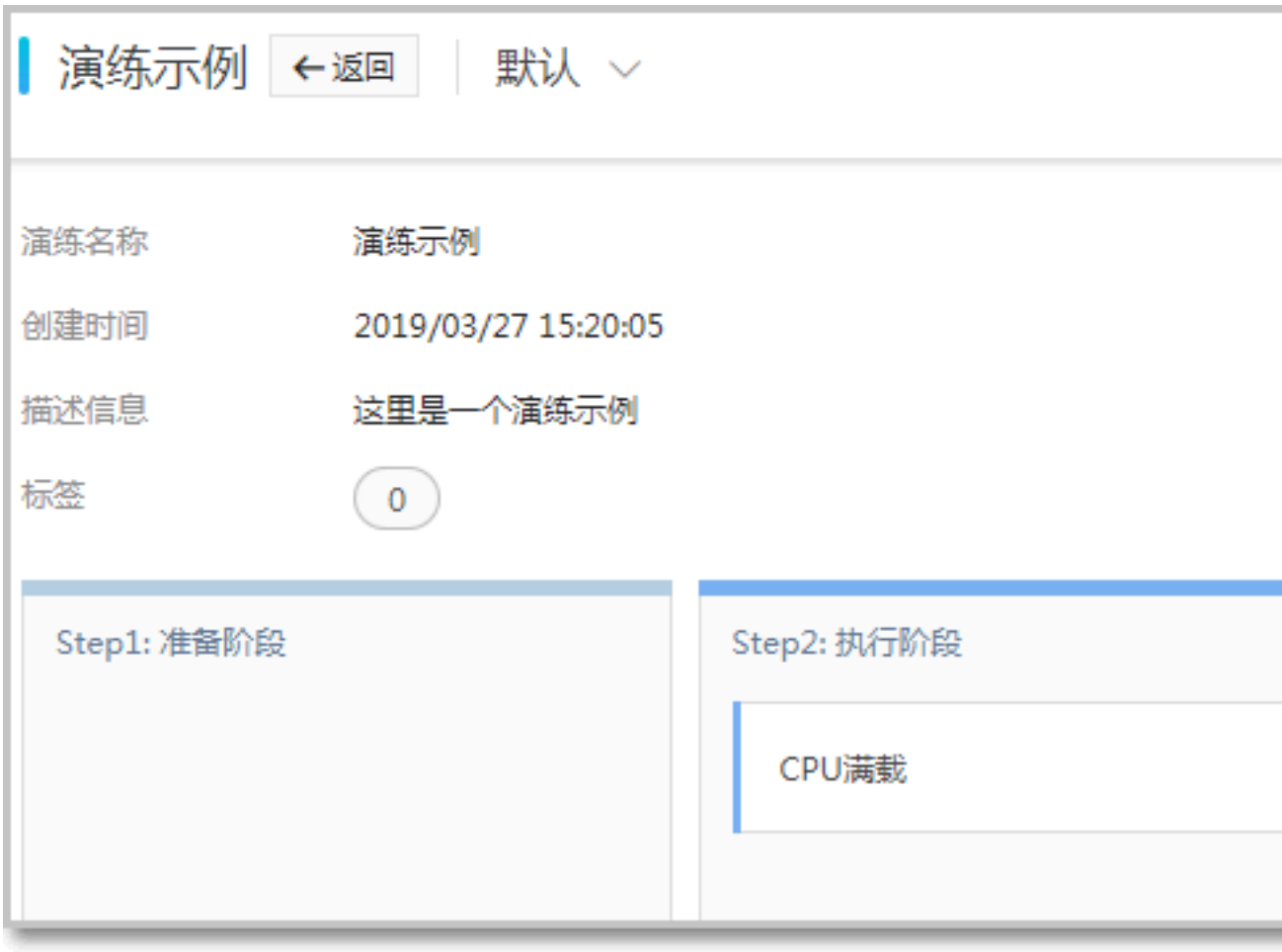
创建演练

## 操作步骤

演练任务创建完成后，您可以直接执行演练。

1. 通过以下任一方式开始演练：

   · 如果您刚刚保存演练任务，可在页面右上角，单击开始演练。




   · 在AHAS 控制台左侧导航栏中，选择故障演练 > 演练列表。在演练列表中，找到目标演练任务，单击其操作列的开始。


应用高可用服务
故障演练
默认
概览
架构感知
流控降级
故障演练
演练列表
探针管理
进行中的演练
0
请输入
演练名
演练示例
测试进行中
成功
服务注册异常
成功

2. 单击确定，开始执行演练。

演练开始后，您可以随时关注演练进度、演练时长、活动运行的结果等。


在演练中，您还可以进行以下操作：

- 单击右上角结束，随时一键结束演练。即立刻终止当前运行的活动，并且跳过 Step1 ~ 3 阶段的所有活动，同时开始执行Step4：恢复阶段的活动。
- 单击 Step4：恢复阶段的活动卡片上的执行，随时开始恢复。

