

应用运维管理

产品介绍

文档版本 01
发布日期 2024-05-11



版权所有 © 华为云计算技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为云计算技术有限公司

地址：贵州省贵安新区黔中大道交兴功路华为云数据中心 邮编：550029

网址：<https://www.huaweicloud.com/>

目录

1 什么是应用运维管理	1
2 产品优势	3
3 应用场景	4
4 与 AOM 1.0 对比	6
5 与其他服务的关系	9
6 AOM 与 CES 的云服务监控功能对比	13
7 约束与限制	15
8 指标总览	20
8.1 简介	20
8.2 基础指标：虚机指标	21
8.3 基础指标：容器指标	31
8.4 基础指标：Modelarts 指标	61
8.5 基础指标：IEF 指标	71
8.6 基础指标：CSE 指标	77
8.7 基础指标：IoTDA 指标	85
8.8 基础指标：node-exporter 指标	91
8.9 指标维度	94
9 安全	97
9.1 责任共担	97
9.2 身份认证与访问控制	98
9.2.1 服务的访问控制	98
9.3 数据保护技术	98
9.4 审计与日志	99
9.5 服务韧性	100
9.6 监控安全风险	100
9.7 认证证书	101
10 基本概念	103
10.1 应用资源管理	103
10.2 资源监控	104
10.3 自动化运维	105

10.4 采集管理.....	106
11 权限管理.....	107
12 隐私声明.....	120
13 修订记录.....	121

1 什么是应用运维管理

应用运维管理（Application Operations Management，简称AOM）是云上应用的一站式立体化运维管理平台，融合云监控、云日志、应用性能、真实用户体验、后台链接数据等多维度可观测性数据源，提供应用资源统一管理、一站式可观测性分析和自动化运维方案，帮助用户及时发现故障，全面掌握应用、资源及业务的实时运行状况，提升企业海量运维的自动化能力和效率。

图 1-1 AOM 产品结构图



- 托管&运行态**

AOM可无缝对接多个上层运维服务，支持快速从应用管理与运维平台（ServiceStage）、函数工作流（FunctionGraph）和微服务引擎（Cloud Service Engine，CSE）等上层运维服务采集指标数据并实时、可视化呈现。
- 可观测性分析**

基于四层指标体系（基础设施层指标、中间件层指标、应用层指标和业务层指标），通过事务监控、容器监控、普罗监控等功能，提供异常检测、历史数据分析、性能分析、关联性和场景化分析等可观测性分析能力。
- 运维自动化**

提供批量磁盘清理、作业编排、脚本执行等功能，将日常运维操作服务化、标准化、自动化。

- 应用资源管理

通过应用管理、资源搜索等功能，将所有资源对象与应用统一管理，为AOM上层运维场景服务提供准确、及时、一致的资源配置数据。

- 采集管理

统一管理插件，并为AOM提供指令下发功能，如脚本下发和执行。

- 开放

支持原生PromQL数据上报，也可通过API等方式上报数据，并也可通过grafana查看数据，通过kafka转储数据。

2 产品优势

- **兼容开放**
支持各类开源协议低门槛接入，开放了运营、运维数据查询接口和采集标准，提供全托管、免运维、低成本的云原生监控能力。
- **开箱即用**
无需修改业务代码即可接入使用，非侵入式数据采集，安全无忧。
- **丰富的数据源**
融合云监控、云日志、应用性能、真实用户体验、后台链接数据等多维度可观测性数据源。
- **全栈一体化监控**
覆盖用户端、服务端及云产品所有监控数据，提供从数据发现、数据展示到数据异常告警的全方位、可视化监控服务，实现从上到下、从前端到后端的全栈一体化监控。
- **关联分析**
应用和资源层层自动关联，通过应用、组件、实例、主机和事务等多视角下钻分析指标、日志、告警数据，直击异常。
- **精准告警**
构建指标告警、事件告警、日志告警的统一告警体系，提供分组、抑制和静默等告警降噪策略，结合告警通知和订阅等功能，帮助用户从容应对海量告警风暴，快速检测和修复业务告警。
- **统一可视化**
支持对多种数据源，在同一个仪表盘中统一监控与分析，并以折线图、数字图等形式多样化、可视化呈现，帮助用户快速监控资源、及时获取数据发展趋势和制定决策。

3 应用场景

提升用户体验

业务痛点

行业竞争激烈，用户对服务的体验要求高，尤其是用户体验成为互联网企业的核心竞争力。如何能获取用户的真实体验感受，减少流失率，提升用户转化率成为企业的难题。

解决方案

AOM提供应用体验管理能力，实时分析应用事务从用户请求、服务器到数据库，再到服务器、用户请求的完整过程，实时感知用户对应用的满意度，帮助您全面了解用户体验状况。对于用户体验差的事务，通过拓扑和调用链完成事务问题定位。

- 通过前端性能监控，提供页面的性能、JS错误请求、API请求、服务的运营相关指标（PV/UV）情况，实时了解并追踪应用性能情况。
- 通过用户会话追踪，定位使用过程中影响用户体验的慢请求、慢加载、慢交互等问题，实时了解用户使用情况。
- 通过页面加载性能分析，提供多维度页面的首屏时间、白屏时间、可交互时间等页面加载指标数据，全面还原用户的体验感受，定位访问慢的原因。

定位应用性能瓶颈

业务痛点

随着业务不断发展，业务逻辑日益复杂，导致应用性能问题分析与定位日益艰难，给监控运维带来了巨大的挑战：应用之间的依赖关系复杂，难以梳理。调用链路长，排查和定位群体困难。接口调用、数据库调用关系复杂，管理难度大。

解决方案

AOM提供大型分布式应用异常诊断能力，当应用出现崩溃或请求失败时，通过应用拓扑+调用链下钻能力分钟级完成问题定位。

- 基于应用拓扑自助发现，定位性能瓶颈。
- 真实还原应用大规模业务访问场景，帮助用户提前识别应用性能问题。
- 基于关键性能指标对比，优化应用性能。
- 根据指标变化趋势配置告警，及时了解异常。

容器运维场景

业务痛点

Prometheus是容器场景的最佳监控工具，但自建Prometheus对于运维人力有限的中小型企业而言，成本太高。对于业务发展快速的大企业又容易出现性能瓶颈。因而使用云上托管Prometheus已成为越来越多上云企业的第一选择。

解决方案

AOM全面对接开源Prometheus生态，将容器服务Kubernetes集群接入Prometheus监控，通过Grafana大盘监控主机和Kubernetes集群的众多性能指标。

- 通过CCE的kube-prometheus-stack插件、自建K8s集群、ServiceMonitor、PodMonitor等多种方式采集上报指标，监控部署在CCE集群内的业务数据。
- 通过丰富的告警模板，帮助业务快速发现和定位问题。

4 与 AOM 1.0 对比

AOM 2.0基于AOM 1.0原有功能，结合用户常用的应用监控，增加了多种指标和日志数据的采集与监控，并对监控结果可视化呈现。同时，通过自动化运维功能将日常运维操作服务化、自动化，减少运维人员重复性操作。

本章节主要对比AOM 2.0与AOM 1.0版本的功能特性。

表 4-1 AOM 2.0 与 AOM 1.0 版本的功能特性对比

功能项		功能项说明	AOM 1.0	AOM 2.0
应用资源管理	资源检索	提供应用、主机等资源的检索功能，支持通过ID、关键字、名称等方式快速检索资源。	不支持	支持
	应用管理	管理云服务对象与应用之间的关系，遵循“应用 + 子应用（可选）+ 组件 + 环境”的应用管理模型。		
	资源管理	全局查看所有的云服务资源对象与应用的关联关系，便于用户对资源进行分析和管理的。		
	环境标签	为已创建的应用环境添加标签，便于用户快速过滤和查找相同属性的应用环境。		
资源监控	接入中心	快速接入需要监控的业务层、应用层、中间件层、基础设施层指标。	不支持	支持
	仪表盘	将不同图表展示到同一个屏幕上，通过多种图表形式展示资源的指标、日志和性能数据。		

功能项	功能项说明	AOM 1.0	AOM 2.0
告警管理	对服务设置事件条件或者对资源的指标设置阈值条件，当AOM自身或外部服务存在异常或可能存在异常而产生告警时，将告警信息以邮件、短信、企业微信等方式发送给指定的人员。	部分支持 只支持按指标类型和Prometheus命令创建指标告警规则，不支持按全量指标创建。	支持
应用洞察	基于CMDB模型，从应用、组件到环境对资源健康状态分层监控。	部分支持 不支持基于CMDB模型，对指标数据分层监控。	支持
容器洞察	从工作负载和集群2个维度监控CCE的资源使用、状态和告警等信息，以便快速响应，保证工作负载顺畅运行。	支持	支持
指标浏览	实时监控日志数据和各资源的指标数据及趋势，还可对关注的指标进行创建告警规则等操作，以便实时查看业务及分析数据关联分析。	部分支持 只支持监控分析指标数据，不支持监控分析日志数据。	支持
基础设施监控	实时监控主机和云服务运行状态和虚机CPU、内存、磁盘等信息。	支持	支持
Prometheus监控	全面对接开源Prometheus生态，支持类型丰富的组件监控，提供多种开箱即用的预置监控大盘，可灵活扩展云原生组件指标插件。	不支持	支持
业务监控	将ELB上报到LTS的日志数据提取为指标来统一管理，便于后续在指标浏览、仪表盘界面实时监控。	不支持	支持
日志分析	通过日志接入和日志搜索功能帮您快速在海量日志中查询到所需的日志，还可通过分析日志的来源信息和上下文原始数据快速定位问题。	支持	支持
进程监控	通过配置应用发现规则发现和收集主机上部署的应用和关联的指标，并支持从应用下钻到组件，从组件下钻到实例，从实例下钻到容器。通过各层状态，完成对应用或组件的立体监控。	支持	支持

功能项		功能项说明	AOM 1.0	AOM 2.0
	数据订阅	支持用户订阅指标或者告警信息，订阅后可以将数据转发到用户配置的kafka或DMS的Topic中，供消费者消费转发的订阅的信息。	支持	支持
自动化运维	服务场景	提供不同任务类型，并且可以对各原子服务场景卡片进行管理。	不支持	支持
	定时运维	提供创建定时任务等功能，并展示定时任务的执行记录。		
	任务管理	提供执行任务等功能，并展示所有任务的执行记录。		
	参数库	提供创建参数等功能，并展示现有所有参数信息。		
	作业管理	提供创建作业等功能，并支持对作业进行管理。		
	脚本管理	提供创建脚本等功能，并支持对脚本进行版本管理。		
	文件管理	提供创建文件包等功能，并支持对文件包进行版本管理。		
采集管理	UniAgent管理	通过UniAgent调度采集任务对数据进行采集。目前UniAgent安装支持两种方式：手动安装和自动安装。	不支持	支持
	插件市场	提供系统插件、中间件、自定义插件能力。	不支持	支持
	采集任务	UniAgent通过执行插件脚本，采集数据。主要有三大类采集任务分别为系统采集任务、中间件采集任务和自定义采集任务。	不支持	支持

随着AOM 1.0相关功能逐步被AOM 2.0全面取代，AOM 1.0后续会逐步下线，建议用户将AOM 1.0升级到AOM 2.0版本，升级的相关操作请参见[升级AOM](#)。

5 与其他服务的关系

AOM可与消息通知服务、分布式消息服务、云审计等服务配合使用。例如，通过消息通知服务您可将AOM的告警规则状态变更信息通过短信或电子邮件的方式发送给相关人员。同时AOM对接了虚拟私有云、弹性负载均衡等中间件服务，通过AOM您可对这些中间件服务进行监控。AOM还对接了云容器引擎、云容器实例等服务，通过AOM您可对这些服务的基础资源和应用进行监控，并且还可查看相关的日志和告警。

消息通知服务

消息通知服务（Simple Message Notification，简称SMN）可以依据您的需求主动推送通知消息，最终您可以通过短信、电子邮件、应用等方式接收通知信息。您也可以应用之间通过消息通知服务实现应用的功能集成，降低系统的复杂性。

AOM使用SMN提供的消息发送机制，当您因不在现场而无法通过AOM查询阈值规则状态的变更信息时，能及时将该变更信息以邮件或短信的方式发送给相关人员，以便您及时获取资源运行状态等信息并采取相应措施，避免因资源问题造成业务损失。详细内容请参见[创建指标告警规则](#)。

对象存储服务

对象存储服务（Object Storage Service，简称OBS）是一个基于对象的海量存储服务，为客户提供海量、安全、高可靠、低成本的数据存储能力，包括：创建、修改、删除桶，上传、下载、删除对象等。

AOM支持将日志转储到OBS的桶中，以便长期存储。

云日志服务

云日志服务（Log Tank Service，简称LTS）可以提供日志收集、分析、存储等服务。用户可以通过云日志服务快速高效地进行设备运维管理、用户业务趋势分析、安全监控审计等操作。

AOM作为华为云可观测性分析统一入口，自身并不具有日志服务功能，所有的日志服务及相关功能均集成自LTS。

云审计服务

云审计服务（Cloud Trace Service，简称CTS）为您提供云账户下资源的操作记录，通过操作记录您可以实现安全分析、资源变更、合规审计、问题定位等场景。您可以通

过配置OBS对象存储服务，将操作记录实时同步保存至OBS，以便保存更长时间的操作记录。

通过CTS您可记录与AOM相关的操作，便于日后的查询、审计和回溯。

统一身份认证服务

统一身份认证服务（Identity and Access Management，简称IAM）是提供身份认证、权限分配、访问控制等功能的身管理服务。

通过IAM可对AOM进行认证鉴权及细粒度授权。

云监控服务

云监控服务（Cloud Eye）为您提供一个针对弹性云服务器、带宽等资源的立体化监控平台。使您全面了解云上的资源使用情况、业务的运行状况，并及时收到异常告警做出反应，保证业务顺畅运行。

AOM通过调用云监控服务的接口来获取各种云服务监控数据，并展现在AOM界面，以便在AOM界面统一对这些服务进行监控。

应用性能管理服务

应用性能管理服务（Application Performance Management，简称APM）是实时监控并管理云应用性能和故障的云服务，提供专业的分布式应用性能分析能力，可以帮助运维人员快速解决应用在分布式架构下的问题定位和性能瓶颈等难题，为您的体验保驾护航。

为了更好的监控、管理应用，AOM集成了APM的相关功能。

虚拟私有云服务

虚拟私有云服务（Virtual Private Cloud，简称VPC）为弹性云服务器构建隔离的、您自主配置和管理的虚拟网络环境，提升您云中资源的安全性，简化您的网络部署。

当您开通了VPC后，无需额外安装其他插件，即可在AOM界面监控VPC的运行状态及各种指标。

弹性负载均衡服务

弹性负载均衡服务（Elastic Load Balance，简称ELB）是将访问流量根据转发策略分发到后端多台云服务器流量分发控制服务。弹性负载均衡可以通过流量分发扩展应用系统对外的服务能力，通过消除单点故障提升应用系统的可用性。

当您开通了ELB后，无需额外安装其他插件，即可在AOM界面监控ELB的运行状态及各种指标。

关系型数据库服务

关系型数据库服务（Relational Database Service，简称RDS）是一种基于云计算平台的即开即用、稳定可靠、弹性伸缩、便捷管理的在线关系型数据库服务。

当您开通了RDS后，无需额外安装其他插件，即可在AOM界面监控RDS的运行状态及各种指标。

分布式缓存服务

分布式缓存服务（Distributed Cache Service，简称DCS）是华为云提供的一款内存数据库服务，兼容了Redis、Memcached和内存数据网格三种内存数据库引擎，为您提供即开即用、安全可靠、弹性扩容、便捷管理的在线分布式缓存能力，满足高并发及数据快速访问的业务诉求。

当您开通了DCS后，无需额外安装其他插件，即可在AOM界面监控DCS的运行状态及各种指标。

云容器引擎服务

云容器引擎服务（Cloud Container Engine，简称CCE）是提供高性能可扩展的容器服务，基于云服务器快速构建高可靠的容器集群，深度整合网络和存储能力，兼容Kubernetes及Docker容器生态。帮助您轻松创建和管理多样化的容器工作负载，并提供容器故障自愈，监控日志采集，自动弹性扩容等高效运维能力。

通过AOM界面您可监控CCE的基础资源和运行在其上的应用，同时在AOM界面还可查看相关的日志和告警。

云容器实例服务

云容器实例服务（Cloud Container Instance，简称CCI）提供 Serverless Container（无服务器容器）引擎，让您无需创建和管理服务器集群即可直接运行容器。

通过AOM界面您可监控CCI的基础资源和运行在其上的应用，同时在AOM界面还可查看相关的日志和告警。

资源编排服务

资源编排服务（Resource Formation Service，简称RFS）通过图形化设计器，直观便捷的进行云服务资源开通和应用部署，将复杂的云服务资源配置和应用部署配置通过模板描述，从而实现一键式云资源与应用的开通与复制；同时在示例模板中提供了海量的免费应用模板，覆盖各热点云服务应用场景，方便您直接使用或为您设计个性化模板提供参考。

通过AOM界面您可监控RFS的基础资源和运行在其上的应用，同时在AOM界面还可查看相关的日志和告警。

应用管理与运维平台

应用管理与运维平台（ServiceStage）是面向企业的一站式PaaS平台服务，提供应用云上托管解决方案，帮助企业简化部署、监控、运维和治理等应用生命周期管理问题；提供微服务框架，兼容主流开源生态，不绑定特定开发框架和平台，帮助企业快速构建基于微服务架构的分布式应用。

通过AOM界面您可监控ServiceStage的基础资源和运行在其上的应用，同时在AOM界面还可查看相关的日志和告警。

函数 workflow 服务

函数 workflow 服务（FunctionGraph）是一款无服务器（Serverless）计算服务，无服务器计算是一种托管服务，服务提供商会实时为你分配充足的资源，而不需要预留专用的服务器或容量，真正按实际使用付费。

通过AOM界面您可监控FunctionGraph的基础资源和运行在其上的应用，同时在AOM界面还可查看相关的日志和告警。

智能边缘平台

智能边缘平台（Intelligent EdgeFabric，简称IEF）通过纳管您的边缘节点，提供将云上应用延伸到边缘的能力，联动边缘和云端的数据，满足客户对边缘计算资源的远程管控、数据处理、分析决策、智能化的诉求，同时，在云端提供统一的设备/应用监控、日志采集等运维能力，为企业提供完整的边缘和云协同的一体化服务的边缘计算解决方案。

AOM提供对IEF资源的运维能力，无需额外安装其他插件，通过AOM您可监控IEF的资源（例如：边缘节点、应用、函数），同时在AOM还可以查看IEF资源的日志和告警。

弹性云服务器

弹性云服务器（Elastic Cloud Server，简称ECS）是由CPU、内存、镜像、云硬盘组成的一种可随时获取、弹性可扩展的计算服务器，同时它结合虚拟私有云、虚拟防火墙、数据多副本保存等能力，为您打造一个高效、可靠、安全的计算环境，确保您的服务持久稳定运行。弹性云服务器创建成功后，您就可以像使用自己的本地PC或物理服务器一样，在云上使用弹性云服务器。

您购买了弹性云服务器（弹性云服务器操作系统需满足表7-4中的使用限制，且购买后需要给弹性云服务器安装UniAgent，否则无法使用AOM监控）后，在AOM界面可对弹性云服务器的基础资源和运行在其上的应用进行监控，同时在AOM界面还可查看相关的日志和告警。

裸金属服务器

裸金属服务器（Bare Metal Server，简称BMS）为您和您的企业提供专属的云上物理服务器，为核心数据库、关键应用系统、高性能计算、大数据等业务提供卓越的计算性能以及数据安全，结合云中资源的弹性优势，租户可灵活申请，按需使用。

您购买了裸金属服务器（裸金属服务器操作系统需满足表7-4中的使用限制，且购买后需要给裸金属服务器安装UniAgent，否则无法使用AOM监控）后，在AOM界面可对裸金属服务器的基础资源和运行在其上的应用进行监控，同时在AOM界面还可查看相关的日志和告警。

6 AOM 与 CES 的云服务监控功能对比

本章节主要对比AOM 2.0的云服务监控与CES的云服务监控功能差异。

AOM的指标数据来源于CES。AOM的指标数据为Prometheus格式，CES为自定义格式。AOM与CES的云服务监控功能对比如表6-1所示。

表 6-1 AOM 的云服务监控与 CES 的云服务监控功能对比

功能	CES云监控	AOM 2.0云监控
跨账号统一监控	不支持	支持
自定义数据存储时长	不支持（默认可存储3个月）。	支持（最长367天）。
数据导出	<ul style="list-style-type: none">支持导出近3个月聚合数据。支持导出近48个小时原始数据。	<ul style="list-style-type: none">支持仪表盘、API导出。支持数据写入kafka。
聚合查询	不支持聚合查询，仅支持简单查询。	支持多实例聚合查询，例如：按标签、按资源组等聚合。
PromQL语法	不支持	支持，告警规则、仪表盘、指标浏览等监控功能均支持。
仪表盘	支持标准云产品的单实例大盘。	提供丰富的预置模板。
仪表盘支持的图表类型	2	8+
仪表盘支持的监控视图数	50	100+
创建告警规则数	仅支持1000。	默认3000+，最多支持10000+。

