云日志服务

用户操作指南

天翼云科技有限公司

1. 产品介绍

1.1. 产品定义

云日志服务（ALS）是天翼云打造的一款云原生日志观测分析平台产品，可为应用的海量日志数据提供大规模、低成本、集中式的平台化服务，具备一站式的日志采集、加工、查询分析、可视化、告警、投递消费等能力，全面满足应用研发、运维、服务监控与业务分析等应用场景。

基本概念

您在使用云日志服务前，需要了解以下基本概念

术语	说明
日志项目	日志项目是用于管理云日志服务的资源单元，通常可将某个独立项目/业务的日志对应至一个日志项目中。每个日志项目可包含多个日志单元。
日志单元	日志单元是进行日志数据的采集、存储、检索和分析的基本单元，日志数据以日志单元的方式进行管理，通常可将一个应用/服务下的日志采集至一个日志单元中。
日志	日志是应用系统运行过程中产生的记录数据，这些数据包括用户的操作、接口的访问、系统发生的错误等。通常以文本形式存储在应用系统所在的设备上。
采集器	云日志服务提供的日志采集工具
主机组	主机组是一组需要采集日志的云主机列表，是一个虚拟分组，云日志服务通过主机组来管理所有需要通过采集器采集日志的云主机。
采集配置	采集配置是指采集器执行日志采集时的策略，包括文件采集路径、分词模式等。

核心功能概览

- 日志采集

支持从云主机、容器应用中采集日志，实现日志数据统一采集、批量抓取上报、多管道数据处理，提高采集效率，同时支持采集规则动态配置，易于管理

- 日志检索与分析

支持秒级日志查询，提供内容模糊、时间、上下文等多种查询方式，提供各类统计分析场景，并支持可视化仪表盘，实时观察各种数据，提升分析效率。

- 数据加工

支持各种复杂数据的加工，包括数据的规整、脱敏和过滤

- 日志告警

支持设置自定义告警规则，针对日志数据进行监控，支持多个告警渠道，及时发现问题

购买方式

实名注册天翼云账号后，即可在产品详情页点击**立即开通**进行服务开通，开通云日志服务不收取任何费用，公测期间可免费试用。详情请参考快速入门-开通云日志服务

1.2. 功能特性

日志采集

- t 采集器支持无侵入式接入，实现日志实时采集、日志文件批量抓取上报、多管道数据处理，实现高效采集。
- 支持从云主机、容器应用、微服务应用以及其他组件中采集日志
- 采集规则支持动态配置，提供单行/多行全文、正则、分隔符、JSON 等日志结构化解析方式

日志检索与分析

- 对于采集到的日志数据，支持秒级日志查询，从日志采集到可查询的时间<1 分钟
- 支持模糊查询、全文查询、字段查询、时间范围查询
- 支持上下文查询功能
- 支持各类查询分析场景

数据加工

支持对日志数据进行加工，包括数据的规整、脱敏和过滤：

- 数据规整：由于日志数据通常来自于不同的系统组件、应用程序或设备，其格式和结构可能各异，导致在对日志数据进行分析、搜索和可视化时出现困难。数据规整可针对混乱格式的日志进行字段提取、格式转换，统计为一致性格式以便后续的处理与分析。

- 数据脱敏：对日志数据中的敏感信息（如密码、手机号、地址等）行脱敏。
- 数据过滤：针对关键业务或服务的日志进行过滤，用于后续重点分析等场景。

告警

- 告警规则：支持对一个或多个日志单元设置自定义告警规则，告警规则将按照设定的周期执行监控任务，对指定的日志单元执行检索分析，当检索结果满足触发条件时将发送告警通知，以便用户及时发现异常问题。
- 告警通知：支持自定义通知策略与通知对象，可实现邮件、短信、翼连等方式的告警通知
- 静默策略：支持设计告警静默策略

1.3. 产品优势

一站式托管

提供从采集、存储、加工、检索分析、投递的一站式云日志服务，免人工运维。

快速响应

采集端安装便捷，实时采集传输日志，入库后即可检索分析。

海量日志处理

亿级日志秒级检索，轻松应对海量日志。

低成本

日志数据高压缩比存储，自定存储时长，按量付费。

1.4. 应用场景

运维管理

在集群的运维工作中，企业应用的运维日志分散在不同的节点上，通过应用云日志服务提供的采集器将分散在集群各个节点的重要日志数据采集到云日志服务平台进行集中化统一管理，享受全托管式服务。

- 全托管式：日志集中统一管理，提供日志采集、存储、检索与分析能力。
- 海量日志管理：支持每天百 TB 级日志的接入，十亿级日志秒级搜索

- 提升运维效率：通过关键词检索可快速搜索出异常事件的日志，定位问题节点，结合上下文查询能力将异常事件的调用链完整还原，全面提升运维效率。

性能优化和容量规划

通过云日志服务收集和分析系统的性能日志，可以帮助识别性能瓶颈，并进行优化。此外，通过分析日志数据，可以进行容量规划，确保系统资源的合理分配，以满足未来的需求。

用户行为分析

在应用程序或网站中记录用户的行为日志，可以用于分析用户的偏好、行为模式和趋势，从而改进产品和提供更好的用户体验。

1.5. 术语解释

基础资源

术语	说明
日志项目	日志项目是用于管理云日志服务的资源单元，通常可将某个独立项目/业务的日志对应至一个日志项目中。每个日志项目可包含多个日志单元。
日志单元	日志单元是进行日志数据的采集、存储、检索和分析的基本单元，日志数据以日志单元的方式进行管理，通常可将一个应用/服务下的日志采集至一个日志单元中。
日志	日志是应用系统运行过程中产生的记录数据，这些数据包括用户的操作、接口的访问、系统发生的错误等。通常以文本形式存储在应用系统所在的设备上。
日志组	日志组是一个包含多条日志的集合，是写入与读取日志的基本单位，提高数据读写效率。

数据采集

术语	说明
采集器	云日志服务提供的日志采集工具
主机组	主机组是一组需要采集日志的云主机列表，是一个虚拟分组，云日志服务通过主机组来管理所有需要通过采集器采集日志

	的云主机。
采集配置	采集配置是指采集器执行日志采集时的策略，包括文件采集路径、分词模式等。

查询与分析

术语	说明
查询	通过查询条件指定过滤规则，返回符合条件的日志。目前支持关键字模糊查询、全文查询、字段查询
分析	在查询的基础上，通过构造各类分析场景，分析返回分析结果。

数据加工

术语	说明
数据加工	对日志数据进行加工，包括数据的规整、脱敏和过滤的过程，可理解为日志 ETL
DSL	DS 是一种 Python 兼容的脚本语言，用于进行日志加工服务
加工规则	数据加工脚本，是一组 DSL 编排的逻辑代码的集合

1.6. 使用限制

、

2.计费说明

2.1. 计费模式

云日志服务产品目前处于公测阶段，支持免费试用。

限制类别	说明
日志项目	您在 1 个天翼云账号下最多可创建 1 个日志项目。

日志单元	您在 1 个日志项目中最多可创建 5 个日志单元。
采集规则	一个日志单元最多创建 250 个采集规则配置。
日志保存时间	日志保存时间为 15 天。

3. 快速入门

3.1. 开通云日志服务

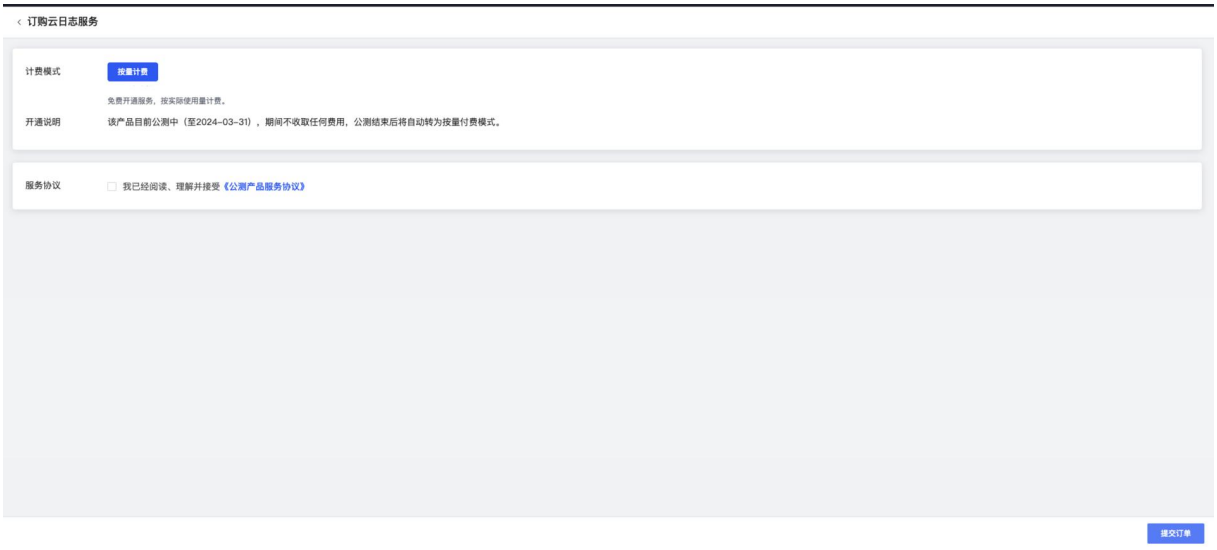
在使用天翼云-云日志服务前，您需要先开通服务。开通服务不收取任何费用，在公测期间，您可免费使用云日志服务。

前提条件

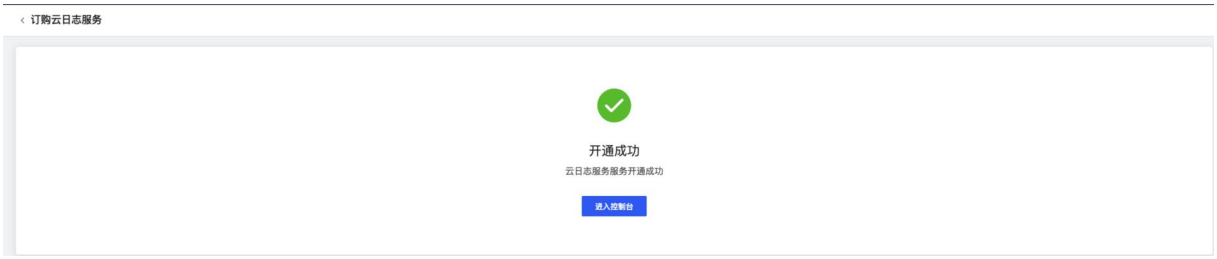
- 具备已通过实名认证的天翼云账号

操作步骤

1、登录天翼云官网，进入云日志服务产品详情页
2、点击【立即开通】。
3、在页面上方选择资源池。
4、勾选用户协议，并点击提交订单。



5、提交后稍等片刻，页面提示开通成功后，即可进入云日志服务控制台。



3.2. 创建日志项目与日志单元

在采集日志前，您需要先创建日志项目与日志单元。日志项目是云日志服务的资源管理单位，您可使用日志项目管理不同的应用、产品或项目中的数据。每个日志项目下可创建多个日志单元，是您访问云日志服务资源的入口。日志单元是云日志服务中日志数据的采集、存储、查询、分析的基本单元。每个日志单元隶属于一个日志项目，每日志项目中可创建多个日志单元。本文主要介绍云日志服务如何创建日志项目与日志单元。

前提条件

- 具备已通过实名认证的天翼云账号
- 已开通可用的天翼云 linux 云主机，且持续产生日志

操作步骤

1、登录云日志服务控制台，

2、创建日志项目

(1) 在概览页右侧，点击创建项目

(2) 在创建项目弹窗中，输入项目的名称，项目名称全局唯一。

(3) 点击确定，完成日志项目创建

3、创建日志单元

(1) 在云日志服务控制台概览页面右侧的日志项目区域，选择日志项目并点击名称，或在左侧菜单栏中点击【日志项目】，选择选择日志项目并点击名称，进入日志项目页面。

(2) 页面左上方点击【+】按钮

(3) 在创建日志单元弹窗，输入日志单元名称与日志保存时间。

(4) 点击确定，完成日志单元创建

3.3. 采集日志

前提条件

- 已开通云日志服务
- 已创建日志项目与日志单元

操作步骤

(1) 在创建好的日志单元下，点击【数据接入】

(2) 选择云主机

a. 在已开通云主机列表中，选择您需要采集日志的目标云主机。选中的云主机将打包作为主机组

b. 输入主机组名称

c. 点击下一步

(3) 安装采集器

a. 安装采集器前，需要先接入云主机所在的 VPC 网络。在 Step1：VPC 接入中，点击【全部接入】按钮，稍等片刻后，确保列表中所有 VPC 都处于已接入状态。
若 VPC 无法接入，请查看“VPC 接入失败”进行问题排查

b. 复制 Step2 中提供的采集器安装命令，在目标云主机中执行该命令安装采集器。

c. 安装完成后需要检查采集器状态。在 Step3：检查采集器状态中，点击【开始检测】按钮，稍等片刻后查看采集器状态，确保所有目标云主机的采集器状态均为已连通。
若多次重试后仍无法连通，请查看“云主机采集器无法连通”进行问题排查

d. 点击下一步

(4) 创建采集配置，请如下说明配置参数。

参数	描述
采集规则名称	采集规则的名称，在所属的日志项目内唯一，
采集路径	根据日志在服务器上的位置，设置日志目录和文件名称。 本案例的日志路径为/var/log/nginx/access.log。
采集策略	选择“全量”
切割模式	针对原始日志执行分词的模式，本案例选择“单行全文” 其它切割模式请查看数据采集—采集文本日志

3.4. 查询与分析日志

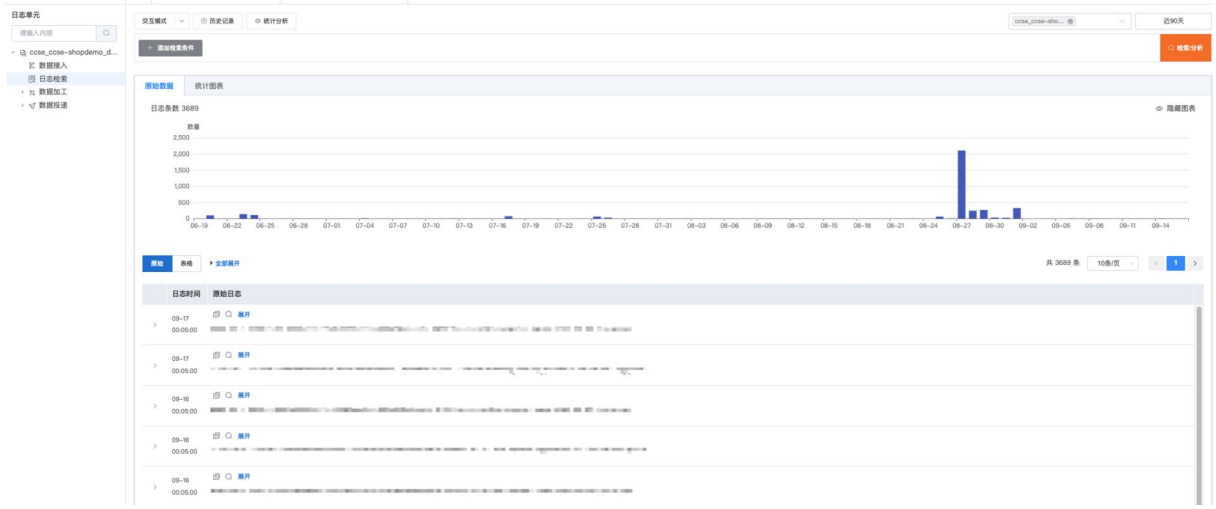
前提条件

- 已根据接入向导成功配置日志采集规则

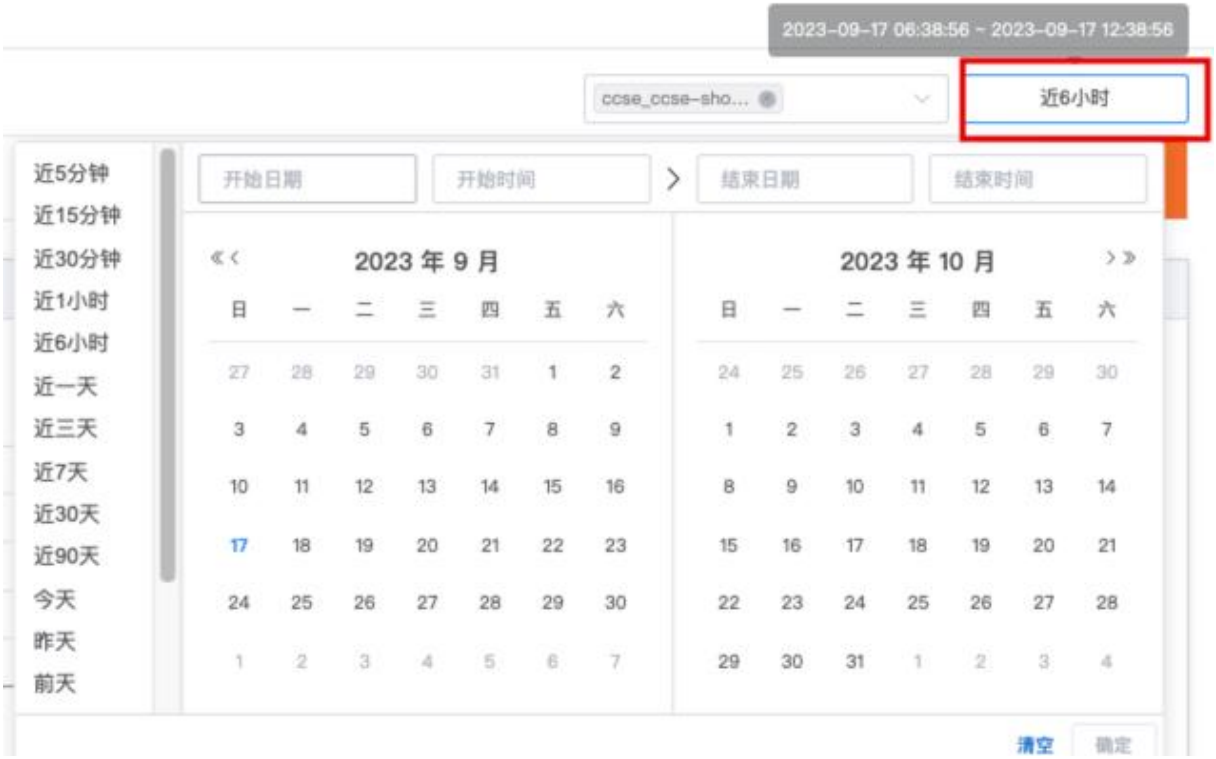
操作步骤

1、在步骤二的配置向导结束步骤中，点击【进行日志检索】。或在左侧日志单元中下方点击【日志检索】，进入检索页面

2、在检索页面中点击【检索/分析】按钮，即可查看已采集到的日志。



3、可在右上方选择时间范围进行日志过滤



4、在页面上方检索输入框中输入关键词，即可进行全文关键词检索，如下所示，查询带有“level”关键词的日志

5、您也可以通过切换至交互搜索模式进行字段搜索、多条件搜索。

6、点击【统计分析】按钮，可进行字段筛选、过滤、指标统计、分组统计等统计分析操作，例如以下案例中，统计各个 pod（podIp 字段）对应的日志数量。

关于查询与分析的更多信息，请参考操作指南-查询与分析

4. 用户指南

4.1. 资源管理

4.1.1. 资源管理概述

在使用云日志服务进行日志采集、加工、查询、分析、可视化、告警、投递等操作前，您需要先创建相关资源，具体流程如下。

1、注册天翼云账号并完成实名认证。

2、开通云日志服务

(1) 进入云日志服务产品详情页，点击【立即开通】

(2) 根据页面提示，开通云日志服务。开通服务不收取费用，且公测期间将提供免费试用。

3、创建日志项目，具体操作请查看管理日志项目

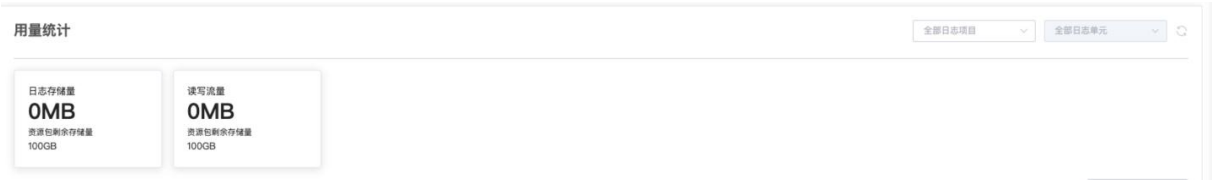
4、创建日志单元，具体惭怍请查看管理日志单元

4.1.2. 控制台首页

云日志服务控制台首页提供用量统计、用量明细、日志项目等信息。

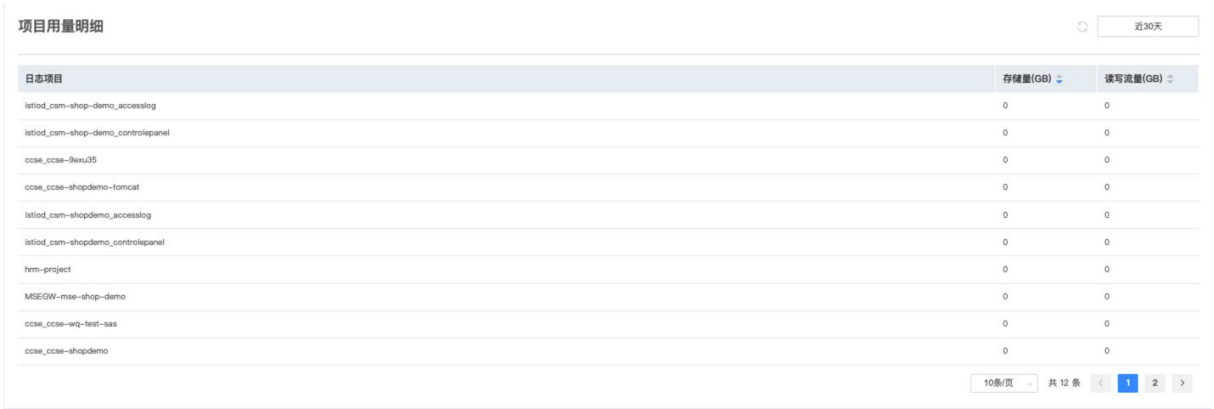
用量统计

提供当前日志存储量与累计读写流量统计，可根据项目与时间范围进行统计。详情请查看资源统计。



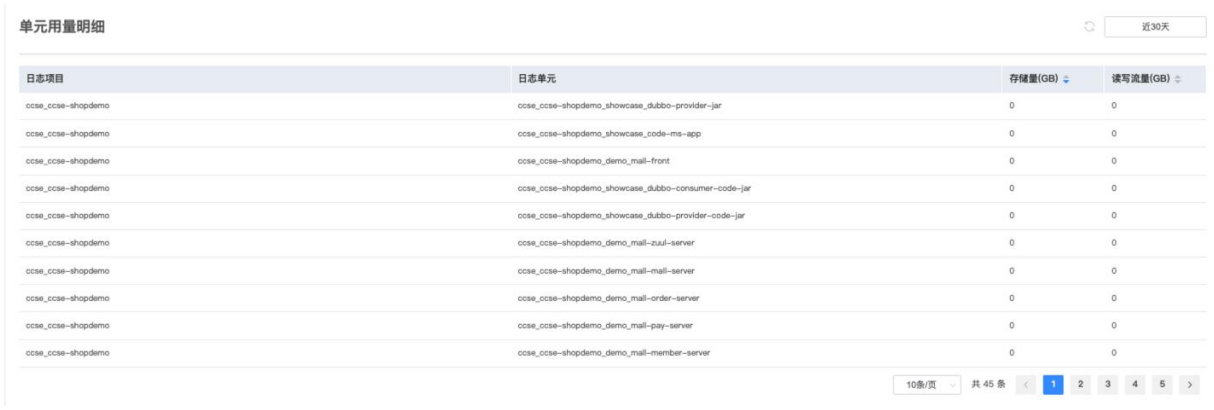
项目用量明细

可查看当前每个日志项目下的存储量与累计读写流量，支持根据时间范围进行统计。详情请查看资源统计。



单元用量明细

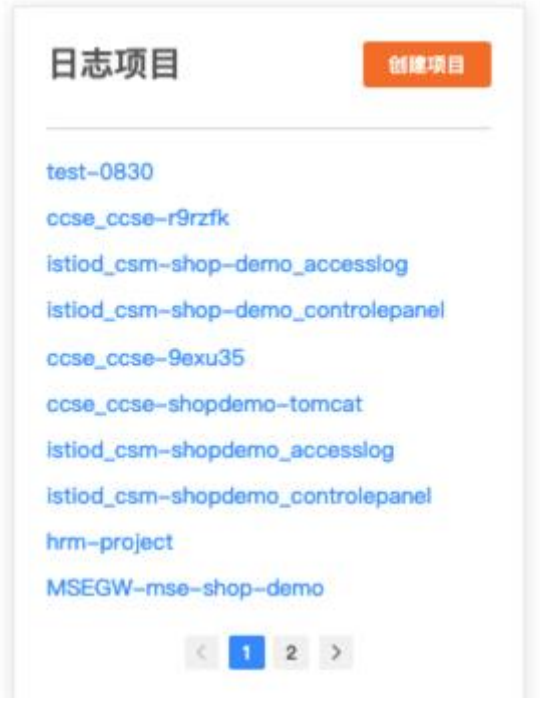
可查看当前每个日志单元下的存储量与累计读写流量，支持根据时间范围进行统计。详情请查看资源统计。



