

流计算 Oceanus

快速入门





【版权声明】

©2013-2025 腾讯云版权所有

本文档(含所有文字、数据、图片等内容)完整的著作权归腾讯云计算(北京)有限责任公司单独所有,未经腾讯云事先明确书面许 可,任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分内容。前述行为构成对腾讯云著作权的侵犯,腾讯云 将依法采取措施追究法律责任。

【商标声明】

🔗 腾讯云

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。未经腾讯云及有关权利人书面许可,任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为,否 则将构成对腾讯云及有关权利人商标权的侵犯,腾讯云将依法采取措施追究法律责任。

【服务声明】

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

【联系我们】

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



文档目录

快速入门

从零开始上手 创建独享集群 创建 SQL 作业 创建 JAR 作业 创建 ETL 作业 创建 Python 作业 Flink 动态 CEP 快速入门



快速入门 从零开始上手

最近更新时间: 2023-08-21 14:08:31

流计算 Oceanus 是位于云端的流式数据汇聚、计算服务,提供全托管的云上服务,您无需关注基础设施的运维,便能便捷对接云上 数据源,获得完善的配套支持。

通过阅读本文,您将可以零基础上手流计算 Oceanus。

步骤1 登录控制台

登录 流计算 Oceanus 控制台。

步骤2 新建集群

选择控制台左侧菜单栏 计算资源,在页面左上角选择地域,然后单击新建。创建独享集群,详情可参见 创建独享集群。



创建成功独享集群后,单击**操作**中的**关联空间**,可以将集群关联到某一工作空间,关联后该工作空间可以使用这一集群的计算资源来 运行作业。

步骤3 新建作业

选择控制台左侧菜单栏工作空间。单击进入空间。

🔗 腾讯云

流计算 Oceanus	工作空门 ③ 广州	v								帮助文档 เ占
器 工作空间 ② 计算资源 □ 角色权限		新 は工作交向 K3 空向名称 K3 / 空向D	更多▼	k2 空間各称 k2 メ 空間D	更多 *	K1 空间2時 空间10	更 8 ▼	输入空间名称关键5 lake1 空间名称 lake1 /* 空间D	Q 更多 ▼	
		空時販送 - * 计算算 先気数 创建時間 2022-10-19 09:56:59 成员数量 42人 作业数量 0个 現入室間		오메麗王 - 计算页源 <u>未文</u> 算 信服时间 2022-10-19 09:56:54 역 전量 0 个 22 入2766		空周描述 - 計算調整 - 式実数 位達封词 2022-10-19 09-56-47 成点数量 42人 作空数量 0个 - 22人空間		空周期送 - * 計算規算 1 介募基 重音 磁振時間 2022-10-17 15-32-37 成点数量 42人 作业数量 1个 現入変層		
			更多 ▼	wedata1 空間名称 wedata1 / 空間の 空間の 空間の ごの間 ボ・・ 仕取買酒 1个集群 豊着 信部対対 2022-09-06 17:01:55 広友質量 43人	更多 ▼	空間を称 空間の 空間加 空間加 空間加 空間加 空間加 空間加 空間加 空間加	更多 ▼	田田 (加) (1) (1) (1) (1) (1) (1) (1) (1) (1) (1	更 \$ ▼	•
		作业较量 0个 进入空间		作业设量 0个 进入空间		作业数量 0个 进入空间		作业装量 0个 进入空间		

进入空间后,单击控制台左侧菜单栏 作业管理,选择新建作业所在集群的地域,然后单击新建,单击新建作业。

作业管理 ◎ 广州 🗸	Default v								*
+新建 ✓批量 口分类 〒 等送	直ぐ			and a star of a star way that the star of a star of the					
▼ 请输入"作业名称"关键词	Q	「高公告」作业	"现已支持"按日求分类" 可切罐	日定又日求对作业进行为	r央,方使多作业的管理 👱	御 洋開			
器 作业目录 (5)		作业概览							
- # S		SQL作业		JAR作业		ETL作业		Python彻业	
S	运行中 2 CU	^{通行中} 5 ↑	[#] [⊥] 156 ∧	道行中 0个	^{#⊭} 7 ↑	运行中 0 个	#⊭ 1 ↑	运行中 0 个	^{停止} 1 ↑
	运行中 2 CU								
> D		快速入门							

选择所需创建的作业,填写作业名称,选择运行集群,以上内容填写完成后,单击确定。



新建作业					>
作业类型	SQL 作业	JAR 作业	ETL 作业	Python 作业	
所属目录	作业目录				•
	在"按目录分类";	模式下,可为作业	选择分类目录		
作业名称	请输入名称				
	长度小于100,支	持中文/英文/数字	/"-"/"_"/"."		
计算资源	请选择集群			•	Φ
标签	暂无标签				
	标签用于从不同组	挂度对资源分类管 ³	理,如需了解更多	,请前往 <mark>标签产品文</mark>	〔档℃
		确定	取消		

后续操作

- 流计算 Oceanus 可创建四种类型的作业,详情可参见 创建 SQL 作业 、创建 JAR 作业 、创建 Python 作业 、创建 ETL 作业 。
- 作业创建成功后,可查看 作业类型 和 作业信息,以及支持的 作业操作。
- 作业创建完成后,即可进行作业开发,更多内容可参见作业开发。
- 如果遇到权限问题或需要对子账户进行权限控制,可以查看文档 权限管理。

创建独享集群

最近更新时间: 2024-11-15 16:24:32

前提条件

- 1. 已创建腾讯云账号,创建账号可参考 注册腾讯云。
- 2. 若使用子账号登录,需要获得用户账户的访问授权,详情请参见 CAM 访问管理 和 流计算服务委托授权。

