

容器服务

TKE Insight

产品文档



腾讯云

【版权声明】

©2013-2024 腾讯云版权所有

本文档著作权归腾讯云单独所有，未经腾讯云事先书面许可，任何主体不得以任何形式复制、修改、抄袭、传播全部或部分本文档内容。

【商标声明】

及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。

【服务声明】

本文档意在向客户介绍腾讯云全部或部分产品、服务的当时的整体概况，部分产品、服务的内容可能有所调整。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。

文档目录

TKE Insight

成本洞察

Node Map

Workload Map

TKE Insight

成本洞察

最近更新时间：2024-05-24 16:01:11

功能介绍

传统的 Kubernetes 集群通常只关注资源层面的数据，例如 Node 中有节点的 Allocatable 标示可分配的资源；Metrics-Server 可以监控节点和 Pod 的利用率数据。TKE Insight 推出成本可视化面板可以帮助用户更好地理解、监控和控制容器环境中的资源使用和成本情况。主要包括如下几个目标：

透明化：用户可以直观地了解应用和服务在容器环境中的资源使用情况，包括不同的集群、命名空间以及工作负载在 CPU、内存方面的成本消耗。这有助于用户了解整体成本结构，提高决策效率。

成本控制与优化：通过成本洞察，用户可以发现成本较高的应用或服务。用户可以针对这些问题进行优化，例如 Request 智能推荐，可抢占式 Job，原生节点专用调度器，QoS Agent，以减少不必要的开支。

预算管理：用户可以根据实际成本情况制定更加合理的预算，并监控预算执行。当资源使用接近或超过预算时，用户可以及时调整策略，避免产生额外支出。

操作步骤

1. 登录 [容器服务控制台](#)。
2. 选择左侧导航中的 **TKE Insight > 成本洞察**。

使用说明

成本洞察的成本数值是一个计算出的参考值，仅供参考，准确费用请登录腾讯云 [费用中心](#) 查看。

超级节点无法区分按量计费具体是普通按量计费还是竞价 Pod。因此目前不支持使用预留券抵消对应规格 Pod 的成本。原因是预留券需要在计算成本时调用接口查询预留券信息，放在监控组件中去调用会对接口造成较大压力。

如果上个月没有成本数据，则当月成本相较于上月成本的百分比将不会展示数据。

在计算成本时，Pod 的 GPU 使用量不进行计算，而是把 GPU 的价格分摊在 CPU 和内存上。即 GPU 节点上的 CPU、内存单价更高。

对于原生节点和普通节点的刊例价指标，在上报时，由于无法获取节点的计费模式，因此统一按照按量计费模式的单价去上报数据。

对于按量计费的超级节点上的 Pod，当 annotation 中指定了多个 GPU 类型时，使用第一个 GPU 类型去计算价格。

价格说明

计费方式

计费项包含了普通节点（CVM）、原生节点和超级节点。

普通节点计费方式

根据节点机型和计费方式（按量计费），找出节点的刊例价。

说明：

机型从 node 的 annotation `node.kubernetes.io/instance-type` 中获取。

原生节点计费方式

根据节点机型和计费方式（按量计费），找出节点的刊例价。

说明：

机型从 node 的 annotation `node.kubernetes.io/instance-type` 中获取。

超级节点计费方式

按量计费

按量计费超级节点刊例价 = 对应超级节点上各个 Pod 所在子机刊例价总和。

Pod 所在子机刊例价计算方式如下：

解析 Pod 所在子机的 CPU 类型和 GPU 类型。如果 CPU 类型指定了多个，则以第一个类型为依据进行价格计算。

如果 GPU 类型指定了多个，则以第一个类型为依据进行价格计算。

如果 GPU 类型不为空，则查找对应 GPU 类型的 CPU 单价、内存单价、GPU 单价和 license 单价（如果 license 不收费，则单价为 0 元）。如果 GPU 类型为空，则查找对应 CPU 类型的 CPU 单价和内存单价。