功能	CES云监控	AOM 2.0云监控
告警模板可添加告警规则数	仅支持50。	支持添加20+云服务，每个云服务支持添加100+告警规则。
历史告警保存时间	7天。	1年。
单次创建告警规则可选择的被监控对象数	5000	不限制，支持全部资源、正则匹配、精准匹配。
聚合告警	不支持	支持基于PromQL语法的聚合告警，例如：按CDN总带宽、EIP总带宽和OBS存储等条件聚合告警。
对接线下Grafana	不支持	支持以Prometheus数据源的方式直接对接到线下Grafana。
对接线下自建Prometheus	不支持	支持数据直接写入自建Prometheus。
业务监控	不支持	支持基于Prometheus、LTS日志和自定义渠道的业务监控。
应用监控	不支持	支持对JVM、Spring、Nginx、Tengine和Tomcat等的监控。
线下IDC监控	不支持	支持基于Prometheus Exporter的线下硬件、存储和网络等监控。
线下中间件监控	不支持	支持对线下MongoDB、Redis和RocketMQ等中间件的监控。

7 约束与限制

应用资源管理使用限制

表 7-1 应用资源管理使用限制

对象	使用限制
应用	租户创建的应用个数 ≤ 1000 。
子应用	应用下创建的子应用数量 ≤ 50 。
组件	子应用下的创建的组件数量 ≤ 50 。
环境	组件下可创建的环境数量 ≤ 20 。
资源	单个环境管理的同类实例资源数量 ≤ 2000 。

资源监控使用限制

表 7-2 资源监控使用限制

分类	对象	使用限制
仪表盘	仪表盘	1个区域中最多可创建1000个仪表盘。
	仪表盘中的图表	1个仪表盘中最多可添加30个图表。
	仪表盘中图表可选资源、阈值规则、组件或主机的个数	<ul style="list-style-type: none">1个数字图最多可添加12个资源，只能展示1个资源，默认展示第一个资源。1个阈值状态图表最多可添加10个阈值规则。1个主机状态图表最多可添加10个主机。1个组件状态图表最多可添加10个组件。

分类	对象	使用限制
指标	指标数据	<ul style="list-style-type: none"> 基础规格：指标数据在数据库中最多保存7天。 专业规格：指标数据在数据库中最多保存30天。
	指标项	资源（例如，集群、组件、主机等）被删除后，其关联的指标项在数据库中最多保存30天。
	维度	每个指标的维度最多为20个。
	指标查询接口	单次最大可同时查询20个指标。
	统计周期	最大统计周期为1小时。
	单次查询返回指标数据	单个指标单次查询最大返回1440个数据点。
	自定义指标	无限制。
	上报自定义指标	单次请求数据最大不能超过40KB，上报指标所带时间戳不能超前于标准UTC时间10分钟，不接收乱序指标，即有新指标上报后，旧指标上报将会失败。
	应用指标 JOB指标	<ul style="list-style-type: none"> 每个主机的容器个数超过1000个时，ICAgent将停止采集该主机应用指标，并发送“ICAgent停止采集应用指标”告警（告警ID：34105）。 每个主机的容器个数缩减到1000个以内时，ICAgent将恢复该主机应用指标采集，并清除“ICAgent停止采集应用指标”告警。 由于JOB在完成任务之后，会自动退出。如果您需要监控JOB指标，要保证存活时间大于90秒才能采集到指标数据。
采集器资源消耗	采集器在采集基础指标时的资源消耗情况和容器、进程数等因素有关，在未运行任何业务的VM上，采集器将消耗30M内存、1% CPU。为保证采集可靠性，单节点上运行的容器个数应小于1000。	
告警规则	告警规则	告警规则（包含指标告警规则和事件告警规则）最多可创建3000个。
	告警模板	告警模板最多可创建150个。
日志	日志功能限制	日志功能限制请参考 LTS日志限制 。
	日志文件	只支持采集文本类型日志文件，不支持采集其他类型日志文件（例如二进制文件）。 每个通过卷挂载日志的路径下，ICAgent最多采集20个日志文件。

分类	对象	使用限制
		每个ICAgent最多采集1000个容器标准输出日志文件，容器标准输出日志只支持json-file类型。
	采集日志文件的资源消耗	日志文件采集时消耗的资源、日志量、文件个数及网络带宽、backend服务处理能力等多种因素强相关。
	日志丢弃	当单行日志长度超过10240字节时，此行会被丢弃。
	日志采集路径	<p>Linux</p> <ul style="list-style-type: none"> 采集路径支持递归路径，**表示递归5层目录。示例：/var/logs/**/a.log。 采集路径支持模糊匹配，匹配目录或文件名中的任何字符。示例：/var/logs/*/a.log、/var/logs/service/a*.log。 采集路径如果配置的是目录，示例：/var/logs/，则只采集目录下后缀为“.log”、“.trace”和“.out”的文件；如果配置的是文件名，则直接采集对应文件，只支持文本类型的文件。 采集路径不能重复配置，即同一主机下的同一路径，即使跨日志组和日志流，也只能配置一次。 <p>Windows</p> <ul style="list-style-type: none"> Windows环境日志采集路径支持递归路径，**表示递归5层目录。配置样例：C:\var\service**\a.log。 Windows环境日志采集路径支持模糊匹配，匹配目录或文件名中的任何字符。配置样例：C:\var\service*\a.log、C:\var\service\a*.log。 采集路径不能重复配置，即同一主机下的同一路径，即使跨日志组和日志流，也只能配置一次。 windows事件日志采集不能重复配置，即同一主机下，即使跨日志组和日志流，也只能配置一次。
	日志重复	当采集器被重启后，重启时间点附近可能会产生一定的数据重复。
	历史日志	日志数据存储时长与您选择的版本有关，且收费不同，详见 价格详情 。
告警列表	告警	可查询最近一年内，时间跨度不超过31天的告警。

分类	对象	使用限制
	事件	可查询最近一年内，时间跨度不超过31天的事件。
应用发现	应用发现规则	应用发现规则最多可创建100个。

自动化运维使用限制

表 7-3 自动化运维使用限制

对象	使用限制
任务	<ul style="list-style-type: none"> 单个操作ECS的任务最多支持选择100个实例。 单个操作RDS的任务最多支持选择20个实例。 单个操作CCE负载的任务最多支持选择10个实例。
作业	<ul style="list-style-type: none"> 单用户支持最多可以创建1000个作业。 每个作业最多支持创建20个全局参数、20个作业步骤、50个执行方案。
脚本	单用户支持最多创建脚本版本合计1000个。
文件包	单用户支持最多创建文件包版本合计1000个。
OS账号	单用户支持最多创建100个账号。
磁盘空间	支持清理1-1000天前的文件。
定时运维	定时运维任务单用户最大配额100个。

采集管理使用限制

- 操作系统使用限制

表 7-4 采集管理支持的操作系统及版本

操作系统	版本				
	Euler OS	1.1 64bit	2.0 64bit		
Cent OS	7.1 64bit	7.2 64bit	7.3 64bit	7.4 64bit	7.5 64bit
	7.6 64bit	7.7 64bit	7.8 64bit	7.9 64bit	8.0 64bit

操作系统	版本			
Ubuntu	16.04 server 64bit	18.04 server 64bit	20.04 server 64bit	22.04 server 64bit

📖 说明

- 对于Linux x86_64服务器，采集管理支持上表中所有的操作系统及版本。
- 对于Linux ARM服务器，当前CentOS操作系统支持7.4/7.5/7.6版本，EulerOS操作系统支持2.0版本，Ubuntu操作系统支持18.04版本。
- 资源使用限制

表 7-5 资源使用限制

对象	使用限制
Agent客户端	当连续两分钟平均CPU使用率大于50%或者内存大于100M时，Agent客户端将自动重启。
Agent安装、升级或卸载	一次最多可对100台主机安装、升级或卸载Agent。
主机删除	一次最多可删除50台卸载了Agent的主机记录。

8 指标总览

8.1 简介

指标是对资源性能的数据描述或状态描述，指标由**命名空间**、**维度**、指标名称和单位组成。

指标命名空间

指标命名空间是对一组资源和对象产生的指标的抽象整合，不同命名空间中的指标彼此独立，因此来自不同应用程序的指标不会被错误地汇聚到相同的统计信息中。

- 系统指标的命名空间：命名空间是固定不变的，均以“PAAS.”开头，如表8-1所示。

表 8-1 系统指标命名空间

命名空间名称	说明
PAAS.AGGR	集群指标的命名空间。
PAAS.NODE	主机指标、网络指标、磁盘指标和文件系统指标的命名空间。
PAAS.CONTAINER	组件指标、实例指标、进程指标和容器指标的命名空间。
PAAS.SLA	SLA指标的命名空间。

- 自定义指标的命名空间：用户自定义，自定义时命名空间必须以字母开头，但不能以“PAAS.”、“SYS.”和“SRE.”开头，且以0~9、a~z、A~Z或下划线（_）组成的格式为XX.XX的3~32位字符串。

指标维度

维度是指标的分类。每个指标都包含用于描述该指标的特定特征，可以将维度理解为这些特征类别。

- 系统指标维度：维度是固定不变的，不同类型的指标维度不同，维度信息的详细说明请参见[指标维度](#)。

- 自定义指标维度：维度为1~32位的字符串，需要您自定义。

8.2 基础指标：虚拟机指标

介绍通过ICAgent上报到AOM的虚拟机指标类别、名称、含义等信息。

表 8-2 虚拟机指标说明

指标类别	指标	指标名称	指标含义	取值范围	单位
网络指标	aom_node_network_receive_bytes	下行Bps	该指标用于统计测试对象的入方向网络流速。	≥0	字节/秒 (Byte/s)
	aom_node_network_receive_packets	下行Pps	每秒网卡接收的数据包个数。	≥0	个/秒 (Packet/s)
	aom_node_network_receive_error_packets	下行错包率	每秒网卡接收的错误包个数。	≥0	个/秒 (Count/s)
	aom_node_network_transmit_bytes	上行Bps	该指标用于统计测试对象的出方向网络流速。	≥0	字节/秒 (Byte/s)
	aom_node_network_transmit_error_packets	上行错包率	每秒网卡发送的错误包个数。	≥0	个/秒 (Count/s)
	aom_node_network_transmit_packets	上行Pps	每秒网卡发送的数据包个数。	≥0	个/秒 (Packet/s)
	aom_node_network_total_bytes	总Bps	该指标用于统计测试对象出方向和入方向的网络流速之和。	≥0	字节/秒 (Byte/s)
磁盘指标	aom_node_disk_read_kilobytes	磁盘读取速率	该指标用于统计每秒从磁盘读出的数据量。	≥0	千字节/秒 (kB/s)
	aom_node_disk_write_kilobytes	磁盘写入速率	该指标用于统计每秒写入磁盘的数据量。	≥0	千字节/秒 (kB/s)
磁盘分区指标	aom_host_diskpartition_thinpool_metadata_percent	Thin pool元数据空间使用率	该指标用于统计CCE节点上thinpool元数据空间使用百分比。	0~100	百分比 (%)

指标类别	指标	指标名称	指标含义	取值范围	单位
	aom_host_diskpartition_thinpool_data_percent	Thin pool 数据空间使用率	该指标用于统计 CCE 节点上 thinpool 数据空间使用百分比。	0~100	百分比 (%)
	aom_host_diskpartition_total_capacity_megabytes	Thin pool 磁盘分区容量	该指标用于统计 CCE 节点上 thinpool 总空间容量。	≥0	兆字节 (MB)
文件系统指标	aom_node_disk_available_capacity_megabytes	可用磁盘空间	还未经使用的磁盘空间。	≥0	兆字节 (MB)
	aom_node_disk_capacity_megabytes	磁盘空间容量	总的磁盘空间容量。	≥0	兆字节 (MB)
	aom_node_disk_rw_status	磁盘读写状态	该指标用于统计主机上磁盘的读写状态。	0、1 • 0表示读写 • 1表示只读	无
	aom_node_disk_usage	磁盘使用率	已使用的磁盘空间占总的磁盘空间容量百分比。	0~100	百分比 (%)
主机指标	aom_node_cpu_limit_core	CPU 内核总量	该指标用于统计测量对象申请的 CPU 核总量。	≥1	核 (Core)
	aom_node_cpu_used_core	CPU 内核占用	该指标用于统计测量对象已经使用的 CPU 核个数。	≥0	核 (Core)
	aom_node_cpu_usage	CPU 使用率	该指标用于统计测量对象的 CPU 使用率。	0~100	百分比 (%)
	aom_node_memory_free_megabytes	可用物理内存	该指标用于统计测量对象上的尚未被使用的物理内存。	≥0	兆字节 (MB)
	aom_node_virtual_memory_free_megabytes	可用虚拟内存	该指标用于统计测量对象上的尚未被使用的虚拟内存。	≥0	兆字节 (MB)

指标类别	指标	指标名称	指标含义	取值范围	单位
	aom_node_gpu_memory_free_megabytes	显存容量	该指标用于统计测量对象的显存容量。	>0	兆字节 (MB)
	aom_node_gpu_memory_usage	显存使用率	该指标用于统计测量对象已使用的显存占显存容量的百分比。	0~100	百分比 (%)
	aom_node_gpu_memory_used_megabytes	显存使用量	该指标用于统计测量对象已使用的显存。	≥0	兆字节 (MB)
	aom_node_gpu_usage	GPU使用率	该指标用于统计测量对象的GPU使用率。	0~100	百分比 (%)
	aom_node_npu_memory_free_megabytes	NPU存储容量	该指标用于统计测量对象的NPU存储容量。	>0	兆字节 (MB)
	aom_node_npu_memory_usage	NPU存储使用率	该指标用于统计测量对象已使用的NPU存储占NPU存储容量的百分比。	0~100	百分比 (%)
	aom_node_npu_memory_used_megabytes	NPU存储使用量	该指标用于统计测量对象已使用的NPU存储。	≥0	兆字节 (MB)
	aom_node_npu_usage	NPU使用率	该指标用于统计测量对象的NPU使用率。	0~100	百分比 (%)
	aom_node_npu_temperature_centigrade	NPU温度	该指标用于统计NPU的温度。	-	摄氏度 (°C)
	aom_node_memory_usage	物理内存使用率	该指标用于统计测量对象已使用内存占申请物理内存总量的百分比。	0~100	百分比 (%)
	aom_node_status	主机状态	该指标用于统计主机状态是否正常。	<ul style="list-style-type: none"> 0表示正常 1表示异常 	无

指标类别	指标	指标名称	指标含义	取值范围	单位
	aom_node_ntp_offset_ms	NTP偏移量	该指标用于统计主机本地时间与NTP服务器时间的偏移量，NTP偏移量越接近于0，主机本地时间与NTP服务器时间越接近。	-	毫秒（ms）
	aom_node_ntp_server_status	NTP服务器状态	该指标用于统计主机是否成功连接上NTP服务器。	0、1 <ul style="list-style-type: none"> 0表示已连接 1表示未连接 	无
	aom_node_ntp_status	NTP同步状态	该指标用于统计主机本地时间与NTP服务器时间是否同步。	0、1 <ul style="list-style-type: none"> 0表示同步 1表示未同步 	无
	aom_node_process_number	进程数量	该指标用于统计测量对象上的进程数量。	≥0	无
	aom_node_gpu_temperature_c entigrade	显卡温度	该指标用于统计显卡的温度。	-	摄氏度（℃）
	aom_node_memory_total_megabytes	物理内存容量	该指标用于统计测量申请的物理内存总量。	≥0	兆字节（MB）
	aom_node_virtual_memory_total_megabytes	虚拟内存容量	该指标用于统计测量对象上的虚拟内存总量。	≥0	兆字节（MB）
	aom_node_virtual_memory_usage	虚拟内存使用率	该指标用于统计测量对象已使用虚拟内存占虚拟内存总量的百分比。	0~100	百分比（%）
	aom_node_current_threads_number	当前线程数量	该指标用于统计主机中当前创建的线程数量。	≥0	无

指标类别	指标	指标名称	指标含义	取值范围	单位
	aom_node_sys_max_threads_num	最大线程数量	该指标用于统计主机最大可创建的线程数量。	≥0	无
	aom_node_phy_disk_total_capacity_megabytes	物理磁盘总容量	该指标用于统计主机的磁盘总容量。	≥0	兆字节 (MB)
	aom_node_physical_disk_total_used_megabytes	物理磁盘已使用总容量	该指标用于统计主机已使用的磁盘总容量。	≥0	兆字节 (MB)
	aom_billing_hostUsed	主机数量	一天内接入的主机数量。	≥0	无
集群指标	aom_cluster_cpu_limit_core	CPU内核总量	该指标用于统计测量对象申请的CPU核总量。	≥1	核 (Core)
	aom_cluster_cpu_used_core	CPU内核占用	该指标用于统计测量对象已经使用的CPU核数。	≥0	核 (Core)
	aom_cluster_cpu_usage	CPU使用率	该指标用于统计测量对象的CPU使用率。	0~100	百分比 (%)
	aom_cluster_disk_available_capacity_megabytes	可用磁盘空间	还未经使用的磁盘空间。	≥0	兆字节 (MB)
	aom_cluster_disk_capacity_megabytes	磁盘空间容量	总的磁盘空间容量。	≥0	兆字节 (MB)
	aom_cluster_disk_usage	磁盘使用率	已使用的磁盘空间占总的磁盘空间容量百分比。	0~100	百分比 (%)
	aom_cluster_memory_free_megabytes	可用物理内存	该指标用于统计测量对象上的尚未被使用的物理内存。	≥0	兆字节 (MB)
	aom_cluster_virtual_memory_free_megabytes	可用虚拟内存	该指标用于统计测量对象上的尚未被使用的虚拟内存。	≥0	兆字节 (MB)
	aom_cluster_gpu_memory_free_megabytes	显存可用量	该指标用于统计测量对象的显存可用量。	>0	兆字节 (MB)

指标类别	指标	指标名称	指标含义	取值范围	单位
	aom_cluster_gpu_memory_usage	显存使用率	该指标用于统计测量对象已使用的显存占显存容量的百分比。	0~100	百分比 (%)
	aom_cluster_gpu_memory_used_megabytes	显存已使用量	该指标用于统计测量对象已使用的显存。	≥0	兆字节 (MB)
	aom_cluster_gpu_usage	GPU使用率	该指标用于统计测量对象的GPU使用率。	0~100	百分比 (%)
	aom_cluster_memory_usage	物理内存使用率	该指标用于统计测量对象已使用内存占申请物理内存总量的百分比。	0~100	百分比 (%)
	aom_cluster_network_receive_bytes	下行Bps	该指标用于统计测试对象的入方向网络流速。	≥0	字节/秒 (Byte/s)
	aom_cluster_network_transmit_bytes	上行Bps	该指标用于统计测试对象的出方向网络流速。	≥0	字节/秒 (Byte/s)
	aom_cluster_memory_total_megabytes	物理内存容量	该指标用于统计测量申请的物理内存总量。	≥0	兆字节 (MB)
	aom_cluster_virtual_memory_total_megabytes	虚拟内存容量	该指标用于统计测量对象上的虚拟内存总量。	≥0	兆字节 (MB)
	aom_cluster_virtual_memory_usage	虚拟内存使用率	该指标用于统计测量对象已使用虚拟内存占虚拟内存总量的百分比。	0~100	百分比 (%)
容器指标	aom_container_cpu_limit_core	CPU内核总量	该指标用于统计测量对象限制的CPU核总量。	≥1	核 (Core)
	aom_container_cpu_used_core	CPU内核占用	该指标用于统计测量对象已经使用的CPU核个数。	≥0	核 (Core)
	aom_container_cpu_usage	CPU使用率	该指标用于统计测量对象的CPU使用率。服务实际使用的与限制的CPU核数量比率。	0~100	百分比 (%)

指标类别	指标	指标名称	指标含义	取值范围	单位
	aom_container_disk_read_kilobytes	磁盘读取速率	该指标用于统计每秒从磁盘读出的数据量。	≥0	千字节/秒 (kB/s)
	aom_container_disk_write_kilobytes	磁盘写入速率	该指标用于统计每秒写入磁盘的数据量。	≥0	千字节/秒 (kB/s)
	aom_container_filesystem_available_capacity_megabytes	文件系统可用	该指标用于统计测量对象文件系统的可用大小。仅支持1.11及其更高版本的kubernetes集群中驱动模式为devicemapper的容器。	≥0	兆字节 (MB)
	aom_container_filesystem_capacity_megabytes	文件系统容量	该指标用于统计测量对象文件系统的容量。仅支持1.11及其更高版本的kubernetes集群中驱动模式为devicemapper的容器。	≥0	兆字节 (MB)
	aom_container_filesystem_usage	文件系统使用率	该指标用于统计测量对象文件系统使用率。实际使用量与文件系统容量的百分比。仅支持1.11及其更高版本的kubernetes集群中驱动模式为devicemapper的容器。	0~100	百分比 (%)
	aom_container_gpu_memory_free_megabytes	显存容量	该指标用于统计测量对象的显存容量。	>0	兆字节 (MB)
	aom_container_gpu_memory_usage	显存使用率	该指标用于统计测量对象已使用的显存占显存容量的百分比。	0~100	百分比 (%)
	aom_container_gpu_memory_used_megabytes	显存使用量	该指标用于统计测量对象已使用的显存。	≥0	兆字节 (MB)