日志项目

可查看您创建的所有日志项目。点击项目名称可进入项目详情页面进行日志数据管理。详情请

查看管理日志项目。



4.1.3. 资源统计

云日志服务资源统计是对日志使用资源进行统计与可视化展示，主要分类有日志存储量与读写流量。因数据上报与存储时间有先后，统计数值与实际情况会有略微差异，因此统计数值仅供参考。

- 日志存储量：指日志压缩后的存储量。
- 读写流量：根据日志数据被压缩后上传到云日志服务所产生的传输流量计算。

用量统计

用量统计主要展示当前日志存储量与累计读写流量，默认情况下展示所有日志项目所产生的用量。您可根据自己的实际需求选择目标日志项目与日志单元。

可查看指定时间范围内日志存储量、读写流量的使用趋势，您可根据自己的实际需求选择时间范围，可选择相对时间、整点时间、自定义时间。

- 相对时间：表示查询距离当前时间 1 天、3 天、1 周等时间区间的用量情况
- 整点时间：表示查询今天、昨天、本周、上周等时间区间的用量情况。
- 自定义时间：表示查询指定时间范围的用量情况。

选择时间范围后，将按照选择时间范围显示数据趋势图。趋势图中每个点表示某时间内的数据统计。

鼠标悬浮至趋势图中可展示具体时间与统计数据。

项目用量明细

项目用量明细主要展示当前所有日志项目所产生的日志存储量与读写流量，默认展示时间为 30 天（相对）的资源数据，您可根据自己的实际需求选择时间范围。

点击目标日志项目，可在下方单元用量明细中查看该项目下所有日志单元所产生的日志存储量与读写流量。

单元用量明细

单元用量明细从日志单元维度展示当前所产生的日志存储量与读写流量，默认展示时间为 30 天（相对）的资源数据，您可根据自己的实际需求选择时间范围。

默认情况下展示所有日志单元的用量明细，您可点击上方项目用量明细中的项目名称，以查看隶属于该项目的日志单元的用量明细。

4.1.4. 管理日志项目

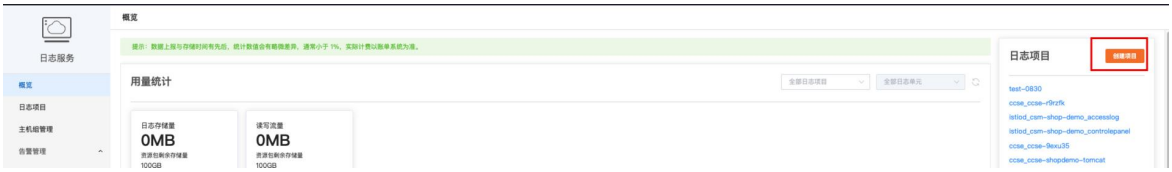
概述

日志项目是云日志服务的资源管理单位，您可使用日志项目管理不同的应用、产品或项目中的数据。每个日志项目下可创建多个日志单元，是您访问云日志服务资源的入口。本文将介绍如何通过控制台管理日志项目

操作步骤

创建日志项目

1. 登录云日志服务控制台
2. 在概览页面右方的日志项目栏目，点击【创建项目】；或在左侧菜单栏中点击【日志项目】，在页面中点击【创建项目】



3. 在创建项目弹窗中，输入项目的名称，项目名称全局唯一。

4. 点击确定，完成日志项目创建

编辑日志项目

1. 在控制台左侧菜单栏中点击【日志项目】。
2. 在页面中找到您需要编辑的日志项目，点击【编辑】。

3. 在弹出的编辑日志项目窗口中，可修改日志项目名称与备注。

4. 点击确定，即可保存编辑内容。

删除日志项目

1. 在控制台左侧菜单栏中点击【日志项目】
2. 在页面中找到您需要删除的日志项目，点击【删除】
3. 在弹出提示框中，单击确定，即可删除当前日志项目。

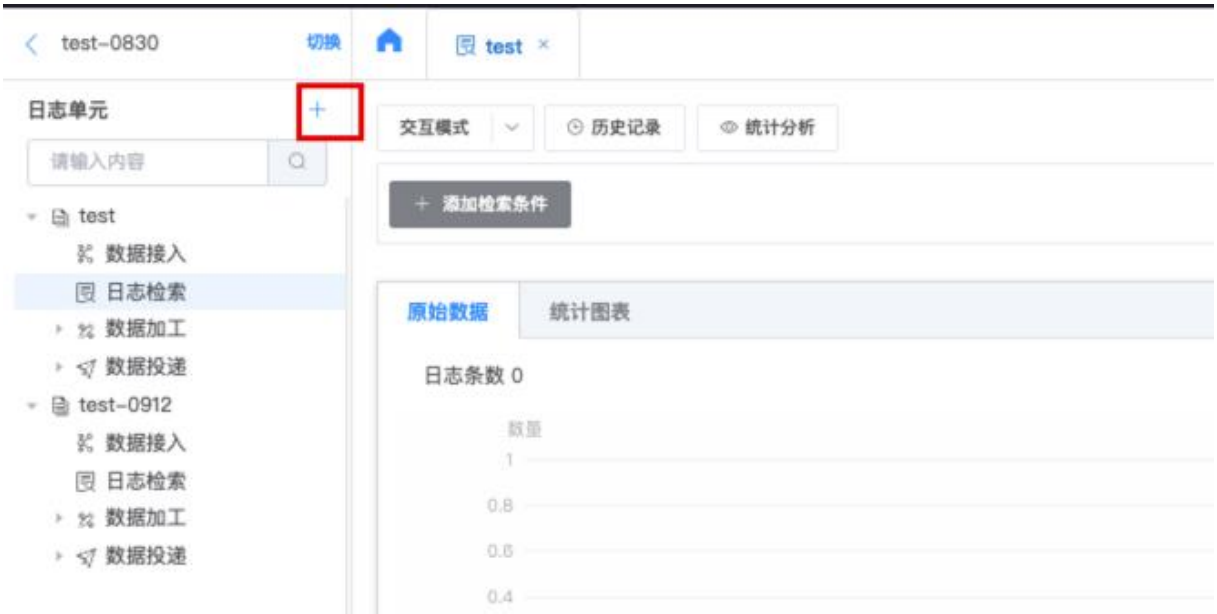
4.1.5. 管理日志单元

概述

日志单元是云日志服务中日志数据的采集、存储、查询、分析的基本单元。每个日志单元隶属于一个日志项目，每日志项目中可创建多个日志单元。本文介绍如何通过控制台管理日志单元。

创建日志单元

1. 进入云日志服务控制台概览页面，在右侧的日志项目区域，单击目标日志项目；或在左侧菜单栏中点击【日志项目】，单击目标日志项目，进入日志项目页面。
2. 在日志项目页面左上方点击【+】按钮



3．在创建日志单元弹窗，输入日志单元名称与日志保存时间。

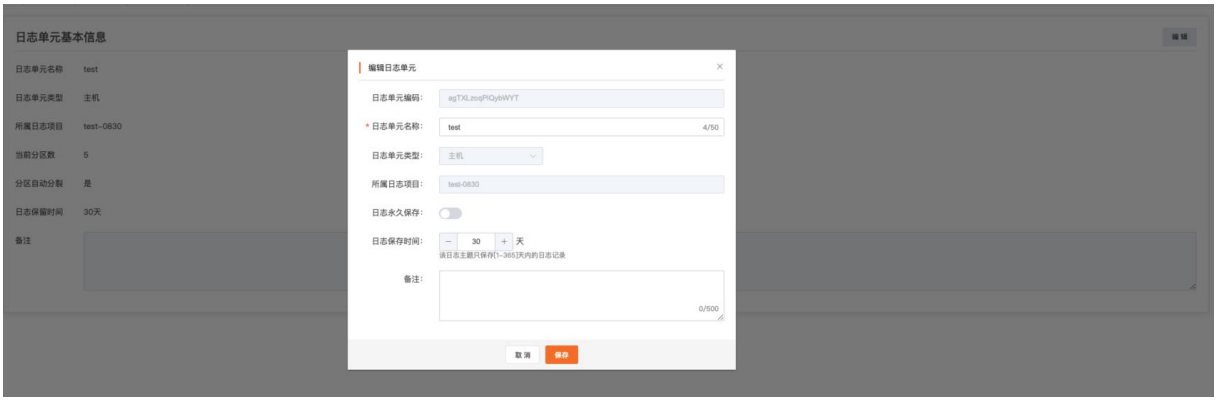
4．点击确定，完成日志单元创建

编辑日志单元

1．在日志项目页面中，点击目标日志单元名称，即可查看日志单元基本信息。

2．在页面右上方点击【编辑】按钮。

3．在弹出的编辑日志单元窗口中，可修改日志单元名称、日志保存时间与备注。



4．点击保存，即可保存编辑内容。

删除日志单元

1. 在日志项目页面中，将鼠标悬浮在目标日志单元上，然后点击删除按钮。
2. 在弹出提示框中，单击确定，即可删除当前日志单元。

4.1.6. 管理主机组

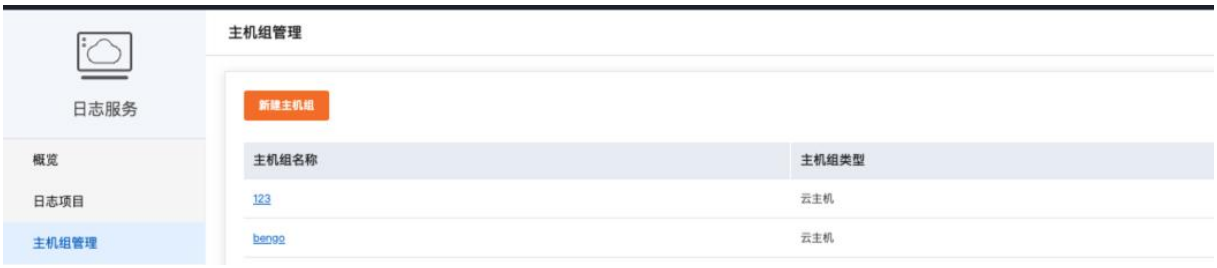
概述

主机组是一组需要采集日志的云主机列表，是一个虚拟分组，云日志服务通过主机组来管理所有需要通过采集器采集日志的云主机。建议您可根据不同的业务场景来划分不同的主机组组，以方便管理云日志服务。

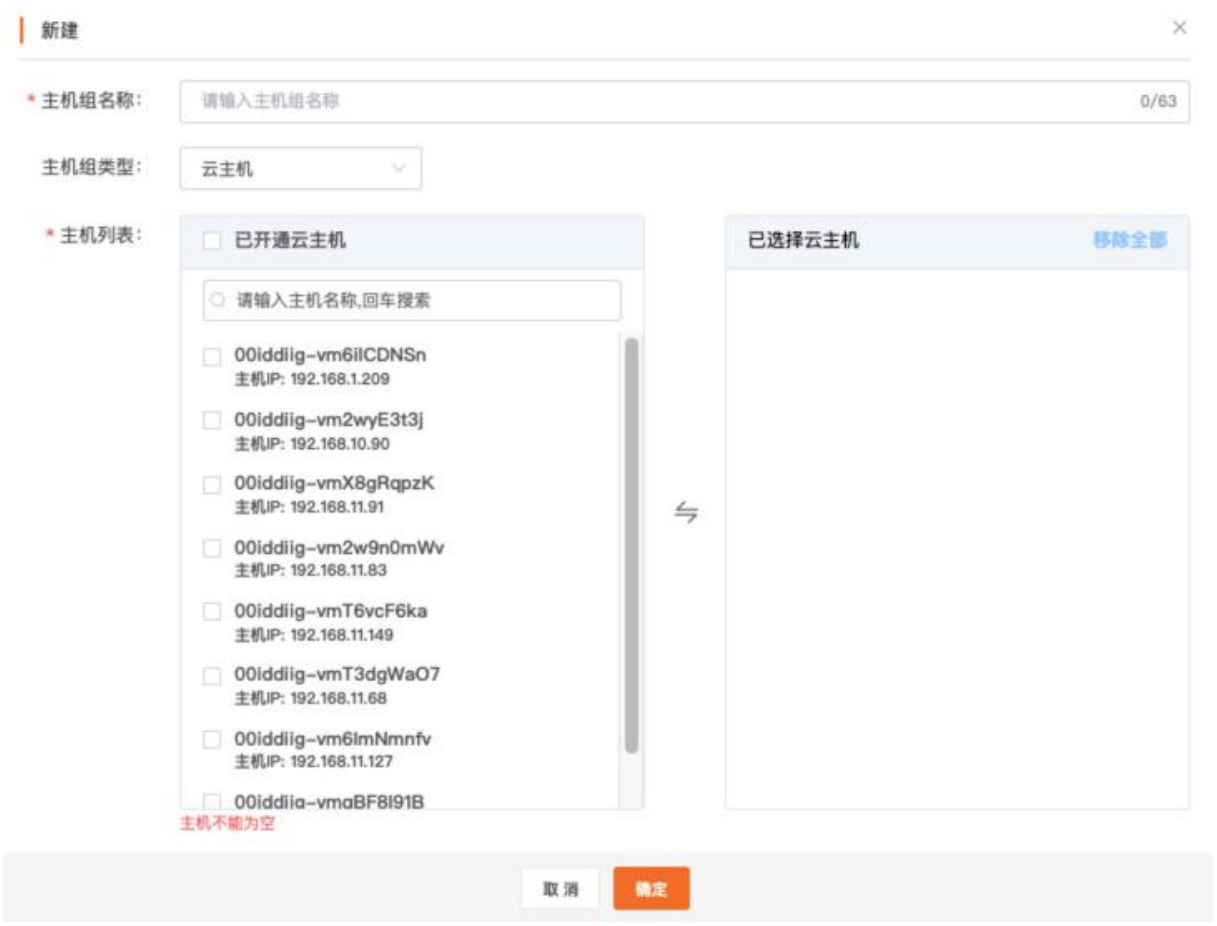
操作步骤

创建主机组

1. 登录云日志服务控制台。
2. 在左侧导航栏中，点击主机组管理。



3. 在页面左上方点击【新建主机组】。
4. 在弹出的窗口中，输入主机组名称，并在云主机列表中选择勾选目标云主机。
   点击【从已有主机组中导入云主机】，可快速导入其他主机组所包含的云主机。



5. 点击确定，即可完成创建。

查看主机组详情

1. 在左侧导航栏中，点击主机组管理
2. 点击目标主机组名称，进入主机组详情页面
3. 查看基本信息，包括主机组名称、主机组类型、云主机个数与创建时间。点击主机组名称编辑按钮，可修改主机组名称。

4. 查看云主机配置详情：

字段	说明
主机名称	属于当前主机组的所有云主机的名称
IP	属于当前主机组的所有云主机的 IP 地址
主机状态	● 运行中：表示该主机正常运行 ● 异常：表示该主机已被冻结或销毁，将无

	法采集该主机的日志数据
采集器状态	● 已连通：表示该主机已安装采集器，且采集器可正常上报心跳数据。 ● 未连通：表示该主机未安装采集器，或网络原因导致无法上报心跳数据。详情请查看“云主机采集器无法连通”进行问题排查 点击列表上方的【检测采集器状态】，可批量执行采集器状态检测。

5．查看主机组已关联的日志单元。点击目标日志单元名称，可跳转至日志单元页面。

编辑主机组

1．在左侧导航栏中，点击主机组管理

2．点击目标主机组名称，进入主机组详情页面

3．新增云主机

a．在云主机配置模块，点击【新增云主机】

b．在弹出的弹窗中，选择目标云主机

c．点击确定，即可在该主机组中添加云主机

4．移除云主机

a．在云主机配置模块，选择目标云主机，点击【移除】按钮

b．在弹出提示框中，单击确定，即可移除当前云主机。

注意：从主机组移除该主机后，关联了该主机组的日志单元将会停止采集该主机的日志

c．点击【批量移除】，勾选目标云主机，可实现批量移除云主机

删除主机组

1．在左侧导航栏中，点击主机组管理

2．选择目标主机组，点击【删除按钮】

3．在弹出提示框中，单击确定，即可移除当前主机组。

注意：删除该主机组后，将会停止采集该主机组的日志数据，该操作将会影响所有与当前主机组关联的日志单元

4.2. 数据采集

4.2.1. 数据采集概述

云日志服务提供采集器的方式进行日志采集，方便用户在各种数据源场景下采集日志并导入到云日志服务中。目前支持Linux 云主机通过采集器的方式进行采集，以及采集容器标准输出或文件日志。

日志结构化解析

日志的结构化解析指云日志服务数据将以 key—value 对的形式存储在云日志服务平台上。日志数据结构化后，您可以在云日志服务控制台根据指定的键值进行日志检索、分析与加工。目前采集器提供多种解析方式，详情如下

解析方式	说明
单行全文	单行全文是指一条日志仅包含一行的内容，在采集的时候，将使用换行符来作为一条日志的结束符，即在日志文件中，以换行符分隔两条日志。日志数据本身不再进行日志结构化处理，也不会提取日志字段。每条日志都会存在一个默认的字段 message，采集器会将日志内容存放在 message 中。 详情请参考采集文本日志—单行全文模式
多行全文	多行全文日志是指一条完整的日志数据可能跨占多行，您需要指定首行正则已进行匹配，当某行日志匹配上预先设置的正则表达式，就认为是一条日志的开头，而下一个行首出现作为该条日志的结束标识符。日志内容同样也会存放在 message 字段中。 详情请参考采集文本日志—多行全文模式
单行正则	单行正则模式用于处理结构化的日志，针对包含一行内容的日志，您需要指定一个正则表达式，采集器按照正则表达式将一条完整日志提取多个值 详情请参考采集文本日志—单行正则模式

多行正则	多行正则模式用于处理结构化的日志，针对包含多行内容的日志，您需要指定一个行首正则表达式用于匹配日志的开头，并指定一个正则表达式用于提取多个值 详情请参考采集文本日志-多行正则模式
单行分隔符	单行分隔符模式支持通过分隔符将一条日志分割成多个值，从而实现结构化处理，该模式仅适用于单行日志，每条完整的日志以换行符为结束标识符。 详情请参考采集文本日志-多行分隔符模式
JSON	支持解析 Object 类型的 JSON 日志，提取 JSON 日志内容作为 Key-Value 对，即 Object 首层的键作为 Key，Object 首层的值作为 Value。 详情请参考采集文本日志-JSON 模式

配置流程

云日志服务提供配置向导，帮助您方便快捷完成日志接入，具体流程请查看采集文本日志

费用说明

公测期间日志采集不收取费用。

4.2.2. 采集器

4.2.2.1. 采集机制

文件监听

在目标云主机上安装采集器，并在控制台中创建并启用采集配置规则后，采集配置规则将会下发至采集器中，采集器会按照设置的采集配置规则以及日志路径/文件名，开始监听目标文件。

文件读取

若采集配置规则中的提取模式为全量模式，则从文件的开头开始读；若为增量模式，则只读取文件内新增的内容。

日志解析

采集器从目标文件中读取到日志后，将根据采集配置规则中切割模式进行分行与解析。切割模式支持单行与多行，并支持根据正则表达式、分隔符模式对日志内容进行切分，具体请查看采集文本日志。

4.2.2.2. 采集器安装（Linux 系统）

天翼云云日志服务提供专门的日志采集器，您需要将它部署安装至目标服务器上，可快速采集日志到云日志服务中。采集器安装操作需要在日志接入流程中完成，本文将介绍如何在目标云主机上进行采集器安装。

安装环境

- 仅支持天翼云云主机上安装
- 仅支持 64 位 Linux 操作系统环境（暂不支持 Windows），并适配主流 Linux 操作系统版本

前提条件

已在云日志服务控制台中新建日志项目与日志单元，详情请查看管理日志项目与管理日志单元。

安装步骤

1、登录云日志服务控制台。

2、在控制台概览页面的日志项目模块，点击目标日志项目名称。

3、在目标日志单元下，点击【数据接入】，若该日志单元初次配置日志采集，则页面会提供采集配置向导。

4、选择云主机

(1) 如果您还没有可用的主机组，请执行以下操作：

a. 输入主机组名称。

b. 在已开通云主机列表中，选择您需要采集日志的目标云主机，作为云主机组进行统一采集管理。

c. 点击下一步。

(2) 如果您已有可用的主机组，请点击【选择已有主机组】，选择目标主机组，并点击下一步：
注：目前仅支持采集天翼云 linux 云主机

5、安装采集器 。

(1) 安装采集器前，需要先接入云主机所在的 VPC 网络。在 Step1：VPC 接入中，点击【全部接入】按钮，稍等片刻后检查 VPC 的网络接入状态，确保列表中所有 VPC 都处于已接入状态。
注：若 VPC 无法接入，请点击对应 VPC 的【接入】按钮重试，若重试仍然失败，请查看“常见问题–VPC 接入失败”进行问题排查。

(2) 复制 Step2 中提供的采集器安装命令，在目标云主机中执行该命令安装采集器。

(3) 安装完成后需要检查采集器状态。
在 Step3：检查采集器状态中，点击【开始检测】按钮，稍等片刻后查看采集器状态，确保所有目标云主机的采集器状态均为“运行中”。

注：若采集器无法连通，请点击对应云主机的【重新检测】按钮重试，若重试仍然失败，请查看“常见问题–云主机采集器无法连通”进行问题排查，或点击【移除主机】从主机组中移除该主机。

4.2.3. 采集文本日志

4.2.3.1. 单行全文模式

概述

单行全文是指一条日志仅包含一行的内容，在采集的时候，将使用换行符来作为一条日志的结束符，即在日志文件中，以换行符分隔两条日志。日志数据本身不再进行日志结构化处理，也不会提取日志字段，每条日志都被作为一个整体被采集到云日志服务中，日志采集流程简单快捷。本文介绍如何通过云日志服务控制台创建单行全文模式的采集配置

前提条件

- 已创建日志项目与日志单元。详情请查看管理日志项目与管理日志单元
- 目标服务器持续产生日志

背景信息

在单行全文模式下，日志数据本身不再进行日志结构化处理，采集器会将日志内容存放在 message 字段中。如采集网站访问日志：

- 原始日志：

```
192.168.1.100 - - [24/Aug/2023:15:42:18 +0000] "GET /example-page HTTP/1.1" 200 1265 "https://www.example.com" "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.159 Safari/537.36"
```

- 采集到云日志服务后的日志：

```
message: 192.168.1.100 - - [24/Aug/2023:15:42:18 +0000] "GET /example-page HTTP/1.1" 200 1265 "https://www.example.com" "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.159 Safari/537.36"
```

操作步骤

1、登录云日志服务控制台

2、在控制台概览页面的日志项目模块，点击目标日志项目名称。

3、在目标日志单元下，点击【数据接入】，若该日志单元初次配置日志采集，则页面会提供采集配置向导。

4、选择云主机

(1) 如果您还没有可用的主机组，请执行以下操作

a. 输入主机组名称

b. 在已开通云主机列表中，选择您需要采集日志的目标云主机，作为云主机组进行统一采集管理。

c. 点击下一步

(2) 如果您已有可用的主机组，请点击【选择已有主机组】，选择目标主机组，并点击下一步

注：目前仅支持采集天翼云 linux 云主机

5、安装采集器

(1) 安装采集器前，需要先接入云主机所在的 VPC 网络。在 Step1：VPC 接入中，点击【全部接入】按钮，稍等片刻后检查 VPC 的网络接入状态，确保列表中所有 VPC 都处于已接入状态。
注：若 VPC 无法接入，请点击对应 VPC 的【接入】按钮重试，若重试仍然失败，请查看“常见问题—VPC 接入失败”进行问题排查

(2) 复制 Step2 中提供的采集器安装命令，在目标云主机中执行该命令安装采集器。

(3) 安装完成后需要检查采集器状态。
在 Step3：检查采集器状态中，点击【开始检测】按钮，稍等片刻后查看采集器状态，确保所有目标云主机的采集器状态均为“运行中”。

注：若采集器无法连通，请点击对应云主机的【重新检测】按钮重试，若重试仍然失败，请查看“常见问题—云主机采集器无法连通”进行问题排查，或点击【移除主机】从主机组中移除该主机

(4) 点击下一步

6、创建采集配置，请按如下说明配置参数。

参数	描述
采集规则名称	采集规则的名称，在所属的日志项目内唯一，
采集路径	根据日志在服务器上的位置，设置日志目录和文件名称。 日志路径必须以正斜线（/）开头，其中目录名和文件名支持完整名称和通配符模式。如 ● /var/log/auth.log：表示/var/log 目录下的 auth.log 日志文件 ● /var/log/*.log：表示/var/log 目录下后缀名为.log 的日志文件 ● /var/log/app_*/**/*.log：表示/var/log 目录下符合 app_*格式的目录中后缀名为.log 的日志文件
采集策略	● 全量：从目标文件的第一行日志开始采集 ● 增量：从采集配置下发的时间对应的日志开始采集，如果下发采集配置后，日志文件无更新，则不会采集该文件中的日志
切割模式	针对原始日志执行分词的模式，选择“单行全文”

