

消息队列 CKafka 版

CKafka 连接器

产品文档





【版权声明】

©2013-2024 腾讯云版权所有

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

【商标声明】



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

【服务声明】

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



文档目录

CKafka 连接器

连接器简介 连接器概述

连接器优势

应用场景

技术原理

使用限制

连接管理开放网段

连接管理

任务管理

新建数据接入任务

HTTP 上报

对象存储 COS

数据传输服务 DTS

MongoDB

新建数据流出任务

数据目标

流出至 Clickhouse 流出至日志服务 CLS

数据流出至对象存储 COS

流出至 Elasticsearch Service

TDSQL-C PostgreSQL

分布式数据仓库(TDW)

简单数据处理

简单数据处理

数据转换

过滤器规则说明

管理任务

Schema 管理

事件中心



CKafka 连接器 连接器简介 连接器概述

最近更新时间:2024-01-09 14:54:11

什么是 CKafka 连接器

CKafka 连接器是腾讯云上 SaaS 化的数据接入和处理工具,可以方便快捷地完成一站式的数据接入、处理和分发。 连接器提供基于 HTTP/TCP 协议的 SDK 协助客户快速完成数据上报、基于 CDC(Change Data Capture)机制快速 订阅、存储多款数据库(MySQL、PostgreSQL、MongoDB 等)变更信息,打通了多款云产品的日志投递。并提供 了简单可配置的数据清洗(ETL)能力,以及丰富的数据流出渠道,协助客户低成本搭建数据流转链路,构建数据源 和数据处理系统间的桥梁。

CKafka 连接器旨在做好云上、云下(跨云、混合云)场景的数据连接,提供 SaaS 化的、简单可配置的、低成本的数据流转能力。打造一条可靠的、稳定的数据链路。

连接器功能

数据上报

CKafka 连接器提供基于 HTTP/TCP 协议的客户端 SDK, 方便客户将数据上报到多款消息队列(Kafka、Pulsar、 RocketMQ 等),屏蔽数据上报 Server 端的开发维护成本和多种消息队列的复杂协议理解和 SDK 使用成本,简化数 据上报流程, SaaS 化地完成数据的上报。

数据库变更信息订阅

CKafka 连接器支持基于 CDC(Change Data Capture)机制订阅多款数据库的变更数据。例如订阅 MySQL的 Binlog, MongoDB 的 Change Stream, PostgreSQL 的行级的数据变更(Row-level Change), SQL Server 的行级 的数据变更(Row-level Change)等,将这些数据订阅到消息队列中,方便业务上对这些数据的处理、分发、查询 等。

集成丰富的数据源

CKafka 连接器旨在打通腾讯公有云、跨云、混合云场景下的数据连接。支持云上多款云产品的日志接入、支持用户 自建场景的数据源接入、支持跨云场景的数据源接入等。目前已打通日志、数据库、中间件、HTTP 等多种场景的数 据接入。

数据清洗/数据流出



当完成数据上报,数据进入到消息队列时,CKafka 连接器提供了强大的、界面化的数据 ETL 配置能力,方便快捷的完成数据的清洗、格式化、数据格式转换等操作。并支持将完成 ETL 处理的数据,转储到下游

Serverless 自定义处理

CKafka 连接器基于 Serverless 平台的优势,依托 Serverless Function 的按量计费、用户自编码能力,支持用户基于 Serverless Function 自定义的完成业务逻辑的编写,完成自定义的数据处理分发功能。





连接器优势

最近更新时间:2024-01-09 14:54:11



易用性

CKafka 连接器需通过简单的界面配置,轻松完成数据上报、清洗和存储的链路搭建。屏蔽数据接入过程中底层复杂的系统搭建、组件运维过程。

实时性

CKafka 连接器在数据采集、上报和流转整条链路过程中,实现秒级的接收、处理、并分发到下游系统。例如业务需要上报某个数据,并在 ES 中查询,如果使用 CKafka 连接器的数据链路,则客户端上报后,即刻(秒级)在 ES 可以查询。

弹性伸缩

CKafka 连接器无需预估业务容量,系统会根据流量规模自动弹性伸缩,保证波峰时系统可用性。按需使用, Serverless 化的完成数据接入、处理、转储的整个流程。

高可用

CKafka 连接器支持接入层、处理层、分发层均为分布式跨可用区部署,遇到故障即可自动切换,服务可用性不低于 99.9%。当系统故障时,数据会暂存在消息队列中,可保证数据100%不丢失。系统恢复后,会继续处理分发消息队 列中的数据到下游。



安全性

CKafka 连接器支持不同租户间网络隔离,支持数据上报集成 CAM 鉴权、数据流转集成 SASL 权限控制,严格控制 访问权限,保证数据安全。

丰富的上下游生态

CKafka 连接器支持云上、云下(跨云、混合云)场景、支持自建和云上服务的数据连接。上游集成日志、数据库、中间件、日志等多种数据源。实现了云上容器、COS、Elasticsearch Service、MySQL 和 PostgreSQL 等15+云上产品打通,一站式实现数据的接入和流动。



应用场景

最近更新时间:2024-01-09 14:54:11

数据上报查询

CKafka 连接器可应用在需要进行数据上报的场景。如手机 App 的操作行为分析、前端页面的 Bug 日志上报、业务 数据的上报等等。一般情况下,这些上报的数据都需要转储到下游的存储分析系统里面进行处理(例如 Elasticsearch, HDFS 等)。在常规操作中,我们需要搭建 Server、购买存储系统、并在中间自定义代码进行数据接 收、处理、转储等。流程繁琐,长期系统维护成本高。

CKafka 连接器旨在用 SaaS 化的思路解决这个问题,目标是通过如下两步:界面配置、SDK 上报,完成整个链路的 搭建,并基于 Serverless 理念,以按量计费,弹性伸缩,无需预估容量等方式,简化客户的研发投入成本和实际的 使用成本。



数据库变更信息订阅

CKafka 连接器支持基于 CDC(Change Data Capture)机制订阅多款数据库的变更数据。例如订阅 MySQL 的 Binlog, MongoDB 的 Change Stream, PostgreSQL 的行级的数据变更(Row-level Change), SQL Server 的行级 的数据变更(Row-level Change)等。例如,在实际业务使用过程当中,业务经常需要订阅 MySQL 的 Binlog 日 志,获取 MySQL 的变更记录(Insert、Update、Delete、DDL、DML 等),并针对这些数据进行对应的业务逻辑处 理,如查询、故障恢复、分析等。

在默认情况下,客户往往需要自定义搭建基于 CDC 的订阅组件(例如 Canal、Debezium、Flink CDC 等)来完成对数据库的数据订阅。而在搭建、运维这些订阅组件的过程中,人力投入成本较高,需要搭建配套的监控体系,保证



订阅组件的稳定运行。

基于此种情况, CKafka 连接器提供 SaaS 化的组件, 通过界面配置化的完成数据的订阅、处理、转储等整个流程。



数据集成

CKafka 连接器支持将不同环境(腾讯公有云、用户自建 IDC、跨云、混合云等)的不同数据源(数据库、中间件、 日志、应用系统等)的数据集成到公有云的消息队列服务中,以便进行数据的处理和分发。在实际业务过程中,用 户经常需要将多个数据源的数据汇总到消息队列中,例如同一个应用业务客户端数据、业务 DB 的数据、业务的运行 日志数据汇总到消息队列中进行分析处理。正常情况下,需要先将这些数据进行清洗格式化后,再做统一的转储、 分析或处理。

CKafka 连接器提供了数据聚合、存储、处理、转储的能力。简而言之,就是提供数据集成的能力,将不同的数据源 连接到下游的数据目标中。





数据清洗/转储

有一些客户的场景中,数据已经存在作为缓存层的消息队列(例如 Kafka)中了,需要进行清洗格式化(ETL)处理 后,存储到下游(例如 Kafka、Elasticsearch、COS)等。常见的情况下,用户需要使用 Lostash、Flink 或自定义编 码进行数据的清洗,并维护这些组件的稳定运行。当数据只需要简单的处理的情况下,学习数据处理组件的语法、 规范、技术原理,维护数据处理组件就显得比较繁琐,额外增加了研发和运维的成本。

CKafka 连接器提供了轻便、界面化、可配置的数据处理(ETL)和转储的能力,支持界面化的配置数据的 ETL,同时将数据转储到下游存储。







技术原理

最近更新时间:2024-01-24 17:22:03

系统架构

CKafka 连接器系统架构如下所示:





主要分为四层:



数据源

数据源指客户数据源所处的位置。数据源可以在云上、自建 IDC、跨云、混合云中。数据可以是业务数据、日志数据或者 DB 中的数据等。

接入层

CKafka 连接器提供了适配多种协议的接入层,如 HTTP Rest, Kafka Protocol, Change Data CapTure 等。接入层 分布式部署,具有弹性伸缩、自动重试等能力。保证数据接入的稳定性。

