检索分析服务

操作指南

腾讯云

【版权声明】

【商标声明】

【服务声明】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

【联系我们】

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或95716。

文档目录

操作指南
访问控制
CAM 访问控制

最近更新时间：2025-09-12 14:17:52

CAM 简介

访问管理（Cloud Access Management，CAM）是腾讯云提供的 Web 服务，主要用于帮助用户安全管理腾讯云账户下的资源的访问权限。用户可以通过 CAM 创建、管理和销毁用户(组)，并使用身份管理和策略管理控制其他用户使用腾讯云资源的权限，CAM 策略的详细信息及使用请参见 CAM 策略 。

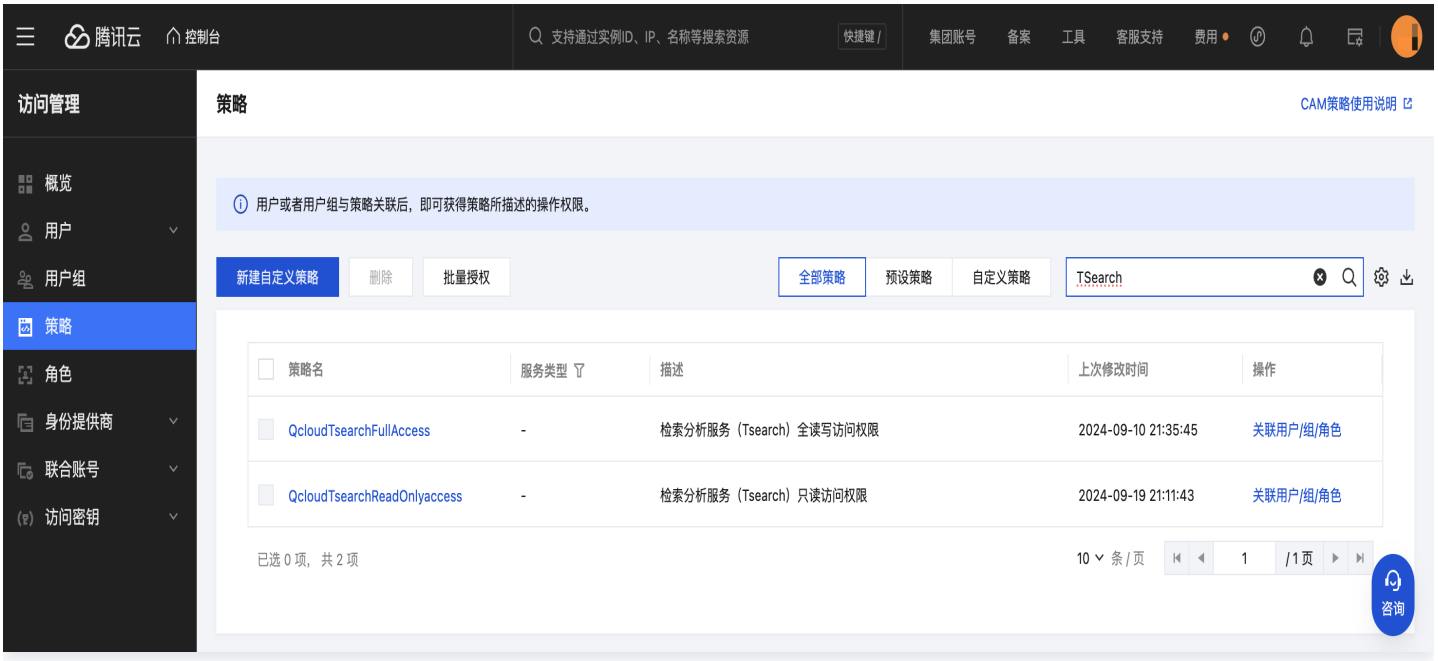
TSearch CAM 策略

通用权限策略

TSearch 默认提供了两种通用策略：

- 全读写策略 QcloudTSearchFullAccess，可以让用户拥有创建和管理 TSearch 所有集群实例的权限。
- 只读策略 QcloudTSearchReadOnlyAccess，可以让用户拥有查看 TSearch 集群实例的权限，但是不具有创建、更新、删除等操作的权限。

您可以登录 访问管理控制台 > 策略，在搜索框中搜索“TSearch”，在列表中会显示默认策略，可对需要授权的账号进行绑定。

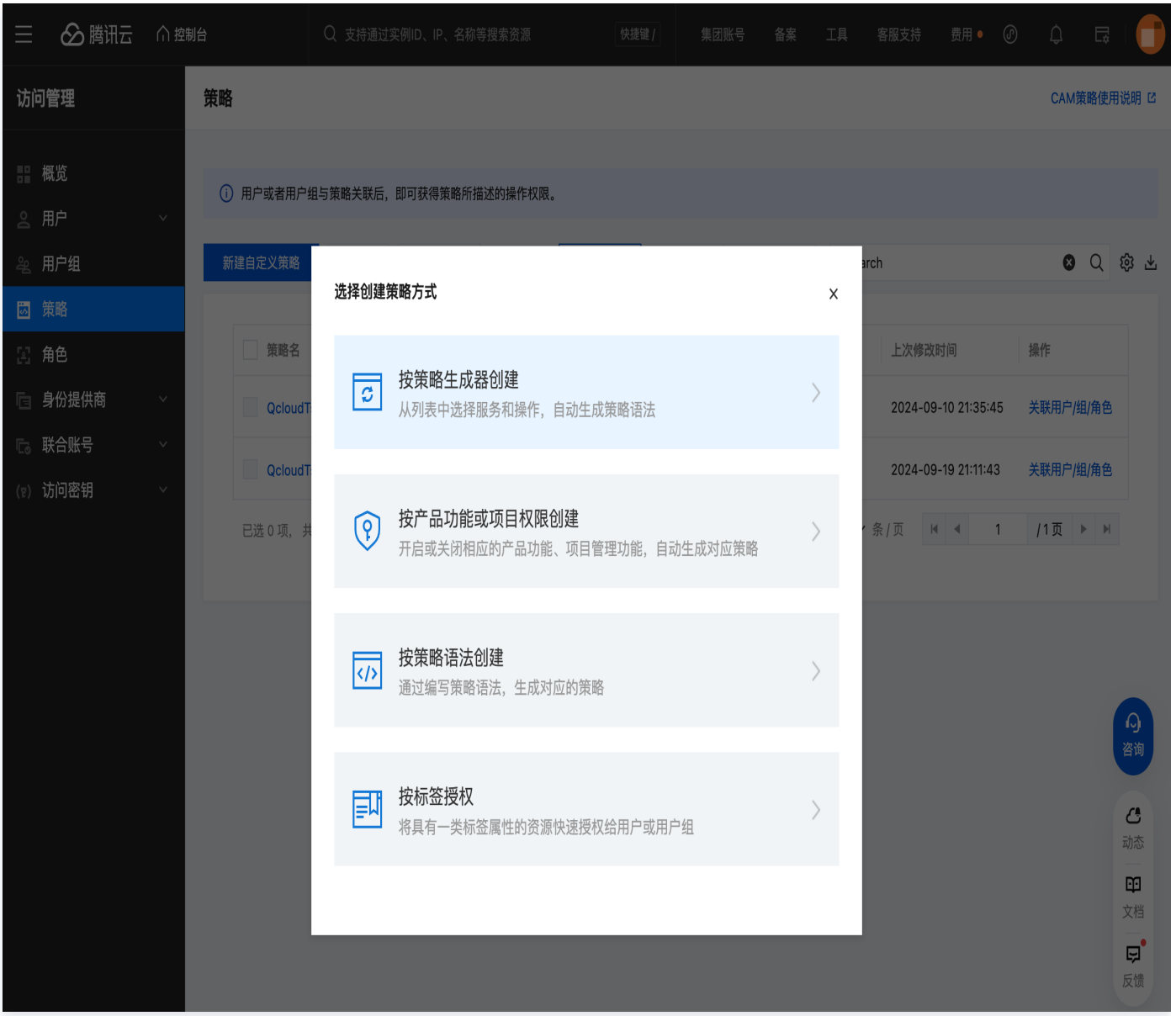


如果默认策略不能满足需求，可单击**新建自定义策略**，进行自定义授权。

自定义权限策略：通过访问管理实现

腾讯云访问管理提供4种自定义策略方式。

- 按策略生成器创建：从策略向导中选择服务、操作、定义资源，自动生成策略语法，可以灵活满足差异化权限管理需求，优先推荐使用。详细的操作步骤，您可以访问 通过策略生成器创建自定义策略 。
- 按产品功能或项目权限创建：通过开启或关闭项目管理、主题模型或内容分发网络相关功能，自动生成对应策略。
- 按策略语法创建：您可以编写策略语法，生成对应的策略，权限粒度灵活，可以解决对权限精细划分有较高要求的用户诉求。详细的操作步骤，您可以访问 通过策略语法创建自定义策略 。
- 按标签授权：通过标签授权创建自定义策略，策略生成后该策略将具有一类标签属性资源的权限。详细的操作步骤，您可以访问 通过标签授权创建自定义策略 。



自定义权限策略：通过策略语法实现

TSearch 可授权的资源类型如下：

资源类型	资源描述
instance	qcs::tsearch:$region:$account:instance/*

各 API 支持资源级权限访问控制详情如下：

接口名	描述	是否关联资源	资源描述
获取集群列表、单个集群信息	DescribeInstances	是	qcs::tsearch:${Region}:uin/${uin}:instance/${instanceId}
创建集群	CreateInstance	否	*
更新集群	UpdateInstance	是	qcs::tsearch:${Region}:uin/${uin}:instance/${instanceId}
重启集群	RestartInstance	是	qcs::tsearch:${Region}:uin/${uin}:instance/${instanceId}
删除集群	DeleteInstance	是	qcs::tsearch:${Region}:uin/${uin}:instance/${instanceId}
更新插件	UpdatePlugins	是	qcs::tsearch:${Region}:uin/${uin}:instance/${instanceId}

支持区域如下：

区域	名称	区域 ID
华南地区	广州	ap-guangzhou
华东地区	上海	ap-shanghai
	南京	ap-nanjing
华北地区	北京	ap-beijing
西南地区	重庆	ap-chongqing

自定义策略的语法如下：

```
{
    "version": "2.0",
```

```
    "statement": [
        {
            "action": [
                "Action"
            ],
            "resource": "Resource",
            "effect": "Effect"
        }
    ]
}
```

- Action：替换成要允许或拒绝的操作。
- Resource：替换成要授权的具体资源。
- Effect：替换成允许或拒绝。

TSearch 目前支持除了 DescribeInstances 接口之外的其他操作接口来访问控制管理，您可以将账号下某个集群的更新、重启、删除等操作赋予某个子账号进行管理。

自定义权限示例

授予某账号指定集群更新操作权限，策略语法如下：

```
{
    "version": "2.0",
    "statement": [
         {
            "action": [
                "tsearch:Describe*"
            ],
            "resource": [
               "qcs::tsearch:ap-guangzhou:uin/$uin:instance/$instanceID"
            ],
            "effect": "allow"
        },
        {
            "action": [
                "vpc:Describe*",
                "vpc:Inquiry*",
                "vpc:Get*"
            ],
            "resource": "*",
            "effect": "allow"
        },
        {
```

```
            "action": [
                "monitor:*",
                "cam:ListUsersForGroup",
                "cam:ListGroups",
                "cam:GetGroup"
            ],
            "resource": "*",
            "effect": "allow"
        },
        {
            "action": [
                "tsearch:Update*"
            ],
            "resource": [
                "qcs::tsearch:ap-
guangzhou:uin/$uin:instance/$instanceID"
            ],
            "effect": "allow"
        }
    ]
}
```

TSearch 集群访问控制

最近更新时间：2025-09-12 14:17:52

TSearch 集群部署在逻辑隔离的私有网络 VPC 中，我们提供了丰富的能力来切实保证云上资源的安全性，包括：

- 对腾讯云账户下资源的 CAM 访问管理（参见 CAM 访问控制）
- TSearch 集群访问密码 / TSearch 集群用户登录认证
- 设置 Web 节点外网访问 IP 黑白名单，或限制 Web 节点仅能通过内网访问。

设置 TSearch 集群访问密码

在创建腾讯云 TSearch 集群时，会要求用户设置默认用户 tsearch 的密码，该账号和密码用于 Web节点可视化组件页面登录，若集群已开启 TSearch 集群用户登录认证，则此用户名和密码还会用于 TSearch 集群的登录认证，提供进一步的安全防护，详情如下：

重置 TSeach 集群访问密码

用户需要调整 TSearch 集群访问密码时，可以通过集群详情页的密码重置功能对 TSearch 集群用户名密码进行重置，操作页面如下：

设置 Web 公网访问 IP 黑白名单

由于 Web 节点是默认通过公网访问，在进行密码校验的基础上，TSearch 还为 Web 节点访问提供了 IP 黑白名单功能，进一步保障用户集群的访问安全性。
Web 节点公网访问白名单默认为 127.0.0.1，表示不允许所有 IPv4 和 IPv6 地址访问。在集群详情的访问控制界面可以设置 Web 节点公网访问 IP 黑白名单。

- 配置规则：支持多个 IP，IP 之间以英文逗号分隔，格式为192.168.0.1,192.168.0.0/24，最多支持50个。
- 白名单配置：支持单击自动获取当前IP地址，单击确认后一键填入。

- 黑白名单设置：客户可以设置任意一个，如果黑白名单都配置，以白名单为准。

限制 Web 节点仅能通过内网访问

如果用户担心公网访问安全性，也可以关闭外网访问，设置仅允许内网访问。

开启公网访问

基于安全考虑，TSearch 集群外网访问是默认关闭的，对于已开启 TSearch 集群用户登录认证的集群，允许用户基于使用便捷性的需要开启外网访问，但必须同时设置 IP 白名单以提供安全防护。


确认开启 Web节点 公网访问地址?
警告：开启公网访问可能给集群引入安全风险，这将允许直接访问、操作甚至删除你在TSearch里的数据，请谨慎开启。
IP白名单可提供一定保护，请确保白名单设置的安全性，并避免泄漏。
可访问IP白名单 *
点击自动获取当前IP地址
支持多个IP，IP之间支持以英文逗号、分号或换行符分隔，格式可以是192.168.0.1,192.168.0.0/24，最多支持500个。
注：设置127.0.0.1代表禁止所有IPv4地址访问，0.0.0.0/0代表允许所有IPv4地址访问。
确定
取消

管理集群
集群状态

最近更新时间：2025-09-12 14:15:02

在集群列表页和集群详情页的基本信息中，均可查看集群的运行状态信息。

单击集群 ID/名称进入集群详情页。

集群状态是反映集群是否在变更中或正常使用的状态，包括正常、处理中，具体含义如下：

状态	含义
正常	集群创建完成，且没有配置变更、集群重启等动作，可以正常访问和使用的状态
处理中	集群创建、集群配置变更、集群重启等操作，需要一定的时间进行处理的状态，期间部分服务访问会受到影响，如 Kibana、数据存储和查询等

其中健康状态是 TSearch 集群众多监控信息中非常重要的一个，用来表征集群总体上是否工作正常。健康状态种类如下：

颜色	健康状态
绿色（green）	所有的主分片和副本分片都正常运行
黄色（yellow）	所有的主分片都正常运行，但不是所有的副本分片都正常运行

红色（red）	有主分片没有正常运行

重启集群

最近更新时间：2025-09-12 14:15:02

操作场景

腾讯云检索分析服务 TSearch 为用户提供了集群重启功能，用户可以根据集群的情况使用此功能对集群进行重启操作。

注意事项

1. 重启集群为耗时操作，建议用户在非必要情况下不要进行该操作。
2. 滚动重启过程会逐个重启各节点，因此用户应保证集群至少有1副本，否则会导致重启时集群变为 RED 状态，甚至有丢失数据的风险。
3. 当集群状态为 YELLOW 或 RED，或集群存在无副本索引等情况下，用户进行重启操作会提示需要进行强制重启。强制重启操作有较高风险，存在丢失数据的可能，在此种情况下建议用户先恢复集群状态为 GREEN，并为所有索引创建至少一个副本，然后再进行重启操作。若当前集群状态无法恢复，用户在了解强制重启风险的前提下仍要进行重启操作，可以勾选强制重启选项进行重启。

操作步骤

1. 登录 检索分析服务 TSearch 控制台 。
2. 重启集群有两个入口，分别位于集群列表页和集群详情页。
   - 在集群列表页，选择相应的集群，选择**操作 > 更多 > 重启**。
   - 单击集群名称进入集群详情页，在右上角单击**重启**，然后在下拉框中选择重启类型，包括集群重启、节点重启、Web 节点重启。
3. 正常集群在单击**重启**后出现如下界面，选择“操作类型”，然后单击**重启**，集群状态变为处理中，等待集群状态恢复为正常，重启操作即完成。当前集群重启支持滚动模式和全量模式，滚动模式会逐个节点进行重启，确保集群持续可用；全量模式则会同时重启所有节点，可在集群hung住等不可用的场景下，快速拉起集群，该操作会同时重启全体节点，重启速度快，但期间业务无法访问。

