

ALIBABA CLOUD

阿里云

专有云企业版

实时计算（流计算）
用户指南

产品版本：v3.16.2

文档版本：20220916



法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
 危险	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 危险 重置操作将丢失用户配置数据。
 警告	该类警示信息可能会导致系统重大变更甚至故障，或者导致人身伤害等结果。	 警告 重启操作将导致业务中断，恢复业务时间约十分钟。
 注意	用于警示信息、补充说明等，是用户必须了解的内容。	 注意 权重设置为0，该服务器不会再接受新请求。
 说明	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 说明 您也可以通过按Ctrl+A选中全部文件。
>	多级菜单递进。	单击设置> 网络> 设置网络类型。
粗体	表示按键、菜单、页面名称等UI元素。	在结果确认页面，单击确定。
Courier字体	命令或代码。	执行 <code>cd /d C:/window</code> 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	<code>bae log list --instanceid</code> <code>Instance_ID</code>
[] 或者 [a b]	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
{ } 或者 {a b}	表示必选项，至多选择一个。	<code>switch {active stand}</code>

目录

1.什么是实时计算	06
2.快速入门	07
2.1. 登录实时计算控制台	07
2.2. 热词统计	07
2.2.1. 概述	07
2.2.2. 代码开发	07
2.2.3. 代码调试	09
2.2.4. 数据运维	10
2.3. 天猫双十一大屏	10
2.3.1. 概述	11
2.3.2. 问题描述	11
2.3.3. 准备工作	12
2.3.4. 注册数据存储	14
2.3.5. 开发	14
2.3.6. 运维	15
3.管理项目	17
4.数据存储	20
4.1. 概述	20
4.2. VPC访问授权	20
4.3. 数据存储概览	21
4.3.1. 概述	21
4.3.2. 存储类别	21
4.3.3. 存储使用	22
4.4. 注册大数据总线（DataHub）	24
4.5. 注册日志服务（Log Service）	26
4.6. 注册表格存储（Tablestore）	27

4.7. 注册云数据库（RDS）	28
5.数据开发	34
5.1. 创建作业	34
5.2. 开发阶段	34
5.2.1. SQL辅助	34
5.2.2. SQL版本管理	34
5.2.3. 数据存储管理	35
5.3. 调试阶段	35
5.4. 上线阶段	38
5.5. 启动作业	39
5.6. 暂停作业	40
5.7. 停止作业	40
5.8. 查看日志	41

1.什么是实时计算

阿里云实时计算（Alibaba Cloud Realtime Compute）是运行在阿里云平台上的流式大数据分析平台，为您提供云上流式数据实时化的分析工具。使用阿里云Flink SQL，您无需完成底层流式处理逻辑的开发工作，即可搭建自己的流式数据分析和计算服务。

目前对信息高时效性、可操作性的需求不断增长，这要求软件系统在更少的时间内能处理更多的数据。传统的大数据处理模型将在线事务处理和离线分析从时序上完全分割，但该架构目前已经越来越落后于人们对于大数据实时处理的需求。

实时计算的产生来源于对于上述数据加工时效性的严苛需求：数据的业务价值随着时间的流失而迅速降低。因此在数据发生后必须尽快对其进行计算和处理。而传统的大数据处理模式对于数据加工均遵循传统日清日毕模式，即以小时甚至以天为计算周期对当前数据进行累计并处理。显然这类处理方式无法满足数据实时计算的需求。在诸如实时大数据分析、风控预警、实时预测、金融交易等诸多业务场景领域，批量（也称为离线）处理。对于数据处理时延要求苛刻的应用领域而言是完全无法胜任业务需求的。而实时计算作为针对流数据的实时计算模型，可有效地缩短全链路数据流时延、实时化计算逻辑和平摊计算成本，最终有效满足实时处理大数据的业务需求。

什么是流数据

从广义上说，所有大数据的生成均可以看作是一连串发生的离散事件。这些离散的事件以时间轴为维度进行观看就形成了一条条事件流或数据流。不同于传统的离线数据，流数据是指由数千个数据源持续生成的数据，流数据通常也以数据记录的形式发送，但相较于离线数据，流数据普遍的规模较小。流数据产生源头来自于源源不断的事件流，例如客户使用您的移动或Web应用程序生成的日志文件、网购数据、游戏内玩家活动、社交网站信息、金融交易大厅或地理空间服务，以及来自数据中心内所连接设备或仪器的遥测数据。

实时计算具备以下特点：

- 实时（Real time）且无界（Unbounded）的数据流

实时计算面对计算的数据源是实时且流式的，流数据是按照时间发生顺序地被实时计算订阅和消费。且由于数据发生的持续性，数据流将长久且持续地集成进入实时计算系统。例如，对于网站的访问单击日志流，只要网站不关闭其单击日志流将一直不停产生并进入实时计算系统。因此，对于流系统而言，数据是实时且不终止（无界）的。

- 持续（Continuous）且高效的计算

实时计算是一种事件触发的计算模式，触发源就是上述的无界流式数据。一旦有新的流数据进入实时计算，实时计算立刻发起并进行一次计算任务，因此整个实时计算是持续进行的计算。

- 流式（Streaming）且实时的数据集成

流数据触发一次实时计算的计算结果，可以被直接写入目的数据存储。例如，将计算后的报表数据直接写入RDS进行报表展示。因此流数据的计算结果可以类似流式数据源一样持续写入目的数据存储。

2. 快速入门

2.1. 登录实时计算控制台

该章节介绍了如何登录阿里云实时计算产品控制台。


前提条件

- 登录Apsara Uni-manager运营控制台前，确认您已从部署人员处获取Apsara Uni-manager运营控制台的服务域名地址。
- 推荐使用Chrome浏览器。

操作步骤

1. 在浏览器地址栏中，输入Apsara Uni-manager运营控制台的服务域名地址，按回车键。
2. 输入正确的用户名及密码。

请向运营管理员获取登录控制台的用户名和密码。

 **说明** 首次登录Apsara Uni-manager运营控制台时，需要修改登录用户名的密码，请按照提示完成密码修改。为提高安全性，密码长度必须为10~32位，且至少包含以下两种类型：

- 英文大写或小写字母（A~Z、a~z）
- 阿拉伯数字（0~9）
- 特殊符号（感叹号（!）、at（@）、井号（#）、美元符号（\$）、百分号（%）等）

3. 单击**账号登录**。
4. 在页面顶部的导航栏中，鼠标悬浮至产品，然后单击**实时计算Realtime Compute**。
5. 选择**组织和地域**。
6. 单击**Blink**。

2.2. 热词统计

2.2.1. 概述

热词统计分析在搜索热词、论坛热词、标签热词等场景有非常广泛的应用。

例如微博搜索的实时热词统计，可以方便引导用户了解微博上最新最火的热词。热词统计分析实际上就是一个简单的wordcount任务，而流式实时热词统计分析将wordcount处理逻辑整体转换为流式实时处理，可以做到实时对热词进行统计分析，并可以实时展现。