操作步骤

登录 <mark>流计算 Oceanus 控制台</mark>,在计算资源中单击**新建**。集群信息与配置填写完成后,单击**立即购买**后将跳转至订单确认页,支付 完成后即成功创建集群。集群列表将出现新购的集群信息,集群状态为初始化中,初始化完成后集群状态将变为运行中,用户即可在 运行中的集群上创建流计算作业。

- 选择地域与可用区,购买的集群需和在流计算用到的其他资源(如消息队列或云数据库等)在同一地域和可用区。
- 设置集群名称和集群描述,便于识别集群用途。
- 选择所需的计算 CU 数,最低12个 CU 起购,单个集群每次最多可购买300个 CU,若有更多需要可通过 售后支持 申请。
- 选择 VPC 与子网。请注意此处需选择您在流计算中需要用到的其他资源如消息队列或云数据库所在的 VPC,流计算独享集群将 在该 VPC 子网下申请弹性网卡并占用一定的 IP 数量,所选的子网的可用 IP 数量需至少大于购买的计算 CU 数。
- 选择 COS 存储桶。流计算独享集群使用对象存储 COS 来保存作业的 checkpoint、jar 包等数据。如本地域无可用存储桶,可 在 对象存储控制台 新建。使用 COS 可能产生一定费用,详情可查看 对象存储 > 计费概述。
- 设置用户访问作业的 Flink UI 的密码,用户名默认为 "admin" 无需设置。
- 可为集群打上标签,标签用于从不同维度对资源分类管理,如需了解更多,可查看标签。
- 可勾选自动续费,避免集群到期后被隔离停服,影响服务。



计费模式•	包年包月 投量计费
lible* .	- 华康地区 - 华东地区 - 华东地区 - 华东地区
12131 -	广州 深圳 深圳金融 福州 杭州 济南 南京 上海 北京 石家庄 天津 长沙 武汉 成都 重庆
	港港台地区 - 亚太东南 - 亚太南部 - 亚太南北 - 美国西部 - 美国东部 - 北美地区 - 欧洲地区 -
	中国香港 中国台北 新加坡 曼谷 孟买 首尔 日本 法兰克福 硅谷 弗吉尼亚 多伦多 莫斯科
	不同地域的云产品之间内网不互通,您购买的集群需要与流计算使用到的其他云产品资源处于同一地域
□用区 *	1 mills
VPC •	vpc-15o3tdwn cluster-n2ettyps v subnet-4xf8jqkw cluster-n2ettyps v 🗘 共 16381 个子网IP, 剩余可用 16375 个
	流计算通过 VPC 和弹性网卡来访问同地域中的其他云产品资源,并需要占用一定的子网 IP 数量,请确保所选子网的可用 IP 数量充足
	如现有网络不符合您的要求,请前往 VPC 控制台 新建私有网络 或 新建子网
集群名称*	请输入集群名称
集群描述 •	清输入集群描述
	支持1-50个英文、汉字、数字、连接线-或下划线
计算 CU 数 ▪	- 12 +
	CU 是流计算的最小计算资源和计费单位。1CU 包含1个 CPU 和 4GB 内存。 当CU数大于等于 48 时,减免管理节点费用。
管理 CU 数	2 CU
	e 管理CU数是指流计算中单个集群额外使用的管理节点资源,固定消耗 2 CU
对家仔宙 605▼	这样好话间 ▼
	流订算使用对象存储 CUS 米保存作业的 cneckpoint、jar 包、或投速日志等,则本地域尤可用存储确,请用任 对象存储控制台 忆 前建
Flink UI 用户名	admin
Flink UI 密码 *	请输入 Flink UI 密码 💋
	用户名与密码将用于登录查看作业的 Flink UI 界面
TALL PROFILE IN CONTRACT	300 (ma) (m. 1) (100)
佣认 Flink UI 密码 *	语书次输入 Flink UI 密約 92
标签	暂无标签 /
	标签用于从不同维度对资源分类管理,如需了解更多,请前往 标签产品文档 忆
购买时长。	
则天时长。	1个月 2 3 6 1年 2年 3年 其它时长
自动续费	账户余额足够时,设备到期后按月自动续费
费用	
	白那田 _ 计描述资源用 → 部 约资源第田 → 特别印 №
	(A) AD = 1) 予以示以言す E III (A) AD 1 = II (A) AD 1 = III (A) AD 1 = II (A) AD 1 = III (A) AD 1 = II (A) AD 1 = II (
	□ 同意 简讯云 流计算 Oceanus 服务等级协议 亿 和 退费规则 亿

▲ 注意

- •为了最大化利用底层机器资源,推荐购买6的倍数CU。
- •出于底层机器性能考虑,单个集群建议最大值为800个CU。

内核版本

- 新建集群目前仅支持提交 Flink-1.13, Flink-1.14, Flink-1.16 版本任务,可以 提交工单 申请支持 Flink-1.11 版本任务。
- 存量集群当前只能提交已有的 Flink 版本任务,可以 提交工单 申请升级。
- Flink-1.13, Flink-1.14, Flink-1.16版本任务默认采用 Application Mode 部署。

创建 SQL 作业

最近更新时间: 2024-08-23 14:38:32

前提条件

流计算作业 SQL 作业需运行于流计算独享集群,若还没有集群,请参考 创建独享集群。

步骤1: 创建 SQL 作业

- 1. 登录 流计算 Oceanus 控制台,进入某一工作空间后,单击左侧导航作业管理。
- 进入作业管理页面,单击新建 > 新建作业,作业类型选中 SQL 作业,输入作业名称,并选择一个运行中的集群,新建的 SQL 作 业将运行于此集群。
- 3. 单击确定后,即成功创建作业。