解析 Pod 所在子机的规格（即 CPU 规格、内存规格、GPU 规格），详情请参见 [资源规格计算方法](#) 和 [资源规格](#)。

计算 Pod 的刊例价： $\text{GPU 规格} \times \text{GPU 单价} + \text{CPU 规格} \times \text{CPU 单价} + \text{内存规格} \times \text{内存单价} + \text{license 单价}$

说明：

CPU 类型从 Pod 的 annotation `eks.tke.cloud.tencent.com/cpu-type` 中获取。

GPU 类型从 Pod 的 annotation `eks.tke.cloud.tencent.com/gpu-type` 中获取。

节点刊例价上报到腾讯云可观测平台

1. 针对每一个节点，会将节点的刊例价信息上报到 [腾讯云可观测平台 TCOP](#)。当前支持 5min 和 1h 粒度，5min 粒度表示节点 5min 的价格，1h 粒度表示节点 1h 的价格。
2. 针对按量计费超级节点上的 Pod，会上报各个 Pod 的刊例价格信息。当前支持 5min 和 1h 粒度，5min 粒度表示 Pod 所在子机 5min 的价格，1h 粒度表示 Pod 所在子机 1h 的价格。

1. 当月成本计算方式

从腾讯云可观测平台 TCOP 拉取最近一个月普通节点、原生节点、超级节点的价格数据，进行相加，以得到一个总的价格。

2. 预估下月成本计算方式

从腾讯云可观测平台 TCOP 以小时粒度拉取最近1h普通节点、原生节点、超级节点价格信息。假设最近1h普通节点、原生节点和超级节点价格之和为x元，则预估的下月成本为 $x \times 24 \times 30$ 元。

3. 成本走势

今天成本：从监控中拉取最近一天普通节点、原生节点和超级节点的价格数据，进行相加。

当月成本走势：从云监控以小时粒度拉取最近一个月的价格数据，前端以小时粒度展示出最近一个月的成本趋势。

说明：

每个小时的数据点表示对应时间点普通节点、原生节点和超级节点刊例价总和。

4. 资源成本分布

分别拉取普通节点、原生节点、超级节点最近1h的成本，然后乘以24和30，得出各个节点的最近一个月的普通节点成本、原生节点成本和超级节点成本。

5. 命名空间成本分布

通过最近一小时的费用预估集群中每个命名空间负载一个月的成本。

1. 获取当前集群的所有节点（普通节点、原生节点和超级节点）。
2. 从腾讯云可观测平台 TCOP 拉取普通节点、原生节点上各个 Pod 的 CPU 使用量和内存使用量。
3. 分别拉取到当前集群中 CVM 节点和原生节点的机型信息，并计算出各个机型的 CPU 单价和内存单价（超级节点 CPU 单价和内存单价是固定的，所以不用计算）。计算方式如下：

3.1 根据全网所有机型的价格预估，得出 CPU 单价为 x 元/核/小时，内存单价为 y 元/G/小时。

3.2 根据3.1得出 CPU 单价/内存单价的比例，然后将整个机器的价格按照内存进行计费，得出该机器内存单价，再将内存单价乘以比例得出 CPU 单价。例如：CPU 单价/内存单价=r，机器为 x 核 yGB，机器价格为 z，则 CPU/内存单价计算方式：