大数据的wordcount任务好比编程教学中的 `Hello World`，通常均作为新手用户必不可少的入门任务。下面就以阿里云实时计算的wordcount为例，讲解如何开发一个流式版本的wordcount。您通过wordcount任务，可以学习基本的Flink SQL语法格式，以及任务编写和发布的基本操作。


2.2.2. 代码开发

下面我们以wordcount作业为例，讲解下如何编写第一个实时计算任务。

前提条件

在进行wordcount作业开发前，请先在外部存储中创建源表stream_source（内部仅包含一个类型为string，名称为word的列）和结果表stream_result（内部包含一个类型为string，名称为word的列，另加一个类型为bigint，名称为cnt的列）。然后将这两个表注册到阿里云实时计算中。此外，已完成了项目创建，详情请参见[创建项目](#)


1. [登录实时计算控制台](#)，进入阿里云实时计算产品首页。
2. 单击头部导航栏开发页签，进入作业开发页面。
3. 右键单击用户创建的文件夹。
4. 选择新建作业，进入新建作业页面。
5. 输入作业信息：
 - 文件名称：wordcount
 - 作业类型：FLINK_STREAM/SQL
 - 存储位置：保持默认设置
6. 在作业开发窗口输入以下代码。

 **说明** 外部表的STRING类型需在实时计算作业中声明为VARCHAR。


```
create table stream_source (word varchar);
create table stream_result (word varchar, cnt bigint);
insert into
    stream_result
select
    t.word,
    count (1)
from
    stream_source t
group by
    t.word;
```

上述SQL代码讲解如下：


代码第1行，声明引用一张流式数据表，该数据表名称为stream_source。

 **说明** 如前所述，实时计算的数据驱动源来自于流式数据。因此这里的stream_source就是数据驱动源，stream_source每条（批）数据均会触发下游实时计算的一次计算。

代码第2行，声明引用一张结果表，用来存放wordcount计算结果。该数据表名称为stream_result。

 **说明** 如前所述，实时计算本身不带有任何数据存储，所有的结果数据存储理论上均为普通的RDS、Tablestore等存储系统。这里声明引用一张结果表，为计算的结果数据存储所用。

代码从第5行开始，进入正式的wordcount计算逻辑，这段SQL的含义：从stream_source表读取数据，针对每条进入的数据，统计各个word出现的次数。

 **说明** 为了尽量减少您学习实时计算的成本，阿里云实时计算提供的Flink SQL和标准SQL格式基本一致，最大限度降低您的学习门槛。

实时计算的wordcount任务和批量任务原理基本一致，仅仅是在于流式的wordcount数据源是持续且无界的，因此实时计算wordcount理论上除了您终止（Kill）作业，否则不会停止运行。

2.2.3. 代码调试

阿里云实时计算提供了强大的调试功能，能够将流式存储、静态存储和结果存储进行模拟调试，便于您验证SQL的正确性。

② 说明

- 为了避免对线上存储系统造成读写影响，实时计算调试过程需要所有的输入表提供测试数据，不允许读取线上存储系统中的数据。
- 所有写入（INSERT）操作仅输出至本地屏幕，不会对线上系统造成影响。

调试方法

1. 单击开发页面顶部的**调试**，启动调试任务。
2. 在调试页面单击**下载调试模板**并根据您的测试策略在测试模板中填写调试数据。

② 说明 实时计算对于上传的调试数据有较严格的定义：

- 调试数据文件最多支持1MB且1K条记录。
- 调试文件仅支持 UTF-8 格式。
- 采用半角逗号（,）分隔的CSV格式，避免内容出现逗号。
- 数值类型仅支持普通格式，不支持科学计数法。

3. 单击**单击上传数据**，上传调试数据。
4. 单击**确定**。
5. 在输出窗口查看调试结果。

wordcount测试模板示例

② 说明 实时计算调试文件类型为CSV，建议使用以下软件打开模板数据，并进行修改。


- Windows平台用户建议使用Excel软件。
- Mac平台用户建议使用VIM/Sublime软件，不建议使用Number软件，避免修改CSV文件时新增无关字段信息。

wordcount测试模板样例

A1						
	A	B	C	D	E	F
1	word(String)					
2	aliyun					
3	aliyun					
4	aliyun					
5						
6						

测试数据示例

您可以下载[测试数据](#)，并通过[调试](#)界面进行上传。

 **说明** PDF版本文档中 [热词统计测试数据](#)无法通过链接下载。您可以通过咨询系统管理员获取测试数据。

调试结果查看

流式数据触发实时计算的计算机制。在调试状态下，测试数据的数据源stream_source的每条数据将会直接触发一次流式计算，并输出计算结果。测试文件有3条数据，每条输出1次计算，所以结果展示页面也有3条数据，运算逻辑分别如下。

- 第1行源头数据（数据为aliyun）到达实时计算时：实时计算检测发现之前不存在aliyun单词，因此计算结果为 `<aliyun, 1>`，输出至屏幕。
- 第2行源头数据（数据为aliyun）到达实时计算时：此实时计算检测发现已经存在 `<aliyun, 1>` 的记录，因此将该值+1，得出结果为 `<aliyun, 2>`，输出至屏幕。
- 第3行源头数据（数据为aliyun）到达实时计算时：此实时计算检测发现已经存在 `<aliyun, 2>` 的记录，因此将该值+1，得出结果为 `<aliyun, 3>`，输出至屏幕。

最终观察结果以最后1条输出结果为准，即 `<aliyun, 3>` 代表本次调试数据最终输出的结果。提供另一份[测试数据](#)供您参考，便于您在不同测试数据下，观察调试界面输出情况。

2.2.4. 数据运维

代码测试完毕经验证准确无误后，即可将其发布到[数据运维](#)模块，提交任务进入实时计算集群进行生产运行。

操作步骤

1. 在开发页面单击[上线](#)按钮，弹出[上线新版本](#)窗口。
2. 在资源配置页面，单击[下一步](#)。
3. 在数据检查页面，单击[下一步](#)。
4. 在上线作业页面，单击[上线](#)。
5. 单击[运维](#)页签，在作业列表中可查看新上线的wordcount任务。
6. 单击wordcount任务对应操作列的[启动](#)按钮，弹出[启动作业](#)窗口。
7. 选择[指定数据读取时间](#)后，单击[确定](#)按钮，实时计算任务即可被生产集群调度起来。

执行结果

任务启动成功后，单击作业名称可显示作业运行信息。

- Q：既然已经在分布式实时计算集群运行起来了，为什么这个计算任务即没有流式数据输入，也没有数据输出？
- A：在定义上述 `my_source`、`my_result` 表时，并未指定因外部引用数据源类型。因此，在这类未指定具体数据源类型的情况下，实时计算将输入的Stream表视作内部随机产生字符串或者数字的随机表，同时将输出的结果表视作直接丢弃数据。