4. 当集群状态为 YELLOW 或 RED，或集群存在无副本索引的情况下，界面中会出现强制重启提示，例如下图只有勾选了强制重启才能对集群发起重启操作。在此种情况下用户可参考 注意事项 第三条中的描述进行操作。重启操作发起后，集群状态会变为处理中，等待集群状态恢复为正常，重启操作即完成。


重启
操作类型
集群重启
节点重启
Web节点重启
集群ID
集群名称
• 此操作需要重启集群，当前集群健康状态为红色，重启可能导致集群数据读写会受到影响，建议调整集群状态正常后，再进行操作。
已了解风险，仍强制滚动重启：期间服务访问性能可能短时间受影响。
全量模式重启：同时重启全部节点，重启速度快，但期间业务无法访问。
不推荐
升级同集群版本内核
重启
取消

销毁集群

最近更新时间：2025-09-12 14:15:02

操作场景

因为业务发展需要，当原有 TSearch 集群无法满足您的需求，如果是集群配置无法满足需求，您可以通过调整集群配置把集群调整到合适的规格，操作方法请参见 调整配置 。如果是需要退货时，您可以在控制台对集群进行销毁，避免服务继续运行而产生费用。

> **注意：**
> - 若您主动销毁包年包月的集群，销毁即刻起7天（7 × 24小时）集群处于隔离期，在隔离期内您可以通过续费方式恢复集群访问；
> - 若您主动销毁按量计费的集群，集群将立刻被销毁，数据无法恢复，请谨慎操作。

不同计费模式退费说明

不同计费模式下的集群，销毁集群的条件有所区别：

- 预付费包年包月的集群，如果集群还未到期，需要提前销毁时，会有条件限制：只支持1个集群的5天无理由退还和3个集群的普通自主退还，详细说明请参见 包年包月退费 。
- 后付费按量计费的集群，根据使用量计费，可以随时销毁集群，销毁后，就不再产生费用。操作步骤

1. 登录 检索分析服务 TSearch 控制台 。
2. 销毁集群有两个入口，分别位于集群列表页和集群详情页。

> **说明：**
> 包年包月的集群销毁，需要满足一定的条件，详细说明请参见 包年包月退费 。

2.1 在集群列表页，选择相应的集群，在“操作”列选择**更多 > 销毁**。

2.2 单击集群名称进入集群详情页，在右上角选择**更多操作 > 销毁**。

3. 在“销毁集群”页面，单击**确定**，系统将清空集群数据，并回收资源，**数据清空后，无法恢复**。包年包月集群会把剩余未使用的费用以赠送金的形式退还到您的账户，折扣和代金券不予退还。


销毁集群
销毁选中的TSearch集群?
• 销毁后，集群将以"已过期"状态保留7天，请提前备份数据。
• 集群销毁后，退款金额将按购买使用的现金和赠送金支付比例退还至您的腾讯云账号。
• 若购买集群时享有折扣或代金券，折扣和代金券不予退还。
集群ID
集群名称
已阅读并同意退费规则
确定
取消

集群多可用区部署

最近更新时间：2025-09-12 14:15:02

创建支持多可用区的集群

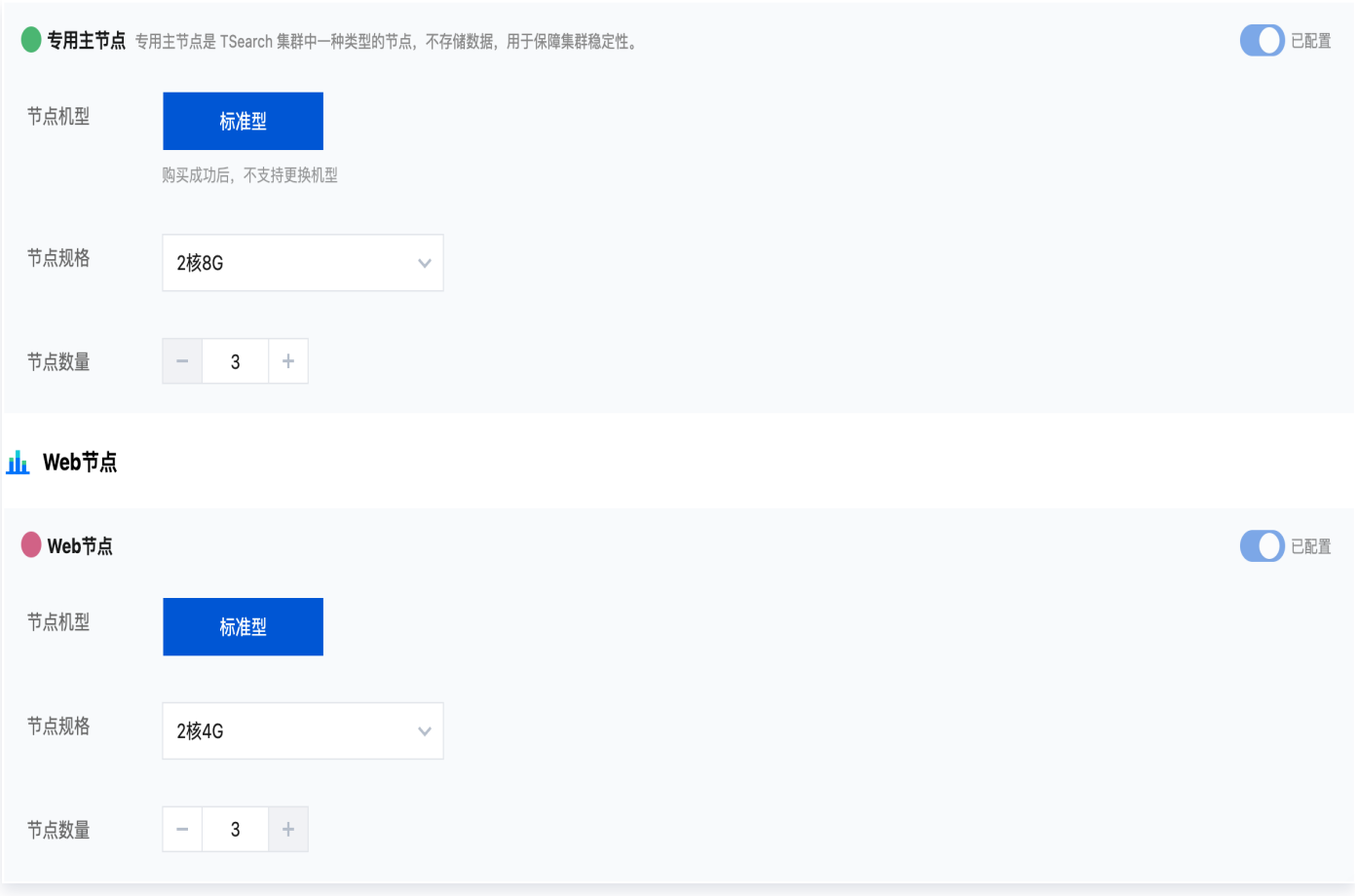
1. 进入 腾讯云检索分析服务 TSearch 购买页 选择“可用区部署模式”，设置多可用区网络。

> **说明：**
> - 由于要开启多可用区容灾的集群，必须先开启专用主节点，且最小为三个，所以能支持多可用区容灾功能的地域必须最少支持三个机房。目前仅有部分大地域如北京、上海、广州、南京支持多可用区容灾的功能，其他暂不开放的地域随着腾讯云机房的建设，我们也会持续的加入这个功能。
> - 其他参数设置与单可用区基本一致。

2. 以广州地域为例，在可用区部署模式中，支持创建双可用区和三可用区集群，用户需要选择和可用区数量相等的可用区及子网。

- 节点数量会自动按可用区的倍数调整。为了保证集群的稳定性及可靠性，当用户选择多个可用区时，专用主节点默认开启，可以选择专用主节点个数为三个或五个。
- 专用主节点会均匀的分布在三个可用区中，保证一个可用区发生不可用的情况下，不会出现超过一半的专用主节点不可用的情况，始终保持集群有超过法定的主节点选择个数，保证了集群的可靠性。

- 对于选择了两个可用区的集群，还有一个隐藏专用主节点会放在该地域非数据节点所在可用区外的别的可用区内，保障集群高可用。



集群多可用区容灾原理

数据节点

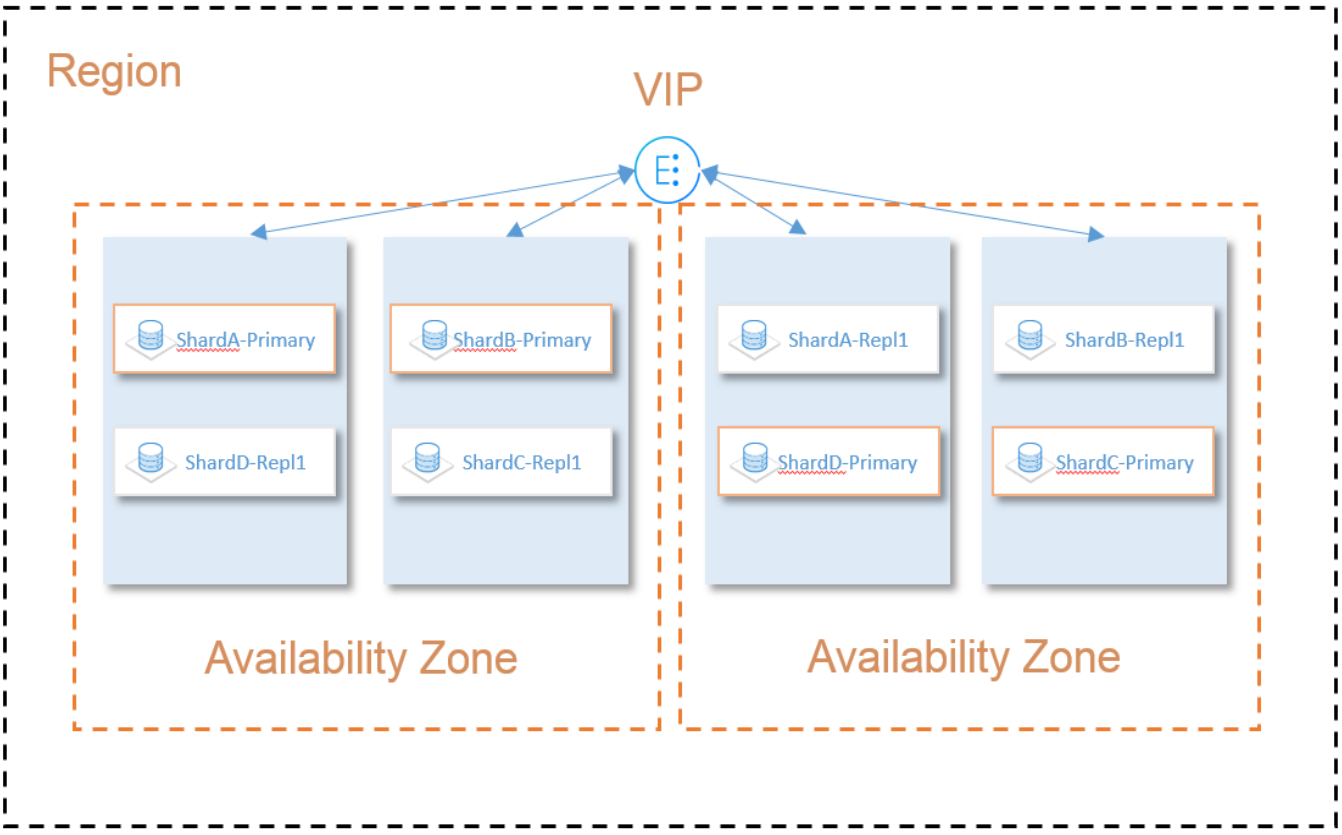
为了保证多可用区容灾的功能生效，用户需要遵守以下原则：

1. 购买集群的数据节点个数为可用区个数的倍数，例如选择两个可用区容灾，那么数据节点个数应该为2、4、6、8...以此类推。
2. 索引分片至少设置1个副本，即保证集群始终有两份以上的数据。

TSearch 会自动的将用户所购买的数据节点均地部署在用户所选择的可用区中，且所部署的数据节点含有可用区感知的功能。该功能使用户数据的副本会分布到多个可用区中，保证单个可用区仅有一份副本。
TSearch 提供了 VPC 内负载均衡功能，用户通过我们提供的 VIP 连接集群，通过 TSearch 的 API 进行数据读写及集群控制操作。

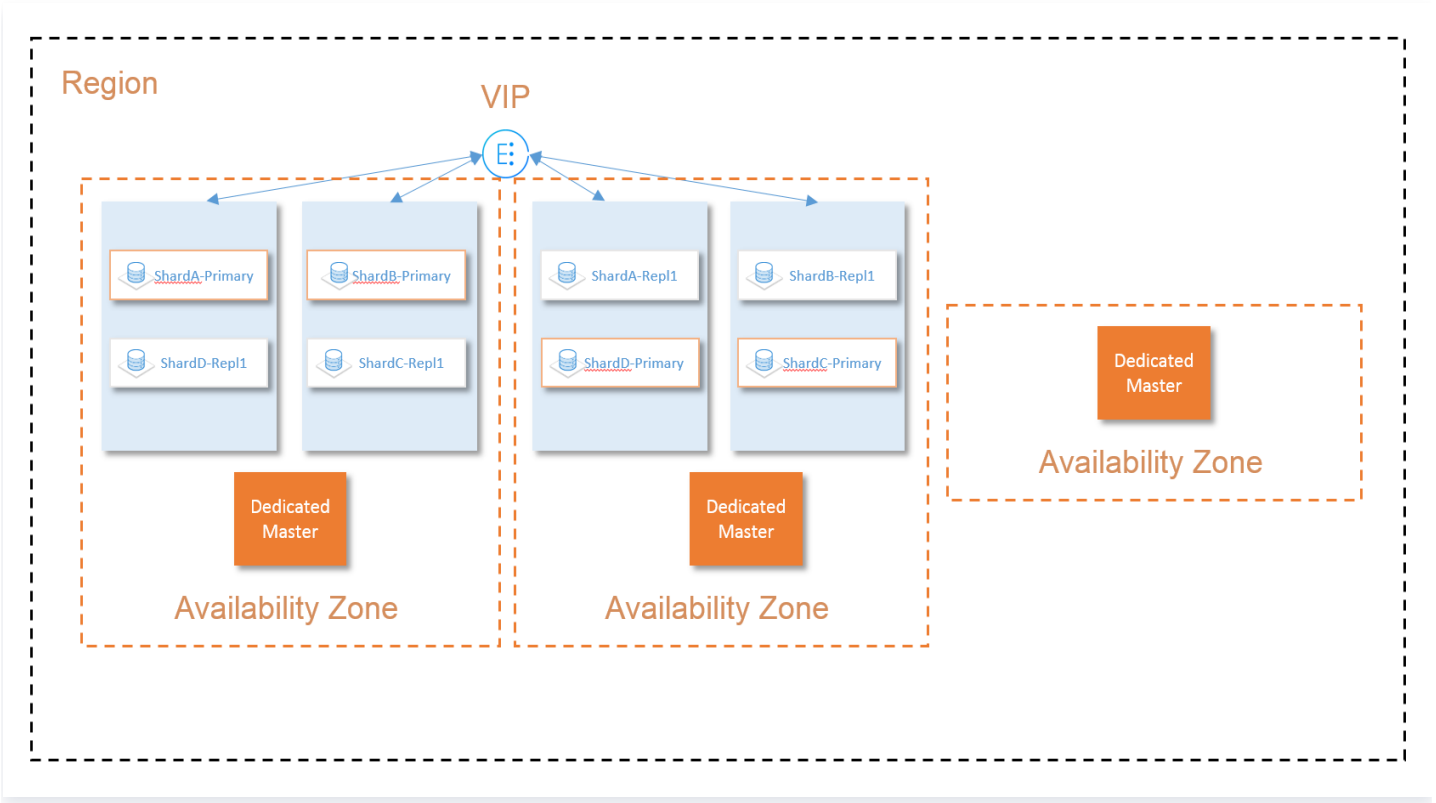
- 此 VIP 绑定了集群内部的所有数据节点，并提供负载均衡功能，用户所有请求会平均分布到集群的各个数据节点上。
- 此 VIP 还带有健康检查功能，如一个周期内多次检查确认某节点没有响应，健康检查功能会暂时从 VIP 的绑定列表中摘除有问题的节点，直到节点恢复正常。

这样就保证了当某个节点宕机，或者某个机房的可用区不可用的情况下，自动地剔除有问题的节点，保证用户的客户端不会请求到有问题的节点。从而在可用区故障的情况下，实现客户业务无感知的切换，提高了客户业务的稳定性。



专用主节点

为了提高集群的可靠性，用户在使用多可用区容灾功能时最少要创建三个专用主节点，且分布在三个不同的可用区中。即便用户选择的是双可用区部署数据节点，我们也会自动的为用户再多选择一个可用区部署专用主节点。这种部署方式，可以保证当一个可用区不可用时，集群依旧有超过半数的法定主节点选举个数，可以保证集群的正常选主。