得出该机器内存单价： $z/(x*r+y)$

CPU 单价： $r*(z/(x*r+y))$

4. 根据2中获得的各个 Pod 的 CPU 使用量和内存使用量，再根据 Pod 所在节点的 CPU 单价和内存单价，得出各个 Pod 的价格。

5. 根据2中获得的按量计费超级节点，从腾讯云可观测平台 TCOP 获取按量计费超级节点上 Pod 的刊例价。

6. 根据4和5中计算出来的 Pod 成本信息，聚合出各个命名空间的成本信息。

7. 空闲成本：用当前所有节点的总成本 - 各个命名空间的成本之和。

6. 工作负载成本分布

通过最近一小时的费用预估集群中每个工作负载一个月的成本。

1. 获取当前集群的所有节点（普通节点、原生节点和超级节点）。

2. 从监控拉取普通节点、原生节点上各个 Pod 的 CPU 使用量和内存使用量。
3. 分别拉取到当前集群中普通节点和原生节点的机型信息，并计算出各个机型的 CPU 单价和内存单价（超级节点 CPU 单价和内存单价是固定的，所以不用计算，CPU 单价：36元/核/小时，内存单价：18元/G/小时）。计算方式如下：
 - 3.1 根据成本大师的预估（根据全网所有机型的价格），得出 CPU 单价为 x 元/核/小时，内存单价为 y 元/G/小时。
 - 3.2 根据3.1得出 CPU 单价/内存单价的比例，然后将整个机器的价格按照内存进行计费，得出该机器内存单价，再将内存单价乘以比例得出 CPU 单价。例如：CPU 单价/内存单价= r ，机器为 x 核 y GB，机器价格为 z ，则 CPU/内存单价计算方式：
得出该机器内存单价： $z/(x*r+y)$
CPU 单价： $r*(z/(x*r+y))$
4. 根据2中获得的各个 Pod 的 CPU 使用量和内存使用量，再根据 Pod 所在节点的 CPU 单价和内存单价，得出各个 Pod 的价格。
5. 根据2中获得的按量计费超级节点，从腾讯云可观测平台 TCOP 获取按量计费超级节点上 Pod 的刊例价。
6. 根据4和5中计算出来的 Pod 成本信息，聚合出各个 Workload 的成本信息。
7. 空闲成本：用当前所有节点的总成本 - 各个 Workload 的成本之和。

Node Map

最近更新时间：2024-05-24 15:28:58

概述

传统的对象管理页面一般都是通过列表形式展开的，例如 TKE 集群里现有的资源对象列表，但列表的展示方式具备如下缺点：每一行都是一些抽象的数字，可读性不高；对象之间没有明确的标识区分；无法支持某种形态的排序。为此，腾讯云容器服务的 **TKE Insight** 的能力，通过可视化的形式，将用户所有的资源对象进行展示。该模块具备丰富的过滤查询、类型聚合、状态展示能力，可以帮助用户快速定位目标对象。

Node Map 主要通过可视化的页面展示节点的各项状态和指标，帮助用户了解当前集群节点的资源利用率、装箱率情况，辅助用户分析节点可能存在的问题。

注意事项

Node Map 不支持统计超级节点的各项指标。

使用方式

1. 登录 [容器服务控制台](#)。
2. 在左侧选择 **TKE Insight > Node Map**。
3. 在 Node Map 页面中，选择**地域**、**集群类型**和**集群**后，进行查看。
4. 右上角可选按照24小时、7天和30天数据，均值或峰值数据筛选，如下图所示：