2.3. 天猫双十一大屏

2.3.1. 概述

双十一大屏是每年天猫双十一购物狂欢节的亮点应用，整个阿里云集团对外实时交易总额的披露均是通过这块大屏完成。

之前天猫双十一大屏后台流式计算使用开源的Storm来进行开发，须考虑各种异常情况、故障情况，使得整个大屏开发时间长达一个月左右。而后阿里数据事业部选择使用阿里云实时计算Flink SQL，将整个双十一大屏的开发周期缩短到三天，并且由于阿里云实时计算底层完全屏蔽了故障处理、执行优化，最终上线的双十一大屏全链路实际上比Storm作业更快、更高效。

天猫双十一大屏



2.3.2. 问题描述

双十一的大屏流式数据来源于天猫的交易订单表。为简化问题，将天猫交易数据抽象简化为如下二维表（`tmall_trade_detail`）：

字段名	类型	注释
tid	BIGINT	交易订单ID
buyer_uid	BIGINT	买家ID
seller_uid	BIGINT	卖家ID
gmtdate	TIMESTAMP	交易时间
payment	DOUBLE	订单金额

针对上述流式数据表，需要计算两个指标，包括截止到当前时间点的交易总笔数以及交易总金额。这两个数据指标经过实时计算写入在线的RDS系统，并通过大屏页面展示出来。RDS的表设计如下
(`tmall_trade_state`)：

字段名	类型	注释
gmtdate	VARCHAR(16)	交易日期
trade_count	BIGINT	交易总数
trade_sum	DOUBLE	交易总量

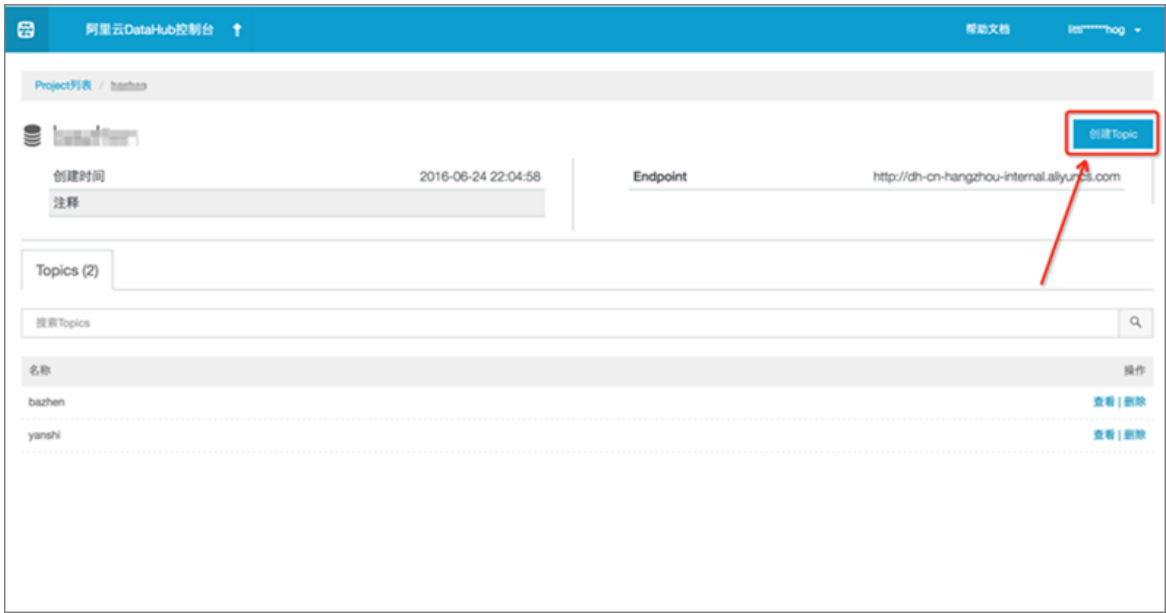
接下来，我们要做的是运用阿里云实时计算，在10分钟之内搭一套全链路双十一大屏。

2.3.3. 准备工作

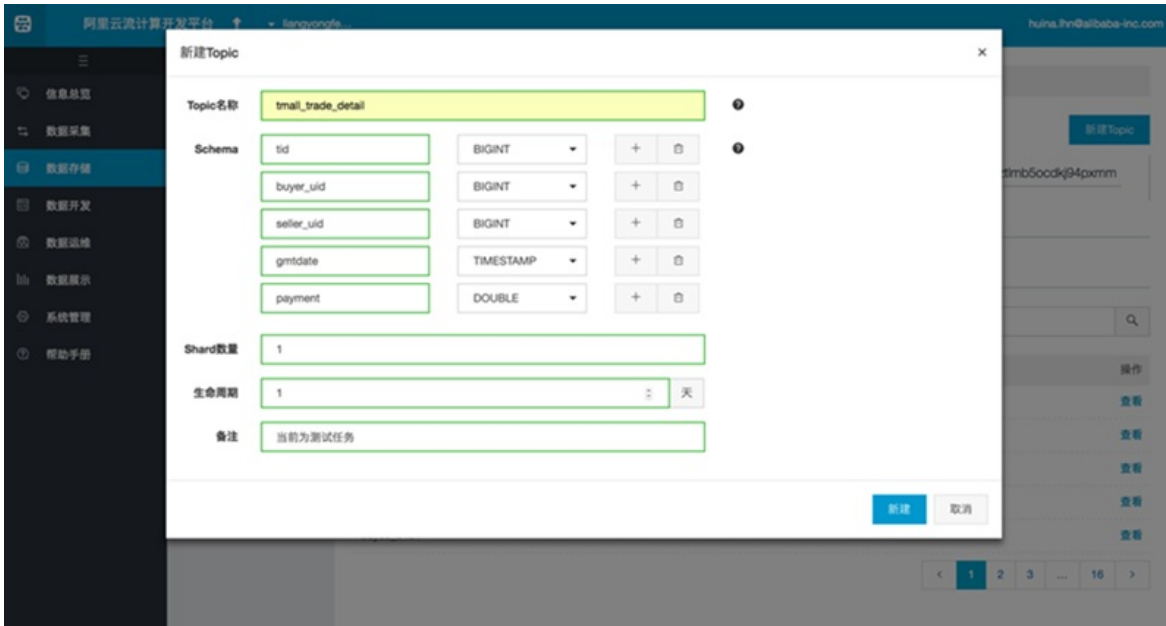
在实时计算任务编写开始前，我们需要创建上下游相关存储系统。下面为您介绍重点介绍如何注册DataHub数据存储。

创建 DataHub Topic

- 1. 登录DataHub。具体步骤请参见《实时数据分发平台DataHub用户指南》中《登录DataHub控制台》章节。
- 2. 进入DataHub WebConsole界面，单击查看进入具体的Project。
- 3. 单击创建Topic，进入Topic创建页面。