存储层

CKafka 连接器的存储层是腾讯云上的消息队列(MQ),默认是 Kafka,也可以支持 Pulsar、RocketMQ 等其他 MQ。MQ 存储层主要起到削峰填谷,数据分发、缓存的作用。

数据处理转出层

CKafka 连接器提供了数据清洗(ETL)引擎,根据配置进行数据清洗,并提供了 Connector 的转储能力,可以消费存储层的数据,将数据导入到下游的各个不同的存储引擎。



使用限制

最近更新时间:2024-01-09 14:54:11

本文列举了CKafka 连接器中对一些指标和性能的限制,请您在使用中注意不要超出对应的限制值,避免出现异常。

限制项	说明
任务并行 度	数据接入、处理、转储底层都是多个子任务(worker)并发运行的。默认情况下,worker并行度为1。系统会自动检查上有数据是否有堆积,当有堆积时会自动扩容运行 worker 数,以提高任务处理能力。当前阶段,worker 数的上限 = kafka Topic 的分区数。worker 数用户不可配置,系统会自动扩缩容 worker。
任务吞吐 性能	CKafka 连接器是以任务形态进行,其任务性能依赖上下游组件的服务能力。如上游是 Kafka > DIP > Elasticsearch 链路。当上下游无性能瓶颈时,DIP 会通过调整任务并行度来提高数据处理能力,当上下游有性能瓶颈时,则数据的流转耗时会相应增加。客户可以通过查看监控指标和配置堆积告警等手段来发现瓶颈。
任务数量	默认情况下,每个客户的任务数量上限是200。如果需要更多的数量,可以提工单申请。
连接数量	默认情况下,每个客户的连接数量上限是100。如果需要更多的数量,可以提工单申请。
主题数量	默认情况下,每个客户的主题数量上限是200。如果需要更多的数量,可以提工单申请。
Schema 数量	默认情况下,每个客户的主题数量上限是100。每个 Schema 最多允许100个字段。如果需要更多的数量,可以提工单申请。
HTTP 接 入 QPS 限 制	默认情况下,每个 HTTP 接入点的 QPS 的限制为2000。如果需要更多的数量,可以提工单 申请。
数据上报 每批次数 据总大小	HTTP 数据上报的每批次上报的数据大小最大为5MB,超过该数据大小就会报错。
数据上报 每批次数 据条数	HTTP 数据上报的每批次上报的数据条数最大为500条,超过该数据大小就会报错。
极端情况 任务数据 丢失	数据处理 和 数据流出 任务本质上是创建 CKafka 生产者或消费者,对选定的任务数据源 topic 进行生产消费。 当消费成功,即数据成功投递到数据目标资源后,任务对应的消费者组(datahub-<任务ID>)才 会提交数据对应的消息 offset。但如果任务在成功投递数据后,在提交 offset 前出现异常情况而 导致重启,此时数据将会被重复投递到目标资源中。因此建议使用者在业务逻辑代码中对于极端 情况做好幂等性等处理。 当消息还未被成功消费,但消息已经到了设置的消息最大保留时间(包括动态消息保留策略,以 及手动设置的 topic 的消息保留策略)而被提前删除时,任务将无法消费这部分过期消息,导致数



据目标缺少这部分消息对应的数据。



连接管理开放网段

最近更新时间:2024-07-19 14:26:17

场景说明

由于部分组件需要在其安全组内配置开放相应的端口及网段才能使用 CKafka 的连接器功能,本文档提供相应地域需要开放的端口及网段。

开放说明

对应组件需要开放的端口

MySQL

统一需要开放端口 TCP:3306(如有自定义端口,请根据实际设置端口进行开放)。

MariaDB

统一需要开放端口 TCP:3306(如有自定义端口,请根据实际设置端口进行开放)。

PostgreSQL

统一需要开放端口 TCP:5432(如有自定义端口,请根据实际设置端口进行开放)。

MongoDB

统一需要开放端口 TCP:27017(如有自定义端口,请根据实际设置端口进行开放)。

SQLServer

统一需要开放端口 TCP:1433(如有自定义端口,请根据实际设置端口进行开放)。

ClickHouse

统一需要开放端口TCP:9000(如有自定义端口,请根据实际设置端口进行开放)。

Kafka (自建)

暂时只支持云梯环境/支撑环境(指腾讯内部自研环境)的 kafka 自建实例,统一需要开放端口 TCP:9092(如有自定 义端口,请根据实际设置端口进行开放),需要根据使用地域开放下面列表中对应的网段。

对应地域需要开放的网段

每个地域都需要开放的网段:

请提交工单联系我们,以获得开放的网络。





连接管理

最近更新时间:2024-01-09 14:54:11

操作场景

当有大量的数据任务共用同一个数据源或者数据目标时,每次创建任务时需要重复填写数据源或者数据目标的配置 信息,操作比较繁琐。

CKafka 连接器支持单独构建连接,连接创建完成后,可以直接关联到具体的数据任务,作为数据源或者数据目标, 无需再重复配置,降低用户的操作成本。

一个连接可以关联多个数据任务。目前 CKafka 连接器支持的连接类型有:数据仓库(Clickhouse),时序数据库(CTSDB),分析型数据库 Doris,数据传输服务(DTS),Elasticsearch Service, MongoDB, MySQL,

PostgreSQL, TDSQL-C, MariaDB和 SQL Server。

本文介绍在 CKafka 连接器中创建, 修改, 删除连接的使用方法。

操作步骤

新建连接

1. 登录 CKafka 控制台。

2. 在左侧导航栏选择连接器 > 连接列表,选择好地域后,单击新建连接。

3. 在弹窗中选择好链接类型后,单击下一步,填写连接配置信息。

数据仓库(Clickhouse)

时序数据库 (CTSDB)

分析型数据库 Doris

数据传输服务 (DTS)

Elasticsearch Service

MongoDB

MySQL

PostgreSQL

TDSQL-C

MariaDB

SQL Server

云数据仓库-PostgreSQL

TDSQL PostgreSQL 版

连接名称:填写连接名称。

描述:选填,填写连接说明。



数据仓库类型:支持云上 Clickhouse 和自建 Clickhouse。

云上 Clickhouse: 云数据仓库 ClickHouse 在创建的时候就已经使用私有连接进行封装,因此可以直接在控制台选择 对应的 **云数据仓库 ClickHouse** 实例,数据流出会自动进行相关 VPC 网络的打通。

自建 Clickhouse:由于 CKafka 实例采用的是托管实例的形式,而 emr ClickHouse 是直接在购买的 CVM 上创建公网 路由,因此需要用户手动在此基础上创建负载均衡服务,才能打通 VPC 服务。下面以 emr ClickHouse 为例创建负载 均衡:

1.1.1 打开 emr 控制台,选取目标集群,依次单击 集群资源,节点状态,在状态页中找到 ClickHouse 的节点 IP。1.1.2 进入负载均衡 控制台,新建一个负载均衡实例后,单击上方导航栏的 监听器管理,在页面中单击新建

TCP/UDP/TCP SSL 监听器,在端口中填写数据流出时用到的端口。

1.1.3 创建监听器后,单击 绑定后台服务,输入 ClickHouse 中的 tcp 端口,默认应为 9000。

1.1.4 绑定完成后,即可在数据流出控制台选择当前创建的负载均衡服务,端口填写的是 负载均衡 服务中监听的端口。

说明:

目前仅支持创建与负载均衡同一地域的数据流出 ClickHouse 服务。

用户名: ClickHouse 的用户名(默认为 default)

密码: ClickHouse 的密码。

出于安全考虑,数据流出 ClickHouse 中密码为必填项。目前实例创建后可能密码为空,此时需要在配置文件 user.xml 中修改密码。具体可参见 ClickHouse 官方文档。