Node Map 功能说明

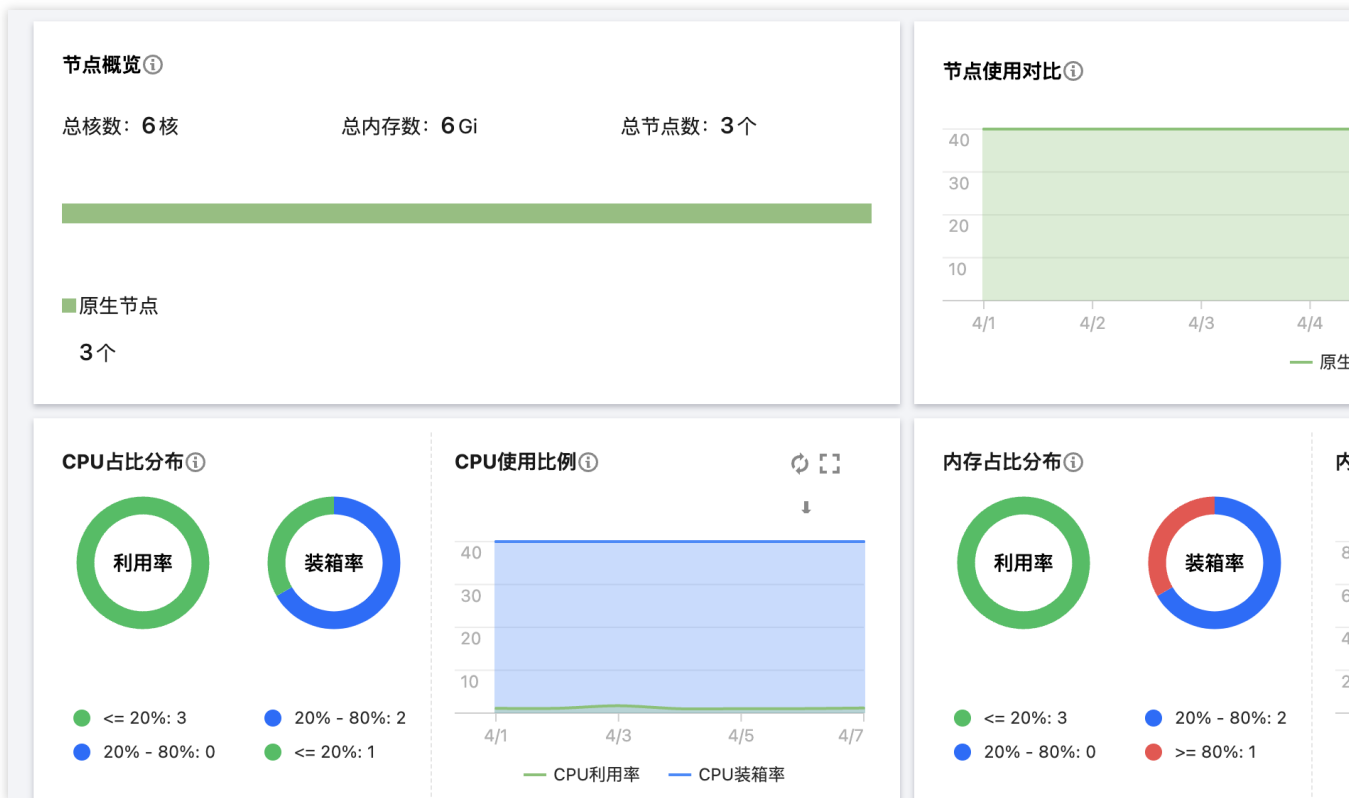
Node Map 分为概览和资源对象热图，如下图所示：



1. 上半区域是集群里的节点概览。
2. 下半区域是集群里的每一个节点资源对象。

集群/节点概览

集群/节点概览如下图所示：



节点概览

展示当前集群节点的总核数和总内存数。

展示当前集群中不同节点类型节点（原生节点、普通节点、注册节点）数量。

注意：

超级节点没有节点的概念，因此不在统计范围之内。

节点使用对比

对比不同类型节点（原生节点、普通节点、注册节点）CPU 装箱率、内存装箱率、CPU 利用率和内存利用率的情况。

节点使用资源分布

展示了集群中节点（原生节点、普通节点、注册节点）CPU /内存使用情况的分布。

注意：

超级节点没有节点的概念，因此不在统计范围之内。

指标	描述	说明
CPU 利用率分布	根据节点近24小时、7天或30天（右上角周期时间段选择）利用率均值、峰值查看集群中节点（原生节点、普通节点、注册节点）CPU 利用率分布情况。展示利用率在某一区间的节点个数。	CPU 利用率定义为节点真实 CPU 资源使用量除以节点规格。超级节点没有节点的概念，因此不在统计范围之内。

