流计算 Oceanus

操作指南

产品文档

腾讯云

文档目录

操作指南
准备上下游数据
上下游数据一览

最近更新时间：2020-06-01 16:13:07

SQL 作业的上下游数据介绍

- 数据源（Source）指的是输入流计算系统的上游数据来源。在当前的流计算 Oceanus SQL 模式的作业中，数据源可以是 CKafka、云数据库 MySQL 等。
- 数据目的（Sink）指的是流计算系统输出处理结果的目的地。在当前的流计算 Oceanus SQL 模式的作业中，数据目的可以是消息队列 CKafka、云数据库 MySQL、云数据库 PostgreSQL、Elasticsearch Service（即将支持）。

对于本文提到的各项概念，例如 Tuple 数据目的和 Upsert 数据目的的区别，请参见 词汇表。

产品名	可用作数据源	可作为 Tuple 数据目的	可作为 Upsert 数据目的
消息队列 CKafka	支持	支持	不支持
云数据库 MySQL	部分支持	支持	支持
云数据库 PostgreSQL	不支持	支持	支持（需9.5及以上版本）
Elasticsearch Service	不支持	即将支持	即将支持
云数据仓库	不支持	支持	不支持
日志服务 CLS	支持	不支持	不支持
对象存储 COS	不支持	支持	不支持

注意：

- 当云数据库 MySQL 用作数据源时，可用于 JOIN 条件的右表、或者作为 `QUERY_DB_STR` 函数（见 SQL 手册）的查询表。除上述两种场景外，暂不可用于其他用途（例如流式数据读取）。
- 云数据仓库由于底层所用的 PostgreSQL 版本过低，目前不支持作为 Upsert 数据目的。如果希望写入 Upsert 数据流，请使用云数据库 PostgreSQL 9.5及以上版本。

JAR 作业的上下游数据介绍

当**独享集群**的 VPC 与用户指定的 VPC 建立互通关系后，JAR 模式的作业即可访问用户特定 VPC 下的所有网络可达的资源，包括但不限于该 VPC 下的各项腾讯云服务，例如消息队列、数据库、API 服务、云服务器 CVM 等。此外，还可以在这个特定 VPC 下搭建代理服务，以访问外部的互联网地址（例如公网 API 等），进一步增强流计算作业的处理能力。

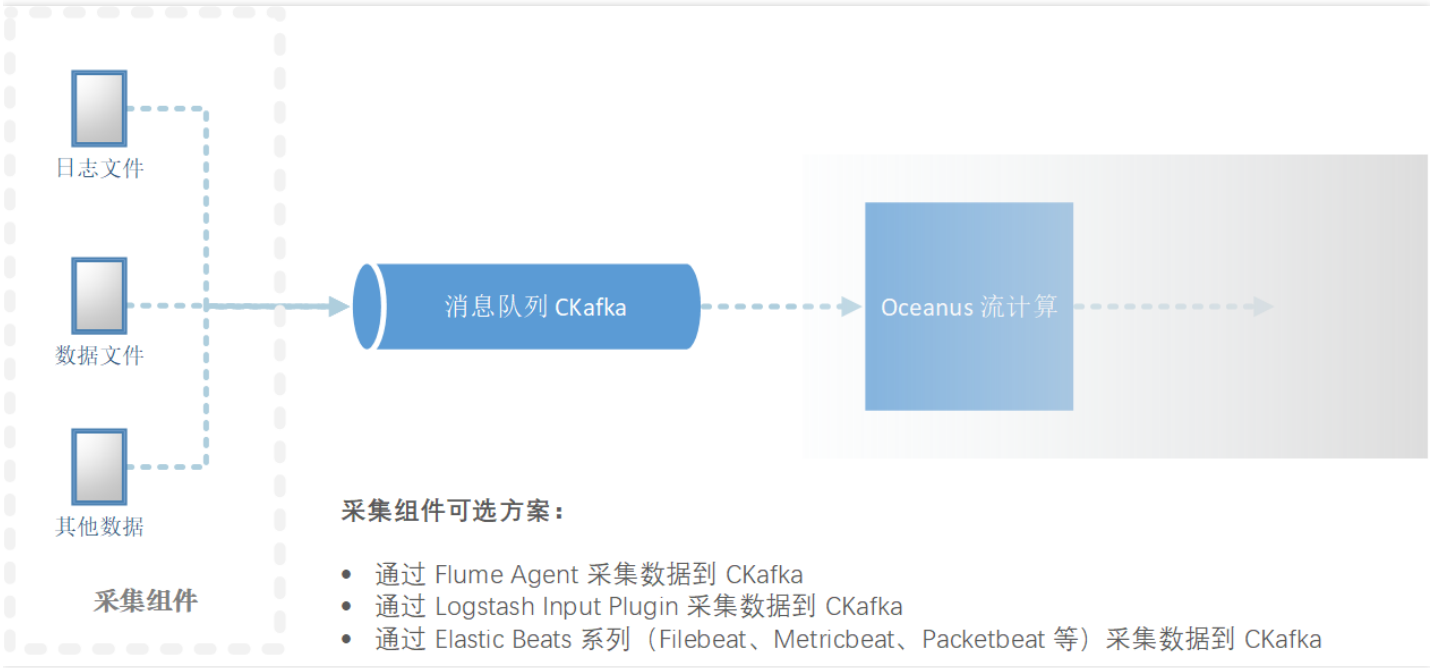
数据采集到 CKafka

最近更新时间：2020-06-01 16:10:58

背景介绍

用户可以将自己的日志、数据等信息通过多种采集组件输入到 CKafka 等产品中，作为数据源被流计算 Oceanus 来消费和处理。

例如，用户可以通过 Apache Flume 提供的 Agent，将日志、事件等数据源源源不断地从各个节点传输到 CKafka 产品中，整体流程如下图：



Apache Flume 数据采集操作指南

请参见 CKafka 产品文档中 Flume 接入 CKafka 最佳实践。

Logstash 数据采集操作指南

请参见 CKafka 产品文档中 Logstash 接入 CKafka 最佳实践。

Filebeat 数据采集操作指南

这里介绍使用 Filebeat 将日志文件导入 CKafka 的方式。

1. 下载 Filebeat 并安装

在官网 Filebeat 下载 页面，选择适合自己操作系统的 Filebeat 安装包，下载并安装到需要采集日志的主机上。以下默认用户使用的是 Linux 版本系统，且使用官方的安装包（rpm、deb）。

2. 配置 YAML 文件

对于 Linux 系统，Filebeat 默认安装后的配置文件位于 `/etc/filebeat/filebeat.yml` 。为了简单起见，我们选择直接编辑这个文件。

采集的文件位于 `/var/log/` 目录下，以 world- 前缀开头，以 .log 后缀结尾，命令如下：

```
[root@9 /var/log/hello]# ls
world-20190920.log  world-20190921.log  world-20190922.log  world-20190923.log  world-20190924.log  world-20190925.log
```

日志文件：每条日志可能有多行，但是每行的开头一定是2019-09-23这样的年-月-日数字格式，格式如下：

```
2019-09-23 14:46:22.025 [main] DEBUG  org.apache.flink.yarn.YarnFactory  -  Starting YARN TaskManager
2019-09-23 14:46:22.025 [main] INFO  org.apache.flink.yarn.YarnFactory  -  OS current user: hadoop
2019-09-23 14:46:22.365 [main] INFO  org.apache.flink.yarn.YarnFactory  -  Current Hadoop/Kerberos user: hadoop
2019-09-23 17:19:39.268 [Fetcher for Kafka] WARN  com.tencent.cloud.HelloThread  - Exception happened while fetching data:
org.apache.http.conn.ConnectTimeoutException: Connect to 10.0.0.1:8080 failed: connect timed out
    at org.apache.http.impl.conn.DefaultHttpClientConnectionOperator.connect(DefaultHttpClientConnectionOperator.java:151)
    at org.apache.http.impl.conn.PoolingHttpClientConnectionManager.connect(PoolingHttpClientConnectionManager.java:359)
    at org.apache.http.impl.execchain.MainClientExec.establishRoute(MainClientExec.java:381)
    at org.apache.http.impl.execchain.MainClientExec.execute(MainClientExec.java:237)
```