7、参数配置完成后，点击【保存并启用】，表示创建采集配置并开始采集日志。点击【保存不启用】，则表示只创建采集配置但暂不下发启用，您稍后可在数据接入管理页面进行启用，详情请查看 采集配置管理

8、在完成页面，点击【日志检索】，将跳转至日志检索分析页面。详情请查看查询与分析日志

4.2.3.2. 多行全文模式

概述

多行全文日志是指一条完整的日志数据可能跨占多行，您需要指定首行正则表达式以进行匹配，当某行日志匹配上预先设置的正则表达式，则认为是一条日志的开头，而下一个行首出现作为该条日志的结束标识符。提取的日志内容同样也会存放在 message 字段中。本文介绍如何通过云日志服务控制台创建多行全文模式的采集配置。

前提条件

- 已创建日志项目与日志单元。详情请查看管理日志项目与管理日志单元
- 目标服务器持续产生日志

背景信息

在多行全文模式下，日志数据本身不再进行日志结构化处理，采集器会将日志内容存放在 message 字段中。如您需要采集的原始数据为：

- 原始日志：

```
“2019-12-15 17:13:06,043 [main] ERROR
com.test.logging.FooFactory:
java.lang.NullPointerException

    at com.test.logging.FooFactory.createFoo(FooFactory.java:15)
    at
com.test.logging.FooFactoryTest.test(FooFactoryTest.java:11)”
```

- 行首正则表达式：

```
\d+-\d+-\d+\s\d+:\d+:\d+,\d+\s.*
```

- 采集到云日志服务后的日志：

```
message:“2019-12-15 17:13:06,043 [main] ERROR
com.test.logging.FooFactory:
java.lang.NullPointerException    at
com.test.logging.FooFactory.createFoo(FooFactory.java:15)    at
com.test.logging.FooFactoryTest.test(FooFactoryTest.java:11)”
```

操作步骤

1、登录云日志服务控制台

2、在控制台概览页面的日志项目模块，点击目标日志项目名称。

3、在目标日志单元下，点击【数据接入】，若该日志单元初次配置日志采集，则页面会提供采集配置向导。

4、选择云主机

(1) 如果您还没有可用的主机组，请执行以下操作

a. 输入主机组名称

b. 在已开通云主机列表中，选择您需要采集日志的目标云主机，作为云主机组进行统一采集管理。

c. 点击下一步

(2) 如果您已有可用的主机组，请点击【选择已有主机组】，选择目标主机组，并点击下一步
注：目前仅支持采集天翼云 linux 云主机

5、安装采集器

(1) 安装采集器前，需要先接入云主机所在的 VPC 网络。在 Step1：VPC 接入中，点击【全部接入】按钮，稍等片刻后检查 VPC 的网络接入状态，确保列表中所有 VPC 都处于已接入状态。
注：若 VPC 无法接入，请点击对应 VPC 的【接入】按钮重试，若重试仍然失败，请查看“常见问题—VPC 接入失败”进行问题排查

(2) 复制 Step2 中提供的采集器安装命令，在目标云主机中执行该命令安装采集器。

(3) 安装完成后需要检查采集器状态。
在 Step3：检查采集器状态中，点击【开始检测】按钮，稍等片刻后查看采集器状态，确保所有目标云主机的采集器状态均为“运行中”。

注：若采集器无法连通，请点击对应云主机的【重新检测】按钮重试，若重试仍然失败，请查看“常见问题—云主机采集器无法连通”进行问题排查，或点击【移除主机】从主机组中移除该主机

(4) 点击下一步

6、创建采集配置，请按如下说明配置参数。

参数	描述
采集规则名称	采集规则的名称，在所属的日志项目内唯一，
采集路径	根据日志在服务器上的位置，设置日志目录和文件名称。 日志路径必须以正斜线（/）开头，其中目录名和文件名支持完整名称和通配符模式。如 ● /var/log/auth.log：表示/var/log 目录下的 auth.log 日志文件 ● /var/log/*.log：表示/var/log 目录下后缀名为.log 的日志文件 ● /var/log/app_*/**/*.log：表示/var/log 目录下符合 app_*格式的目录中后缀名为.log 的日志文件
采集策略	● 全量：从目标文件的第一行日志开始采集 ● 增量：从采集配置下发的时间对应的日志开始采集，如果下发采集配置后，日志文件无更新，则不会采集该文件中的日志
切割模式	针对原始日志执行分词的模式，选择“多行全文”

日志样例	输入您需要采集的日志样例
首行正则表达式	首行正则表达式用于匹配每一条日志的行首，以确认每条日志的开头位置。输入完成后，点击【验证】，系统将根据您输入的日志样例判断表达式是否通过以及成功解析的日志条数

7、参数配置完成后，点击【保存并启用】，表示创建采集配置并开始采集日志。点击【保存不启用】，则表示只创建采集配置但暂不下发启用，您稍后可在数据接入管理页面进行启用，详情请查看 采集配置管理

8、在完成页面，点击【日志检索】，将跳转至日志检索分析页面。详情请查看查询与分析日志

4.2.3.3. 单行分隔符模式

概述

单行分隔符模式支持通过配置的分隔符将一条日志分割成多个 key-value 键值，从而实现结构化处理，该模式仅适用于单行日志，每条完整的日志以换行符为结束标识符。本文介绍如何通过云日志服务控制台创建单行分隔符模式的采集配置。

前提条件

- 已创建日志项目与日志单元。详情请查看管理日志项目与管理日志单元
- 目标服务器持续产生日志

背景信息

如您需要采集的原始数据为：

- 原始日志：

```
10.21.21.1 - ::: [Fri Dec 22 15:49:33 CST 2021 +0800] ::: GET
/online/sample HTTP/1.1 ::: 127.0.0.1 ::: 200 ::: 647 ::: 35 :::
http://127.0.0.1/
```

- 若指定的分隔符为":::"，则该日志会被分割成 8 个字段，您可为这 8 个字段指定对应的 key，如下所示：

```
IP: 10.21.21.1 -
bytes: 35
host: 127.0.0.1
length: 647
referer: http://127.0.0.1/
request: GET /online/sample HTTP/1.1
status: 200
time: [Fri Dec 22 15:49:33 CST 2021 +0800]
message: 10.21.21.1 - ::: [Fri Dec 22 15:49:33 CST 2021 +0800] :::
GET /online/sample HTTP/1.1 ::: 127.0.0.1 ::: 200 ::: 647 ::: 35 :::
http://127.0.0.1/
:
```

操作步骤

1、登录云日志服务控制台

2、在控制台概览页面的日志项目模块，点击目标日志项目名称。

3、在目标日志单元下，点击【数据接入】，若该日志单元初次配置日志采集，则页面会提供采集配置向导。

4、选择云主机

(1) 如果您还没有可用的主机组，请执行以下操作

a. 输入主机组名称

b. 在已开通云主机列表中，选择您需要采集日志的目标云主机，作为云主机组进行统一采集管理。

c. 点击下一步

(2) 如果您已有可用的主机组，请点击【选择已有主机组】，选择目标主机组，并点击下一步
注：目前仅支持采集天翼云 linux 云主机

5、安装采集器

(1) 安装采集器前，需要先接入云主机所在的 VPC 网络。在 Step1：VPC 接入中，点击【全部接入】按钮，稍等片刻后检查 VPC 的网络接入状态，确保列表中所有 VPC 都处于已接入状态。
注：若 VPC 无法接入，请点击对应 VPC 的【接入】按钮重试，若重试仍然失败，请查看“常见问题—VPC 接入失败”进行问题排查

(2) 复制 Step2 中提供的采集器安装命令，在目标云主机中执行该命令安装采集器。

(3) 安装完成后需要检查采集器状态。
在 Step3：检查采集器状态中，点击【开始检测】按钮，稍等片刻后查看采集器状态，确保所有目标云主机的采集器状态均为“运行中”。

注：若采集器无法连通，请点击对应云主机的【重新检测】按钮重试，若重试仍然失败，请查看“常见问题—云主机采集器无法连通”进行问题排查，或点击【移除主机】从主机组中移除该主机

(4) 点击下一步

6、创建采集配置，请按如下说明配置参数。

参数	描述
采集规则名称	采集规则的名称，在所属的日志项目内唯一，
采集路径	根据日志在服务器上的位置，设置日志目录和文件名称。 日志路径必须以正斜线（/）开头，其中目录名和文件名支持完整名称和通配符模式。如 ● /var/log/auth.log：表示/var/log 目录下的 auth.log 日志文件 ● /var/log/*.log：表示/var/log 目录下后缀名为.log 的日志文件 ● /var/log/app_*/**/*.log：表示/var/log 目录下符合 app_* 格式的目录中后缀名为.log 的日志文件

采集策略	● 全量：从目标文件的第一行日志开始采集 ● 增量：从采集配置下发的时间对应的日志开始采集，如果下发采集配置后，日志文件无更新，则不会采集该文件中的日志
切割模式	针对原始日志执行分词的模式，选择“单行分隔符”
日志样例	输入您需要采集的日志样例
分隔符	选择空格、竖线或输入其他自定义符号（数字、英文和中文）
日志提取内容	选择分隔符后，系统将自动根据分隔符对日志样例进行切割，结果会展示在日志提取内容中，您需要为每个字段定义唯一的key。

7、参数配置完成后，点击【保存并启用】，表示创建采集配置并开始采集日志。点击【保存不启用】，则表示只创建采集配置但暂不下发启用，您稍后可在数据接入管理页面进行启用，详情请查看 采集配置管理

8、在完成页面，点击【日志检索】，将跳转至日志检索分析页面。详情请查看查询与分析日志

4.2.3.4. 单行正则模式

概述

单行正则模式用于处理结构化的日志，针对仅包含一行内容的日志，您需要指定一个正则表达式，采集器按照正则表达式将一条完整日志提取为多个 key-value 键值。本文介绍如何通过云日志服务控制台创建单行正则模式的采集配置。

前提条件

- 已创建日志项目与日志单元。详情请查看管理日志项目与管理日志单元
- 目标服务器持续产生日志

背景信息

如您需要采集的原始数据为：

- 原始日志：

```
10.145.32.100 - - [17/May/2023:13:21:30 +0800] "GET /my/course/1
HTTP/1.1" 127.0.0.1 200 782 9703
"http://127.0.0.1/course/explore?filter%5Btype%5D=all&filter%5Bpr
ice%5D=all&filter%5BcurrentLevelId%5D=all&orderBy=studentNum"
"Mozilla/5.0 (Windows NT 10.0; WOW64; rv:64.0) Gecko/20100101
Firefox/64.0"  0.354 0.354
```

- 配置自定义表达式为：

```
(\S+)[^\[]+(\[[^:]+:\d+:\d+:\d+\s\S+)\s"(\w+)\s(\S+)\s([^"]+)"\s(
\S+)\s(\d+)\s(\d+)\s(\d+)\s"([^"]+)"\s"([^"]+)"\s+(\S+)\s(\S+).*
```

- 系统将根据正则表达式提取键值对，您需要为每个提取出来的值指定 key 名称，如下所示：

```
body_bytes_sent: 9703
http_host: 127.0.0.1
http_protocol: HTTP/1.1
http_referer:
http://127.0.0.1/course/explore?filter%5Btype%5D=all&filter%5Bpri
ce%5D=all&filter%5BcurrentLevelId%5D=all&orderBy=studentNum
http_user_agent: Mozilla/5.0 (Windows NT 10.0; WOW64; rv:64.0)
Gecko/20100101 Firefox/64.0
remote_addr: 10.145.32.100
request_length: 782
request_method: GET
request_time: 0.354
request_url: /my/course/1
status: 200
time_local: [17/May/2023:13:21:30 +0800]
upstream_response_time: 0.354
message: 10.145.32.100 - - [17/May/2023:13:21:30 +0800] "GET
/my/course/1 HTTP/1.1" 127.0.0.1 200 782 9703
"http://127.0.0.1/course/explore?filter%5Btype%5D=all&filter%5Bpr
ice%5D=all&filter%5BcurrentLevelId%5D=all&orderBy=studentNum"
"Mozilla/5.0 (Windows NT 10.0; WOW64; rv:64.0) Gecko/20100101
Firefox/64.0"  0.354 0.354
```

操作步骤

1、登录云日志服务控制台

2、在控制台概览页面的日志项目模块，点击目标日志项目名称。

3、在目标日志单元下，点击【数据接入】，若该日志单元初次配置日志采集，则页面会提供采

集配置向导。

4、选择云主机

(1) 如果您还没有可用的主机组，请执行以下操作

a. 输入主机组名称

b. 在已开通云主机列表中，选择您需要采集日志的目标云主机，作为云主机组进行统一采集管理。

c. 点击下一步

(2) 如果您已有可用的主机组，请点击【选择已有主机组】，选择目标主机组，并点击下一步
注：目前仅支持采集天翼云 linux 云主机

5、安装采集器

(1) 安装采集器前，需要先接入云主机所在的 VPC 网络。在 Step1：VPC 接入中，点击【全部接入】按钮，稍等片刻后检查 VPC 的网络接入状态，确保列表中所有 VPC 都处于已接入状态。
注：若 VPC 无法接入，请点击对应 VPC 的【接入】按钮重试，若重试仍然失败，请查看“常见问题—VPC 接入失败”进行问题排查

(2) 复制 Step2 中提供的采集器安装命令，在目标云主机中执行该命令安装采集器。

(3) 安装完成后需要检查采集器状态。
在 Step3：检查采集器状态中，点击【开始检测】按钮，稍等片刻后查看采集器状态，确保所有目标云主机的采集器状态均为“运行中”。

注：若采集器无法连通，请点击对应云主机的【重新检测】按钮重试，若重试仍然失败，请查看“常见问题—云主机采集器无法连通”进行问题排查，或点击【移除主机】从主机组中移除该主机

(4) 点击下一步

6、创建采集配置，请按如下说明配置参数。

参数	描述
采集规则名称	采集规则的名称，在所属的日志项目内唯一，
采集路径	根据日志在服务器上的位置，设置日志目录和文件名称。 日志路径必须以正斜线（/）开头，其中目录名和文件名支持完整名称和通配符模式。如 ● /var/log/auth.log：表示/var/log 目录下的 auth.log 日志文

	件 ● /var/log/*.log：表示/var/log 目录下后缀名为.log 的日志文件 ● /var/log/app_*/**/*.log：表示/var/log 目录下符合 app_*格式的目录中后缀名为.log 的日志文件
采集策略	● 全量：从目标文件的第一行日志开始采集 ● 增量：从采集配置下发的时间对应的日志开始采集，如果下发采集配置后，日志文件无更新，则不会采集该文件中的日志
切割模式	针对原始日志执行分词的模式，选择“单行正则”
日志样例	输入您需要采集的日志样例
正则表达式	输入正则表达式，点击【验证】按钮，系统将根据您输入的正则表达式对日志样例进行字段切割
日志提取内容	根据正则表达式切割的结果会展示在日志提取内容中，您需要为每个字段定义唯一的 key。

7、参数配置完成后，点击【保存并启用】，表示创建采集配置并开始采集日志。点击【保存不启用】，则表示只创建采集配置但暂不下发启用，您稍后可在数据接入管理页面进行启用，详情请查看 采集配置管理

8、在完成页面，点击【日志检索】，将跳转至日志检索分析页面。详情请查看查询与分析日志

4.2.3.5. 多行正则模式

概述

多行正则模式用于处理结构化的日志，针对包含多行内容的日志，您需要指定一个行首正则表达式用于匹配日志的开头，并指定一个正则表达式用于提取多个值，采集器按照该正则表达式将一条完整日志提取为多个 key-value 键值。本文介绍如何通过云日志服务控制台创建多行正则模式的采集配置。

前提条件

- 已创建日志项目与日志单元。详情请查看管理日志项目与管理日志单元

- 目标服务器持续产生日志

背景信息

如您需要采集的原始数据为：

- 原始日志：

```
[2023-04-02T14:29:01,000] [INFO] java.lang.Exception: exception
happened
    at TestPrintStackTrace.f(TestPrintStackTrace.java:3)
    at TestPrintStackTrace.g(TestPrintStackTrace.java:7)
    at TestPrintStackTrace.main(TestPrintStackTrace.java:16)
```

- 配置行首正则表达式为：

```
\[\d+-\d+-\w+:\d+:\d+,\d+]\s\[\w+]\s.*
```

- 配置自定义正则表达式为：

```
\[(\d+-\d+-\w+:\d+:\d+,\d+)\]\s\[(\w+)\]\s(.*)
```

- 系统将根据行首正则表达式匹配每条日志的开头，并根据自定义正则表达式提取键值对，您需要为每个提取出来的值指定 key 名称，如下所示：

```
time:   2023-04-02T14:29:01,000`
level:  INFO`
msg: java.lang.Exception: exception happened
    at TestPrintStackTrace.f(TestPrintStackTrace.java:3)
    at TestPrintStackTrace.g(TestPrintStackTrace.java:7)
    at TestPrintStackTrace.main(TestPrintStackTrace.java:16)
```

操作步骤

1、登录云日志服务控制台

2、在控制台概览页面的日志项目模块，点击目标日志项目名称。

3、在目标日志单元下，点击【数据接入】，若该日志单元初次配置日志采集，则页面会提供采集配置向导。

4、选择云主机

(1) 如果您还没有可用的主机组，请执行以下操作

a. 输入主机组名称

b. 在已开通云主机列表中，选择您需要采集日志的目标云主机，作为云主机组进行统一采集管理。

c. 点击下一步

(2) 如果您已有可用的主机组，请点击【选择已有主机组】，选择目标主机组，并点击下一步

注：目前仅支持采集天翼云 linux 云主机

5、安装采集器

(1) 安装采集器前，需要先接入云主机所在的 VPC 网络。在 Step1：VPC 接入中，点击【全部接入】按钮，稍等片刻后检查 VPC 的网络接入状态，确保列表中所有 VPC 都处于已接入状态。
注：若 VPC 无法接入，请点击对应 VPC 的【接入】按钮重试，若重试仍然失败，请查看“常见问题—VPC 接入失败”进行问题排查

(2) 复制 Step2 中提供的采集器安装命令，在目标云主机中执行该命令安装采集器。

(3) 安装完成后需要检查采集器状态。
在 Step3：检查采集器状态中，点击【开始检测】按钮，稍等片刻后查看采集器状态，确保所有目标云主机的采集器状态均为“运行中”。

注：若采集器无法连通，请点击对应云主机的【重新检测】按钮重试，若重试仍然失败，请查看“常见问题—云主机采集器无法连通”进行问题排查，或点击【移除主机】从主机组中移除该主机

(4) 点击下一步

6、创建采集配置，请按如下说明配置参数。

参数	描述
采集规则名称	采集规则的名称，在所属的日志项目内唯一，
采集路径	根据日志在服务器上的位置，设置日志目录和文件名称。 日志路径必须以正斜线（/）开头，其中目录名和文件名支持完整名称和通配符模式。如 ● /var/log/auth.log：表示/var/log 目录下的 auth.log 日志文件 ● /var/log/*.log：表示/var/log 目录下后缀名为.log 的日志文件 ● /var/log/app_*/**/*.log：表示/var/log 目录下符合 app_*格式的目录中后缀名为.log 的日志文件
采集策略	● 全量：从目标文件的第一行日志开始采集 ● 增量：从采集配置下发的时间对应的日志开始采集，如果下发采集配置后，日志文件无更新，则不会采集该文件中的日志
切割模式	针对原始日志执行分词的模式，选择“多行正则”

日志样例	输入您需要采集的日志样例
首行正则表达式	首行正则表达式用于匹配每一条日志的行首，以确认每条日志的开头位置。输入完成后，点击【验证】，系统将根据您输入的日志样例判断表达式是否通过以及成功解析的日志条数
正则表达式	输入正则表达式，点击【验证】按钮，系统将根据您输入的正则表达式对日志样例进行字段切割
日志提取内容	根据正则表达式切割的结果会展示在日志提取内容中，您需要为每个字段定义唯一的 key。

7、参数配置完成后，点击【保存并启用】，表示创建采集配置并开始采集日志。点击【保存不启用】，则表示只创建采集配置但暂不下发启用，您稍后可在数据接入管理页面进行启用，详情请查看 采集配置管理

8、在完成页面，点击【日志检索】，将跳转至日志检索分析页面。详情请查看查询与分析日志

4.2.3.6. JSON 模式

概述

支持解析 Object 类型的 JSON 日志，提取 JSON 日志内容作为 Key-Value 对，即 Object 首层的键作为 Key，Object 首层的值作为 Value。

前提条件

- 已创建日志项目与日志单元。详情请查看管理日志项目与管理日志单元
- 目标服务器持续产生日志

背景信息

如您需要采集的原始数据为：

- 原始日志：

```
{"remote_ip":"10.133.47.111","time_local":"12/Jan/2023:21:33:18
+0800","body_sent":23,"responsetime":0.232,"upstreamtime":"0.232"
,"upstreamhost":"unix:/tmp/php-cgi.sock","http_host":"127.0.0.1",
"method":"POST","url":"/event/dispatch","request":"POST
/event/dispatch
HTTP/1.1","xff":"-","referer":"http://127.0.0.1/my/course/4","age
nt":"Mozilla/5.0 (Windows NT 10.0; WOW64; rv:64.0) Gecko/20100101
Firefox/64.0","response_code":"200"}
```


- 经过结构化处理后，采集到云日志服务后的日志如下：

```
agent: Mozilla/5.0 (Windows NT 10.0; WOW64; rv:64.0) Gecko/20100101
Firefox/64.0
body_sent: 23
http_host: 127.0.0.1
method: POST
referer: http://127.0.0.1/my/course/4
remote_ip: 10.133.47.111
request: POST /event/dispatch HTTP/1.1
response_code: 200
responsetime: 0.232
time_local: 12/Jan/2023:21:33:18 +0800
upstreamhost: unix:/tmp/php-cgi.sock
upstreamtime: 0.232
url: /event/dispatch
xff: -
```

操作步骤

1、登录云日志服务控制台

2、在控制台概览页面的日志项目模块，点击目标日志项目名称。

3、在目标日志单元下，点击【数据接入】，若该日志单元初次配置日志采集，则页面会提供采集配置向导。

4、选择云主机

(1) 如果您还没有可用的主机组，请执行以下操作

a. 输入主机组名称

b. 在已开通云主机列表中，选择您需要采集日志的目标云主机，作为云主机组进行统一采集管理。

c. 点击下一步

(2) 如果您已有可用的主机组，请点击【选择已有主机组】，选择目标主机组，并点击下一步
注：目前仅支持采集天翼云 linux 云主机

5、安装采集器

(1) 安装采集器前，需要先接入云主机所在的 VPC 网络。在 Step1：VPC 接入中，点击【全部接入】按钮，稍等片刻后检查 VPC 的网络接入状态，确保列表中所有 VPC 都处于已接入状态。
注：若 VPC 无法接入，请点击对应 VPC 的【接入】按钮重试，若重试仍然失败，请查看“常见问题—VPC 接入失败”进行问题排查

(2) 复制 Step2 中提供的采集器安装命令，在目标云主机中执行该命令安装采集器。

(3) 安装完成后需要检查采集器状态。
在 Step3：检查采集器状态中，点击【开始检测】按钮，稍等片刻后查看采集器状态，确保所有目标云主机的采集器状态均为“运行中”。

注：若采集器无法连通，请点击对应云主机的【重新检测】按钮重试，若重试仍然失败，请查看“常见问题—云主机采集器无法连通”进行问题排查，或点击【移除主机】从主机组中移除该主机

(4) 点击下一步

6、创建采集配置，请按如下说明配置参数。

参数	描述
采集规则名称	采集规则的名称，在所属的日志项目内唯一，
采集路径	根据日志在服务器上的位置，设置日志目录和文件名称。 日志路径必须以正斜线（/）开头，其中目录名和文件名支持完整名称和通配符模式。如 ● /var/log/auth.log：表示/var/log 目录下的 auth.log 日志文件 ● /var/log/*.log：表示/var/log 目录下后缀名为.log 的日志文件 ● /var/log/app_*/**/*.log：表示/var/log 目录下符合 app_*格式的目录中后缀名为.log 的日志文件
采集策略	● 全量：从目标文件的第一行日志开始采集 ● 增量：从采集配置下发的时间对应的日志开始采集，如果下发采集配置后，日志文件无更新，则不会采集该文件中的日

	志
切割模式	针对原始日志执行分词的模式，选择“JSON”

7、参数配置完成后，点击【保存并启用】，表示创建采集配置并开始采集日志。点击【保存不启用】，则表示只创建采集配置但暂不下发启用，您稍后可在数据接入管理页面进行启用，详情请查看 采集配置管理