新建作业					
作业类型	SQL 作业	JAR 作业	ETL 作业	Python 作业	
所属目录	作业目录				•
	在"按目录分类";	模式下,可为作业	/选择分类目录		
作业名称	请输入名称				
	长度小于100,支	持中文/英文/数字	-/"-"/"_"/"."		
计算资源	请选择集群				φ
标签	暂无标签				
	标签用于从不同维	主度对资源分类管	理,如需了解更多	8,请前往 <mark>标签产品</mark> 文	て档 🖸
		确定	取消		

步骤2:流计算服务委托授权

选择作业管理中刚新建的作业,单击**开发调试**。在未授权时,弹出访问授权对话框如下,单击**前往授权**,授权流计算作业访问您的消 息队列或云数据库等资源。此授权的详细说明参见 流计算服务委托授权 。

流计算角色授权	×
"流计 算服务"需要访问CKafka、COS等资源,系约 予流计算服务。	充将创建CAM角色,统一整合相关资源权限授
说明:未来流计算支持的更多资源将添加到此CAM	M角色进行自动授权。
前往授权	取消



步骤3:编写 SQL 语句

授权完成后,可在开发调试的代码编辑框中输入 SQL 语句,可无需另外准备数据快速创建作业。示例语句具体执行的内容如下:

- 1. 使用内置 Connector "datagen"创建数据源表 Data_Input,包含字段 age(BIGINT型)、score(BIGINT型)。 datagen 是位于集群本地的数据源,可以不断生成随机数据。
- 使用内置 Connector "blackhole"创建数据结果表 Data_Output,包含字段 avg_age(BIGINT型)、avg_score (BIGINT型)。blackhole 是位于集群本地的数据目的,可以不断接收数据。
- 将 Data_Input 中的 age 和 score 取平均数之后存储于 Data_Output。
 若要使用其他数据源或数据目的的 Connector,如 Ckafka 或 Elasticsearch,可参考 上下游开发指南,并自行准备数据。

```
CREATE TABLE 'Data_Input' ( --步骤 1 : 创建数据源表 (Source) Data_Input
age BIGINT,
score BIGINT
) WITH (
 'connector' = 'datagen',
 'rows-per-second'='100', -- 每秒产生的数据条数
 'fields.age.kind'='random', -- 无界的随机数
 'fields.age.min'='1', -- 随机数的最小值
 'fields.score.kind'='random', -- 无界的随机数
 'fields.score.kind'='random', -- 死界的随机数
 'fields.score.kind'='random', -- 死界的随机数
 'fields.score.min'='1', -- 随机数的最大值
 'fields.score.min'='1', -- 随机数的最大值
 'fields.score.max'='1000' -- 随机数的最大值
);
CREATE TABLE 'Data_Output' ( --步骤 2 : 创建数据结果表 (Sink ) Data_Output
 'avg_score' BIGINT,
 'avg_score' BIGINT
) WITH (
 'connector' = 'blackhole'
);
INSERT INTO 'Data_Output' --步骤 3 : 将数据源表 (Source) Data_Input 中的 age 和
score 取平均数之后存储于数据结果表 (Sink ) Data_Output
 SELECT AVG(age), AVG(score) FROM 'Data_Input';
```

步骤4:设置作业参数

在作业参数中设置 Checkpoint 和算子默认并行度等参数的值,使用其他上下游数据则需选择相应的内置 Connector。

步骤5:发布运行 SQL 作业



单击**发布草稿**,将进行作业运行检查,检查通过后将进入发布确认。发布将生成新的作业版本,版本号由系统自动生成。

发布新版本	本 V19	×
发布者	angener mer er ser me	
版本说明*	不超过50个字符	
	确认发布	

发布草稿后,单击版本管理,可以查看并切换当前作业的不同版本。

÷	运行中 (V3)	批量操作					
作业概览	开发调试 监控 事件 日志 快照 调优	版本	版本描述	发布者	发布时间	最近启动时间	操作
	版本管理(M3) 作业学数 表交量智快 保存 进行成本 V3 停止	〇 草稿	作业草稿		-	-	
	■ < 用約 → Ⅱ位 + 脳入現版 ● 复制到草稿 ● 依式化 Q 查找 ✓ 语法检查 ● 谓述 表	〇 • V3 (线上	.) 3		2023-02-13 19:18:16	2023-02-13 20:19:45	
	引 用 ① 只读模式,已发布的版本不可编辑,可"复制到草稿"再编辑	○ V2	2	主账户	2023-02-13 19:17:37		删除
	ee 支 1 Datagen Connector 可以随机生成一些数据用于测试 量 2 参見 https://cloud.tencent.com/document/product/849/58713	O V1		主账户	2023-02-13 19:17:37		#18
	<pre>3 CREATE VALUE VAL</pre>						

切换到期望运行的作业版本后,单击**运行版本**,再单击**确认**即可启动作业。

运行版本 V18			×
 不使用任何快照,直接运行 使用历史快照运行 			
选择快照			
	确定	取消	

步骤6:查看作业运行情况

作业发布并启动运行后,将变为操作中的状态,成功启动后将变为运行中的状态。作业运行中时,可以通过监控、日志、Flink UI 等 功能查看作业运行的情况。



创建 JAR 作业

最近更新时间: 2024-12-19 17:29:11

前提条件

流计算作业 JAR 作业需运行于流计算独享集群,若还没有集群,请参考 创建独享集群。

步骤1: JAR 包开发

创建 JAR 作业需要上传已开发好的 JAR 包,您可以单击 GitHub 链接,下载我们提供的 WordCount 示例 JAR 包(文件名为 flink-hello-world-1.0.0.jar)来完成。