那么可以将上述的 `/etc/filebeat/filebeat.yml` 替换为如下内容：

```
filebeat.inputs:
- type: log
enabled: true
paths:
- /var/log/hello/world-*.log # 日志文件的路径, 路径支持通配符 *, 自动发现新增的日志
multiline.pattern: '^\d{4}-\d{2}-\d{2}' # 每条日志的起始格式（正则表达式），根据实际情况进行调整
multiline.negate: true
multiline.match: after
exclude_lines: ['DEBUG'] # 排除含有 DEBUG 字符串的行, 避免采集到大量调试日志
output.kafka:
enabled: true
hosts: ["10.0.0.1:9092"] # CKafka 的上报 IP 和端口，可以在 CKafka 的详情页查看
topic: '您要导入的 CKafka Topic' # 请先在 CKafka 中创建 Topic 并在这里填写
partition.hash:
```

```
hash: ['beat.hostname']
reachable_only: false
random: true
required_acks: 1
compression: gzip
version: 0.10.2.1
max_message_bytes: 10000
timeout: 10
```

更多的参数和解释，请参见 Filebeat 官网的 配置 Kafka 输出。

配置完成后，运行如下命令来启动 Filebeat 采集服务。随后即可通过消费数据导入的 CKafka Topic 来查看日志是否已经成功被采集。如果采集成功，则可以用于流计算的后续处理。

```
service filebeat start
```

查看 Filebeat 的运行日志时，运行如下命令，以判断 Filebeat 的上报是否遇到了问题。

```
tail -f /var/log/filebeat/filebeat
```

创建和管理集群

最近更新时间：2020-03-12 09:41:17

创建和管理集群

流计算 Oceanus 支持共享集群和独享集群两种模式：

1. 共享集群：由系统管理，用户不用关注它的创建和管理维护。
2. 独享集群：用户自己创建，由系统独立部署在一个专属的 VPC 中，与其他租户完全隔离，具有极高的安全性，只有独享集群才能支持提交 JAR 作业。

本文介绍内容仅针对独享集群。

创建集群

使用流计算 Oceanus 的独享模式，首先需要创建一个独享集群。

1. 在 流计算 Oceanus 控制台 的左侧菜单栏，单击【集群管理】，进入【我的集群】页面。

2. 在【我的集群】页面单击【新建集群】，进入【创建集群】界面。

根据需求，选择相应的配置，单击【立即购买】完成购买，相关参数说明如下：

参数	说明
地域/可用区	建议选择最近的地域，可降低访问时延、提高访问速度。详细说明请参见 地域和可用区
集群名称	自定义集群名称
集群描述	帮助识别此集群的描述
计算资源（CU）	1CU = 1核4G，请根据您的计算需要选择
VPC	这是独享集群关联的 VPC 和子网，独享集群通过网络打通，访问该网络环境下的资源和服务，详细的网络架构可以查看 了解共享和独享集群

3. 集群创建完成后，在【集群管理】中查看集群列表中新建的集群，等待集群状态从“未初始化”变为“运行中”，代表集群创建成功。目前集群创建是半自动的，通常在24小时内完成创建。

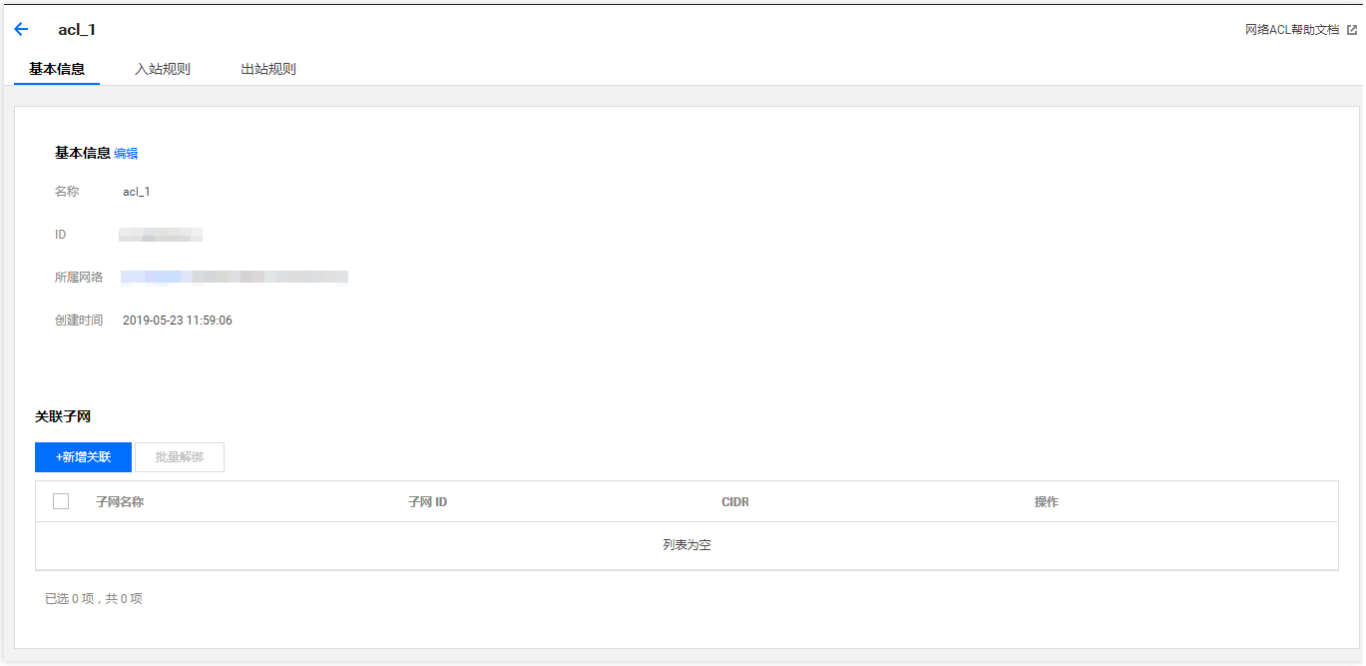
创建对等连接

1. 集群创建完成后，后台会自动帮您完成对等连接的创建工作，保证您的 VPC 与创建集群 VPC 的网络互通。选择【私有网络】>【对等连接】，查看已创建的对等连接。

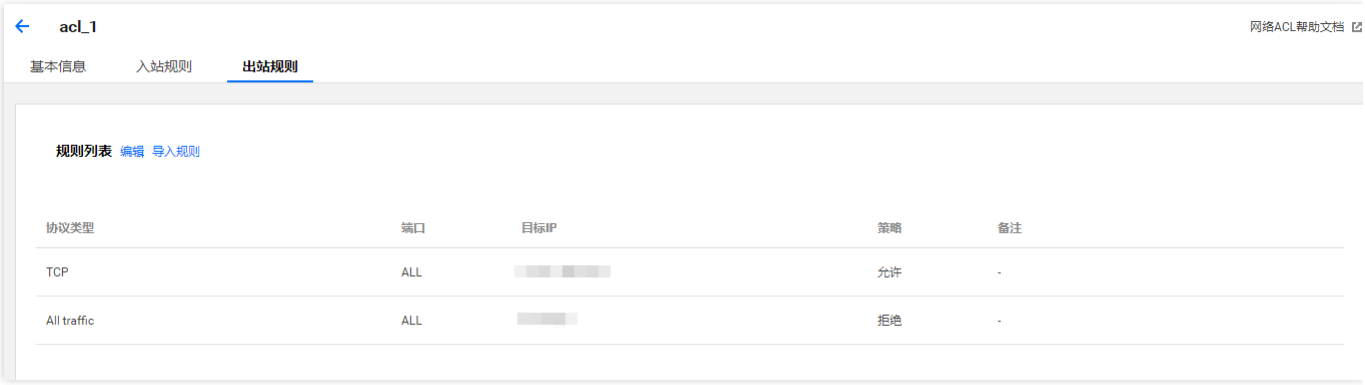
2. 创建对等连接时，也会创建相关的安全组、网络 ACL、以及添加相应的路由信息，以允许您的独享集群中的作业能访问您的 VPC 中的资源，如 CKafka、TencentDB 等。
   - 查看安全组：选择【私有网络】>【安全】>【安全组】查看新建的安全组。