8、在完成页面，点击【日志检索】，将跳转至日志检索分析页面。详情请查看查询与分析日志

4.2.4. 采集容器日志

4.2.4.1. 通过 DaemonSet 方式采集容器日志

容器云服务引擎 CCSE 提供日志中心服务，本文介绍如何将应用服务的日志统一上报到云日志服务中心，从而进行检索查询。

前提条件

- 已开通云日志服务
- 已开通容器云服务引擎 CCSE 并创建容器集群，同时安装日志插件。详情请参考容器云服务引擎 CCSE 帮助文档 https://www.ctyun.cn/document/10083472/10102632

在发布应用时配置日志收集

以发布一个 tomcat 服务应用为例进行介绍，其他应用与此类似。

1、登录容器云服务引擎 CCSE 控制台。

2、点击目标集群名称，进入集群管理页面。

3、选择"工作负载"菜单栏下的"无状态"菜单，点击"新增"按钮。

4、工作负载名称、数据卷、伸缩方式、实例数量、实例内容器信息根据应用实际情况进行配置，详情请查看容器云服务引擎 CCSE 帮助文档。

https://www.ctyun.cn/document/10083472/10102312

5、点击“实例内容器”正下方的“显示高级设置”，展示出“Pod 注解”部分。

6、输入 Pod 注解信息

- 注解名：ctyun.sls.logs
- 注解值

```
[
    {
        "sls.capture.type":"stdout",
        "sls.app.name":"tomcat",
        "sls.container.name":"tomcat-demo",
        "sls.log.project":"ccse_ccse-als-test",

"sls.log.rule":"ccse_ccse-als-test_default_deployment_tomcat",

"sls.log.unit":"ccse_ccse-als-test_default_deployment_tomcat"
    }
]
```

(1) sls.log.project　sls.log.rule　sls.log.rule　等值需要根据集群名称调整

(2) sls.capture.type 是云日志服务的输出类型，包括标准输出：stdout、文件日志：applog，如果是文件日志还需要通过注解 key：sls.log.path 指定日志路径 。注解值示例是标准输出 stdout。

(3) 注解值是 json 数组形式，支持多容器实例日志采集。

(4) 工作负载类型根据实际情况填写，statefulset deployment 使用小写

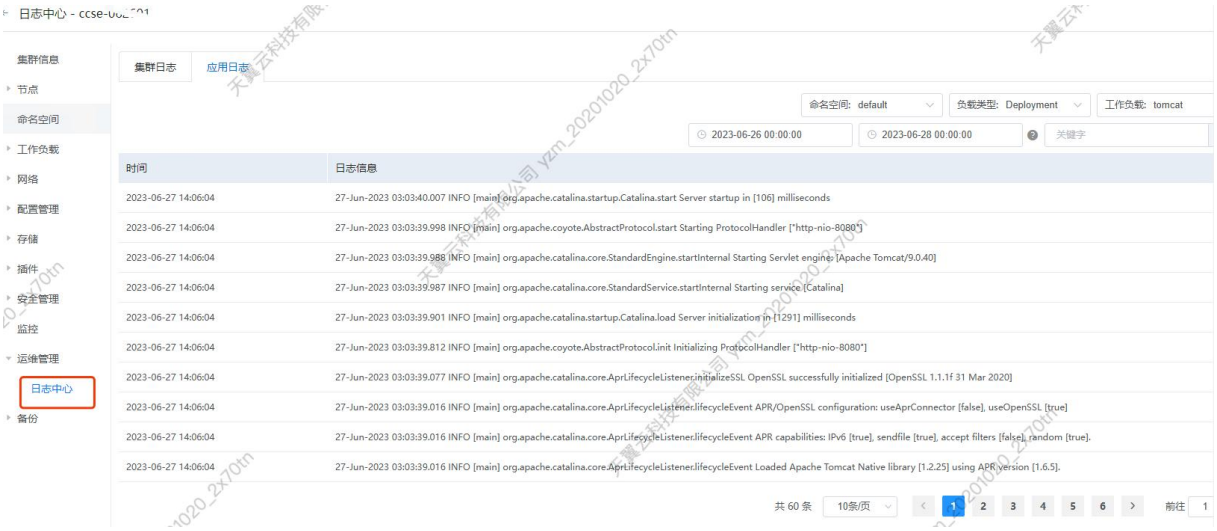
注解 key	拼写规则	参考示例
sls.log.project	ccse_ ＋ "ccse 集群名称"	ccse_“ccse 集群名称”
sls.log.rule	ccse_ ＋ "ccse 集群名称" ＋ “_” ＋ "命名空间" ＋ “_“+工作负载类型+”_” ＋ “应用名称”	ccse_“ccse 集群名称”_default_deployment_tomcat
sls.log.unit	ccse_ ＋ "ccse 集群名称" ＋ “_” ＋ "命名空间" ＋ “_“+工作负载类型+”_” ＋ “应用名称”	ccse_“ccse 集群名称”_default_deployment_tomcat
sls.app.name	应用的名字	tomcat
sls.container.name	容器实例的名字	tomcat—demo
sls.capture.type	容器日志的输出类型	stdout 或者 applog
sls.log.path	容器日志的输出类型为 applog 时，日志所在容器的路径	/path/to/your/app.log

7、提交发布应用

CCSE 控制台查看日志

1、点击左侧的运维管理，进入日志中心页面，选择“云日志服务”tab 页。

2、选择发布工作负载所用的命名空间和和工作负载名称，如果能检索出日志则日志功能正常。



云日志服务控制台查看日志

1、登录云日志服务控制台。

2、选择容器集群所对应的日志项目。在容器集群中安装插件时，将会默认创建名称为“k8s-log-{集群 id}”的日志项目。

3、选择目标日志单元，点击日志检索，即可查看采集的日志，并可进行检索与分析。

采集配置规则管理

在 CCSE 中完成日志采集配置后，您可在云日志服务控制台查看并管理相应的采集配置。选择目标日志项目与日志单元，在数据接入栏目下，即可查看并管理当前的采集规则。

4.2.4.2. 通过 MSAP 配置日志收集

在容器云服务引擎 CCSE 中创建或部署应用时，您可将业务文件日志、容器标准输出日志输入至云日志服务中进行检索、统计分析、加工、投递等操作，本文介绍 MSAP 发布应用时如何配置日志收集。

前提条件

- 已开通云日志服务

- 已开通容器云服务引擎 CCSE，并已安装日志插件。详情请参考容器云服务引擎 CCSE 帮助文档 https://www.ctyun.cn/document/10083472/10102632
- 已开通微服务云应用平台

在创建应用时配置日志收集

1、登录 MSAP 控制台，在左侧菜单栏，单击持续交付-应用管理，在应用列表页中点击【新增】按钮。

2、在选择应用分类页面中，选择微服务，点击下一步。

3、选择对应语言，点击下一步。

3、在应用基本信息页面，输入应用名称、接入方式等信息，具体请参考微服务云应用平台帮助文档 https://www.ctyun.cn/document/10038027/10087158。配置完成后点击下一步。

4、在应用配置页面，设置应用的部署环境、镜像信息、部署版本、资源参数等，设置完成后单击下一步。

5、在应用高级设置页面，展开日志收集管理，点击【开通日志收集到 ALS 云日志服务】，并根据您的业务需求设置日志收集信息。

注：采集模式默认为单行全文，若您需要指定其他切割模式，请查看下文修改采集配置。

配置项	描述
日志项目	在 CCSE 中安装日志插件时将自动根据 CCSE 集群名创建日志项目，采集的容器日志将存放在该日志项目下，不支持修改。
日志单元	输入日志单元的名称，如果对应的日志项目下不存在该日志单元，则会新建一个日志单元。
采集日志类型	您可选择文件日志（容器内日志路径）或容器标准输出日志。
采集路径	输入容器内的日志路径

6、设置完成后点击下一步。

7、在应用创建完成页面，确认应用基本信息、应用配置和应用高级设置等信息，确认完毕后点击确定创建并部署。

在更新应用时配置日志收集

1、登录 MSAP 控制台，在左侧菜单栏，单击监控运维-已发布应用，点击目标应用名称

2、在应用总览页面上方点击【部署】。

3、在**选择部署模式**页面，选择具体的部署方式，点击**【开始部署】。**

4、设置应用的环境和部署包信息，展开**日志收集管理**，点击**【开通日志收集到 ALS 云日志服务】**，并根据您的业务需求设置日志收集信息。然后单击【提交】。

注：采集模式默认为单行全文，若您需要指定其他切割模式，请查看下文修改采集配置

配置项	描述
日志项目	在安装日志插件时将自动创建日志项目，采集的容器日志将存放在该日志项目下，不支持修改
日志单元	输入日志单元的名称，如果对应的日志项目下不存在该日志单元，则会新建一个日志单元。
采集日志类型	您可选择文件日志（容器内日志路径）或容器标准输出日志。
采集路径	输入容器内的日志路径

结果验证

应用部署完成后，MSAP 将根据所配的日志收集规则收集日志并存放到指定的文件内。

1. 登录 MSAP 控制台。
2. 在左侧菜单栏，单击监控运维-已发布应用，点击目标应用名称。
3. 在左侧导航栏，单击**日志中心**，即可查看当前配置的日志采集规则。点击【查看日志】，跳转云日志服务控制台进行日志检索，若存在日志数据，则表示日志收集配置成功。您可对日志数据进行检索、统计分析、加工、投递等操作。
4. 您也可以直接登录云日志服务控制台，选择 CCSE 集群对应的日志项目，进入目标日志单元，即可进行日志数据进行查询统计、加工、投递等操作。

采集配置规则管理

在 MSAP 中完成日志采集配置后，您可在云日志服务控制台查看并管理相应的采集配置。

新增采集配置

1、登录云日志服务控制台。

2、在左侧菜单栏中点击【日志项目】，点击目标日志项目。

3、在左侧日志单元列表中选择目标日志单元，点击数据接入。

4、点击新增采集配置。

5、选择日志类型为标准输出或文件日志，设置其它基本信息（例如采集规则名称、采集路径、切割模式），上述配置项与采集服务器文本日志的配置相同，配置说明请参考采集文本日志

6、点击创建并启用 ，完成采集配置创建。创建完成后稍等片刻，即可在日志检索中查看采集的日志。

修改采集配置

1、在左侧日志单元列表中选择目标日志单元，点击数据接入。

2、选择目标采集配置规则，点击【修改】。

3、在修改页面，您可修改采集路径（标准输出无法修改采集路径）、采集策略、切割模式等配置项。

4、点击确定，完成采集配置修改。修改完成后稍等片刻，即可在日志检索中查看采集的日志。

删除采集配置

您可在云日志服务控制台删除容器日志采集规则，删除后将不再采集对应容器日志。

4.3. 查询与分析

4.3.1. 日志查询

4.3.1.1. 日志检索

云日志服务支持秒级查询上亿条日志数据，可快速定位目标字段并筛选日志查询结果，为具有复杂搜索功能和要求的应用程序提供支持。

查询基本操作

云日志服务提供简单易用的查询方式，服务使用快捷模式与交互模式作为查询接口，支持 INFO、DEBUG 等常用日志级别的快速查询。日志检索中可选择不同的日志单元以及当前日志单元下的采集规则，支持多选，默认为全选。

快捷模式：快速筛选列出包含目标日志字段或内容的语句并高亮显示。

交互模式：日志查询时对多个字段设置过滤规则，返回符合条件的日志。

操作方式

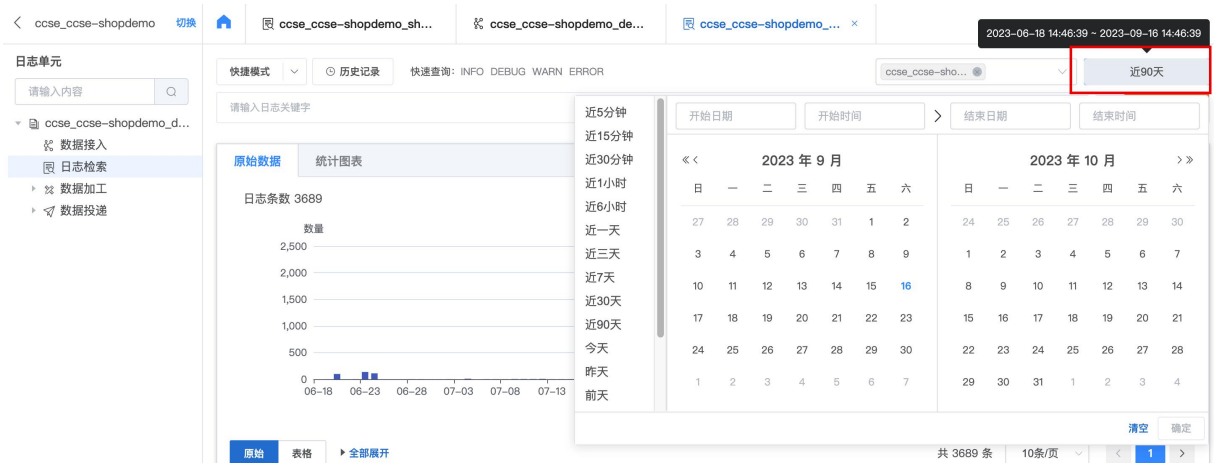
控制台方式：登录云日志服务控制台，选择日志项目、日志单元和采集规则，选择快捷模式或交互模式进行日志检索。

快捷模式

1. 登录云日志服务控制台。选择目标日志项目与日志单元，并点击日志检索进入检索页面。

2. 选择模式为快捷模式，在输入框中填写日志字段或内容，也可点击快速查询中常用日志字段如 INFO、DEBUG、WARN、ERROR 执行字段查询。

3. 设置查询时间范围，可快速选择近 30 分钟、近 1 小时等时间段，也可自定义时间段查询日志。

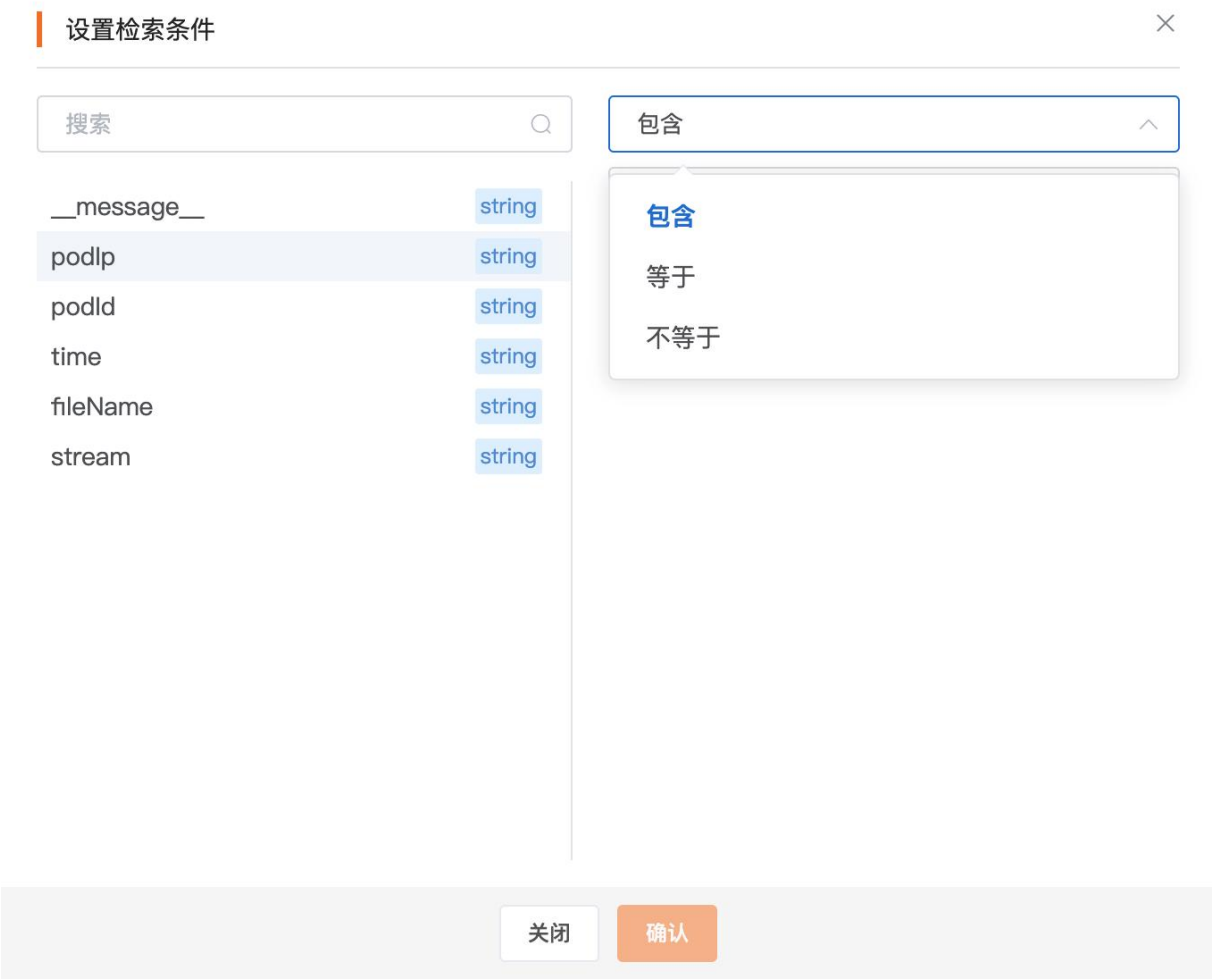


交互模式

1. 登录云日志服务控制台。选择目标日志项目与日志单元，并点击日志检索进入检索页面。

2. 查询模式选择交互模式。

3. 点击添加检索条件，弹出设置检索条件框。



4. 择字段名后可以设置三种查询规则，分别是：包含：当字段内容含有当前值时返回查询语句。

   a. 等于：当字段内容与当前值内容相同时返回查询语句。

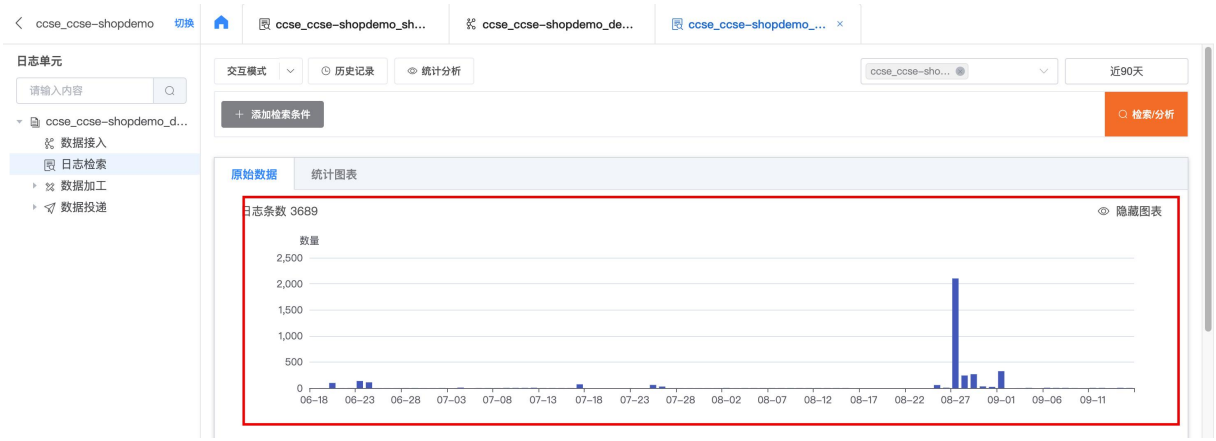
   b. 不等于：当字段内容与当前值内容不相同时返回查询语句。

5. 交互模式支持多字段查询，通过 and 或者 or 选择需要满足的字段条件，进行多字段条件查询。

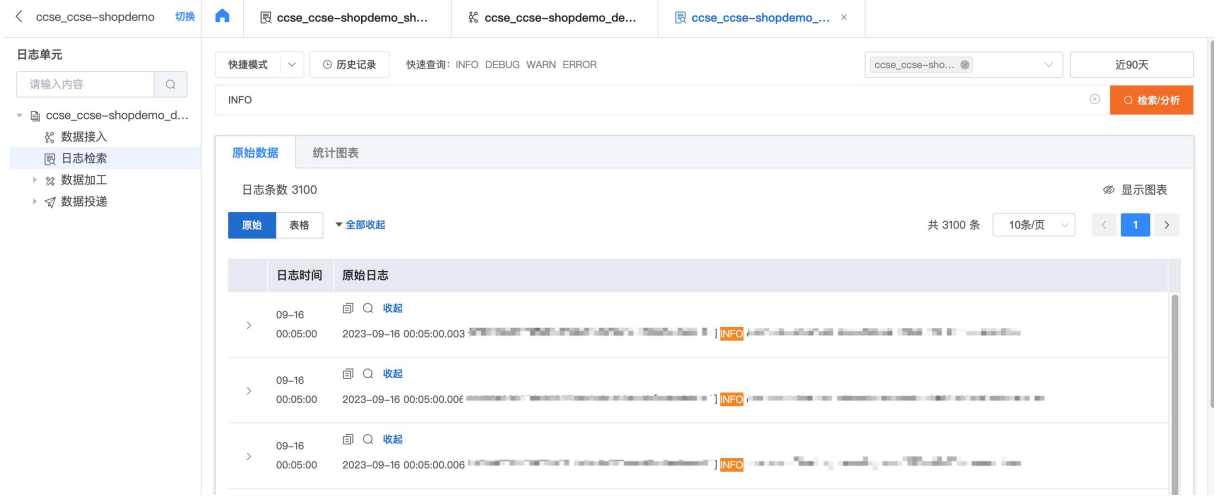
查询结果

云日志服务提供日志条数统计、原始数据与表格数据展示查询结果。

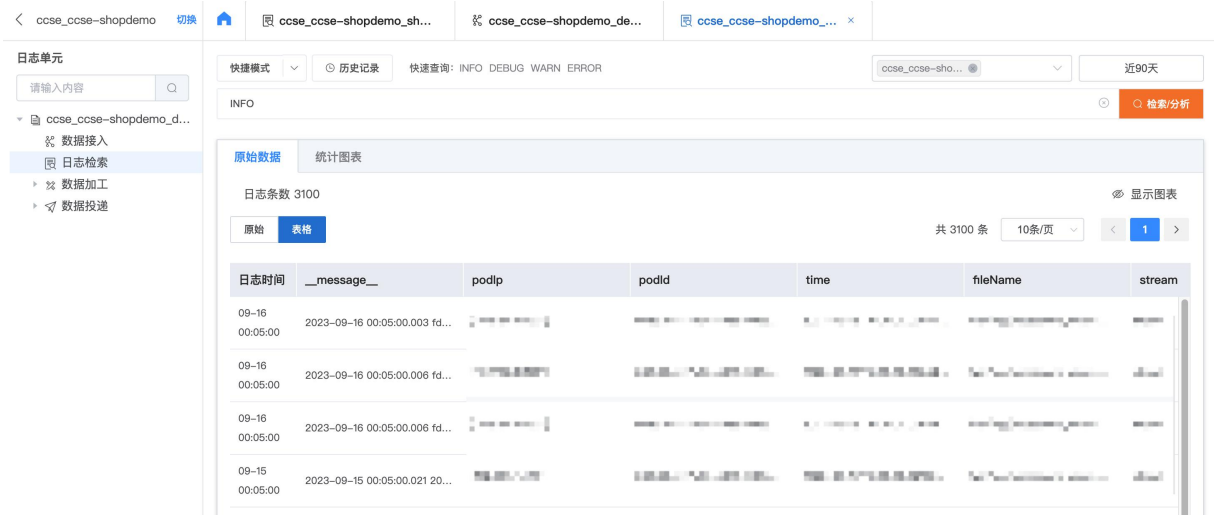
● 日志条数统计：原始数据的日志条数直方图主要展示查询到的日志在不同时间段的分布情况。将鼠标悬浮于数据块上时，可以查看该数据块代表的时间范围和日志命中次数。



● 原始日志：原始日志页签中展示当前查询的结果，点击原始日志可以查看日志详细信息。其中查询字段或内容被高亮，您可以基于查询结果快速定位到日志信息并做进一步的处理。



● 表格日志：表格日志页签展示当前查询结果的详细信息



4.3.1.2. 上下文查询

本小节介绍如何在云日志服务控制台查看指定日志在原始日志文件内的上下文信息。

前提条件

- 已创建日志项目、日志单元与采集规则
- 已采集到日志
-

背景信息

日志中的上下文查询是指定日志来源和其中一条日志，将该日志在原始文件中的前若干条（上文）或后若干条日志（下文）也同时检索出来。通过查看指定日志的上下文信息，您可以在业务故障排查过程中快速查找相关故障信息，方便您快速定位问题与解决各类故障。

功能优势

上下文查询的功能优势主要体现在日志数据的理解和分析。当您对日志进行检索时，提供上下文信息可以显著改进查询的效果和数据分析的质量。

1. 在进行故障排查时，上下文查询可以帮助您理解事件或错误发生的背景和前因后果。
2. 云日志服务具有独立性，无需改动日志文件格式也无需登录服务器即可查看文件的指定日志的上下文信息。
3. 无须担心由于服务器存储空间不足或日志文件覆盖造成的数据丢失，在云日志服务控制台可以随时查看历史数据。