步骤2: 上传 JAR 包

登录 流计算 Oceanus 控制台,选择依赖管理,单击新建>新建依赖。在新建程序包中选择地域,该地域需与独享集群所在地域一 致,默认使用本地上传的方式。单击选择依赖将打开本地文件选择窗口,选择步骤1中下载的 JAR 包并上传。填写依赖描述(如程序 包名称、程序包描述、版本说明)后,单击确定完成新建程序包。完成新建后程序包列表将出现刚新建的程序包。

新建依赖			×
地域			
依赖类型 🛈	提醒: 流计算只能引用相同地域的依赖 ○ JAR 程序包 ○ Python 程序包 ○ 配置文件 ○ 数据文件		
所属目录	依赖目录		
上传依赖	● 本地上传 ── 使用 COS 资源		
		选择依赖	
	请上传 JAR 文件,大小不超过 400M,文件过大建议使用COS方式上传。		
依赖描述	请输入依赖的内容或用途,不超过50个字符		
	确定取消		

步骤3: 创建 JAR 作业

登录 流计算 Oceanus 控制台,进入某一工作空间后,单击左侧导航**作业管理**,进入作业管理页面,单击新建 > 新建作业,作业类 型选中 JAR 作业,输入作业名称,并选择一个运行中的集群,新建的 JAR 作业将运行于此集群,单击确定后即成功创建作业。



建作业					
业类型	SQL 作业	JAR 作业	ETL 作业	Python 作业	
属目录	作业目录				•
	在"按目录分类" 植	莫式下,可为作业	选择分类目录		
业名称	请输入名称				
	 长度小于100,支	持中文/英文/数字	/"-"/"_"/"."		
算资源	请选择集群				- ¢
签	暂无标签				
	标签用于从不同维	挂度对资源分类管 ³	理,如需了解更多	,请前往 <mark>标签产品</mark> :	文档 🖸
		确定	取消		

步骤4:流计算服务委托授权

选择**作业管理**中刚新建的作业,单击**开发调试**。在未授权时,弹出访问授权对话框如下,单击**前往授权**,授权流计算作业访问您的 CKafka、TencentDB 等资源。此授权的详细说明参见 服务委托授权 。



步骤5: 配置 JAR 作业

在**开发调试**中,在**主程序包**的下拉框中选择刚才新建的程序包名称,在**主类**输入

com.tencent.cloud.oceanus.wordcount.WordCount 。您可以在作业参数设置中调整作业的算子默认并行度、添加外部依



赖。		
← jar 未初始化		
作业概览 开发调试	式 监控	事件 日志 快照
	版本管理 <mark>(草</mark> 和	高) 作业参数 保存 发布草稿
	主程序包 *	flink-connector-redis-1.14.0.jar v1 💌
	主类 *	com.tencent.cloud.oceanus.wordcount.WordCount
	主类入参	请输入主类参数

步骤6:发布运行 JAR 作业

1. 单击**发布草稿**,将进行作业运行检查,检查通过后将进入发布确认。发布将生成新的作业版本,版本号由系统自动生成。

发布新版本	SV19	×
发布者		
版本说明*	不超过50个字符	
	确认发布	



2. 发布草稿后,单击版本管理,可以查看并切换当前作业的不同版本。

← 作业概覚	停止▲ 开发调试 盜控 事件 日志 快照		批量操作	版本描述	发布者	发布时间	最近启动时间	操作
3	1846年1872(1948) 1752(1952 1947 1947 1947 1947 1947 1947 1947 1947	* v1 *	 ● 単稿 ○ V1 (线上) 	作业草稿 1		- 2023-01-05 21:16:50	- 2023-01-09 17:52:54	•
	主京・ com.tencent.clos 主気入争 (mox)ココールの		v3v2	v3_128 3		2023-01-05 21:40:47 2023-01-05 21:30:33	2023-01-05 21:40:58 2023-01-06 14:39:04	删除

3. 切换到期望运行的作业版本后,单击运行版本,再单击确认即可启动作业。

○ 不使用任何快照。直接运行			
● 使用历史快照运行			
选择快照		▼	

步骤7: 查看作业运行情况

作业发布并启动运行后,将变为操作中的状态,成功启动后将变为运行中的状态。作业运行中时,可以通过监控、日志、Flink UI 等功能查看作业运行的情况。



创建 ETL 作业

最近更新时间: 2025-02-14 09:39:42

▲ 注意:

流计算 Oceanus 不再支持 ETL 的作业类型,用户存量作业可以继续运行,推荐使用 SQL 作业类型 或者 WeData 实时 集成 功能来完成 ETL 业务。

前提条件

流计算作业 ETL 作业需运行于流计算独享集群,若还没有集群,请参考 创建独享集群。

步骤1: 创建 ETL 作业

登录 流计算 Oceanus 控制台,进入某一工作空间后,单击左侧导航**作业管理**,进入作业管理页面,单击新建 > 新建作业,作业类 型选中 ETL 作业,输入作业名称,并选择一个运行中的集群,新建的 ETL 作业将运行于此集群,单击确定后即成功创建作业。

<u>1</u>	SQL 作业	JAR 作业	ETL 作业	Python 作业				
ł	作业目录							
	在 "按目录分类" 札	莫式下,可为作业	选择分类目录					
R	请输入名称							
	长度小于100,支持中文/英文/数字/"-"/"_"/"."							
Im/	请选择集群 🔹 🗘							
	暂无标签							
	标签用于从不同维	E度对资源分类管理	浬,如需了解更多	8,请前往 <mark>标签产</mark> 8	品文档 🖸			

△ 注意

注意: 当前ETL作业暂时只支持1.13版Flink,尚不支持1.13版本内核的集群需要先提 工单 升级后再使用。

步骤2:流计算服务委托授权

选择**作业管理**中刚新建的作业,单击**开发调试**。在未授权时,弹出访问授权对话框如下,单击**前往授权**,授权流计算作业访问您的消 息队列或云数据库等资源。此授权的详细说明参见 流计算服务委托授权 。