指标类别	指标	指标名称	指标含义	取值范围	单位
	aom_container_gpu_usage	GPU使用率	该指标用于统计测量对象的GPU使用率。	0~100	百分比 (%)
	aom_container_npu_memory_free_megabytes	NPU存储容量	该指标用于统计测量对象的NPU存储容量。	>0	兆字节 (MB)
	aom_container_npu_memory_usage	NPU存储使用率	该指标用于统计测量对象已使用的NPU存储占NPU存储容量的百分比。	0~100	百分比 (%)
	aom_container_npu_memory_used_megabytes	NPU存储使用量	该指标用于统计测量对象已使用的NPU存储。	≥0	兆字节 (MB)
	aom_container_npu_usage	NPU使用率	该指标用于统计测量对象的NPU使用率。	0~100	百分比 (%)
	aom_container_memory_request_megabytes	物理内存总量	该指标用于统计测量对象限制的物理内存总量。	≥0	兆字节 (MB)
	aom_container_memory_usage	物理内存使用率	该指标用于统计测量对象已使用内存占限制物理内存总量的百分比。	0~100	百分比 (%)
	aom_container_memory_used_megabytes	物理内存使用量	该指标用于统计测量对象实际已经使用的物理内存 (Resident Set Size)。	≥0	兆字节 (MB)
	aom_container_network_receive_bytes	下行Bps	该指标用于统计测试对象的入方向网络流速。	≥0	字节/秒 (Byte/s)
	aom_container_network_receive_packets	下行Pps	每秒网卡接收的数据包个数。	≥0	个/秒 (Packet/s)
	aom_container_network_receive_error_packets	下行错包率	每秒网卡接收的错误包个数。	≥0	个/秒 (Count/s)
	aom_container_network_rx_error_packets	容器错包个数	该指标用于统计测量对象收到错误包的数量。	≥0	个 (Count)

指标类别	指标	指标名称	指标含义	取值范围	单位
	aom_container_network_transmit_bytes	上行Bps	该指标用于统计测试对象的出方向网络流速。	≥0	字节/秒 (Byte/s)
	aom_container_network_transmit_error_packets	上行错包率	每秒网卡发送的错误包个数。	≥0	个/秒 (Count/s)
	aom_container_network_transmit_packets	上行Pps	每秒网卡发送的数据包个数。	≥0	个/秒 (Packet/s)
	aom_process_status	状态	该指标用于统计 Docker 容器状态是否正常。	0、1 • 0表示正常 • 1表示异常	无
	aom_container_memory_workingset_usage	workingset的使用率	该指标用于统计该内存workingset的使用率。	0~100	百分比 (%)
	aom_container_memory_workingset_used_megabytes	内存workingset使用量	该指标用于统计该内存workingset的使用量。	≥0	兆字节 (MB)
进程指标	aom_process_cpu_limit_core	CPU内核总量	该指标用于统计测量对象申请的CPU核总量。	≥1	核 (Core)
	aom_process_cpu_used_core	CPU内核占用	该指标用于统计测量对象已经使用的CPU核个数。	≥0	核 (Core)
	aom_process_cpu_usage	CPU使用率	该指标用于统计测量对象的CPU使用率。服务实际使用的与申请的CPU核数量比率。	0~100	百分比 (%)
	aom_process_handle_count	句柄数	该指标用于统计测量对象使用的句柄数。	≥0	无
	aom_process_max_handle_count	最大句柄数	该指标用于统计测量对象使用的最大句柄数。	≥0	无

指标类别	指标	指标名称	指标含义	取值范围	单位
	aom_process_memory_request_megabytes	物理内存总量	该指标用于统计测量对象申请的物理内存总量。	≥0	兆字节 (MB)
	aom_process_memory_usage	物理内存使用率	该指标用于统计测量对象已使用内存占申请物理内存总量的百分比。	0~100	百分比 (%)
	aom_process_memory_used_megabytes	物理内存使用量	该指标用于统计测量对象实际已经使用的物理内存 (Resident Set Size)。	≥0	兆字节 (MB)
	aom_process_status	状态	该指标用于统计进程状态是否正常。	0、1 • 0表示正常 • 1表示异常	无
	aom_process_thread_count	线程数	该指标用于统计测量对象使用的线程数。	≥0	无
	aom_process_virtual_memory_total_megabytes	虚拟内存总量	该指标用于统计测量对象申请的虚拟内存总量。	≥0	兆字节 (MB)

📖 说明

- 当主机类型为“CCE”时，可以查看磁盘分区指标，支持的系统为：CentOS 7.6版本、EulerOS 2.5。
- 以root用户登录CCE节点，执行`docker info | grep 'Storage Driver'`命令查看docker存储驱动类型。如果执行结果为“Device Mapper”，表示驱动类型为Device Mapper，则支持查看磁盘分区thinpool指标。如果执行结果不为“Device Mapper”，则不支持查看磁盘分区thinpool指标。
- 内存使用率 = (物理内存容量 - 可用物理内存) / 物理内存容量；虚拟内存使用率 = ((物理内存容量 + 虚拟内存总量) - (可用物理内存 + 可用虚拟内存)) / (物理内存容量 + 虚拟内存总量)。目前创建的虚拟机默认虚拟内存为0，在未配置虚拟内存的情况下，监控页面内存使用率，虚拟内存使用率相同。
- 物理磁盘已使用总容量和物理磁盘总容量指标仅统计本地磁盘分区的文件系统类型，不统计主机通过网络形式挂载的文件系统(juicefs、nfs、smb等)。
- 集群指标由AOM通过主机指标汇聚，汇聚的集群指标不包含master节点的主机指标。

8.3 基础指标：容器指标

介绍通过CCE的kube-prometheus-stack插件或者自建K8s集群上报到AOM集群容器指标的类别、名称、含义等信息。

表 8-3 CCE/自建 K8s 集群容器指标

Target名称	Job名称	指标	指标含义
<ul style="list-style-type: none"> • serviceMonitor/monitoring/coredns/0 • serviceMonitor/monitoring/node-local-dns/0 	coredns 和node-local-dns	coredns_build_info	CoreDNS的构建信息
		coredns_cache_entries	缓存中元素的数量
		coredns_cache_size	缓存大小
		coredns_cache_hits_total	缓存命中次数
		coredns_cache_misses_total	缓存未命中次数
		coredns_cache_requests_total	不同维度下DNS解析请求的计数
		coredns_dns_request_duration_seconds_bucket	DNS请求耗时直方图 bucket
		coredns_dns_request_duration_seconds_count	DNS请求耗时直方图 count
		coredns_dns_request_duration_seconds_sum	DNS请求耗时直方图 sum
		coredns_dns_request_size_bytes_bucket	DNS请求大小直方图 bucket
		coredns_dns_request_size_bytes_count	DNS请求大小直方图 count
		coredns_dns_request_size_bytes_sum	DNS请求大小直方图 sum
		coredns_dns_requests_total	DNS请求次数
		coredns_dns_response_size_bytes_bucket	DNS响应大小直方图 bucket
coredns_dns_response_size_bytes_count	DNS响应大小直方图 count		

Target名称	Job名称	指标	指标含义
		coredns_dns_response_size_bytes_sum	DNS响应大小直方图sum
		coredns_dns_responses_total	DNS响应码及数量
		coredns_forward_con_n_cache_hits_total	每个协议和数据流的缓存命中次数
		coredns_forward_con_n_cache_misses_total	每个协议和数据流的缓存未命中次数
		coredns_forward_healthcheck_broken_total	健康检查broken次数
		coredns_forward_healthcheck_failures_total	健康检查失败次数
		coredns_forward_max_concurrent_rejects_total	并发超限拒绝次数
		coredns_forward_request_duration_seconds_bucket	forward请求耗时直方图bucket
		coredns_forward_request_duration_seconds_count	forward请求耗时直方图count
		coredns_forward_request_duration_seconds_sum	forward请求耗时直方图sum
		coredns_forward_requests_total	每个数据流的请求次数
		coredns_forward_responses_total	每个数据流的响应次数
		coredns_health_request_duration_seconds_bucket	health请求耗时直方图bucket
		coredns_health_request_duration_seconds_count	health请求耗时直方图count
		coredns_health_request_duration_seconds_sum	health请求耗时直方图sum

Target名称	Job名称	指标	指标含义
		coredns_health_request_failures_total	health请求失败总数
		coredns_hosts_reloaded_timestamp_seconds	最近一次加载host文件的时间戳
		coredns_kubernetes_dns_programming_duration_seconds_bucket	DNS编程耗时直方图 bucket
		coredns_kubernetes_dns_programming_duration_seconds_count	DNS编程耗时直方图 count
		coredns_kubernetes_dns_programming_duration_seconds_sum	DNS编程耗时直方图 sum
		coredns_local_localhost_requests_total	localhost请求数量
		coredns_nodecache_setup_errors_total	nodecache设置错误次数
		coredns_dns_response_rcode_count_total	每个Zone和Rcode的响应次数
		coredns_dns_request_count_total	DNS请求次数
		coredns_dns_request_do_count_total	设置了DO标志位的请求总数
		coredns_dns_do_requests_total	设置了DO标志位的请求总数
		coredns_dns_request_type_count_total	每个Zone和Type的请求次数
		coredns_panics_total	发生Panic的总数
		coredns_plugin_enabled	各插件的启用状态
		coredns_reload_failed_total	reload失败次数

Target名称	Job名称	指标	指标含义
serviceMonitor/ monitoring/ kube- apiserver/0	apiserver	aggregator_unavailable_apiservice	不可用的APIService数量
		apiserver_admission_controller_admission_duration_seconds_bucket	准入控制器 (Admission Controller) 的处理延时
		apiserver_admission_webhook_admission_duration_seconds_bucket	准入Webhook (Admission Webhook) 的处理延时
		apiserver_admission_webhook_admission_duration_seconds_count	准入Webhook (Admission Webhook) 的处理请求统计
		apiserver_client_certificate_expiration_seconds_bucket	证书剩余有效时长
		apiserver_client_certificate_expiration_seconds_count	证书剩余有效时长
		apiserver_current_inflight_requests	在处理读请求数量
		apiserver_request_duration_seconds_bucket	客户端对APIServer的访问时延
		apiserver_request_total	对APIServer不同请求的计数
		go_goroutines	协程数量
		kubernetes_build_info	构建信息
		process_cpu_seconds_total	进程CPU累计时长
		process_resident_memory_bytes	进程常驻内存集的大小
		rest_client_requests_total	REST请求数量
workqueue_adds_total	工作队列Add次数		
workqueue_depth	工作队列深度		

Target名称	Job名称	指标	指标含义
		workqueue_queue_duration_seconds_bucket	任务在工作队列中存在的时长
		aggregator_unavailable_apiservice_total	不可用的APIService总数
		rest_client_request_duration_seconds_bucket	REST请求耗时
serviceMonitor/ monitoring/ kubelet/0	kubelet	kubelet_certificate_manager_client_expiration_renew_errors	证书续期错误次数
		kubelet_certificate_manager_client_ttl_seconds	Kubelet客户端证书的TTL
		kubelet_cgroup_manager_duration_seconds_bucket	控制组管理器操作的持续时间bucket
		kubelet_cgroup_manager_duration_seconds_count	控制组管理器操作的持续时间count
		kubelet_node_config_error	如果节点遇到与配置相关的错误，则此指标为true（1），否则为false（0）
		kubelet_node_name	节点名称，值始终为1
		kubelet_pleg_relist_duration_seconds_bucket	PLEG中relist Pod耗时bucket
		kubelet_pleg_relist_duration_seconds_count	PLEG中relist Pod耗时count
		kubelet_pleg_relist_interval_seconds_bucket	PLEG中relist时间间隔bucket
		kubelet_pod_start_duration_seconds_count	Pod启动时间count
kubelet_pod_start_duration_seconds_bucket	Pod启动时间bucket		

Target名称	Job名称	指标	指标含义
		kubelet_pod_worker_duration_seconds_bucket	同步单个Pod的持续时间。按操作类型细分：create、update、sync
		kubelet_running_containers	当前运行的容器数
		kubelet_running_pods	当前运行的pod数
		kubelet_runtime_operations_duration_seconds_bucket	运行时操作耗时bucket
		kubelet_runtime_operations_errors_total	按操作类型列出的运行时操作错误的累积数
		kubelet_runtime_operations_total	按操作类型列出的运行时操作的累积数
		kubelet_volume_statistics_available_bytes	volume可用量
		kubelet_volume_statistics_capacity_bytes	volume容量
		kubelet_volume_statistics_inodes	volume inode总数
		kubelet_volume_statistics_inodes_used	volume 已使用inode数量
		kubelet_volume_statistics_used_bytes	volume已用量
		storage_operation_duration_seconds_bucket	存储操作耗时bucket
		storage_operation_duration_seconds_count	存储操作耗时count
		storage_operation_errors_total	存储操作错误次数
		volume_manager_total_volumes	Volume Manager中的volumes总数
		rest_client_requests_total	HTTP请求数，维度包含Status code、Method和Host

Target名称	Job名称	指标	指标含义
		rest_client_request_duration_seconds_bucket	请求时延bucket
		process_resident_memory_bytes	进程常驻内存集的大小
		process_cpu_seconds_total	进程CPU累计时长
		go_goroutines	协程个数
serviceMonitor/ monitoring/ kubelet/1	kubelet	container_cpu_cfs_periods_total	容器已经执行的CPU时间周期数
		container_cpu_cfs_throttled_periods_total	容器被限流的CPU时间周期数
		container_cpu_cfs_throttled_seconds_total	容器被限流的CPU时间
		container_cpu_load_average_10s	容器过去10秒内的CPU负载的平均值
		container_cpu_usage_seconds_total	容器在所有CPU内核上的累积占用时间
		container_file_descriptors	容器打开的文件描述符数量
		container_fs_inodes_free	文件系统的可用inode数量
		container_fs_inodes_total	文件系统的总计inode数量
		container_fs_io_time_seconds_total	磁盘/文件系统花费在I/O上的累计秒数
		container_fs_limit_bytes	容器可以使用的磁盘/文件系统总量
		container_fs_read_seconds_total	容器累积读取磁盘/文件系统数据的秒数
		container_fs_reads_bytes_total	容器累积读取磁盘/文件系统数据的总量
		container_fs_reads_total	容器已完成读取磁盘/文件系统的累积计数
container_fs_usage_bytes	文件系统的使用量		

Target名称	Job名称	指标	指标含义
		container_fs_write_seconds_total	容器累计写入磁盘/文件系统的秒数
		container_fs_writes_bytes_total	容器累积写入磁盘/文件系统数据的总量
		container_fs_writes_total	容器已完成写入磁盘/文件系统的累积计数
		container_memory_cache	容器总页缓存内存
		container_memory_failcnt	容器内存使用达到限制的次数
		container_memory_max_usage_bytes	容器历史最大内存使用量
		container_memory_rss	容器常驻内存集的大小
		container_memory_swap	容器虚拟内存使用量
		container_memory_usage_bytes	容器当前的内存使用量
		container_memory_working_set_bytes	容器工作集内存使用量
		container_network_receive_bytes_total	容器网络累积接收数据总量
		container_network_receive_errors_total	接收时遇到的错误累积计数
		container_network_receive_packets_dropped_total	接收时丢弃的数据包的累积计数
		container_network_receive_packets_total	接收数据包的累积计数
		container_network_transmit_bytes_total	容器网络累积传输数据总量
		container_network_transmit_errors_total	传输时遇到的错误累积计数
		container_network_transmit_packets_dropped_total	传输时丢弃的数据包的累积计数
		container_network_transmit_packets_total	传输数据包的累积计数

Target名称	Job名称	指标	指标含义
		container_spec_cpu_quota	容器分配的CPU配额
		container_spec_memory_limit_bytes	容器可以使用的总内存量限制
		machine_cpu_cores	逻辑CPU个数
		machine_memory_bytes	内存总数
serviceMonitor/ monitoring/ kube-state-metrics/0	kube-state-metrics-prom	kube_cronjob_status_active	当前活动的周期任务
		kube_cronjob_info	周期任务
		kube_cronjob_labels	周期任务标签
		kube_configmap_info	ConfigMap信息
		kube_daemonset_created	Daemonset创建时间
		kube_daemonset_status_current_number_scheduled	当前调度的DaemonSet数量
		kube_daemonset_status_desired_number_scheduled	期望调度的DaemonSet数量
		kube_daemonset_status_number_available	期望运行DaemonSet且已经至少运行一个Pod的节点数
		kube_daemonset_status_number_misscheduled	期望之外运行DaemonSet的节点数
		kube_daemonset_status_number_ready	期望运行DaemonSet且Pod已经就绪的节点数
		kube_daemonset_status_number_unavailable	期望运行DaemonSet但没有Pod运行的节点数
		kube_daemonset_status_updated_number_scheduled	正在运行已更新的DaemonSet的Pod的节点数
		kube_deployment_created	Deployment创建时间戳
kube_deployment_labels	Deployment标签		

Target名称	Job名称	指标	指标含义
		kube_deployment_metadata_generation	Deployment期望状态的generation
		kube_deployment_spec_replicas	Deployment期望的副本数
		kube_deployment_spec_strategy_rollingupdate_max_unavailable	滚动升级时最大不可用副本数
		kube_deployment_status_observed_generation	Deployment控制器观察到的generation
		kube_deployment_status_replicas	Deployment当前的副本数
		kube_deployment_status_replicas_available	Deployment可用的replicas数量
		kube_deployment_status_replicas_ready	Deployment已Ready的replicas数量
		kube_deployment_status_replicas_unavailable	Deployment不可用的replicas数量
		kube_deployment_status_replicas_updated	Deployment已更新的replicas数量
		kube_job_info	job信息
		kube_namespace_labels	命名空间标签
		kube_node_labels	节点标签
		kube_node_info	节点信息
		kube_node_spec_taint	节点污点信息
		kube_node_spec_unchedulable	节点是否可调度
		kube_node_status_allocatable	节点可分配资源
		kube_node_status_capacity	节点资源总量
		kube_node_status_condition	节点状态condition

Target名称	Job名称	指标	指标含义
		kube_node_volcano_oversubscription_status	节点超卖状态
		kube_persistentvolume_status_phase	PV状态的phase
		kube_persistentvolumeclaim_status_phase	PVC状态的phase
		kube_persistentvolume_info	PV信息
		kube_persistentvolumeclaim_info	PVC信息
		kube_pod_container_info	Pod的容器信息
		kube_pod_container_resource_limits	容器的资源limits
		kube_pod_container_resource_requests	容器的资源requests
		kube_pod_container_status_last_terminated_reason	上次terminated的原因
		kube_pod_container_status_ready	容器的readiness是否成功
		kube_pod_container_status_restarts_total	容器重启次数
		kube_pod_container_status_running	容器是否running
		kube_pod_container_status_terminated	容器是否terminated
		kube_pod_container_status_terminated_reason	terminated容器的reason
		kube_pod_container_status_waiting	容器是否waiting
		kube_pod_container_status_waiting_reason	waiting容器的reason
		kube_pod_info	Pod信息
		kube_pod_labels	Pod标签

Target名称	Job名称	指标	指标含义
		kube_pod_owner	Pod的Owner信息
		kube_pod_status_phase	Pod当前的phase
		kube_pod_status_ready	Pod状态是否Ready
		kube_secret_info	Secret信息
		kube_statefulset_created	StatefulSet的创建时间戳
		kube_statefulset_labels	StatefulSet的标签信息
		kube_statefulset_metadata_generation	StatefulSet期望状态的generation
		kube_statefulset_replicas	StatefulSet期望的Pod数
		kube_statefulset_status_observed_generation	StatefulSet控制器观察到的generation
		kube_statefulset_status_replicas	StatefulSet的replicas数量
		kube_statefulset_status_replicas_ready	StatefulSet已ready的replicas数量
		kube_statefulset_status_replicas_updated	StatefulSet已更新的replicas数量
		kube_job_spec_completions	成功完成所需Pod数量
		kube_job_status_failed	失败的Job
		kube_job_status_succeeded	成功的Job
		kube_node_status_allocatable_cpu_cores	节点可分配CPU核数
		kube_node_status_allocatable_memory_bytes	节点可分配内存
		kube_replicaset_owner	RS的所有者信息
		kube_resourcequota	ResourceQuota信息

Target名称	Job名称	指标	指标含义
		kube_pod_spec_volumes_persistentvolumeclaims_info	Pod关联的PVC信息
serviceMonitor/ monitoring/ prometheus- lightweight/ 0	prometheus- lightweight	vm_persistentqueue_blocks_dropped_total	发送队列Block丢弃数量
		vm_persistentqueue_blocks_read_total	发送队列Block读取数量
		vm_persistentqueue_blocks_written_total	发送队列Block写入数量
		vm_persistentqueue_bytes_pending	发送队列Pending字节数
		vm_persistentqueue_bytes_read_total	发送队列读取的字节数
		vm_persistentqueue_bytes_written_total	发送队列写入的字节数
		vm_promscrape_active_scrapers	活跃的采集数量
		vm_promscrape_con_n_read_errors_total	采集读取错误次数
		vm_promscrape_con_n_write_errors_total	采集写入错误次数
		vm_promscrape_max_scrape_size_exceeded_errors_total	采集大小超限错误次数
		vm_promscrape_scrape_duration_seconds_sum	采集耗时sum
		vm_promscrape_scrape_duration_seconds_count	采集耗时count
		vm_promscrape_scrapes_total	采集次数
vmagent_remotewrite_bytes_sent_total	远程写发送字节数		
vmagent_remotewrite_duration_seconds_sum	远程写耗时sum		

