ALIBABA CLOUD

阿里云

应用高可用服务

网关防护

文档版本：20220530

阿里云

法律声明

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

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

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

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

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

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

6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

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

目录

1.什么是网关防护

AHAS可以对网关进行流量控制，从流量入口处拦截骤增的流量，防止下游服务被压垮。
网关防护的主要功能如下：

- 针对路由配置中的某个路由进行流量控制，或者自定义一组API进行流量控制。
- 针对请求的客户端IP、Header或者URL参数进行流控。
- 限制某个API的调用频率，支持秒、分钟、小时、天等多个时间维度。

网关防护支持以下接入方式：

- 接入容器服务Kubernetes版应用。
- 接入Spring Cloud Gateway应用。
- 接入Spring Cloud Zuul应用。
- 通过Agent接入。
- 将Nginx接入流量防护。

关于网关防护的规则配置有如下功能：

- 如何查看网关所有接口的QPS、RT等数据，请参见接口详情。
- 如何查看所有节点的QPS、RT等数据，请参见机器监控。
- 如何创建网关流控规则，请参见API流控规则。
- 如何自定义API，请参见API管理。

2.接入网关

2.1. 接入容器服务Kubernetes版应用

对于部署在容器服务Kubernetes版中的Java应用，可以使用AHAS应用防护对其配置流控、降级和系统规则来保证系统稳定性。本文将介绍如何将容器服务Kubernetes版中的应用接入AHAS应用防护。

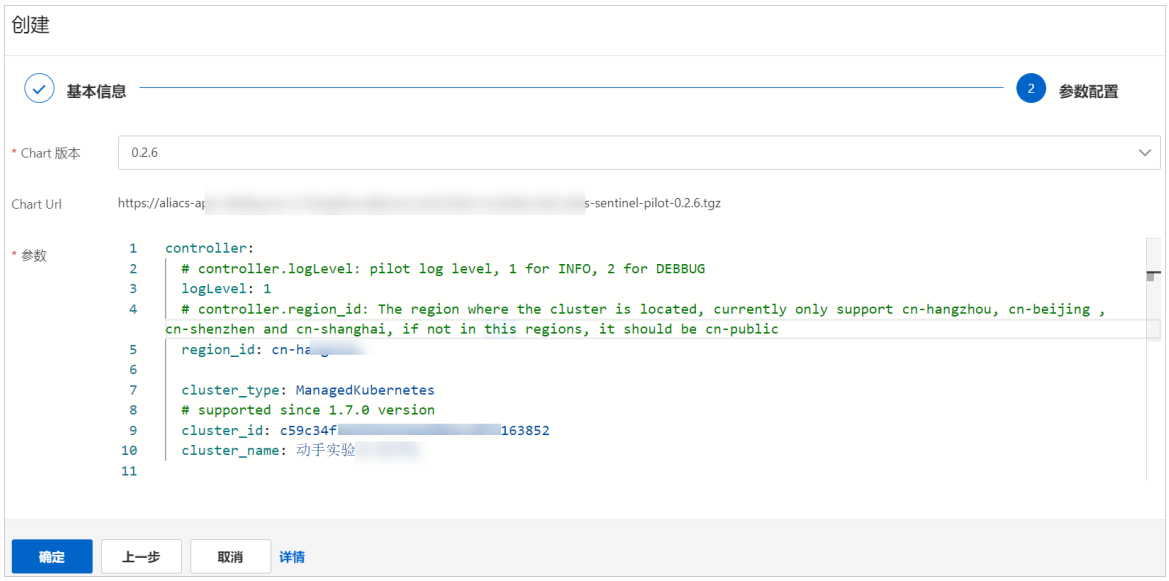
前提条件

快速创建Kubernetes托管版集群

步骤一：安装AHAS组件

在容器服务Kubernetes中安装AHAS组件后才能将Java应用接入AHAS应用防护。

1. 登录容器服务管理控制台。
2. 在控制台左侧导航栏中选择**市场 > 应用市场**。
3. 在**应用目录**页面的搜索框中输入**ack-ahas-sentinel-pilot**应用并单击。
4. 在**ack-ahas-sentinel-pilot**页面右上角单击**一键部署**，在弹出的**创建**面板选择目标集群，单击**下一步**然后单击**确定**。



参数	描述	备注
region_id	○ 如果集群和VPC之间有专线，该参数为专线连接的region。 ○ 如果集群和VPC之间没有专线，该参数填入**cn-public**。	根据所选集群自动生成。
cluster_id	您的集群ID。	
cluster_name	您的集群名称。	