操作步骤

1. 登录到云日志服务控制台。选择目标日志项目与日志单元，并点击日志检索进入检索页面。
2. 通过快捷模式或交互模式查询目标字段，例如 ERROR 信息。
3. 在原始页的原始日志标签下，找到目标日志并点击上下文检索图标。
4. 使用鼠标在当前页面上下滚动查看指定日志的上下文信息。选择时间范围，可以快速筛选前后时间段内的日志信息。
    a. 搜索框输入关键字可以高亮显示的字符串，便于观察需要的日志内容
    b. 提示：上下文查询最多加载五百条数据。

5. 您可以选择时间范围，并可自由查看当前日志上下文的日志内容。

4.3.2.日志分析

4.3.2.1. 概述

云日志服务提供分析功能，该功能结合了查询功能和统计分析功能。本文介绍分析功能的基本用法、使用限制等信息。

基础用法

统计分析基于快捷模式查询的结果，也可独立添加字段筛选场景，并由筛选场景结果做其他统计分析。

功能类型	说明
日志查询	使用日志查询提供的快捷模式与交互模式进行查询，返回符合条件的日志信息。
统计分析	添加筛选场景并完成统计分析，或对查询结果进行计算和统计。

日志分析功能

云日志服务支持以下统计分析场景。您可以使用控制台统计分析功能进行日志分析。

- 基础分析字段筛选&过滤
- 基础统计指标统计
  - 分组统计
- 高级统计日志占比
  - TopN
  - 时间趋势

4.3.2.2. 基础分析

4.3.2.2.1. 字段筛选&过滤

日志基础分析主要提供字段筛选&过滤功能，提供最基础的日志过滤操作，本文介绍该功能的基本用法。

函数原型

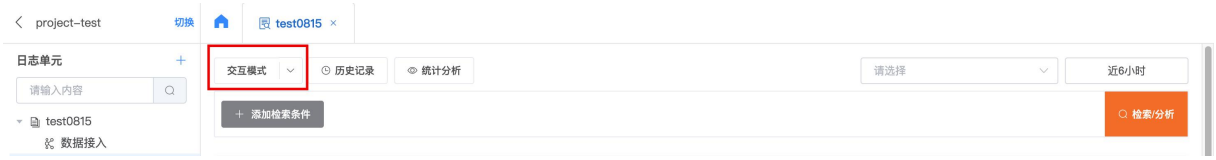
不同于繁琐的 SQL 语句，统计分析功能提取了 SQL 保留字并将业务字段提供给您用于快速筛选和过滤，该功能对应的原始 SQL 语句为：

```
SELECT field1, field2... (FROM log_table) where condition1,
condition2... ORDER BY field3 ASC/DESC, field4 ASC/DESC... LIMIT
number;
```

您可以自由填写 field1、field2、condition1、condition2、field3、number 等内容即可实现字段筛选&过滤。

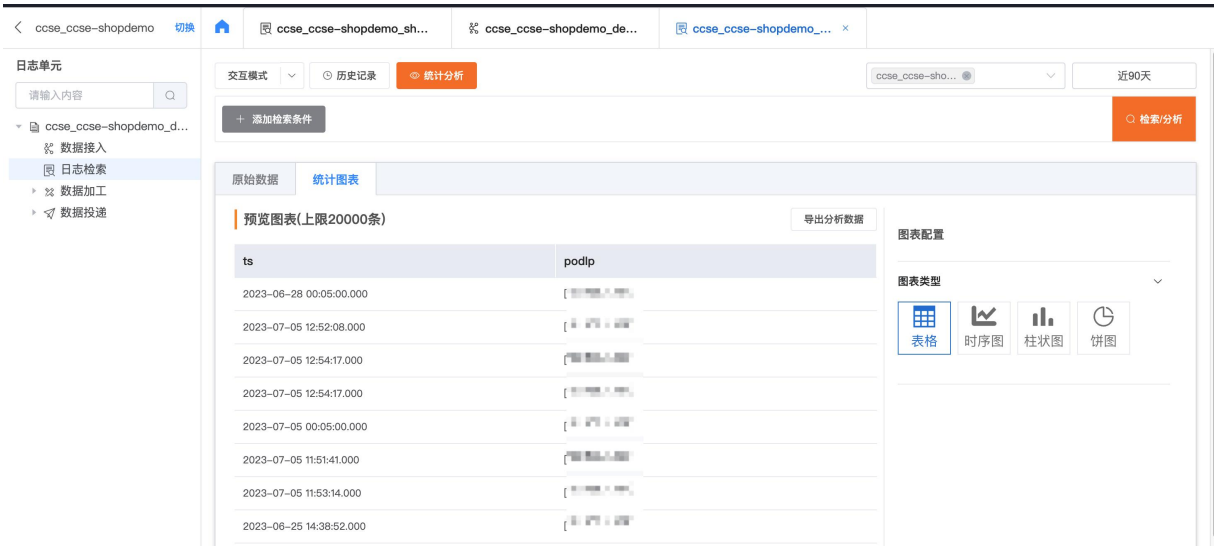
操作步骤

1. 登录云日志服务控制台。选择目标日志项目与日志单元，并点击日志检索进入检索页面。
2. 使用交互模式过滤日志（可选）。



3. 点击统计分析，选择基础分析中的字段筛选&过滤作为分析场景。

4. 设置条件，其中筛选字段为必填条件，其他条件可选填。
5. 设置完条件后点击检索分析，即可按条件进行基础分析。



4.3.2.3. 基础统计

4.3.2.3.1. 指标统计

基础统计提供指标统计与分组统计两个使用场景。本文介绍指标统计的基本用法。

支持的统计场景

- 日志条数总条数：统计日志总条数。
  - 字段值非空的条数：统计指定字段值不为 NULL 的日志条数。
  - 字段值非零的条数：统计指定字段值不为 0 的日志条数
- 日志字段不同字段值：统计指定字段的所有不同取值。
  - 不同字段值数量：统计指定字段的所有不同取值数量。
  - 随机字段值：给定字段随机挑选一个字段值。
- 数学计算最大值：计算指定字段的取值的最大值。
  - 最小值：计算指定字段取值的最小值。
  - 平均值：计算指定字段取值的平均值。

- 求和：计算指定字段所有取值的总和。
- 平方和：计算指定字段所有取值的平方和。
- 极差：计算指定字段的最大值和最小值的差值。

- 数学统计总体方差：统计指定字段的取值的总体方差。
  - 样本方差：统计指定字段的取值的样本方差。
  - 总体标准差：统计指定字段的取值的总体标准差。
  - 样本标准差：统计指定字段的取值的样本标准差。
- 估算函数中位数：估算指定字段的取值的中位数。即：对字段所有取值进行正序排列，返回大约处于50%位置的字段值
  - 不同字段值数量：估算给定字段所有不同取值的数量，默认存在2.3%的标准误差。

操作步骤

1. 登录云日志服务控制台。选择目标日志项目与日志单元，并点击日志检索进入检索页面。
2. 使用交互模式过滤日志（可选）。
3. 点击统计分析，您可以选择基于字段筛选&过滤场景添加指标统计场景，也可以直接选择指标统计场景。以统计字段值非空的条数为例，如下图所示为具体场景。

4. 点击检索/分析即可获得统计结果。

如上图所示为字段值非空的条数，您也可以继续添加场景或选择其他统计函数对日志查询结果做进一步分析。

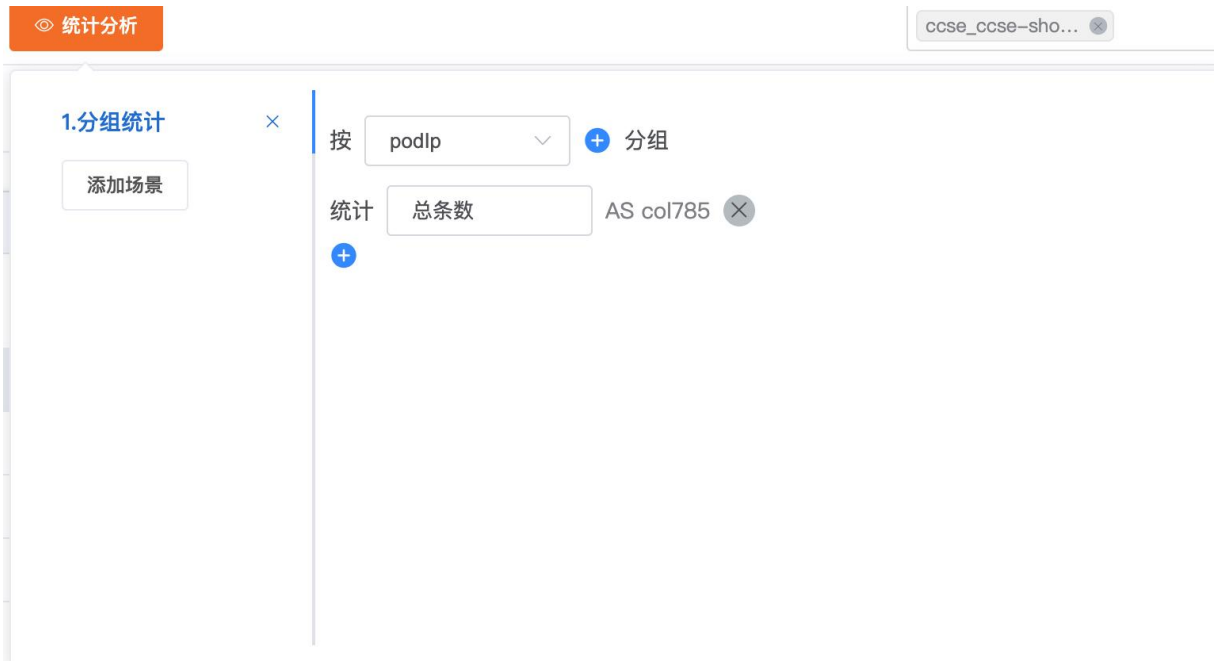
4.3.2.3.2. 分组统计

基础统计提供指标统计与分组统计两个使用场景。本文介绍分组统计的基本用法。

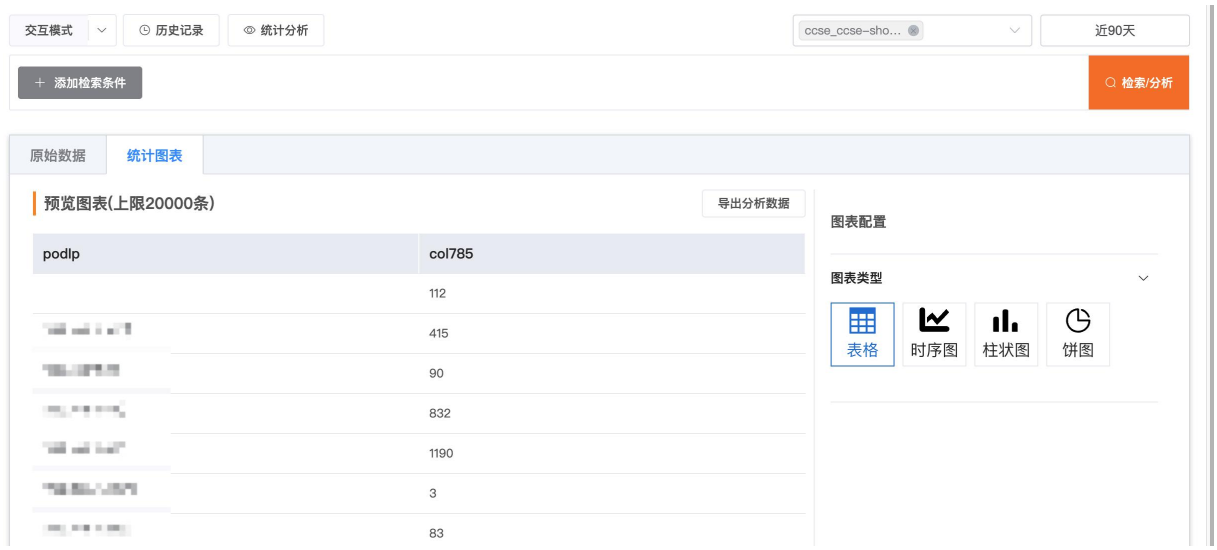
与指标统计类似，分组统计可以在选定多个字段后进行指标统计。

操作步骤

1. 登录云日志服务控制台。选择目标日志项目与日志单元，并点击日志检索进入检索页面。
2. 使用交互模式过滤日志（可选）。
3. 点击统计分析，您可以选择基于字段筛选&过滤场景添加分组统计场景，也可以直接选择分组统计场景。如下图所示，点击请选择框可以后的下拉选项为当前可进行分组的字段，点击 ⊕ 按钮可添加多个分组字段，鼠标移入也可移除特定字段。每个分组字段均为必填。
4. 以统计不同 podip 的总条数为例，如下图所示为具体场景。



5. 应用上图所示分组统计条件，下图为搜索结果。被分组字段与指标统计字段作为结果列展示在统计图表中。



4.3.2.4. 高级统计

4.3.2.4.1. 日志占比

日志占比用于统计单个或多个经由字段过滤后的日志数量与占比统计。

操作步骤

1. 登录云日志服务控制台。
2. 使用交互模式过滤日志（可选）。
3. 点击统计分析，您可以选择日志占比作为场景，并选择统计字段、统计条件与内容，如下图所示。

4．点击检索/分析即可获得统计结果

函数原型

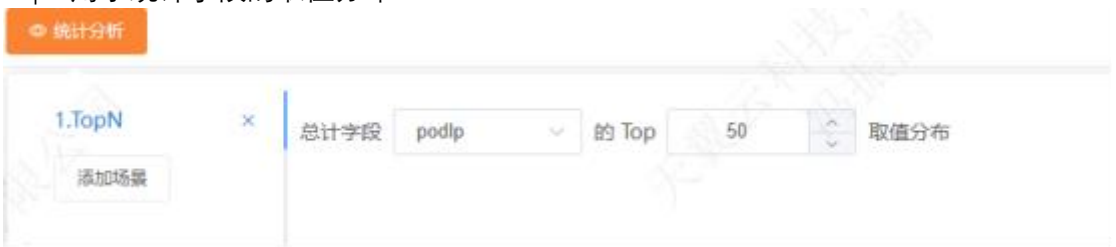
```
SELECT
    countIf(URL LIKE '%http%') AS count,
    COUNT(*) AS total,
    (count / total) AS percent
FROM
    your_table;
```

分析结果

如图所示为日志占比统计结果，其中 count 为满足条件的日志数量，total 为所有日志总数，percent 为满足条件日志数量的占比。

4.3.2.4.2. TopN

TopN 用于统计字段的取值分布



（含水印图）

如上图所示，您可以选择字段，并填入百分比的取值分布。

其中第一列为您选择的字段，count 表示字段的数量，percent 表示字段对应的日志数量在日志总数中的占比。

4.3.2.4.3. 时间趋势

时间趋势用于按时间粒度统计每个时间段内的函数指标。

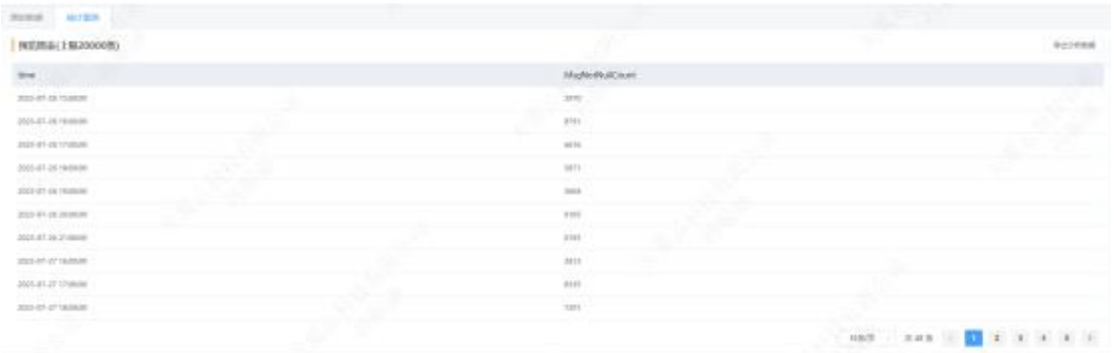
时间粒度

当前统计的时间粒度可以选择秒、分钟、小时、天、周、月、季度、年作为聚合单位，并可以选择函数统计指标作为时间粒度的统计值。

函数原型

```
SELECT
    toStartOfxxx(column) AS xxx_of_day,
    funcName(field) AS xxxxx
FROM
    log_table
GROUP BY
    xxx_of_day
ORDER BY
    xxx_of_day;
```

分析结果



分析结果以时间为粒度，如上图所示为统计到每小时内的字段值非空的条数。

4.4. 数据加工

4.4.1. 数据加工概述

云日志服务提供可扩展、高可用的数据加工服务。数据加工服务可用于日志的规整、富化、流转、脱敏和过滤等。

加工流程

日志加工服务通过如下三个步骤完成加工处理。

1. 通过消费组对源日志单元的已分词日志进行读取。
2. 通过加工规则对读取到的每一条分词日志进行加工处理。
3. 将加工后的日志写入目标日志单元。加工完成后，您可以在目标日志单元中查看加工后的日志。

加工语法

加工 DSL（Domain Specific Language）提供了 200 多个内置函数。

4.4.2. 基本概念

基本概念

ETL

ETL 是指将对业务系统的数据进行抽取、清洗、转换、加载的过程，从而整合零散、不标准、不统一的数据。云日志服务支持加载源日志单元数据，将数据转换后输出到目标日志单元，同时也支持加载 OSS 或其他日志单元的数据。

事件、数据、日志

在数据加工功能中，事件、数据都表示日志，例如事件时间就是日志时间，丢弃事件字段函数 e_drop_fields 就是用于丢弃特定日志字段的函数。

日志时间

日志时间指事件所发生的时间，也称事件时间。在云日志服务中的保留字段为__time__，一般由日志中的时间信息直接提取生成。数据类型为整数字符串，Unix 标准时间格式，单位为秒，表示从 1970-1-1 00:00:00 UTC 计算起的秒数。

日志接收时间

日志到达云日志服务的服务器被接收时的时间，默认不保存在日志中，但是如果日志单元开启了记录外网 IP 地址，则该时间会保留在日志标签字段的__receive_time__中。数据加工中时间的完整字段名是__tag__:__receive_time__。数据类型为整型，Unix 标准时间格式。单位为秒，表示从 1970-1-1 00:00:00 UTC 计算起的秒数。

日志标签

日志存在标记，区别于其他字段，在数据加工中，标签字段以__tag__:作为前缀。包括：
用户自定义标签：用户通过 API PutLogs 写入数据时添加的标签。
系统标签：云日志服务为用户添加的标签，包括__client_ip__和__receive_time__。

配置相关概念

源日志单元

数据加工中，从中读取数据再进行加工的日志单元是源日志单元。

一个加工任务仅支持一个源日志单元，但可以对一个源日志单元配置多个加工任务。

目标日志单元
数据加工中，数据写入的日志单元是目标日志单元。

一个加工任务可以配置多个目标日志单元，可以是静态配置，也可以是动态配置。具体配置方法，请参见多目标日志单元数据分发。

DSL
DSL（Domain Specific Language）是云日志服务数据加工使用的一种 Python 兼容的脚本语言。DSL 基于 Python 提供内置两百多个函数，简化常见的数据加工模式。也支持用户自定义的扩展 Python 脚本。更多信息，请参见语言简介。

加工规则
数据加工脚本， DSL 编排的逻辑代码的集合。

加工任务
数据加工最小调度单元，由源日志单元、目标日志单元、加工规则、加工时间范围以及其他配置项组成

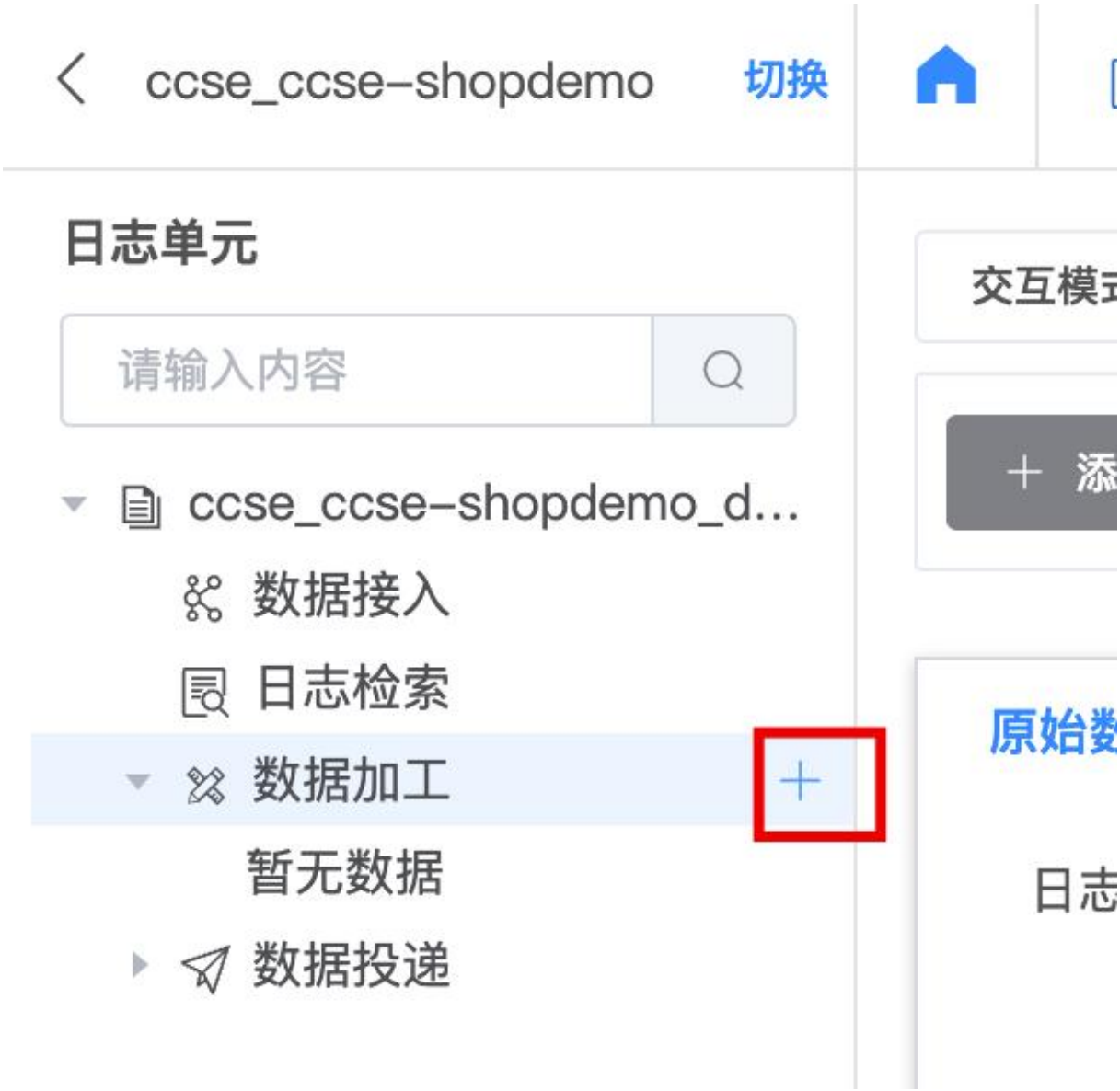
4.4.3. 创建数据加工任务

云日志服务支持您使用数据加工读取源日志单元中的数据，对数据进行加工处理后，写入到不同的目标日志单元中。您也可以对加工后的数据进行查询和分析，进一步发掘数据价值。本文介绍如何在云日志服务控制台上创建数据加工任务。