> 说明:
> 在恢复阶段的每个活动上，都可以立刻运行。原因是从系统安全性的考虑，无论在什么情况下，哪怕演练已经执行完毕或者演练未进入恢复阶段，您都可以执行恢复操作。

3. 查看单个演练活动运行结果

演练活动运行结束后，单击该活动卡片，可看到当前活动上每台机器的运行结果。

机器信息一共有三种状态：

- 绿色：标明了当前机器运行成功
- 红色：标明了当前机器运行失败
- 灰色：标明了当前机器没有运行

单击具体的颜色框，可以看到当前运行的具体结果和错误。

4. 如果演练活动配置为手工推进，活动运行结束后，该活动卡片出现一个感叹号，说明当前活动需要二次确认。如下图所示：

**说明:**

非手工推进的演练活动，会自动执行，直至演练结束。

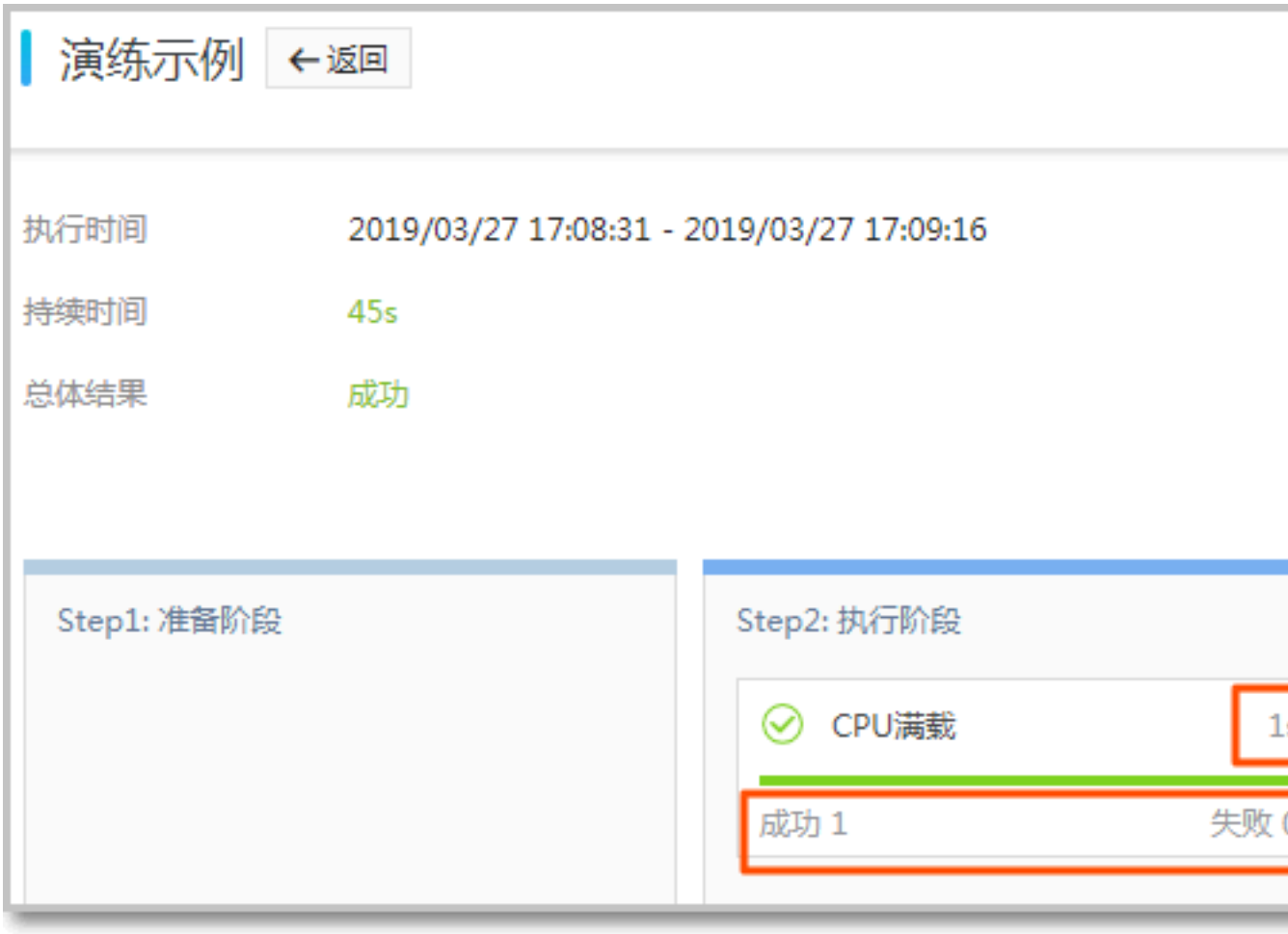

单击该活动卡片，在右侧框中，选择继续或终止：


- 继续执行：无论当前活动是否运行成功，单击继续执行，就会根据演练任务中定义的活动顺序，继续下一个演练活动。

> **说明:**
> 手工推进情况下，无论该活动本身是否成功，只要您确认继续执行，系统将默认该活动执行成功，以此统计最终演练状态。

- 终止演练：即结束演练，跳过 Step1 ~ 3 阶段的所有活动，同时开始执行Step4：恢复阶段的活动。

5. 演练结束后，单击演练详情页右上角演练报告，可查看其具体信息。




状态说明

单个演练活动的状态说明如下：

- 运行成功：表示演练活动执行成功。
- 不符合预期：表示演练活动在检查阶段时，检查结果不符合用户期望的指标，但是执行时候未发生系统异常。
- 异常：表示活动执行过程中发生了系统错误或者其他异常。
- 待执行：剩余待执行的活动。

整个演练任务的状态说明如下：

> **说明:**
> 在无手工推进的情况下，如果有一个演练活动异常，整个演练任务就异常。如果有手工推进的话，以人工选择结果为准。

- 进行中
- 等待用户确认：需用户推进演练
- 成功
- 不符合预期：表示演练活动在检查阶段时，检查结果不符合用户期望的指标，但是执行时候未发生系统异常。如果演练不包含检查阶段，则演练不会出现该状态。
- 异常：表示任务执行过程中发生了系统错误或者其他异常。
- 中断：表示用户主动结束演练。

常见问题

参见故障演练常见问题。

# 6 演练活动参数说明

故障演练活动是演练任务的核心，本文介绍了主要的演练活动及其参数配置说明。

每一个执行阶段的演练活动往往对应一个恢复阶段的演练任务。恢复阶段的演练活动目的是清除故障演练的影响，使应用或服务恢复正常，通常不需要配置参数。本文不再介绍恢复阶段的演练活动。

### CPU 满载

CPU 负载达到 100%。