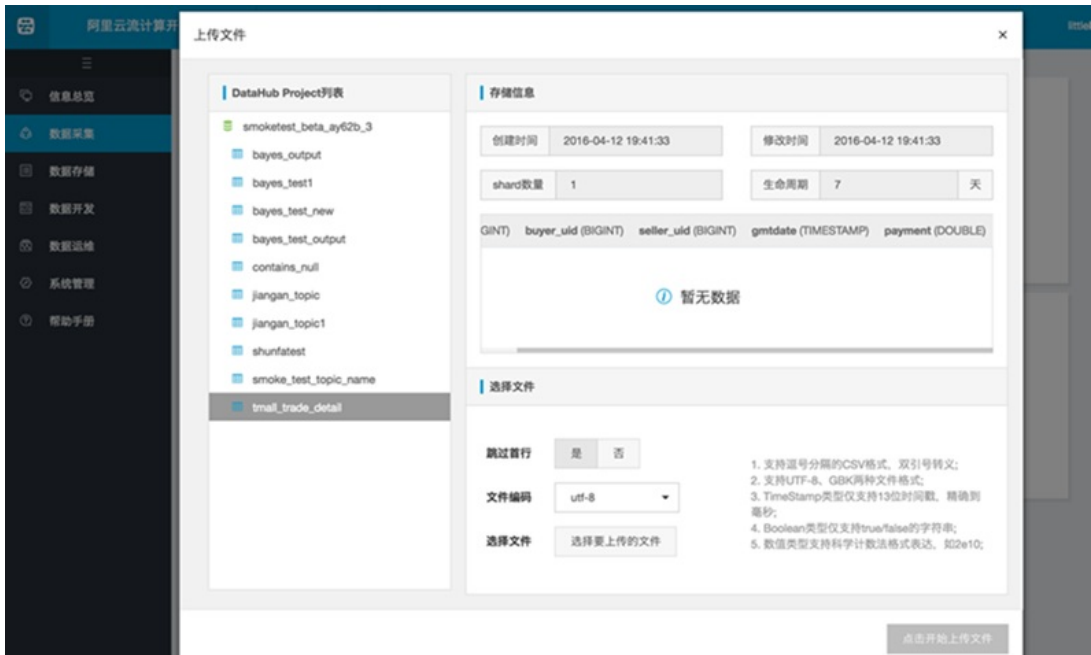
- 4. 根据上述 RDS 表结构设计表。



准备就绪，您可以开始进行Flink SQL编写了。

上传DataHub数据

进入DataHub 界面，选择上述创建完成的DataHub Topic进行数据上传。



1. 登录DataHub控制台。
2. 单击左侧导航栏数据采集。
3. 单击文件上传。
4. 双击已经创建的项目名称。
5. 单击选择文件。
6. 单击点击开始上传文件。

为了简化您的操作，提供[双十一大屏案例数据](#)，单击下载后利用 DataHub 的文件上传工具即可完成数据采

集。

2.3.4. 注册数据存储

使用阿里云实时计算内置的数据存储功能可以方便的添加 DataHub 的Topic 信息、表的创建或者数据源的引用。

操作步骤

1. 登录[实时计算控制台](#)，进入阿里云实时计算产品首页。
2. 单击头部导航栏开发页签，进入作业开发页面。
3. 单击左侧导航栏的数据存储页签。
4. 选中 DataHub数据存储文件夹。
5. 单击顶部的+注册与网络。
6. 将DataHub的Project注册到实时计算，具体参数配置请参见[注册大数据总线\(DataHub\)](#)。

选择使用 RDS（MySQL）为数据做可视化展现的存储，同时需要在阿里云实时计算注册RDS的连接信息，具体注册方法请参见[注册云数据库（RDS）](#)。

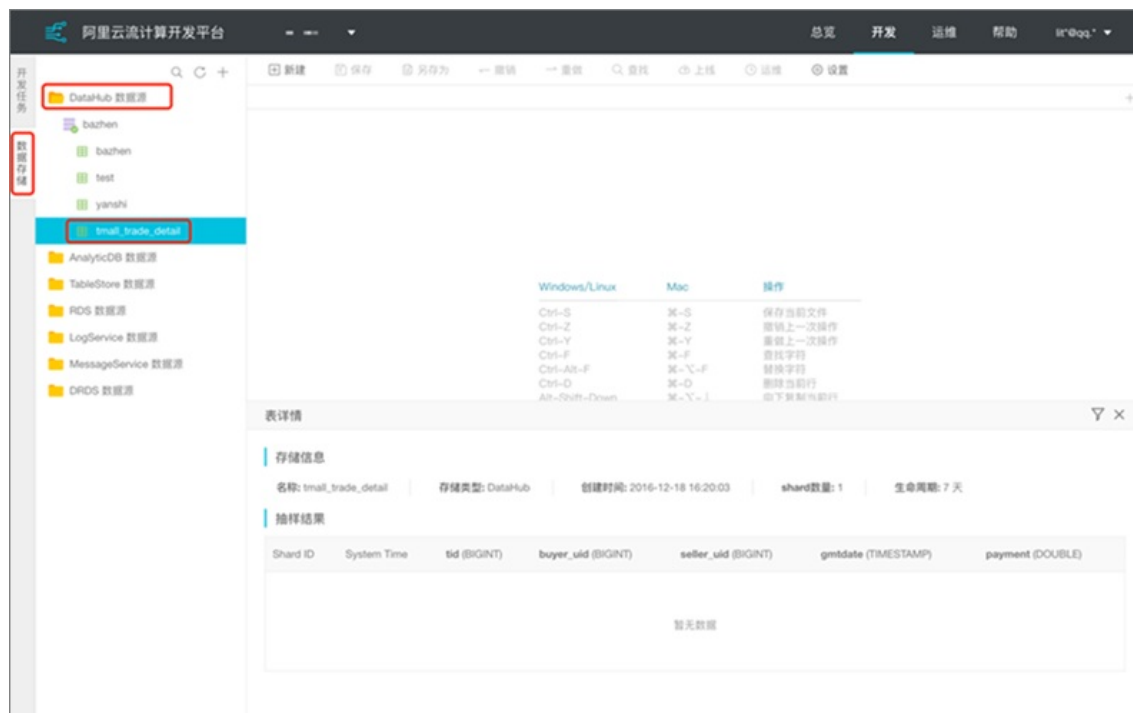
2.3.5. 开发

一旦完成数据采集工作，我们就可以专心来研究 Flink SQL 的编写工作。

1. 引用数据源。

编写 SQL 第一步，我们应该为实时计算声明数据的输入表（这里是 DataHub）以及数据的输出表（RDS）。在开发页面选择数据存储，分别找到对应的DataHub的Topic和RDS的表信息，选择性引用：

- DataHub的Topic作为我们数据输入源，单击选择作为输入表引用，实时计算将自动解析Topic的Schema，并自动添加对应的SQL到IDE：




- RDS 的表作为数据结果输出，单击选择作为结果表引用，实时计算将自动解析RDS的表Schema，并自动添加对应的SQL到IDE。

2. 编写Flink SQL。

如果我们严格按照上述教程说明的Topic /表名称进行创建工作，那么tmall_d11 任务中包含的Flink SQL已经能够直接运行，否则请根据您的实际建表情况调整tmall_d11任务中有关DataHub Topic、RDS Table名称。代码如下：

```
replace into tmall_trade_state
select
    from_unixtime(FLOOR(tmall_trade_detail.gmtdate/1000), 'yyyy-MM-dd') as gmt_date
,
    count(tid) as trade_count,
    sum(payment) as trade_sum
from
    tmall_trade_detail
group by
    from_unixtime(FLOOR(tmall_trade_detail.gmtdate/1000), 'yyyy-MM-dd');
```