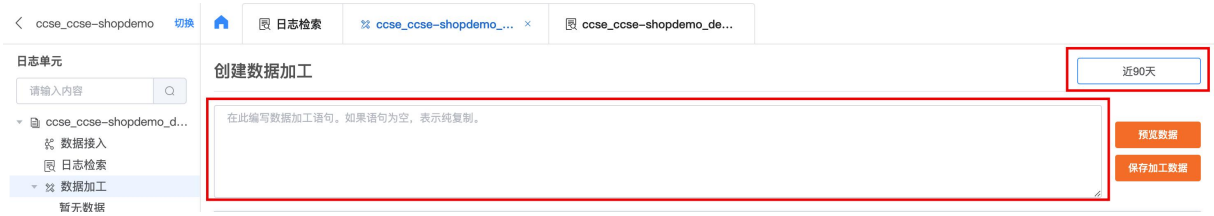
操作步骤

加工数据预览

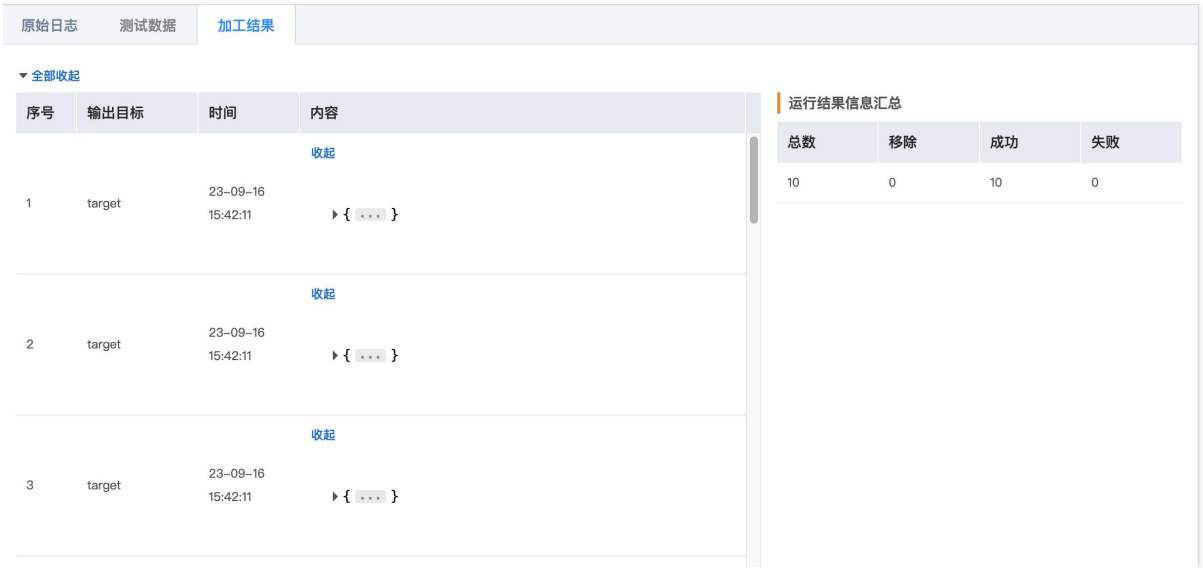
1. 登录云日志服务控制台。
2. 选择目标日志项目与日志单元，在左侧导航栏的数据加工，点击【+】按钮



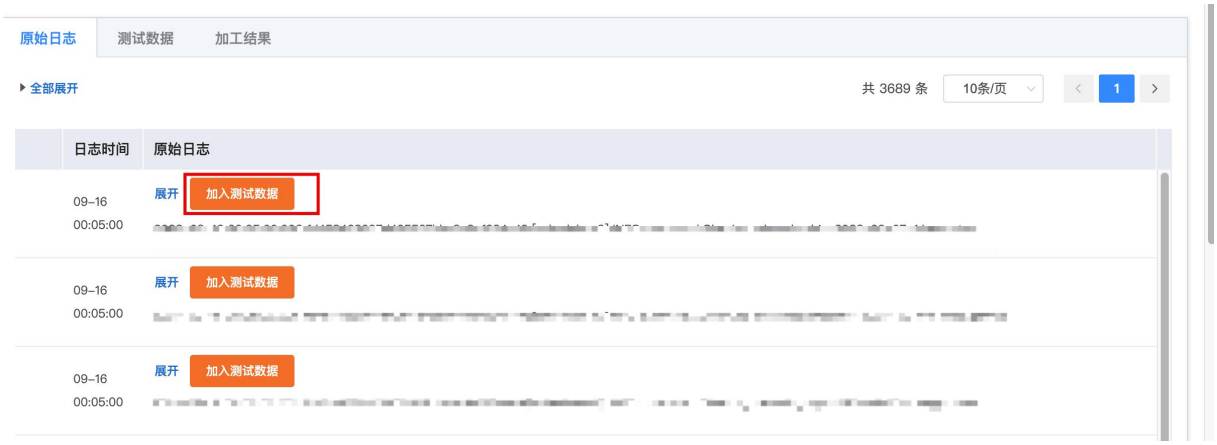
3. 选择所需要加工的日志的时间范围，并在页面中输入加工语句



4. 输入加工语句后，点击预览数据，即可预览数据加工结果以及结果汇总。

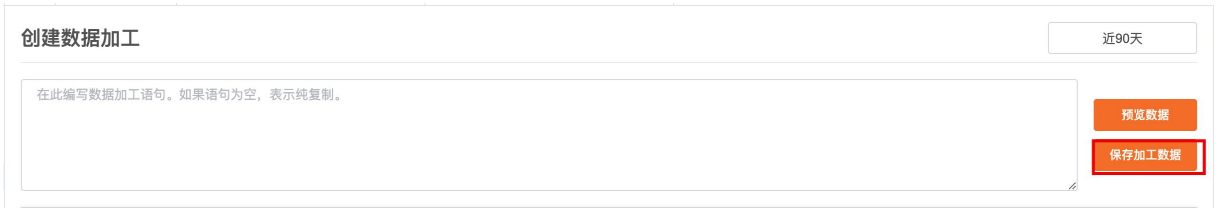


5. 可点击原始日志，查看加工前的原始日志。默认取原始日志中的前 10 条日志进行加工预览，您也可点击【加入测试数据】，针对特定的日志进行加工预览。



创建加工任务

1. 当您确认加工预览结果符合预期时，您可点击【保存加工数据】



在创建数据加工任务窗口中，根据以下信息进行配置，然后点击确定。

参数	说明
任务名称	数据加工任务的名称。

存储目标名称	存储目标的名称。存储目标中包括日志项目、日志单元等配置。
目标日志项目	用于存储数据加工结果的目标日志项目。
目标日志单元	用于存储数据加工结果的目标日志单元。
加工范围	持续加工：数据加工任务将持续进行。指定结束时间：您可指定数据加工任务的结束时间。

后续步骤

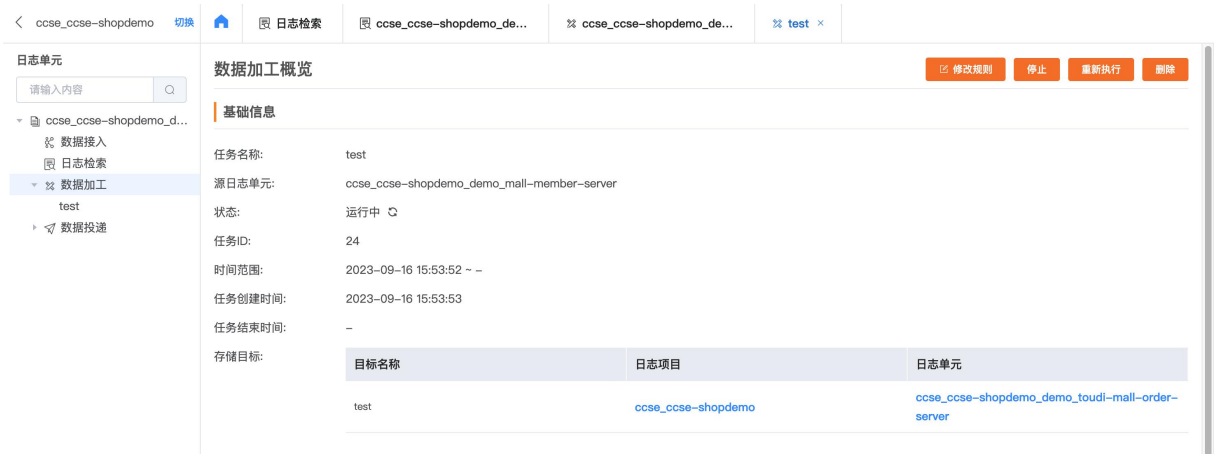
1. 在左侧导航栏的数据加工概览页面中，您可查看数据加工任务详情，修改加工任务，暂停加工任务等操作。
2. 在目标日志项目、日志单元中进行查询与分析操作。

4.4.4. 管理数据加工任务

本文介绍如何在云日志服务控制台上管理数据加工任务，包括查看任务详情与状态，修改、启动、停止和删除任务，设置告警等操作。

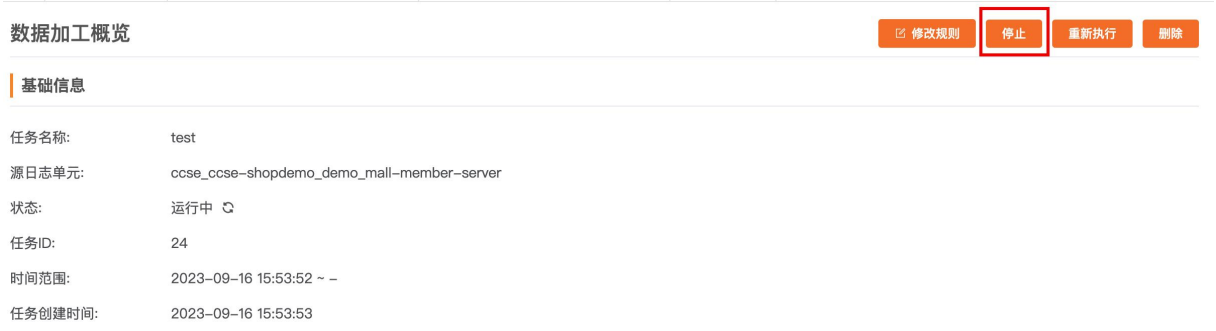
查看任务详情

1. 登录云日志服务控制台，选择目标日志项目与日志单元。
2. 在左侧导航栏中，点击数据加工。
3. 加工任务列表中，选择目标加工任务。
4. 在数据加工概览页面，查看加工任务详情。



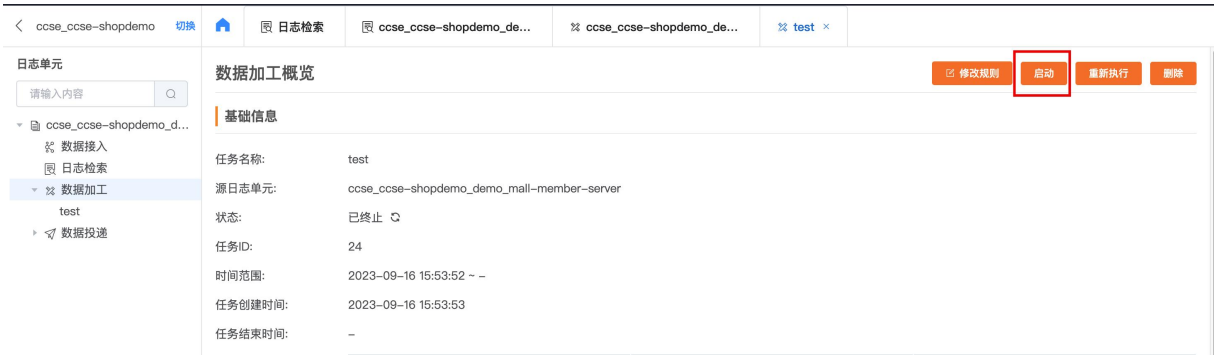
停止任务

对于状态为运行中的任务，您可在数据加工概览页面，单击停止，即可停止当前加工任务。



启动任务

对于状态为已终止的任务，您可在数据加工概览页面，单击启动，即可启动当前加工任务。



重新执行

任何状态的任务都支持重新执行。您可在在数据加工概览页面，点击重新执行按钮，数据加工任务将会重新执行。

修改加工规则

1. 在数据加工概览页面，点击修改规则按钮。

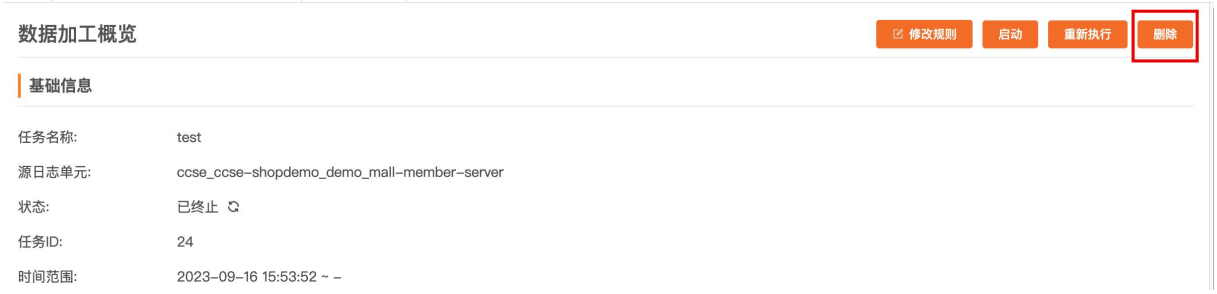
2. 在修改页面，您可根据需求修改数据加工规则，并进行数据预览，数据预览步骤请参考创建数据加工任务。

3. 确认数据加工语法后，点击修改数据加工。

4. 修改加工配置，相关参数说明请参考创建数据加工任务。点击确定，即可完成修改。

删除任务

在数据加工概览页面，点击删除，即可删除当前数据加工任务。



数

4.5. 日志投递

4.5.1. 概述

云日志服务提供投递功能，支持通过控制台将数据准实时投递至对象存储、消息队列等云产品。日志投递适用于数据存储、离线分析等场景。数据投递的实时性较弱，通常为 5 分钟~30 分钟。数据延迟依赖于存储系统。

4.5.2. 投递至 Kafka

4.5.2.1. 创建投递任务

云日志服务采集到数据后，支持将数据投递至 Kafka 进行后续的存储或分析。

前提条件

1. 已创建日志项目和日志单元
2. 已采集到日志

投递数据

登录云日志服务控制台。

1. 在日志项目列表区域，单击目标日志项目。
2. 在日志单元页中，单击目标日志单元左侧的>，选择数据投递 > Kafka。
3. 将鼠标悬浮 Kafka 上，单击+。
4. 在投递 Kafka 功能面板，配置如下参数，然后单击确定。

参数	说明
投递任务名称	投递任务的名称
Kafka 实例	Kafka 实例 broker 地址 **注**：需要提供外网可访问的、支持 SASL_PLAINTEXT 认证方式的 broker 地址，以逗号分隔 比如 10.1.1.2:9092,10.1.1.3:9092
Topic	Kafka 生产数据指定的 topic

	注：如果 Kafka 实例不支持自动创建 Topic，请提前创建
用户名	SASL_PLAINTEXT 认证方式的用户名
密码	SASL_PLAINTEXT 认证方式的密码
投递时间范围	仅支持持续投递和指定结束时间投递，不能指定开始时间，即只有投递任务成功启动以后新产生的日志才能被投递

4.5.2.2. 管理投递任务

您可以在数据投递概览页面管理 Kafka 投递任务，包括查看 Kafka 投递任务的基础信息、统计报表，修改配置并重启任务、删除 Kafka 投递任务等操作

前提条件

1．已创建 Kafka 投递任务

操作入口

登录云日志服务控制台。

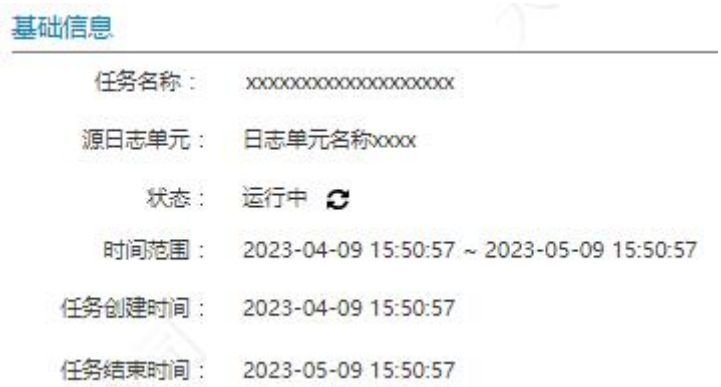
1．在日志项目列表区域，单击目标日志项目。

2．在日志单元页中，单击目标日志单元左侧的>，选择数据投递 ＞ Kafka。

3．单击目标 kafka 投递任务。

查看 Kafka 投递任务基础信息

您可以在基础信息区域查看 Kafka 投递任务的任务名称、源日志单元、任务状态、任务时间范围等信息



查看统计图表

创建 Kafka 投递任务后，云日志服务默认创建一个仪表盘，您可以在仪表盘中查看 Kafka 投递任务的运行指标

（统计图表）

图表名称	说明
读流量总计	从源日志单元中读取成功的日志流量
读数据条数	从源日志单元中读取成功的日志条数
投递成功条数	投递到目标Kafka成功的日志条数
投递失败条数	投递到目标Kafka失败的日志条数
投递速率	每秒处理日志条数

删除任务

如果您不再需要运行该Kafka投递任务，您可以在数据投递概览页面的右上角选择 删除

注意：投递任务被删除后，不可恢复，请谨慎操作

4.6. 告警

4.6.1. 概述

云日志服务支持在控制台设置自定义告警规则，支持针对不同的日志项目、日志单元进行规则配置，设置好的告警规则将周期性执行监控任务，对指定的日志单元执行检索分析，当检索结果满足触发条件时将发送告警通知，以便用户及时发现异常问题。

基本概念

术语	说明
告警规则	监控告警的管理单元，一条完成的告警规则包含监控日志单元对象、触发条件、检查频率、通知策略与告警内容等信息
监控对象	需要进行告警监控的日志单元，您可以针对该日志单元设置检索分析条件，系统将会检查结果是否满足触发条件
触发条件	当监控日志单元的检索结果满足您设置的触发条件时，将会触发告警通知，若不再满足触发条件，则告警恢复
检查频率	即告警规则的执行周期，支持固定频率与固定时间
通知组	即通知对象，支持设置短信、邮件联系人，翼连群，WebHook集成，您可以通过通知策略将目标告警通知发送给目标通知对象
通知策略	通知策略包含通知对象、通知模板与通知时段，当告警触发时，只要不

	符合静默策略，就会依照通知策略在对应渠道发送告警信息给对应的通知对象。

产品优势

快速响应：支持全面的监控能力，出现业务问题可快速响应，有效提高问题解决的速度，减少业务故障异常带来的损失。

免运维：通过云日志服务控制台即可快速创建告警规则并管理各种告警事件，降低系统运维成本与运维人员时间成本

灵活配置：可结合日志检索分析结果设置告警规则

多种通知方式：通知渠道支持邮件、翼连群、Webook 集成飞书、企微、钉钉

低成本：公测期间不收取通知费用，且无额外的管理费用

费用说明

产品公测期间不收取费用

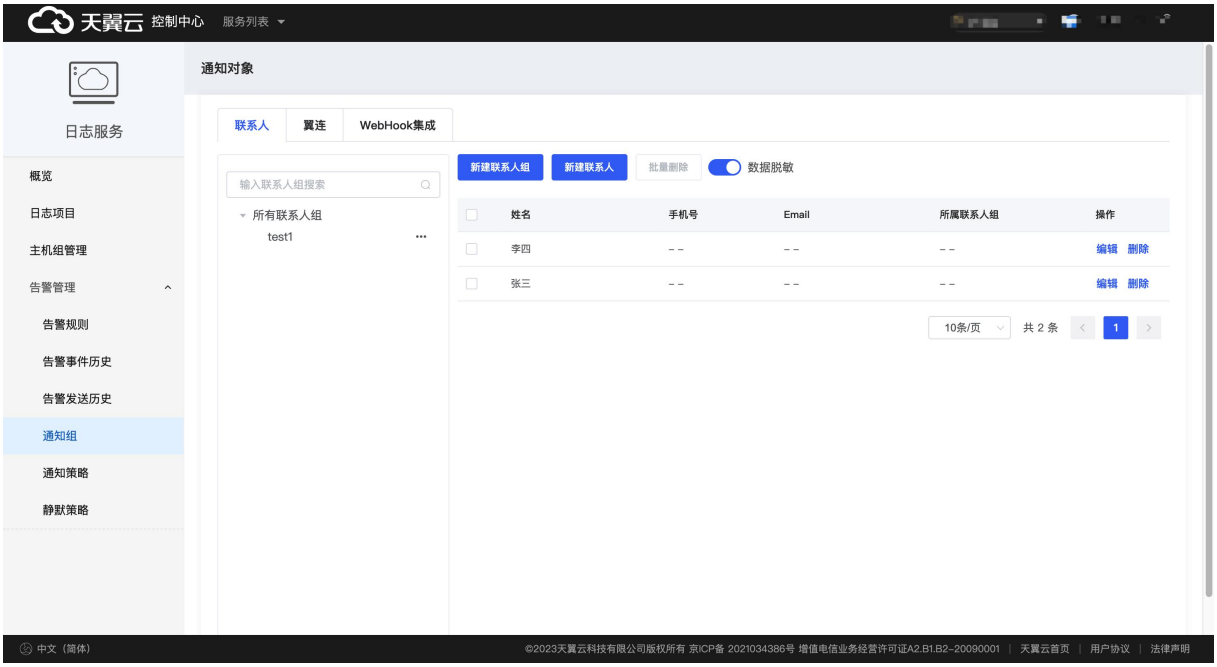
4.6.2. 通知组管理

告警支持通过邮箱、翼连、WebHook 集成等方式将告警信息通知给对应接收人，您可以在通知组管理中设置通知渠道与通知人信息。本文将介绍如何新增通知组。

操作步骤

1、登录云日志服务控制台

2、左侧导航栏点击通知组模块，进入通知组管理页面。



3、新增通知对象

- 新增联系人
  - 点击【新建联系人】
  - 输入姓名、手机号码、邮箱等相关信息。点击确定，完成新建联系人

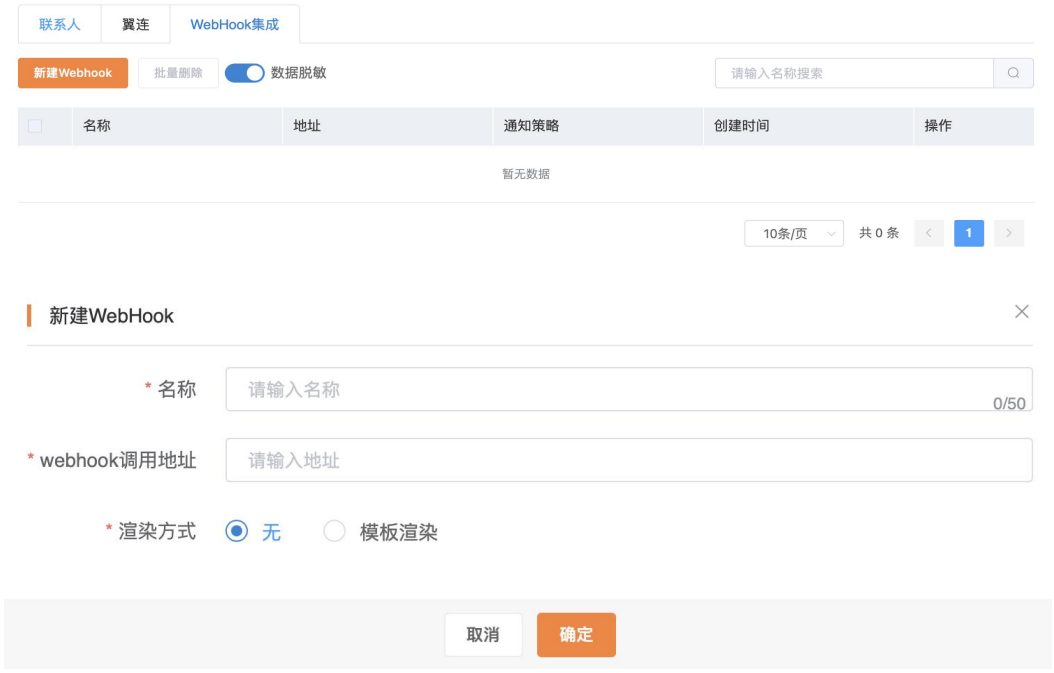
  - 您可将多个联系人添加到一个联系人组中。点击【新建联系人组】，输入联系人组名并在下方选择对应的联系人，点击确定按钮即可完成联系人组创建

- 新建翼连群
  - 点击翼连标签页，单击【新建翼连群】

  - 输入名称与翼连群号，将某个翼连群作为一个“联系人组”。

- WebHook 集成

- 支持以 WebHook 的方式对第三方通知对象（企微、飞书、钉钉）发送告警信息。
- 点击【新建 Webhook】



- 输入通知组名称与 webhook 调用地址。您需要在各个 Webhook 通知渠道侧完成相关配置，并获取 Webhook URL 地址。详情如下：

通知渠道	说明
钉钉	在钉钉中创建自定义机器人，获取 Webhook URL 地址。详情请参考自定义机器人接入。
企业微信	在企业微信中创建自定义机器人，获取 Webhook URL 地址。详情请参考群机器人配置说明。
飞书	在飞书中创建自定义机器人，并获取 Webhook URL 地址。详情请参考如何在群组中创建机器人。

- 渲染方式选择模板渲染，点击【选择模板自动填入】，选择对应的通知渠道
- 点击确定，完成创建

4.6.3. 通知策略管理

通知策略包含通知对象、通知模板与通知时段，当告警触发时，只要不符合静默策略，就会根据所配置的通知策略在对应渠道发送告警信息给对应的通知对象。本文介绍如何创建通知策略。