连接名称:填写连接名称。

描述:选填,填写连接说明。

CTSDB 地址:选择连接的 CTSDB 数据库地址。

用户名:填写数据库用户名。

密码:填写数据库的密码。

连接名称:填写连接名称。

描述:选填,填写连接说明。

源数据库类型:默认自建/EMR Doris。

CLB 实例: 仅支持内网的 CLB, 且该 CLB 需要同时挂载 Doris FE 和 BE 的端口。

FE 端口:FE jdbc 端口,默认 jdbc 端口为 9030。

BE 端口:BE http 端口,默认 http 端口为 8040。

用户名:填写数据库用户名。

密码:填写数据库的密码。

连接名称:填写连接名称。

描述:选填,填写连接说明。

DTS 实例:选择 DTS 实例, DTS 订阅 Topic 分区数量要与目标转储 Kafka 的 Topic 分区数量一致。

DTS 消费组:选择 DTS 消费组。

消费者账号:DTS 消费组的账号。

消费组密码:DTS 消费组的密码。



连接名称:填写连接名称。 描述:选填,填写连接说明。 ES 实例集群:选取腾讯云 Elasticsearch Service 实例集群。 实例用户名:输入 Elasticsearch 实例用户名、腾讯云 Elasticsearch 默认用户名为 elastic,且不可更改。 实例密码:输入 Elasticsearch 实例密码。 连接名称:填写连接名称。 描述:选填,填写连接说明。 源数据库类型: 腾讯云 MongoDB:选择数据库实例。 自建 MongoDB:选择用户 CLB 实例并指定端口。 用户名:源 MongoDB 的用户名。 密码:源 MongoDB 的密码。 连接名称:填写连接名称。 描述:选填,填写连接说明。 源数据库类型: 腾讯云 MySQL:选择数据库实例。 自建 MySQL:选择用户 CLB 实例并指定端口。仅支持内网的 CLB, 且受 MySQL 同步机制限制, 基于 CLB 方案的 数据接入仅支持 CLB 挂载 1 台主库。 用户名:填写 MySQL 的用户名。 密码:填写 MySQL 的密码。 连接名称:填写连接名称。 描述:选填,填写连接说明。 源数据库类型: 腾讯云 PostgreSQL:选择数据库实例。 自建 PostgreSQL:选择用户 CLB 实例并指定端口。仅支持内网的 CLB, 且受 PostgreSQL 同步机制限制, 基于 CLB 方案的数据接入仅支持 CLB 挂载 1 台主库。 用户名:填写 PostgreSQL 的用户名。 密码:填写 PostgreSQL 的密码。 连接名称:填写连接名称。 描述:选填,填写连接说明。 数据库类型:支持 PostgreSQL 和 MySQL 两种类型。 数据库实例:选择数据库实例,目前仅支持运行中的实例。 用户名:填写数据库用户名。 密码:填写数据库密码。 连接名称:填写连接名称。 描述:选填,填写连接说明。 数据库实例:选择数据库实例。 用户名:填写数据库用户名。



密码:填写数据库密码。 连接名称:填写连接名称。 描述:选填,填写连接说明。 数据库实例:选择数据库实例。 用户名:填写数据库用户名。 密码:填写数据库密码。 连接名称:填写连接名称。 描述:选填,填写连接说明。 数据库实例:选择数据库实例。 用户名:填写数据库用户名。 密码:填写数据库密码。 连接名称:填写连接名称。 描述:选填,填写连接说明。 数据库实例:选择数据库实例。 用户名:填写数据库用户名。 密码:填写数据库密码。

4. 连接配置信息填写后,单击**下一步**,开始进行连接校验,校验成功后,连接创建完成,在连接列表可以看到创建 好的连接。

编辑配置

1. 登录 CKafka 控制台。

2. 在左侧导航栏选择连接器 > 连接列表, 单击目标连接的"ID", 进入基本信息页面。

3. 单击基本信息模块右上角的**编辑配置**,修改连接配置信息,您可以选择是否开启**更新并重启所有关联任务**,开后 后,连接关联的的所有任务都将更新并重启。

查看关联任务

1. 登录 CKafka 控制台。

2. 在左侧导航栏选择连接器 > 连接列表, 单击目标连接的"ID", 进入基本信息页面。

3. 在页面上方选择关联任务页签,可以查看到该连接关联的任务列表。您可以通过数据源或者数据目标来筛选关联 的任务类型。

删除连接

在连接列表页面,单击操作列的删除,在弹窗中二次确认后,即可删除连接。

注意:

只有关联任务数为0时连接才可以被删除。



任务管理 新建数据接入任务 HTTP上报

最近更新时间:2024-01-09 14:54:12

操作场景

Datahub 支持接入各种数据源产生的不同类型的数据,统一管理,再分发给下游的离线/在线处理平台,构建清晰的数据通道。

本文以 HTTP 数据为例介绍如何在 CKafka 控制台创建数据主动上报任务,并对任务进行修改配置,帮助您更好地了 解数据接入功能。

操作步骤

创建数据接入任务

前提条件:已创建好目标 CKafka 实例和 Topic。

- 1. 登录 CKafka 控制台。
- 2. 在左侧导航栏单击数据接入,选择好地域后,单击新建任务。
- 3. 在弹窗中数据源类型选择主动上报。

Create Task			×
1 Select Data Type	Source > (2	Configure Task	
Data Source Type	Actively reported data	Asynchronously pulled data	
	НТТР		
		Next	



4. 单击下一步,填写任务名称,选择创建好的目标 CKafka 实例和 Topic。

Create Task	
Select Da Type	ata Source > 2 Configure Task
Task Name	Please enter a name
	It can only contain letters, digits, underscores, or symbols ("-" and ".").
CKafka Instance	Please select
Target Topic	Please select
	The data access feature cannot be normally used if ACL policies are configured for the selected topic.
Schema	Please select
	Create Schema 🔀
	Show advanced configuration
	Previous

任务名称:填写任务名称,只能包含字母、数字、下划线、"-"、"."。

CKafka 实例:选择目标 CKafka 实例。

目标 CKafka Topic:选择数据投递的目标 CKafka Topic,若选择的 Topic 设置了ACL 策略,会影响功能正常使用。 Schema:绑定的 Schema 后,将会按该 Schema 对数据进行格式校验。若无合适的 Schema 可以单击 新建 schema 跳转至新建页面。

QPS 限制:填写 QPS 限制。

5. 单击**提交**,任务创建成功后会生成接入点信息。

6. 复制接入点信息到 SDK 中使用, 用于写入数据。

说明:

详细说明请参见 数据上报 SDK。

修改数据目标

1. 登录 CKafka 控制台。

2. 在左侧导航栏单击数据接入,单击目标任务的 ID,进入任务基本信息页面。



3. 单击数据接入模块右上角的更改数据目标,修改数据接入目标。 说明: 仅支持切换目标 CKafka Topic,不支持修改 CKafka 实例。 切换 Topic 会有一定延时,新数据目标会在大约一分钟内生效。 更改数据目标不会重新生成接入点。

4. 单击提交, 完成数据目标修改。

绑定/解绑 Schema

若用户在新建任务时没有绑定 Schema,后续也可以再进行绑定。同时也支持解绑 Schema。操作路径如下: 1. 登录 CKafka 控制台。

2. 在左侧导航栏单击**数据接入**,单击目标任务的 ID,进入任务基本信息页面,在基本信息模块可以绑定/解绑 Schema。

查看监控

1. 登录 CKafka 控制台。

2. 在左侧导航栏单击数据接入,单击目标任务的 ID,进入任务基本信息页面。

3. 选择**监控**页签,可查看目标 Topic 监控数据。

暂停任务

在 数据接入页面,单击目标任务的操作栏的暂停,可暂停任务。

说明:

当您发现数据接入任务影响了 CKafka 正常服务时,可以暂停数据接入。

恢复任务

在数据接入页面,单击目标任务的操作栏的恢复,可将暂停任务恢复。

说明:

处于暂停状态的任务可以重新启动,将继续转储数据。

删除任务

在 数据接入 页面,单击目标任务的操作栏的删除,在二次确认弹窗中单击确认,可删除任务。 说明:

删除任务表示停止数据接入并删除任务记录,不会影响到已经转储的数据和相关的 CKafka 实例。 任务一旦删除不可恢复,请您谨慎操作。



对象存储 COS

最近更新时间:2024-01-09 14:54:11

操作场景

Datahub 支持接入各种数据源产生的不同类型的数据,统一管理,再分发给下游的离线/在线处理平台,构建清晰的数据通道。

本文以 COS 数据为例介绍如何在 CKafka 控制台创建数据异步拉取任务,并对任务进行修改配置,帮助您更好地了 解数据接入功能。

操作步骤

创建数据接入任务

前提条件

已创建好目标 CKafka 实例和 Topic。 已创建好存储桶和对象。

操作步骤

1. 登录 CKafka 控制台。

2. 在左侧导航栏单击数据流入,选择好地域后,单击新建任务。

3. 在弹窗中数据源类型选择**异步拉取 > COS**。



Create Task				×
1 Select Data Type	Source > (2	Configure Tasl	k	
Data Source Type	Actively reported data	Asynchronously p	pulled data	
	MongoDB		DTS	
	© cos			
		Next		

4. 单击**下一步**,填写任务详情。

×



Create Task	
Select Data So Type	ource > 2 Configure Task
Task name	Please enter a name
	It can only contain letters, digits, underscores, or symbols ("-" and".").
Target CKafka Instance	Please select 🔹
Target Topic	Please select 🔹
	The data access feature cannot be normally used if ACL policies are configured for the selected topic.
Source Bucket	Please select V
Source Object	Please select 💌
	The current topic supports a max message size of 12 MB. If the size of a single data row exceeds this limit, data will not be written to the topic.
Role Authorization (✓ Authorize COS
	You need to grant permissions to a third-party role to access COS.
	Previous Submit