> 说明：
> 如果您的 VPC 中的资源，由于 IP 出入站规则的限制，导致集群中的作业无法访问，只需将该安全组配置到资源上即可。

- 查看网络 ACL 规则：选择【私有网络】>【安全】>【网络ACL】，查看是否针对您的 VPC 设置网络 ACL。

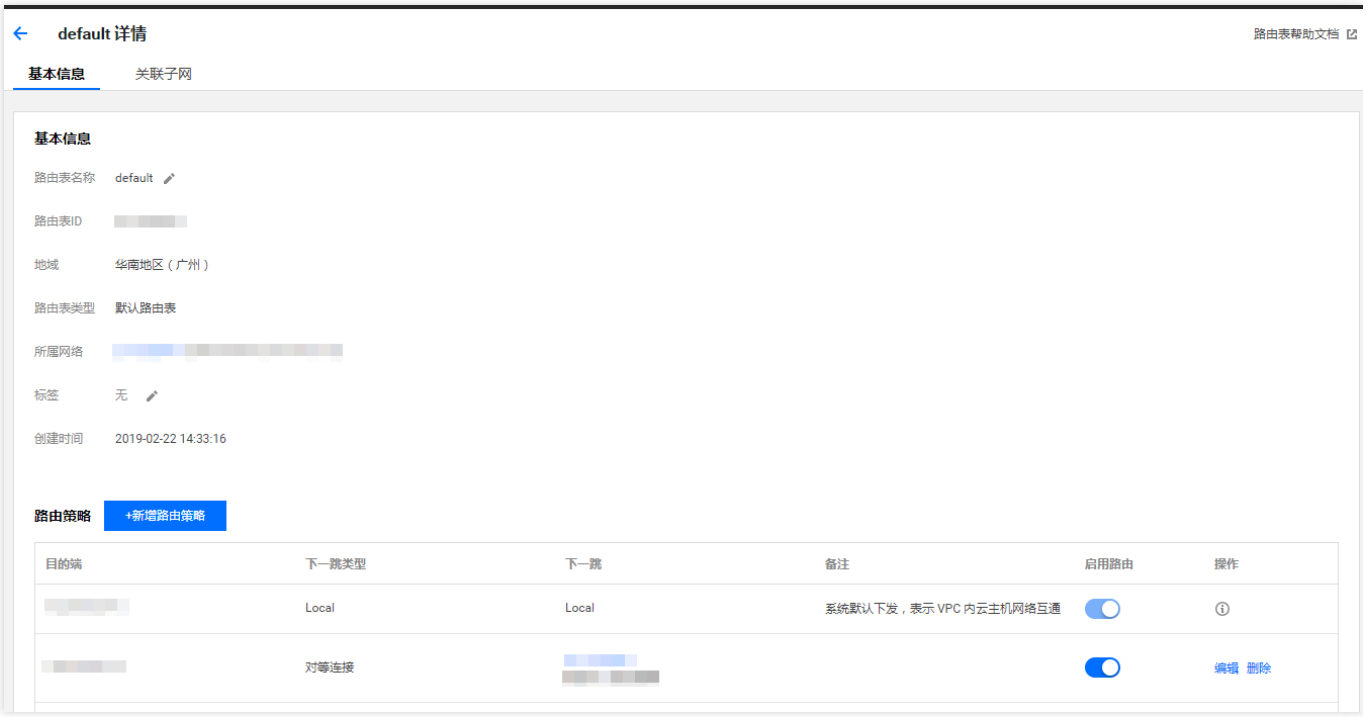


在入站/出站规则中，查看是否有允许独占集群访问的规则。



> 说明：
> 如果网络 ACL 中有关于用户 VPC 的限制，在创建对等连接时，会在对应的网络 ACL 中加入独占集群 VPC 的出站/入站访问规则。若网络 ACL 中没有关于用户 VPC 的限制，则不会修改网络 ACL。

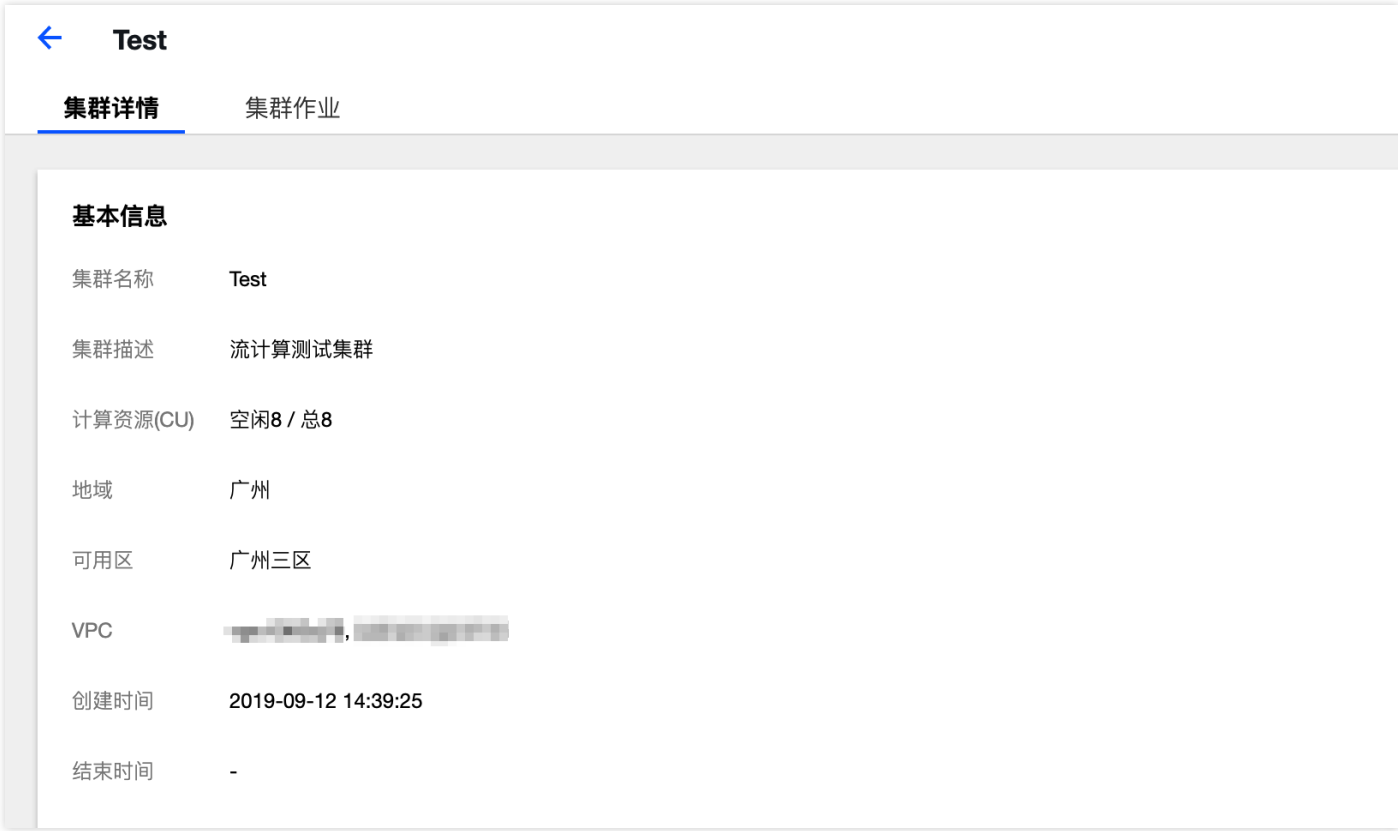
- 查看路由策略：选择【私有网络】>【路由表】，进入用户 VPC 对应的路由表，查看基于用户的对等连接创建出来的路由策略。



查看集群详情

1. 查看集群详情

   在【我的集群】页面，单击想要查看的集群名称，可进入集群详情页面：



相关参数如下：

参数	说明
集群名称	自定义集群名称
集群描述	帮助识别此集群的描述
计算资源（CU）	集群空闲的 CU/集群总共拥有的 CU
地域/可用区	此集群所在的地域/可用区
VPC	这是独享集群关联的 VPC 和子网，独享集群通过网络打通，访问该网络环境下的资源和服务，详细的网络架构可以查看 了解共享和独享集群
创建时间	集群创建时间
结束时间	集群付费的到期时间

2. 查看集群下的作业

单击【集群作业】，这里展示了所有在此集群中提交的作业。

创建 SQL 作业

最近更新时间：2020-03-10 18:33:24

SQL 作业是通过 SQL 语句直接编写业务逻辑的方式。对于流计算开发人员来说，是一种实现简单高效、体验友好的作业开发方式。

前提条件

创建 SQL 作业时，需要先准备好上下游数据，具体请参见 上下游数据一览。

操作步骤

1. 创建 SQL 作业

登录 流计算 Oceanus 控制台，单击左侧菜单栏【流计算 Oceanus】下的【作业管理】，进入作业管理页面。单击【新建SQL作业】，进入创建 SQL 作业页面，输入相关信息购买和创建作业。