前提条件：

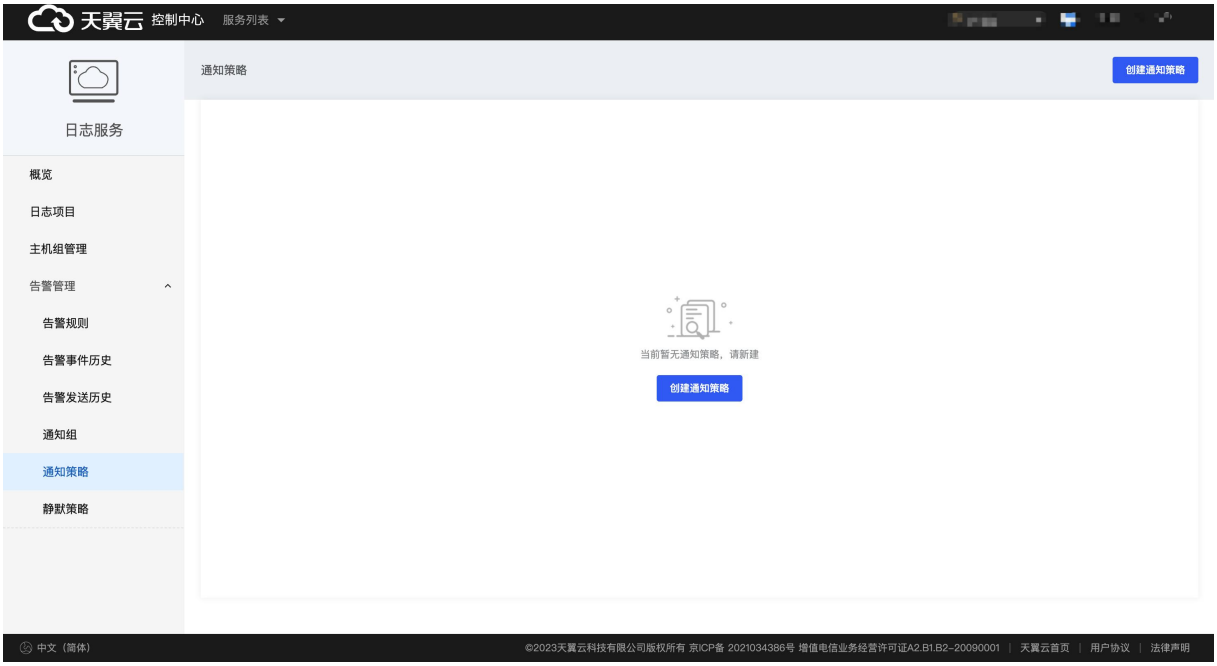
已完成通知组配置，详情请查看通知组管理

操作步骤

1、登录云日志服务控制台

2、左侧导航栏点击告警管理—通知策略模块

3、点击【创建通知策略】



4、在页面中输入通知策略名称。

5、点击【添加通知对象】，可添加联系人、联系人组、翼连、webhook 四种类型的通知对象，每种类型下可添加具体通知对象。

6、检查通知模板，您可根据实际情况修改不同渠道的通知模板内容。

7、设置通知时段，告警信息只在您配置的时段进行通知，不配置默认全天通知时段。

8、点击【确定】，完成通知策略创建。

4.6.4. 告警规则

监控告警的管理单元，一条完成的告警规则包含监控日志单元对象、触发条件、检查频率、通知策略与告警内容等信息。本文将介绍如何创建告警规则

前提条件

1. 已创建日志项目、日志单元并完成日志数据接入。详情请查看数据采集
2. 已创建通知策略，详情请查看 通知策略管理

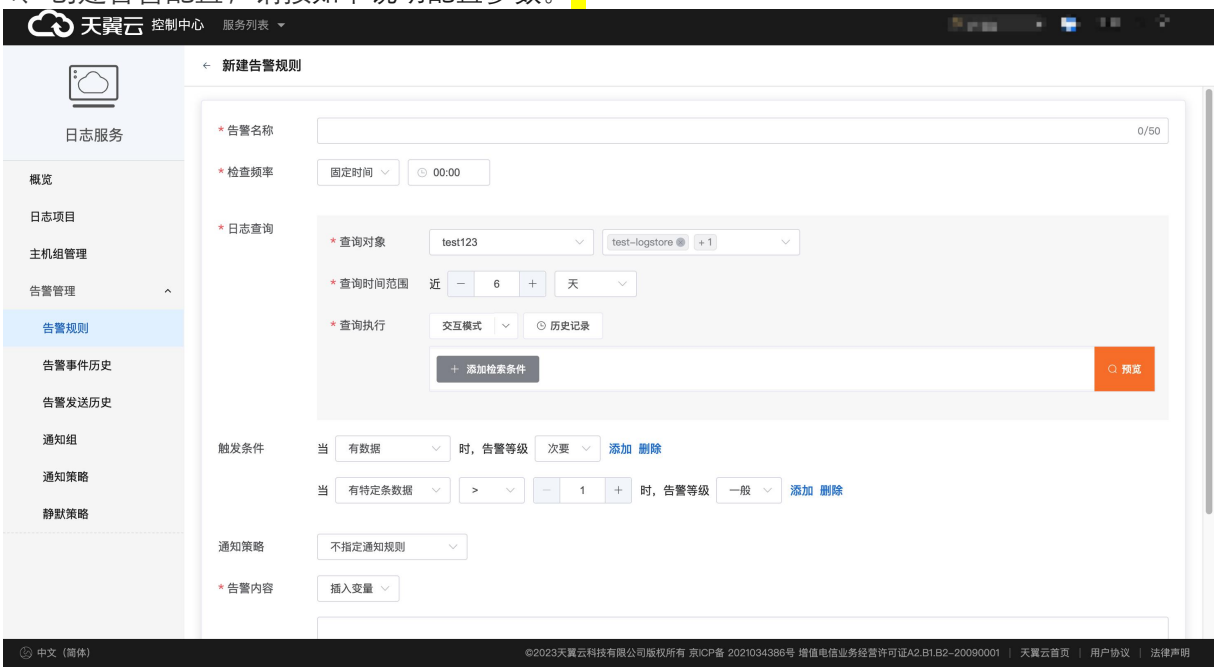
操作步骤

1、登录云日志服务控制台

2、左侧导航栏点击告警规则模块，进入告警规则页面。

3、点击创建告警规则

4、创建告警配置，请按如下说明配置参数。



参数	说明
告警名称	设置告警名称
检查频率	告警规则的执行周期，支持固定频率与固定时间 ● 固定频率：按固定的时间间隔执行一次监控任务时间间隔，支持设置分钟、小时、天的频率 ● 固定时间：按固定的时间点，每天执行一次监控任务
日志查询	● 查询对象：选择需要进行监控的日志项目与日志单元，可同时针对多个

	日志单元进行监控 ● 查询时间范围：针对目标日志单元进行检索时的时间范围 ● 查询执行：针对目标日志单元的检索条件，点击【预览】按钮可预览检索结果，当检索结果满足触发条件时，则触发告警。
触发条件	当目标日志单元的检索结果满足触发条件时，则触发告警通知。可同时设置多条触发条件，目前支持两类触发条件： ● 检索结果有数据：当检索结果有日志数据时即触发告警 ● 检索结果有特定条数据：当检索结果中的日志数据满足一定条件时即触发告警，可设置条件表达式。
通知策略	选择已创建好的通知策略
告警内容	告警内容将作为告警信息中的一个字段，告警内容支持插入各类变量。

5、点击【新建】，完成告警规则设置

4.6.5. 告警历史

4.6.5.1. 告警事件历史

告警事件历史展示当前租户下所有告警事件历史，支持筛选及查看详情。

功能入口

1、登录云日志服务控制台

2、在左侧导航栏中选择**告警管理—告警事件历史**

功能说明

展示当前租户下所有告警事件信息。

（1）支持对事件名称、事件状态、事件对象和对象类型进行筛选。

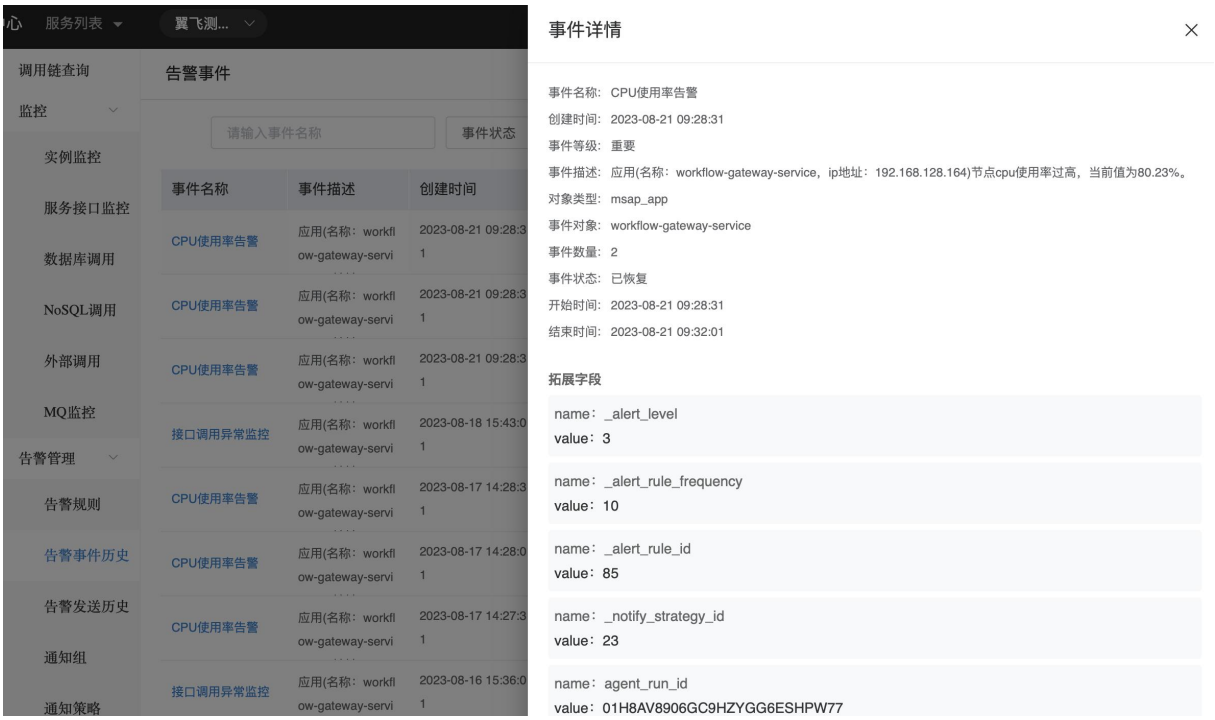
事件名称：显示告警规则名称

事件状态：显示事件当前状态是告警中、已恢复、静默

事件对象：监控对象，比如应用名称、集群名称等

对象类型：告警事件对象的类型

（2）支持查看事件详情



（3）支持查看告警详情

4.6.5.2. 告警发送历史

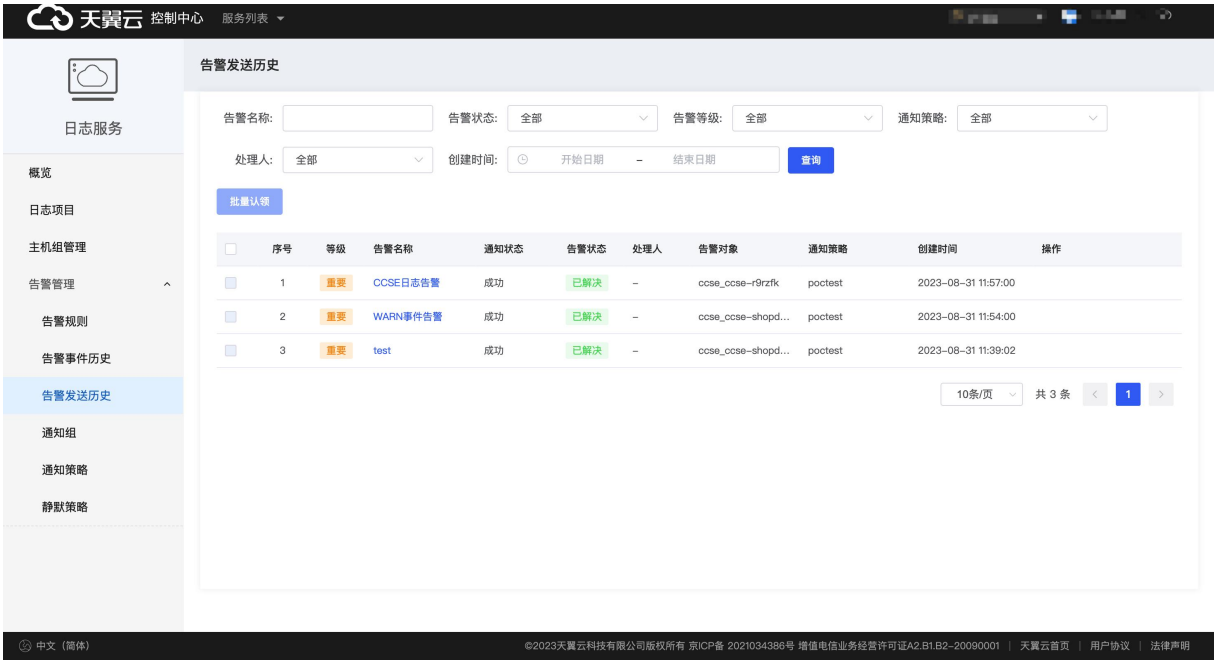
告警发送历史支持展示当前租户下所有发送的告警历史，支持筛选和对告警信息进行操作。

功能入口

1、登录云日志服务控制台

2、在左侧导航栏中选择**告警管理—告警发送历史**

功能说明



展示当前租户下所有告警发送信息。

（1）支持对告警名称、告警状态、告警等级、通知策略和创建时间进行筛选。

告警名称：显示告警规则名称

告警状态：显示事件当前状态是待认领、已解决、处理中

告警等级：显示告警的重要层级分布是一般、次要、重要、紧急

通知策略：告警对应的通知策略

处理人：告警的最新解决/认领人

创建时间：告警产生的时间

（2）支持认领、解决、指定处理人操作

认领：当告警处于待认领状态时，可主动领取当前告警

解决：当告警处于待认领/处理中状态时，点击解决可更新告警状态为已解决

指定处理人：指定他人处理该告警

（3）支持查看告警详情

5. 最佳实践

将本地日志迁移到云日志服务

操作场景

应用程序通常首先直接将运行日志输出到本地的文件中。当运维人员需要查询日志时，登录服务器，通过操作系统的文件查看工具，比如 more，vim，grep 等查询需要关注的日志内容。这种场景下，由于日志分布在各个服务器上，命令行操作易出错、服务器权限等限制因素造成日志

查找效率低下。

利用云日志服务，可以获得以下优势：

- 数据集中存储，无需登录多台服务器查询，这在微服务架构下尤为重要；
- 快速检索日志，告别繁琐的命令行操作，提升故障处理效率；
- 实时检测异常日志，设置告警，提升故障响应时效。

前置条件

- 开通云日志服务；
- 应用服务部署在天翼云的云主机上。

操作步骤

1、登录云日志服务控制台

2、在控制台概览页面的日志项目模块，点击目标日志项目名称。

3、在目标日志单元下，点击【数据接入】，若该日志单元初次配置日志采集，则页面会提供采集配置向导。

4、选择云主机

(1) 如果您还没有可用的主机组，请执行以下操作

a. 输入主机组名称

b. 在已开通云主机列表中，选择您需要采集日志的目标云主机，作为云主机组进行统一采集管理。

c. 点击下一步

(2) 如果您已有可用的主机组，请点击【选择已有主机组】，选择目标主机组，并点击下一步
注：目前仅支持采集天翼云 linux 云主机

5、安装采集器

(1) 安装采集器前，需要先接入云主机所在的 VPC 网络。在 Step1：VPC 接入中，点击【全部接入】按钮，稍等片刻后检查 VPC 的网络接入状态，确保列表中所有 VPC 都处于已接入状态。
注：若 VPC 无法接入，请点击对应 VPC 的【接入】按钮重试，若重试仍然失败，请查看“常见问题—VPC 接入失败”进行问题排查

(2) 复制 Step2 中提供的采集器安装命令，在目标云主机中执行该命令安装采集器。

(3) 安装完成后需要检查采集器状态。
在 Step3：检查采集器状态中，点击【开始检测】按钮，稍等片刻后查看采集器状态，

确保所有目标云主机的采集器状态均为“运行中”。

注：若采集器无法连通，请点击对应云主机的【重新检测】按钮重试，若重试仍然失败，请查看“常见问题–云主机采集器无法连通”进行问题排查，或点击【移除主机】从主机组中移除该主机

(4) 点击下一步

6、创建采集配置，请按如下说明配置参数。

参数	描述
采集规则名称	采集规则的名称，在所属的日志项目内唯一，
采集路径	根据日志在服务器上的位置，设置日志目录和文件名称。 日志路径必须以正斜线（/）开头，其中目录名和文件名支持完整名称和通配符模式。如 ● /var/log/auth.log：表示/var/log 目录下的 auth.log 日志文件 ● /var/log/*.log：表示/var/log 目录下后缀名为.log 的日志文件 ● /var/log/app_*/**/*.log：表示/var/log 目录下符合 app_* 格式的目录中后缀名为.log 的日志文件
采集策略	● 全量：从目标文件的第一行日志开始采集 ● 增量：从采集配置下发的时间对应的日志开始采集，如果下发采集配置后，日志文件无更新，则不会采集该文件中的日志
切割模式	针对原始日志执行分词的模式，选择“单行全文”

7、参数配置完成后，点击【保存并启用】，表示创建采集配置并开始采集日志。点击【保存不启用】，则表示只创建采集配置但暂不下发启用，您稍后可在数据接入管理页面进行启用，详情请查看 采集配置管理

8、在完成页面，点击【日志检索】，将跳转至日志检索分析页面。

优化查询性能

本文介绍优化查询的方法，用于提高查询效率。

缩减查询的时间范围和数据量

所选时间范围越大，查询消耗资源越多，响应时间越长。建议缩小时间范围提升响应速度。

数据量越大，查询速度越慢。尽可能地进行精确匹配，输入的关键词不建议过短导致扫描的数据量大，响应也就越慢。

选中区分度较大的关键词

查询关键词时，填写区分度较大的关键词，能快速定位到所需的数据，避免大量无用的数据扫描。填写关键词时建议输入大于 5 个字符。

查询 Log4j 日志

本文以最常见的 Java 项目为例，介绍 Log4j 的日志分析操作流程。

前提条件

- 已采集 Log4j 日志。
- 已定义采集规则配置。

【等录入采集规则模板图片】

背景信息

Log4j 是 Apache 的一个开放源代码项目，通过 Log4j 可以控制日志的优先级、输出目的地和输出格式。日志级别从高到低为 ERROR、WARN、INFO、DEBUG，日志的输出目的地指定了将日志打印到控制台还是文件中，输出格式控制了输出的日志内容格式。类似的组件还用 logback 等。

例如某个 Java 项目，希望查询某段时间的异常信息，来查看是否存在相关系统错误，以便快速

定位故障。针对此需求，云日志服务提供快捷的查询功能。比如某 Java 的日志信息如下：

level：ERROR

location：cn.ctyun.log4jtest.getUserInfo

message：get connection from pool timeout，pool is busy，reject task

time：2023—07—20 10:20:30.437

操作步骤

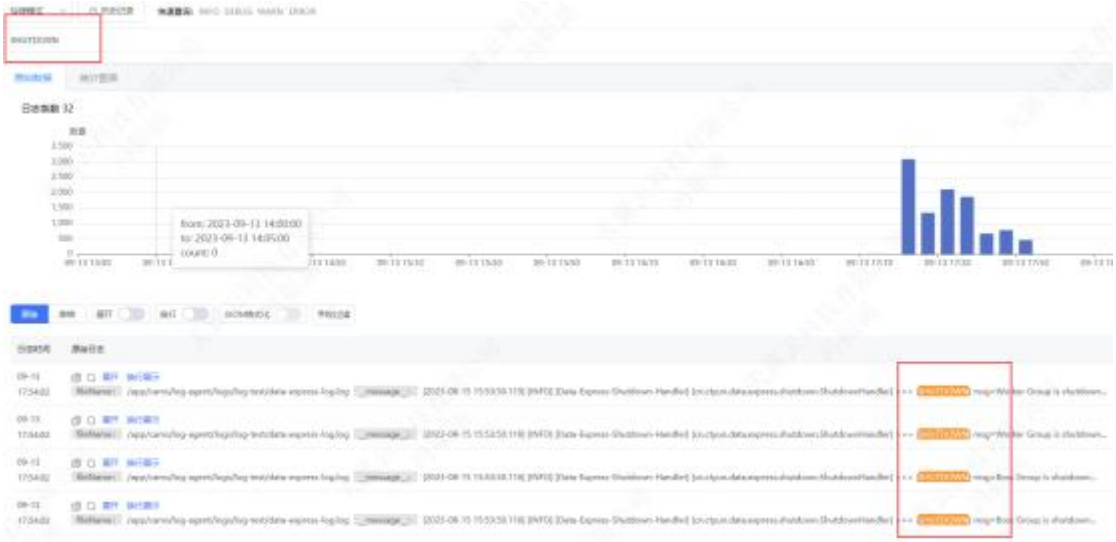
1．登录日志服务控制台。

2．在日志单元项目中，选择日志存储所属的项目。

3．在所属项目中，选择或者过滤出对应的日志单元。

4．通过【快捷模式】选择 ERROR，设置查询分析的时间范围，即可统计所选时间段的错误信息。

5．或者在检索框输入其他业务相关的特定关键字进行快速检索



点击上下文检索图标可以快速定位到日志所在行的上下文信息

以便于排查日志信息得到系统运行状态。

查询分析 Nginx 访问日志

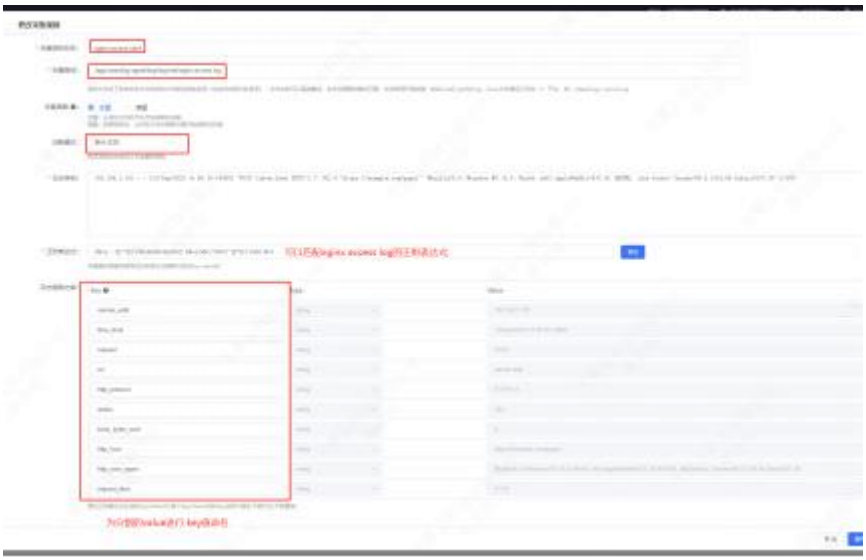
日志服务 LTS 支持采集 Nginx 日志，并进行多维度查询。本文介绍分析网站访问情况、异常和重要场景告警的分析案例。

背景信息

Nginx 是一款主流的网站服务器，当您选用 Nginx 搭建网站时，Nginx 日志是运维网站的重要信息。云日志服务支持通过数据接入向导一站式采集 Nginx 日志，并为 Nginx 日志创建索引和仪表盘。nginx_Nginx 访问日志仪表盘包括请求状态占比、请求方法占比、访问 PV/UV 统计、流入流出流量统计等信息，全方位展示网站访问情况。您还可以使用云日志服务的查询分析语句，分析网站的延时情况，及时调优网站。针对性能问题、服务器错误、流量变化等重要场景，您还可以设置告警，当满足告警条件时给您发送告警信息。