步骤二：为Java应用开启AHAS应用防护

您可以按需为新建的应用或已有的应用开启AHAS应用防护。

- 如需在创建新应用的同时开启AHAS应用防护，具体步骤如下：
  i. 登录容器服务管理控制台。
  ii. 在控制台左侧导航栏中，单击**集群**。
  iii. 在**集群列表**页面中，单击目标集群名称或者目标集群右侧**操作**列下的**详情**。
  iv. 在集群管理页左侧导航栏中，选择**工作负载 > 无状态**。
  v. 在**无状态 Deployment**页面右上角单击**使用YAML创建资源**。
  vi. 在**创建**页面上方选择**示例模板**，在**模板**中替换 `image` 为Java应用的image，并在**模板**中将以下 `annotations` 添加到**spec > template > metadata**层级下，然后单击**创建**。
  如需修改YAML文件中其它配置项，配置项说明如下：
  配置项

Parameter	Description	Default
image.imageTag	pilot镜像tag。	0.1.1
image.imagePullPolicy	镜像拉取策略，必须是Always、IfNotPresent、Never三者中的一个。	Always
controller.logLevel	pilot日志级别，1表示INFO，2表示DEBBUG。	1
controller.region_id	目标集群所在的region，如cn-hangzhou、cn-beijing、cn-shenzhen、cn-shanghai。如果是公网，则为cn-public。	cn-hangzhou

  完整YAML示例模板如下：

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: agent-foo
  labels:
    name: agent-foo
spec:
  replicas: 1
  selector:
    matchLabels:
      name: agent-foo
  template:
    metadata:
      labels:
        name: agent-foo
      annotations:
        ahasPilotAutoEnable: "on"
        ahasAppName: "K8sFooTest"
        ahasNamespace: "default"
    spec:
      containers:
      - name: foo
        image: registry.cn-hangzhou.aliyuncs.com/sentinel-docker-repo/foo:0.1.1
        imagePullPolicy: Always
```

公网 非公网

- 如需为现有应用开启AHAS应用防护，操作步骤如下。
  i. 登录容器服务管理控制台。
  ii. 在控制台左侧导航栏中，单击**集群**。
  iii. 在**集群列表**页面中，单击目标集群名称或者目标集群右侧**操作**列下的**详情**。
  iv. 在集群管理页左侧导航栏中，选择**工作负载 > 无状态**或**工作负载 > 有状态**。
  v. 在**无状态（Deployment）**或**有状态（StatefulSet）**页面上，单击目标应用右侧**操作**列中选择**更多 > 查看YAML**。
  vi. 在**编辑YAML**对话框中将以下 `annotations` 添加到**spec > template > metadata**层级下，并单击**更新**。

公网 非公网

```
annotations:
  # 是否开启AHAS插件，on、true表示开启，off、false表示关闭。
  ahasPilotAutoEnable: "on"
  #将 <your-deployment-name> 替换为应用名称。
  ahasAppName: "<your-deployment-name>"
  #命名空间默认为default。
  ahasNamespace: "default"
  #将 <your-license> 替换为真实license，在公网环境中AHAS控制台 > 流量防护 >应用防护 > 新应用接入页面获取。
  ahasLicenseKey: "<your-license>"
```

结果验证

1. 登录容器服务管理控制台。
2. 在控制台左侧导航栏中，单击**集群**。
3. 在**集群列表**页面中，单击目标集群名称或者目标集群右侧**操作**列下的**详情**。
4. 在集群管理页左侧导航栏中，选择**工作负载 > 无状态**或**工作负载 > 有状态**。

在目标应用的**操作**列将出现**应用流控**按钮。单击**应用流控**即可跳转至AHAS控制台。

2.2. 接入Spring Cloud Gateway应用

Spring Cloud Gateway应用可以通过SDK接入的方式接入AHAS网关防护。将Spring Cloud Gateway应用接入AHAS网关防护后，可以对其配置流控规则来保证系统稳定性。本文介绍如何使用SDK方式将Spring Cloud Gateway应用接入网关防护。

操作步骤

1. 登录AHAS控制台，然后在页面左上角选择地域。
2. 在控制台左侧导航栏中选择**流量防护 > 网关防护**。
3. 在**网关防护**页面右上角单击**网关接入**，然后单击**Spring Cloud Gateway网关接入**页签。
4. 在Spring Cloud Gateway应用的Pom文件中添加以下依赖：

```
<dependency>
   <groupId>com.alibaba.csp</groupId>
   <artifactId>spring-cloud-gateway-starter-ahas-sentinel</artifactId>
   <version>x.y.z</version>
</dependency>
```

在**Spring Cloud Gateway网关接入**页签查看Pom依赖的最新版本，将 `x.y.z` 替换为最新的版本号。

5. 通过以下任意一种方式，配置应用的启动参数。
   - 添加JVM -D参数。

- 非公网环境下添加以下参数：