Region
VIP
ShardA-Primary
ShardB-Primary
ShardD-Repl1
ShardC-Repl1
Dedicated Master
Availability Zone
ShardA-Repl1
ShardB-Repl1
ShardD-Primary
ShardC-Primary
Dedicated Master
Availability Zone
Dedicated Master
Availability Zone

可视化配置—Web 节点

最近更新时间：2025-09-12 14:15:02

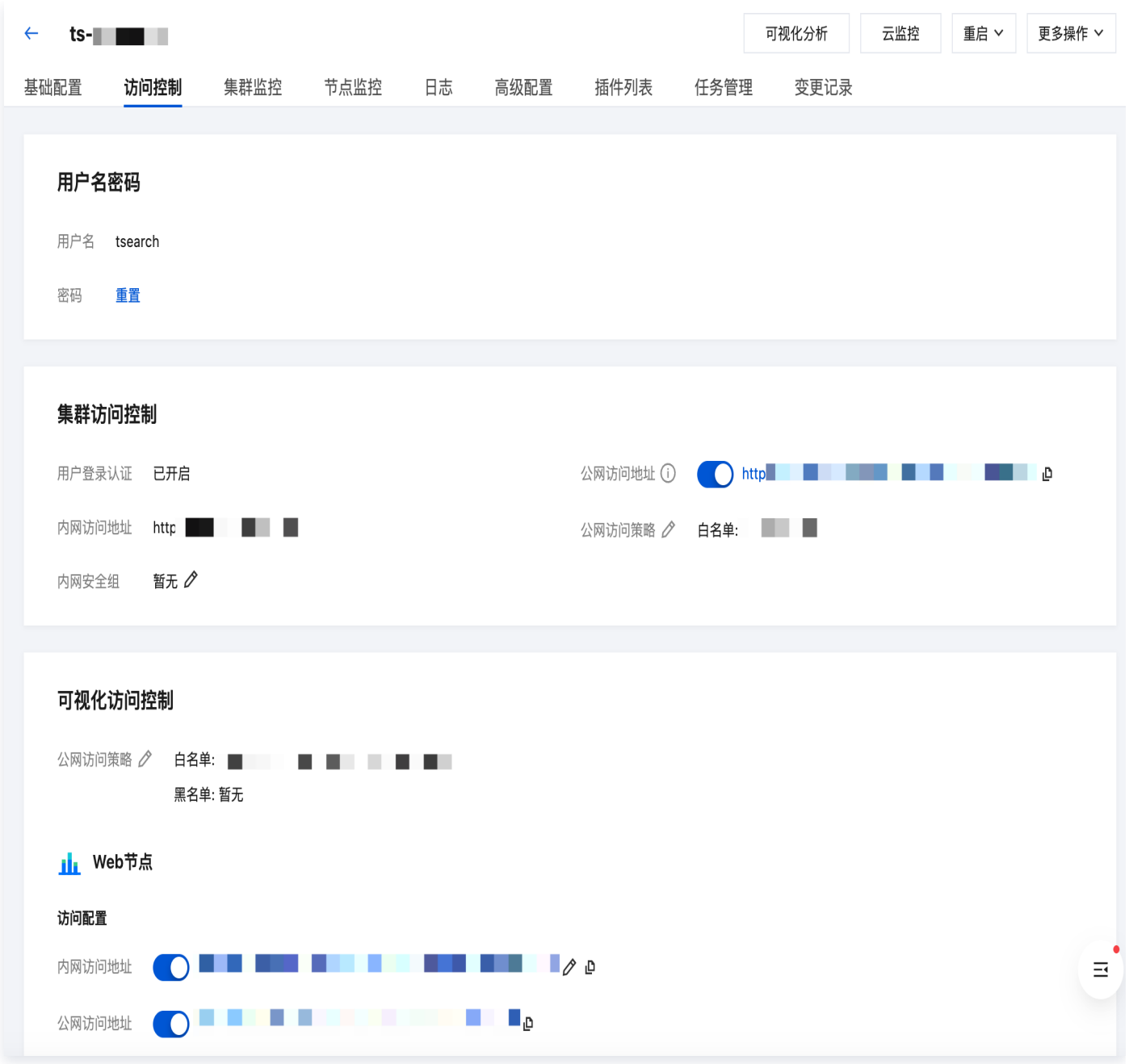
购买腾讯云检索分析服务 TSearch 集群时，您需要同步购买 Web 节点，此节点将支持 Kibana 组件的使用。您可以根据业务需要，升级该 Web 节点至更高规格。

背景信息

Kibana 能够帮助用户实时监测和管理集群状态。腾讯云检索分析服务 TSearch 提供 Kibana 登录入口，能够调整 Kibana 基础配置，支持动态调整 Kibana 节点资源，为用户业务提供扩展性。

操作步骤

1. 登录腾讯云 检索分析服务 TSearch 控制台，单击**ID/名称**访问目标集群，在访问控制页的“可视化访问控制”以及页面右上角 **可视化分析** 按钮直接进入 Kibana 控制台。



2. 访问 Kibana 控制台，在登录界面输入用户名和密码登录。用户名默认为 tsearch，密码为集群创建时设置的密码。
3. 支持通过公网和内网访问 Kibana，并在**公网访问策略**模块进行配置，配置方法参考 TSearch 集群访问控制 。
4. 登入 Kibana 后，支持在控制台进行数据查询、仪表板制作等操作，详细指引请参见 Kibana Guide 。

集群扩缩容
调整配置

最近更新时间：2025-09-12 14:15:02

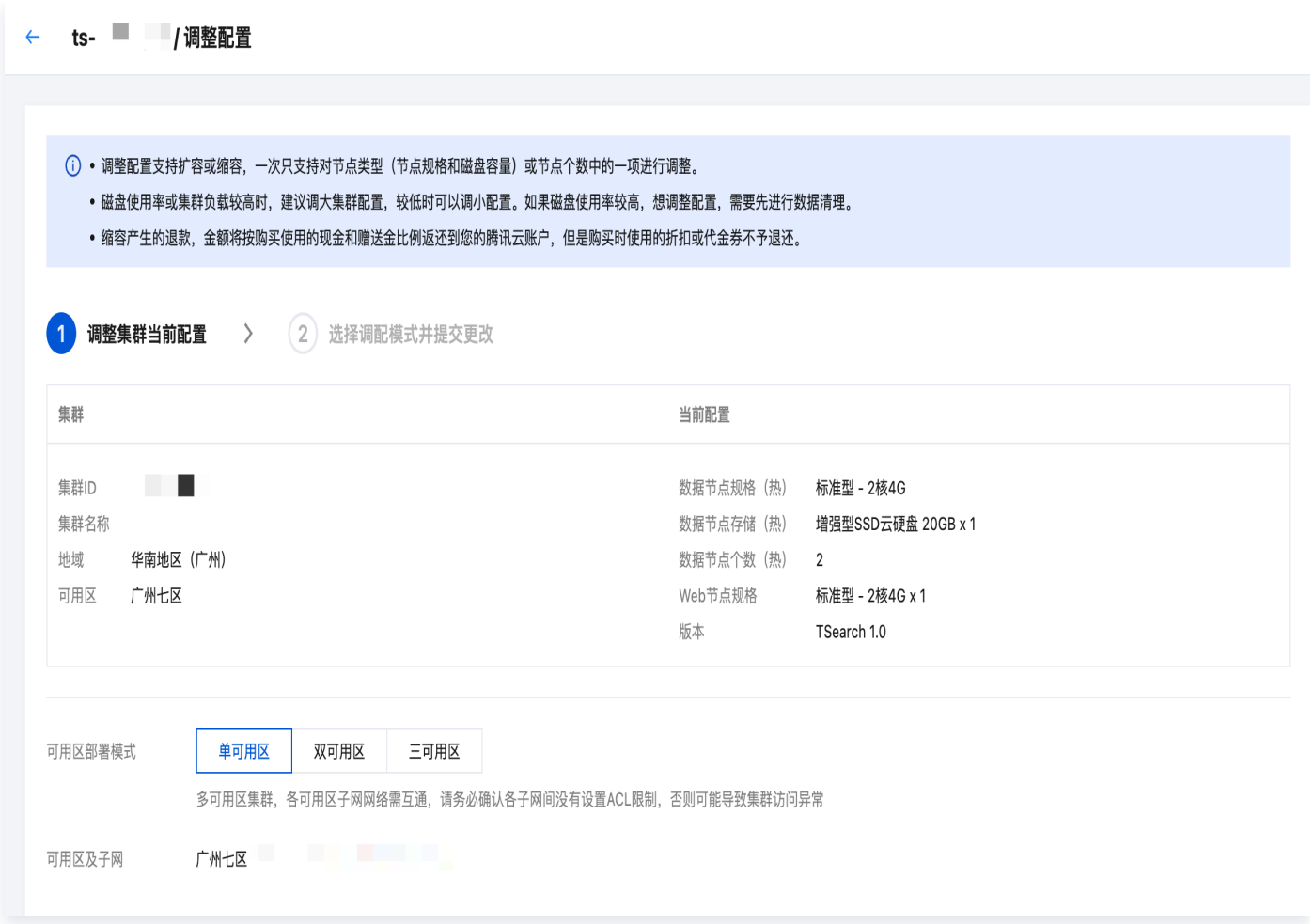
应用场景

在业务发展过程中，集群的数据量和访问量是动态的。随着业务的发展，数据量和访问量都在增大。当集群规模跟实际业务需求不太匹配时，可以动态调整集群的配置，对集群进行扩容；也可以在数据量和访问量不大时，暂时调小集群的规模。腾讯云 TSearch 支持对节点个数、节点类型和单个数据节点的存储量等进行配置调整。此外，也支持对集群的专用主节点和 Web 节点（ Kibana 节点）进行调整，具体见以下说明。

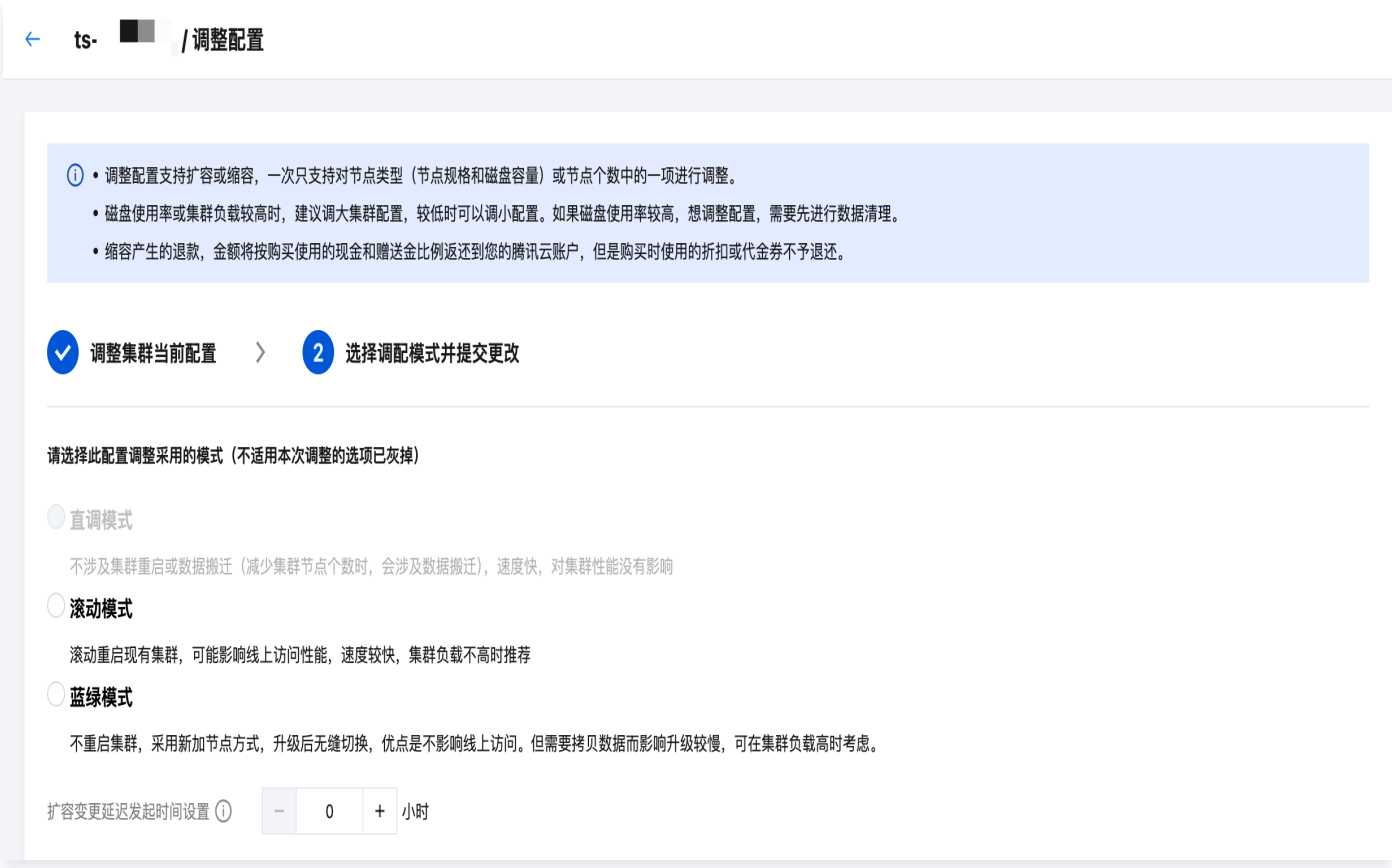
操作步骤

1. 登录 检索分析服务 TSearch 控制台，进入集群列表页。
2. 集群调整配置有两个入口，分别位于集群列表页和集群详情页。
   - 在集群列表页，选择需要扩容的集群，选择**操作 > 更多 > 调整配置**。
   - 单击集群名称进入集群详情页，在右上角选择**更多操作 > 调整配置**。
3. 在弹出的调整配置中，根据业务需求对集群节点规格或节点数量进行调整，单击**确定**。

> **说明：**
> 一次操作仅支持对节点类型（节点规格和存储）或节点数量中的一项进行扩容或缩容调整，不支持同时调整。



4. 支持根据业务需要选择不同的调配模式，了解调配模式原理查看 集群变配建议和原理介绍 。



5. 配置调整开始后，集群状态为处理中，待集群状态变为正常，即可正常使用。
6. 可以在集群详情页，查看集群的变更进度。

调整配置费用说明

不同计费模式下的集群，调整配置时，费用结算方式会有所不同。

- 预付费包年包月的集群，在调整配置时，会根据集群剩余有效期以及新的配置的定价，计算需要的费用，具体可以参见 调整配置费用说明 。
- 后付费按量计费的集群，计费周期为秒，当配置调整完成后，下一秒费用按新的配置定价进行结算。

集群变配建议和原理介绍

最近更新时间：2025-09-12 14:15:02

随着业务的不断发展，TSearch 集群的数据规模和访问量也越来越大，当集群规模与配置逐渐和业务实际需求不匹配时，您可选择及时扩容 TSearch 集群以满足业务的需要。

变配建议综述

当业务遇到瓶颈时，采取何种方式变配是我们需要面对的问题，您可参考下表结合业务情况进行判断（表中提到的滚动模式和蓝绿模式见后文说明）：

云盘机型

变配方式	适用场景	实现原理介绍	特点说明
增加磁盘容量	计算资源充足，磁盘存储空间不足	直接调整，逐个对集群中节点上挂载的磁盘扩容容量（对加密云盘将采取蓝绿模式）	平滑扩容不停服，流程时间短，每个节点预计30秒
增加磁盘数量	计算资源充足，磁盘存储空间不足，或 IOPS 和吞吐量不足	蓝绿模式	蓝绿模式耗时长但对线上业务无影响
添加更多节点	节点规格较高，但集群整体计算资源不足	直接调整，向集群中加入更多同等规格的节点	平滑扩容不停服，流程时间短，不受节点数量影响，预计5－10分钟完成
提升节点规格	节点规格较低，且集群整体计算资源不足	可选择滚动模式或蓝绿模式	滚动模式较快，蓝绿模式耗时长但对线上业务无影响

本地盘机型

变配方式	适用场景	实现原理介绍	特点说明
添加更多节点	节点规格较高，但集群整体计算资源不足	直接调整，向集群中加入更多同等规格的节点	平滑扩容不停服，流程时间短，不受节点数量影响，预计5－10分钟完成
提升节点规格	节点规格较低，且集群整体计算资源不足	蓝绿模式	蓝绿模式耗时长但对线上业务无影响

说明：

- 上表侧重说明场景和变配方式的对应关系，其中实现原理介绍是针对单独调整当前这一项配置的情况，当同时调整多项配置时（如同时调整节点规格、磁盘容量、磁盘数量），实际的方式会有差异。
- 一次扩容操作只支持对纵向（节点规格、磁盘容量、磁盘数量）或横向（节点个数）的其中一个方向进行调整，不支持纵向横向同时调整。
- 对本地盘机型而言，其挂载的本地盘容量大小和个数是固定的，不支持如云盘机型那样单独对节点磁盘进行调整，而是对节点整体进行调整。