Target名称	Job名称	指标	指标含义
		vmagent_remotewrite_duration_seconds_count	远程写耗时count
		vmagent_remotewrite_packets_dropped_total	远程写丢弃包次数
		vmagent_remotewrite_pending_data_bytes	远程写Pending字节数
		vmagent_remotewrite_requests_total	远程写请求次数
		vmagent_remotewrite_retries_count_total	远程写重试次数
		go_goroutines	协程个数
serviceMonitor/ monitoring/ node-exporter/0	node-exporter	node_boot_time_seconds	节点开机时间
		node_context_switches_total	上下文切换次数
		node_cpu_seconds_total	节点CPU时间
		node_disk_io_now	当前运行的I/O数量
		node_disk_io_time_seconds_total	I/O总耗时
		node_disk_io_time_weighted_seconds_total	执行I/O所花费的加权秒数
		node_disk_read_bytes_total	读取成功的字节总数
		node_disk_read_time_seconds_total	读取花费的总时长
		node_disk_reads_completed_total	读取完成的次数
		node_disk_write_time_seconds_total	写入花费的总时长
		node_disk_writes_completed_total	写入完成的次数
	node_disk_written_bytes_total	写入成功的字节总数	

Target名称	Job名称	指标	指标含义
		node_docker_thinpool_data_space_available	docker thinpool数据可用空间
		node_docker_thinpool_metadata_space_available	docker thinpool元数据可用空间
		node_exporter_build_info	NodeExporter构建信息
		node_filefd_allocated	已分配的文件描述符
		node_filefd_maximum	最大文件描述符
		node_filesystem_available_bytes	挂载的文件系统可用空间
		node_filesystem_device_error	statfs获取文件信息失败
		node_filesystem_free_bytes	挂载的文件系统剩余空间
		node_filesystem_readonly	只读挂载的文件系统
		node_filesystem_size_bytes	挂载的文件系统占用空间
		node_forks_total	节点fork总数
		node_intr_total	节点中断总数
		node_load1	节点1分钟CPU负载
		node_load15	节点15分钟CPU负载
		node_load5	节点5分钟CPU负载
		node_memory_Buffers_bytes	节点缓冲区的内存
		node_memory_Cached_bytes	节点页面缓存中的内存
		node_memory_MemAvailable_bytes	节点可用内存
		node_memory_MemFree_bytes	节点空闲内存
		node_memory_MemTotal_bytes	节点内存总量

Target名称	Job名称	指标	指标含义
		node_network_receive_bytes_total	累积接收数据总量
		node_network_receive_drop_total	接收时丢弃的累积计数
		node_network_receive_errs_total	接收时遇到的错误累积计数
		node_network_receive_packets_total	接收数据包的累积计数
		node_network_transmit_bytes_total	累积传输数据总量
		node_network_transmit_drop_total	传输时丢弃的累积计数
		node_network_transmit_errs_total	传输时遇到的错误累积计数
		node_network_transmit_packets_total	传输数据包的累积计数
		node_procs_blocked	当前blocked的进程
		node_procs_running	当前运行的进程
		node_sockstat_sockets_used	正在使用的套接字总量
		node_sockstat_TCP_allocated	已分配的TCP套接字数量
		node_sockstat_TCP_inuse	正在使用的TCP套接字数量
		node_sockstat_TCP_orphan	无主的TCP连接数
		node_sockstat_TCP_tw	等待关闭的TCP连接数
		node_sockstat_UDPLITE_inuse	UDP-Lite 套接字当前使用量
		node_sockstat_UDP_inuse	UDP 套接字当前使用量
		node_sockstat_UDP_mem	UDP 套接字缓冲区使用量
		node_timex_offset_seconds	时钟时间偏移

Target名称	Job名称	指标	指标含义
		node_timex_sync_status	节点时钟同步状态
		node_uname_info	节点uname信息
		node_vmstat_oom_kill	/proc/vmstat中的oom_kill
		process_cpu_seconds_total	进程CPU累计时长
		process_max_fds	进程最大文件描述符
		process_open_fds	进程当前打开的文件描述符
		process_resident_memory_bytes	进程常驻内存集的大小
		process_start_time_seconds	进程启动时间
		process_virtual_memory_bytes	进程虚拟内存大小
		process_virtual_memory_max_bytes	进程虚拟内存最大大小
		node_netstat_Tcp_ActiveOpens	从CLOSED 状态直接转换到SYN-SENT 状态的TCP连接数
		node_netstat_Tcp_PassiveOpens	从LISTEN 状态直接转换到SYN-RCVD 状态的TCP连接数
		node_netstat_Tcp_CurrEstab	当前状态为ESTABLISHED 或 CLOSE-WAIT 的TCP 连接数
		node_vmstat_pgmajfault	/proc/vmstat中的pgmajfault
		node_vmstat_pgpgout	/proc/vmstat中的pgpgout
		node_vmstat_pgfault	/proc/vmstat中的pgfault
		node_vmstat_pgpgin	/proc/vmstat中的pgpgin
		node_processes_max_processes	PID限制值

Target名称	Job名称	指标	指标含义
		node_processes_pids	PID个数
		node_nf_contrack_entries	链接状态跟踪表分配的数量
		node_nf_contrack_entries_limit	链接状态跟踪表总量
		promhttp_metric_handler_requests_in_flight	当前正在处理的metrics数量
		go_goroutines	NodeExporter协程个数
podMonitor/ monitoring/ nvidia-gpu- device- plugin/0	monitoring/ nvidia-gpu- device-plugin	cce_gpu_utilization	gpu卡算力使用率
		cce_gpu_memory_utilization	gpu卡显存使用率
		cce_gpu_encoder_utilization	gpu卡编码使用率
		cce_gpu_decoder_utilization	gpu卡解码使用率
		cce_gpu_utilization_process	gpu各进程算力使用率
		cce_gpu_memory_utilization_process	gpu各进程显存使用率
		cce_gpu_encoder_utilization_process	gpu各进程编码使用率
		cce_gpu_decoder_utilization_process	gpu各进程解码使用率
		cce_gpu_memory_used	gpu显存使用量
		cce_gpu_memory_total	gpu显存总量
		cce_gpu_memory_free	gpu显存空闲量
		cce_gpu_bar1_memory_used	gpu bar1 内存使用量
		cce_gpu_bar1_memory_total	gpu bar1 内存总量
		cce_gpu_clock	gpu时钟频率
cce_gpu_memory_clock	gpu显存频率		

Target名称	Job名称	指标	指标含义
		cce_gpu_graphics_clock	gpu图形处理器频率
		cce_gpu_video_clock	gpu视频处理器频率
		cce_gpu_temperature	gpu温度
		cce_gpu_power_usage	gpu功率
		cce_gpu_total_energy_consumption	gpu总能耗
		cce_gpu_pcie_link_bandwidth	gpu pcie 带宽
		cce_gpu_nvlink_bandwidth	gpu nvlink 带宽
		cce_gpu_pcie_throughput_rx	gpu pcie 接收带宽
		cce_gpu_pcie_throughput_tx	gpu pcie 发送带宽
		cce_gpu_nvlink_utilization_counter_rx	gpu nvlink 接收带宽
		cce_gpu_nvlink_utilization_counter_tx	gpu nvlink 发送带宽
		cce_gpu_retired_pages_sbe	gpu 单比特错误隔离页数量
		cce_gpu_retired_pages_dbe	gpu 双比特错误隔离页数量
		xgpu_memory_total	xgpu显存总量
		xgpu_memory_used	xgpu显存使用量
		xgpu_core_percentage_total	xgpu算力总量
		xgpu_core_percentage_used	xgpu算力使用量
		gpu_schedule_policy	gpu模式分成0、1、2三种：0为显存隔离算力共享模式；1为显存算力隔离模式；2为默认模式表示当前卡还没被用于xgpu设备分配。

Target名称	Job名称	指标	指标含义
		xgpu_device_health	xgpu设备的健康情况。0表示xgpu设备为健康状态，1表示为非健康状态。
serviceMonitor/ monitoring/ prometheus- server/0	prometheus- server	prometheus_build_info	Prometheus构建信息
		prometheus_engine_query_duration_seconds	查询时间
		prometheus_engine_query_duration_seconds_count	查询次数
		prometheus_sd_discovered_targets	各个 job 发现的采集目标数
		prometheus_remote_storage_bytes_total	发送字节数
		prometheus_remote_storage_enqueue_retries_total	入队列重试次数
		prometheus_remote_storage_highest_timestamp_in_seconds	存储在WAL中的任何样本的最大时间戳
		prometheus_remote_storage_queue_highest_sent_timestamp_seconds	远程写入成功发送的最大时间戳
		prometheus_remote_storage_samples_dropped_total	远程写丢弃sample数量
		prometheus_remote_storage_samples_failed_total	远程写失败sample数量
		prometheus_remote_storage_samples_in_total	远程写写入sample数量
		prometheus_remote_storage_samples_pending	远程写pending sample数量
prometheus_remote_storage_samples_retried_total	远程写重试sample数量		

Target名称	Job名称	指标	指标含义
		prometheus_remote_storage_samples_total	远程写sample总数
		prometheus_remote_storage_shard_capacity	用于并行发送到远程存储的队列中每个分片的容量
		prometheus_remote_storage_shards	当前用于并行发送到远程存储的分片数
		prometheus_remote_storage_shards_desired	分片队列期望基于输入样本和输出样本的比率运行的分片数
		prometheus_remote_storage_shards_max	可用于并行发送到远程存储的分片数的最大值
		prometheus_remote_storage_shards_min	可用于并行发送到远程存储的分片数的最小值
		prometheus_tsdb_wal_segment_current	TSDB当前正在写入的WAL段索引
		prometheus_tsdb_head_chunks	head中保存的chunk数量
		prometheus_tsdb_head_series	head中保存的series数量
		prometheus_tsdb_head_samples_appended_total	head中添加的samples数量
		prometheus_wal_watcher_current_segment	每个远程写入实例当前正在读取的WAL段文件
		prometheus_target_interval_length_seconds	抓取时间间隔
		prometheus_target_interval_length_seconds_count	抓取时间间隔count
		prometheus_target_interval_length_seconds_sum	抓取时间间隔sum
		prometheus_target_scrapes_exceeded_body_size_limit_total	采集body大小超限次数

Target名称	Job名称	指标	指标含义
		prometheus_target_scrapes_exceeded_sample_limit_total	采集sample超限次数
		prometheus_target_scrapes_sample_duplicate_timestamp_total	时间戳重复Samples数量
		prometheus_target_scrapes_sample_out_of_bounds_total	时间戳超限Samples数量
		prometheus_target_scrapes_sample_out_of_order_total	乱序Samples数量
		prometheus_target_sync_length_seconds	同步scrape pool的间隔
		prometheus_target_sync_length_seconds_count	同步scrape pool的间隔count
		prometheus_target_sync_length_seconds_sum	同步scrape pool的间隔sum
		promhttp_metric_handler_requests_in_flight	当前正在处理的metrics数量
		promhttp_metric_handler_requests_total	metrics处理次数
		go_goroutines	协程个数
podMonitor/ monitoring/ virtual-kubelet-pods/0	monitoring/ virtual-kubelet-pods	container_cpu_load_average_10s	容器过去10秒内的CPU负载的平均值
		container_cpu_system_seconds_total	容器系统CPU总时长
		container_cpu_usage_seconds_total	容器在所有CPU内核上的累积占用时间
		container_cpu_user_seconds_total	容器用户CPU总时长
		container_cpu_cfs_periods_total	容器已经执行的CPU时间周期数
		container_cpu_cfs_throttled_periods_total	容器被限流的CPU时间周期数

Target名称	Job名称	指标	指标含义
		container_cpu_cfs_throttled_seconds_total	容器被限流的CPU时间
		container_fs_inodes_free	文件系统的可用inode数量
		container_fs_usage_bytes	文件系统的使用量
		container_fs_inodes_total	文件系统的总计inode数量
		container_fs_io_current	磁盘/文件系统当前正在进行的 I/O 数量
		container_fs_io_time_seconds_total	磁盘/文件系统花费在 I/O 上的累计秒数
		container_fs_io_time_weighted_seconds_total	磁盘/文件系统累积加权 I/O 时间
		container_fs_limit_bytes	容器可以使用的磁盘/文件系统总量
		container_fs_reads_bytes_total	容器累积读取磁盘/文件系统数据的总量
		container_fs_read_seconds_total	容器累积读取磁盘/文件系统数据的秒数
		container_fs_reads_merged_total	容器合并读取磁盘/文件系统的累积计数
		container_fs_reads_total	容器已完成读取磁盘/文件系统的累积计数
		container_fs_sector_reads_total	容器已完成扇区读取磁盘/文件系统的累积计数
		container_fs_sector_writes_total	容器已完成扇区写入磁盘/文件系统的累积计数
		container_fs_writes_bytes_total	容器累积写入磁盘/文件系统数据的总量
		container_fs_write_seconds_total	容器累计写入磁盘/文件系统的秒数
		container_fs_writes_merged_total	容器合并写入磁盘/文件系统的累积计数

Target名称	Job名称	指标	指标含义
		container_fs_writes_total	容器已完成写入磁盘/文件系统的累积计数
		container_blkio_device_usage_total	容器区分IO操作对磁盘的使用总量
		container_memory_failures_total	容器内存分配失败的累积计数
		container_memory_failcnt	容器内存使用达到限制的次数
		container_memory_cache	容器总页缓存内存
		container_memory_mapped_file	容器内存映射文件的大小
		container_memory_max_usage_bytes	容器历史最大内存使用量
		container_memory_rss	容器常驻内存集的大小
		container_memory_swap	容器虚拟内存使用量
		container_memory_usage_bytes	容器当前的内存使用量
		container_memory_working_set_bytes	容器工作集内存使用量
		container_network_receive_bytes_total	容器网络累积接收数据总量
		container_network_receive_errors_total	接收时遇到的错误累积计数
		container_network_receive_packets_dropped_total	接收时丢弃的数据包的累积计数
		container_network_receive_packets_total	接收数据包的累积计数
		container_network_transmit_bytes_total	容器网络累积传输数据总量
		container_network_transmit_errors_total	传输时遇到的错误累积计数
		container_network_transmit_packets_dropped_total	传输时丢弃的数据包的累积计数

Target名称	Job名称	指标	指标含义
		container_network_transmit_packets_total	传输数据包的累积计数
		container_processes	容器当前运行的进程数
		container_sockets	容器当前打开套接字的个数
		container_file_descriptors	容器当前打开文件描述符的个数
		container_threads	容器内当前运行的线程数
		container_threads_max	容器内允许运行的最大线程数
		container_ulimits_soft	容器内1号进程的soft ulimit 值。如果为-1，则无限制，优先级和nice除外
		container_tasks_state	处于给定状态的任务数 (sleeping, running, stopped, uninterruptible, or ioawaiting)
		container_spec_cpu_period	容器分配的CPU周期
		container_spec_cpu_shares	容器分配的CPU份额
		container_spec_cpu_quota	容器分配的CPU配额
		container_spec_memory_limit_bytes	容器可以使用的总内存量限制
		container_spec_memory_reservation_limit_bytes	容器可以使用的预留内存限制
		container_spec_memory_swap_limit_bytes	容器可以使用的虚拟内存限制
		container_start_time_seconds	容器已经运行的时间
		container_last_seen	最近一次监控采集器感知到容器的时间

Target名称	Job名称	指标	指标含义
		container_accelerator_memory_used_bytes	容器正在使用的GPU加速卡内存量
		container_accelerator_memory_total_bytes	总GPU加速卡可用内存量
		container_accelerator_duty_cycle	GPU加速卡实际运行时间百分比
podMonitor/ monitoring/ everest-csi-controller/0	monitoring/ everest-csi-controller	everest_action_result_total	action结果数量
		everest_function_duration_seconds_bucket	action耗时直方图bucket
		everest_function_duration_seconds_count	action耗时直方图count
		everest_function_duration_seconds_sum	action耗时直方图sum
		everest_function_duration_quantile_seconds	action所需的时间分位数
		node_volume_read_completed_total	读操作完成次数
		node_volume_read_merged_total	读操作merged次数
		node_volume_read_bytes_total	扇区读字节总数
		node_volume_read_time_milliseconds_total	读操作总耗时
		node_volume_write_completed_total	写操作完成次数
		node_volume_write_merged_total	写操作merged次数
		node_volume_write_bytes_total	扇区写字节总数
node_volume_write_time_milliseconds_total	写操作总耗时		

Target名称	Job名称	指标	指标含义
		node_volume_io_now	当前正在进行的I/O次数
		node_volume_io_time_seconds_total	IO操作总耗时
		node_volume_capacity_bytes_available	可用容量
		node_volume_capacity_bytes_total	总容量
		node_volume_capacity_bytes_used	已用容量
		node_volume_inodes_available	可用inode
		node_volume_inodes_total	总inode
		node_volume_inodes_used	已用inode
		node_volume_read_transmissions_total	读取传输次数
		node_volume_read_timeouts_total	读取超时次数
		node_volume_read_sent_bytes_total	读取字节数
		node_volume_read_queue_time_milliseconds_total	读取队列等待耗时
		node_volume_read_rtt_time_milliseconds_total	读取rtt时长
		node_volume_write_transmissions_total	写入传输次数
		node_volume_write_timeouts_total	写入超时次数
		node_volume_write_queue_time_milliseconds_total	写入队列等待耗时
		node_volume_write_rtt_time_milliseconds_total	写入rtt时长

Target名称	Job名称	指标	指标含义
		node_volume_localvolume_stats_capacity_bytes	本地存储容量
		node_volume_localvolume_stats_available_bytes	本地存储可用量
		node_volume_localvolume_stats_used_bytes	本地存储已用量
		node_volume_localvolume_stats_inodes	本地存储inodes数量
		node_volume_localvolume_stats_inodes_used	本地存储inodes已用量
podMonitor/ monitoring/ nginx-ingress-controller/0	monitoring/ nginx-ingress-controller	nginx_ingress_controller_bytes_sent	发送到客户端的字节数
		nginx_ingress_controller_connect_duration_seconds	与上游服务器建立连接所花费的时间
		nginx_ingress_controller_header_duration_seconds	从上游服务器接收第一个报头所花费的时间
		nginx_ingress_controller_ingress_upstream_latency_seconds	上行服务延迟
		nginx_ingress_controller_request_duration_seconds	请求处理时间（以毫秒为单位）
		nginx_ingress_controller_request_size	请求长度（包括请求行、请求头和请求体长度）
		nginx_ingress_controller_requests	Nginx Ingress Controller自启动以来处理的所有HTTP请求的总数
		nginx_ingress_controller_response_duration_seconds	从上游服务器接收响应所花费的时间
		nginx_ingress_controller_response_size	响应长度（包括请求行、头和请求体长度）

Target名称	Job名称	指标	指标含义
		nginx_ingress_controller_nginx_process_connections	当前处于活动、读取、写入、等待状态，每种状态各自的客户端连接数
		nginx_ingress_controller_nginx_process_connections_total	自启动以来处于接受、已处理状态，每种状态各自的客户端连接数
		nginx_ingress_controller_nginx_process_cpu_seconds_total	Nginx进程消耗的CPU时间总量（以秒为单位）
		nginx_ingress_controller_nginx_process_num_procs	进程数
		nginx_ingress_controller_nginx_process_oldest_start_time_seconds	自1970/01/01以来的开始时间（以秒为单位）
		nginx_ingress_controller_nginx_process_read_bytes_total	读取的字节数
		nginx_ingress_controller_nginx_process_requests_total	自启动以来，Nginx处理的请求总数
		nginx_ingress_controller_nginx_process_resident_memory_bytes	进程的常驻内存使用量，即实际占用的物理内存大小
		nginx_ingress_controller_nginx_process_virtual_memory_bytes	进程的虚拟内存使用量，即分配给进程的总内存大小，包括实际物理内存和虚拟交换空间
		nginx_ingress_controller_nginx_process_write_bytes_total	Nginx进程向磁盘或其他长期存储设备写入的数据量
		nginx_ingress_controller_build_info	记录Nginx Ingress Controller的构建信息，包括版本号、编译时间等
		nginx_ingress_controller_check_success	Nginx Ingress Controller的健康检查结果。1为正常，0为异常

Target名称	Job名称	指标	指标含义
		nginx_ingress_controller_config_hash	运行配置的哈希值
		nginx_ingress_controller_config_last_reloaded_successful	Nginx Ingress Controller的配置重新加载是否成功
		nginx_ingress_controller_config_last_reloaded_successful_timestamp_seconds	Nginx Ingress Controller最后一次配置重新加载是否成功的时间戳
		nginx_ingress_controller_ssl_certificate_info	Nginx Ingress Controller的证书信息
		nginx_ingress_controller_success	Nginx Ingress Controller的重新加载操作的累计次数
		nginx_ingress_controller_orphan_ingress	入口孤立状态，1 表示孤立。0 表示正常。 “namespace”为标识入口命名空间的字符串，“ingress”表示入口名称，“type”表示孤立类型为“no-service”或“no-endpoint”。
		nginx_ingress_controller_admission_config_size	Admission准入控制器的配置大小
		nginx_ingress_controller_admission_rendertime_duration	Admission准入控制器配置渲染所花费的时间
		nginx_ingress_controller_admission_rendertime_ingresses	Admission准入控制器配置渲染的入口长度
		nginx_ingress_controller_admission_roundtrip_duration	Admission准入控制器处理新事件所花费的时间
		nginx_ingress_controller_admission_tested_duration	Admission准入控制器测试所花费的时间
		nginx_ingress_controller_admission_tested_ingresses	Admission准入控制器测试的入口长度