```
//将AppName替换为自定义的应用名称
-Dproject.name=AppName
```

- 公网环境下添加以下参数：

```
/将AppName替换为自定义的应用名称，将 <license> 替换为真实值。
-Dproject.name=AppName
-Dahas.license=<license>
```

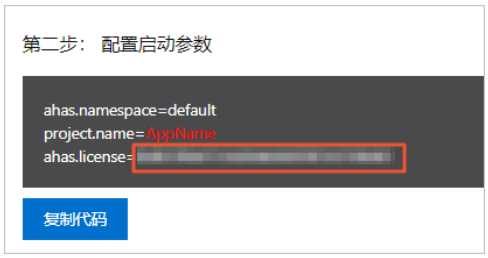
- 修改Spring Property配置文件。 在application.properties配置文件中，配置如下：
  - 非公网环境下添加以下参数：

```
#指定您要接入的特定的AHAS环境ahas.namespace=default
#自定义您的应用名称project.name=AppName
```

  - 公网环境下添加以下参数：

```
#指定您要接入的特定的AHAS环境ahas.namespace=default
#自定义您的应用名称project.name=AppName
#配置license信息ahas.license=<license>
```

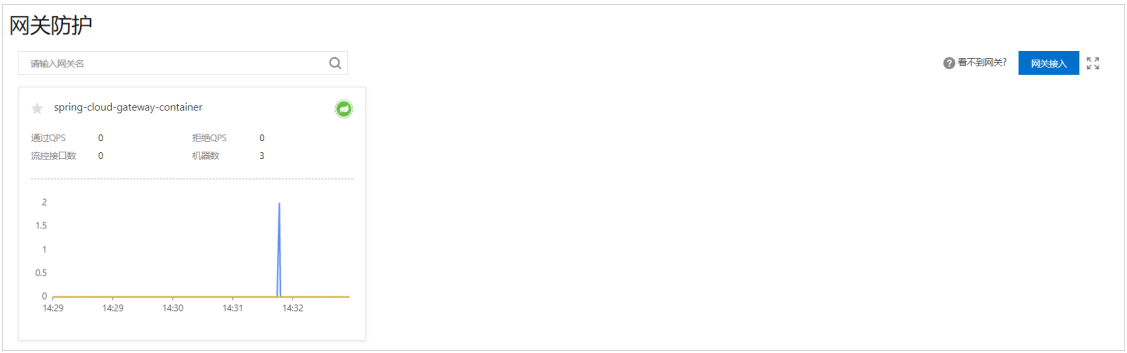
若在公网地域，需要查看License信息。请在**第二步：配置启动参数**区域查看（非公网地域不需要），具体请参见查看License。



6. 重启网关应用。

结果验证

登录AHAS控制台，在左侧导航栏选择**流量防护 > 网关防护**，在**网关防护**页面出现该网关应用的资源卡片，则说明接入成功。



后续步骤

接入网关应用后，可以为该应用配置网关流控规则。

- API流控规则

- API管理

2.3. 接入Spring Cloud Zuul应用

Spring Cloud Zuul应用可以通过SDK接入的方式接入AHAS网关防护。将Spring Cloud Zuul应用接入AHAS网关防护后，可以对其配置流控规则来保证系统稳定性。本文介绍如何使用SDK方式将Spring Cloud Zuul应用接入网关防护。

操作步骤

1. 登录AHAS控制台，然后在页面左上角选择地域。
2. 在控制台左侧导航栏中选择**流量防护 > 网关防护**。
3. 在**网关防护**页面右上角单击**网关接入**，然后单击**Zuul（1.x） 网关接入**页签。
4. 在Spring Cloud Zuul应用的Pom文件中添加以下依赖：

```
<dependency>
   <groupId>com.alibaba.csp</groupId>
   <artifactId>spring-cloud-zuul-starter-ahas-sentinel</artifactId>
   <version>x.y.z</version>
</dependency>
```

> 说明 在Zuul（1.x） 网关接入页签查看Pom依赖最新版本，将 `x.y.z` 替换为新版本的版本号。

5. 通过以下任意一种方式，配置应用的启动参数。
   - 添加JVM -D参数。
     - 非公网环境下添加以下参数：

```
//将AppName替换为自定义的应用名称。
-Dproject.name=AppName
```

     - 公网环境下添加以下参数：

```
/将AppName替换为自定义的应用名称，将 <license> 替换为真实值。
-Dproject.name=AppName
-Dahas.license=<license>
```

   - 修改Spring Property配置文件。 在application.properties配置文件中，配置如下：

- 非公网环境下添加以下参数：