CPU 装箱率分布	根据节点近24小时、7天或30天（右上角周期时间段选择）装箱率均值、峰值查看集群中节点 CPU 装箱率分布情况。展示装箱率在某一区间的节点个数。	CPU 装箱率定义为节点上所有 Pod 的 CPU Request 之和除以节点规格。
内存利用率分布	根据节点近24小时、7天或30天（右上角周期时间段选择）利用率均值、峰值查看集群中节点内存利用率分布情况。展示利用率在某一区间的节点个数。	内存利用率定义为节点真实内存资源使用量除以节点规格。
内存装箱率分布	根据节点近24小时、7天或30天（右上角周期时间段选择）装箱率均值、峰值查看集群中节点内存装箱率分布情况。展示装箱率在某一区间的节点个数。	内存装箱率定义为节点上所有 Pod 的内存 Request 之和除以节点规格。
CPU 使用比例	根据页面右上角的指定周期时间段查看集群整体 CPU 装箱率和利用率的走势情况。	利用率统计了集群中原生节点、普通节点、注册节点、超级节点的利用率情况。 装箱率统计了集群中原生节点、普通节点、注册节点的装箱率情况，超级节点没有节点的概念，因此不在统计范围之内。
内存使用比例	根据页面右上角的指定周期时间段查看集群整体内存装箱率和利用率的走势情况。	利用率统计了集群中原生节点、普通节点、注册节点、超级节点的利用率情况。 装箱率统计了集群中原生节点、普通节点、注册节点的装箱率情况，超级节点没有节点的概念，因此不在统计范围之内。

节点对象查看

节点对象如下图所示，您可以通过节点筛选、分组、状态进行节点的过滤和聚合展示。



筛选：根据相关指标筛选得到目标节点，若您不选择，则是默认选择了所有的节点。支持过滤多个属性，取交集。

聚合：对筛选出来的节点进行分组，每一组节点在同一个浅色框内，一个框内的每个节点具有相同的属性值。

状态：您可以通过**节点状态**进行筛选，支持选择 **CPU 使用率**、**内存使用率**、**CPU 装箱率**和**内存装箱率**。

CPU 使用率：节点近24小时、7天或30天 CPU 利用率均值、峰值。

内存使用率：节点近24小时、7天或30天内存利用率均值、峰值。

CPU 装箱率：节点近24小时、7天或30天 CPU 装箱率均值、峰值。

内存装箱率：节点近24小时、7天或30天内存装箱率均值、峰值。

注意：

在节点对象热图中的节点，默认按照当前选择的“节点状态”属性值进行**升序排列**。例如，您选择的节点状态是“CPU使用率”，则节点对象热图里的节点默认是按照节点近期（右上角周期时间段选择）CPU 使用率均值或峰值升序排列。

筛选后的节点由**绿、蓝、红**三色区分。您可以单击页面右下角的



，调整三种状态的阈值范围。如下图所示：



节点操作

当您的鼠标悬浮在节点对象上后，您可以查看节点详情，并进行如下操作：



1. 单击**详情**，查看当前节点的详情信息。如下图所示：



节点详情

Pod 详情

功能开关

Request 推荐：集群级别的功能开关。开启后，可以根据 Workload 的使用情况，给 Workload 推荐合适的 Request。详情请参见 [Request 推荐](#)。

原生节点专用调度器：集群级别的功能开关。开启后，可以虚拟放大原生节点规格，调度更多的 Pod。详情请参见 [原生节点专用调度器](#)。

节点信息

节点规格：节点的 CPU 和内存量。

装箱率：节点上所有 Pod 的 Request 之和除以节点规格（依据页面右上角的页面选择）。

24h 利用率均值：节点 CPU 和内存24小时均值利用率。

24h 利用率峰值：节点 CPU 和内存24小时峰值利用率。

节点走势图信息

CPU 使用比例：

根据右上角的指定周期时间段查看集群中节点 CPU 使用比例走势情况。

内存使用比例：

根据右上角的指定周期时间段查看集群中节点内存使用比例走势情况。

CPU 用量走势图

申请量：Pod 的 Request 数值。

实际用量：Pod 实际的用量。

推荐量：Pod 推荐的用量。需要提前开启 [Request 推荐](#)。

Pod 信息

Pod 名称：Pod 的名字。

Pod 状态：Pod 的状态。

所属 Workload：当前 Pod 所属的 Workload。

Request：黑字部分是当前 Pod 的 Request 配置，蓝字部分是当前 Pod 通过 Request 推荐，推荐的建议数值（需要打开 [Request 推荐组件](#)）。