相关信息如下：

参数	说明
地域	作业所在地域
作业名称	给作业取一个名字

2. 流计算服务委托授权

在【作业管理】页面单击某个作业名称打开【作业详情】页，单击【分析开发】，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。

> 说明：
> 此授权的详细说明参见 流计算服务委托授权。

3. 编写 SQL 语句

在【作业管理】页面单击某个作业名称打开【作业详情】页，单击【分析开发】，进入作业开发页。在代码编辑框输入 SQL 语句进行业务逻辑开发，SQL 语句编写请参见 SQL 手册-术语表。单击【语法检查】检查语法是否完全正确，如果提示错误，请根据提示修改 SQL 语句。

编写 SQL 语句方式如下：

1. 直接编写 SQL 语句
   可直接在代码编辑框中输入 SQL 语句。

```
CREATE TABLE `page_visits` (
  `record_time` VARCHAR,
  `user_id` VARCHAR,
  `page_id` VARCHAR
) WITH (
  `type` = 'ckafka',
  `instanceId` = 'ckafka-2b6b5m11',
  `encoding` = 'json',
  `topic` = 'page_visits'
);

CREATE TABLE `page_info` (
  `page_id` VARCHAR,
  `page_name` VARCHAR,
  PRIMARY KEY ( `page_id` )
) WITH (
  `type` = 'mysql',
  `instanceId` = 'cdb-o3ubl2ss',
  `user` = 'root',
  `password` = 'hd514256525',
  `database` = 'my_database',
  `table` = 'page_info'
);

CREATE TABLE `pv_uv_results` (
  `record_time` VARCHAR,
```

2. 通过【分析模板】生成 SQL 语句

单击代码编辑框上方的【分析模板】，在弹出的【选择模板】对话框中选择一个合适的模板，将模板代码添加到代码编辑框中，然后根据自己的业务实际情况修改代码逻辑。

3. 通过【选择数据流】生成 SQL 语句

单击代码编辑框上方的【选择数据流】，通过【添加数据流DDL】对话框自动生成 SQL 代码，然后编写计算逻

辑。

示例如下：类型选择 CKafka，实例 ID 选择您在准备工作中创建的 CKafka，Topic 选择作为数据源的 Topic，数据格式选择 json，输入字段结构，单击【预校验】来校验输入的字段结构是否与 CKafka 中一致，最后单击【添加】。

添加完数据流以后，在代码编辑框中会生成如下代码：

```
CREATE TABLE `page_visits` (
`record_time` VARCHAR,
`user_id` VARCHAR,
```

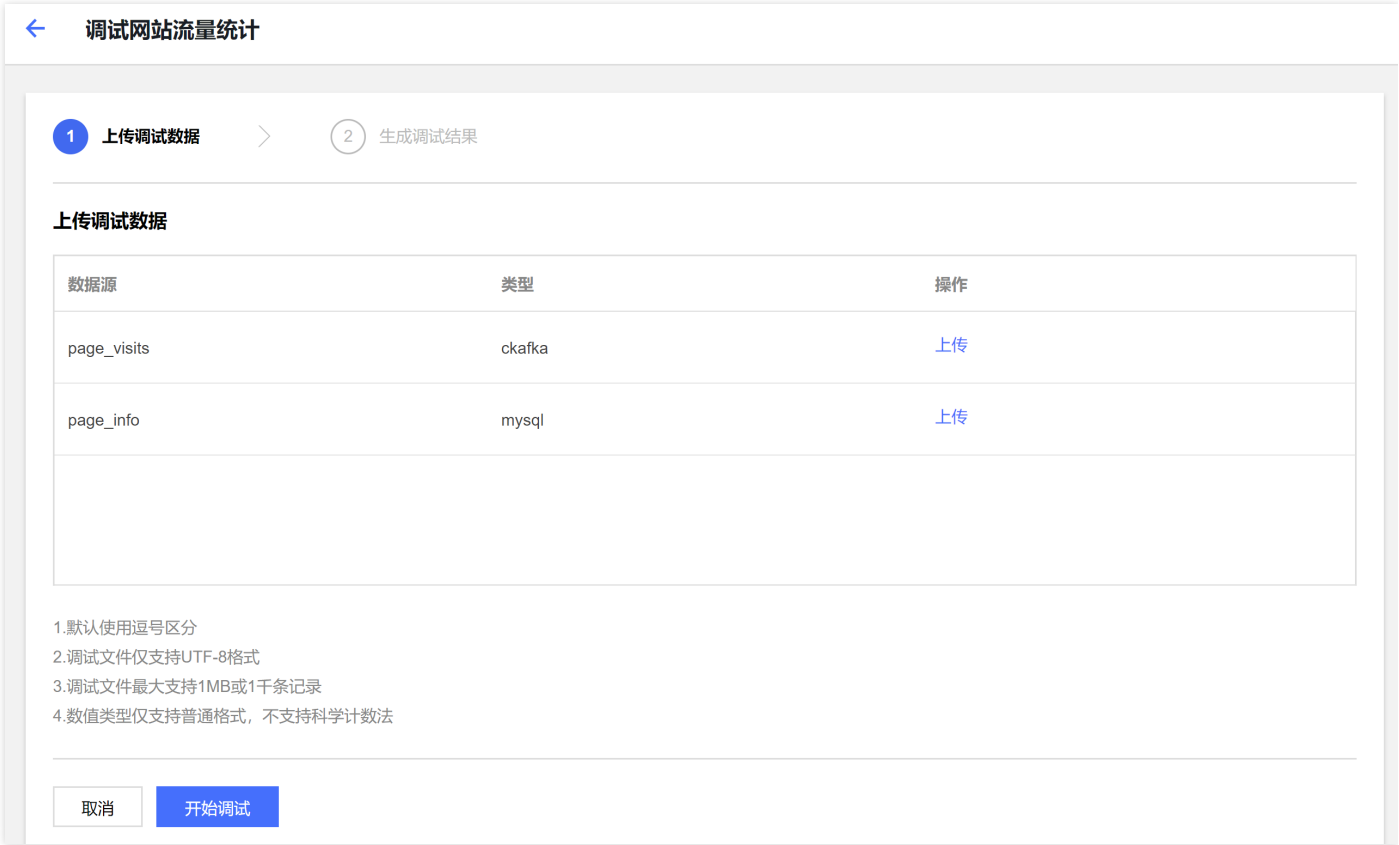
```
`page_id` VARCHAR
) WITH (
`type` = 'ckafka',
`instanceId` = 'ckafka-xxxxxxxx',
`encoding` = 'json',
`topic` = 'page_visits'
);
```