变配基本概念

当集群变配时，常见的有直接调整、滚动模式、蓝绿模式等，下面将分别介绍这几种变配模式：

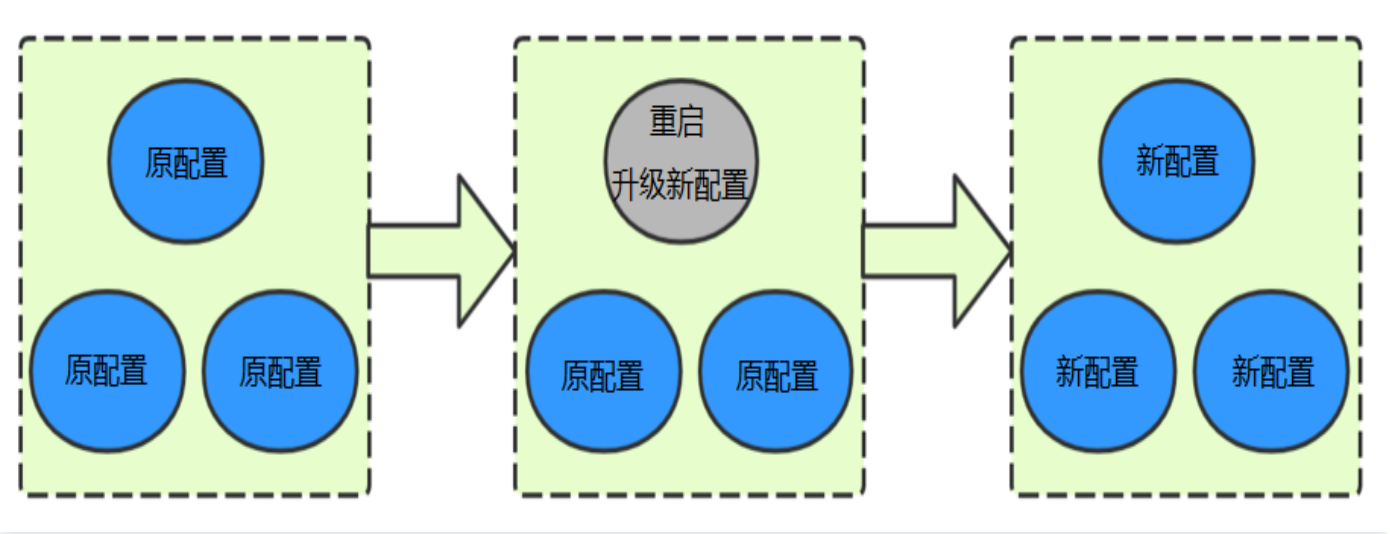
直接调整

直接调整不涉及对集群较重的操作（例如重启或拷贝数据），一般来说对线上业务无影响，且变配速度较快。例如，集群横向添加更多节点、或纵向增加磁盘容量等都是直接调整。

滚动模式

对集群中节点逐个滚动重启完成变配，期间系统服务不间断，但可能影响线上访问性能。

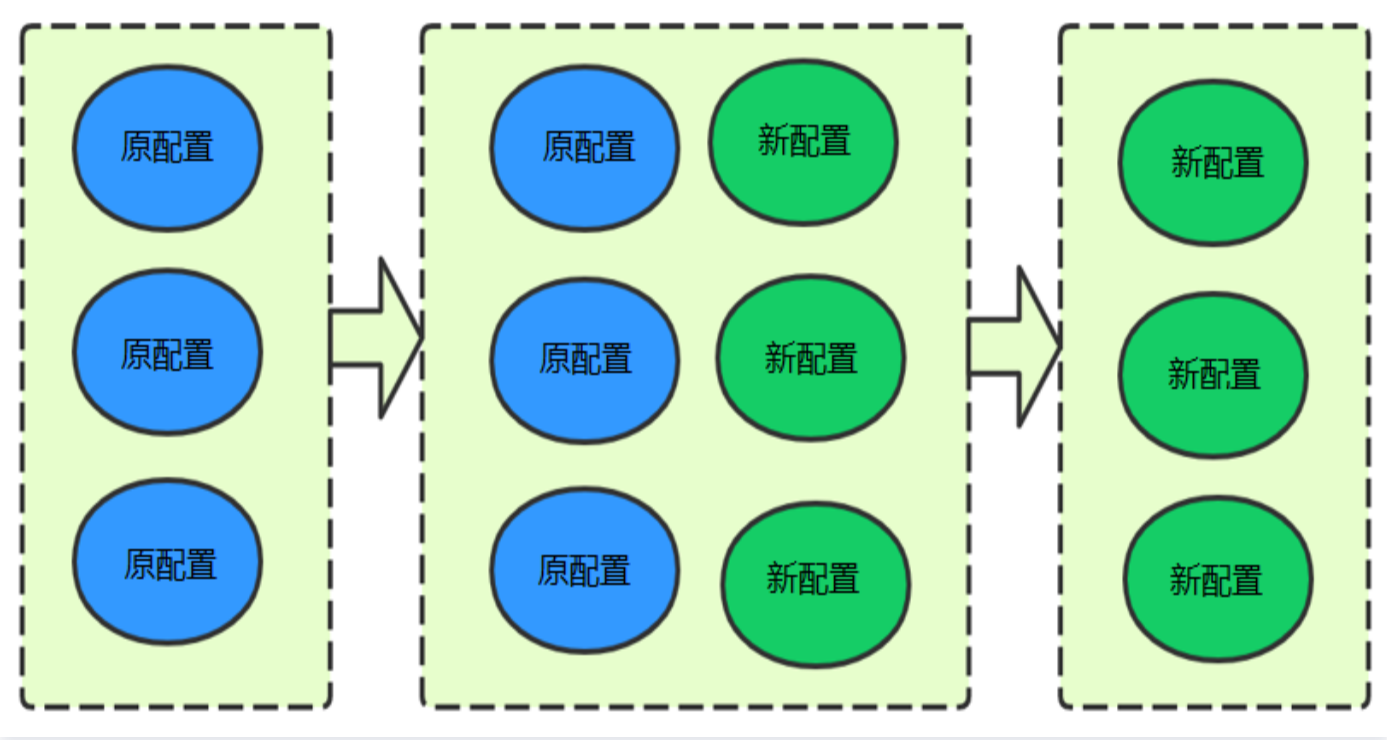
- **优点：**由于不需要迁移数据，扩容的时间不受集群数据规模影响，变配能够快速完成。适合于集群遇到性能瓶颈，期望快速完成扩容变配的场景。
- **缺点：**需要滚动重启集群的每一个数据节点，因此如果集群存在无副本情况下会有丢失数据的风险，且在变配过程中会不停的有节点离开集群和加入集群，会对集群的性能有一定的影响，因此不建议在业务高峰期间选择此种变配方案。



蓝绿模式

不重启集群，为原集群添加相同数量的新节点（配置为需要的节点且拷贝原节点数据），完成配置后，无缝切换并剔除掉原节点，**变配完成后节点 IP 会发生变化**。

- **优点：**变配过程中业务不停服，且扩容非常平滑，适合对集群可用性较高的场景。
- **缺点：**因涉及数据拷贝，变配耗时和集群数据量成正比，变配时间从数分钟到数天或更长。

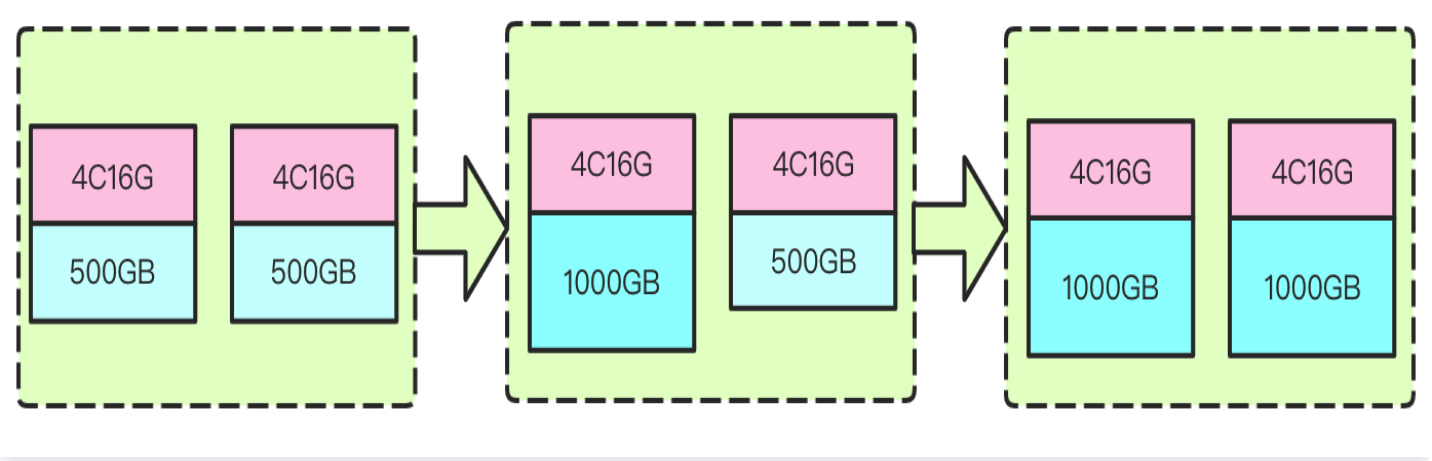


主要变配场景的原理详述（以云盘机型为例）

增加磁盘容量

不改变计算资源配置的情况下提升每个节点的磁盘容量，以提升集群整体的存储量。

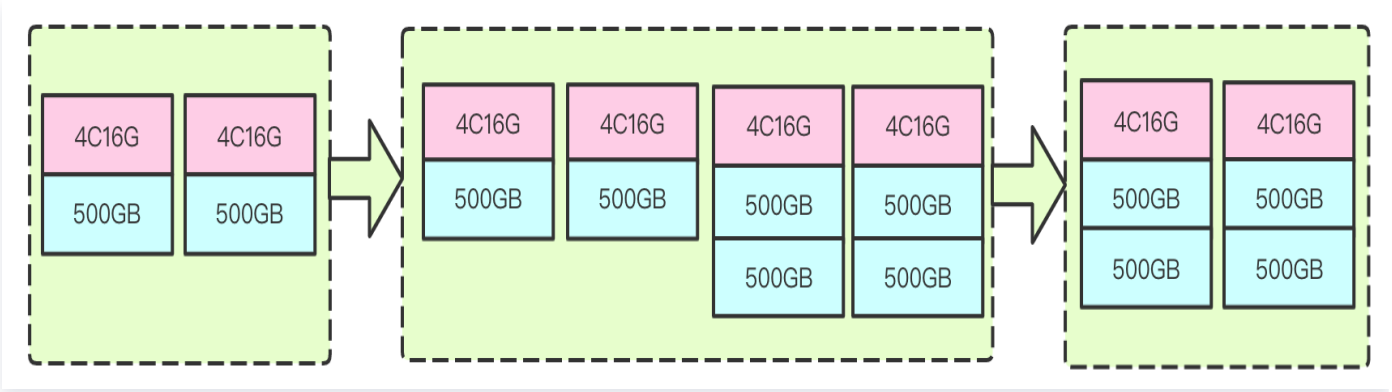
原理：扩容单盘容量的流程是对集群中每个数据节点上挂载的磁盘依次进行纵向扩容操作。以达到提升整个集群磁盘存储容量的目的。例如，磁盘原有容量为1000GB，将该磁盘容量提升至5000GB。此种扩容方式无需重启节点和集群，对业务使用集群不造成任何影响。由于集群磁盘容量的扩容是采用数据节点滚动扩容的方式，因此扩容时间与集群节点数量成正相关。预计一个节点扩容需要10秒，以10个数据节点的集群为例，预计扩容磁盘的时间在2分钟左右。流程原理示意图如下图所示：



适用场景：适用于计算资源充足，但是磁盘容量不足的情形。例如，客户集群 CPU 负载和内存使用率都较低，但是磁盘平均利用率较高，且数据都比较重要，不能通过删除数据的方式来释放磁盘空间。此时，客户可选择只扩容磁盘容量的方式来扩容集群

增加磁盘数量

不改变计算资源配置的情况下提升每个节点的磁盘数量，以提升集群整体的存储量、IOPS 和吞吐量。
原理：将节点上挂载的云盘块数进行提升，例如，原集群中每个数据节点上只挂载了一块盘，通过此流程，能够实现集群中每个数据节点上挂载多块云盘，目前仅支持**蓝绿模式**这一种变配方式。
流程原理示意图如下所示：向集群中加入挂载了多块云盘的同等数量的节点，然后将旧节点中的数据迁移至新节点，最后再将旧节点剔除集群。

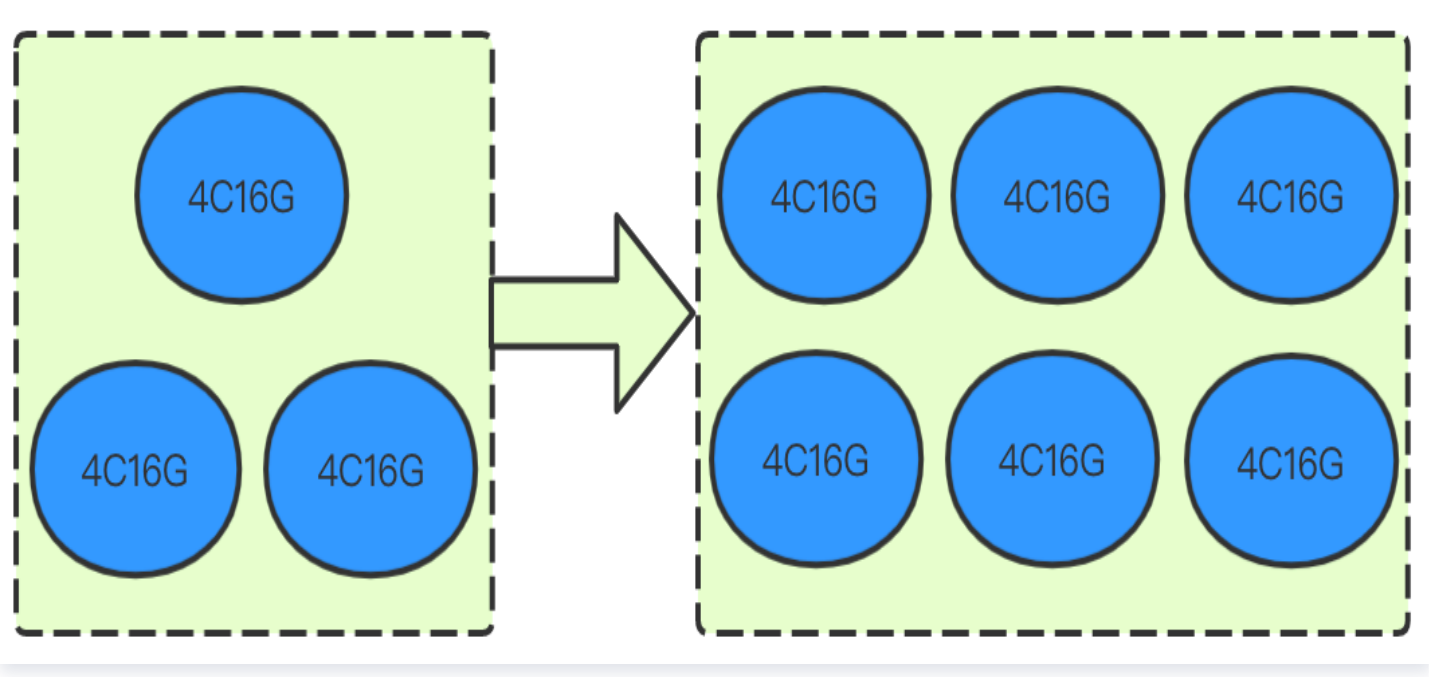


优势：在整个扩容流程中，集群平滑扩容，适用于对业务稳定性和可用性要求较高的场景。
不足：变配的时间受到集群数据量大小的影响，数据量越大，扩容流程时间越长。变配完成后，节点IP会发生变化。

添加更多节点

适用于集群的计算资源遇到瓶颈，如 CPU 负载持续较高，内存使用率居高不下，多次触发内存熔断，甚至内存出现 Out Of Memory 现象。此时您可通过对集群节点进行扩容，以提升集群整体性能。