8.4 基础指标：Modelarts 指标

介绍Modelarts通过Agent上报到AOM的指标。

表 8-4 Modelarts 通过 Agent 上报到 AOM 的指标

指标类别	指标	指标名称	指标含义	取值范围	单位
CPU	ma_container_cpu_util	CPU 使用率	该指标用于统计测量对象的CPU使用率。	0~100	百分比 (%)
	ma_container_cpu_used_core	CPU 内核占用量	该指标用于统计测量对象已经使用的CPU核个数	≥0	核 (Core)
	ma_container_cpu_limit_core	CPU 内核总量	该指标用于统计测量对象申请的CPU核总量。	≥1	核 (Core)
内存	ma_container_memory_capacity_megabytes	内存总量	该指标用于统计测量对象申请的物理内存总量。	≥0	兆字节 (MB)
	ma_container_memory_util	物理内存使用率	该指标用于统计测量对象已使用内存占申请物理内存总量的百分比。	0~100	百分比 (%)
	ma_container_memory_used_megabytes	物理内存使用量	该指标用于统计测量对象实际已经使用的物理内存(对应 container_memory_working_set_bytes当前内存工作集 (working set) 使用量。(工作区内存使用量=活跃的匿名与和缓存, 以及file-baked页 <=container_memory_usage_bytes))	≥0	兆字节 (MB)
存储IO	ma_container_disk_read_kilobytes	磁盘读取速率	该指标用于统计每秒从磁盘读出的数据量。	≥0	千字节/秒 (kB/s)
	ma_container_disk_write_kilobytes	磁盘写入速率	该指标用于统计每秒写入磁盘的数据量。	≥0	千字节/秒 (kB/s)

指标类别	指标	指标名称	指标含义	取值范围	单位
GPU显存	ma_container_gpu_mem_total_megabytes	显存容量	该指标用于统计训练任务的显存容量。	>0	兆字节 (MB)
	ma_container_gpu_mem_util	显存使用率	该指标用于统计测量对象已使用的显存占显存容量的百分比。	0~100	百分比 (%)
	ma_container_gpu_mem_used_megabytes	显存使用量	该指标用于统计测量对象已使用的显存。	≥0	兆字节 (MB)
GPU	ma_container_gpu_util	GPU使用率	该指标用于统计测量对象的GPU使用率。	0~100	百分比 (%)
	ma_container_gpu_mem_copy_util	GPU内存带宽利用率	表示内存带宽利用率。以英伟达GPU V100为例，其最大内存带宽为900 GB/sec，如果当前的内存带宽为450 GB/sec，则内存带宽利用率为50%。	0~100	百分比 (%)
	ma_container_gpu_enc_util	GPU编码器利用率	表示编码器利用率	0~100	百分比 (%)
	ma_container_gpu_dec_util	GPU解码器利用率	表示解码器利用率	0~100	百分比 (%)
	DCGM_FI_DEV_GPU_TEMP	GPU温度	表示GPU温度。	>0	摄氏度 (°C)
	DCGM_FI_DEV_POWER_USAGE	GPU功率	表示功率。	>0	瓦 (W)
	DCGM_FI_DEV_MEMORY_TEMP	内存温度	表示内存温度。	>0	摄氏度 (°C)

指标类别	指标	指标名称	指标含义	取值范围	单位
	DCGM_FI_PROF_GR_ENGINE_ACTIVE	Graphics Engine Activity	表示在一个时间间隔内，Graphics或Compute引擎处于Active的时间占比。该值表示所有Graphics和Compute引擎的平均值。Graphics或Compute引擎处于Active是指Graphics或Compute Context绑定到线程，并且Graphics或Compute Context处于Busy状态。	0~1.0	占比 (Fraction)
	DCGM_FI_PROF_SM_OCCUPANCY	SM Occupancy	<p>表示在一个时间间隔内，驻留在SM上的线程束与该SM最大可驻留线程束的比例。</p> <p>该值表示一个时间间隔内的所有SM的平均值。</p> <p>占用率越高不代表GPU使用率越高。只有在GPU内存带宽受限的工作负载 (DCGM_FI_PROF_DRAM_ACTIVE) 情况下，更高的占用率表示更有效的GPU使用率。</p>	0~1.0	占比 (Fraction)

指标类别	指标	指标名称	指标含义	取值范围	单位
	DCGM_FL_PROF_PIPE_TENSOR_ACTIVE	Tensor Activity	<p>表示Tensor (HMMA/IMMA) Pipe处于Active状态的周期分数。</p> <p>该值表示一个时间间隔内的平均值，而不是瞬时值。</p> <p>较高的值表示Tensor Cores的利用率较高。</p> <p>该值为1 (100%) 表示在整个时间间隔内每隔一个指令周期发出一个Tensor指令 (两个周期完成一条指令)。</p> <p>假设该值为0.2 (20%)，可能有如下情况：</p> <p>在整个时间间隔内，有20%的SM的Tensor Core以100%的利用率运行。</p> <p>在整个时间间隔内，有100%的SM的Tensor Core以20%的利用率运行。</p> <p>在整个时间间隔的1/5时间内，有100%的SM上的Tensor Core以100%利用率运行。</p> <p>其他组合模式。</p>	0~1.0	占比 (Fraction)

指标类别	指标	指标名称	指标含义	取值范围	单位
	DCGM_FI_PROF_DRAM_ACTIVE	Memory BW Utilization	<p>表示在一个时间间隔内，向设备内存发送数据或者从设备内存接收数据的时间占比。</p> <p>该值表示一个时间间隔内的平均值，而不是瞬时值。</p> <p>较高的值表示设备内存的利用率较高。</p> <p>该值为1（100%）表示在整个时间间隔内每个周期执行一次DRAM指令（实际上，约0.8(80%)的峰值是可达到的最大峰值)。</p> <p>假设该值为0.2 (20%)表示在时间间隔内，有20%的周期从设备内存读取或写入设备内存。</p>	0~1.0	占比 (Fraction)

指标类别	指标	指标名称	指标含义	取值范围	单位
	DCGM_FL_PROF_PIPE_FP16_ACTIVE	FP16 Engine Activity	<p>表示FP16（半精度）管道处于Active的周期分数。</p> <p>该值表示一个时间间隔内的平均值，而不是瞬时值。</p> <p>较高的值代表FP16 Cores有较高的利用率。</p> <p>该值为 1 (100%) 表示在整个时间间隔内上每两个周期（Volta类型卡为例）执行一次FP16指令。</p> <p>假设该值为0.2（20%），可能有如下情况：</p> <p>在整个时间间隔内，有20%的SM的FP16 Core以100%的利用率运行。</p> <p>在整个时间间隔内，有100%的SM的FP16 Core以20%的利用率运行。</p> <p>在整个时间间隔的1/5时间内，有100%的SM上的FP16 Core以100%利用率运行。</p> <p>其他组合模式。</p>	0~1.0	占比（Fraction）

指标类别	指标	指标名称	指标含义	取值范围	单位
	DCGM_FL_PROF_PIPE_FP32_ACTIVE	FP32 Engine Activity	<p>表示乘加操作FMA (Fused Multiply-Add) 管道处于Active的周期分数，乘加操作包括FP32 (单精度) 和整数。</p> <p>该值表示一个时间间隔内的平均值，而不是瞬时值。</p> <p>较高的值代表FP32 Cores有较高的利用率。</p> <p>该值为1 (100%) 表示在整个时间间隔内上每两个周期 (Volta类型卡为例) 执行一次FP32指令。</p> <p>假设该值为0.2 (20%) ，可能有如下情况：</p> <p>在整个时间间隔内，有20%的SM的FP32 Core以100%的利用率运行。</p> <p>在整个时间间隔内，有100%的SM的FP32 Core以20%的利用率运行。</p> <p>在整个时间间隔的1/5时间内，有100%的SM上的FP32 Core以100%利用率运行。</p> <p>其他组合模式。</p>	0~1.0	占比 (Fraction)

指标类别	指标	指标名称	指标含义	取值范围	单位
	DCGM_FL_PROF_PIPE_FP64_ACTIVE	FP64 Engine Activity	<p>表示FP64（双精度）Pipe处于Active状态的周期分数。</p> <p>该值表示一个时间间隔内的平均值，而不是瞬时值。</p> <p>较高的值代表FP64 Cores有较高的利用率。</p> <p>该值为 1（100%）表示在整个时间间隔内上每四个周期（以Volta类型卡为例）执行一次FP64指令。</p> <p>假设该值为0.2（20%），可能有如下情况：</p> <p>在整个时间间隔内，有20%的SM的FP64 Core以100%的利用率运行。</p> <p>在整个时间间隔内，有100%的SM的FP64 Core以20%的利用率运行。</p> <p>在整个时间间隔的1/5时间内，有100%的SM上的FP64 Core以100%利用率运行。</p> <p>其他组合模式。</p>	0~1.0	占比（Fraction）

指标类别	指标	指标名称	指标含义	取值范围	单位
	DCGM_FL_PROF_SM_ACTIVE	SM Activity	<p>表示在一个时间间隔内，至少一个线程束在一个SM（Streaming Multiprocessor）上处于Active的时间占比。</p> <p>该值表示所有SM的平均值，且该值对每个块的线程数不敏感。</p> <p>线程束处于Active是指一个线程束被调度且分配资源后的状态，可能是在Computing、也可能是非Computing状态（例如等待内存请求）。</p> <p>该值小于0.5表示未高效利用GPU，大于0.8是必要的。</p> <p>假设一个GPU有N个SM： 一个核函数在整个时间间隔内使用N个线程块运行在所有的SM上，此时该值为1（100%）。 一个核函数在一个时间间隔内运行N/5个线程块，此时该值为0.2。 一个核函数使用N个线程块，在一个时间间隔内，仅运行了1/5个周期的时间，此时该值为0.2。</p>	0~1.0	占比（Fraction）

指标类别	指标	指标名称	指标含义	取值范围	单位
	DCGM_FI_PROF_PCIE_TX_BYTES DCGM_FI_PROF_PCIE_RX_BYTES	PCIe Bandwidth	表示通过PCIe总线传输/接收的数据速率，包括协议标头和数据有效负载。 该值表示一个时间间隔内的平均值，而不是瞬时值。 该速率在时间间隔内平均。例如，在1秒内传输1 GB数据，则无论以恒定速率还是突发传输数据，速率都是1 GB/s。理论上的最大PCIe Gen3带宽为每通道985 MB/s。	≥0	字节/秒 (Byte/s)
	DCGM_FI_PROF_NVLINK_RX_BYTES DCGM_FI_PROF_NVLINK_TX_BYTES	NVLink Bandwidth	表示通过NVLink传输/接收的数据速率，不包括协议标头。 该值表示一个时间间隔内的平均值，而不是瞬时值。 该速率在时间间隔内平均。例如，在1秒内传输1 GB数据，则无论以恒定速率还是突发传输数据，速率都是1 GB/s。理论上，最大NVLink Gen2带宽为每个方向每个链路25 GB/s。	≥0	字节/秒 (Byte/s)
网络IO	ma_container_network_receive_bytes	下行 Bps	该指标用于统计测试对象的入方向网络流速。	≥0	字节/秒 (Byte/s)
	ma_container_network_receive_packets	下行 Pps	每秒网卡接收的数据包个数。	≥0	个/秒 (Packet/s)
	ma_container_network_receive_error_packets	下行 错包率	每秒网卡接收的错误包个数。	≥0	个/秒 (Count/s)
	ma_container_network_transmit_bytes	上行 Bps	该指标用于统计测试对象的出方向网络流速。	≥0	字节/秒 (Byte/s)

指标类别	指标	指标名称	指标含义	取值范围	单位
	ma_container_network_transmit_error_packets	上行错包率	每秒网卡发送的错误包个数。	≥0	个/秒 (Count/s)
	ma_container_network_transmit_packets	上行Pps	每秒网卡发送的数据包个数。	≥0	个/秒 (Packets/s)
NPU	ma_container_npu_util	NPU使用率	该指标用于统计测量对象的NPU使用率。	0~100	百分比 (%)
	ma_container_npu_memory_util	NPU显存使用率	该指标用于统计测量对象已使用的NPU显存占NPU存储容量的百分比。	0~100	百分比 (%)
	ma_container_npu_memory_used_megabytes	NPU显存使用量	该指标用于统计测量对象已使用的NPU显存。	≥0	兆字节 (MB)
	ma_container_npu_memory_total_megabytes	NPU显存容量	该指标用于统计测量对象的NPU显存容量。	≥0	兆字节 (MB)

8.5 基础指标：IEF 指标

介绍IEF服务上报到AOM的监控指标，包括指标的类别、名称、含义等信息。

说明

IEF服务的指标上报到AOM后，系统会根据指标映射规则将IEF指标转换后，呈现在AOM控制台的“指标浏览”界面。

表 8-5 IEF 服务监控指标

指标类别	分类	AOM呈现的指标	IEF上报的指标	指标名称	指标含义	取值范围	单位
主机指标	CPU	aom_node_cpu_limit_core	cpuCoreLimit	CPU内核总量	该指标用于统计测量对象申请的CPU核总量。	≥1	核 (Core)

指标类别	分类	AOM呈现的指标	IEF上报的指标	指标名称	指标含义	取值范围	单位
		aom_node_cpu_used_core	cpuCoreUsed	CPU内核占用量	该指标用于统计测量对象已经使用的CPU核个数。	≥0	核 (Core)
		aom_node_cpu_usage	cpuUsage	CPU使用率	该指标用于统计测量对象的CPU使用率。	0 ~ 100	百分比 (%)
	内存	aom_node_memory_total_megabytes	totalMem	物理内存容量	该指标用于统计测量对象申请的物理内存总量。	≥0	兆字节 (MB)
		aom_node_memory_free_megabytes	freeMem	可用物理内存	该指标用于统计测量对象上的尚未被使用的物理内存。	≥0	兆字节 (MB)
		aom_node_memory_usage	memUsedRate	物理内存使用率	该指标用于统计测量对象已使用内存占申请物理内存总量的百分比。	0 ~ 100	百分比 (%)
		aom_node_virtual_memory_usage	virMemUsedRate	虚拟内存使用率	该指标用于统计测量对象已使用虚拟内存占虚拟内存总量的百分比。	≥0	兆字节 (MB)
	网络	aom_node_network_receive_bytes	recvBytesRate	下行 Bps	该指标用于统计测试对象的入方向网络流速。	≥0	字节/秒 (Byte/s)
		aom_node_network_transmit_bytes	sendBytesRate	上行 Bps	该指标用于统计测试对象的出方向网络流速。	≥0	字节/秒 (Byte/s)
	磁盘	aom_node_disk_capacity_megabytes	diskCapacity	磁盘空间容量	总的磁盘空间容量。	≥0	兆字节 (MB)

指标类别	分类	AOM呈现的指标	IEF上报的指标	指标名称	指标含义	取值范围	单位
		aom_node_disk_available_capacity_megabytes	diskAvailableCapacity	可用磁盘空间	还未经使用的磁盘空间。	≥0	兆字节 (MB)
		aom_node_disk_usage	diskUsageRate	磁盘使用率	已使用的磁盘空间占总的磁盘空间容量百分比。	0 ~ 100	百分比 (%)
		aom_node_disk_read_kilobytes	diskReadRate	磁盘读取速率	该指标用于统计每秒从磁盘读出的数据量。	≥0	千字节/秒 (kB/s)
		aom_node_disk_write_kilobytes	diskWriteRate	磁盘写入速率	该指标用于统计每秒写入磁盘的数据量。	≥0	千字节/秒 (kB/s)
	GPU	aom_node_gpu_memory_free_megabytes	gpuMemoryCapacity	显存容量	该指标用于统计测量对象的显存容量。	≥0	兆字节 (MB)
		aom_node_gpu_memory_usage	gpuMemoryUsage	显存使用率	该指标用于统计测量对象已使用的显存占显存容量的百分比。	0 ~ 100	百分比 (%)
		aom_node_gpu_memory_used_megabytes	gpuMemoryUsed	显存使用量	该指标用于统计测量对象已使用的显存。	≥0	兆字节 (MB)
		aom_node_gpu_usage	gpuUtil	GPU使用率	该指标用于统计测量对象显卡使用率。	0 ~ 100	百分比 (%)
	主机	aom_node_process_number	processNum	进程数量	该指标用于统计测量对象上正在运行的进程个数。	≥0	无

指标类别	分类	AOM呈现的指标	IEF上报的指标	指标名称	指标含义	取值范围	单位
	Atlas 500 智能小站	aom_node_npu_temperature_centigrade	node_temperature	节点温度	Atlas 500小站节点温度，由用户调用edgecore接口上报。	≥0	摄氏度（℃）
		node_power	node_power	节点功率	Atlas 500小站节点功率，由用户调用edgecore接口上报。	≥0	瓦（W）
		node_voltage	node_voltage	节点电压	Atlas 500小站节点电压，由用户调用edgecore接口上报。	≥0	伏（V）
		npu_temperature	npu_temperature	芯片温度	Atlas 500小站节点NPU卡温度，由用户调用edgecore接口上报。	≥0	摄氏度（℃）
		npu_health	npu_health	芯片健康状态	Atlas 500小站节点NPU卡的健康状态，由用户调用edgecore接口上报。	≥0	无
		ai_cpu_rate	ai_cpu_rate	AI cpu 占用率	昇腾AI加速卡的AI CPU使用率，由用户调用edgecore接口上报。	0 ~ 100	百分比（%）
		ai_core_rate	ai_core_rate	AI core 占用率	昇腾AI加速卡的AI Core使用率，由用户调用edgecore接口上报。	0 ~ 100	百分比（%）
		ctrl_cpu_rate	ctrl_cpu_rate	控制cpu 占用率	昇腾AI加速卡在控制CPU使用率，由用户调用edgecore接口上报。	0 ~ 100	百分比（%）
		ddr_cap_rate	ddr_cap_rate	ddr 内存占用率	Atlas 500小站节点的ddr内存占用率，由用户调用edgecore接口上报。	0 ~ 100	百分比（%）
		ddr_bw_rate	ddr_bw_rate	ddr 带宽占用率	Atlas 500小站节点的ddr带宽占用率，由用户调用edgecore接口上报。	0 ~ 100	百分比（%）

指标类别	分类	AOM呈现的指标	IEF上报的指标	指标名称	指标含义	取值范围	单位
容器指标	CPU	aom_container_cpu_limit_core	cpuCoreLimit	CPU内核总量	该指标用于统计测量对象申请的CPU核总量。	≥1	核 (Core)
		aom_container_cpu_used_core	cpuCoreUsed	CPU内核占用量	该指标用于统计测量对象已经使用的CPU核个数。	≥0	核 (Core)
		aom_container_cpu_usage	cpuUsage	CPU使用率	该指标用于统计测量对象的CPU使用率。	0 ~ 100	百分比 (%)
	内存	aom_container_memory_request_megabytes	memCapacity	物理内存容量	该指标用于统计测量对象申请的物理内存总量。	≥0	兆字节 (MB)
		aom_container_memory_used_megabytes	memUsed	物理内存使用量	该指标用于统计测量对象上已使用的物理内存。	≥0	兆字节 (MB)
		memUsedRate	memUsedRate	物理内存使用率	该指标用于统计测量对象已使用内存占申请物理内存总量的百分比。	0 ~ 100	百分比 (%)
	磁盘	aom_container_disk_read_kilobytes	diskReadRate	磁盘读取速率	该指标用于统计每秒从磁盘读出的数据量。	≥0	千字节/秒 (kB/s)
		aom_container_disk_write_kilobytes	diskWriteRate	磁盘写入速率	该指标用于统计每秒写入磁盘的数据量。	≥0	千字节/秒 (kB/s)
	网络	aom_container_network_receive_bytes	recvBytesRate	下行Bps	该指标用于统计测试对象的入方向网络流速。	≥0	字节/秒 (Byte/s)

指标类别	分类	AOM呈现的指标	IEF上报的指标	指标名称	指标含义	取值范围	单位
		aom_container_network_transmit_bytes	sendBytesRate	上行Bps	该指标用于统计测试对象的出方向网络流速。	≥0	字节/秒 (Byte/s)
	GPU	aom_container_gpu_memory_free_megabytes	gpuMemCapacity	显存容量	该指标用于统计测量对象的显存容量。	≥0	兆字节 (MB)
		aom_container_gpu_memory_usage	gpuMemUsage	显存使用率	该指标用于统计测量对象已使用的显存占显存容量的百分比。	0 ~ 100	百分比 (%)
		aom_container_gpu_memory_used_megabytes	gpuMemUsed	显存使用量	该指标用于统计测量对象已使用的显存。	≥0	兆字节 (MB)
		aom_container_gpu_usage	gpuUtil	GPU使用率	该指标用于统计测量对象显卡使用率。	0 ~ 100	百分比 (%)
容器状态	aom_container_status	status	容器状态	该指标用于统计容器的状态。	≥0	无	
进程指标	CPU	aom_process_cpu_usage	cpuUsage	CPU使用率	该指标用于统计测量对象的CPU使用率。	0 ~ 100	百分比 (%)
	内存	aom_process_memory_used_megabytes	memUsed	物理内存使用量	该指标用于统计测量对象上已使用的物理内存。	≥0	兆字节 (MB)
	进程状态	aom_process_status	status	进程状态	该指标用于统进程的状态。	≥0	无

指标类别	分类	AOM呈现的指标	IEF上报的指标	指标名称	指标含义	取值范围	单位
	GPU	gpuMemCapacity	gpuMemCapacity	显存容量	该指标用于统计测量对象的显存容量。	≥0	兆字节 (MB)
		gpuMemUsage	gpuMemUsage	显存使用率	该指标用于统计测量对象已使用的显存占显存容量的百分比。	0 ~ 100	百分比 (%)
		gpuMemUsed	gpuMemUsed	显存使用量	该指标用于统计测量对象已使用的显存。	≥0	兆字节 (MB)
		gpuUtil	gpuUtil	GPU使用率	该指标用于统计测量对象显卡使用率。	0 ~ 100	百分比 (%)