 **说明** 以上表、字段信息请根据实际情况进行修改。

3. 调试Flink SQL。


提供**双十一大屏案例数据**，下载后使用调试功能上传该测试数据进行数据调试。

4. 任务上线。

调试完成后，经验证逻辑无误后，在数据开发中单击**上线任务**，您即可完成任务上线工作。上线任务操作将您的改动提交到数据运维中，可在生产环境下进行任务启动等生产运维工作。

2.3.6. 运维

在数据运维页面，选中已创建的作业任务（例如tmall_d11），启动并完成参数配置后即可启动流式任务。

 **说明** 实时计算任务启动时，需要您指定启动时间，实际上就是从源头数据存储的指定时间点开始读取数据。

指定的时间需要在上述上传时间点之前，例如，设置启动时间为一个小时之前。当前时间点是14点10分，源头数据上传时间为10分钟前，因此设置启动时间为13点整。

启动作业

启动参数

指定读取数据时间:

2016-09-08 13:00:00

代码WITH指定时间参数优先于当前选择的时间值

透明升级

启用透明升级:

可升级时间:

每日

00:27

至

00:30

Code Rolling 数据位点:

从升级前

0

天

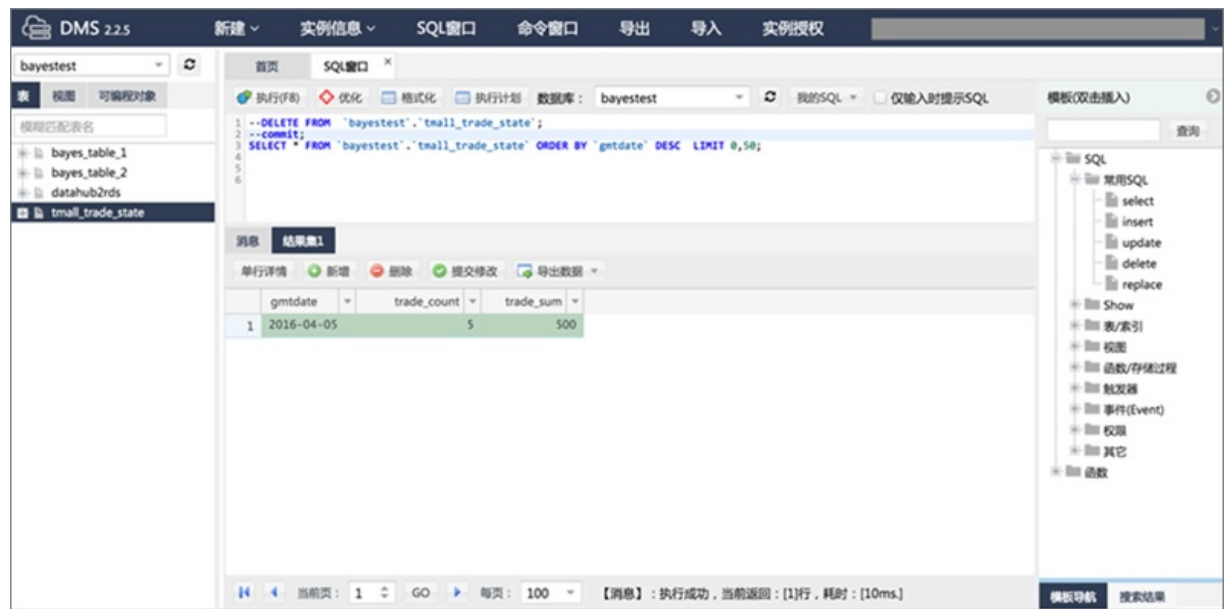
00:00

开始

确定

取消

实时计算开始运行后，既可在RDS数据库查看最终数据输出，构造数据为5笔交易，总交易额为500RMB，和最终流式计算结果一致。从端到端验证了业务代码的正确性。



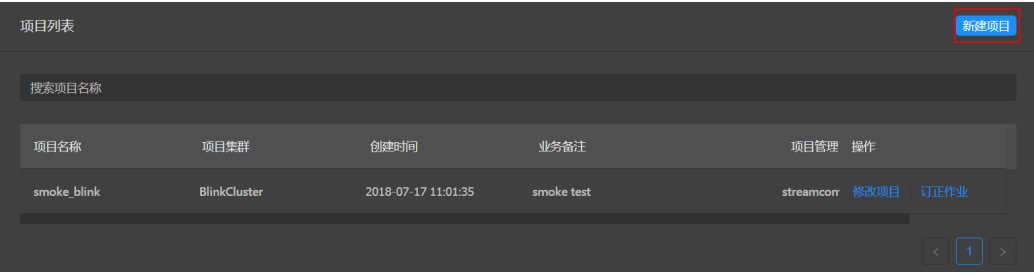
3.管理项目

本文为您介绍如何创建和搜索项目。

创建项目

- 1. 登录实时计算控制台。
- 2. 在页面顶部菜单栏上，鼠标悬浮在用户头像后，单击项目管理。
- 3. 在项目管理区域，单击右上角的新建项目。

创建项目



- 4. 填写项目配置信息。

创建项目

*

项目名称

请输入项目名称

*

项目类型

Blink项目

*

项目集群

选择项目集群

*

项目管理员

选择项目管理员

*

项目备注

请输入项目备注

GPU数量

SLOTS数量

50

创建

取消

配置项说明

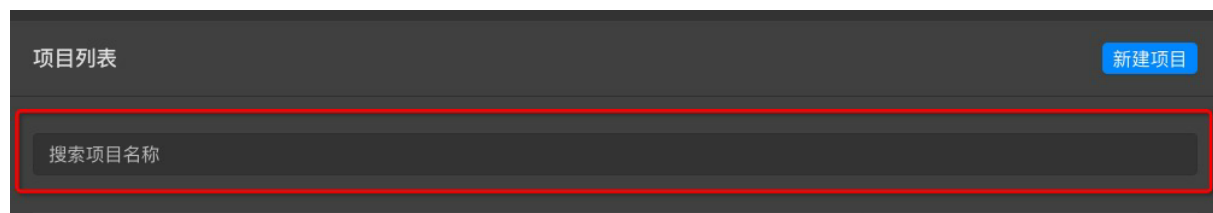
配置	说明
项目名称	输入新创建的项目名称。
项目类型	专有云默认选择Blink项目。