原理：集群添加节点是指在不改变集群节点机型配置的情况下，通过向集群中加入更多节点来完成集群的扩容操作，其流程原理示意图如下所示：



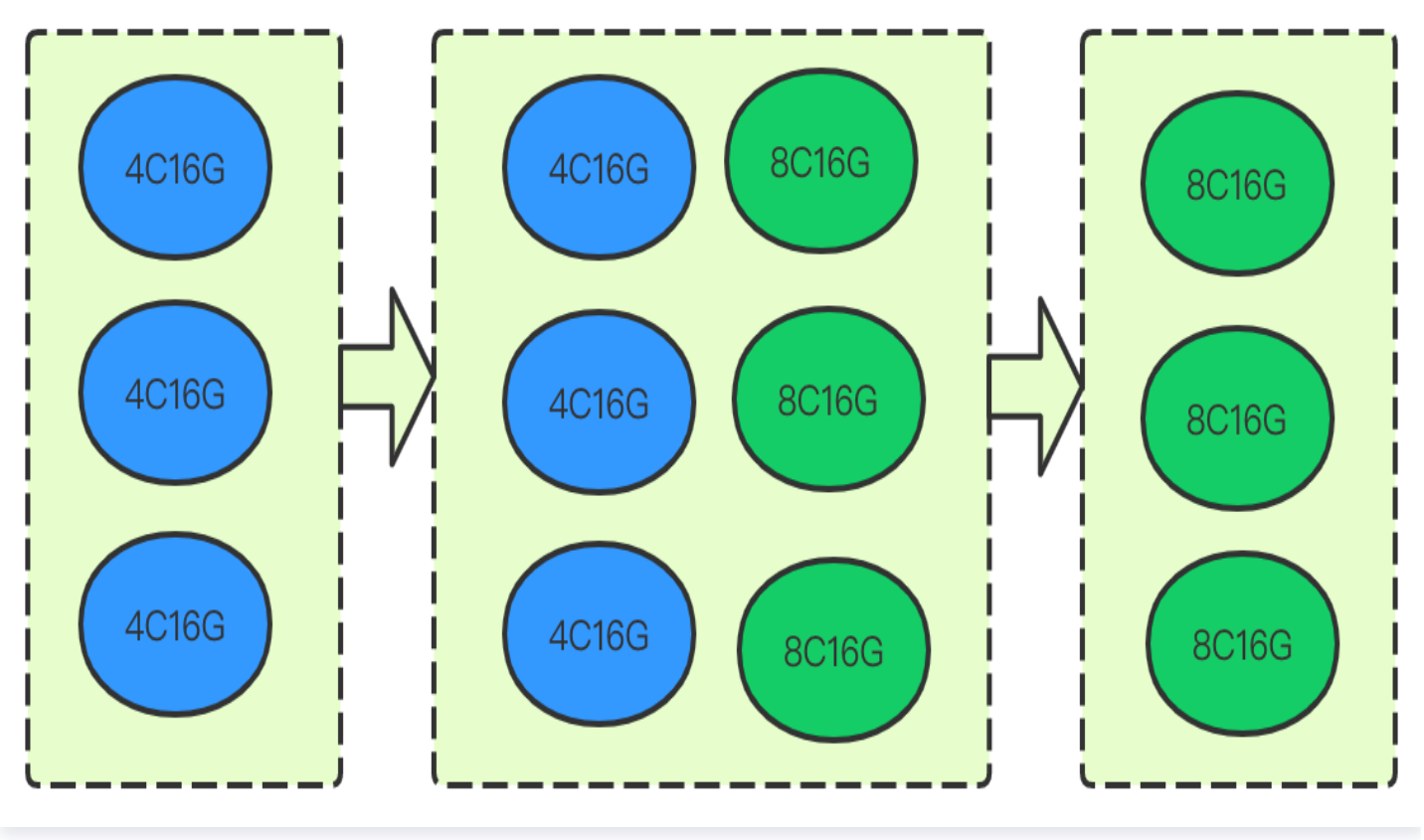
优势：主要是能够做到平滑扩容，扩容过程中不影响业务使用集群，并且由于在扩容流程中不涉及到新老数据节点间的数据搬迁，因此扩容时间不受节点数量影响，通常在5 – 10分钟完成。
适用场景：节点的配置已经很高，且期望进一步提升集群整体的读写性能，同时对扩容期间集群的稳定性要求较高。

提升节点规格

同样适用于集群的计算资源遇到瓶颈，如 CPU 负载持续较高，内存使用率居高不下，多次触发内存熔断，甚至内存出现 Out Of Memory 现象。可通过对集群节点提升规格，以提升集群整体性能。
原理：节点提升规格是指在不改变集群节点个数的情况下，整体提升集群中数据节点的计算资源配置，目前支持蓝绿模式和滚动模式两种变配方式。例如，将数据节点的配置从4核16G提升到8核16G。

采用蓝绿模式

原理是先将数量大小相等且更高规格的节点加入集群，然后将原节点上的数据全部搬迁到新节点上后，再将原节点剔除集群，以完成集群的扩容操作。其流程原理示意图如下所示：



这种方式的优势主要是能够做到平滑扩容，扩容过程不影响集群的使用和可用性。不足之处在于这种扩容流程需要将原节点中的数据迁移到新节点上，迁移时间受数据量的影响较大，因此如果集群的数据量在 TB 以上，且希望能够尽快完成扩容流程，可选择节点滚动模式，或者通过在 kibana 中设置如下属性，提升数据搬迁的速度：

```
PUT _cluster/settings  {     "persistent": {
"cluster.routing.allocation.node_concurrent_recoveries":  "8",
"indices.recovery.max_bytes_per_sec": "80mb"    }  }
```

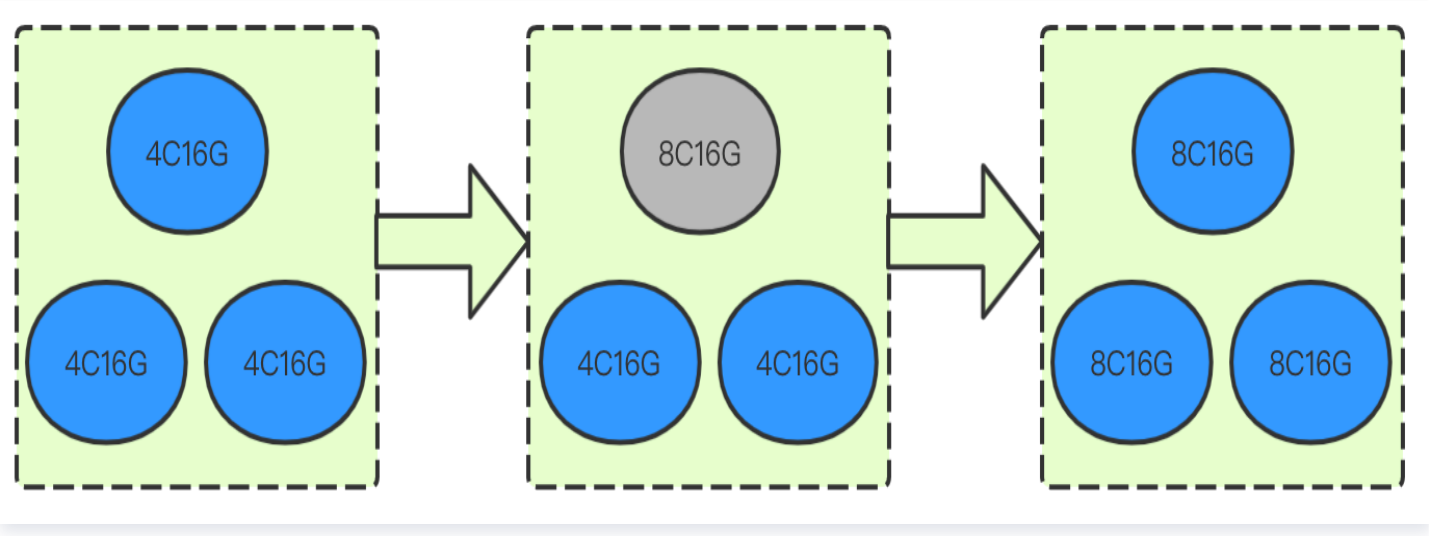
其中：

- **cluster.routing.allocation.node_concurrent_recoveries** 属性表示集群中每个节点上分片并发恢复的个数，默认是2。可根据老数据节点的 CPU 核数 × 4 来确定具体的值，但不要超过50。例如老数据节点为4核16G，则建议该值设置为16，老数据节点为16核64G，则建议该值设置为50，如果发现调大了该值后集群的稳定性受到影响，可适当减小该值。
- **indices.recovery.max_bytes_per_sec** 属性表示节点之间数据传输的最大带宽限制，默认是40mb。该值不宜设置的过高，否则会破坏集群的稳定性，客户可以5mb为步长，逐步调整该限制值，并持续观察集群的稳定性，最终选择一个相对平衡的值。

适用场景：节点的配置较低，期望提升集群整体的读写性能，但是对集群在扩容期间的稳定性要求较高，且对扩容时间不是很紧急，这种情况下可选择数据搬迁的方式进行扩容。

采用滚动模式

原理是逐个对集群中的节点通过重启的方式完成规格配置的扩容操作。由于该变配方式不涉及到数据迁移，因此变配流程时间不受集群数据规模的影响。变配时间与节点个数成正比，每个节点预计需要3 – 5分钟。其流程原理示意图如下所示（灰色表示重启中）：



适用场景：适合集群节点配置较低，遇到性能瓶颈，对集群稳定性要求不高，期望快速提升集群性能的场景，由于集群在变配过程中节点滚动重启，会有节点不停的离开集群和加入集群，因此建议在业务低峰期操作。

提升节点规格方案对比：

提升节点规格方案	优势	不足
蓝绿模式	平滑变配，业务无感知	• 蓝绿模式，流程时间不可控 • 节点 IP 发生变化
滚动模式	• 变配流程时间短，和集群数据量无关 • 节点 IP 不变	• 节点滚动重启，对业务有一定影响 • 无副本索引有丢数据风险

集群配置—YML 文件配置

最近更新时间：2025-09-12 14:15:02

1. 您可以通过修改腾讯云检索分析服务 TSearch 实例的 YML 参数配置，配置常用的参数。

使用步骤

查看配置项

登录 检索分析服务 TSearch 控制台，单击需要修改配置的**集群 ID/名称**，进入集群详情页，然后单击**高级配置**，即可查看配置项参数。
在如下列表中可见当前值和集群配置值，原因是存量集群可能进行过各类非标操作，导致集群实际配置值与默认值不一致，因此控制台作出展示。

参数名	参数说明	当前值	集群配置值	取值说明
indices.queries.cache.size	设置过滤器缓存的内存大小	未配置	10%	正整数，范围为[1, 100]，默认值为10%
indices.queries.cache.count	缓存条目的总大小	未配置	10000	正整数，范围为[1, 2147483647]，默认值为10000
http.max_initial_line_length	HTTP请求URL的最大长度	未配置	4kb	正整数，范围为[1, 20000]，默认值为4KB
http.max_header_size	请求头的最大大小	未配置	8kb	正整数，范围为[1, 20000]，默认值为8KB
http.max_content_length	设置请求内容的最大大小	未配置	100mb	正整数，范围为[1, 2000]，默认值为100MB
reindex.remote.whitelist	远程TSearch集群访问地址白名单	未配置	[]	填入合法的白名单域名或IP+端口(可选)，如:127.0.0.1:9200,127.0.0.1，多个域名或IP之间支持以英文逗号分隔
thread_pool.write.queue_size	文档写入队列大小	未配置	10000	正整数，范围为[1000, 2147483647]，默认值为10000
thread_pool.search.queue_size	文档搜索队列大小	未配置	1000	正整数，范围为[1000, 100000]，默认值为1000
thread_pool.get.queue_size	文档实时获取队列大小	未配置	1000	正整数，范围为[1000, 2147483647]，默认值为1000

修改配置项

1. 单击**修改配置**，可对相应的配项进行修改，配置项的输入限制，见表格最后一列“取值说明”。

2. 单击**确定修改**，新的配置项参数将会应用到您的集群，会滚动重启集群。如果您集群中的索引有副本，重启后不会对您的业务造成影响，但建议在业务低峰期操作。

> **注意：**
> 如果集群健康状态为 YELLOW 或 RED，或集群存在无副本索引时，修改集群配置项对话框会有强制重启的提示和选项框，此时进行更新配置操作有较大风险，建议修复集群状态，为所有索引添加副本，再进行修改配置操作。

如果用户已了解该操作风险，仍要进行更新配置操作，单击**确定**进行重启。详情参考下图：

确定后，腾讯云 TSearch 实例会进行重启，重启过程中可在集群变更记录中查看进度。重启成功后即可完成 YML 文件的配置。

支持的参数

参数	说明	默认值
indices.queries.cache.size	设置过滤器缓存的内存大小	10%
indices.queries.cache.count	缓存条目的总大小	10000

配置 reindex 白名单

参数	说明	默认值
reindex.remote.whitelist	远程 TSearch 集群访问地址白名单	`[]`

自定义队列大小

参数	说明	默认值
thread_pool.write.queue_size	文档写入队列大小	10000
thread_pool.search.queue_size	文档搜索队列大小	1000
thread_pool.get.queue_size	文档实时获取队列大小	1000

自定义请求配置

参数	说明	默认值
http.max_initial_line_length	HTTP请求URL的最大长度	4kb
http.max_header_size	请求头的最大大小	8kb
http.max_content_length	设置请求内容的最大大小	100mb

如果有其他配置项自定义设置需求，可通过 售后支持 进行反馈。

插件配置

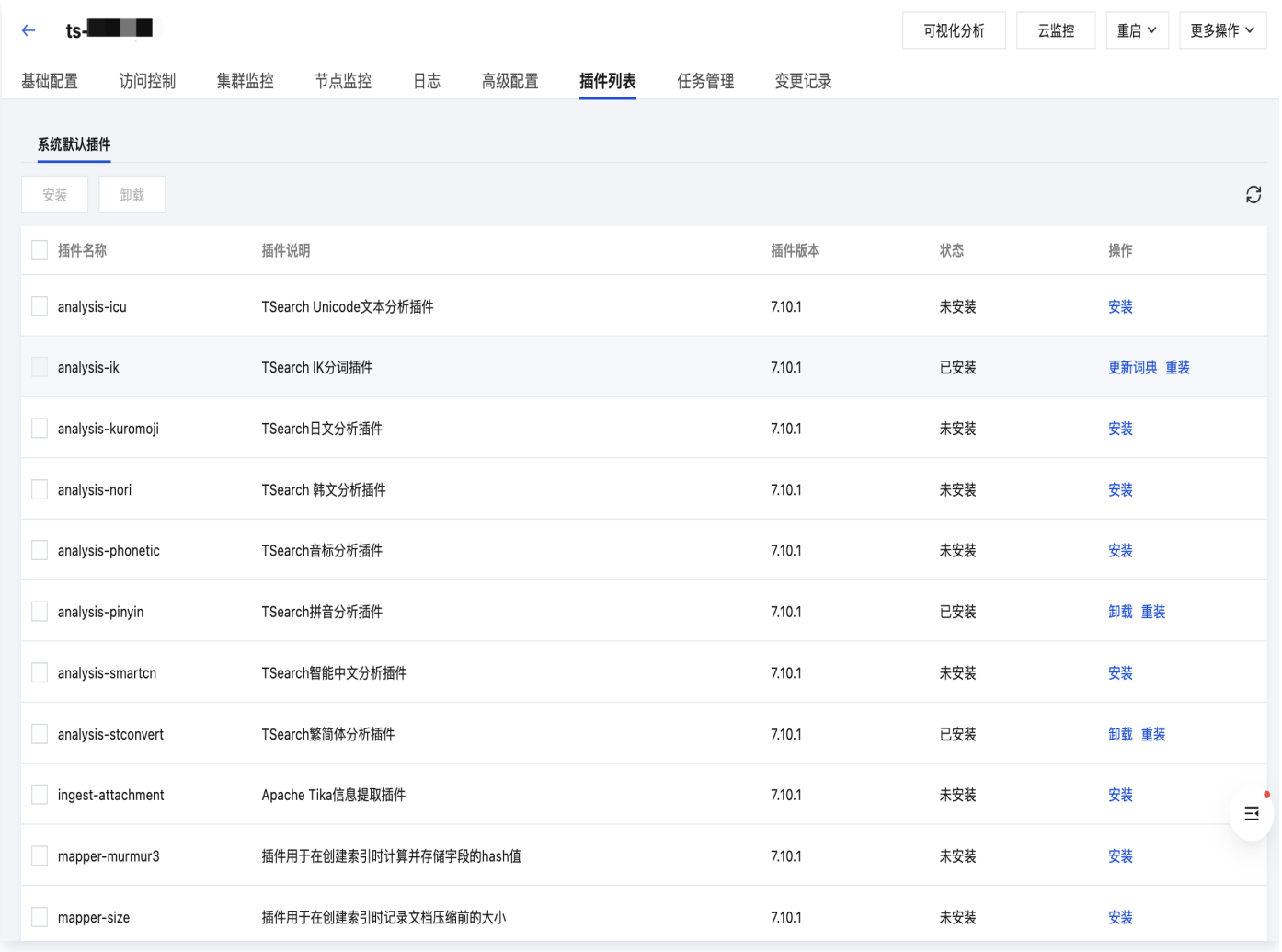
最近更新时间：2025-09-12 14:15:02

腾讯云检索分析服务 TSearch 提供了插件功能。当您购买了腾讯云 TSearch 实例后，您可以根据需求在插件列表页面安装或者卸载这些插件。本文介绍安装或卸载腾讯云 TSearch 插件的方法。