您可以用同样的方式添加其他数据流如 TencentDB For MySQL 等。

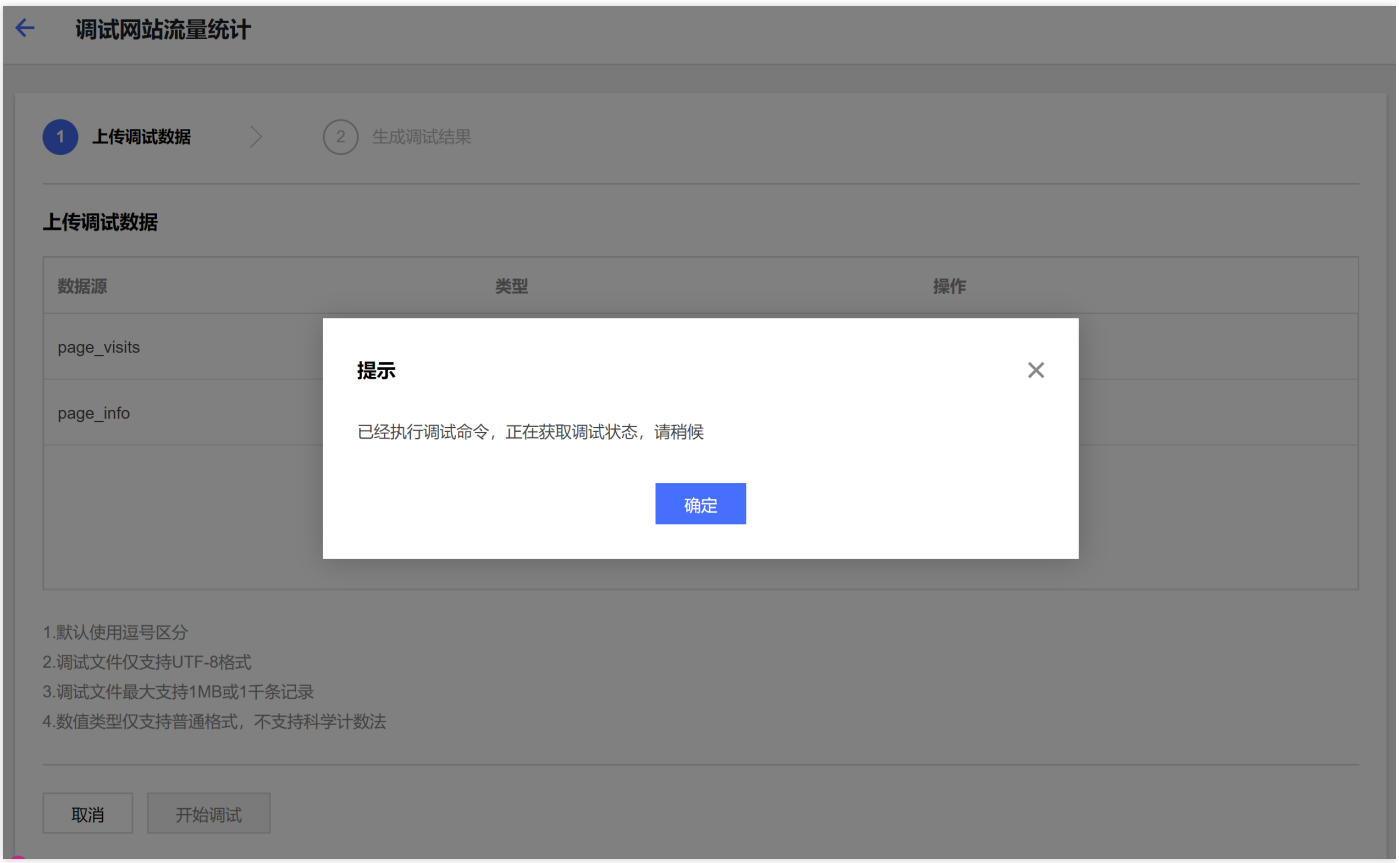
4. 调试 SQL 作业

编写完代码后，可选择在发布之前先进行调试。我们为用户提供了一套独立的模拟环境，用户可以在该环境中上传数据，模拟运行并检查输出结果。

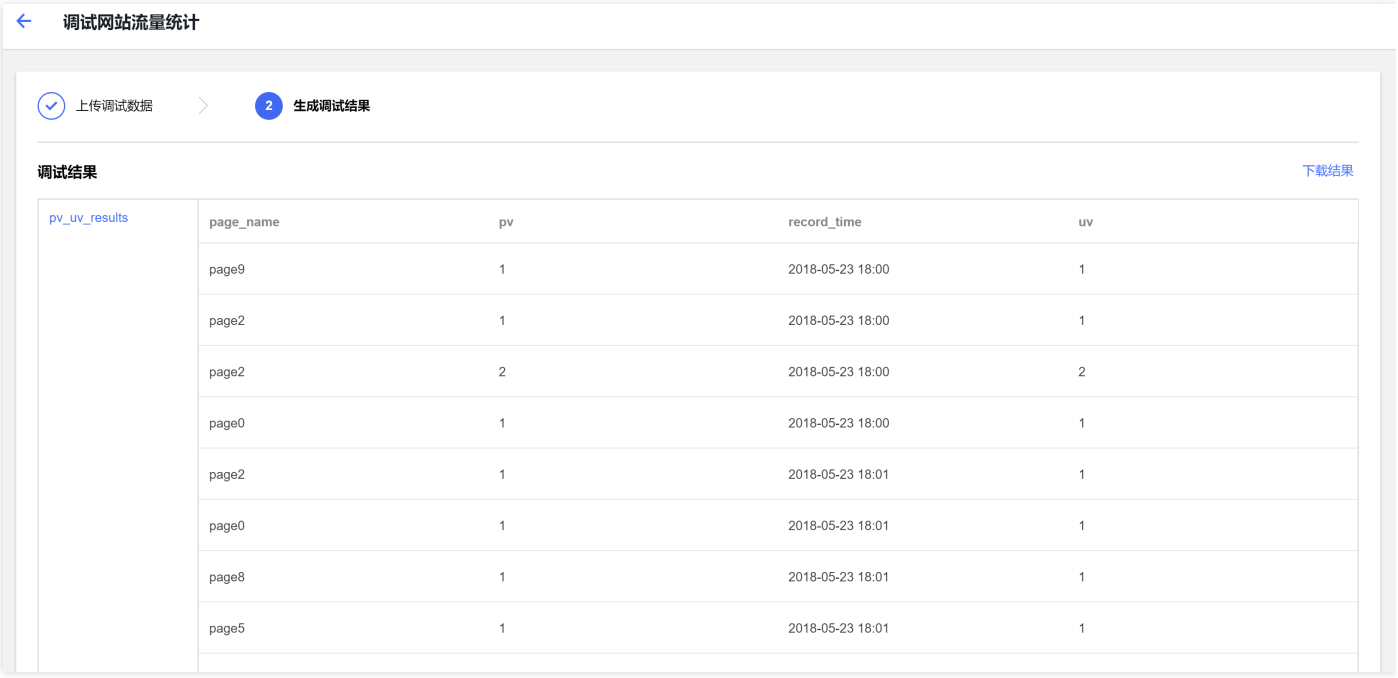
1. 选择【分析开发】>【调试】，会进入调试页面，如下所示：



2. 单击对应数据源的【上传】，从本地上传数据文件，单击【开始调试】，界面显示如下：



3. 单击【确定】，等待片刻后，会显示调试结果，可单击【下载结果】下载调试结果。

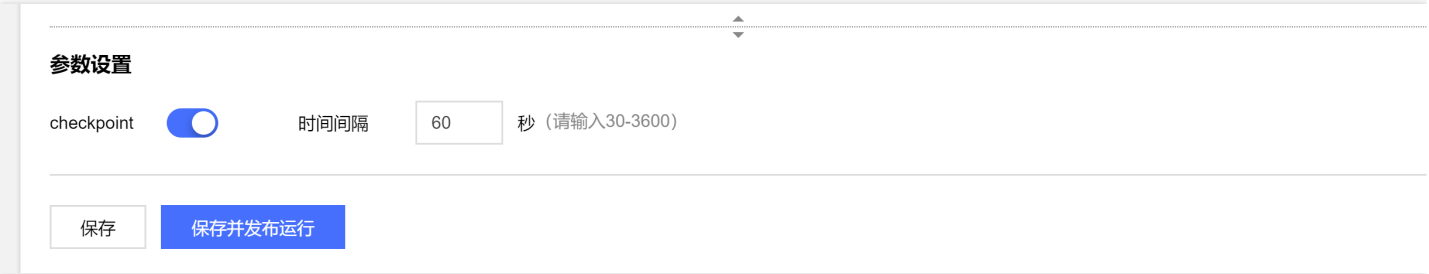


根据调试结果判断 SQL 代码的执行达到预期效果，即可进行发布。

5. 发布运行 SQL 作业

参数设置

在【分析开发】页的【参数设置】区域，设置是否开启 checkpoint 以及时间间隔。



参数	说明
checkpoint	定时快照，用来保存作业的运行状态，当作业意外崩溃时可从最近快照点恢复
时间间隔	进行 checkpoint 操作的时间间隔

发布运行

单击【保存并发布运行】，弹出如下对话框，允许指定数据消费时间点来运行作业，可以选择您期望的时间点，单击【确定】，提交运行作业。

> 说明：
> 消费时间点指定流计算作业从何时开始消费数据。

创建 JAR 作业

最近更新时间：2020-03-10 18:32:41

在 JAR 作业方式下，是用户自己基于 Flink 的 API 来开发业务逻辑，形成 JAR 包，然后提交到云端流计算集群上运行。这种方式适用于熟悉 Flink 开发的开发人员，支持他们通过更自由的开发实现，去应对复杂度更高的需求场景。