任务名称:只能包含字母、数字、下划线、"-"、"."。 CKafka 实例:选择 CKafka 实例。

UNAIKA 头例,远择 UNAIKA 头例。

目标 CKafak topic:选择数据投递的目标 CKafka Topic。

源存储桶:选择源数据存储桶。

源对象:选择源数据对象,当前 Topic 支持的最大消息大小为12MB,单行文本数据超过12MB 会写入失败。 角色授权:使用对象存储(COS)产品功能,您需要授予一个第三方角色代替您执行访问相关产品权限。 5.单击**提交**,完成任务创建。

更改数据源和数据目标

1. 登录 CKafka 控制台。

- 2. 在左侧导航栏单击数据接入,单击目标任务的ID,进入任务基本信息页面。
- 3. 单击数据目标模块右上角的更改数据目标,修改数据目标信息。



查看监控

1. 登录 CKafka 控制台。

2. 在左侧导航栏单击**数据接入**,单击目标任务的ID,进入任务基本信息页面。 3. 选择**监控**页签,可查看目标 Topic 监控数据。

暂停任务

在 数据接入页面,单击目标任务的操作栏的暂停,可暂停任务。

说明:

当您发现数据接入任务影响了 CKafka 正常服务时,可以暂停数据接入。

恢复任务

在数据接入页面,单击目标任务的操作栏的恢复,可将暂停任务恢复。

说明:

处于暂停状态的任务可以重新启动,将继续转储数据。

删除任务

在 数据接入 页面,单击目标任务的操作栏的删除,在二次确认弹窗中单击确认,可删除任务。 说明:

删除任务表示停止数据接入并删除任务记录,不会影响到已经转储的数据和相关的 CKafka 实例。 任务一旦删除不可恢复,请您谨慎操作。



数据传输服务 DTS

最近更新时间:2024-01-09 14:54:12

操作场景

Datahub 支持接入各种数据源产生的不同类型的数据,统一管理,再分发给下游的离线/在线处理平台,构建清晰的数据通道。

本文以 DTS 数据为例介绍如何在 CKafka 控制台创建数据异步拉取任务,并对任务进行修改配置,帮助您更好地了 解数据接入功能。

操作步骤

创建数据接入任务

前提条件

已创建好目标 CKafka 实例和 Topic。

已创建 DTS 实例 和 消费组。

操作步骤:

1. 登录 CKafka 控制台。

2. 在左侧导航栏单击数据流入,选择好地域后,单击新建任务。

3. 在弹窗中数据源类型选择**异步拉取 > DTS**。

Create Task			×
1 Select Data Type	Source > (2	Configure Task	
Data Source Type	Actively reported data	Asynchronously pulled data	
	MongoDB		
	log cos		
		Next	

4. 单击下一步,填写任务详情。

>



Create Task	
Select Data Source Type	Configure Task
Task name	Please enter a name
	It can only contain letters, digits, underscores, or symbols ("-" and".").
Target CKafka Instance	Please select
Target Topic	Please select 🔹
	The data access feature cannot be normally used if ACL policies are configured for the selected topic.
DTS Instance	Please select
	The partition count of the topics that are subscribed to in DTS must be set to the same as that of the target Kafka topics.
DTS Consumer Group	Please enter
Consumer Group Account	Please enter
Consumer Group Password	Please enter
	Previous Submit

任务名称:只能包含字母、数字、下划线、"-"、"."。

CKafka 实例:选择 CKafka 实例。

目标 CKafak topic:选择数据投递的目标 CKafak Topic。

DTS 实例:选择 DTS 实例, DTS 订阅 Topic 分区数量要与目标转储 Kafka 的 Topic 分区数量一致。

DTS 消费组:选择 DTS 消费组。

消费者账号:DTS 消费组的账号。

消费组密码:DTS 消费组的密码。

5. 单击**提交**,完成任务创建。

更改数据源和数据目标



1. 登录 CKafka 控制台。

2. 在左侧导航栏单击数据接入,单击目标任务的 ID,进入任务基本信息页面。

3. 单击数据目标模块右上角的更改数据目标,修改数据目标信息。

查看监控

1. 登录 CKafka 控制台。

2. 在左侧导航栏单击**数据接入**,单击目标任务的 ID,进入任务基本信息页面。

3. 选择**监控**页签,可查看目标 Topic 监控数据。

暂停任务

在 数据接入页面,单击目标任务的操作栏的暂停,可暂停任务。

说明:

当您发现数据接入任务影响了 CKafka 正常服务时,可以暂停数据接入。

恢复任务

在 数据接入页面,单击目标任务的操作栏的恢复,可将暂停任务恢复。

说明:

处于暂停状态的任务可以重新启动,将继续转储数据。

删除任务

在 数据接入页面,单击目标任务的操作栏的删除,在二次确认弹窗中单击确认,可删除任务。 说明:

删除任务表示停止数据接入并删除任务记录,不会影响到已经转储的数据和相关的 CKafka 实例。 任务一旦删除不可恢复,请您谨慎操作。



MongoDB

最近更新时间:2024-01-09 14:54:11

操作场景

Datahub 支持接入各种数据源产生的不同类型的数据,统一管理,再分发给下游的离线/在线处理平台,构建清晰的数据通道。

本文以 MongoDB 为例介绍如何在 CKafka 控制台创建数据异步拉取任务,并对任务进行修改配置,帮助您更好地了 解数据接入功能。

操作步骤

创建数据接入任务

- 1. 登录 CKafka 控制台。
- 2. 在左侧导航栏单击数据流入,选择好地域后,单击新建任务。
- 3. 在弹窗中数据源类型选择 异步拉取 > MongoDB。

Create Task			×
1 Select Data Type	Source > (2	Configure Task	
Data Source Type	Actively reported data	Asynchronously pulled data	
	MongoDB	E DTS	
	left cos		
		Next	

4. 单击**下一步**,填写任务详情。



Create Task	
Select Data So Type	ource > 2 Configure Task
Task name	Please enter a name
	It can only contain letters, digits, underscores, or symbols ("-" and".").
Target CKafka Instance	Please select
Target Topic	Please select 🔹
	The data access feature cannot be normally used if ACL policies are configured for the selected topic.
Source Database Type	TencentDB for MongoDB Self-built MongoDB
Database Instance	Please select 🔹
Username	Enter the user name
Password	Please enter
Database	Asterisk (*) by default
	You cannot select the default MongoDB database for data import.
Collection	Asterisk (*) by default
	Show advanced configuration
	Previous

任务名称:只能包含字母、数字、下划线、"-"、"."。 CKafka 实例:选择 CKafka 实例。 目标 CKafak topic:选择数据流入的目标 CKafka topic。 源数据库类型: 腾讯云 MongoDB:选择数据库实例。 自建 MongoDB:选择用户 CLB 实例并指定端口。 用户名:源 MongoDB 的用户名。





密码:源 MongoDB 的密码。

database:源 MongoDB 的数据库名,不支持导入 MongoDB 默认数据库数据。 collection:源 MongoDB 的集合,默认是监听所有即可,即"",也可以指定某个集合。 复制存量数据:是否复试源 MongoDB 的存量数据。 5.单击**提交**,完成任务创建。

更改数据源和数据目标

1. 登录 CKafka 控制台。

2. 在左侧导航栏单击数据接入,单击目标任务的 ID,进入任务基本信息页面。

3. 单击数据源模块右上角的更改数据源,修改数据源信息。

4. 单击数据目标模块右上角的更改数据目标,修改数据目标信息。

查看监控

1. 登录 CKafka 控制台。

2. 在左侧导航栏单击数据接入,单击目标任务的 ID,进入任务基本信息页面。
 3. 选择监控页签,可查看目标 Topic 监控数据。

暂停任务

在数据接入页面,单击目标任务的操作栏的暂停,可暂停任务。

说明:

当您发现数据接入任务影响了 CKafka 正常服务时,可以暂停数据接入。

恢复任务

在 数据接入 页面,单击目标任务的操作栏的恢复,可将暂停任务恢复。

说明:

处于暂停状态的任务可以重新启动,将继续转储数据。

删除任务

在 数据接入页面,单击目标任务的操作栏的删除,在二次确认弹窗中单击确认,可删除任务。 说明:

删除任务表示停止数据接入并删除任务记录,不会影响到已经转储的数据和相关的 CKafka 实例。 任务一旦删除不可恢复,请您谨慎操作。



新建数据流出任务 数据目标 流出至 Clickhouse

最近更新时间:2024-01-09 14:54:11

操作场景

DataHub 提供数据流出能力,您可以将 CKafka 数据分发至数据仓库 ClickHouse 以对数据进行存储、查询和分析。

前提条件

若使用腾讯云维护的 ClickHouse 产品,使用时需开通相关产品功能。同时也支持数据流出至自建 ClickHouse。 在 ClickHouse 建好表,建表的时候需要指定好 Column 和 Type。