```
#指定您要接入的特定的AHAS环境。
ahas.namespace=default
#自定义您的应用名称。
project.name=AppName
```

- 公网环境下添加以下参数：

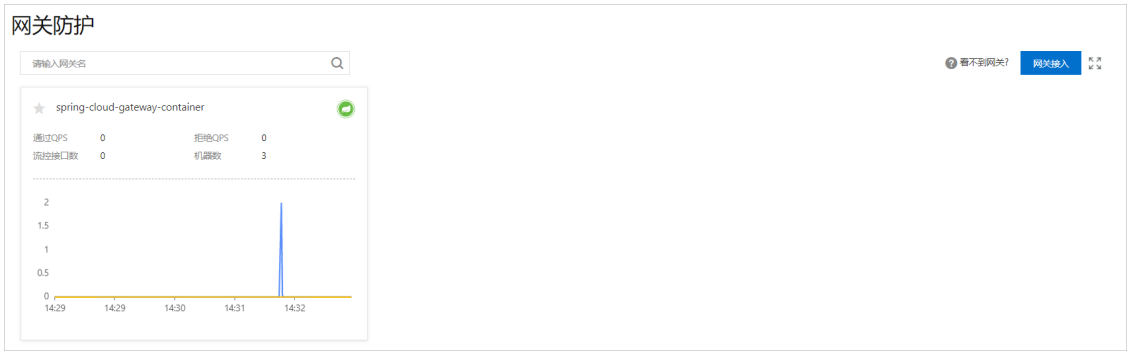
```
#指定您要接入的特定的AHAS环境。
ahas.namespace=default
#自定义您的应用名称。
project.name=AppName
#配置license信息。
ahas.license=<license>
```

若在公网地域，需要查看License信息。请在**第二步：配置启动参数**区域查看（非公网地域不需要），具体请参见查看License。

6. 重启网关应用。

结果验证

登录AHAS控制台，在左侧导航栏选择**流量防护 > 网关防护**，在**网关防护**页面出现该网关应用的资源卡片，则说明接入成功。



后续步骤

接入网关应用后，可以为该应用配置网关流控规则。

- API流控规则
- API管理

2.4. 通过Agent接入

若您的网关应用使用AHAS支持的第三方组件和框架，则可以使用Agent接入方式，无需修改代码即可接入AHAS应用防护。本文介绍如何通过Agent接入网关应用。

前提条件

确认应用使用的第三方组件和框架在支持列表中，详细信息，请参见支持组件列表。

操作步骤

1. 登录AHAS控制台，然后在页面左上角选择地域。
2. 在控制台左侧导航栏中选择**流量防护 > 网关防护**。
3. 在**网关防护**右上角单击**网关接入**。
4. 单击**Agent接入**页签。
5. 选择以下任意一种方式下载Agent。
   - 执行以下命令下载Agent。

     ```
     wget https://ahasoss-cn-hangzhou.oss-cn-hangzhou.aliyuncs.com/agent/prod/latest/ahas-java-agent.jar
     ```

   - 在Agent接入页面，单击**此处链接下载**，下载ahas-java-agent.jar安装包。
6. 执行以下命令，启动应用并挂载Agent。

   若在公网地域，需要查看License信息。请在**Agent接入**页签查看（非公网地域不需要），具体操作，请参见查看License。

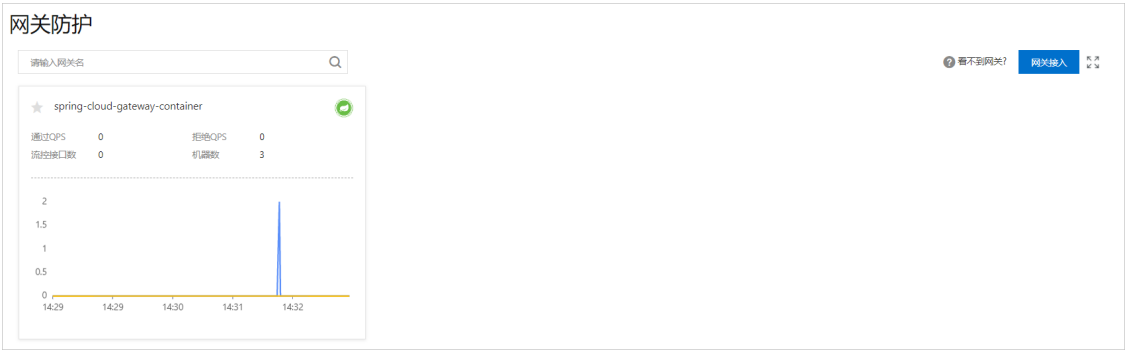
   非公网 公网

   ```
   # 添加启动参数。
   -Dahas.namespace=default -Dproject.name=AppName  -javaagent:<workdir>/agent/ahas-java-agent.jar
   ```

7. 重启网关。

结果验证

登录AHAS控制台，在左侧导航栏选择**流量防护 > 网关防护**，在**网关防护**页面出现该网关应用的资源卡片，则说明接入成功。



3.控制台操作

3.1. 接口详情

在接口详情页面，主要展示该应用所有接口的通过QPS、限流QPS、异常QPS指标、RT、并发数据等，还可以管理网关接口的流控规则。本文介绍网关防护的接口详情页的主要功能。

功能入口

1. 登录AHAS控制台，然后在页面左上角选择地域。
2. 在控制台左侧导航栏中选择**流量防护 > 网关防护**。
3. 在**网关防护**页面单击目标应用卡片。
4. 在左侧导航栏选择**接口详情**。

功能介绍

接口详情页面展示了该网关的所有接口的详细信息，包括统计的QPS、RT、并发等数据。

您还可以在此页面进行以下操作：

- （图标①）在页面右上角选择**展示模式**，默认**详情展示**。
  - **详情展示**：以时序图和时序列表的形式展现接口的通过QPS、限流QPS、RT等信息。
  - **统计展示**：以列表的形式展现某一天接口的指标占比、通过总请求数、拒绝总请求数等信息。
- （图标②）在页面右上角可以选择**回放时间**，查看接口的历史数据。

  > **说明** 高级防护最多保留7天的历史数据，入门级防护仅保留半小时的历史数据。

- （图标③）在接口列表区域，单击接口名称，可以具体查看该接口QPS数据时序图、RT数据时序图、并发数据时序图以及防护事件等，以及该接口在不同节点上的流量情况。
- （图标④）在时序图区域，可以选择要展示或隐藏的指标，还可以选择接口指标的展现形式。