8.6 基础指标：CSE 指标

介绍CSE服务上报到AOM的监控指标，包括指标的类别、名称、含义等信息。

表 8-6 CSE 服务监控指标

指标类别	分类	指标	指标名称	指标含义	取值范围	单位
ServiceComb	registry	servicecomb_service_center_db_service_total	微服务版本数	微服务版本数。	≥0	个
		servicecomb_service_center_db_instance_total	微服务实例数	微服务实例数。	≥0	个
		servicecomb_service_center_http_request_total	http请求次数	http请求次数，包括多种url、方法、code。	≥0	次

指标类别	分类	指标	指标名称	指标含义	取值范围	单位
		servicecomb_service_center_http_request_durations_microseconds	http请求总耗时	http请求总耗时，包括多种url、方法、code。	≥0	微秒 (μs)
	config	servicecomb_kie_request_count	http请求次数	http请求次数，包括多种url、方法、code。	≥0	次
		servicecomb_kie_request_process_duration	http请求总耗时	http请求总耗时，包括多种url、方法、code。	≥0	毫秒 (ms)
		servicecomb_kie_config_count	配置个数	ServiceComb配置个数。	≥0	个
Nacos	config	nacos_config_Count	Nacos config配置个数	Nacos集群内各节点的配置个数。	≥0	个
		nacos_getConfig	Nacos config读配置请求数	Nacos集群内各节点的读配置请求次数。	≥0	次
		nacos_longPolling	Nacos config HTTP长连接数 (监听者数)	Nacos config HTTP长连接数。	≥0	个
		nacos_publish	Nacos config写配置请求数	Nacos集群内各节点的写配置请求次数。	≥0	次
		nacos_subscriberCount	Nacos config订阅者数量	Nacos config订阅者数量。	≥0	个
		nacos_configPushCost	Nacos config推送时延	Nacos config推送时延。	≥0	毫秒 (ms)

指标类别	分类	指标	指标名称	指标含义	取值范围	单位
	http	nacos_http_server_requests_seconds_count	http请求次数	http请求次数, 包括多种url、方法、code。	≥0	次
		nacos_http_server_requests_seconds_max	http请求总耗时最大值	统计周期内http请求时延的最大值, 包括多种url、方法、code, 使用Nacos-Client 1.x时上报。	≥0	秒 (s)
		nacos_http_server_requests_seconds_sum	http请求总耗时	http请求总耗时, 包括多种url、方法、code。	≥0	秒 (s)
	naming	nacos_avgPushCost	Nacos naming 平均推送耗时	Nacos naming 平均推送耗时 (ms)。	≥0	毫秒 (ms)
		nacos_maxPushCost	Nacos naming 最大推送耗时	Nacos naming 最大推送耗时 (ms)。	≥0	毫秒 (ms)
		nacos_failedPush	Nacos naming 推送失败数	Nacos naming 推送失败数。	≥0	次
		nacos_ipCount	Nacos naming ip个数	微服务注册实例数。	≥0	个
		nacos_serviceSubscriberCount	Nacos naming 订阅者数量	Nacos naming 订阅者数量。	≥0	个
		nacos_serviceCount	Nacos naming 域名个数 (2.x 版本)	Nacos集群内各节点的服务个数。	≥0	个

指标类别	分类	指标	指标名称	指标含义	取值范围	单位
应用网关	envoy	cpuUsage	CPU使用率	该指标用于统计测量对象的CPU使用率。	0~100	百分比(%)
		envoy_http_downstream_cx_active	活跃连接总数	该指标用于统计活跃连接数。	≥0	个
		downstream_cx_delayed_close_timeout	延迟关闭连接总数	该指标用于统计延迟关闭的连接数。	≥0	个
		envoy_http_downstream_cx_destroy	摧毁连接总数	该指标用于统计摧毁的连接数。	≥0	个
		envoy_http_downstream_cx_destroy_active_rq	摧毁活跃连接数	该指标用于统计摧毁的活跃连接数。	≥0	个
		envoy_http_downstream_cx_destroy_local	摧毁本地的连接数	该指标用于统计摧毁的本地连接数。	≥0	个
		envoy_http_downstream_cx_destroy_local_active_rq	摧毁本地活跃连接数	该指标用于统计摧毁本地活跃连接数。	≥0	个
		envoy_http_downstream_cx_destroy_remote	摧毁远端关闭的连接数	该指标用于统计因为远端关闭而摧毁的连接数。	≥0	个
		envoy_http_downstream_cx_destroy_remote_active_rq	摧毁远端关闭的活跃连接数	该指标用于统计因为远端关闭而摧毁的活跃连接数。	≥0	个
		envoy_http_downstream_cx_drain_close	驱逐关闭连接数	该指标用于统计因为驱逐而关闭的连接数。	≥0	个
		envoy_http_downstream_cx_http1_active	http1获取连接数	该指标用于统计http1的连接数。	≥0	个

指标类别	分类	指标	指标名称	指标含义	取值范围	单位
		envoy_http_downstream_cx_max_duration_reached	超时连接数	该指标用于统计因超过最大连接时间而关闭的连接数。	≥0	个
		envoy_http_downstream_cx_tx_bytes_total	总发送字节数	该指标用于统计总发送的字节数。	≥0	字节 (Byte)
		envoy_http_downstream_rq	总请求数	该指标用于统计总请求数。	≥0	个
		envoy_http_downstream_rq_http1_total	http1总请求数	该指标用于统计http1总请求数。	≥0	个
		envoy_http_downstream_rq_http2_total	http2总请求数	该指标用于统计http2总请求数。	≥0	个
		envoy_http_downstream_rq_idle_timeout	空闲时间超时关闭请求数	该指标用于统计因空闲时间超时的请求数。	≥0	个
		envoy_http_downstream_rq_too_large	请求体过大请求数	该指标用于统计请求body过大返回413的请求数。	≥0	个
		downstream_rq_ws_on_no_n_ws_route	websocket没有路由的请求数	该指标用于统计因为没有路由而被拒绝的请求数。	≥0	个
		envoy_http_local_rate_limiter_http_local_rate_limit_enforced	限流请求数	该指标用于统计被限流的请求数。	≥0	个
		envoy_cluster_circuit_breakers_default_cx_open	连接断路器触发状态	0: 连接断路器低于其并发限制。 1: 连接断路器已达到容量限制不再接纳。	0、1	无

指标类别	分类	指标	指标名称	指标含义	取值范围	单位
		envoy_cluster_circuit_breakers_high_cx_open				
		envoy_cluster_circuit_breakers_default_cx_pool_open	连接池断路器触发状态	0: 连接池断路器低于其并发限制。 1: 连接池断路器已达容量限制不再接纳。	0、1	无
		envoy_cluster_circuit_breakers_high_cx_pool_open				
		envoy_cluster_circuit_breakers_default_remaining_cx	连接断路器的剩余连接数	断路器达到并发限制之前的剩余连接数。	≥0	个
		envoy_cluster_circuit_breakers_high_remaining_cx				
		envoy_cluster_circuit_breakers_default_remaining_cx_pools	连接池断路器的剩余连接数	断路器连接池达到并发限制之前的剩余连接数。	≥0	个
		envoy_cluster_circuit_breakers_high_remaining_cx_pools				
		envoy_cluster_circuit_breakers_default_remaining_pending	断路器达到并发限制之前剩余的待处理请求数	断路器达到并发限制之前剩余的待处理请求数。	≥0	个
		envoy_cluster_circuit_breakers_high_remaining_pending				

指标类别	分类	指标	指标名称	指标含义	取值范围	单位
		envoy_cluster_circuit_breakers_default_remaining_retries	断路器达到并发限制之前的剩余重试次数	断路器达到并发限制之前的剩余重试次数。	≥0	次
		envoy_cluster_circuit_breakers_high_remaining_retries				
		envoy_cluster_circuit_breakers_default_remaining_rq	断路器达到并发限制之前剩余的请求数	断路器达到并发限制之前剩余的请求数。	≥0	个
		envoy_cluster_circuit_breakers_high_remaining_rq				
		envoy_cluster_circuit_breakers_default_rq_open	请求断路器触发状态	0: 请求断路器低于其并发限制。 1: 请求断路器已达容量限制不再接纳。	0、1	无
		envoy_cluster_circuit_breakers_high_rq_open				
		envoy_cluster_circuit_breakers_default_rq_retry_open	重试断路器触发状态	0: 重试断路器低于其并发限制。 1: 重试断路器已达容量限制不再接纳。	0、1	无
		envoy_cluster_circuit_breakers_high_rq_retry_open				
		envoy_cluster_ejections_overflow	驱逐达到最大驱逐阈值的次数	主机因请求异常率达到最大驱逐比例而被限制驱逐的次数。	≥0	次
		envoy_cluster_ejections_consecutive_5xx	连续5xx驱逐次数	主机因连续返回5xx导致的驱逐次数。	≥0	次

指标类别	分类	指标	指标名称	指标含义	取值范围	单位
		envoy_cluster_ejections_detected_consecutive_5xx	连续5xx驱逐次数（即使未强制执行）	主机因连续返回5xx导致的驱逐次数（即使未强制执行）。	≥0	次
		envoy_cluster_ejections_detected_consecutive_gateway_failure	检测到的连续网关故障的驱逐次数	检测到的连续网关故障的驱逐次数（即使未强制执行）。	≥0	次
		envoy_cluster_ejections_detected_consecutive_local_origin_failure	连续本地源故障的驱逐次数	检测到的连续本地源故障导致的驱逐次数（即使未强制执行）。	≥0	次
		envoy_cluster_ejections_enforced_consecutive_local_origin_failure	连续本地源故障强制驱逐次数	主机因检测到的连续本地源故障导致的驱逐次数。	≥0	次
		envoy_cluster_ejections_detected_failure_percentage	请求失败率阈值的驱逐次数	主机因请求失败率达到失败率阈值而被驱逐的次数。	≥0	次
		envoy_cluster_ejections_detected_local_origin_failure_percentage	本地主动请求检测失败率阈值的驱逐次数	主机因本地主动请求检测失败率阈值的驱逐次数。	≥0	次
		envoy_cluster_ejections_detected_local_origin_success_rate	本地主动请求检测成功率未达阈值的驱逐次数	主机因本地主动请求检测成功率未达到阈值而触发的驱逐次数（即使未强制执行）。	≥0	次
		envoy_cluster_ejections_detected_success_rate	本地主动请求检测成功率未达阈值的驱逐次数	主机因本地主动请求检测成功率未达到阈值而触发的驱逐次数。	≥0	次

指标类别	分类	指标	指标名称	指标含义	取值范围	单位
		envoy_cluster_ejections_enforced_consecutive_5xx	连续 5xx 强制驱逐次数	主机因连续 5xx 强制驱逐次数。	≥0	次
		envoy_cluster_ejections_enforced_consecutive_gateway_failure	连续网关故障强制驱逐次数	主机因连续网关故障强制驱逐次数。	≥0	次
		envoy_cluster_ejections_enforced_failure_percentage	请求失败率超阈值而被强制驱逐次数	主机因请求失败率超阈值而被强制驱逐的次数。	≥0	次
		envoy_cluster_ejections_enforced_local_origin_failure_percentage	本地主动请求失败率超阈值而被强制驱逐次数	主机因本地主动请求失败率超过阈值而被强制驱逐的次数。	≥0	次
		envoy_cluster_ejections_enforced_local_origin_success_rate	本地主动请求成功率未达阈值驱逐次数	主机因本地主动请求成功率未达阈值而强制驱逐的次数。	≥0	次
		envoy_cluster_ejections_enforced_success_rate	请求成功率未达阈值驱逐次数	主机因请求成功率未达阈值而强制驱逐的次数。	≥0	次
		envoy_cluster_ejections_enforced_total	强制驱逐的次数	由于任何异常类型而强制驱逐的次数。	≥0	次
		envoy_http_downstream_cx_rx_bytes_total	请求收到的总字节数	请求收到的总字节数。	≥0	字节 (Byte)

8.7 基础指标：IoTDA 指标

介绍IoTDA服务上报到AOM的监控指标，包括指标的类别、名称、含义等信息。

表 8-7 IoTDA 服务监控指标

指标类别	指标	指标名称	指标含义	取值范围	单位
设备总数	iotda_device_status_onlineCount	在线设备数	统计租户在当前实例/资源空间下的在线的设备数量	≥0	个
	iotda_device_status_totalCount	设备总数	统计租户在当前实例/资源空间下的总的设备数量	≥0	个
	iotda_device_status_offlineCount	离线设备数	统计租户在当前实例/资源空间下的离线的设备数量	≥0	个
	iotda_device_status_abnormalCount	异常设备数	统计租户在当前实例/资源空间下的异常的设备数量	≥0	个
	iotda_device_status_inactiveCount	未激活设备数	统计租户在当前实例/资源空间下的未激活的设备数量	≥0	个
NB数据上报数	iotda_south_dataReport_totalCount	NB数据上报总数	统计当前实例/资源空间下NB数据上报总次数	≥0	次
	iotda_south_dataReport_successCount	NB数据上报成功次数	统计当前实例/资源空间下NB数据上报成功次数	≥0	次
	iotda_south_dataReport_failedCount	NB数据上报失败次数	统计当前实例/资源空间下NB数据上报失败次数	≥0	次
MQTT事件上报数	iotda_south_eventUp_totalCount	MQTT事件上报总数	统计当前实例/资源空间下MQTT事件上报总数	≥0	次
	iotda_south_eventUp_successCount	MQTT事件上报成功次数	统计当前实例/资源空间下MQTT事件上报成功次数	≥0	次
	iotda_south_eventUp_failedCount	MQTT事件上报失败次数	统计当前实例/资源空间下MQTT事件上报失败次数	≥0	次
MQTT属性上报数	iotda_south_propertiesReport_totalCount	MQTT属性上报总数	统计当前实例/资源空间下MQTT属性上报总数	≥0	次
	iotda_south_propertiesReport_successCount	MQTT属性上报成功次数	统计当前实例/资源空间下MQTT属性上报成功次数	≥0	次

指标类别	指标	指标名称	指标含义	取值范围	单位
	iotda_south_propertiesReport_failedCount	MQTT属性上报失败次数	统计当前实例/资源空间下MQTT属性上报失败次数	≥0	次
MQTT消息上报数	iotda_south_messageUp_totalCount	MQTT消息上报总数	统计当前实例/资源空间下MQTT消息上报总数	≥0	次
	iotda_south_messageUp_successCount	MQTT消息上报成功次数	统计当前实例/资源空间下MQTT消息上报成功次数	≥0	次
	iotda_south_messageUp_failedCount	MQTT消息上报失败次数	统计当前实例/资源空间下MQTT消息上报失败次数	≥0	次
AMQP流转次数	iotda_amqp_forwarding_totalCount	AMQP流转次数	统计当前实例/资源空间下AMQP流转次数	≥0	次
	iotda_amqp_forwarding_successCount	AMQP流转成功次数	统计当前实例/资源空间下AMQP流转成功次数	≥0	次
	iotda_amqp_forwarding_failedCount	AMQP流转失败次数	统计当前实例/资源空间下AMQP流转失败次数	≥0	次
Function Graph流转次数	iotda_functionGraph_forwarding_totalCount	FunctionGraph流转次数	统计当前实例/资源空间下FunctionGraph流转次数	≥0	次
	iotda_functionGraph_forwarding_successCount	FunctionGraph流转成功次数	统计当前实例/资源空间下FunctionGraph流转成功次数	≥0	次
	iotda_functionGraph_forwarding_failedCount	FunctionGraph流转失败次数	统计当前实例/资源空间下FunctionGraph流转失败次数	≥0	次
MRS Kafka流转次数	iotda_mrsKafka_forwarding_totalCount	MRS Kafka流转次数	统计当前实例/资源空间下MRS Kafka流转次数	≥0	次
	iotda_mrsKafka_forwarding_successCount	MRS Kafka流转成功次数	统计当前实例/资源空间下MRS Kafka流转成功次数	≥0	次
	iotda_mrsKafka_forwarding_failedCount	MRS Kafka流转失败次数	统计当前实例/资源空间下MRS Kafka流转失败次数	≥0	次
MQTT流转次数	iotda_mqtt_forwarding_totalCount	Mqtt流转次数	统计当前实例/资源空间下MQTT流转次数	≥0	次

指标类别	指标	指标名称	指标含义	取值范围	单位
	iotda_mqtt_forwaring_successCount	Mqtt流转成功次数	统计当前实例/资源空间下MQTT流转成功次数	≥0	次
	iotda_mqtt_forwaring_failedCount	Mqtt流转失败次数	统计当前实例/资源空间下MQTT流转失败次数	≥0	次
MySQL流转次数	iotda_mysql_forwaring_totalCount	MySQL流转次数	统计当前实例/资源空间下MySQL流转次数	≥0	次
	iotda_mysql_forwaring_successCount	MySQL流转成功次数	统计当前实例/资源空间下MySQL流转成功次数	≥0	次
	iotda_mysql_forwaring_failedCount	MySQL流转失败次数	统计当前实例/资源空间下MySQL流转失败次数	≥0	次
InfluxDB流转次数	iotda_influxDB_forwaring_totalCount	InfluxDB流转次数	统计当前实例/资源空间下InfluxDB流转次数	≥0	次
	iotda_influxDB_forwaring_successCount	InfluxDB流转成功次数	统计当前实例/资源空间下InfluxDB流转成功次数	≥0	次
	iotda_influxDB_forwaring_failedCount	InfluxDB流转失败次数	统计当前实例/资源空间下InfluxDB流转失败次数	≥0	次
HTTP推送流转次数	iotda_http_forwaring_totalCount	HTTP推送流转次数	统计当前实例/资源空间下HTTP推送流转次数	≥0	次
	iotda_http_forwaring_successCount	HTTP推送流转成功次数	统计当前实例/资源空间下HTTP推送流转成功次数	≥0	次
	iotda_http_forwaring_failedCount	HTTP推送流转失败次数	统计当前实例/资源空间下HTTP推送流转失败次数	≥0	次
OBS流转次数	iotda_obs_forwarding_totalCount	OBS流转次数	统计当前实例/资源空间下OBS流转次数	≥0	次
	iotda_obs_forwarding_successCount	OBS流转成功次数	统计当前实例/资源空间下OBS流转成功次数	≥0	次
	iotda_obs_forwarding_failedCount	OBS流转失败次数	统计当前实例/资源空间下OBS流转失败次数	≥0	次

指标类别	指标	指标名称	指标含义	取值范围	单位
DMS Kafka流转次数	iotda_dmsKafka_forwarding_totalCount	DMS Kafka流转次数	统计当前实例/资源空间下DMS Kafka流转次数	≥0	次
	iotda_dmsKafka_forwarding_successCount	DMS Kafka流转成功次数	统计当前实例/资源空间下DMS Kafka流转成功次数	≥0	次
	iotda_dmsKafka_forwarding_failedCount	DMS Kafka流转失败次数	统计当前实例/资源空间下DMS Kafka流转失败次数	≥0	次
DIS流转次数	iotda_dis_forwarding_totalCount	DIS流转次数	统计当前实例/资源空间下DIS流转次数	≥0	次
	iotda_dis_forwarding_successCount	DIS流转成功次数	统计当前实例/资源空间下DIS流转成功次数	≥0	次
	iotda_dis_forwarding_failedCount	DIS流转失败次数	统计当前实例/资源空间下DIS流转失败次数	≥0	次
ROMA流转次数	iotda_roma_forwarding_totalCount	ROMA流转次数	统计当前实例/资源空间下ROMA流转次数	≥0	次
	iotda_roma_forwarding_successCount	ROMA流转成功次数	统计当前实例/资源空间下ROMA流转成功次数	≥0	次
	iotda_roma_forwarding_failedCount	ROMA流转失败次数	统计当前实例/资源空间下ROMA流转失败次数	≥0	次
LTS流转次数	iotda_lts_forwarding_totalCount	LTS流转次数	统计当前实例/资源空间下LTS流转次数	≥0	次
	iotda_lts_forwarding_successCount	LTS流转成功次数	统计当前实例/资源空间下LTS流转成功次数	≥0	次
	iotda_lts_forwarding_failedCount	LTS流转失败次数	统计当前实例/资源空间下LTS流转失败次数	≥0	次
查询设备影子	iotda_query_shadow_totalCount	查询设备影子总数	统计当前实例/资源空间下查询设备影子总数	≥0	次
	iotda_query_shadow_successCount	查询设备影子成功数	统计当前实例/资源空间下查询设备影子成功数	≥0	次
	iotda_query_shadow_failedCount	查询设备影子失败数	统计当前实例/资源空间下查询设备影子失败数	≥0	次
消息下发	iotda_message_down_totalCount	消息下发总数	统计当前实例/资源空间下消息下发总数	≥0	次