操作步骤

创建任务

- 1. 登录 CKafka 控制台。
- 2. 在左侧导航栏单击数据流出,选择好地域后,单击新建任务。
- 3. 目标类型选择数据仓库 ClickHouse。

Create Task	×
Select Target Type	Configure Task
Task name	Please enter a name
	It can only contain letters, digits, underscores, or symbols ("-" and".").
Target CKafka Instance	Please select 🔹
Source Topic	Please select
	The data distribution feature cannot be used if the selected tonic is cont



	with ACL policies. Data in this the same format.	topic can be successfully du	mped only if it
Source Data	Preview Data		
	The source data must be in JS	ON format.	
Starting Position	O Start consumption from the	e latest position	
	Start consumption from the	e starting position	
	 Start consumption from the 	e time-in-point position	٦
Data Warehouse Type	Cloud Data Warehouse	Self-built ClickHouse	
Data Warehouse Instance	Please select		•
Specified Port	TCP port required		
Username	Enter the user name		
Password	Please enter		
Cluster	Enter a ClickHouse cluster		*
Database	Required field		*
Table	Required field		v
Discard Message with Parsing Failure ()			
	Previous Submit		

配置 CKafka 数据源

在页面中设置 CKafka 数据源相关设置,相关配置如下:

1. 任务名称:只能包含字母、数字、下划线、"-"、"."。

2. CKafka 实例:选择数据源 CKafka。

3. **源 Topic**:选择实例下的 Topic,一条数据流出任务最多支持选择 5 个源 Topic,选中的 Topic 内的数据格式需要保持一致方可转储成功。

消息解析

上述设置后,单击**预览数据**,将会选取**源 Topic**中的第一条消息进行解析。



说明:

目前解析消息需要满足以下条件:

消息为 JSON 字符串结构。

解析后的消息为单层 JSON,即目前无法解析嵌套结构的 JSON。

当消息不为单层 JSON 结构时,推荐使用数据处理进行简单的消息格式转换。

单击解析后,控制台将会出现解析后的消息字段,可以通过修改预览结果中的 type 属性来确定投递到目标对应列的 类型。

当选择 type 为 Date 或 DateTime 时,如果源消息格式为整型,将会尝试使用 unix timestamp 格式解 析;如果源消息格式为字符串,将会尝试用常用的时间格式模式串解析。

流出配置

根据流出目标 ClickHouse 不同,分为了 云数据仓库 ClickHouse 和 自建 ClickHouse 两类。

云数据仓库 ClickHouse

自建或 emr ClickHouse

由于 **云数据仓库 ClickHouse** 在创建的时候就已经使用私有连接进行封装,因此可以直接在控制台选择对应的 **云数 据仓库 ClickHouse** 实例,数据流出会自动进行相关 VPC 网络的打通。

由于 CKafka 实例采用的是托管实例的形式,而 emr ClickHouse 是直接在购买的 CVM 上创建公网路由,因此需要用 户手动在此基础上创建负载均衡服务,才能打通 VPC 服务。下面以 emr ClickHouse 为例创建负载均衡:

1. 打开 emr 控制台,选取目标集群,依次单击 集群资源,节点状态,在状态页中找到 ClickHouse 的节点 IP。

2. 进入负载均衡控制台,新建一个负载均衡实例后,单击上方导航栏的监听器管理,在页面中单击新建

TCP/UDP/TCP SSL 监听器,在端口中填写数据流出时用到的端口。

3. 创建监听器后,单击绑定后台服务,输入 ClickHouse 中的 tcp 端口,默认应为 9000。

4. 绑定完成后,即可在数据流出控制台选择当前创建的负载均衡服务,端口填写的是 负载均衡 服务中监听的端口。
 说明:

目前仅支持创建与负载均衡同一地域的数据流出 ClickHouse 服务。

网络打通后,还需要配置数据流出实例其余选项,说明如下:

用户名:目标 ClickHouse 的用户名(默认为 default)。

密码:目标 ClickHouse 的密码。

注意:

出于安全考虑,数据流出 ClickHouse 中密码为必填项。

目前实例创建后可能密码为空,此时需要在配置文件 user.xml 中修改密码。具体可参考 ClickHouse 官方文档

0

cluster: **ClickHouse** 的集群名称(默认为 default_cluster)。

database: ClickHouse 设置的数据库名称。

table:在该数据库内构建的表名称,目前数据流出 ClickHouse 服务不会自动创建表,**需要客户手动创建当前** ClickHouse 目标表。

CIICKHOUSE 日你衣。

丢弃解析消息:消息解析失败原因一般是消息字段与目标库字段 type 不一致。若不丢弃解析失败消息,则任务异常,转储不再继续。



最后单击**提交**,即可完成任务创建。

监控配置

1. 登录 CKafka 控制台。

- 2. 在左侧导航栏单击数据流出,单击目标任务的 ID,进入任务基本信息页面。
- 3. 在任务详情页顶部,单击**监控**,选择要查看资源,设置好时间范围,可以查看对应的监控数据。

产品限制和费用计算

转储速度与 CKafka 实例峰值带宽上限有关,如出现消费速度过慢,请检查 CKafka 实例的峰值带宽或增加 CKafka partition 数。



流出至日志服务 CLS

最近更新时间:2024-01-09 14:54:11

操作场景

Datahub 提供数据流出能力,您可以将 CKafka 数据分发至日志服务 CLS 便于解决业务问题定位,指标监控,安全审计等日问题。

前提条件

该功能目前依赖 CLS 服务,使用时需开通相关产品功能。

操作步骤

1. 登录 CKafka 控制台。

- 2. 在左侧导航栏单击数据流出,选择好地域后,单击新建任务。
- 3. 目标类型选择**日志服务(CLS)**,单击**下一步**。





Create Task		×
Select Target	Type > 2 Configure Task	
Task name	Please enter a name	
	It can only contain letters, digits, underscores, or symbols ("-" and".").	
Target CKafka Instance	Please select	
Source Topic	Please select	
	The data distribution feature cannot be used if the selected topic is configured with ACL policies.	
Source Data	Preview Data	
	Parse the data into JSON and ship it	
KEY ()	Default: content	
Starting Position ()	 Start consumption from the latest position 	
	Start consumption from the starting position	
	Start consumption from the time-in-point position	
Logset	Please select	
Log Topic	Please select 👻	
	If there is no suitable logset or log topic, you can go to CLS 🛂 to create one.	
Role Authorization (i)	✓ Authorize CLS	
	You need to grant permissions to a third-party role to access CLS.	
	Previous Submit	

任务名称:只能包含字母、数字、下划线、"-"、"."。

CKafka 实例:选择数据源 CKafka。

源 Topic:选择源 Topic, 一条数据流出任务最多支持选择5个源 Topic。

起始位置:转储时历史消息的处理方式, topic offset 设置。

日志集:选择日志集,日志集日志服务的项目管理单元,用于区分不同项目的日志。



日志主题:自动创建日志主题或者选择已有日志主题。一个日志集可以包含多个日志主题,一个日志主题对应一类应用或服务,建议将不同机器上的同类日志收集到同一个日志主题。 角色授权:使用日志服务(CLS)产品功能,您需要授予一个第三方角色代替您执行访问相关产品权限。 4.单击**提交**,完成任务创建。

查看监控

1. 登录 CKafka 控制台。

2. 在左侧导航栏单击数据流出,单击目标任务的 ID,进入任务基本信息页面。
 3. 在任务详情页顶部,单击监控,选择要查看资源,设置好时间范围,可以查看对应的监控数据。

产品限制和费用计算

转储速度与 CKafka 实例峰值带宽上限有关,如出现消费速度过慢,请检查 CKafka 实例的峰值带宽或增加 CKafka partition 数。

转储速度与 CKafka 单个文件大小相关,如超过该500M,会自动分包上传。

该功能基于云函数 SCF 服务提供。SCF 为用户提供了一定 免费额度 ,超额部分产生的收费,请以 SCF 服务的 计费 规则 为准。



数据流出至对象存储 COS

最近更新时间:2024-01-09 14:54:11

操作场景

Datahub 提供数据流出能力,您可以将 CKafka 数据分发至 COS 以便于对数据进行分析与下载等操作。

前提条件

该功能目前依赖对象存储(COS)产品,使用时需开通相关产品功能。

操作步骤

创建任务

- 1. 登录 CKafka 控制台。
- 2. 在左侧导航栏单击数据流出,选择好地域后,单击新建任务。
- 3. 目标类型选择**对象存储(COS)**,单击**下一步**。



🔗 腾讯云	
-------	--

Create Task		×
Select Target	Type > 2 Configure Task	
Task name	Please enter a name	
	It can only contain letters, digits, underscores, or symbols ("-" and".").	
Target CKafka Instance	Please select	
Source Topic	Please select 🔹	
	The data distribution feature cannot be used if the selected topic is configured with ACL policies.	
Source Data	Preview Data	
Starting Position (O Start consumption from the latest position	
	Start consumption from the starting position	
	Start consumption nom the time-in-point position	
Target Bucket	Please select C •	
Aggregation Mode(24 or 50 MB	
	Range: 1-24 hours, 1-50 MB	
Storage Format	Please select	
Role Authorization (i)	✓ Authorize COS	
	You need to grant permissions to a third-party role to access COS.	
	Previous Submit	