步骤3: 配置数据源表

授权完成后,单击作业进入**作业开发 > 草稿**,从左侧列表拖入 MySQL 数据源。

← myma	← 停止							
作业开发	草稿	Ŀ	作业参数	检查	保存		运行	
数据源								
MySQL								
Elasticsearch								
数据目的								
MySQL								
PostgreSQL							MuSQL	
ClickHouse							 ■ ■	
Elasticsearch								
Doris								
& Kafka								



单击 MySQL 数据源可以进行数据源的配置。

数据源		v
1 连接参数	> 2 字段选择	
数据类型	Mysql	1
数据名称 🔸	MySQL	
数据库实例 *	请选择 ▼ 管理连接信息	
数据库时区★	Asia/Shanghai	
Server-id	5010 — 5020	
过滤操作 🛈	插入 更新 删除	
其他参数	Ŧ	=
下一步		

如果当前没有可用的数据库实例,可以单击管理连接信息,并单击新建连接信息增加新的数据源。

存储类型	MvSQL				
	,				•
连接方式	腾讯云实例	IP连接			
连接名称 *	请输入连接名称				
套接实例 ★	请选择			▼	Ø
用户名(选填)	示例: root				
密码(选填)	请输入密码				Ø
连通性测试	¢				

腾讯云实例可以使用当前账号下的腾讯云产品实例,但注意账号需要拥有对应产品的查询列表权限。

IP连接则直接用填入的 IP 进行连接,需要确保 IP 与 Oceanus 集群处于同一个 VPC 或是有进行过相关的网络打通操作。 选择好数据源之后,可以通过下拉框选择需要进行同步的数据库与表。



数据源		
1 连接参数	▶ (2) 字段选择	
数据名称*	MySQL	
连接实例 *	dennis-56test ▼ 管理连接信息	
库表选择*	● 单表同步	
	请选择 ▼	
		C
数据库时区 *	Asia/Shanghai	εp
Server-id (j)	5010 — 5020	=
下一步		

单击下一步,配置需要进行同步的字段,并单击确认完成配置。

数据源	
✓ 连接参数 〉 2 字段选择	
✓ 字段名称	字段类型
✓ f_sequence	INT
✓ f_random	INT
✓ f_random_str	VARCHAR
✓ Name	VARCHAR
上一步 确定	



再次返回画布,可以看到数据源已经配置完成。

← 停止							
作业开发 草稿	线上作业参数	检查保存停止	运行				
数据源							
MySQL							
Elasticsearch							
数据目的							
MySQL							
PostgreSQL			MySQL				
ClickHouse			配置完成 ⊘				
Elasticsearch							
Doris							
& Kafka							

步骤4: 配置数据目的

从左侧拖入数据目的,并与数据源相连。





单击数据目的对象,配置好数据库实例与库表选择:



数据目的		
1 连接参数	 > 2 字段映射 	
数据类型	Mysql	
数据名称◆	MySQL	
数据库实例 *	MySQL007 管理连接信息	
库表选择◆	● 单表同步	
	test / my_test2 v	
数据库时区★	Asia/Shanghai	
保存间隔 🛈 •	- 100 +	
下一步		

单击下一步进行字段映射配置:

数据目的				Ŧ
文 连接参数 > 2	字段映射			
字段名称	字段类型	映射字段名称	映射字段类型	是否为主键
f_sequence	INT	f_sequence	✓ INT	
f_random	INT	f_random		
f_random_str	VARCHAR	f_random_str	▼ VARCHAR	
Name	VARCHAR	Name	✓ VARCHAR	
共 4 条			10 ▼ 条/页	4 1 /1页 ► H E
上一步 确定				

完成后,单击确认完成数据目的的配置:

🔗 腾讯云

数据源				
MySQL		4	•	+
数据目的				
MySQL	NySQL NySQL 配置完成			
PostgreSQL				
ClickHouse				
ElasticSearch				

步骤5:设置作业参数

在作业参数中可以设置作业相关的参数,详情可以查看文档作业高级参数。

作业参数			
link 配置 heckpoint	时间间隔 — 600	+ 秒 (30 到 3600秒)	
业失败恢复 🛈			
5级参数 访	1 pipeline.max-parallel 2	ism: 2048	
志記署			
志配置 5日志采集 ①	采集到对象存储 COS(默认)	▼ 设置默认选项	
志配置 行日志采集 ①	采集到对象存储 COS(默认) autotest-cos-1257058945	● 设置默认选项	
志配置 行日志采集 ①	采集到对象存储 COS (默认) autotest-cos-1257058945 日志采集级别: INFO	 ✓ 设置默认选项 采集高级配置 	



步骤6:发布运行 ETL 作业

单击运行,会进行作业预检查,并启动作业。

← 日志 未初始化							作业操作	•	Flink UI		云监控	
作业开发 草稿	作业参数	检查保存	U									
数据源					4	+	_	67	-	A		
MySQL						т						
数据目的												
MySQL		MySQL	MySQL									
PostgreSQL	预检查											•
	检查对象		检查结果						ł	金童状态		
ClickHouse	MySQL		检查结果:成功	I				Þ		\odot		
E	MySQL		检查结果:成功	I				Þ		\odot		
ElasticSearch												