指标类别	指标	指标名称	指标含义	取值范围	单位
	iotda_message_down_successCount	消息下发成功数	统计当前实例/资源空间下消息下发成功数	≥0	次
	iotda_message_down_failedCount	消息下发失败数	统计当前实例/资源空间下消息下发失败数	≥0	次
属性设置	iotda_properties_set_totalCount	属性设置总数	统计当前实例/资源空间下属性设置总数	≥0	次
	iotda_properties_set_successCount	属性设置成功数	统计当前实例/资源空间下属性设置成功数	≥0	次
	iotda_properties_set_failedCount	属性设置失败数	统计当前实例/资源空间下属性设置失败数	≥0	次
属性查询	iotda_properties_query_totalCount	属性查询总数	统计当前实例/资源空间下属性查询总数	≥0	次
	iotda_properties_query_successCount	属性查询成功数	统计当前实例/资源空间下属性查询成功数	≥0	次
	iotda_properties_query_failedCount	属性查询失败数	统计当前实例/资源空间下属性查询失败数	≥0	次
命令数	iotda_command_totalCount	命令总数	统计当前实例/资源空间下命令总数	≥0	次
	iotda_command_successCount	命令成功数	统计当前实例/资源空间下命令成功数	≥0	次
	iotda_command_failedCount	命令失败数	统计当前实例/资源空间下命令失败数	≥0	次
流控数据	iotda_flowcontrol_north_failedCount	API调用流控失败数	统计当前实例下调用API流控失败次数	≥0	次
	iotda_flowcontrol_south_failedCount	南向调用流控失败数	统计当前实例下数据上报流控失败次数	≥0	次
	iotda_flowcontrol_forwarding_failedCount	流转流控失败数	统计当前实例下流转流控失败数	≥0	次
同时在线设备数	iotda_max_online_devices_totalCount	每天最大同时在线设备数	统计当前实例下每天最大同时在线设备数	≥0	个
	iotda_max_online_devices_percent	最大同时在线设备数百分比	统计当前实例下每天最大同时在线设备数百分比	0~100	百分比(%)

8.8 基础指标：node-exporter 指标

介绍通过中间件Exporter上报到AOM的node-exporter指标的类别、名称、含义等信息。

表 8-8 CCE/自建 K8s 集群容器指标

Job名称	指标	指标说明
node-exporter	node_filesystem_size_bytes	挂载的文件系统占用空间
	node_filesystem_readonly	只读挂载的文件系统
	node_filesystem_free_bytes	挂载的文件系统剩余空间
	node_filesystem_avail_bytes	挂载的文件系统可用空间
	node_cpu_seconds_total	节点CPU时间
	node_network_receive_bytes_total	累积接收数据总量
	node_network_receive_errors_total	接收时遇到的错误累积计数
	node_network_transmit_bytes_total	累积传输数据总量
	node_network_receive_packets_total	接收数据包的累积计数
	node_network_transmit_drop_total	传输时丢弃的累积计数
	node_network_transmit_errors_total	传输时遇到的错误累积计数
	node_network_up	网卡的状态
	node_network_transmit_packets_total	传输数据包的累积计数
	node_network_receive_drop_total	接收时丢弃的累积计数

Job名称	指标	指标说明
	go_gc_duration_seconds	数据来自调用 debug.ReadGCStats(), 调用该函数时, 会将传入参数GCStats结构体的 PauseQuantile字段设置为5, 这样函数将会返回最小、25%、50%、75%和最大, 这5个GC暂停时间百分位数。然后prometheus Go客户端根据返回的GC暂停时间百分位数、以及NumGC和PauseTotal变量创建摘要类型指标。
	node_load5	节点5分钟CPU负载
	node_filefd_allocated	已分配的文件描述符
	node_exporter_build_info	NodeExporter构建信息
	node_disk_written_bytes_total	写入成功的字节总数
	node_disk_writes_completed_total	写入完成的次数
	node_disk_write_time_seconds_total	写入花费的总时长
	node_nf_conntrack_entries	链接状态跟踪表分配的数量
	node_nf_conntrack_entries_limit	链接状态跟踪表总量
	node_processes_max_processes	PID限制值
	node_processes_pids	PID个数
	node_sockstat_TCP_alloc	已分配的TCP套接字数量
	node_sockstat_TCP_inuse	正在使用的TCP套接字数量
	node_sockstat_TCP_tw	等待关闭的TCP连接数
	node_timex_offset_seconds	时钟时间偏移
	node_timex_sync_status	节点时钟同步状态
	node_uname_info	节点uname信息
	node_vmstat_pgfault	/proc/vmstat中的pgfault
	node_vmstat_pgmajfault	/proc/vmstat中的pgmajfault
	node_vmstat_pgpgin	/proc/vmstat中的pgpggin
	node_vmstat_ppggout	/proc/vmstat中的pgpgout

Job名称	指标	指标说明
	node_disk_reads_completed_total	读取完成的次数
	node_disk_read_time_seconds_total	读取花费的总时长
	process_cpu_seconds_total	该指标计算使用到utime（Go进程执行在用户态模式下的滴答数）和stime（Go进程执行在内核态时候的滴答数，例如系统调用时），参数的单位为jiffies，jiffy描述了两次系统定时器中断之间的滴答时间。 process_cpu_seconds_total等于utime和stime之和除以USER_HZ。即将程序滴答总数除以Hz（每秒滴答数）得到就是操作系统运行该进程的总时间（以秒为单位）。
	node_disk_read_bytes_total	读取成功的字节总数
	node_disk_io_time_weighted_seconds_total	执行I/O所花费的加权秒数
	node_disk_io_time_seconds_total	I/O总耗时
	node_disk_io_now	当前运行的I/O数量
	node_context_switches_total	上下文切换次数
	node_boot_time_seconds	节点开机时间
	process_resident_memory_bytes	即RSS（Resident Set Size），指的是常驻内存集，是进程实际使用的内存，他不包括分配但未使用的内存，也不包括换出的内存页面，但包含共享内存。
	node_intr_total	节点中断总数
	node_load1	节点1分钟CPU负载
	go_goroutines	通过runtime.NumGoroutine()调用获取，基于调度器结构sched和全局allglen变量计算得来。由于sched结构体的所有字段可能并发的更改，因此最后会检查计算的值是否小于1，如果小于1，那么返回1。
	scrape_duration_seconds	采集scrape target花费的时间
	node_load15	节点15分钟CPU负载

Job名称	指标	指标说明
	scrape_samples_post_metric_relabeling	metric被重设标签后，剩余sample数量
	node_netstat_Tcp_PassiveOpens	从LISTEN 状态直接转换到SYN-RCVD 状态的TCP连接数
	scrape_samples_scraped	scrape target暴露的sample数量
	node_netstat_Tcp_CurrEstab	当前状态为ESTABLISHED 或CLOSE-WAIT 的TCP 连接数
	scrape_series_added	scrape target新增加的系列数
	node_netstat_Tcp_ActiveOpens	从CLOSED 状态直接转换到SYN-SENT 状态的TCP连接数
	node_memory_MemTotal_bytes	节点内存总量
	node_memory_MemFree_bytes	节点空闲内存
	node_memory_MemAvailable_bytes	节点可用内存
	node_memory_Cached_bytes	节点页面缓存中的内存
	up	scrape target的状态
	node_memory_Buffers_bytes	节点缓冲区的内存

8.9 指标维度

通过 ICAgent 上报的虚机指标的指标维度

表 8-9 通过 ICAgent 上报的虚机指标的指标维度

指标类别	指标维度	说明
网络指标	clusterId	集群ID。
	hostID	主机ID。
	nameSpace	集群的命名空间。
	netDevice	网卡名称。
	nodeIP	主机IP。
	nodeName	主机名称。

指标类别	指标维度	说明
磁盘指标	clusterId	集群ID。
	diskDevice	磁盘名称。
	hostID	主机ID。
	nameSpace	集群的命名空间。
	nodeIP	主机IP。
	nodeName	主机名称。
磁盘分区指标	diskPartition	磁盘分区。
	diskPartitionType	磁盘分区类型。
文件系统指标	clusterId	集群ID。
	clusterName	集群名称。
	fileSystem	文件系统。
	hostID	主机ID。
	mountPoint	挂载点。
	nameSpace	集群的命名空间。
	nodeIP	主机IP。
	nodeName	主机名称。
主机指标	clusterId	集群ID。
	clusterName	集群名称。
	gpuName	GPU名称。
	gpuID	GPU ID。
	npuName	NPU名称。
	npuID	NPU ID。
	hostID	主机ID。
	nameSpace	集群的命名空间。
	nodeIP	主机IP。
	hostName	主机名称。
集群指标	clusterId	集群ID。
	clusterName	集群名称。
	projectId	项目ID。
容器指标	appID	服务ID。

指标类别	指标维度	说明
	appName	服务名称。
	clusterId	集群ID。
	clusterName	集群名称。
	containerID	容器ID。
	containerName	容器名称。
	deploymentName	工作负载名称。
	kind	应用类型。
	nameSpace	集群的命名空间。
	podID	实例ID。
	podIP	实例IP。
	podName	实例名称。
	serviceID	存量ID。
	nodename	主机名称。
	nodeIP	主机IP。
	virtualServiceName	Istio虚拟服务名。
	gpuID	GPU ID。
	npuName	NPU名称。
	npuID	NPU ID。
进程指标	appName	服务名称。
	clusterId	集群ID。
	clusterName	集群名称。
	nameSpace	集群的命名空间。
	processID	进程ID。
	processName	进程名称。
	serviceID	存量ID。

9 安全

9.1 责任共担

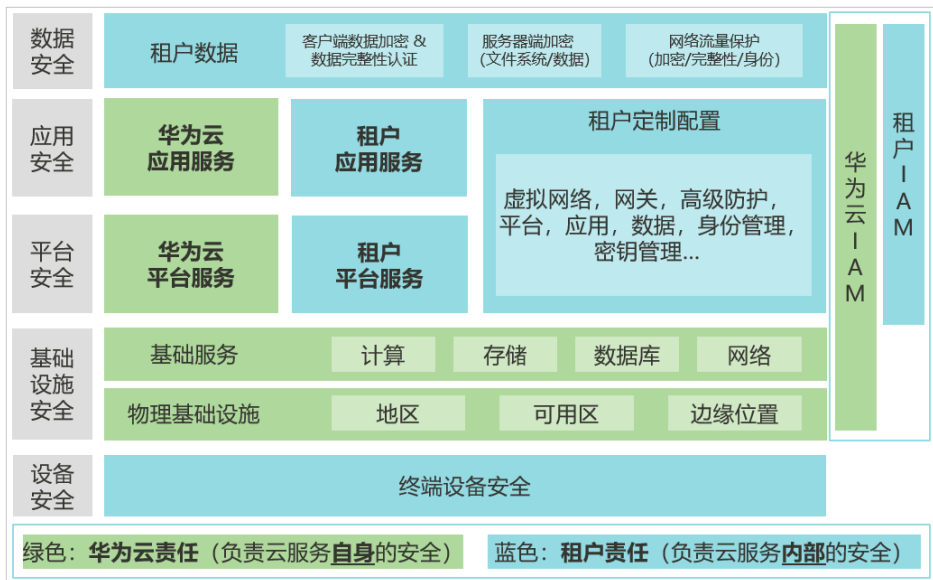
华为云秉承“将对网络和业务安全性保障的责任置于公司的商业利益之上”。针对层出不穷的云安全挑战和无孔不入的云安全威胁与攻击，华为云在遵从法律法规业界标准的基础上，以安全生态圈为护城河，依托华为独有的软硬件优势，构建面向不同区域和行业的完善云服务安全保障体系。

安全性是华为云与您的共同责任，如图9-1所示。

- **华为云**：负责云服务自身的安全，提供安全的云。华为云的安全责任在于保障其所提供的 IaaS、PaaS 和 SaaS 类云服务自身的安全，涵盖华为云数据中心的物理环境设施和运行其上的基础服务、平台服务、应用服务等。这不仅包括华为云基础设施和各项云服务技术的安全功能和性能本身，也包括运维运营安全，以及更广义的安全合规遵从。
- **租户**：负责云服务内部的安全，安全地使用云。华为云租户的安全责任在于对使用的 IaaS、PaaS 和 SaaS 类云服务内部的安全以及对租户定制配置进行安全有效的管理，包括但不限于虚拟网络、虚拟主机和访客虚拟机的操作系统，虚拟防火墙、API 网关和高级安全服务，各项云服务，租户数据，以及身份账号和密钥管理等方面的安全配置。

《[华为云安全白皮书](#)》详细介绍华为云安全性的构建思路与措施，包括云安全战略、责任共担模型、合规与隐私、安全组织与人员、基础设施安全、租户服务与租户安全、工程安全、运维运营安全、生态安全。

图 9-1 华为云安全责任共担模型



9.2 身份认证与访问控制

9.2.1 服务的访问控制

身份认证

无论用户通过AOM控制台还是API访问AOM，都会要求访问请求方出示身份凭证，并进行身份合法性校验，同时提供登录保护和登录验证策略加固身份认证安全。AOM服务基于统一身份认证服务（IAM），支持三种身份认证方式：[用户名密码](#)、[访问密钥](#)、[临时访问密钥](#)。同时还提供[登录保护](#)及[登录验证策略](#)。

访问控制

对企业中的员工设置不同的AOM访问权限，以达到不同员工之间的权限隔离，使用统一身份认证服务（Identity and Access Management，简称IAM）进行精细的权限管理。该服务提供用户身份认证、权限分配、访问控制等功能，可以帮助您安全的控制华为云资源的访问。详情请参考[AOM权限管理](#)。

9.3 数据保护技术

AOM通过多种数据保护手段和特性，保障AOM的数据安全可靠。

表 9-1 AOM 的数据保护手段和特性

数据保护手段	简要说明	详细介绍
传输加密 (HTTPS)	AOM支持HTTPS传输协议，保证数据传输的安全性。	构造请求

数据保护手段	简要说明	详细介绍
数据冗余存储	指标、告警、配置等数据以多副本方式存储，保障数据可靠性。	/
数据订阅	AOM支持客户配置数据订阅，把客户的指标和告警数据发送到客户指定的DMS实例中，客户可以自由处理转储出来的这部分数据。	数据订阅

9.4 审计与日志

审计

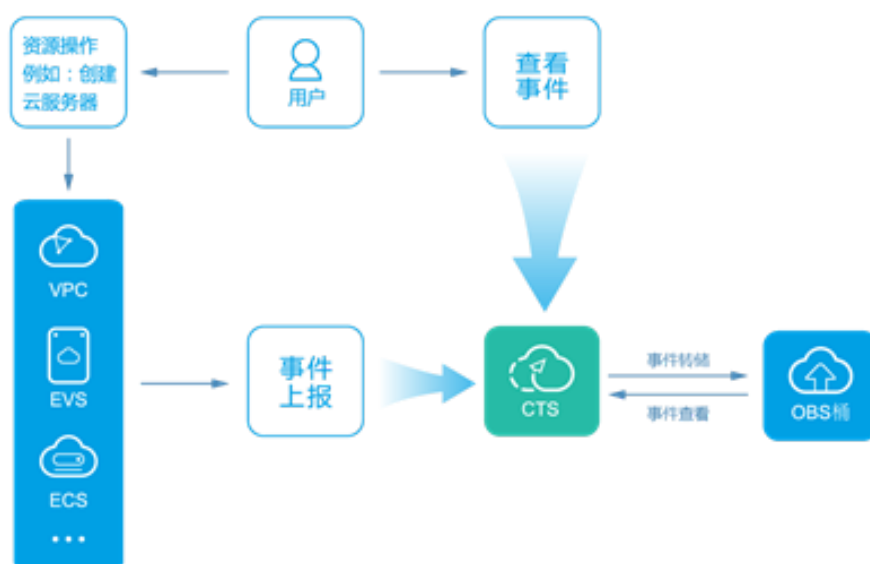
云审计服务（Cloud Trace Service，CTS），是华为云安全解决方案中专业的日志审计服务，提供对各种云资源操作记录的收集、存储和查询功能，可用于支撑安全分析、合规审计、资源跟踪和问题定位等常见应用场景。

用户开通云审计服务并创建和配置追踪器后，CTS可记录AOM的管理事件用于审计。

CTS的详细介绍和开通配置方法，请参见[CTS快速入门](#)。

CTS支持追踪的AOM管理事件列表，请参见[审计](#)。

图 9-2 云审计服务



日志

AOM支持采集容器服务日志和虚拟机（这里的虚拟机指操作系统为Linux的弹性云服务器或裸金属服务器）日志采集，并展现在AOM界面中供用户搜索和查看，详细介绍请参见[日志分析](#)。

9.5 服务韧性

AOM服务提供了多种可靠性容灾能力，可以通过AZ内实例容灾、多AZ容灾、多集群容灾、数据多副本等技术方案，保障服务的持久性和可靠性。

表 9-2 AOM 服务可靠性架构

可靠性方案	简要说明
AZ内实例容灾	单AZ内，AOM实例通过多实例方式实现实例容灾，快速剔除故障节点，保障实例持续提供服务。
多AZ容灾	AOM支持跨AZ容灾，当一个AZ异常时，不影响其他AZ下的实例持续提供服务。
多集群容灾	AOM支持多集群方案，当一个集群异常时，可以继续提供服务。
数据容灾	AOM的配置、指标、告警等数据均是多副本存储，保证了数据可靠性。

9.6 监控安全风险

AOM通过多种方式监控安全风险，保障数据安全可靠，具体请参见[表9-3](#)。

表 9-3 AOM 的监控安全风险

监控安全风险	简要说明	详细介绍
资源监控	AOM提供应用监控、工作负载监控、集群监控、主机监控、指标浏览等功能，支持实时监控您的应用及相关云资源，并将数据可视化展示，帮助用户快速分析应用健康状态。	<ul style="list-style-type: none">● 应用监控● 工作负载监控● 集群监控● 主机监控● 指标浏览
告警管理	AOM支持通过告警规则对应用、资源及业务设置告警条件，当AOM自身或外部服务存在异常或可能存在异常时，将告警信息通过邮件、短信、企业微信等方式发送给指定人员，帮助用户及时发现故障。	告警管理使用说明

9.7 认证证书

合规证书

华为云服务及平台通过了多项国内外权威机构（ISO/SOC/PCI等）的安全合规认证，用户可自行[申请下载](#)合规资质证书。

图 9-3 合规证书下载



资源中心

华为云还提供以下资源来帮助用户满足合规性要求，具体请查看[资源中心](#)。

图 9-4 资源中心



销售许可证&软件著作权证书

另外，华为云还提供了以下销售许可证及软件著作权证书，供用户下载和参考。具体请查看[合规资质证书](#)。

图 9-5 销售许可证&软件著作权证书



10 基本概念

10.1 应用资源管理

表 10-1 应用资源管理基本概念

术语	说明
应用资源管理	简称CMDB，在AOM2.0中管理应用与云资源的关系，并为AOM的监控、自动化运维、APM服务提供统一、及时的资源环境管理服务。
应用TOPO结构	资源与应用关系的层次结构。CMDB遵循“应用 + 子应用（可选）+ 组件 + 环境”的应用管理模型。
企业项目	华为云企业项目，一个项目可以包含一个或者多个应用。
应用	应用资源管理的基础对象，资源管理模型的根节点。一个应用代表一个逻辑单元，可以理解为项目、产品、业务。应用创建后，每个region都可以看到相同的应用拓扑信息。 例如一个商城应用包含用户管理服务、登录验证服务、商品列表、订单管理服务等。
子应用	应用管理的可选节点，同一个应用下，最多可创建三层子应用。子应用可以理解为服务，对组件/微服务的归组分类。
组件	构成应用的最小单元，可以理解为应用程序或者应用依赖的一个中间件云服务组件，例如RDS，DMS。组件一般跟下面的环境一起组合使用，一个组件可以包含一个或者多个环境。比如一个订单的应用程序，包含功能测试环境，压力测试环境，预发环境以及现网环境等。
环境	用于区分一个组件的不同环境或者一个组件的多个部署区域。一个组件或者程序，由于部署不同的配置参数，形成多个环境。每个环境都有region属性，可以通过region信息实现环境的过滤，也可以在创建环境时打上一个或多个标签，通过标签进行环境过滤。例如按照环境类型区分包括：正式、测试。
环境标签	为环境设置的一个属性，多个环境可能具有相同的的标签，通过标签可过滤显示需要的环境。同一个标签只能添加在本应用下的不同环境，不能跨应用添加。

术语	说明
资源绑定	将资源对象与应用下的环境建立关联关系，在同一个应用下资源对象实例可属于多个环境。
解绑资源	关联资源后，当组件或环境信息变化，不需要资源时，可将资源从原应用节点解除绑定。
资源转移	关联资源后，当组件或环境信息变化，需要及时转移资源时，可将资源从原应用节点转移到目标应用的节点下。

10.2 资源监控

表 10-2 资源监控基本概念

术语	说明
指标	<p>指标是对资源性能的数据描述或状态描述，指标由命名空间、维度、指标名称和单位组成。</p> <p>其中，命名空间特指指标的命名空间，可将其理解为存放指标的容器，不同命名空间中的指标彼此独立，因此来自不同应用程序的指标不会被错误地聚合到相同的统计信息中。维度是指标的分类，每个指标都包含用于描述该指标的特定特征，可以将维度理解为这些特征的类别。</p>
主机	AOM的每一台主机对应一台虚拟机或物理机。主机可以是您自己的虚拟机或物理机，也可以是您通过华为云购买的虚拟机（例如：弹性云服务器，简称ECS）或物理机（例如：裸金属服务器，简称BMS）。只要主机的操作系统满足AOM支持的操作系统，且主机已安装ICAgent，即可将主机接入到AOM中进行监控。
日志	<p>AOM提供了海量运行日志的检索和分析功能，支持日志采集、下载、转储、搜索，并提供报表分析、SQL查询、实时监控、关键词告警等能力。</p> <p>AOM的基础版和按需版所对应的日志存储时长、大小和计费方式不同，详见收费详情。</p>
日志流量	日志流量指的是每秒上报的日志大小。每个租户在每个Region的日志流量不能超过10MB/s。如果超过10MB/s，则可能导致日志丢失。
告警	告警是指AOM、ServiceStage、CCE、APM等服务在异常情况或在可能导致异常情况下上报的信息，告警会引起业务异常，您需要对告警进行处理。
事件	事件是指AOM、ServiceStage、CCE、APM等服务发生了某种变化，但不一定会引起业务异常，事件一般用来表达一些重要信息。您不用对事件进行处理。