  - **卡片模式**：各接口以卡片的形式展现各接口的数据。
  - **概览模式**：以QPS、RT、并发各数据的统计维度展现接口的数据。

> **说明** 模式的切换仅在全部接口场景下支持。

- （图标⑤）在时序图区域，还可以对各接口设置流控规则等操作。
  - 单击 图标，可以将该接口添加至流量大盘，便于在流量大盘中观测系统整体流量，请参见创建流量大盘。
  - 单击 或 图标，进入管理规则对话框，可以新增或删除流控规则，也可以编辑已有的规则或开启关闭规则。详情请参见API流控规则。
  - 单击 图标，可以查看该接口指标的历史数据。

> **说明** 高级防护最多保留7天的历史数据，入门级防护仅保留半小时的历史数据。

3.2. 机器监控

在机器监控页面，主要展示了所有节点的通过QPS、限流QPS、异常QPS、RT 、并发等指标，还可以在此页面为接口管理流控规则。本文介绍机器监控页的主要功能。

功能入口

1. 登录AHAS控制台，然后在页面左上角选择地域。
2. 在控制台左侧导航栏中选择**流量防护 > 网关防护**。
3. 在**网关防护**页面单击目标应用卡片。
4. 在左侧导航栏中选择**机器监控**。

功能介绍

机器监控页面展示了应用的所有节点详细信息以及这些节点的QPS、CPU、LOAD时序图。

您可以在此页面进行以下操作：

- （图标①）在页面右上角选择**回放时间**，查看历史数据。

> **说明** 高级防护最多保留7天的历史数据，入门级防护仅保留半小时的历史数据。

- （图标②）在**节点名称**区域，罗列了全部节点和对应的通过QPS、限流QPS、异常QPS、RT等信息。单击节点名称可以查看对应的各数据时序图。

- （图标③）在时序图区域，可以进行以下操作：
  - 单击**QPS**、**CPU**、**LOAD**页签，可以分别查看全部节点相关指标的时序图，还可以选择要展示或隐藏的指标。
  - 单击 图标，可以查看该接口指标的历史数据。

    > **说明** 高级防护最多保留7天的历史数据，入门级防护仅保留半小时的历史数据。

  - 单击节点名称后，会在右侧**节点概览**页展示该节点对应的各数据时序图。可以单击**分接口详情**页签，筛选查看不同接口的数据。还可以单击**callstack信息**页签，查看所有接口的信息，并可以设置该接口的限流规则、查看历史数据。
    - 平铺视图：不区分调用链路关系，平铺展示接口的运行情况。
    - 树状视图：根据接口的调用链路关系，展示树状结构。
  - 单击目标接口**操作**列中的**流控**、**隔离**或**降级**，可以快速管理限流规则。详情请参见配置流控规则、配置隔离规则和配置熔断规则。
  - 单击目标接口**操作**列中的**查看监控**，可查看该接口指标的历史数据。

3.3. API管理

在AHAS网关防护中，您可以创建API分组，并自定义每个API下面的URL路径匹配规则。AHAS网关防护可以针对自定义的API分组进行流量控制。本文介绍如何在网关防护中管理API。

新建自定义API

1. 登录AHAS控制台，然后在页面左上角选择地域。
2. 在控制台左侧导航栏中选择**流量防护 > 网关防护**。
3. 在**网关防护**页面单击目标应用卡片。
4. 在左侧导航栏选择**API管理**，单击页面右上角的**新增API**。
5. 在**新建自定义API**对话框中，填写API名称。

   > **说明** 该名称需要全局唯一，并且不能与路由配置文件中的路由ID重复。

6. 填写URL路径匹配规则，先选择**匹配模式**，再根据匹配模式的要求填写**匹配串**。
   - **匹配模式**分为以下三类：
     - **精确模式**：严格按照给定的匹配串来匹配URL路径。示例：`/foo` 代表严格按照 `/foo` 这个路径来匹配。
     - **前缀模式**：按照给定的匹配串来进行前缀匹配，匹配串需符合Spring Web风格。示例：`/foo/**` 代表匹配以 `/foo/` 开头的所有URL，像 `/foo/22` 这种URL都可以匹配。
     - **正则模式**：按照给定的正则表达式匹配串来进行匹配。
   - **匹配串**：根据匹配模式的要求填写匹配串。
7. 单击**+新增匹配规则**，可添加多个URL路径匹配规则。
8. 单击**新增**，完成自定义API的创建。
   新增的API将出现在**API管理**页面。