4. 在任务设置页面,填写任务详情。

任务名称:只能包含字母、数字、下划线、"-"、"."。 CKafka 实例:选择数据源 CKafka。 源 Topic:选择源 Topic, 一条数据流出任务最多支持选择5个源 Topic。

起始位置:转储时历史消息的处理方式, topic offset 设置。

目标存储桶:对不同的 Topic,选取相应的 COS 中 Bucket,则请求消息会自动在 Bucket 下创建 instance-id/topic-id/date/timestamp 为名称的文件路径进行存储。



聚合方式:请至少填写一种聚合方式,文件将根据指定方式聚合进入 COS 存储桶。如果指定了两种聚合方式,则会同时生效。例:指定每1h或1GB聚合一次,若在1h之前达到1GB,则文件会聚合,同时在1h时也会聚合一次。 注意:

当单条消息大小大于设置的聚合文件大小时,消息可能会被截断。

存储格式:支持 csv 和 json 格式。

角色授权:使用对象存储(COS)产品功能,您需要授予一个第三方角色代替您执行访问相关产品权限。 5.单击**提交**,完成任务创建。

查看监控

1. 登录 CKafka 控制台。

2. 在左侧导航栏单击**数据流出**,单击目标任务的 ID,进入任务基本信息页面。

3. 在任务详情页顶部,单击监控,选择要查看资源,设置好时间范围,可以查看对应的监控数据。

产品限制和费用计算

转储速度与 CKafka 实例峰值带宽上限有关,如出现消费速度过慢,请检查 CKafka 实例的峰值带宽或增加 CKafka partition 数。

转储速度与 CKafka 单个文件大小相关,如超过该500M,会自动分包上传。

当前仅支持和 CKafka 实例同个地域的 COS 进行消息存储,为保证延时,不支持跨地域存储。

使用 COS 消息转储, 文件内容是 CKafka 消息里的 value 用 utf-8 String 序列化拼接而成, 暂不支持二进制的数据格 式。

开启转 COS 的操作账号必须对目标 COS Bucket 具备写权限。

使用 COS 消息转储必须至少拥有一个 VPC 网络环境,如在创建时选择基础网络请参见 路由接入方式 绑定 VPC 网络。



流出至 Elasticsearch Service

最近更新时间:2024-01-09 14:54:11

操作场景

Datahub 提供数据流出能力,您可以将 CKafka 数据分发至 Elasticsearch Service (ES)便于海量数据存储搜索、实时日志分析等操作。

说明:

只支持7.0以上版本的 Elasticsearch Service。

前提条件

该功能目前依赖 Elasticsearch Service 服务,使用时需开通相关产品功能。

操作步骤

创建数据流出任务

- 1. 登录 CKafka 控制台。
- 2. 在左侧导航栏单击数据流出,选择好地域后,单击新建任务。
- 3. 目标类型选择 Elasticsearch Service, 单击下一步。



Create Task		×
Select Target Type >	2 Configure Task	
Task name	Please enter a name	
	It can only contain letters, digits, underscores, or symbols ("-" and".").	
Target CKafka Instance	Please select 🔹	
Source Topic	Please select 🗸	
	The data distribution feature cannot be used if the selected topic is configured with ACL policies.	S
Source Data	Preview Data	
	The source data must be in JSON format.	
Starting Position	O Start consumption from the latest position	
	 Start consumption from the starting position Start consumption from the time-in-point position 	
Instance Cluster	Please select	
Instance Username	elastic	
Instance Password	Enter the instance password	
	The max message size supported by ES is 32 KB.	
Discard Message with Parsing Failure		
	Previous Submit	

任务名称:只能包含字母、数字、下划线、"-"、"."。 CKafka 实例:选择数据源 CKafka。 源 Topic:选择源 Topic。 源数据:支持拉取源数据。 起始位置:转储时历史消息的处理方式,topic offset 设置。 实例集群:选取腾讯云 Elasticsearch Service 实例集群信息。





实例用户名:输入 Elasticsearch 实例用户名,腾讯云 Elasticsearch 默认用户名为 elastic,且不可更改。 实例密码:输入 Elasticsearch 实例密码。 丢弃解析失败消息:消息解析失败原因一般是转储到 ES 的消息超过了32kb 或者消息与目标 index 的 schema 不一 致。若不丢弃解析失败消息,则任务异常,转储不再继续。 4.单击**提交**,完成任务创建。

查看监控

1. 登录 CKafka 控制台。

2. 在左侧导航栏单击**数据流出**,单击目标任务的ID,进入任务基本信息页面。

3. 在任务详情页顶部,单击监控,选择要查看资源,设置好时间范围,可以查看对应的监控数据。

产品限制和费用计算

转储速度与 CKafka 实例峰值带宽上限有关,如出现消费速度过慢,请检查 CKafka 实例的峰值带宽。



TDSQL-C PostgreSQL

最近更新时间:2024-01-09 14:54:12

操作场景

CKafka 连接器提供数据流出能力,您可以将 CKafka 数据分发至 TDSQL-C PostgreSQL以对数据进行存储、查询和分析。

前提条件

该功能目前依赖 TDSQL TDSQL-C 版产品,使用时需开通相关产品功能。

操作步骤

1. 登录 CKafka 控制台。

- 2. 在左侧导航栏单击连接器 > 任务管理 > 任务列表,选择好地域后,单击新建任务。
- 3. 填写任务名称,任务类型选择数据流出,数据目标类型选择 TDSQL-C PostgreSQL,单击下一步。
- 4. 配置数据源信息。

源 Topic 类型:选择数据源 Topic

弹性 Topic:选择提前创建好的弹性 Topic,详情参见 Topic 管理。

CKafka 实例内 Topic:选择在 CKafka 创建好的实例和 Topic,若实例设置了ACL 策略,请确保选中的 topic 有读写 权限,详情参见 Topic 管理。

起始位置:选择转储时历史消息的处理方式, topic offset 设置。

5. 设置上述信息后,单击**下一步**,单击**预览 Topic 数据**,将会选取**源 Topic**中的第一条消息进行解析。 **说明:**

目前解析消息需要满足以下条件:

消息为 JSON 字符串结构。

源数据必须为单层 JSON 格式, 嵌套 JSON 格式可使用使用 数据处理 进行简单的消息格式转换。

6. (可选)开启数据处理规则,具体配置方法请参见简单数据处理。

7. 单击**下一步**, 配置数据目标信息。

数据目标:选择提前创建好的 TDSQL-C 连接。

database:选择数据流出的数据库。

table:选择数据流出的表。

数据库同步模式:



默认字段匹配:本选项仅用于:

源 topic 数据为 DIP 从 MySQL/PostgreSQL 中订阅的、单张表的 Binlog/row-level changes 数据(增删改)。 源 topic 数据必须有主键,且必须包含 schema。

字段逐一匹配:

源数据:点击拉取源 Topic 数据,需要逐一给消息字段选择匹配的目标表字段。

插入模式:支持 **insert** 或者 **upsert**,选择 upsert 时需选择**主键**(当插入行冲突时,任务将更新冲突行除主键之外其 余的列)。

上游数据格式:支持 JSON 和 Debezium。

说明:

当上游 MySQL Binlog/PostgreSQL row-level changes 数据表结构变化时,变化可以同步更新到下游 PostgreSQL。 失败消息处理:失败消息处理:选择投递失败的消息的处理方式,支持**丢弃、保留**和**投递至 CLS**(需指定投递到的 日志集和日志主题并授权访问日志服务 CLS)三种方式。

保留:适合用于测试环境,任务运行失败时将会终止任务不会重试,并且在事件中心中记录失败原因。

丢弃:适合用于生产环境,任务运行失败时将会忽略当前失败消息。建议使用 "保留" 模式测试无误后,再将任务编辑成 "丢弃" 模式用于生产。

投递至 CLS:适合用于严格生产环境,任务运行失败时会将失败消息及元数据和失败原因上传到指定 CLS 主题中。 8.单击**提交**,可以在任务列表看到刚刚创建的任务,在状态栏可以查看任务创建进度。



分布式数据仓库 (TDW)

最近更新时间:2024-07-12 11:47:01

操作场景

Datahub 提供数据流出能力,您可以将 CKafka 数据分发至分布式数据仓库 TDW 以对数据进行存储、查询和分析。

前提条件

该功能目前依赖分布式数据仓库(TDW)产品,使用时需开通相关产品功能。

操作步骤

创建任务

1. 登录 CKafka 控制台。

- 2. 在左侧导航栏单击数据流出,选择好地域后,单击新建任务。
- 3. 目标类型选择**分布式数据仓库(TDW)**,单击**下一步**。



Create Task		×		
Select Target 1	ype > 2 Configure Task			
Task name	Please enter a name			
	It can only contain letters, digits, underscores, or symbols ("-" and".").			
Target CKafka Instance	Please select			
Source Topic	Please select 🔻			
	The data distribution feature cannot be used if the selected topic is configured with ACL policies.			
Source Data	Preview Data			
Starting Position	osition (i) O Start consumption from the latest position			
	 Start consumption from the starting position 			
	Start consumption from the time-in-point position			
TDW BID	Example: b_teg_dw_b_test_ss_share_requ			
TDW TID	Enter the TID (API name)			
	Previous Submit			