注意：

1. 单击蓝字部分推荐的 Request，您可以一键更新该 Pod 所属的 Deployment 的 Request 数值。
2. 一键更新 Request 能力属于原生节点专有功能，如果 Workload 有部分 Pod 不在原生节点上，需要先将 Pod 迁移到原生节点。详情请参见 [Request 推荐](#)。
2. 单击**驱逐**，将节点内的 Pod 从节点中驱逐到集群内其他节点。
3. 单击**移出**，将节点移出节点池。

若节点在节点池：当前节点归属节点池管理，需要前往节点池页面处理。

若节点不在节点池：可以将节点从集群中删除。若节点是按量计费的节点，支持同时销毁节点。

4. 单击**取消封锁**，将节点调度进新的 Pod。

说明：

节点在驱逐之后，会自动封锁。单击**取消封锁**后，该节点可以继续调度进新的 Pod。

Workload Map

最近更新时间：2023-05-06 17:36:46

概述

传统的对象管理页面一般都是通过列表形式展开的，例如 TKE 集群里现有的资源对象列表，但列表的展示方式具备如下缺点：每一行都是一些抽象的数字，可读性不高；对象之间没有明确的标识区分；无法支持某种形态的排序。为此，TKE 推出了云原生资产管理的能力，通过可视化的形式，将用户所有的资源对象进行展示，该平台具备丰富的过滤查询、类型聚合、状态展示能力，可以帮助用户快速定位目标对象。

Workload Map 主要通过可视化的页面展示工作负载的各项状态和指标，帮助用户了解当前工作负载的配置量和实际使用情况，辅助用户分析工作负载可能存在的问题。

注意事项

Workload Map 不支持提供超级节点上的 Pod 推荐值的监控数据。在聚合监控指标时，如果有 Workload 的 Pod 在超级节点上，则在统计 Workload 和 Cluster 级别的推荐值时，会缺少部分数据。

使用方式

1. 登录 [容器服务控制台](#)。
2. 在左侧选择云原生资产管理 > Workload Map。
3. 在“Workload Map”页面中，选择地域、集群类型和集群后，进行查看。

Workload Map 功能说明

Workload Map 分为概览和资源对象热图

上半区域是集群里的 Workload 概览。

1. 下半区域是集群里的每一个 Workload 资源对象。

工作负载概览

工作负载概览：展示当前集群中是有推荐 Request 的工作负载的数量和所占比例；展示当前集群节点的总核数和总内存数。

注意

首先需要打开 Request 推荐功能，才会推荐工作负载调整 Request 数值。

CPU 用量：集群中工作负载的 CPU 用量情况。

申请量：集群中所有工作负载中所有 Pod 的 Request 的 CPU 之和。

实际用量：集群中所有工作负载中所有 Pod 实际的 CPU 用量之和。

推荐量：集群中所有工作负载中所有 Pod 推荐的 CPU 用量之和。需要提前开启 Request 推荐。

内存用量：集群中工作负载的内存用量情况。

申请量：集群中所有工作负载中所有 Pod 的 Request 的内存之和。

实际用量：集群中所有工作负载中所有 Pod 实际的内存用量之和。

推荐量：集群中所有工作负载中所有 Pod 推荐的内存用量之和。需要提前开启 Request 推荐。

工作负载对象查看

工作负载对象，您可以通过工作负载的筛选、聚合、状态进行 Workload 的过滤和聚合展示。

筛选：根据相关指标筛选得到目标 Workload，若您不选择，则是默认选择了所有的 Workload。支持过滤多个属性，取交集。

聚合：对筛选出来的 Workload 进行分组，每一组节点在同一个浅色框内，一个框内的每个 Workload 具有相同的属性值。

状态：您可以通过**状态**进行筛选，支持选择**CPU 利用率峰值**、**CPU 利用率均值**和 **CPU 装箱率**。

CPU 利用率峰值：Workload CPU 峰值利用率，分24小时，7天，30天维度。

CPU 利用率均值：节点 CPU 均值利用率，分24小时，7天，30天维度。

CPU 装箱率：节点 CPU 均值装箱率，分24小时，7天，30天维度。

注意

在工作负载对象热图里的节点，默认按照当前选择的“状态”属性值进行**升序排列**。例如，您选择的工作负载状态是“24h CPU 利用率均值”，则工作负载对象热图里的工作负载默认是按照工作负载24h CPU 利用率均值升序排列。