步骤7:查看作业日志

单击日志按钮可以查看作业日志。

← 日志 汕	运行中	作业操作 ▼ Fli
作业开发 草	稿 线上 作业参数 监控 日志 停止	
		+ -
日志		
运行日志 👻	实例 1 ▼ cql-qj2u25fp-55162-taskmanager-1-1 [TaskManager] ▼	3
近 1 小时	近 24 小时 近 7 天 2022-04-20 10:09:26 ~ 2022-04-21 10:09:26 🖶 请输入关键词,可以使用 AND, OR 搜索 快捷诊断	诊断指南 🖸
cql-qj2u25fp-55	5162-78d4b88dd9-sblks [JobManager] 启动时间: 2022-04-20 20:30:08,结束时间: -	
pos=-9223372036 2022-04-20 20:3	5550020, grius-fucbuist-csof-fied-bruf-525400540154:1-455500004, fow-o, event-off, endofiset-bintogoffset-cris_set-c 6854775808, row=0, event=0}}}] 36:28.825 [Source Data Fetcher for Source: TableSourceScan(table=[[default catalog, default database, mysgl cdc source 165	50456611615]],
<pre>fields=[f_seque split fetcher 1</pre>	ence, f_random, f_random_str, Name]) (1/1)#13] INFO org.apache.flink.connector.base.source.reader.fetcher.SplitFetcher - 1	Starting
2022-04-20 20:3 fields=[f_seque	- 3 6:28.827 [Source Data Fetcher for Source: TableSourceScan(table=[[default_catalog, default_database, mysql_cdc_source_165 ence, f_random, f_random_str, Name]) (1/1)#13] INFO com.ververica.cdc.connectors.mysql.source.reader.MySqlSplitReader - (50456611615]], Create binlog
2022-04-20 20:3 fields=[f_seque MBean 'debezium	36:28.832 [Source Data Fetcher for Source: TableSourceScan(table=[[default_catalog, default_database, mysql_cdc_source_165 ence, f_random, f_random_str, Name]) (1/1)#13] WARN io.debezium.relational.history.DatabaseHistoryMetrics - Unable to reg m.mysql:type=connector-metrics,context=schema-history,server=mysql_binlog_source': debezium.mysql:type=connector-metrics,	50456611615]], gister the
Context=schema- 2022-04-20 20:3 fields=[f_seque	-history,server=mysql_binlog_source 36:28.833 [Source Data Fetcher for Source: TableSourceScan(table=[[default_catalog, default_database, mysql_cdc_source_165 ence, f_random, f_random_str, Name]) (1/1)#13] INFO com.ververica.cdc.connectors.mysql.debezium.task.context.StatefulTask GTLD set 7bc5b13f_c367_11ea_bfd7_55240054ef54:1_4555568677 does contain the GTLD set 7bc5b13f_c367_11ea_bfd7_53540054ef54:1	50456611615]], kContext – 1–455568664
required by the 2022-04-20 20:3	e connector. 36:28.834 [Source Data Fetcher for Source: TableSourceScan(table=[[default_catalog, default_database, mysql_cdc_source_165	50456611615]],
Te cus=[1_seque		CUILEXL -



创建 Python 作业

最近更新时间: 2023-08-30 16:58:11

前提条件

流计算作业 Python 作业需运行于流计算独享集群,若还没有集群,请参考 创建独享集群 。目前流计算 Oceanus 支持运行基于开 源 Flink V1.13 开发的 Python 作业,且预装了 Python 3.7 版本的环境。

步骤1: Python 作业开发

创建 Python 作业需要上传已开发好的 Python 文件。

步骤2: 上传 Python 文件

- 1. 登录 流计算 Oceanus 控制台,进入某一工作空间后,选择依赖管理,单击新建 > 新建依赖。
- 2. 在新建程序包中选择地域,该地域需与独享集群所在地域一致,默认使用本地上传的方式。
- 3. 单击选择依赖将打开本地文件选择窗口,选择步骤1中下载的 Python 文件并上传。
- 填写依赖描述,如程序包名称、程序包描述、版本说明后,单击确定完成新建程序包。完成新建后程序包列表将出现刚新建的程序包。

域	◎ 成都 ▼		
主義类型 🛈	JAR 程序包 O Python 程序包 配置文件 数据文件		
属目录	依赖目录 🔻		
传依赖	● 本地上传 ── 使用 COS 资源		
		选择依赖	
	请上传py或zlp类型的文件,大小不超过 50M,文件过大建议使用COS方式上传。		
、赖描述	请输入依赖的内容或用途,不超过50个字符		

步骤3: 创建 Python 作业

- 1. 登录 流计算 Oceanus 控制台,进入某一工作空间后,单击左侧导航作业管理。
- 2. 进入作业管理页面,单击新建 > 新建作业。
- 3. 作业类型选中 Python 作业,输入作业名称,并选择一个运行中的集群,新建的 Python 作业将运行于此集群。
- 4. 单击确定后即成功创建作业。



英型 Sultru JAN TF型 ETLTF型 Fython TF型 目录 作业目录 在 "按目录分类" 模式下,可为作业选择分类目录 名称 请输入名称	
 禄 作业目录 在 "按目录分类" 模式下,可为作业选择分类目录 请输入名称 	
在 "按目录分类" 模式下,可为作业选择分类目录 际 请输入名称	
请输入名称	
请选择集群	- ¢
新开栏袋 ♥	

步骤4:流计算服务委托授权

- 1. 选择作业管理中刚新建的作业,单击开发调试。
- 2. 在未授权时,弹出访问授权对话框如下,单击**前往授权**,授权流计算作业访问您的 CKafka、TencentDB 等资源。此授权的详 细说明参见 流计算服务委托授权 。

流计算角色授权	×
"流计算服务"需要访问CKafka、COS等资源,系统料 予流计算服务。	将创建CAM角色,统一整合相关资源权限授
说明:未来流计算支持的更多资源将添加到此CAM的	角色进行自动授权。
前往授权	取消

步骤5: 配置 Python 作业

在**开发调试**中,在**主程序包**的下拉框中选择刚才新建的程序包名称,在入口类输入对应的参数,如果入口类中有参数可以在入口参数 中加上,目前 Oceanus 中内置了 Python 3.7 版本的环境,可以在 Python 环境中选择。您可以在作业参数设置中调整作业的算