任务名称:只能包含字母、数字、下划线、"-"、"."。 CKafka 实例:选择数据源 CKafka。 源 Topic:选择源 Topic。

源数据:支持拉取源数据。

起始位置:转储时历史消息的处理方式, topic offset 设置。

TDW BID:填写 TDW 业务 BID。

TDW TID:填写 TDW 业务 TID。

4. 单击提交,完成任务创建。

查看监控

1. 登录 CKafka 控制台。

2. 在左侧导航栏单击数据流出,单击目标任务的 ID,进入任务基本信息页面。

3. 在任务详情页顶部,单击监控,选择要查看资源,设置好时间范围,可以查看对应的监控数据。



更改数据源和数据目标

1. 登录 CKafka 控制台。

2. 在左侧导航栏单击数据流出,单击目标任务的 ID,进入任务基本信息页面。

3. 单击数据源模块右上角的更改数据源,修改数据源信息。

4. 单击数据目标模块右上角的更改数据目标,修改数据目标信息。

说明:

更改数据目标,不会重置消费组 offset。

任务暂停过程中,不支持修改数据源和数据目标。

产品限制和费用计算

转储速度与 CKafka 实例峰值带宽上限有关,如出现消费速度过慢,请检查 CKafka 实例的峰值带宽或增加 CKafka partition 数。

转储速度与 CKafka 单个文件大小相关,如超过该500M,会自动分包上传。



简单数据处理 简单数据处理

最近更新时间:2024-01-09 14:54:11

操作场景

DataHub 提供简单的数据处理功能,通过传入数据和配置项,可以实现对数据格式化处理,然后返回处理完成的结构化数据,分发给离线/在线处理平台,构建数据源和数据处理系统间的桥梁。

操作步骤

创建规则

- 1. 登录 CKafka 控制台。
- 2. 在左侧导航栏单击数据处理,选择好地域后,单击新建任务。
- 3. 填写任务基本信息。



 Create Data Processing 	Task	
1 Basic Settings >	2 Data Processing Rule	
Task name	Only contain lattern digita yndernaaren de	
Source Topic	Please select	Please select
Target Topic	Please select	▼ Please select ▼
Starting Position ()	 Start consumption from the latest positi Start consumption from the starting pos Start consumption from the time-in-poir 	on sition nt position
Use EventBridge as Underlying Engine		
Cancel Next		

任务名称:只能包含字母、数字、下划线、"-"、"."。

源 Topic:源数据 CKafka Topic。

目标 Topic:目标数据 CKafka Topic。

起始位置:选择转储时历史消息的处理方式, topic offset 设置。

底层引擎使用事件总线(Event Bridge):用户可以选择底层引擎使用EB。

说明:

仅北京、上海和广州地域支持。

角色授权:使用事件总线(EventBridge)作为底层引擎,您需要授予一个第三方角色代替您执行访问相关产品权限。

4. 单击下一步,设置数据处理规则。



← Create Data Processing Task				
Basic Settings > 2 Data Processing Rule				
Original Data Original Data Original Data				
{"password":"123456"}				
Parsing Mode JSON - OK				
Submit Previous				

原始数据:支持**从源 Topic 拉取**或者自定义。

解析格式:支持 JSON, **分隔符**和**正则提取**三种解析方式。

 JSON_\circ

分隔符:支持 空格 、 制表符 、 , 、 ; 、 | 、 自定义 。 正则提取:需填写正则表达式。 5.选择好解析模式后,单击**确认**,开始解析数据。 6.解析完成后,设置好过滤器规则和数据处理方式。 **说明:**

输出格式目前支持 JSON。



Create Da	ta Processing Task			
Basic Settin	gs 〉 2 Data F Rule	rocessing		
Original Data	Pulled from the source topic	O Custom		
	{"password":"123456"}			
Parsing Mode	JSON 🔻 OK			
Parsing Result	KEY	VALU	JE	
	password	123	3456	
Filter	KEY	Match Mode	VALUE	
	+ Add			
	Only the data that hits the filte	r rule (such as "ErrorCode contair	ns 404") can be output. For details	, see Filter Rule Des
Data Processing	KEY	TYPE	VALUE	
	password	Mapping 🔻	Please select	▼ <u>∎</u>
	+ Add			
	Test Currently, only	JSON is supported as an output f	format.	
Submit	Previous			

过滤器: 仅输出符合过滤器规则的数据。过滤器的匹配模式支持**前缀匹配、后缀匹配、包含匹配(contains)、除 外匹配(except)、数值匹配**和IP**匹配**。详情参见 过滤器规则说明。

数据处理:TYPE 有默认、系统预设、映射、自定义和 JSONPATH 五种。

TYPE = 默认: VALUE 从解析结果中映射,不可编辑。

TYPE = 系统预设:可以选择系统预设的 VALUE,目前支持 DATE(时间戳)。

TYPE = 映射:可以选择已有的 KEY,最终输出的 VALUE 值由指定的 KEY 映射而来。



TYPE = 自定义:可以输入自定义 VALUE。

TYPE = JSONPATH:解析多层嵌套的 JSON 数据,用 \$ 符号开头, . 符号定位到多层 JSON 的具体字段。

7. 单击测试,查看测试结果。

8. 设置失败处理规则。

重试间隔:处理失败重试间隔时间,重试间隔时间的取值范围为60s~1h。

重试次数:处理失败最大重试次数,超过最大次数后将作为失败消息处理。重试次数*时间间隔 <= 6h。 失败消息处理:处理失败消息的处理方式,支持**丢弃、保留**和投递到**死信队列**(需指定死信队列 Topic)。 9.单击**提交**,完成数据处理规则创建。

编辑规则

在 数据处理 任务列表页面,单击目标任务的 ID,进入任务基本信息页面。
 单击处理规则模块右上角的编辑规则,可修改数据处理规则。

更改配置

1. 在数据处理任务列表页面,单击目标任务的 ID,进入任务基本信息页面。

2. 单击**配置信息**模块右上角的**更改配置**,修改数据处理规则配置。

查看监控

1. 在 数据处理 任务列表页面,单击目标任务的 ID,进入任务基本信息页面。

2. 进入任务基本信息页面。

3. 在任务详情页顶部,单击监控,选择要查看资源,设置好时间范围,可以查看对应的监控数据。

暂停任务

在 数据处理 页面,单击目标任务的操作栏的暂停,可暂停任务。

说明:

该操作是异步任务,有延迟,任务状态可能不会立刻改变。

恢复任务

在 数据处理 页面,单击目标任务的操作栏的恢复,可将暂停任务恢复。

说明:

处于暂停状态的任务可以重新启动,并继续处理数据。

删除任务

在 数据处理 页面,单击目标任务的操作栏的删除,在二次确认弹窗中单击确认,可删除任务。 说明:

删除任务表示停止数据处理并删除任务记录,不会影响到已经处理的数据和相关的 CKafka 实例。 任务一旦删除不可恢复,请您谨慎操作。



数据转换

最近更新时间:2024-01-09 14:54:11

在通过 CKafka 连接器进行数据处理的过程中, DIP 支持多种方式对解析后的数据进行格式转换,本文介绍相关的操作方法。

数据转换

输入原始数据,以下为一个示例。





```
{
    "@timestamp": "2022-02-26T22:25:33.210Z",
    "beat": {
        "hostname": "test-server",
        "ip": "6.6.6.6",
        "version": "5.6.9"
    },
    "input_type": "log",
    "message": "{\\"userId\\":888,\\"userName\\":\\"testUser\\"}",
    "offset": 3030131
}
```



DIP 处理方式如下:

方式一:通过选择相应的处理 value 功能一键定义规则。

方式二:通过选择数据类型一键更改相应字段的数据格式。

方式三:通过 JSONPATH 语法实现 join 的拼接功能。

如是

用 \$.concat(\$.data.Response.SubnetSet[0].VpcId,"#",\$.data.Response.SubnetSet[0].SubnetId,"#",\$.data.Response.SubnetSet[0].CidrBlock)) 语法拼接 Vpc 和子网的属性,并且通过 # 字符加以分割。



过滤器规则说明

最近更新时间:2024-01-09 14:54:11

过滤器可以对数据进行规则过滤,例如过滤字段的大小等,符合规则的数据才会被保留。

注意事项

过滤器匹配是逐个字符精确匹配的,需注意大小写,匹配过程中不会对字符串进行任何标准化的操作。 要匹配的值遵循 JSON 规则:用引号引起来的字符串、数字以及不带引号的关键字 true、false 和 null。