筛选后的 Workload 由**绿、蓝、红**三色区分。您可以单击页面右下角的



，调整三种状态的阈值范围。

下载：您可以单击工作负载概览右侧的下载标识



，下载本页 Workload 资源列表信息：

下载列表的字段说明：

名称：工作负载的名称

命名空间：工作负载所属的命名空间

工作负载类型：工作负载所属的 K8s 类型

24h CPU 利用率均值：工作负载近24小时 CPU 利用率的均值。单位：%

7d CPU 利用率均值：工作负载近7天 CPU 利用率的均值。单位：%

30d CPU 利用率均值：工作负载近30天 CPU 利用率的均值。单位：%

24h CPU 利用率峰值：工作负载近24小时 CPU 利用率的峰值。单位：%

7d CPU 利用率峰值：工作负载近7天 CPU 利用率的峰值。单位：%

30d CPU 利用率峰值：工作负载近30天 CPU 利用率的峰值。单位：%

24h 内存利用率均值：工作负载近24小时内内存利用率的均值。单位：%

7d 内存利用率均值：工作负载近7天内内存利用率的均值。单位：%

30d 内存利用率均值：工作负载近30天内内存利用率的均值。单位：%

24h 内存利用率峰值：工作负载近24小时内内存利用率的峰值。单位：%

7d 内存利用率峰值：工作负载近7天内内存利用率的峰值。单位：%

30d 内存利用率峰值：工作负载近30天内内存利用率的峰值。单位：%

工作负载操作

当您的鼠标悬浮在工作负载对象上后，您可以查看当前工作负载的详情信息，并进行如下操作：

1. 单击**详情**，查看当前工作负载的详情信息。

工作负载详情

Pod 详情

功能开关

Request 推荐：集群级别的功能开关。开启后，可以根据 Workload 的使用情况，给 Workload 推荐合适的 Request。

推荐值（蓝色字体）：标识当前工作负载的 Request 推荐值，单击后支持更新（仅在开启 Request 推荐且有推荐值的情况下展示）。

工作负载信息

24h 利用率均值：工作负载 CPU 和内存24小时均值利用率。

24h 利用率峰值：工作负载 CPU 和内存24小时峰值利用率。

工作负载走势图信息

CPU 用量：当前工作负载的 CPU 用量情况。

申请量：当前工作负载中所有 Pod 的 Request 的 CPU 数值之和。

实际用量：当前工作负载中所有 Pod 实际的 CPU 用量之和。

推荐量：当前工作负载中所有 Pod 推荐的 CPU 用量之和。需要提前开启 Request 推荐。

内存用量：当前工作负载的内存用量情况。

申请量：当前工作负载中所有 Pod 的 Request 的内存数值之和。

实际用量：当前工作负载中所有 Pod 实际的内存用量之和。

推荐量：当前工作负载中所有 Pod 推荐的内存用量之和。需要提前开启 Request 推荐。

CPU 用量走势图

申请量：Pod 的 Request 数值。

实际用量：Pod 实际的用量。

推荐量：Pod 推荐的用量。

Pod 详情

Pod 名称：Pod 的名字。

Pod 状态：Pod 的状态。

所在节点：当前 Pod 所属的节点。

节点类型：当前 Pod 所属节点的节点类型。

2. 单击**推荐**，展示当前工作负载的 Request 推荐值。（仅开启 Request 推荐且有推荐值的情况下，支持单击）
3. 单击**删除**，将删除该工作负载。