子默认并行度、添加外部依赖。

版本管理 (草稿)	作业参数 保存 发布草稿
主程序包·	.py * v1 *
Python环境•	Python-3.7 ×
	+ 添加 Python 程序包
入口类	ml.lightgbm.pyflink_inference.kafka2mysql
入口学致	lp 172.xxxport 3306database xxxsource_table sourcesink_table sinkusername rootpassword xxx
数据文件 🛈	+ 添加政務文件

步骤6:发布运行 Python 作业

1. 单击**发布草稿**,将进行作业运行检查,检查通过后将进入发布确认。发布将生成新的作业版本,版本号由系统自动生成。

发布新版本 V3		\times
发布者		
版本说明 · python-job-test		
	确认发布	



2. 发布草稿后,单击版本管理,可以查看并切换当前作业的不同版本。

版本信息	
发布者	
版本说明 2	
配置信息	
主程序包	.py (V1)
λ□类 ·	
入口參数 •	
Python 环境 P	ytton-3.7
rython (EUS) ES	
作业争数 SI用 JAR 程序包	
内置 Connector	
進行日志采集 🛈	日志思: 5,2_merge 日志主题: Oceanus, Following-VS2
高级参数 🛈	pipeline.max-parallelism: 2048
作业失败恢复 🕜	开启
JobManager 规格	1 GU
TaskManager 规格	1 GU
算子默认并行度 🛈	1

3. 切换到期望运行的作业版本后,单击运行版本,再单击确认即可启动作业。

运行版本线上 V2			×
 不使用任何快照,直接运行 使用历史快照运行 	确定	取消	

步骤7: 查看作业运行情况

作业发布并启动运行后,将变为操作中的状态,成功启动后将变为运行中的状态。作业运行中时,可以通过监控、日志、Flink UI 等功能查看作业运行的情况。·



Flink 动态 CEP 快速入门

最近更新时间: 2024-02-27 14:33:21

前提条件

Flink 动态 CEP 作业需运行于流计算独享集群,若还没有集群,请参考 创建独享集群。

操作流程

步骤1: 准备测试数据

准备上游 Kafka Topic

- 1. 登录 消息队列 CKafka 控制台,新建或选择可用的 Ckafka 实例。
- 2. 创建一个名称为 topic_cep_demo 的 Topic ,存放模拟的用户行为日志。

准备 MySQL 数据库

- 1. 登录 云数据库 MySQL 控制台,新建或选择可用的 MySQL 实例。
- 2. 创建 t_mysql_demo 规则表,用来记录 Flink CEP 作业中需要应用的规则。



步骤2:开发并启动 Flink CEP 作业

- 1. 联系在线客服获取 flink-cep 的 jar 包,并把 jar 包添加到您的 Maven 项目的依赖库中。
- 2. 开发作业代码。
- 3. 在流计算 Oceanus 控制台上,上传作业 JAR 包,部署并启动 JAR 作业。

主类: com.tencent.cloud.oceanus.cep.demo.FlinkCepDemo,主类入参如下。

```
--bootstrap-server 172.28.22.5:9092 --topic topic_cep_demo --properties.group.id
flink_cep_demo --url jdbc:mysql://172.28.28.24:3306/d_cep_demo?
user=root&password=waze6011601 --table-name t_mysql_demo --pattern-update-interval-ms
3000
```

步骤3: 插入规则

1. 插入动态更新规则到 MySQL 表中。



INSERT INTO t_mysql_demo (
`id`,	
`version`,	
`pattern`,	
`function`	
'{"name":"end","quantifier":{"consumingStrategy":"SKIP_TILL_NEXT","properties":	
["SINGLE"],"times":null,"untilCondition":null},"condition":null,"nodes":	
[{"name":"end","quantifier":{"consumingStrategy":"SKIP_TILL_NEXT","properties":	
["SINGLE"],"times":null,"untilCondition":null},"condition":	
{"className":"com.tencent.cloud.oceanus.cep.demo.condition.EndCondition","type":"	
LASS"},"type":"ATOMIC"},{"name":"start","quantifier":	
{"consumingStrategy":"SKIP_TILL_NEXT","properties":["LOOPING"],"times":	
{"from":3,"to":3,"windowTime":null},"untilCondition":null},"condition":	
{"expression":"action == 0","type":"AVIATOR"},"type":"ATOMIC"}],"edges":	
[{"source":"start","target":"end","type":"SKIP_TILL_NEXT"}],"window":null,"afterMa tchStrategy":	
{"type":"SKIP_PAST_LAST_EVENT","patternName":null},"type":"COMPOSITE","version":1	
<pre>'com.tencent.cloud.oceanus.cep.demo.dynamic.DemoPatternProcessFunction')</pre>	

为了方便使用并提高数据库中的 Pattern 字段的可读性,流计算 Oceanus 定义了一套 JSON 格式的规则描述。上述 SQL 语 句中的 pattern 字段的值就是按照 JSON 格式的规则,给出的序列化后的 pattern 字符串。它的物理意义是去匹配这样的模 式: 连续3条 action 为 0 的事件发生后,下一条事件的 action 仍非 1。

- 2. 往 Kafka Topic topic_cep_demo 中发送消息。
 - 1,ZhangSan,0,1,1697181992000
 - 1, ZhangSan, 0, 1, 1697181993000
 - 1, ZhangSan, 0, 1, 1697181994000
 - 1, ZhangSan, 0, 1, 169718199500