前提条件

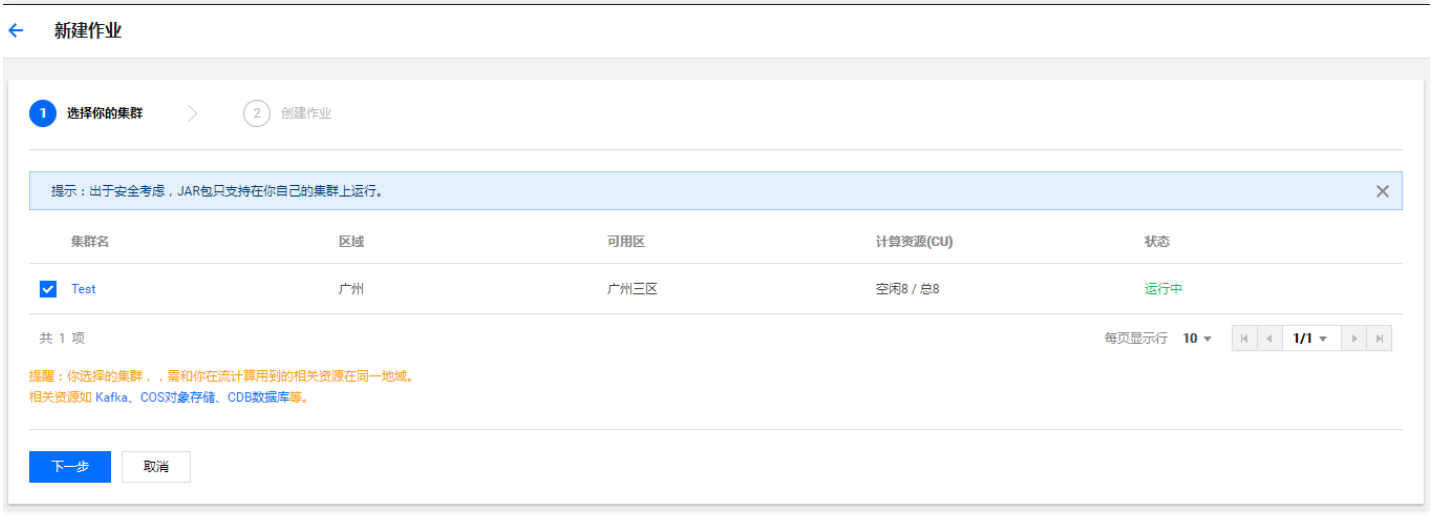
- 需要根据业务需求情况，基于 Flink 的 API 开发好自己的 JAR 包。
- 需要先创建好自己的独享集群，请参见 创建和管理集群。
- 需要准备好上下游数据，具体请参见 上下游数据一览。

操作步骤

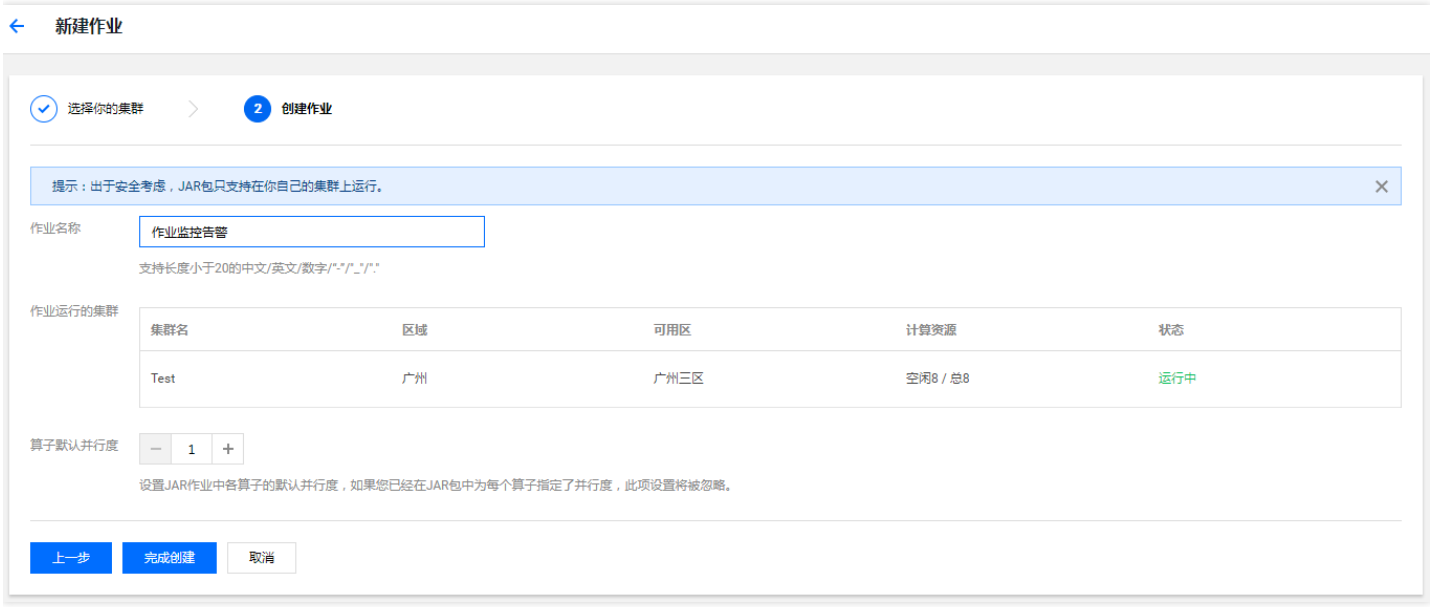
1. 创建 JAR 作业

登录 流计算 Oceanus 控制台，单击左侧菜单栏【流计算】下的【作业管理】，进入作业管理页面。单击【新建JAR作业】，进入【新建作业】，选择您的独享集群，单击【下一步】。

> 说明：
> 如果这一步您没有可选的独享集群，请参见 创建和管理集群 创建独享集群。



输入相关信息后单击【完成创建】，即可在【作业管理】页面看到新创建的作业。



相关信息如下：

参数	说明
作业名称	给作业取一个名字
作业运行的集群	您在上一步选择的自己创建的独享集群
算子默认并行度	请参见 作业并行度设置

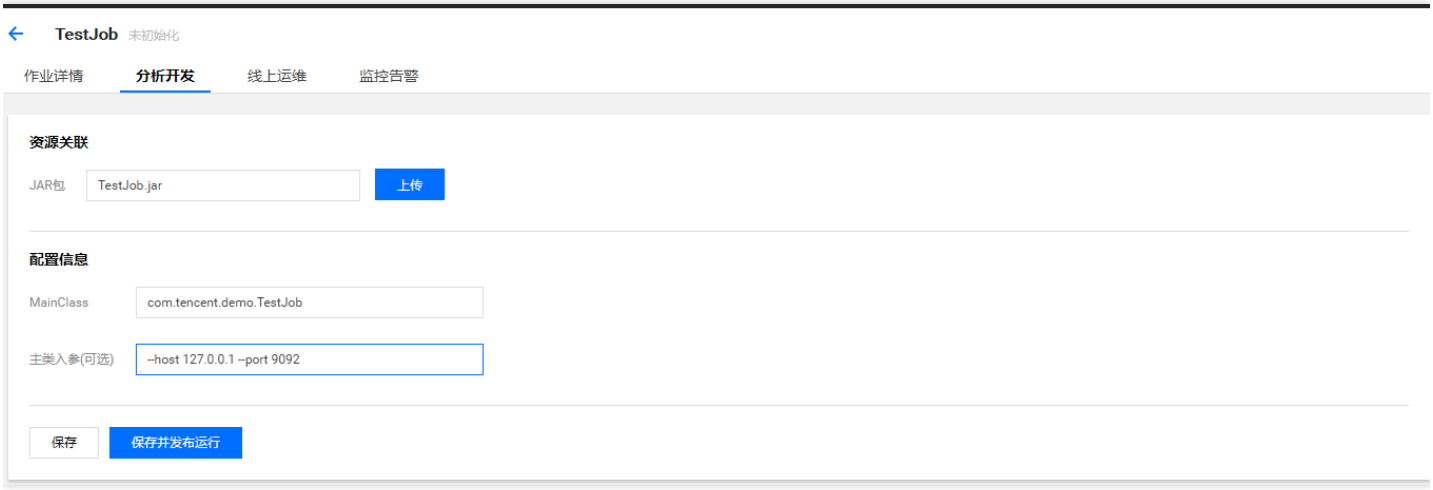
2. 流计算服务委托授权

在【作业管理】页面单击某个作业名称打开【作业详情】页，单击【分析开发】，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。

说明：
此授权的详细说明参见 流计算服务委托授权。

3. JAR 作业开发和发布

在【作业管理】页面单击某个作业名称打开【作业详情】页，单击【分析开发】，进入作业开发页面。上传准备好的 JAR 包，输入配置信息，单击【保存并发布运行】，提交作业。