| 参数名称 | 参数说明 |
| --- | --- |
| CPU 核心数 | 当被演练机器是多核 CPU 时，指定需要提高负载的 CPU 核心数目。 |

### 杀进程

杀掉指定进程。

| 参数名称 | 参数说明 |
| --- | --- |
| 进程名 | 要结束的进程的名称 |

### 域名访问不通

指定域名无法访问。

| 参数名称 | 参数说明 |
| --- | --- |
| 域名 | 不包含协议头（http:// 或 https://），不包含端口号 |
| IP 地址 | 域名的 IP 地址 |

### 网络延迟

制造网络延迟。

| 参数名称 | 参数说明 |
| --- | --- |
| 时长 | 单位是毫秒（ms），填写需要延迟的持续时间 |
| 偏移量 | 基于设置的时长上下浮动。例如，时长为 10 ms，偏移量为 2 ms，那么演练中实际生效时长是 8 ms ~ 12 ms。 |

| 参数名称 | 参数说明 |
| --- | --- |
| 网络设备 | 网卡 ID，例如：eth0 |
| 服务端口号 | 本地服务监听的端口号 |
| 访问的端口号 | 访问的外部服务的端口号 |

### 网络丢包

网络请求的数据包按比例丢失。

| 参数名称 | 参数说明 |
| --- | --- |
| 百分比 | 丢包的比例，范围是 0 ~ 100 |
| 网络设备 | 网卡 ID，例如：eth0 |
| 服务端口号 | 指定的服务的端口号 |
| 排除的端口号 | 排除的本地服务监听的端口号 |

### 填充磁盘

使用文件填充磁盘，占用磁盘空间。

| 参数名称 | 参数说明 |
| --- | --- |
| 空间大小 | 单位是 MB，要占用的磁盘空间大小。 |
| 挂载的磁盘 | 指定挂载的磁盘，例如：*/dev* |

### 磁盘 IO 高

通过读写文件使磁盘 IO 负载增加。

| 参数名称 | 参数说明 |
| --- | --- |
| 读写次数 | 读写的文件总次数，默认 1024 次。 |
| 文件大小 | 单位是 MB，单次读写的文件大小，默认是 1 MB。 |
| 读文件 | 通过读文件的方式增大磁盘 I/O 负载 |
| 写文件 | 通过写文件的方式增大磁盘 I/O 负载 |
| 挂载的磁盘 | 指定增大哪个磁盘的 I/O，例如：*/dev* |

### 删除 Docker 容器

删除 Docker 容器。

| 参数名称 | 参数说明 |
| --- | --- |
| 容器 ID | 要删除的 Docker 容器 ID |
| 是否强制删除 | 强制删除是指允许删除运行状态的容器。 |