配置	说明
项目集群	整个项目中的作业要运行在的集群。
项目管理员	选择项目的管理员。
项目备注	为该项目添加说明信息。
GPU数量	指定该项目要占用集群的GPU的数量。
SLOTS数量	指定该项目要占用的计算单元（CU）的数量，1CU包括1核CPU加4G内存。
支持报警方式	当作业运行出现异常时，可以通过设置的报警方式得到报警信息，包括短信告警和旺旺告警。
支持作业类型	项目支持的作业类型。保持默认选项即可。
支持存储类型	项目支持的存储类型。保持默认选项即可。
最大数据存储数	可添加的数据存储数量。保持默认选项即可。
最大文件版本数	作业代码文件的版本保存数量。保持默认选项即可。
最大文件夹数	该项目中可创建的文件夹的数量。保持默认选项即可。
最大文件层级	该项目中可创建的文件夹的层级。保持默认选项即可。
最大文件数	该项目中可创建的作业代码文件的数量。保持默认选项即可。
最大资源数	用户可上传的JAR包和DICTIONARY资源的最大数量。保持默认选项即可。
最大资源引用数	用户可引用的JAR包和DICTIONARY资源的最大数量。保持默认选项即可。
监控报警	对作业是否启用监控报警功能。保持默认即可。
数据收集	对作业运行期间的数据是否进行收集。保持默认即可。
数据展示	是否开启数据显示功能。保持默认即可。
元数据	是否开启元数据显示功能。保持默认即可。
数据存储	是否开启注册数据存储功能。默认开启，保持默认即可。
数据引擎	是否开启数据引擎功能。保持默认即可。
线上日志	是否记录作业运行的日志。默认开启，保持默认即可。
资源管理	是否可以上传JAR包等资源。默认开启，保持默认即可。
版本切换	是否开启作业版本切换功能。默认开启，保持默认即可。
项目保护	是否开启项目锁定功能。保持默认即可。

5. 单击**确定**完成项目配置。

搜索项目

您可以在项目列表区域上方搜索栏中，输入项目名称的关键字或者全称来快速定位指定的项目。



4.数据存储

4.1. 概述

本小节主要介绍阿里云实时计算支持的各种外部数据存储。

4.2. VPC访问授权

实时计算访问阿里云专有网络（VPC）中的存储资源前，需要进行VPC访问授权。本文为您介绍VPC访问授权的流程。

VPC访问授权步骤

- 1. 登录实时计算控制台。
- 2. 将鼠标悬停至页面右上角账号名称。
- 3. 在下拉菜单中，单击项目管理。
- 4. 在左侧导航栏中，单击VPC访问授权。
- 5. 在VPC访问授权页面的右上角，单击新增授权。
- 6. 在授权流计算访问VPC页面，输入相应的配置参数。

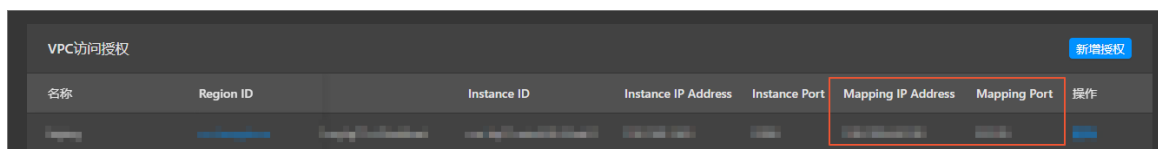
参数	说明
名称	VPC的名称。
地域	VPC中存储设备所在的区域。
VPC ID	VPC中存储设备的VPC网络ID。RDS VPC网络ID查看步骤如下： <ul style="list-style-type: none">i. 登录RDS管理控制台。ii. 在页面左上角，选择实例所在地域。iii. 单击目标实例ID。iv. 单击左侧导航栏中的数据库连接。v. 在实例连接 > 数据库连接 > 网络类型中，查看RDS的VPC ID。例如，<code>vpc-bp11ysht98wrvl9n3****</code>。
Instance ID	VPC中存储设备的实例ID。RDS中实例ID查看步骤如下： <ul style="list-style-type: none">i. 登录RDS管理控制台。ii. 在页面左上角，选择实例所在地域。iii. 单击目标实例ID，进入基本信息页面。iv. 在基本信息 > 实例ID中，查看RDS实例的ID。
Instance Port	VPC中存储设备的端口ID。

常见问题

Q：明文方式中，如何添加专有网络存储设备的URL参数？

A：在使用明文方式引用VPC中的存储时，DDL WITH参数中的URL参数值需填写VPC访问授权页面中的Mapping IP和Mapping Port参数，例如，url='jdbc:mysql://<mappingIP>:<mappingPort>/<databaseName>'。Mapping IP和Mapping Port信息查看步骤如下：

1. 登录实时计算控制台。
2. 将鼠标悬停至页面右上角账号名称。
3. 在下拉菜单中，单击项目管理。
4. 在左侧导航栏中，单击VPC访问授权。
5. 在VPC访问授权页面，查看Mapping IP和Mapping Port信息。



4.3. 数据存储概览

4.3.1. 概述

为方便您管理数据存储，通过提前注册数据存储，您能够享受到更多一站式实时计算开发平台提供的便利性。阿里云实时计算提供包括RDS、AnalyticDB for MySQL、Table Store等各类数据存储系统的管理界面。让您无需跨越多种产品的管理页面，使用阿里云实时计算平台，即可让您一站式管理您的云上数据存储。