相关信息如下：

参数	说明
MainClass	JAR 包运行时的主类
主类入参	JAR 包运行时主类的输入参数

作业并行度设置

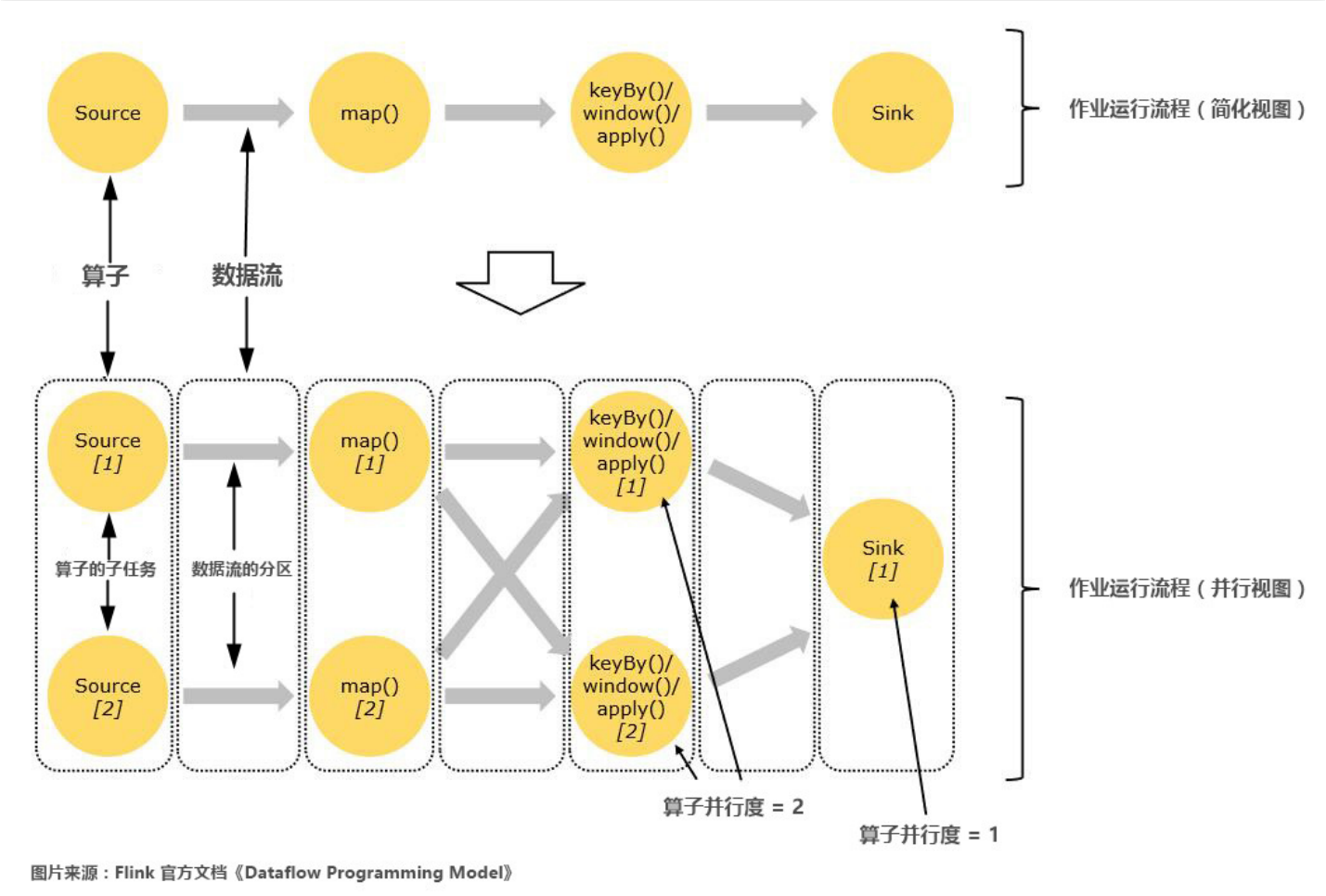
最近更新时间：2020-07-28 09:51:44

作业并行度说明

什么是作业的并行度

流计算作业在运行时是并行执行的。简单来讲，所有算子及其之间的连接关系，组成了整个作业的运行图，而运行图中的各个算子组成了一个个的任务（Task）。每个任务可以分为若干个并行执行的子任务（Subtask），这些子任务的个数就是算子的并行度（Parallelism）。

对同一个作业而言，所有算子中最大的并行度，就是整个作业的并行度。例如下图中，有4个算子，前3个算子的并行度为2，最后一个算子的并行度为1，那么整个作业的并行度取最大值2。



图片来源：Flink 官方文档《Dataflow Programming Model》

为什么要调整并行度

由上图可以看到，理论上来讲，算子的并行度越大，每个算子同时处理数据的能力相对就越强，同时整个作业的吞吐量也会提升，相应地资源占用也会上升；反之，降低并行度，那么每个算子处理能力减少，但整个作业的资源占用也会降低。当作业资源不足时，提升并行度，可以实现扩容；当作业资源冗余时，降低并行度，可以实现缩容。

因此，可以根据作业的实际运行情况，得到一个理想的并行度，从而在保证处理能力的情况下，避免资源浪费，节省成本。

作业并行度和计算资源 CU 的关系

在流计算 Oceanus 中，目前作业的并行度与 CU 之间的比例为1:1。例如，如果一个作业运行后，实际的并行度是8，那么它所消耗的计算资源 CU 数就是8。

在未来的版本中，会开放并行度与 CU 之间的比例调整，例如允许一个计算资源 CU 对应2个及以上并行度，以更灵活地适配多种场景（例如对 CPU 要求不高的 I/O 密集型作业）。

作业并行度与默认并行度的关系

流计算 Oceanus 作业在【购买】和【变配】时，允许设置“默认并行度”。默认并行度只是一个默认值，代表如果程序里未显式指定某个算子的并行度，那么就以默认并行度为准。

- 对于 **SQL 作业**，由于目前不提供运行图算子并行度的设置，那么可以认为购买或变配时设置的“默认并行度”就是最终的作业并行度。
- 对于 **JAR 作业**，用户可以在代码里显式设置每个算子的并行度信息，因此作业并行度 = 提交 JAR 作业时生成的运行图中所有算子的并行度中最大的值。

作业默认并行度初始设置

- 选择【作业管理】>【新建SQL作业】，进入作业创建页面。在这个页面，用户需要选择“计算资源”。由于 SQL 作业目前不提供单个算子的并行度调整选项，这里的计算资源数，可以认为是默认并行度。
- 选择【作业管理】>【新建JAR作业】，首先要选择集群，集群选择后，会进入作业创建界面。在这个界面，用户需要选择“算子默认并行度”。

作业默认并行度查看和调整

对于 SQL 作业，如果要查看它的默认并行度，可以在作业列表页，查看作业“集群”列表内的已购 CU 数。例如一个作业显示为“已购3CU”，则表明默认并行度为3。

对于 JAR 作业，目前界面上还无法查看默认并行度。暂时需要通过腾讯云 API 的 DescribeJobs 接口，先查询作业当前所用的 JobConfig 版本号，然后通过 DescribeJobConfigs 接口，查询这个 JobConfig 版本的 DefaultParallelism 字段值。具体用法可 提交工单 咨询。

运维和监控
作业查看和操作

最近更新时间：2020-06-01 18:08:44

流计算作业发布成功后，会进入“运行中”状态。此时，可以对其进行运维操作（例如暂停、停止等），以及查看各项运行时的监控指标。当作业被暂停后，可以后续从暂停的状态恢复；当作业被停止后，也可以重新点击启动。

暂停

概念解析

作业的“暂停”，表示创建一个快照点，以保存作业当前的运行状态（数据源消费的偏移量、运行中的状态数据，以及数据目的的一些信息等），然后结束这个作业的运行。用户可以稍后从这个暂停点进行恢复。

操作步骤

当一个作业处于运行中状态时，可在流计算的【作业管理】>【操作】中单击【更多】，然后在下拉菜单中选择【暂停】。随后，作业的状态会变成“操作中”。一段时间后，最终状态会变成“暂停”，此时表示作业已经成功被暂停。

继续

概念解析

作业的“继续”操作与“暂停”相对应，表示将作业从上次暂停的快照点恢复，作业的各种状态也会和暂停前保持一致，然后开始继续数据处理过程。

> 注意：
> 如果上游数据源的消费偏移量所处的位置已经无法消费（例如，在消费 CKafka 时，超过了消息过期时间，之前的旧数据已经过期），那么仍然会导致部分数据丢失。因此，请确保暂停的时间不要超过外部数据源中数据的过期时间。