> **注意：**
> 安装或者卸载插件将触发集群滚动重启，请确认后操作。

操作步骤

1. 登录 检索分析服务 TSearch 控制台 。
2. 在集群列表页，单击 **ID/名称** 进入集群详情页。
3. 单击**插件列表**，进入插件列表管理页面。



4. 单击对应插件右侧操作栏下的**安装**或者**卸载**。

5. 在弹出的对话框中，勾选注意事项，确认无误后单击确定。确认后，插件变更操作开始，集群将会重启，可以在**集群变更记录**中查看变更进度。

插件信息

腾讯云检索分析服务 TSearch 支持的插件如下：

插件名称	默认状态	描述	支持的操作	使用限制
analysis-icu	未安装	TSearch Unicode文本分析插件	安装、卸载	仅 ES 索引支持
analysis-ik	已安装	TSearch IK 分析插件，默认不能卸载。	更新词典、重装	仅 ES 索引支持
analysis-kuromoji	未安装	TSearch 日文分析插件	安装、卸载	仅 ES 索引支持
analysis-nori	未安装	TSearch 韩文分析插件	安装、卸载	仅 ES 索引支持
analysis-phonetic	未安装	TSearch 音标分析插件	安装、卸载	仅 ES 索引支持
analysis-pinyin	已安装	TSearch 拼音分析插件	安装、卸载	仅 ES 索引支持
analysis-smartcn	未安装	TSearch 智能中文分析插件	安装、卸载	仅 ES 索引支持
analysis-stconvert	已安装	TSearch 繁简体分析插件	卸载、重装	仅 ES 索引支持
ingest-attachment	未安装	Apache Tika 信息提取插件	安装、卸载	仅 ES 索引支持
mapper-murmur3	未安装	插件用于在创建索引时计算并存储字段的 hash 值	安装、卸载	仅 ES 索引支持
mapper-size	未安装	插件用于在创建索引时记录文档压缩前的大小	安装、卸载	仅 ES 索引支持

repository-cos	已安装	插件用于存储 Elasticsearch Snapshot 到腾讯云 COS，默认不能卸载	无	ES 索引和 TSearch 索引均支持
repository-hdfs	未安装	插件提供了对 Hadoop 分布式文件系统（HDFS）存储库的支持	安装、卸载	仅 ES 索引支持
sql	已安装	开源 SQL 解析插件，基础版、白金版集群已带 sql 解析功能，不需安装	安装、卸载	ES 索引和 TSearch 索引均支持

监控与告警
查看监控

最近更新时间：2025-09-12 14:15:02

操作场景

腾讯云检索分析服务 TSearch 对运行中的集群，提供了多项监控指标，用以监测集群的运行情况，如存储、IO、CPU、内存使用率等。您可以根据这些指标实时了解集群服务的运行状况，针对可能存在的风险及时处理，保障集群的稳定运行。本文为您介绍通过 TSearch 控制台查看集群监控的操作。

操作步骤

1. 登录 检索分析服务 TSearch 控制台，在集群列表单击集群 **ID/名称**，进入集群详情页。
2. 选择**集群监控**页，可以查看集群整体的运行情况，选择**指标分组**，支持拆分查看数据节点（热）、数据节点（温）、专用主节点的集群监控指标。
3. 选择**节点监控**页，可以查看集群内各节点的运行情况和性能指标。

集群监控

在集群监控页，可以进行告警策略设置，同时也可以看到集群的监控数据信息。可通过选择不同的时间范围、指标分组和时间粒度查看集群总体状态和性能指标。

> **说明：**
> 也可通过 腾讯云可观测平台控制台 查看 TSearch 集群完整的监控指标。

ts-[redacted]

可视化分析 | 云监控 | 重启 | 更多操作

基础配置 | 访问控制 | 集群监控 | 节点监控 | 日志 | 高级配置 | 插件列表 | 任务管理 | 变更记录

告警策略

已配置自定义告警策略 去查看

监控数据

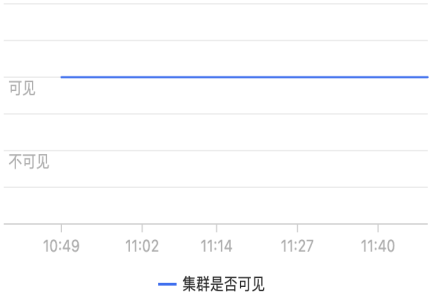
健康状态：绿色　在线节点数：3/3　总分片数：2　未分配分片数：0

时间范围 1小时 2025-05-22 10:49:30 ~ 2025-05-22 11:49:30 数据对比　导出数据 刷新

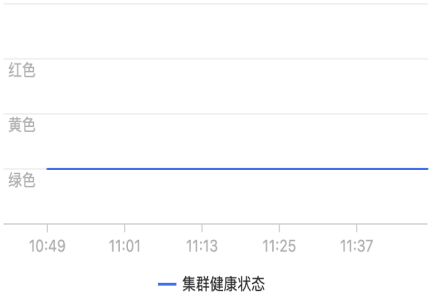
指标分组 数据节点　时间粒度 1分钟粒度

集群总体状态

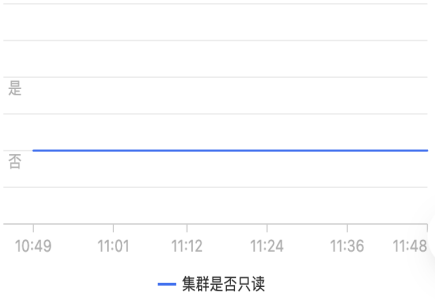
集群是否可见



集群健康状态



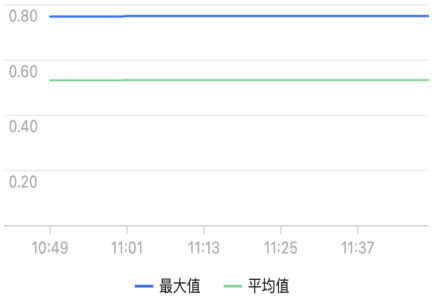
集群是否只读



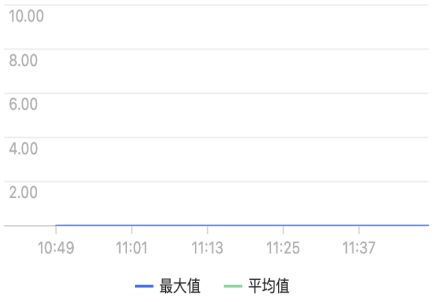
集群性能指标

JVM 使用率 (%)

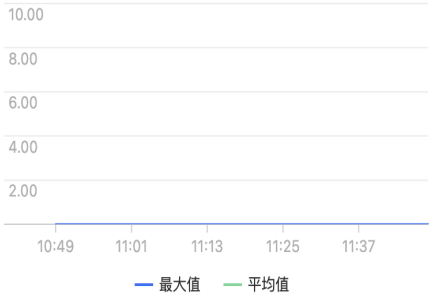
JVM OLD 区使用率 (%)



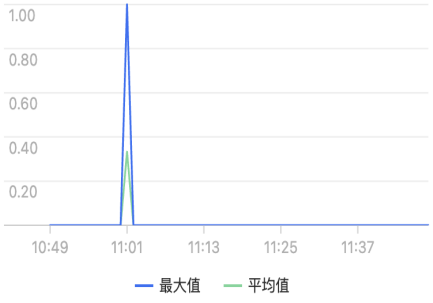
Old GC 次数



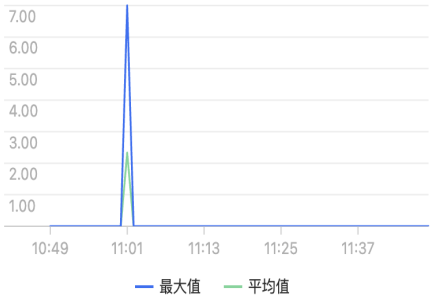
Old GC 耗时 (ms)

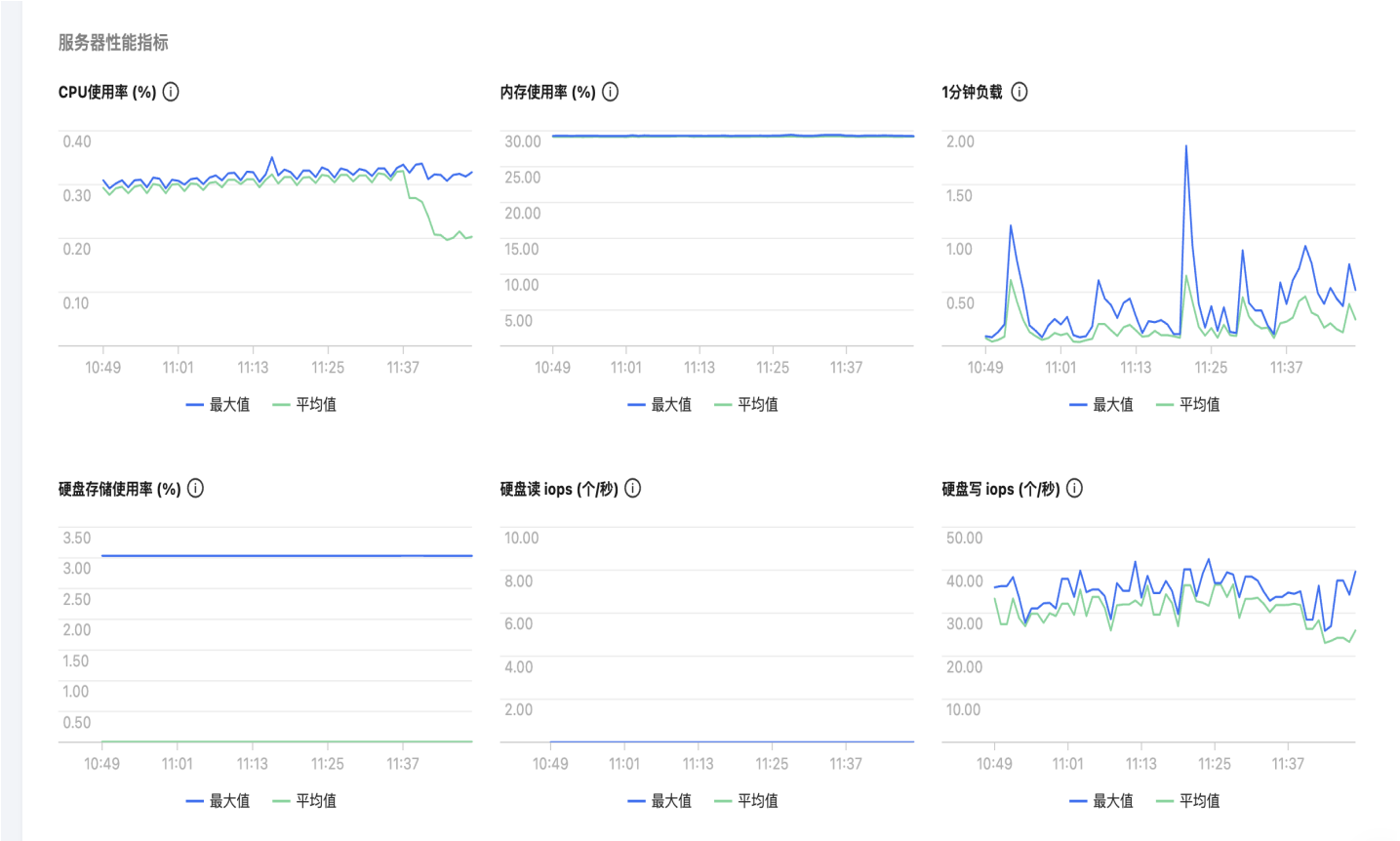


YOUNG GC次数



YOUNG GC耗时 (ms)





节点监控

- **节点列表**

  展示集群各个节点部分实时运行指标。

- **单个节点状态**

  提供每个节点，各项指标详细的历史运行情况。并支持导出数据到本地。

部分指标含义及说明

TSearch 集群一般由多个节点构成，为反映集群整体的运行情况，部分监控指标提供了两类值：平均值、最大值。

- 平均值表示集群所有节点该指标值的平均数。
- 最大值表示集群所有节点该指标的最大值。

所有指标的统计周期均为1分钟，即每1分钟对集群的指标采集1次。您可以通过控制台选择时间范围和不同的时间粒度查看。具体各指标含义说明如下：

监控指标	统计方式	详情
集群健康状态	TSearch 集群健康状态： 0：表示绿色，集群正常 1：表示黄色，告警，部分副本分片不可用 2：表示红色，异常，部分主分片不可用	green：表示所有的主分片和副本分片都可用，集群处于最健康的状态。 yellow：表示所有的主分片均可用，但部分副本分片不可用，此时搜索结果仍然是完整的。但集群的高可用性在一定程度上受到影响，数据面临较高的丢失风险。在集群健康状态变为 yellow 后，应及时调查和定位问题，并修复，防止数据丢失。 red：表示至少一个主分片以及它的全部副本均不可用。集群处于 red 状态意味着已有部分数据丢失：搜索只能返回部分数据，而分配到丢失分片上的写入请求会返回异常。在集群健康状态变为 red 后，应及时定位异常分片，并进行修复。

平均硬盘存储使用率	每单位统计周期内（1分钟），集群各个节点的硬盘存储使用率的平均值	磁盘使用率过高会导致数据无法正常写入。解决方法：及时清理无用的索引。对集群进行扩容，增加单节点的磁盘容量或增加节点个数。
最大硬盘存储使用率	每单位统计周期内（1分钟），集群各个节点中最大硬盘存储使用率	-
平均JVM内存使用率	每单位统计周期内（1分钟），集群各个节点的JVM 内存使用率的平均值	该值过高会导致集群节点 GC 频繁，甚至有出现 OOM。导致该值过高的原因，一般是节点上 TSearch 处理任务超出节点 JVM 的负载能力。您需要注意观察集群正在执行的任务，或调整集群的配置。
最大JVM内存使用率	每单位统计周期内（1分钟），集群各个节点中最大 JVM 内存使用率	-
平均CPU使用率	每单位统计周期内（1分钟），集群各个节点的CPU 使用率的平均值	当集群各节点处理的读写任务超出节点 CPU 的负载能力时，该指标就会过高，CPU 使用率过高会导致集群节点处理能力下降，甚至宕机。您可以从以下几点解决平均 CPU 使用率过高的问题： 观察该指标是持续性较高，还是临时飙升。若是临时飙升，确定是否有临时性复杂任务正在执行。 若该指标持续较高，分析业务对集群的读写操作是否可以优化，降低读写频率，减小数据量，从而减轻节点负载。 对于节点配置无法满足业务吞吐量的情况，建议对集群节点进行纵向扩容，提高单节点的负载能力。
最大CPU使用率	每单位统计周期内（1分钟），集群各个节点中最大 CPU 使用率	-
集群1分钟平均负载	集群1分钟所有节点的平均负载	1分钟平均负载过高时，建议降低集群负载或调大集群节点规格
集群1分钟最大负载	集群1分钟所有节点的最大平均负载	-

平均写入延迟	写入延迟（index_latency），指单次 index 请求耗时（ms/次），集群平均写入延迟是统计周期内（1分钟）所有节点单次 index 请求耗时的平均值。 节点单次 index 请求耗时计算规则：每隔一个统计周期（1分钟）记录一次节点的两个指标，节点历史 index 总次数（_nodes/stats/indices/indexing/index_total），节点历史 index 总耗时（_nodes/stats/indices/indexing/index_time_in_millis），取相邻两次记录的差值，即一个周期内的绝对值并进行计算：index 耗时 / index 次数，得出统计周期内（1分钟）单次 index 平均耗时。	写入延迟，是指单个文档写入平均耗时。集群平均写入延迟，是指统计周期内，所有节点的写入用时的平均值。写入延迟过高时，建议调大节点规格或增加节点个数。
最大写入延迟	写入延迟（index_latency），指单次 index 请求耗时（ms/次），集群最大写入延迟是统计周期内（1分钟）所有节点中单次 index 请求耗时的最大值。 节点单次 index 请求耗时计算规则：见平均写入延迟。	-
平均查询延迟	查询延迟（search_latency），指单次查询请求耗时（ms/次），集群平均查询延迟是统计周期内（1分钟）所有节点单次查询请求耗时的平均值。 节点单次查询请求耗时计算规则：每隔一个统计周期（1分钟）记录一次节点的两个指标，节点历史查询总次数（_nodes/stats/indices/search/query_total），节点历史查询总耗时（_nodes/stats/indices/search/query_time_in_millis），取相邻两次记录的差值，即一个周期内的绝对值并进行计算：query 耗时 / query 次数，得出统计周期内（1分钟）单次查询平均耗时。	查询延迟，是指单个查询平均耗时。集群平均查询延迟，就是统计周期内，所有节点查询用时的平均值。查询延迟过高时，建议调大节点规格或增加节点个数。
最大查询延迟	查询延迟（search_latency），指单次查询请求耗时（ms/次），集群最大查询延迟是统计周期内（1分钟）所有节点中单次查询请求耗时的最大值。 节点单次查询请求耗时计算规则：见平均查询延迟。	-
平均每秒写入	集群所有节点接收到的每秒 index 请求次数的平均值。节点每秒 index 请求次数计算规则：	-