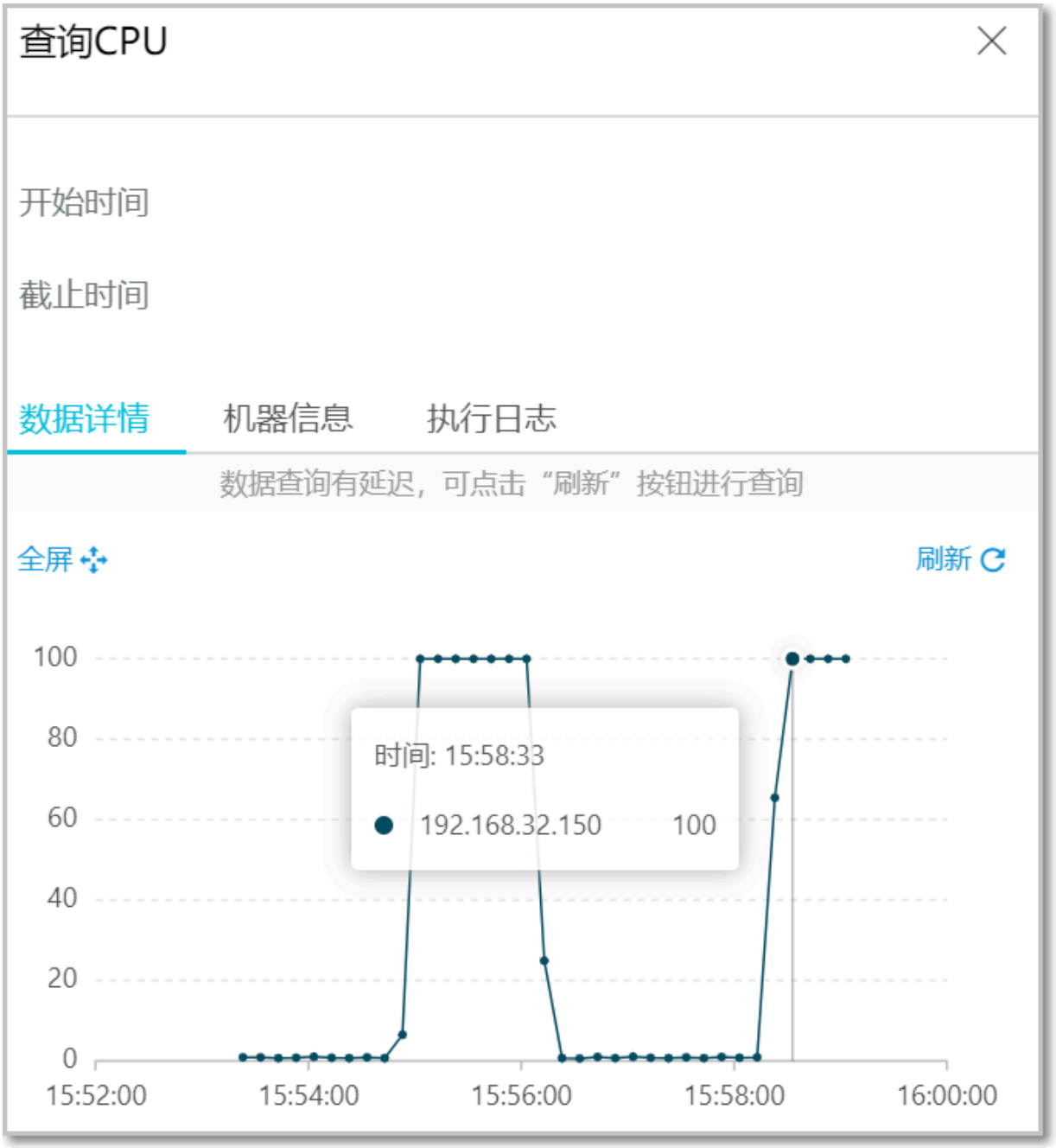
检查探针版本

检查探针版本是否满足演练需要，最低版本为 1.4.3。如不符合，需要重新安装探针。

查询 CPU

查询被演练机器的 CPU 指标 `system.cpu.util`。

在演练执行阶段结束后，可以查看 CPU 指标变化情况、机器信息和执行日志，以确保达到了演练目的，如下图所示。




**查询磁盘 I/O**

在演练执行阶段结束后，查询被演练机器的 I/O，以确保达到了演练目的。

**查询网络丢包率**

在演练执行阶段结束后，查询被演练机器的网络丢包率 `system.net.packet.dropped`，以确保达到了演练目的。

**每秒读完成次数**

在演练执行阶段结束后，查询磁盘读指标 `disk.io.read`，以确保达到了演练目的。

**每秒写次数**

**在演练执行阶段结束后，查询磁盘写指标 `disk.io.write`，以确保达到了演练目的。**