4.3.2. 存储类别

阿里云实时计算支持的外部数据存储包括两大类：流式存储和静态存储。

流式存储

流式存储为下游实时计算提供数据驱动，同时也可以为实时计算作业提供数据输出。

流式存储

支持情况	输入	输出
DataHub	支持	支持
日志服务（LogService）	支持	支持
消息队列（MQ）	支持	支持

静态存储

静态存储为实时计算提供了数据关联查询，同时也可以作为实时计算作业数据输出。

静态存储

支持情况	维表	输出
云数据库（RDS）	支持	支持
表格存储（TableStore）	支持	支持

4.3.3. 存储使用

本文为您介绍如何在实时计算上注册和使用外部存储。

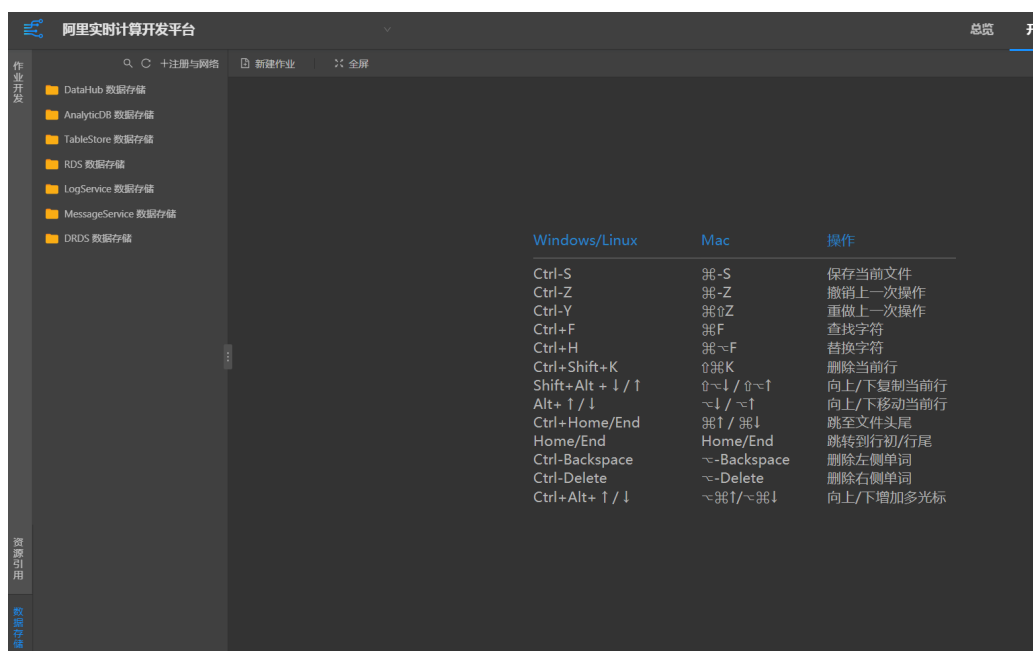
说明 对于不属于同一的主账号下资源的调用，您可以直接在DDL定义语句中写出Access ID和Access Key，此时您无法界面化操作数据源，但是作业可以直接运行。

数据注册

进入数据存储注册页面步骤如下：

1. 登录实时计算控制台。
2. 单击开发页签。
3. 单击数据存储之后，选择对应的数据存储文件夹后，再单击+注册与网络。

数据存储注册页面



4. 设置页面上出现的参数后，单击注册。

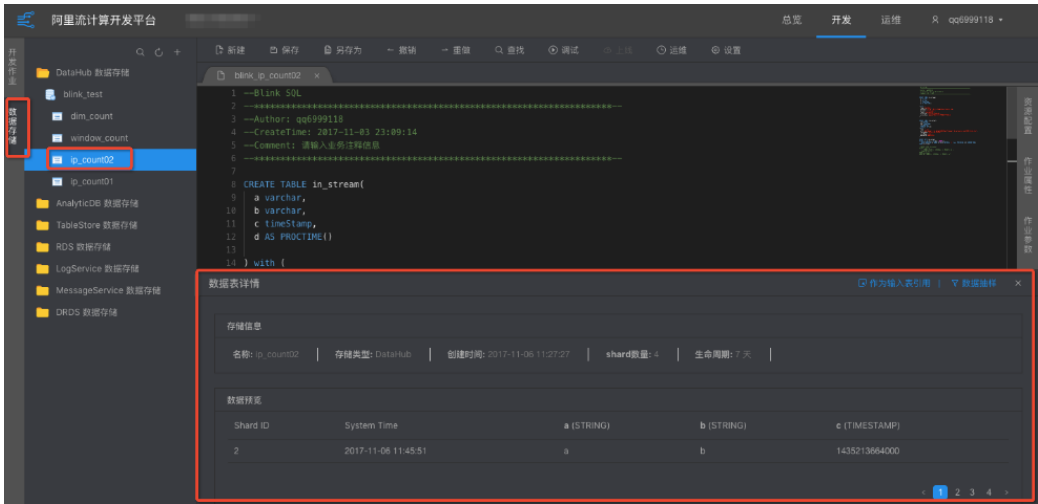
说明 实时计算数据存储功能，当前仅支持相同组织账号属主下的存储资源，不支持跨组织账号授权。

数据预览

实时计算为每个已经注册的数据存储提供了数据预览功能，单击数据存储，选择某个数据存储类型，即可预览数据。本节以DataHub为例为您介绍数据预览功能。

1. 选择数据存储 > DataHub存储。
2. 选择具体的Project和需要预览的Topic，双击即可进入查看数据存储。

数据表详情



自动生成DDL

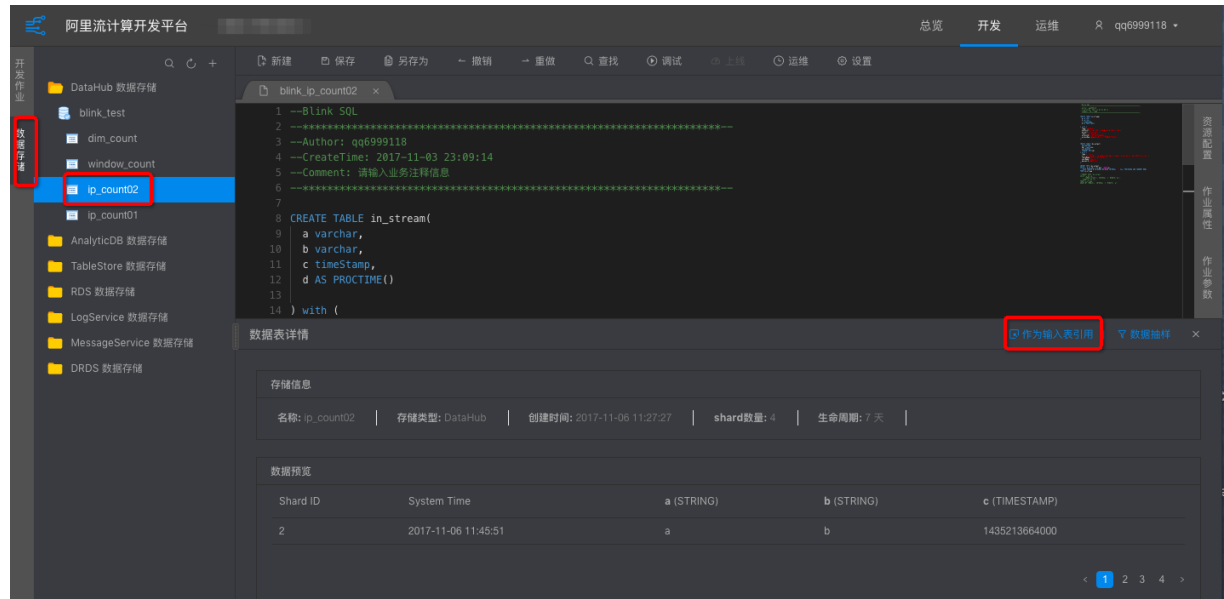
实时计算在引用外部存储时候，需要提前对于外部存储进行声明工作，例如对于声明一个流式输入引用。

```
CREATE TABLE in_stream( a varchar, b varchar, c timeStamp) with ( type='datahub', endPoint='http://dh-cn-hangzhou.aliyuncs.com', project='blink_test', topic='ip_count02', accessId='LTAIYtaf*****', accessKey='gUqyVwfkK2vfJI7jF90*****');
```

实时计算要求声明的表字段名称与源DataHub Topic保持一致，字段类型需要根据DataHub和实时计算之间的字段映射声明字段类型。实时计算提供了辅助生成DDL功能，帮助您一键生成建表DDL语句。