相关操作

新增API后，您可以编辑、删除API。

- **编辑API**
  i. 在**API管理**页面，在目标API的**操作**列，单击**编辑**。
  ii. 在**编辑自定义API** 对话框中，修改URL匹配规则，也可以新增URL匹配规则。
- **删除API**
  i. 在**API管理**页面，在目标API的**操作**列，单击**删除**。
  ii. 在提示框中，单击**确定**，将该API分组删除。

3.4. 集群流控

相较于普通的单机流控，集群流控可以精确控制集群内某个服务的实时调用总量。在网关防护中采用集群流控，用户可无需关心负载均衡状况和网关数量，只需配置总阈值即可完成操作。本文主要介绍设置集群流控的操作步骤。

计费说明

自2021年03月22日起，集群流控功能公测期结束，正式开始收费。集群流控功能按应用申请的QPS量级收费，具体收费方式，请参见价格页面。

> **说明** 在2021年03月22日之前创建，且在2021年03月22日之后未进行变更的集群暂不开启计费。

集群流控试用档位的Token Server可供您继续测试使用，不会产生额外费用。试用档位单个应用QPS阈值之和不超过2000，接口总流量不超过3000。

> **说明** 试用档位仅供测试效果使用，不保证稳定性，请勿在生产环境使用。

前提条件

- 已开通AHAS专业版，若没有开通，请进入开通页面。
- Spring-Cloud-Gateway-Starter-AHAS-Sentinel和Spring-Cloud-Zuul-Starter-AHAS-Sentinel版本≥v1.3.7，对应AHAS-Sentinel-Client版本≥v1.8.8。Agent版本≥v1.8.6。

步骤一：选择档位创建集群

1. 登录AHAS控制台，然后在页面左上角选择地域。
2. 在控制台左侧导航栏中选择**流量防护 > 网关防护**。
3. 在**网关防护**页面单击目标应用卡片。
4. 在网关防护管理页左侧导航栏，单击**集群流控**。
5. 在**集群流控资源配置**区域内，单击右下角的**编辑**。
6. 选择**集群类型**为**生产**，滑动指针选择集群流控的总配置量级，单击**保存**，然后在对话框中单击**确认**。

总配置量级即最大QPS，表示需要流控的接口所能承载的预估的最大QPS，代表可能到来的最大流量。

> **说明** 实际流量（无论是否被流控）超出配置的最大QPS后，流控策略会退化到单机模式。为保证流控效果，阈值之和上限为配置最大QPS的95%，例如最大QPS选择100000，则所有规则阈值之和最大值为95000。

选定总配置量级档位并创建集群后，系统会自动为该应用分配集群的Token Server。

7. （可选）单击**Token Client设置**区域**操作**列的**编辑**，设置**Token请求超时时间**，然后单击**确定**。

   在某些场景下，集群流控Client与Token Server之间的网络通信时延较高，需要调整超时时间。

   > **说明**
   > - AHAS Sentinel Client 1.6.0及以上版本支持设置Token Client。
   > - Token请求超时时间单位为ms，取值范围为(0,10000]，一般不建议超过20 ms。公网环境网络延时较高，建议设置超时时长约为50 ms，但不建议超过80 ms。

步骤二：设置集群流控规则

1. 在网关防护管理页左侧导航栏，单击**规则管理**。
2. 单击**集群流控规则**页签，然后单击**新增集群流控规则**。
3. 在**新增集群流控规则**对话框，设置相关参数。

参数	描述	示例
接口名称	设置接口名称，为对应网关的Route ID或自定义API分组名称。	httpbin_route
是否开启	开启此开关，规则即生效；关闭此开关，规则不生效。	开启
集群阈值	表示该接口的限流阈值。	100
统计窗口时长	集群流控统计的时间窗口长度，取值范围为1秒~24小时。	1秒
失败退化策略	当出现连接失败、通信失败或Token Server不可用等情况时，流控规则是退化到单机限流的模式或是直接通过忽略失败情况： ○ 退化到单机限流：当出现通信失败的情况时，退化到设置的单机阈值来进行流控。需要在规则中配置单机退化阈值，代表单机的兜底阈值。 ○ 直接通过：当出现通信失败的情况时，请求直接通过。	退化到单机限流

参数	描述	示例