查看 JobManager 日志,搜索 JDBCPeriodicPatternProcessorDiscoverer,查看最新的规则。

"{\"name\":\"end\",\"quantifier\":
{\"consumingStrategy\":\"SKIP_TILL_NEXT\",\"properties\":
[\"SINGLE\"],\"times\":null,\"untilCondition\":null},\"condition\":null,\"nodes\":
[{\"name\":\"end\",\"quantifier\":
{\"consumingStrategy\":\"SKIP_TILL_NEXT\",\"properties\":
[\"SINGLE\"],\"times\":null,\"untilCondition\":null},\"condition\":
{\"className\":\"com.tencent.cloud.oceanus.cep.demo.condition.EndCondition\",\"typ
e\":\"CLASS\"},\"type\":\"ATOMIC\"},{\"name\":\"start\",\"quantifier\":
{\"consumingStrategy\":\"SKIP_TILL_NEXT\",\"properties\":[\"LOOPING\"],\"times\":
{\"from\":3,\"to\":3,\"windowTime\":null},\"untilCondition\":null},\"condition\":



{\"expression\":\"action ==
0\",\"type\":\"AVIATOR\"},\"type\":\"ATOMIC\"}],\"edges\":
[{\"source\":\"start\",\"target\":\"end\",\"type\":\"SKIP_TILL_NEXT\"}],\"window\"
:null,\"afterMatchStrategy\":
{\"type\":\"SKIP_PAST_LAST_EVENT\",\"patternName\":null},\"type\":\"COMPOSITE\",\"
version\":1}"

查看 TaskManager 中日志,搜索 A match for Pattern of,查看 CEP 匹配的日志打印。

A match for Pattern of (id, version): (1, 1) is found. The event sequence: start: [Event(1, ZhangSan, 0, 1, 1697181992000), Event(1, ZhangSan, 0, 1, 1697181993000), Event(1, ZhangSan, 0, 1, 1697181994000)]end: [Event(1, ZhangSan, 0, 1, 1697181995000)]

步骤4:更新匹配规则,并查看更新的规则是否生效。

- 1. 更新 MySQL 表中的规则。
 - 1.1 将 StartCondition 中的 action == 0 修改为 action == 0 || action == 2,并且我们将重复出现的次数从 >=3 改为 >=5,对应 SQL 语句如下。

'{"name":"end","quantifier":{"consumingStrategy":"SKIP_TILL_NEXT","properties": ["SINGLE"],"times":null,"untilCondition":null},"condition":null,"nodes": [{"name":"end","quantifier":{"consumingStrategy":"SKIP_TILL_NEXT","properties": ["SINGLE"],"times":null,"untilCondition":null},"condition": {"className":"com.tencent.cloud.oceanus.cep.demo.condition.EndCondition","type" :"CLASS"},"type":"ATOMIC"},{"name":"start","quantifier": {"consumingStrategy":"SKIP_TILL_NEXT","properties":["LOOPING"],"times": {"from":5,"to":5,"windowTime":null},"untilCondition":null},"condition": {"expression":"action == 0 || action == 2","type":"AVIATOR"},"type":"ATOMIC"}],"edges": [{"source":"start","target":"end","type":"SKIP_TILL_NEXT"}],"window":null,"afte rMatchStrategy": {"type":"SKIP_PAST_LAST_EVENT","patternName":null},"type":"COMPOSITE","version" :1}','com.tencent.cloud.oceanus.cep.demo.dynamic.DemoPatternProcessFunction');

1.2 再插入一条记录的 id 为 2 新规则。它和规则 1 的版本 1 一样,其 StartCondition 仍为 action == 0 且重复出现的次数为 >=3。



{"expression":"action == 0","type":"AVIATOR"},"type":"ATOMIC"}],"edges":
[{"source":"start","target":"end","type":"SKIP_TILL_NEXT"}],"window":null,"afte
rMatchStrategy":
{"type":"SKIP_PAST_LAST_EVENT","patternName":null},"type":"COMPOSITE","version"
:1}','com.tencent.cloud.oceanus.cep.demo.dynamic.DemoPatternProcessFunction');

2. 往 Kafak Topic 发送数据。

- 1, ZhangSan, 0, 1, 1697181992000
- 1,ZhangSan,0,1,1697181993000
- 1, ZhangSan, 0, 1, 1697181994000
- 1,ZhangSan,2,1,1697181995000
- 1,ZhangSan,0,1,1697181996000
- 1, ZhangSan, 0, 1, 1697181997000
- 1, ZhangSan, 0, 1, 1697181998000
- l,ZhangSan,2,1,1697181999000
- 3. 查看 TaskManager 中日志,搜索 A match for Pattern of,查看 CEP 匹配的日志打印。

ID 为1的规则匹配日志:

A match for Pattern of (id, version): (1, 2) is found. The event sequence: start: [Event(1, ZhangSan, 0, 1, 1697181992000), Event(1, ZhangSan, 0, 1, 1697181993000), Event(1, ZhangSan, 0, 1, 1697181994000), Event(1, ZhangSan, 2, 1, 1697181995000), Event(1, ZhangSan, 0, 1, 1697181996000)]end: [Event(1, ZhangSan, 0, 1, 1697181997000)]

ID 为2的规则匹配日志:

```
A match for Pattern of (id, version): (2, 1) is found. The event sequence: start:

[Event(1, ZhangSan, 0, 1, 1697181992000), Event(1, ZhangSan, 0, 1, 1697181993000),

Event(1, ZhangSan, 0, 1, 1697181994000)]end: [Event(1, ZhangSan, 2, 1,

1697181995000)]

A match for Pattern of (id, version): (2, 1) is found. The event sequence: start:

[Event(1, ZhangSan, 0, 1, 1697181996000), Event(1, ZhangSan, 0, 1, 1697181997000),

Event(1, ZhangSan, 0, 1, 1697181998000)]end: [Event(1, ZhangSan, 2, 1,

1697181999000)]
```