前提条件

根据配置向导，创建对应的日志单元、安装采集器、配置采集规则，生成日志字段。

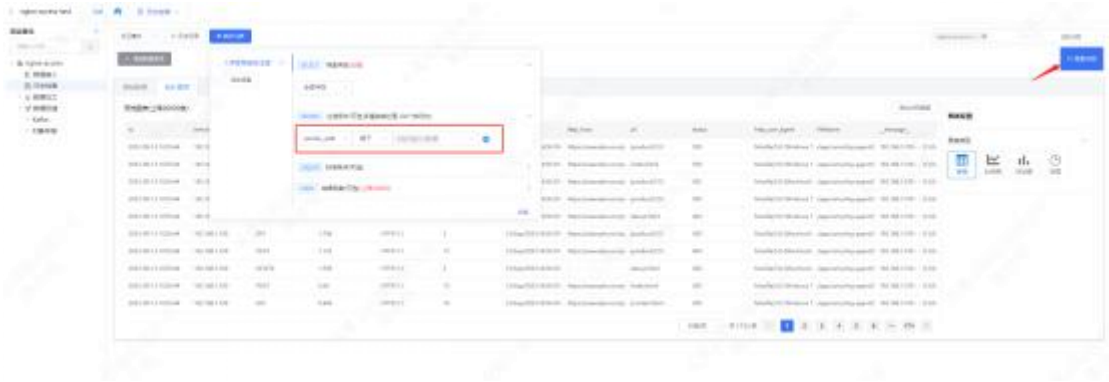


如上图，可以使用正则表达式来分割这个 Nginx 访问日志行。用户可以输入日志样例预览匹配结果，并为匹配结果进行字段命名

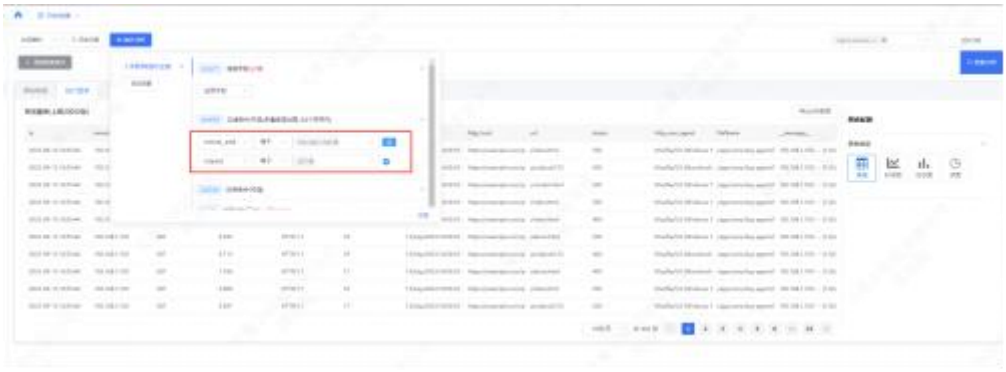
完成上述步骤后，采集器将开始在指定的主机组中进行日志采集，并将采集的日志信息展示在控制台上。用户点击日志检索之后再点击检索分析，便可以在控制台看到检索到的日志信息。

在展示页面上，用户可以根据自己的需求选择不同的展示方式。如对日志进行展开，JSON 格式化等操作

样例 1：查询 IP 192.168.1.100 发过来的请求日志



样例 2：查询 IP 192.168.1.100 发过来的 GET 请求日志



分析网站访问情况

1、登录云日志服务控制台。

2、在日志项目区域选择目标日志单元。

3、点击【日志检索】菜单，选择【交互模式】。

4、点击【统计分析】菜单。

- 查询相应接口的请求状态。

  【录入图片】

- 查询请求状态占比。

  【录入图片】

- 请求方法占比

展示最近一天各请求方法的占比情况，所关联的查询设置如下所示。

【录入图片】

- 流入流出流量统计

【录入图片】

您可以通过云日志服务数据加工函数清洗您所采集的海量日志数据，实现数据格式标准化。

数据加工：过滤日志（e_keep 函数和 e_drop 函数）

您可以使用 e_drop 函数或 e_keep 函数过滤日志，也可以使用 e_if 函数与 e_drop()参数、e_if_else 函数与 e_drop()参数过滤日志。

常用规则如下所示：

- e_keep(e_has(...))：满足条件时保留，不满足条件时丢弃。
- e_drop(e_has(...))：满足条件时丢弃，不满足条件时保留。
- e_if_else(e_has("..."), e_keep(), e_drop())：满足条件时保留，不满足条件时丢弃。
- e_if(e_has("not ..."), e_drop())：满足条件时丢弃，不满足条件时保留。
- e_if(e_has("..."), e_keep())：无意义的加工规则。

示例如下所示：

- 原始日志

```
#日志1
__tag__:observed_ts:  1597214851
entry: app_view
id: 8412
self_tag:  test_ok

#日志2
entry: h5_view
id:  8415
self_tag:  test_ok2
```

- 加工规则

丢弃没有 entry 字段和__tag__:observed_ts 字段的日志。

```
e_if(e_not_has("entry"),e_drop())
e_if(e_not_has("__tag__:observed_ts"),e_drop())
```

- 加工结果

```
__tag__:observed_ts:  1597214851
entry: app_view
id: 8412
self_tag:  test_ok
```

数据加工：日志空缺字段赋值（e_set 函数）

您可以使用 e_set 函数为日志空缺字段赋值。

- 子场景 1：原字段不存在或者为空时，为字段赋值。

```
e_set("result", "......value......", mode="fill")
```

mode 参数取值见下表：

参数值	说明
fill	当目标字段不存在或者值为空时，设置目标字段。
fill—auto	当新值非空，且目标字段不存在或者值为空时，设置目标字段。
add	当目标字段不存在时，设置目标字段。
add—auto	当新值非空，且目标字段不存在时，设置目标字段。
overwrite	总是设置目标字段。
overwrite—auto	当新值非空，设置目标字段。

- 示例如下所示：原始日志

```
name:
```

- 加工规则

```
e_set("name", "aspara2.0", mode="fill")
```

- 加工结果

```
name:  aspara2.0
```

- 子场景 2：为多个字段赋值。

```
e_set("k1", "v1", "k2", "v2", "k3", "v3", ......)
```

- 示例如下所示：
- 原始日志

```
__source__:  192.168.0.1
__topic__:
__tag__:
__receive_time__:
id:  7990
test_string:  <function test1 at 0x1020401e0>
```

- 加工规则 为__topic__字段、__tag__字段和__receive_time__字段赋值。

```
e_set("__topic__","app", "__tag__","stu","__receive_time__","1597214851")
```

- 加工结果

```
__source__:  192.168.0.1
__topic__:  app
__tag__:  stu
__receive_time__:  1597214851
id:  7990
test_string:  <function test1 at 0x1020401e0>
```

数据加工：为日志不存在的字段填充默认值（default 传参）

部分 DSL 表达式函数对输入的参数有一定要求，如果不满足，数据加工窗口会报错或返回默认值。当日志中存在必要而残缺字段时，您可以在 op_len 函数中填充默认值。**注意** 传递默认值给后续的函数时可能会进一步报错，因而需要及时处理函数返回的异常。

- 原始日志

```
data_len: 1024
```

- 加工规则

```
e_set("data_len", op_len(v("data", default="")))
```

- 加工结果

```
data: 0
```

```
data_len: 0
```

基于日志关键字设置告警

将日志采集到云日志服务后，您可以通过云日志服务告警系统实现基于日志关键字的告警。

背景信息

日志记录了系统的运行过程及异常信息，例如 warning 日志、error 日志、Go 语言中的 panic 错误日志、Java 语言中的 java.lang.StackOverflowError 错误日志。日志关键字的检索和监控告警是系统运行中比较常见的需求，您可以通过检索日志中的关键字和设置告警，快速感知和定位问题。云日志服务提供免运维、高性能、配置灵活的告警方案，帮助您实现基于日志关键字的告警。

出现关键字即触发告警

如果您希望日志中出现目标关键字时，就能触发告警，则您可以参考本案例设置查询语句和告警监控规则。

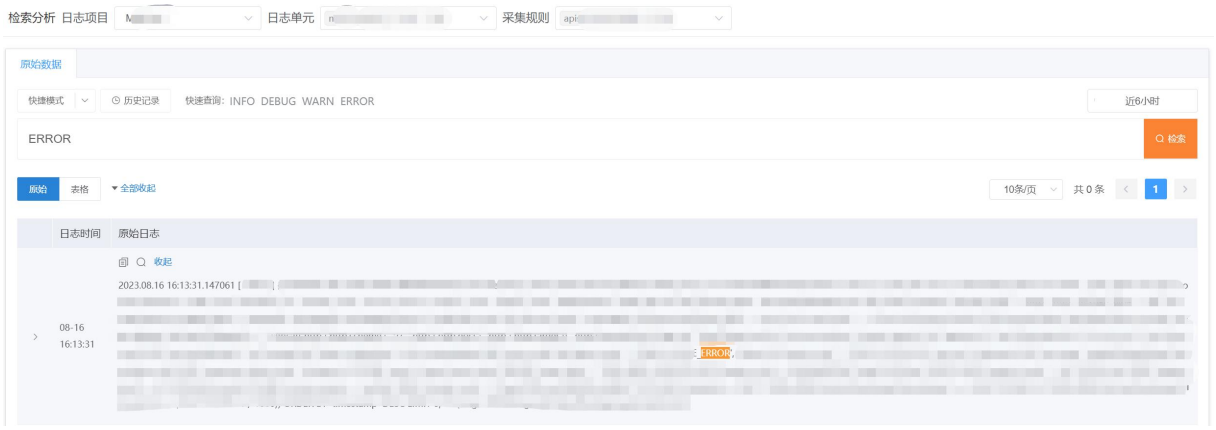
- 查询语句

选择查询时间范围为 **15 分钟（相对）**，然后执行如下语句，查询包含 ERROR 关键字的日志。具体操作，请参见<查询和分析>。

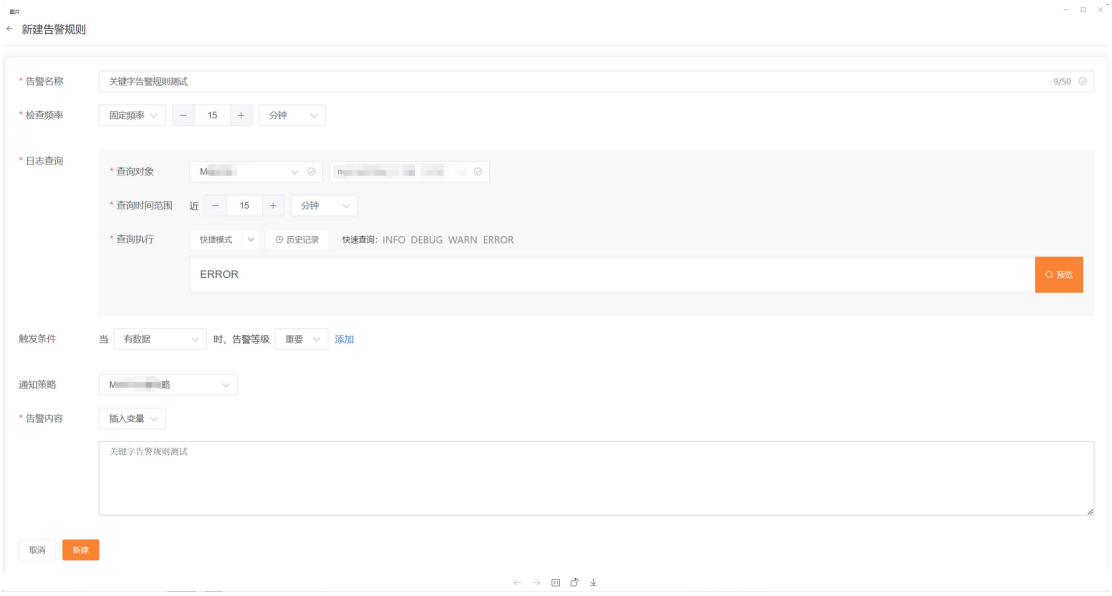
```
ERROR
```

- 查询结果

根据下述查询结果可知，当前 15 分钟内出现一次 ERROR 关键字。



- 告警监控规则配置
  - 基于上述查询结果创建告警监控规则。具体操作，请参见 <创建日志告警规则>。重要配置项说明如下：设置**触发条件**为**有数据**，则当日志中出现 ERROR 关键字时，触发告警。



- 告警通知

创建上述告警监控规则后，只要日志中出现 ERROR 关键字，您就可以收到告警通知。您还可以查看告警详情，进行溯源。

根据关键字出现的次数设置告警

如果您希望在一定时间范围内日志关键字出现的次数达到指定次数时，才触发告警，则您可以参考本案例设置查询分析语句和告警监控规则。

- 查询分析语句

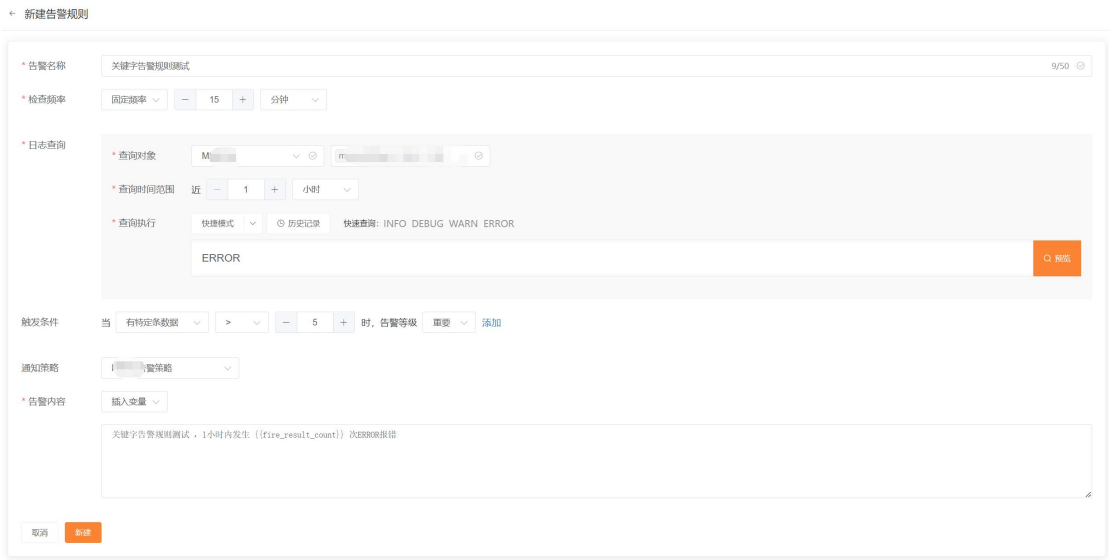
选择查询时间范围为 **1 小时（相对）**，然后执行如下语句，统计 1 小时内出现 ERROR 关键字的次数。具体操作，请参见<查询和分析>。

- 查询分析结果

根据下述查询分析结果可知，当前 1 小时内出现一次 ERROR 关键字 11 次。

- 告警监控规则配置

- 基于上述查询分析结果创建告警监控规则。具体操作，请参见<创建日志告警规则>。重要配置项说明如下：设置**触发条件**为**有数据匹配**，cnt ＞ 5，则 1 小时内日志中出现 ERROR 关键字大于 5 次时，触发告警。
- 设置**标注**中的**描述**为 **1 小时内发生 {{fire_result_count}} 次 ERROR 报错**，则告警通知中将显示当前 1 小时内日志中出现 ERROR 关键字的次数。



- 告警通知

创建上述告警监控规则后，只要当前 1 小时内日志中出现 ERROR 关键字的次数超过 5 次，您就可以收到告警通知。您还可以查看告警详情，进行溯源。

6. 常见问题

6.1. 售前常见问题

什么是云日志服务？

云日志服务是天翼云打造的一款云原生日志观测分析平台产品，可为应用的海量日志数据提供大规模、低成本、集中式的平台化服务，具备一站式的日志采集、加工、查询分析、可视化、告警、投递消费等能力，全面满足应用研发、运维、服务监控与业务分析等应用场景。

云日志服务包含什么功能？

- 日志采集：支持从云主机、容器应用、微服务应用以及其他组件中采集日志。采集规则支持动态配置，提供单行/多行全文、正则、分隔符、JSON 等日志结构化解析方式
- 日志检索：对于采集到的日志数据，支持模糊查询、全文查询、字段查询、时间范围、上下文查询
- 统计分析：支持各类统计分析场景
- 数据加工：支持对日志数据进行加工，包括数据的规整、脱敏和过滤
- 日志投递：支持日志数据投递至 kafka、对象存储
- 告警支持对一个或多个日志单元设置自定义告警规则，告警规则将按照设定的周期执行监控任务

如何开始使用云日志服务？

在使用云日志服务前，您需要完成天翼云账号注册并实名认证。完成实名认证后，打开产品详情页额，点击【立即开通】，按照指引开通云日志服务。详情请查看快速入门

云日志服务支持哪些数据接入方式？

6.2. 计费类

云日志服务如何计费？

计费模式包含资源包与按需计费。当前为公测期间，公测期间您可免费使用云日志服务。

资源包包含哪些费用？

日志资源包可用于抵扣所有按量计费项。详情请查看计费说明–计费项。

资源包如何抵扣？

资源包的规格单位为资源额度 CU（Cost Unit），每个计费项消耗抵扣资源包额度（CU）的比例，均与该计费项的按量付费的单价完全一致。如云日志服务的读写流量按量付费单价为 0.18 元/GB，则每 GB 读写流量将抵扣资源包 0.18CU。

资源包购买后多久生效？

资源包购买后立即生效

如何查看资源包使用明细？

您可在云日志服务控制台首页查看使用明细，详情请参考操作指南–资源管理–资源统计

资源包是否可以叠加使用？

多个资源包可以叠加使用。但仅可叠加资源包规格，不叠加有效时长。

资源包是否支持续费？

在资源包有效期内，您可以随时对资源包进行续费。详情请查看资源包续费

资源包到期后会丢失日志数据吗？

资源包在有效期内，每月有相同的资源额度。当月额度被用完后，自动转为按量付费方式。因此不会导致数据丢失

如何关闭云日志服务？

暂不提供关闭服务功能，若您不再使用云日志服务，您可以选择将您的日志项目全部删除，无需注销账号。详情请查看操作指南–管理日志项目

为什么购买了 1 年期的资源包，1 个月就被用完了？

天翼云日志服务提供的资源包是周期型资源包。根据您购买的时间按月提供固定的额度容量，次月恢复满额，当月剩余额度不可累计到下月。当月额度已用完后，下个月将会恢复额度。

6.3. 数据采集

VPC 无法接入如何排查？

请检查您云主机所在的 VPC 状态是否正常。

云主机采集器无法连通如何排查？

检查云主机所在的 VPC 是否已接入。

检查云主机是否已按步骤正常安装采集。

如何在云主机上安装采集器？

云日志服务控制台提供采集器安装命令，您需要在目标云主机中执行该命令安装采集器。详情请查看采集器安装。

配置数据采集时支持哪种分词方式？

日志的结构化解析指云日志服务数据将以 key—value 对的形式存储在云日志服务平台上。日志数据结构化后，您可以在云日志服务控制台根据指定的键值进行日志检索、分析与加工。目前采集器提供多种解析方式，详情如下：

解析方式	说明
单行全文	单行全文是指一条日志仅包含一行的内容，在采集的时候，将使用换行符来作为一条日志 构化处理，也不会提取日志字段。每条日志都会存在一个默认的字段 message，采集器会
多行全文	多行全文日志是指一条完整的日志数据可能跨占多行，您需要指定首行正则已进行匹配， 出现作为该条日志的结束标识符。日志内容同样也会存放在 message 字段中。详情请参考
单行正则	单行正则模式用于处理结构化的日志，针对包含一行内容的日志，您需要指定一个正则表 单行正则模式
多行正则	多行正则模式用于处理结构化的日志，针对包含多行内容的日志，您需要指定一个行首正 集文本日志—多行正则模式
单行分	单行分隔符模式支持通过分隔符将一条日志分割成多个值，从而实现结构化处理，该模式

隔符	志-多行分隔符模式
JSON	支持解析 Object 类型的 JSON 日志，提取 JSON 日志内容作为 Key-Value 对，即 Object 首层

如何查看日志采集器的运行状态？

在机器上执行 service log-agent-user status

日志采集器在机器上的安装位置？

/app/cams/log-agent/

如何判断日志采集器是否在上报数据？

在日志采集器的 log 中搜索 self monitor，观察其中的 sendBytes 指标值。如果在上报数据，则这个值不为 0。

如何判断某个采集规则是否生效？

在日志采集器 log 中搜索文件名，如果在 self monitor 日志中有这个文件，则说明文件被正常采集。

如何在容器集群中部署日志采集插件？

登录 CCSE 控制台，在左侧导航栏中选择“集群”并选择一个可用的集群。

在左侧二级导航栏中选择“插件”-“插件市场”。

在插件列表中选择 ctg-log-operator，点击“安装”按钮。

在右侧的弹窗中选择“插件版本”，选择“超时时间”，点击“安装”按钮。

6.4. 数据存储

云日志服务中的数据可以保存多久？

云日志服务的数据支持 1—365 天的保存自定义设置，超过保存时间设定数据将会被丢弃。

日志项目或日志单元数据在删除后是否可找回？

日志项目和日志单元的数据在执行删除操作后，会进行任务调度删除，不可找回。请您谨慎操作，并注意保管平台的账号密码防止恶意破坏。

日志数据是否安全存储？

云日志服务底层采用双副本，不同物理机不同硬盘存储不同副本，极大的保证数据丢失风险，同时，在硬盘损坏时会及时替换并做副本拷贝，保证数据的双副本状态。

如果数据量突然激增，云日志服务如何保证服务不受影响？

云日志服务提供弹性伸缩、灵活适配的数据基础框架，实时监测数据流量，削峰填谷，智能数据写入与查询控制，保障服务的稳定运行。

天翼云云会使用我在云日志服务上存储的数据吗？

用户的业务数据，天翼云除执行您的服务要求或者法律法规要求外，不进行任何未获授权的使用及披露。

6.5. 查询与分析

为什么检索不到日志？

1．无日志

（1）可能为日志上报失败

（2）日志时间错误问题，日志上报至未来时间或已超出日志主题存储时长的时间。

2．部分时间段有日志，部分时间段无日志：

（1）可能为日志上报失败，请检查 LogAgent 输出日志是否存在错误。

（2）可能为该时间段确实无日志，请检查自身业务在指定时间段是否产生日志。

为什么关键字查询的检索时间很长？

检索性能和检索的数据规模有关，良好的索引结构有助于大大减少检索的数据规模，对于频繁查询到的字段，建议单独建立字段索引，并使用字段查询而非关键字查询。

如何完成双重条件查询？

需要使用两个条件查询日志时，通过 and 或者 or 选择需要满足的字段条件，进行多字段条件查询。

如何进行上下文查询？

当您检索到日志后，找到目标日志并点击上下文检索图标，即可查看日志上下文。

支持哪种统计分析场景？

- 基础分析
  - 字段筛选&过滤
- 基础统计指标统计
  - 分组统计
- 高级统计日志占比
  - TopN
  - 时间趋势

6.6. 数据加工

常见错误排查

- 基本语法错误
  - 问题：编写了不符合日志加工 DSL 语法的加工规则，例如：多或者少写括号、逗号（,），漏打回车符等
  - 排查方法：通过加工预览检测语法错误，页面会提示 SyntaxError 等错误信息
- 非法运算符
  - 问题：日志加工 DSL 中所有的操作都需要通过 DSL 提供的函数来完成。比如数值运算、比较等操作都需要通过 op_*函数完成，而不能直接使用 python 运算符
  - 排查方法：通过加工预览检测运算符等错误；可将算术运算符和比较操作符等运算符替换为 DSL 提供的函数完成

- 调用不存在的函数
  - 问题：调用了不存在的函数，通过加工预览即可检测出来，并提示 unknown function
  - 排查方法：通过加工预览即可检测出是否调用了不存在的函数；建议检查是否拼写错误。
- 函数参数传递错误
  - 问题：参数类型错误或参数个数错误等，加工预览会有错误结果输出，比如 TypeError、ValueError、"xx takes at least x arguments (xx given)"等
  - 排查方法：通过加工预览和构造测试日志发现参数传递问题

6.7. 数据投递

投递日志到 Kafka 失败

- 问题现象
  - 投递到 Kafka 失败可能报错如下：
  - 网络不可达
    - 错误信息：connect to kafka error
  - SASL 认证失败
    - 错误信息：Authentication failed：Invalid username or password
- 可能原因
  - 根据报错信息，可能的错误原因如下：
  - kafka broker 地址没有配置为 Advertised listener，或者不是外网可访问地址，或者配置了特定的安全组，没有对云日志服务开通访问
  - 填写的 kafka 地址没有开通 SASL_PLAINTEXT 认证，或者用户名/密码填写错误
- 解决方法
  - 投递任务配置里面，需要确保 Kafka broker 地址为外网地址，且配置了 Advertised listener
  - 投递任务配置里面，需要确保 Kafka broker 地址已开通了 SASL_PLAINTEXT 认证，且填写的用户名密码无误

6.8. 告警

如何基于关键字设置告警？

您将日志采集到云日志服务后，可通过日志告警功能实现基于日志关键字的告警。详情可查看最佳实践-出现关键字触发告警。

为什么设置了多个触发条件，只有一个生效？

告警服务将按照触发条件的顺序逐条匹配。若您设置了多个触发条件，当查询结果符合第一个触发条件后，将不会再匹配后面的触发条件。建议您根据告警严重程度，从较高级别的严重度开始配置。

为什么出现漏告警或者误告警？

- 漏告警：例如告警触发条件是错误日志数大于 10 就触发告警，而在日志单元查询分析页面查询时某个时间段内错误日志数实际大于 10 ，却没有触发告警。
- 误告警：例如告警触发条件是 QPS 低于 100 就触发告警，而在日志单元查询分析页面查询时某个时间段内 QPS 实际大于 100，却触发了告警。

出现漏告警或者误告警，一般是由于数据写入到日志单元到可查询存在一定的延迟，当告警监控规则中的查询时间范围设置为相对时间时，会导致告警的查询不完全准确。为了避免这两种情况，建议扩大告警监控规则中的查询时间范围或者将查询时间范围设置为整点时间。

触发告警成功，但是通知失败，如何处理？

可以在 告警发送历史 功能中，点击进入具体的告警发送历史记录中，点击活动 Tab 页，查看通知失败的原因。