操作步骤

当一个作业进入“暂停”状态后，如果需要继续运行，可在流计算的【作业管理】>【操作】中单击【更多】，然后在下拉菜单中选择【继续】。随后，作业的状态会变成“操作中”。一段时间后，最终状态会变成“运行中”，此时表示作业已经恢复成功，正在从上次停下的快照点开始，继续开始处理数据。

停止

概念解析

作业的“停止”操作表示终止它的执行，并**丢弃**所有运行时的状态。

> 注意：
> 如果您希望保留作业当前的运行状态，并让作业下次启动时从上次停下的地方开始消费，请不要选择停止操作，而要选择上文提到的“暂停”。

操作步骤

当一个作业进入“运行中”状态后，如果需要停止，可在流计算的【作业管理】>【操作】中单击【更多】，然后在下拉菜单中选择【停止】。随后，作业的状态会变成“操作中”。一段时间后，最终状态会变成“停止”，此时表示作业已经完全停止运行。

运行

概念解析

作业的“运行”操作与“停止”相对应，表示全新启动一个新的作业运行实例。此时，对于 SQL 作业，会弹出一个对话框，用户可以选择消息队列 CKafka 开始消费的时间点，可以选择从最新的时间点消费，也可以从最早的时间点消费，还可以从指定的时间点开始消费；对于 JAR 作业，则直接开始新的一次运行。

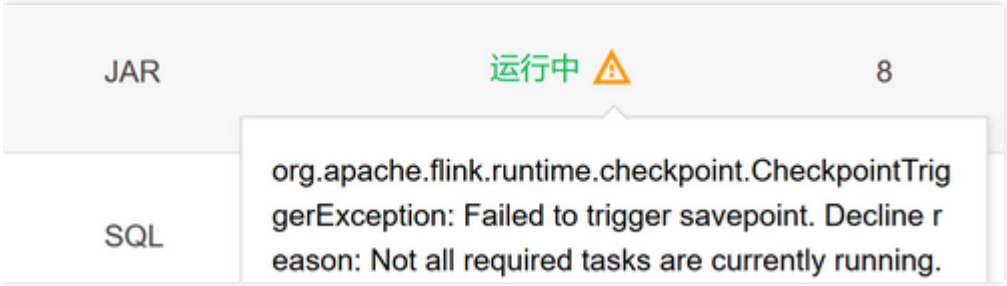
> 注意：
> 消费时间点的选择依赖于上游消息队列，如果设置的消息过期时间很短，那么即使选择从最早的时间点开始消费，那么也有概率无法读取到所有预期的历史数据。

操作步骤

当一个作业进入“停止”状态后，如果需要继续运行，可在流计算的【作业管理】>【操作】中单击【更多】，然后在下拉菜单中选择【运行】。随后，作业的状态会变成“操作中”。一段时间后，最终状态会变成“运行中”，此时表示作业已经启动成功。

特别提示

在作业的操作过程中，如果遇到任何异常情况，则会回退到作业的实际状态（例如，如果作业在暂停时异常退出，则状态会显示为“停止”；如果作业暂停不成功，且仍在运行，则显示为“运行中”），并在状态栏右侧显示一个三角形的叹号⚠，当鼠标移过时，会显示出具体的报错信息，如下图：



在作业的运行过程中，请务必谨慎修改上下游对接产品的配置，包括但不限于对 CKafka 数据源和数据目的所使用的 Topic 做删除、扩容；以及对 MySQL 数据源和数据目的所使用的库表做锁表、修改表结构、新增约束、停机等，否则会对正在运行的流计算作业造成影响，导致数据不完整或作业异常。

作业详情页

在【作业管理】页，单击作业列表中的作业名，可以打开【作业详情】页查看作业的详情信息。

项	解释
作业名称	该作业的名称
作业 ID	该作业的 Serial ID 信息，通常以 cql- 开头
作业类型	作业的类型，目前有 SQL 和 JAR 两种类型
运行状态	作业的当前状态，例如运行中、停止、暂停等等
地域	作业运行的集群所在的地理大区，例如广州、上海、北京等等
可用区	作业运行的集群的可用区，例如上海三区
创建时间	作业被创建的时间点
累计运行时长	作业历史上总共运行的时长
开始运行时间	作业本次开始运行的时间点
运行时长	作业本次运行所持续的时长
运行 CU	作业本次运行所占用的 CU 数

线上运维页

SQL 作业的线上运维

对于 SQL 作业的线上运维页，可以查看本作业的 SQL 代码，还可以查看作业的实时运行结果、源数据等内容。

> 注意：
> 作业的实时运行结果展示、源数据展示功能目前只支持 CKafka，对于 MySQL、PostgreSQL 等其他数据源暂不支持。

JAR 作业的线上运维

对于 JAR 作业的线上运维页，目前可以查看“资源关联”信息，即该作业所绑定的 JAR 包名；还可以查看“配置信息”，目前有 MainClass（JAR 包的主类），主类入参等信息。更多功能即将推出。

监控告警

最近更新时间：2020-06-02 09:53:44

监控指标查看

对于正在运行（或者曾经成功运行过）的流计算作业，用户有两种方式查看监控信息。

1. 单击作业名称，在跳转页面中选择【监控告警】（如下图），在该页面可以查看作业的总览，例如启动时间、运行时长、平均处理耗时等关键指标，也可以查看数据每秒的输入输出条数、CPU 与内存的使用率等情况。

2. 选择【作业管理】>【操作】>【云监控】，即可进入 云监控控制台。在这里可以查看更为详细的监控指标，也可以配置作业专属的监控告警策略。
监控指标列表：

> 注意：
> 部分指标仅在上述两个界面中的其中一个有提供。

指标中文名	指标含义	示例值
启动时间	当前作业启动的时间点	2019-09-10 14:46:26
运行时长	当前作业本次启动的运行时长	1天2时20分
处理平均耗时（计算耗时）	作业中各个算子延时的平均值	3ms
数据平均延时（数据延时）	作业当前 Watermark 水位线与系统时间戳之间的差值	0ms
每秒输入数据	当前作业的所有数据源，每秒输入的总数据流量	16038.23字节/秒
每秒输出数据	当前作业的所有数据目的，每秒输出的总数据流量	8753.82字节/秒
每秒记录输入	当前作业的所有数据源，每秒输入的总数据条数	9854条/秒

每秒记录输出	当前作业的所有数据目的，每秒输出的总数据条数	4875条/秒
CPU 使用率	当前作业的所有计算节点之平均 CPU 使用率	12.6%
内存使用率	当前作业的所有计算节点之平均内存使用率	87.5%
作业堆外直接内存用量	当前作业所有计算节点之平均 JVM 堆外直接内存用量	428781Byte
作业堆内存用量	当前作业所有计算节点之平均 JVM 堆内内存用量	1392410832Byte
作业堆外映射内存用量	当前作业所有计算节点之平均 JVM 堆外内存映射占用内存用量	0Byte
作业非堆内存用量	当前作业所有计算节点之平均 JVM 堆外内存用量（不含 JNI 分配）	105658232Byte

告警服务配置

流计算 Oceanus 产品的告警策略是通过云监控服务来实现的。下面我们针对一些常见的场景进行描述，详情请参见告警概述。

查看作业现有告警策略

用户可以在云监控中选择【告警配置】>【告警策略】，默认可以查看所有产品的告警策略配置。通过在“产品/策略类型”选项框中选择“流计算 Oceanus”，并单击【查询】，则可以查看所有为流计算作业配置的告警项。

新增作业告警策略

1. 新增流计算 Oceanus 作业的告警策略时，可在上述界面单击【新增】，输入策略名称，并填写可选的备注信息。
2. 在“策略类型”下拉框中选择“流计算 Oceanus”，下面即会提示选择“告警对象”。这里可以针对特定作业，或者所有作业进行策略配置，按 Shift 键即可多选。
3. 当告警对象选择完毕，则可以选择“触发条件”。可以在 触发条件模板 中选择已经配置好的模板，或者新增模板。另外，如果不需要使用模板，则可以选择“配置触发条件”，这里可以对上述的多项监控指标做阈值配置和告警。
4. 选择“告警渠道”及消息接收人、接收时段和渠道等信息，并可以配置接口回调（可选）。
5. 当所有内容配置完毕，单击【完成】，则新的告警策略会立刻生效。