次数	每隔一个统计周期（1分钟）记录一次节点历史 index 总次数（_nodes/stats/indices/indexing/index_total），取相邻两次记录的差值，即一个周期内的绝对值并进行计算：index 次数 / 60 秒，得出统计周期内每秒 index 请求次数的平均值。	
平均每秒完成查询次数	集群所有节点接收到的每秒查询请求次数的平均值。节点每秒查询请求次数计算规则：每个统计周期（1分钟）记录一次节点历史查询总次数（_nodes/stats/indices/search/query_total），取相邻两次记录的差值，即一个周期内的绝对值并进行计算：query 次数 / 60秒，得到统计周期内每秒查询请求次数的平均值。	-
写入拒绝率	单位周期内，集群写入请求被拒绝次数 ÷ 总写入请求次数，得到的比率。具体计算规则：每隔一个统计周期采集两个指标：历史写入请求被拒绝次数（_nodes/stats/thread_pool/write/rejected）、历史写入请求次数（_nodes/stats/thread_pool/write/completed），取相邻两次记录的差值，即一个周期内的绝对值并进行计算，写入请求拒绝次数 / 写入请求完成次数。	写入 QPS 过大，CPU、内存、磁盘使用率过高时，可能会造成集群写入拒绝率会增加。一般地，是集群当前配置无法满足业务写操作需求。对于节点配置过低的场景，可以通过提高节点规格或降低写入操作次数来解决。对于磁盘使用率过高的情况，可以通过扩容集群磁盘或删除无用数据来解决。
查询拒绝率	单位周期内，集群查询请求被拒绝次数 ÷ 总查询请求次数，得到的比率。具体计算规则：每隔一个统计周期采集两个指标：历史查询请求被拒绝次数（_nodes/stats/thread_pool/search/rejected）、历史查询请求次数（_nodes/stats/thread_pool/search/completed），取相邻两次记录的差值，即一个周期内的绝对值并进行计算：查询请求拒绝次数 / 查询请求完成次数。	写入 QPS 过大，CPU、内存使用率过高，可能会造成集群查询拒绝率上升。一般地，是集群当前配置无法满足业务读操作需求，该值过高时建议对集群调大集群节点配置，提高集群节点的处理能力。
分片中最大文档数	文档数目最多的分片的文档个数	-
集群最大分片存储量	存储量最大的分片的存储容量	-

活跃查询 context 数量	集群下全部节点统计活跃查询context的均值	-
P90查询任务耗时	每单位统计周期内，集群最大节点执行查询任务的90分位时延	-
P90写入任务耗时	每单位统计周期内，集群最大节点执行写入任务的90分位时延	-

配置告警

最近更新时间：2025-09-12 14:17:52

操作场景

腾讯云检索分析服务 TSearch 提供一些关键指标的配置告警功能，配置告警可帮助您及时发现集群问题并进行处理。本文为您介绍通过控制台配置告警的操作。目前仅支持集群级别的告警能力，节点告警会在后续支持。

操作步骤

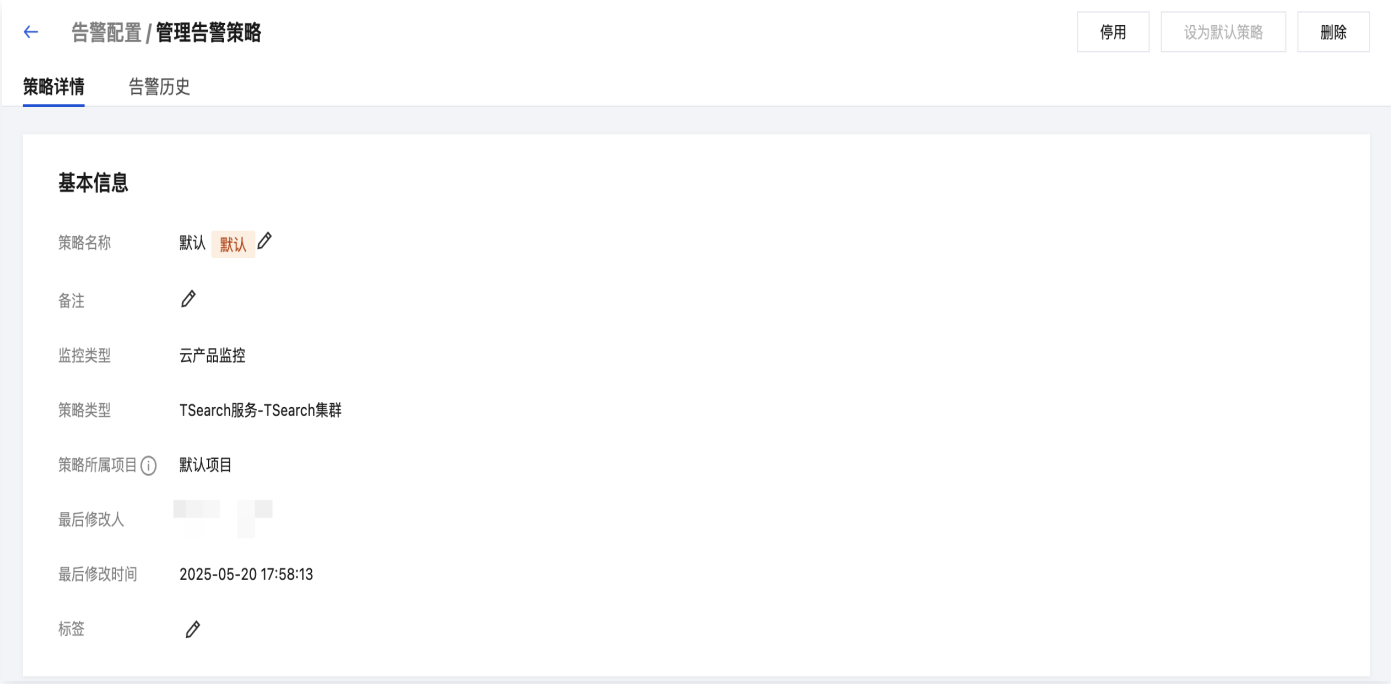
查看集群是否已配置告警

1. 登录 检索分析服务 TSearch 控制台，在集群列表单击集群 ID/名称，进入集群详情页。
2. 在 TSearch 集群详情页面，选择**集群监控 > 告警策略**，可查看集群是否已经配置了告警。

> **警告：**
> 如果您暂未配置告警，强烈建议进行告警策略配置，以便及时获取并处理集群运行的状况及风险，保障服务的稳定。

3. 单击**去查看**。

4. 进入策略详情页面，您可以查看告警策略的基本信息、触发条件、告警对象、告警通知和高级配置，同时也可以修改告警策略。

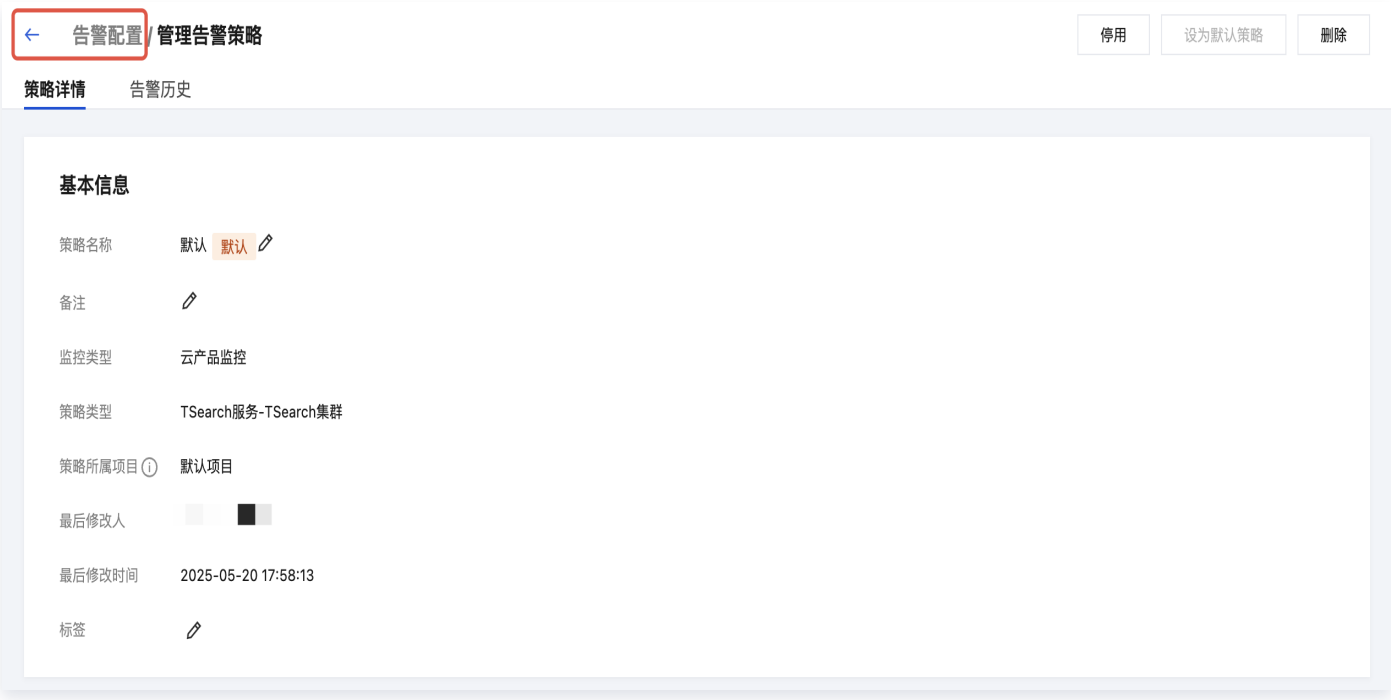


说明：

您也可登录 腾讯云可观测平台控制台 ，在**告警管理 > 监控类型 >云产品监控**，查询某个集群是否已经配置了告警策略。

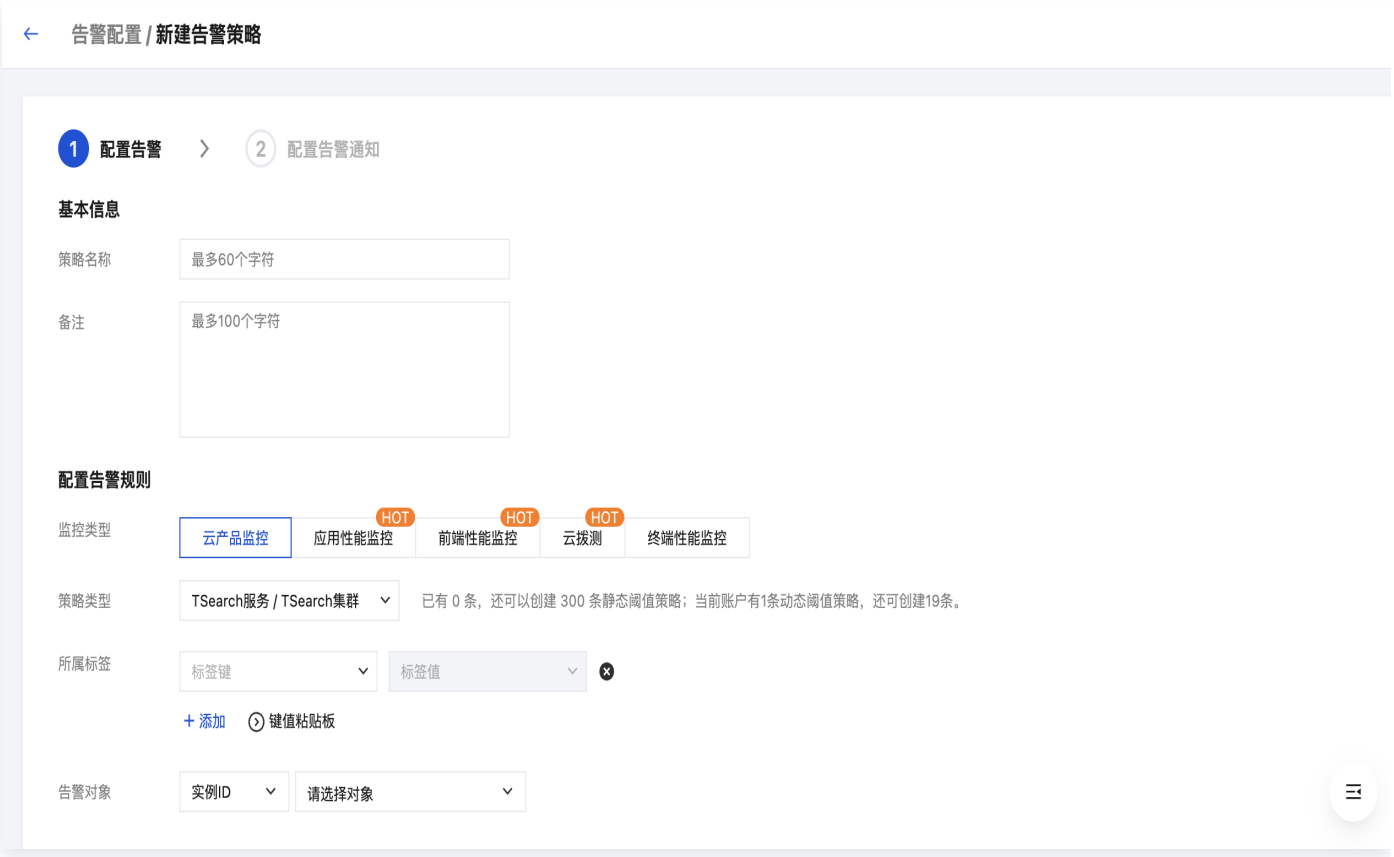
自定义告警配置

1. 在管理告警策略界面，单击**返回上一级**。

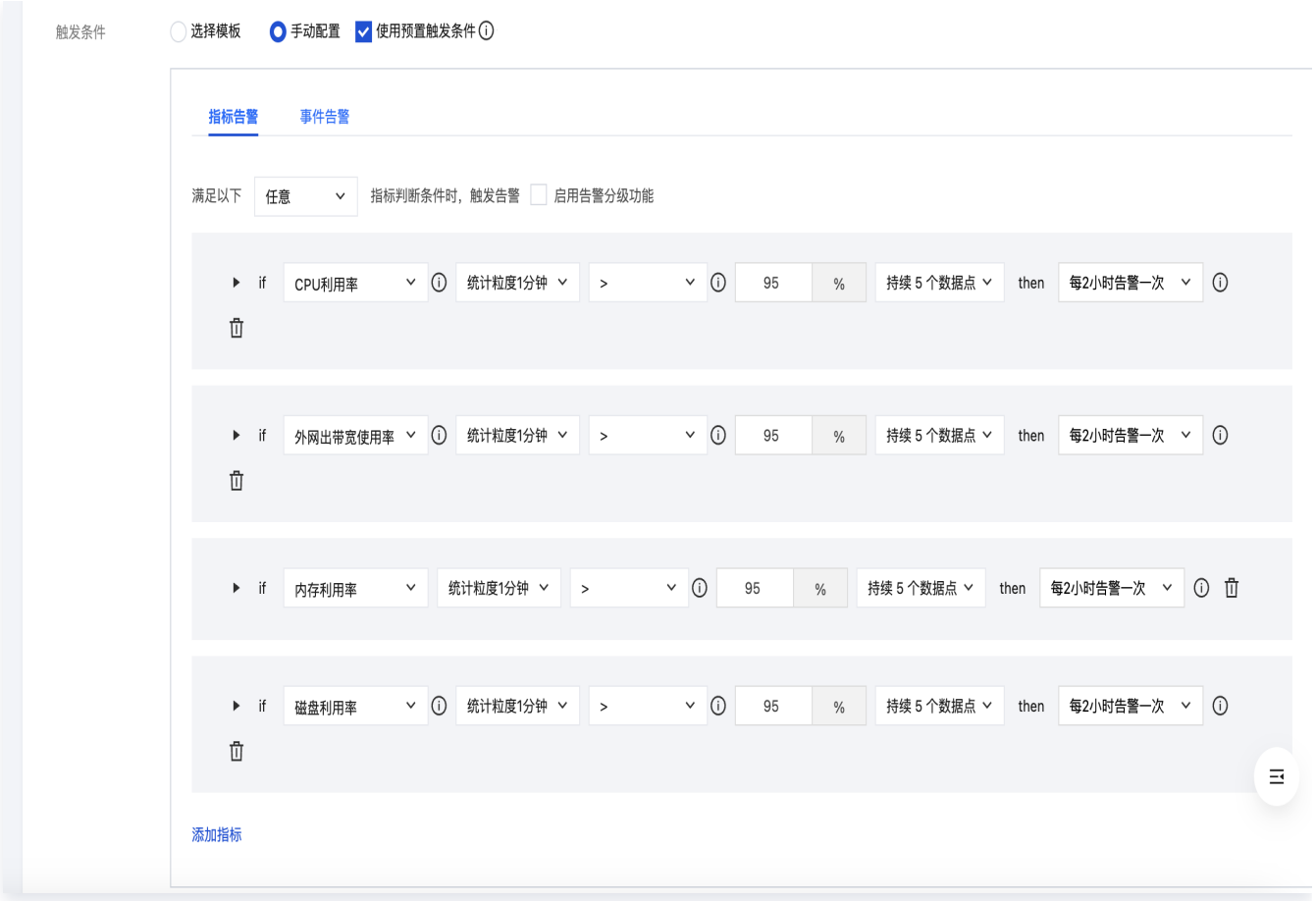


2. 在**告警管理**单击**新建策略**。

3. 在新建策略页面，配置策略参数。
   - **监控类型：选择云产品监控。**
   - **策略类型：选择 TSearch 服务 / TSearch集群。**
   - **告警对象：选择需要配置告警策略的集群。**