退化阈值自动调整	开启后会自动调整退化阈值，默认关闭。 说明 此功能需要SDK版本≥1.8.6。	关闭
退化单机阈值	代表单机的兜底阈值，当失败退化策略选择退化到单机限流时，需要设置此选项。 说明 只有在没有开启**退化阈值自动调整**的情况下，才需要手动填写**退化单机阈值**。若开启**退化阈值自动调整**，您无需填写**退化单机阈值**，而需设置**自动调整增量值**。	10
自动调整增量值	当开启退化阈值自动调整时，需要设置自动调整的增量。这是在根据接口阈值与应用机器数量计算出的单机均摊流量基础上，用来提供保护退化阈值的一个增量。即单机均摊流量加上增量值为实际生效退化阈值。	2

4. 单击**新建**，完成规则创建。
   创建规则完成后，可以在**规则设置**页面查看到创建的集群流控规则，**阈值模式**为**集群总体**。

3.5. API流控规则

为网关应用配置网关流控规则后，AHAS将从流量入口处拦截激增的流量，防止下游服务被压垮。本文将介绍如何为已接入AHAS的网关应用配置网关流控规则。

新建网关流控规则

1. 登录AHAS控制台，然后在页面左上角选择地域。
2. 在控制台左侧导航栏中选择**流量防护 > 网关防护**。
3. 在**网关防护**页面单击目标应用卡片。
4. 单击目标网关应用卡片，然后任选一种方式进入**API流控规则**的配置页面：
   - 在**接口详情**页面，单击API资源卡片右上角的加号图标。
   - 在左侧导航栏中单击**API流控规则**，然后在页面右上角单击**新增流控规则**。
5. 在**新增流控规则**对话框中，配置流控规则。

参数	描述
API	选择适用该规则的自定义API，或者手动输入路由配置文件中的Route ID。
针对请求属性	○ 关闭**针对请求属性**开关：不针对请求属性（如Client IP，URL参数等）进行限流，直接针对该API的所有请求进行流量控制。 ○ 开启**针对请求属性**开关：针对该API的某个请求属性进行限流，可以选择参数属性。 可以根据以下属性进行流量控制： ■ **Client IP**：请求端的IP地址。 ■ **Remote Host**：请求端的Host Header。 ■ **Header**：根据指定的HTTP Header进行解析，匹配对应的Header Key。选择**Header**后，可以配置请求属性值的匹配策略，只有匹配该模式的请求属性值会纳入统计和流控。 ■ **URL参数**：根据指定的HTTP URL参数进行解析，需要填写对应的参数名称。选择**URL参数**后，可以配置请求属性值的匹配策略，只有匹配该模式的请求属性值会纳入统计和流控。 ○ 匹配模式 ■ **精确**：严格按照给定的匹配串来匹配值。 ■ **子串**：若请求属性值包含该子串则匹配成功，如子串匹配 `ab` ，则 `aba` 和 `cabc` 都可以匹配，而 `cba` 则不能匹配。 ■ **正则**：按照给定的正则表达式匹配串来进行匹配。
阈值类型	○ **QPS**：应用或服务流量的QPS指标。选择QPS后，还需设置QPS阈值和统计间隔（支持秒、分钟、小时、天4种维度）。 例如，QPS阈值填写10，统计间隔选择**分**，则代表每分钟对应的请求数目不超过10个。 ○ **线程数**：资源的并发线程数，即该资源正在执行的线程数。 **说明** 开启**针对请求属性**开关后，暂时不支持线程数作为阈值类型。

参数	描述
流控方式	○ **快速失败**：当**阈值类型**为**QPS**时，被拦截的流量将快速失败。即达到阈值时，立即拦截请求。 ○ **匀速排队**：当**阈值类型**为**QPS**时，被拦截的请求将匀速通过，允许排队等待。 需设置具体的超时时间，预计达到超时时间的请求会立即失败，而不会排队。 例如，QPS配置为10，则代表请求每100 ms才能通过一个，多出的请求将排队等待通过。超时时间代表最大排队时间，超出最大排队时间的请求将会直接被拒绝。 **说明** 匀速排队时，QPS不要超过1000（请求间隔1 ms）。
Burst size	当流控方式为**快速失败**时，可以额外设置一个Burst Size，即针对突发请求额外允许的请求数目。
超时时间	当流控方式为**匀速排队**时，需设置具体的超时时间，达到超时时间后请求会失败。例如，QPS配置为5，则代表请求每200 ms才能通过一个，多出的请求将排队等待通过。超时时间代表最大排队时间，超出最大排队时间的请求将会直接被拒绝。