术语	说明
告警清除	告警清除方式包括自动清除和手动清除两种。 <ul style="list-style-type: none">自动清除：产生告警的故障消除后，AOM会自动清除告警，您不需要做任何操作。手动清除：产生告警的故障消除后，AOM不会自动清除告警，您需要手动清除告警。
告警规则	告警规则分为指标告警规则和事件告警规则两种。 <ul style="list-style-type: none">通过指标告警规则，实时监控环境中主机、组件等资源使用情况。当资源使用告警过多，告警通知过于频繁时，通过事件告警规则，简化告警通知，快速识别服务的某一类资源使用问题并及时解决。
告警通知	告警通知有2种方式： <ul style="list-style-type: none">直接告警：在配置告警规则的时候，可以配置告警通知规则，将告警信息通过相关人，以便提醒相关人员及时采取措施清除故障。告警方式包括邮件、短信、钉钉、企业微信、语音等方式。告警降噪：选择告警降噪的分组规则进行告警降噪。
告警行动规则	告警行动规则定义产生告警之后，按照规则做何种动作。包括消息发送到哪里和以什么形式发送。消息发送到哪里通过 华为云服务SMN主题 设置。
Prometheus实例	Prometheus监控功能提供的管理Prometheus数据采集和数据存储分析的逻辑单元。
Prometheus探针	部署在用户侧或者云产品侧Kubernetes集群。负责自动发现采集目标、采集指标和远程写到其他库。
Exporter	一个采集监控数据并通过Prometheus监控功能规范对外提供数据的组件。目前有上百个官方或者第三方 Exporter可供使用，具体请参见 Exporter详情 。
Job	一组Target的配置集合。定义了抓取间隔，访问限制等作用于一组Target的抓取行为。

10.3 自动化运维

表 10-3 自动化运维基本概念

术语	说明
脚本管理	支持Shell、Python、Bat、Powershell脚本语言以及单个脚本的多版本管理。
作业管理	将脚本和文件原子操作进行多步骤编排形成作业模板，用于完成特定运维自动化场景操作。例如初始化业务环境。

术语	说明
执行方案	从作业模板中挑选1个或多个步骤组合组成执行方案。是作业模板的实例化对象。
云服务场景	云服务提供的原子化的变更场景。例如重启ECS。
标准化运维	将脚本、文件管理、执行方案、云服务场景组合合成一套操作流程，用于标准化特定场景的运维场景变更。
服务场景	将作业、标准化运维发布成服务，用户无需感知底层操作逻辑，仅输入简单的信息即可自助完成特定运维场景。
参数库	在作业、标准化运维中定义的全局共享参数，支持字符串，主机列表类型。除了达到多步骤参数共享目的，可清晰的看到整个流程涉及的全部参数和用途。
OS账号	用于执行脚本和文件管理的操作系统账户。
工具市场	服务默认提供的按照场景分类展现各功能的工具卡片，用户可根据需要控制场景从服务市场中上下架和基于安全考虑配置任务的审批流程。

10.4 采集管理

表 10-4 采集管理基本概念

术语	说明
UniAgent	统一数据采集Agent，完成统一插件生命周期管理，并为AOM提供指令下发，如脚本下发和执行。它自身不提供数据采集能力，运维由不同的插件分工采集，插件按需安装、升级和卸载。后续逐步上线其它插件（如云监控和主机安全），统一规范管理。
AK/SK	访问密钥。通过提供租户级别的AK/SK来安装ICAgent，以便于采集日志数据。
ICAgent	ICAgent用于采集指标、日志和应用性能数据，对于在ECS、BMS控制台直接购买的主机，您需手动安装ICAgent。对于通过CCE间接购买的主机，ICAgent会自动安装。
安装机	在AOM界面上，安装机支持批量下发安装UniAgent指令到主机，因此需要将VPC中某一台主机设置为安装机，该VPC其它主机均可通过界面远程安装。
代理区/代理机	为解决多云之间网络互通，需要在华为云购买和配置ECS主机为代理机，同时代理机上需要绑定公网IP，AOM通过该代理机下发部署控制命令到远程主机，运维数据也将经过该代理机至AOM。代理区由多个代理机组组成，主要是考虑代理机的高可用性。

11 权限管理

如果您需要对AOM资源给企业中的员工设置不同的访问权限，以达到不同员工之间的权限隔离，您可以使用统一身份认证服务（Identity and Access Management，简称IAM）进行精细的权限管理。该服务提供用户身份认证、权限分配、访问控制等功能，可以帮助您安全的控制AOM资源的访问。

通过IAM，您可以在华为云账号中给员工创建IAM用户，并使用策略来控制其对AOM资源的访问范围。例如您的员工中有负责软件开发的人员，您希望人员拥有AOM的使用权限，但是不希望其拥有删除应用发现规则等高危操作的权限，那么您可以使用IAM为开发人员创建用户，通过授予仅能使用应用发现规则，但是不允许删除应用发现规则的权限策略，控制其对应用发现规则资源的使用范围。

如果华为云账号已经能满足您的使用需求，不需要创建独立的IAM用户进行权限管理，您可以跳过本章节，不影响您使用AOM的其它功能。

IAM是华为云提供权限管理的基础服务，无需付费即可使用，您只需要为您账号中的资源进行付费。关于IAM的详细介绍，请参见《[IAM产品介绍](#)》。

AOM 权限

默认情况下，管理员创建的IAM用户没有任何权限，您需要将其加入用户组，并给用户组授予策略或角色，才能使得用户组中的用户获得对应的权限，这一过程称为授权。授权后，用户就可以基于被授予的权限对AOM进行操作。

AOM部署时通过物理区域划分，为项目级服务。授权时，“作用范围”需要选择“区域级项目”，然后在指定区域对应的项目中设置相关权限，并且该权限仅对此项目生效；如果在“所有项目”中设置权限，则该权限在所有区域项目中都生效。访问AOM时，需要先切换至授权区域。

根据授权精细程度分为角色和策略。

- **角色：** IAM最初提供的一种根据用户的工作职能定义权限的粗粒度授权机制。该机制以服务为粒度，提供有限的服务相关角色用于授权。由于华为云各服务之间存在业务依赖关系，因此给用户授予角色时，可能需要一并授予依赖的其他角色，才能正确完成业务。角色并不能满足用户对精细化授权的要求，无法完全达到企业对权限最小化的安全管控要求。
- **策略：** IAM最新提供的一种细粒度授权的能力，可以精确到具体服务的操作、资源以及请求条件等。基于策略的授权是一种更加灵活的授权方式，能够满足企业对权限最小化的安全管控要求。例如：针对ECS服务，管理员能够控制IAM用户仅能对某一类云服务器资源进行指定的管理操作。多数细粒度策略以API接口为粒度进行权限拆分，AOM支持的API授权项请参见[策略和授权项说明](#)。

如表11-1所示，包括了AOM的所有系统权限。

表 11-1 AOM 系统权限

子服务名称	策略名称	描述	类别	依赖系统权限
应用资源管理/监控中心/采集管理	AOM FullAccess	AOM2.0管理员权限，拥有该权限的用户可以操作并使用AOM。	系统策略	CCE ReadOnly Access、DMS ReadOnly Access
	AOM ReadOnlyAccess	AOM2.0只读权限，拥有该权限的用户仅能查看AOM数据。	系统策略	
自动化运维	CMS FullAccess	自动化运维管理员权限，拥有该权限的用户可以操作并使用自动化运维。	系统策略	-
	CMS ReadOnlyAccess	自动化运维只读权限，拥有该权限的用户仅能查看自动化运维数据。	系统策略	

应用资源管理常用操作与系统权限

表11-2列出了应用资源管理常用操作与系统权限的授权关系，您可以参照该表选择合适的系统权限。

表 11-2 应用资源管理常用操作与系统权限的关系

操作	AOM FullAccess	AOM ReadOnlyAccess
查询应用详情	√	√
查询子应用详情	√	√
查询组件详情	√	√
查询环境详情	√	√
查询环境标签	√	√
查询资源详情	√	√
新建应用	√	x
更新应用	√	x
删除应用	√	x
新建子应用	√	x
更新子应用	√	x
删除子应用	√	x

操作	AOM FullAccess	AOM ReadOnlyAccess
转移子应用	√	x
新建组件	√	x
更新组件	√	x
删除组件	√	x
转移组件	√	x
新建环境	√	x
更新环境	√	x
删除环境	√	x
创建环境标签	√	x
更新环境标签	√	x
删除环境标签	√	x
导入资源	√	x
更新资源	√	x
删除资源	√	x
转移资源	√	x
资源同步	√	x
绑定资源	√	x
解绑资源	√	x
开通资源授权	√	x
取消资源授权	√	x
获取应用列表	√	√
获取子应用列表	√	√
获取组件列表	√	√
获取应用下标签列表	√	√
获取资源列表	√	√
查询节点拓扑关系	√	√
查询操作记录	√	√

资源监控常用操作与系统权限

表11-3列出了资源监控常用操作与系统权限的授权关系，您可以参照该表选择合适的系统权限。

表 11-3 资源监控常用操作与系统权限的关系

操作	AOM FullAccess	AOM ReadOnlyAccess
创建告警规则	√	x
修改告警规则	√	x
删除告警规则	√	x
创建告警模板	√	x
修改告警模板	√	x
删除告警模板	√	x
创建告警行动规则	√	x
修改告警行动规则	√	x
删除告警行动规则	√	x
创建消息模板	√	x
修改消息模板	√	x
删除消息模板	√	x
创建分组规则	√	x
修改分组规则	√	x
删除分组规则	√	x
创建抑制规则	√	x
修改抑制规则	√	x
删除抑制规则	√	x
创建静默规则	√	x
修改静默规则	√	x
删除静默规则	√	x
创建仪表盘	√	x
修改仪表盘	√	x
删除仪表盘	√	x
创建Prometheus实例	√	x
修改Prometheus实例	√	x

操作	AOM FullAccess	AOM ReadOnlyAccess
删除Prometheus实例	√	x
创建应用发现规则	√	x
修改应用发现规则	√	x
删除应用发现规则	√	x
订阅阈值告警	√	x
配置虚机日志采集路径	√	x

自动化运维常用操作与系统权限

表11-4列出了自动化运维常用操作与系统权限的授权关系，您可以参照该表选择合适的系统权限。

表 11-4 自动化运维常用操作与系统权限的关系

操作	CMS FullAccess	CMS ReadOnlyAccess
创建脚本	√	x
编辑脚本	√	x
复制并新建脚本	√	x
编辑版本	√	x
查看脚本下的版本	√	√
创建文件包	√	x
查看文件包	√	√
编辑文件包	√	x
查看文件包版本列表	√	√
修改文件包版本	√	x
删除文件包	√	x
创建任务	√	x
编辑任务	√	x
删除任务	√	x
查看任务列表	√	√
查看任务详情	√	√
执行任务	√	x

采集管理常用操作与系统权限

表11-5列出了采集管理常用操作与系统权限的授权关系，您可以参照该表选择合适的系统权限。

表 11-5 采集管理常用操作与系统权限的关系

操作	AOM FullAccess	AOM ReadOnlyAccess
查询代理区域	√	√
编辑代理区域	√	x
删除代理区域	√	x
新建代理区域	√	x
查询某代理区域内所有的代理机	√	√
查询全部代理区域	√	√
安装Agent结果详情	√	√
获取对应主机的安装命令	√	√
获取对应主机的心跳及和server的连接状况	√	√
批量卸载运行中的Agent	√	x
查询Agent首页	√	√
测试安装机和目标机的连通性	√	x
批量导入安装Agent	√	x
获取Agent最近一次操作的执行日志相关信息	√	√
获取Agent安装时可选的版本列表	√	√
获取当前项目ID下所有Agent版本号列表	√	√
删除多条Agent主机	√	x

操作	AOM FullAccess	AOM ReadOnlyAccess
根据ecs_id查询Agent信息	√	√
删除单条Agent主机	√	x
设置安装机	√	x
重置安装机参数	√	x
查询当前登录用户的项目能够被设置成安装机的列表	√	√
查询Agent安装机列表	√	√
删除安装机	√	x
批量升级Agent	√	x
查询历史任务详情日志	√	√
查询历史任务详情	√	√
查询全部历史任务	√	√
获取所有种类的执行状态和任务类型	√	√
获取历史任务详情里Agent的执行状态种类	√	√
编辑代理机	√	x
删除Agent代理主机	√	x
设置代理主机	√	x
查询当前用户能够被设置成代理机的列表	√	√
批量更新插件	√	x
批量卸载插件	√	x
批量安装插件	√	x
查询插件历史任务详情日志	√	√
查询插件执行历史记录的分页列表	√	√

操作	AOM FullAccess	AOM ReadOnlyAccess
根据任务id, 查询插件执行记录详情的分页列表	√	√
获取历史任务详情里插件的执行状态的种类	√	√
获取全部插件列表	√	√
查询插件对应的版本号	√	√
查询当前支持的插件列表	√	√
获取租户下CCE集群信息列表	√	√
获取租户下某个CCE集群下的Agent信息列表	√	√
给租户下某个CCE集群安装ICAgent	√	x
给租户下某个CCE集群升级ICAgent	√	x
给租户下某个CCE集群卸载ICAgent	√	x
获取CCE集群列表	√	√
获取ICAgent机器列表	√	√
CCE集群机器安装ICAgent	√	x
CCE集群机器升级ICAgent	√	x
CCE集群机器卸载ICAgent	√	x

细粒度权限说明

使用自定义细粒度策略, 请使用管理员用户进入统一身份认证 (IAM) 服务, 按需选择AOM的细粒度权限进行授权操作。AOM细粒度权限说明请参见[表11-6](#)。

表 11-6 AOM 细粒度权限说明

权限名称	权限描述	权限依赖	应用场景
cms:workflow:create	创建任务	<ul style="list-style-type: none"> ecs:cloudServers:list ecs:cloudServers:listServerInterfaces rds:instance:list 	创建任务
cms:workflow:update	更新任务	<ul style="list-style-type: none"> functiongraph:function:create functiongraph:function:updateCode 	更新任务
cms:workflow:list	获取任务列表	无	查询任务列表
cms:execution:get	获取任务执行信息		查询执行详情
cms:execution:create	执行任务	<ul style="list-style-type: none"> functiongraph:function:create functiongraph:function:invoke functiongraph:function:list 	执行任务，如执行脚本，作业，安装和卸载文件包等
cms:template:get	获取模板信息	无	查询模板详情，查询作业执行方案详情
cms:template:list	获取模板列表		查询作业执行方案列表，查询可用于创建任务的模板列表
cms:script:get	查询脚本信息		查询脚本详情
cms:script:list	查询脚本列表		查询脚本列表
cms:job:list	查询作业列表		查询作业列表
aom:cmdbApplication:get	获取应用详情	无	根据应用ID获取应用详情、根据应用名称获取应用详情
aom:cmdbApplication:update	修改应用		修改应用
aom:cmdbApplication:delete	删除应用		删除应用
aom:cmdbApplication:get	获取应用详情		获取应用详情

权限名称	权限描述	权限依赖	应用场景
aom:cmdbComponent:get	查询组件详情		根据组件ID查询组件详情、根据组件名称查询组件详情
aom:cmdbComponent:create	新增组件		新增组件
aom:cmdbComponent:update	更新组件		更新组件
aom:cmdbComponent:delete	删除组件		删除组件
aom:cmdbComponent:move	转移组件		转移组件
aom:cmdbComponent:list	查询组件列表		查询组件列表
aom:cmdbEnvironment:create	创建环境		创建环境
aom:cmdbEnvironment:update	修改环境		修改环境
aom:cmdbEnvironment:get	获取环境详情		根据环境ID获取环境详情、根据环境名称Region和组件ID获取环境详情
aom:cmdbEnvironment:delete	删除环境		删除环境
aom:cmdbSubApplication:get	查询子应用详情		查询子应用详情
aom:cmdbSubApplication:update	修改子应用		修改子应用
aom:cmdbSubApplication:move	转移子应用		转移子应用
aom:cmdbSubApplication:delete	删除子应用		删除子应用
aom:cmdbSubApplication:create	创建子应用		创建子应用
aom:cmdbSubApplication:list	查询子应用列表		查询子应用列表

权限名称	权限描述	权限依赖	应用场景
aom:cmdbResources:unbind	解绑资源		解绑资源
aom:cmdbResources:bind	绑定资源		绑定资源
aom:cmdbResources:move	转移资源		转移资源
aom:cmdbResources:get	查询资源详情		查询资源详情
aom:alarm:put	上报告警		上报自定义告警
aom:event2AlarmRule:create	新增一条事件类告警规则		新增一条事件类告警规则
aom:event2AlarmRule:set	更新事件类告警规则		更新事件类告警规则
aom:event2AlarmRule:delete	删除事件类告警规则		删除事件类告警规则
aom:event2AlarmRule:list	查询全部事件类告警规则		查询全部事件类告警规则
aom:actionRule:create	新增告警行动规则		新增告警行动规则
aom:actionRule:delete	删除告警行动规则		删除告警行动规则
aom:actionRule:list	获取告警行动规则列表		获取告警行动规则列表
aom:actionRule:update	修改告警行动规则		修改告警行动规则
aom:actionRule:get	通过规则名称获取告警行动规则		通过规则名称获取告警行动规则
aom:alarm:list	获取告警发送结果		获取告警发送结果
aom:alarmRule:create	创建阈值规则		创建阈值规则
aom:alarmRule:set	修改阈值规则		修改阈值规则
aom:alarmRule:get	查询阈值规则		根据ID查询单个阈值规则或者查询所有阈值规则

权限名称	权限描述	权限依赖	应用场景
aom:alarmRule:delete	删除阈值规则		批量删除阈值规则或者根据ID删除单个阈值规则
aom:discoveryRule:list	查看应用发现规则		查询系统中已有应用发现规则
aom:discoveryRule:delete	删除应用发现规则		删除应用发现规则
aom:discoveryRule:set	添加应用发现规则		添加应用发现规则
aom:metric:list	查询时间序列		查询时间序列
aom:metric:list	查询时序数据		查询时序数据
aom:metric:get	查询指标		查询指标
aom:metric:get	查询监控数据		查询监控数据
aom:muteRule:delete	删除静默规则	无	删除静默规则
aom:muteRule:create	新增静默规则		新增静默规则
aom:muteRule:update	修改静默规则		修改静默规则
aom:muteRule:list	获取静默规则列表		获取静默规则列表

AOM 控制台功能依赖的角色或策略

如果IAM用户需要在AOM控制台拥有相应功能的查看或使用权限，请确认已经对该用户所在的用户组设置了AOM FullAccess或AOM ReadOnlyAccess策略的集群权限，再按如下[表11-7](#)增加依赖服务的角色或策略。

📖 说明

用户首次开通AOM服务，AOM会为其创建服务委托，用户除需授权AOM FullAccess外，还需要授予Security Administrator。

表 11-7 AOM 控制台依赖服务的角色或策略

控制台功能	依赖服务	需配置角色/策略
<ul style="list-style-type: none"> 工作负载 集群监控 	云容器引擎 CCE	如果使用工作负载监控和集群监控，需要设置CCE ReadOnlyAccess权限。

控制台功能	依赖服务	需配置角色/策略
数据订阅	分布式消息服务Kafka版	如果使用数据订阅功能，需要设置DMS ReadOnlyAccess权限。

12 隐私声明

由于AOM会将运维数据内容展示到AOM控制台，请您在使用过程中，注意您的隐私及敏感信息数据保护，不建议将隐私或敏感数据上传到AOM，必要时请加密保护。

采集器部署

在弹性云服务器 ECS上手动部署ICAgent过程中，安装命令中会使用到您的AK/SK作为输入参数，安装前请您关闭系统的历史记录收集，以免泄露隐私。安装后ICAgent会加密存储您的AKSK，有效保护敏感信息。

容器监控

在CCE容器监控场景下，AOM的采集器（ICAgent）必须以特权容器的方式运行，请合理的评估特权容器的安全风险，谨慎识别您的容器业务场景。如：节点对外提供的业务属于逻辑多租的共享容器方式，建议采用开源Prometheus等工具进行监控，避免使用AOM的采集器监控您的业务。

13 修订记录

表 13-1 版本说明列表

日期	修订记录
2023-09-30	<ul style="list-style-type: none">• 新增以下章节：<ul style="list-style-type: none">- AOM与CES的云服务监控功能对比- 安全• 优化以下章节：<ul style="list-style-type: none">- 基础指标：容器指标
2023-09-15	删除“计费说明”章节。
2023-08-04	删除“基础指标-云服务指标”章节。
2023-06-30	<ul style="list-style-type: none">• 新增以下章节：<ul style="list-style-type: none">- 与AOM 1.0对比- 基础指标：IoTDA指标• 优化以下章节：<ul style="list-style-type: none">- 基础指标：容器指标- 基础指标-云服务指标
2022-12-30	<ul style="list-style-type: none">• 新增以下章节：<ul style="list-style-type: none">- 基础指标：虚拟机指标- 基础指标：容器指标- 基础指标-云服务指标- 指标维度• 优化以下章节：<ul style="list-style-type: none">- 简介- 计费说明
2022-06-30	第一次发布。