   - **触发条件：支持选择模板和手动配置，默认选择自定义配置触发条件。下图以手动配置示例。新建触发条件模板请继续后述指引。**



说明：

- 指标：例如“CPU 使用率”，统计周期为1分钟或5分钟。因 TSearch 集群的各项指标都是1分钟采集1次，所以选择统计周期是1分钟时，当集群出现一次超过阈值就会触发告警，如果选择5分钟，则5分钟内，连续超过阈值才会触发告警。
- 告警频次：例如“每30分钟警告一次”，指每30分钟内，连续多个统计周期指标都超过了阈值，如果有一次告警，30分钟内就不会再次进行告警，直到下一个30分钟，如果指标依然超过阈值，才会再次告警。

- **配置告警通知**：选择接收人、接收组。

4. 配置完成后，单击**完成**，返回到**告警策略**列表，即可查看到刚配置的告警策略。

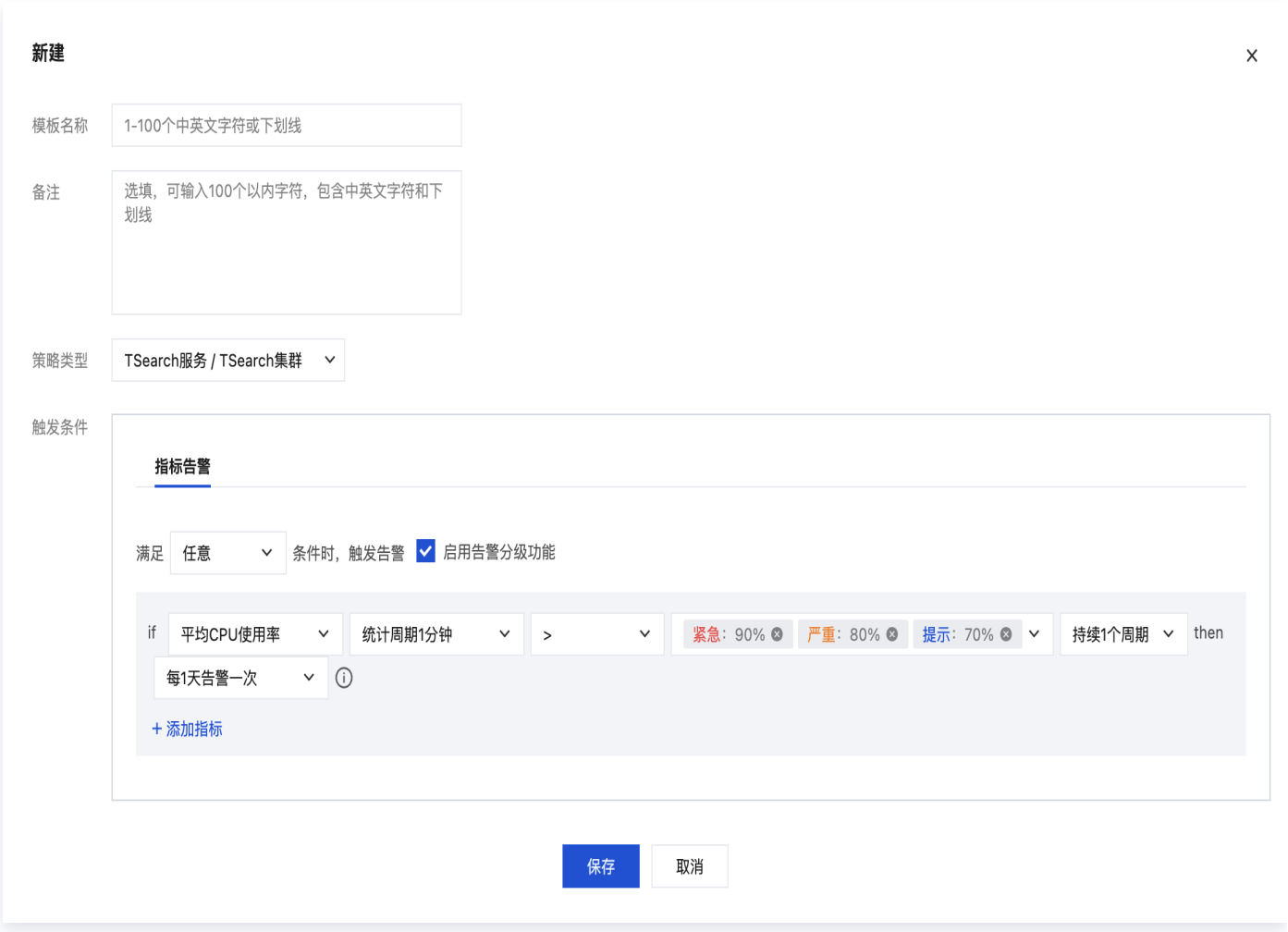
> **说明:**
> 告警策略更详细配置教程可参见 告警策略配置 。

新建触发条件模板

1. 在触发条件页面，单击**新增触发条件模板**。

2. 在触发条件模板页，单击**新建触发条件模板**。

3. 在新建模板页，配置策略类型。
   - **策略类型**：选择 **TSearch 服务 / TSearch集群**。
   - **启用告警分级功能**：勾选此选项，支持紧急、严重和提示三个等级的告警。
4. 确认无误后，单击**保存**。



5. 返回触发条件模板列表页，刷新页面，即可出现刚配置的告警策略模板。

监控告警配置建议

最近更新时间：2025-09-12 14:15:02

腾讯云检索分析服务 TSearch 不仅为运行中的 TSearch 集群提供了多项监控指标，用于监测集群的运行情况，还提供了一些关键指标的配置告警功能，帮助您及时发现集群问题并进行处理。具体使用方法可参考 查看监控 和 配置告警 。

本文为您介绍在使用 TS 集群过程中需要重点关注的一些指标及其告警建议配置：

指标	告警建议配置	详细说明
集群健康状态	统计周期1分钟，≥1，持续5个周期，每30分钟告警一次	集群健康状态取值为： • 0：绿色，表示集群所有主分片和副本分片都可用，集群处于最健康的状态。 • 1：黄色，表示所有的主分片均可用，但存在不可用副本分片。此时，搜索结果仍然是完整的，但集群的高可用性在一定程度上受到影响，数据面临较高的丢失风险。 • 2：红色，表示至少一个主分片以及它的全部副本分片均不可用。集群处于红色状态意味着已有部分数据不可用，搜索只能返回部分数据，而分配到丢失分片上的请求会返回异常。 • 集群健康状态是集群当前运行情况的最直接体现，当集群处于黄色或红色状态时，应立即排查产生原因，并及时修复，防止数据丢失和服务不可用。
平均磁盘使用率	统计周期1分钟，>80%，持续5个周期，每30分钟告警一次	平均磁盘使用率表示集群各节点磁盘使用率的平均值。磁盘使用率过高会导致节点没有足够的磁盘空间容纳分配到该节点上的分片，从而导致创建索引，添加文档等基本操作执行失败。建议在平均磁盘使用率超过75%时及时清理数据或扩容集群。
平均JVM内存使用率	统计周期1分钟，>85%，持续5个周期，每30分钟告警一次	平均 JVM 堆内内存使用率表示集群各节点 JVM 内存使用率的平均值。JVM 内存使用率过高会导致读写操作被拒绝，集群 GC 频繁，甚至出现 OOM 等问题。当发现 JVM 内存使用率超过阈值时，建议通过纵向扩容的方式提高集群节点的规格。
平均CPU使用率	统计周期1分钟，>90%，持续5个周期，每30分钟告警一次	平均 CPU 使用率表示集群各节点 CPU 使用率的平均值。该值过高会导致集群节点处理能力下降，甚至宕机。发现 CPU 过高时，应根据集群当前节点配置情况和业务情况，提高节点规格或降低业务请求量。
bulk拒绝率	统计周期1分钟，>0%，持续1个周期， 每30分钟告警一次	bulk 拒绝率表示单周期内集群执行 bulk 操作被拒绝次数占 bulk 总操作次数的百分比。当 bulk 拒绝率大于0%，即出现 bulk 拒绝时，说明集群已经达到了 bulk 操作处理能力的上限，或集群出现异常，应及时

		排除出现 bulk 拒绝的原因并及时解决，否则会影响业务的 bulk 操作，甚至出现数据丢失情况。
查询拒绝率	统计周期1分钟，>0%，持续1个周期，每30分钟告警一次	查询拒绝率表示单周期内集群执行查询操作被拒绝次数占查询总操作数的百分比。当查询拒绝率大于0%，即出现查询拒绝时，说明集群已经达到了查询操作处理能力的上限，或集群出现异常，应及时排查出现查询拒绝的原因并及时解决，否则会影响业务的查询操作。

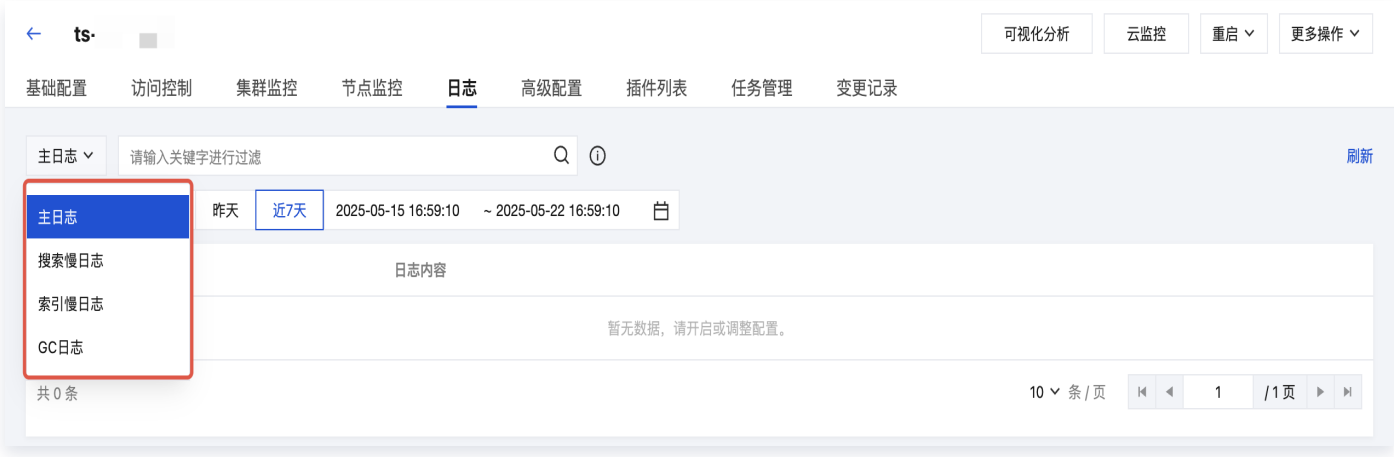
查询集群日志

最近更新时间：2025-09-12 14:15:02

本文为您介绍腾讯云检索分析服务 TSearch 集群运行日志的使用说明。用户可以通过集群的运行日志，了解集群的运行状况，定位问题，辅助集群的应用开发和运维。

查询集群日志

1. 登录 检索分析服务 TSearch 控制台，单击集群 ID/名称，进入集群详情页。
2. 选择**日志**页签，可查看集群的运行日志。
   - TSearch 共有四种类型的日志：主日志、搜索慢日志、索引慢日志和 GC 日志，日志内容包括日志时间、日志级别，以及具体信息等。
   - TSearch 默认提供集群7天内的运行日志，按时间倒序展示，用户可以按时间和关键字进行查询。此外，用户也可调用 TSearch API 调整日志相关配置，例如对于慢日志，设定查询或索引数据时，认为响应较慢的时间阈值。
3. 在日志页面的搜索框，可以按照时间范围和关键字查询相关日志，关键字查询语法同 lucene 查询语法一致。
   - 输入关键词查询，例如：“YELLOW”。
   - 指定字段设置关键词，例如：`message:YELLOW`。
   - 多个条件组合：`level:INFO and ip:10.0.1.1`，可以查询相关日志。



日志说明

主日志

展示集群运行产生日志的时间、级别、信息等，有 INFO、WARN、DEBUG 等不同级别。

```
2019-2-14 08:00:00 10.0****
INFO
[o.e.c.r.a.AllocationService] [1550199698000783811] Cluster health
status changed from [YELLOW] to [GREEN] (reason: [shards started
```

```
[[filebeat-2019.02.19][2]] ...]).

2019-2-14 02:30:02 10.0****
DEBUG
[o.e.a.a.i.d.TransportDeleteIndexAction] [1550199698000783811] failed to
delete indices [[[filebeat-2019.02.09/7VZM6Fa-Twmj8pVaAyyrxg]]]
org.elasticsearch.index.IndexNotFoundException: no such index
    at
org.elasticsearch.cluster.metadata.MetaData.getIndexSafe(MetaData.java:4
75) ~[elasticsearch-5.6.4.jar:5.6.4]
    at
org.elasticsearch.cluster.metadata.MetaDataDeleteIndexService.lambda$del
eteIndices$0(MetaDataDeleteIndexService.jav
```

慢日志

慢日志的目的是捕获超过指定时间阈值的查询和索引请求，用于跟踪分析由用户产生的很慢的请求。

```
2018-10-28 12:04:17
WARN
[index.indexing.slowlog.index] [1540298502000001009] [pmc/wCALr6BfRm-
sr3qOQuGX
Xw] took[18.6ms], took_millis[18], type[articles], id[AWa41-
J9c0s1mOPvR6F3], routing[] , source[]
```

开启和调整慢日志

默认情况，慢日志不开启。开启慢日志需要定义具体动作（query、fetch 或 index）、期望的事件记录等级（INFO、WARN、DEBUG 等）、以及时间阈值。用户可以根据业务场景，开启和调整相关配置。
如需开启慢日志，可在集群列表右侧点击可视化分析或集群详情页的右上角单击可视化分析进入 Kibana 页面，通过 Dev Tools 调用 TSearch 相关 API，或通过客户端调用配置修改 API。
配置所有索引：

```
PUT */_settings
{
"index.indexing.slowlog.threshold.index.debug" : "5ms",
"index.indexing.slowlog.threshold.index.info" : "50ms",
"index.indexing.slowlog.threshold.index.warn" : "100ms",
"index.search.slowlog.threshold.fetch.debug" : "10ms",
"index.search.slowlog.threshold.fetch.info" : "50ms",
"index.search.slowlog.threshold.fetch.warn" : "100ms",
"index.search.slowlog.threshold.query.debug" : "100ms",
"index.search.slowlog.threshold.query.info" : "200ms",
```

```
"index.search.slowlog.threshold.query.warn" : "1s"
}
```

配置单个索引：

```
PUT /my_index/_settings
{
"index.indexing.slowlog.threshold.index.debug" : "5ms",
"index.indexing.slowlog.threshold.index.info" : "50ms",
"index.indexing.slowlog.threshold.index.warn" : "100ms",
"index.search.slowlog.threshold.fetch.debug" : "10ms",
"index.search.slowlog.threshold.fetch.info" : "50ms",
"index.search.slowlog.threshold.fetch.warn" : "100ms",
"index.search.slowlog.threshold.query.debug" : "100ms",
"index.search.slowlog.threshold.query.info" : "200ms",
"index.search.slowlog.threshold.query.warn" : "1s"
}
```

GC 日志

TSearch 默认开启 GC 日志，以下两条具体的 GC 日志，分别展示日志的时间、节点 IP、日志级别等。

```
2019-2-14 20:48:22
10.0.***
INFO
[o.e.m.j.JvmGcMonitorService] [1550199698000783711] [gc][380573]
overhead, spent [307ms] collecting in the last [1s]
2019-2-14 10:04:09
10.0.***
WARN
[o.e.m.j.JvmGcMonitorService] [1550199698000784111] [gc][341943]
overhead, spent [561ms] collecting in the last [1s]
```