前缀匹配

您可以对比数据中的前缀进行键值匹配。

```
例如数据 {"password":"topicname"}, 指定 password 的前缀匹配值为 top ,则 {"password":"topicname"} 可以被正常匹配。
```

后缀匹配

您可以对比数据中的后缀进行键值匹配。 例如数据 {"password":"topicname"} ,指定 password 的后缀匹配值为 name ,则 {"password":"topicname"} 可以被正常匹配。

包含匹配

您可以指定数据中存在的某个字段进行匹配。 例如数据 {"password":"topicname"},指定 password 的包含匹配值为 na ,则 {"password":"topicname"} 可以被正常匹配。

除外匹配

您可以指定某个字段除了提供的值之外的任何值进行匹配。

```
例如数据 {"password":"topicname"},指定 password 的除外匹配值为 topicname,则只有 {"password":"topicname"} 不可以被正常匹配。
```



数值匹配

您可以指定某个字段的数值或范围。

例如数据 { "numeric": 10}, 指定 numeric 的数值匹配为 <15,则 { "numeric": 10} 可以被正常匹 配。

数值匹配规则举例:

大于10输入:>10 大于等于10的输入:>=10 大于等于10且小于等于20输入:>=10&<=20 大于等于10或小于等于5输入:>=10|<=5

IP 地址匹配

您可以指定数据中字段的 IP 地址。输入内容为 CIDR 地址,比如输入1.2.3.4/24即匹配前24位所有1.2.3.开头的 IP 地址。



管理任务

最近更新时间:2024-01-09 14:54:11

本文介绍在 CKafka 控制台管理任务的操作步骤,包含暂停、启动、重启、重建、复制和删除任务等。

编辑数据源

1. 登录 CKafka 控制台。

2. 在 任务列表 页面,单击目标任务的 ID,进入任务基本信息页面。

3. 单击数据源模块右上角的更改数据源,修改数据源信息。MySQL 和 TDSQL-MySQL 数据库订阅任务支持新增订 阅的表。当用户勾选新增订阅的表时:

若原任务在配置数据目标时选择了分发到多个 Topic,则新增订阅表后,需要编辑数据目标,为新增订阅的表指派分发到的 Topic。

若新增的订阅表需要复制存量数据,可以在编辑时打开"复制存量数据"开关。需要设置信息如下:

复制存量数据:选择是否开启复制存量数据开关。开关只对新增监听的表生效。对任务中已经监听的表,仍维持原 有采集逻辑。

存放信令表的数据库名称:开启复制存量数据开关时需要勾选存放信令表的数据库。请确保连接管理中配置的用户 拥有该数据库的创建、修改、删除表的权限(仅用于对信令表的操作)。

编辑数据目标

1. 登录 CKafka 控制台。

2. 在 任务列表 页面,单击目标任务的 ID,进入任务基本信息页面。

3. 单击数据目标模块右上角的更改数据目标,修改数据目标信息。

暂停任务

在任务列表页面,单击目标任务的操作栏的更多 > 暂停任务,二次确认后可暂停任务。

说明:

当您发现数据任务已经影响了 CKafka 正常服务时,可以暂停数据任务。

启动任务

在任务列表页面,单击目标任务的操作栏的更多 > 启动任务,二次确认后可将暂停任务恢复。

说明:

处于暂停状态的任务可以重新启动,将继续转储数据。

重启任务

在任务列表页面,单击目标任务的操作栏的**更多 > 重启任务**,二次确认后可以重新开始任务。 说明:



任务在**异常**状态时,可以重启任务,重启任务表示重新开始任务,不会影响到已经转储的数据和相关的 CKafka 实例。

重建任务

创建失败的任务可能是因为创建任务时的配置失误,用户可以手动重建任务。 1. 在 任务列表 页面,单击目标任务的操作栏的更多 > 重建任务,进入任务设置页面。 2. 指定新的任务名称并编辑数据目标后,单击提交,完成任务重建。

复制任务

当您有大量配置相似的任务时,在第一个任务创建成功后可以通过复制任务功能将任务进行复制。

注意:

新建或者复制任务时不支持创建数据源和数据目标相同的数据任务。

1. 在任务列表页面,单击目标任务的操作栏的更多 > 复制任务,进入任务设置页面。

2. 指定新的任务名称并编辑数据目标后,单击**提交**,完成任务重建。

删除任务

在 任务列表 页面,单击目标任务的操作栏的**删除**,在二次确认弹窗中单击**确认**,可删除任务。 说明:

删除任务表示停止数据接入并删除任务记录,不会影响到已经转储的数据和相关的 CKafka 实例。 任务一旦删除不可恢复,请您谨慎操作。



Schema 管理

最近更新时间:2024-01-09 14:54:11

操作场景

CKafka 连接器支持 Schema 管理功能,您可以将创建好的 Schema 绑定到具体的数据流入任务上,将会按该 Schema 对流入的数据进行格式校验。

操作步骤

创建 Schema

接下来以订阅 MySQL 数据到 ES 为例为您介绍新建链路的操作步骤。

- 1. 登录 CKafka 控制台。
- 2. 在左侧导航栏选择连接器 > Schema 管理,选择好地域后,单击新建 Schema,填写 Schema 信息。

Create Schema				
Schema Name	Please enter a na			
	It can only contain I	etters, digits, unders	cores, or symbols ("-" and".").	
Description	Please enter			
Field Configuration	Field Name	Туре	allowNull (j)	Description
	Field Name	Field Type 🔻	false 👻	Field Descripti
	+ Add			
		Confirm	Cancel	

Schema 名称:填写 Schema 名称,只能包含字母、数字、下划线、"-"、"."。

描述:可选,填写 Schema 描述信息。

字段配置:添加字段,最多支持添加100条字段。



字段名:填写字段名称。 类型:支持 BOOLEAN, INT8, INT16, INT32, INT64, FLOAT32, FLOAT64, STRING 八种类型。 允许为 NULL:校验上游是否存在该字段。允许为 NULL 代表:如果上游没有该字段,自动添加设置的默认值字段。 **说明:** 关于字段配置中是否为 NULL 的详细说明: true: 上游字段存在且符合要求:写入。 上游字段存在且不符合要求:不写入。 上游字段不存在:写入默认值。 false:

上游字段存在且符合要求:写入。 上游字段存在且不符合要求:不写入。 上游字段不存在:不写入。

描述:填写字段描述。

3. 单击**确定**,完成 Schema 创建。

删除 Schema

在 Schema 管理列表页面,单击操作列的删除,在弹窗中二次确认后,即可删除。

绑定任务

Schema 创建完成后,可以绑定到具体的数据流入任务上。

1. 在 Schema 管理列表页面,单击目标 Schema 的"ID",进入基本信息页面。

2. 在页面上方选择关联任务页签,单击绑定任务,勾选具体的数据流入任务,完成绑定。

解绑任务

1. 在 Schema 管理列表页面,单击目标 Schema 的"ID",进入基本信息页面。

2. 在页面上方选择关联任务页签,单击绑定任务操作栏的解绑,完成解绑。

说明:

解绑后,将不会按该 Schema 对数据进行格式校验。



事件中心

最近更新时间:2024-01-09 14:54:11

事件中心将 CKafka 连接器所生成的事件数据进行统一管理、存储、分析和展示,方便您查看和分析,您可以在事件中心查看事件数据详情,同时您也可以为事件配置告警通知规则,及时发现问题并进行处理。 CKafka 连接器当前支持的事件类型有: 更改数据源

更改数据目标

更改数据处理规则

连接刷新

暂停任务

恢复任务

重启任务

本文介绍如何在 CKafka 连接器中查看事件详情和配置事件通知规则。

进入事件中心

1. 登录 CKafka 控制台,

2. 在连接器 > 任务列表页面找到要查看的任务,单击任务的"ID",进入基本信息页面。

3. 在页面上方选择**事件中心**页签,设置好时间范围(支持近7天、近**30**天和自定义时间范围),选择需要查看的事件 类型。单击操作列的**查看详情**,您可以在右侧栏查看事件详情。

配置事件告警规则

1. 在事件详情页面,单击右上角的配置告警跳转至事件规则配置页面。

2. 单击新建事件规则, 配置事件模式信息。

规则名称:填写事件规则名称,规则名称一旦创建无法修改。

事件模式:选择云服务预设事件。

云服务类型:选择数据接入平台 Dip。

事件类型:勾选您要配置告警规则的事件类型。

3. 单击**下一步**,填写事件目标信息。

触发方式:选择消息推送。

消息模板:根据事件类型,提供两种不同类型推送模板,您可以根据实际情况选择,对于告警事件,建议统一选择 告警通知模板。

用户通知:根据您的需要自定义选择通知方式和告警接收对象。



消息队列 CKafka 版

4. 单击完成,完成事件告警规则配置。