6. 单击**新增**。
   新增的规则将出现在**API流控规则**页面。

管理流控规则

在**流控规则**页面，您可以启用、禁用、编辑或删除流控规则。

- 单流控规则启用或禁用：
  在**流控规则**页面，找到目标资源下对应的流控规则，单击状态**状态**栏的启用开关，可快速启用或禁用该规则。
- 多流控规则批量启用或禁用：
  在**流控规则**页面，勾选多个流控规则，单击**批量启用**或**批量禁用**，可快速启用或禁用多个规则。
- 编辑规则：
  在**流控规则**页面，找到目标资源下对应的流控规则，单击**操作**栏的**编辑**，可修改该规则的相关信息。
- 删除规则：
  在**流控规则**页面，找到目标资源下对应的流控规则，单击**操作**栏的删除**删除**。

4.SDK使用手册

4.1. 触发网关防护规则后的限流策略

若默认配置不能满足您的需求时，您可以自定义应用触发流控、降级或系统规则后的逻辑。本文将介绍适用于SDK接入方式的逻辑配置方法。

Spring Cloud Gateway

若您的网关是Spring Cloud Gateway，则默认的限流处理逻辑是返回默认的流控文本`Blocked by Sentinel`，返回`status code`为`429 Too Many Requests`。您可以通过以下Spring配置项来配置限流后的处理策略。

- `spring.cloud.sentinel.scg.fallback.mode`：限流处理策略，目前支持跳转`redirect`和自定义返回`response`两种策略。
- `spring.cloud.sentinel.scg.fallback.redirect`：限流之后的跳转URL，仅在mode=redirect的时候生效。
- `spring.cloud.sentinel.scg.fallback.response-body`：限流之后的返回内容，仅在mode=response的时候生效。
- `spring.cloud.sentinel.scg.fallback.response-status`：限流之后的返回`status code`，仅在mode=response的时候生效。

除此之外，您也可以在GatewayCallbackManager上通过setBlockHandler注册函数实现自定义的逻辑处理被限流的请求，对应接口为`BlockRequestHandler`，编写逻辑可参考`DefaultBlockRequestHandler`默认实现类。

> **说明** YAML文件请注意转成YAML配置的形式。

Zuul 1.x

若您的网关是Zuul 1.x，则默认的限流处理逻辑是返回默认的流控文本，返回`status code`为`429 Too Many Requests`。

您可以通过注册回调的方式定制处理异常，示例如下。

```
// 自定义FallbackProvider。
public class MyBlockFallbackProvider implements ZuulBlockFallbackProvider {
    @Override
    public String getRoute() {
        // 对应的route或API group。
        return "book-service";
    }
    @Override
        public BlockResponse fallbackResponse(String route, Throwable cause) {
            if (cause instanceof BlockException) { // AHAS流控、降级、系统保护异常。
                return new BlockResponse(429, "Blocked by AHAS Sentinel", route);
            } else {
                return new BlockResponse(500, "System Error", route);
            }
        }
 }
 // 注册FallbackProvider。
 ZuulBlockFallbackManager.registerProvider(new MyBlockFallbackProvider());
```