1. 在开发，单击左侧导航栏底部的数据存储。
2. 在数据存储区域，双击数据存储类型文件夹以及文件夹下的各节点，直至目标数据表。
3. 在数据表详情，单击作为输入表引用、作为结果表引用或作为维表引用，即可自动生成DDL。

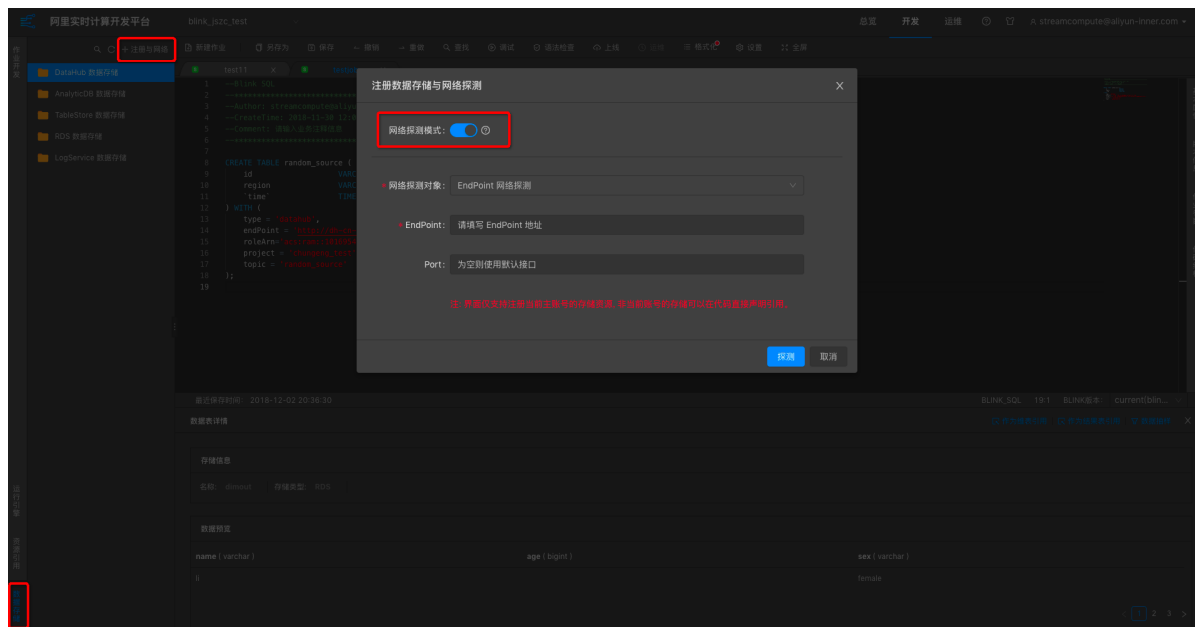
进入数据开发页面中需要编辑的任务，单击数据存储，选择作为输入表进行引用，单击作为输入表引用。此时实时计算系统会在当前光标界面生成上述DDL信息。



网络探测

实时计算的数据存储功能提供网络探测功能，用于探测实时计算产品与被探测的存储资源的网络连通性。网络探测功能开启方式如下：

1. 在开发，单击左侧导航栏底部的数据存储。
2. 在数据存储页签的右上角，单击+注册与网络。
3. 在注册数据存储与网络探测页面，打开网络探测模式开关。



举例：跨一级组织账号的资源引用

当前实时计算界面不支持跨一级组织账号数据存储注册和使用，实时计算数据存储功能仅支持同一个一级组织账号属主下的存储资源，不支持跨一级组织账号授权。如果您需要跨一级组织账号使用存储资源，可以考虑直接在DDL语句中手写外部数据引用。例如A组织用户需要使用B组织用户的资源，则可以在DDL定义如下内容。

```
CREATE TABLE in_stream( a varchar, b varchar, c timeStamp) with ( type='datahub', endPoint='http://dh-cn-hangzhou.aliyuncs.com', project='blink_test', topic='ip_count02', accessId='B用户授权的AccessId', accessKey='B用户授权的AccessKey');
```

4.4. 注册大数据总线（DataHub）

阿里云流式数据服务DataHub是流式数据（Streaming Data）的处理平台，提供对流式数据的发布（Publish），订阅（Subscribe）和分发功能，让您可以轻松构建基于流式数据的分析和应用。阿里云实时计算通常使用DataHub作为流式数据源头和输出目的端。

注册

1. 登录实时计算控制台。
2. 单击顶部导航栏开发页签，进入数据开发页面。
3. 单击左侧菜单栏的数据存储页签。
4. 选择Dat aHub数据存储后单击右键，在弹出的菜单中单击注册数据存储，将Dat aHub的Project注册到阿里云实时计算。

配置说明

配置	说明
网络探测模式	对于已支持注册存储的数据源，将自动进行网络探测；对于暂不支持注册存储的数据源，您可以使用网络探测检测它们的连通性。
数据存储类型	默认选择DataHub数据存储。
Endpoint	<p>填写DataHub的Endpoint，不同的地域下DataHub有不同的Project。如需了解更多Endpoint情况，请联系您的管理员。</p> <p> 说明 有关专有云的Endpoint填写，请联系您的专有云系统管理员咨询有关DataHub Endpoint的地址。</p>
Project	<p>填写DataHub的Project名称。</p> <p> 说明 跨一级部门属主的数据存储不能注册。例如A部门用户拥有DataHub的ProjectA，但B部门用户希望在实时计算使用ProjectA，目前实时计算暂不支持这类使用情况。</p>
AccessKey ID	填写当前账户的AccessKey ID。
AccessKey Secret	填写当前账户的AccessKey Secret，以便实时计算可以访问DataHub的Project。

使用

由于DataHub本身是流数据存储，实时计算只能将其作为流式数据输入和输出，无法作为维表引用。

常见问题

Q: 为什么我注册失败？

A: 实时计算的数据存储页面仅提供协助您完成数据管理，其本身就是使用相关存储SDK代为访问各类存储。因此很多情况下可能是您注册过程出现疏忽导致，请排查如下原因：

- 请确认是否已经开通并拥有DataHub的Project。请登录DataHub控制台，您可以访问DataHub控制台看您是否有权限访问您的Project。
- 请确认您是DataHub Project的属主，特别注意，跨一级部门属主的数据存储不能注册。例如A部门用户拥有DataHub的ProjectA，但B部门用户希望在实时计算使用ProjectA，目前实时计算暂不支持这类使用情况。
- 请确认您填写的DataHub的Endpoint和Project完全正确。
- 请确认您填写的DataHub Endpoint是经典网络地址，而非VPC地址。目前实时计算暂不支持VPC内部地址。
- 不要重复注册，实时计算提供注册检测机制，避免您重复注册。

Q: 为什么数据抽样仅仅针对时间抽样，不支持其他字段抽样吗？

A: DataHub定位是流数据存储，对外提供的接口也仅仅只有时间参数，因此实时计算也只能提供基于时间的抽样。

4.5. 注册日志服务（Log Service）

日志服务Log Service（原SLS）是针对日志场景的一站式解决方案，解决海量日志数据采集/订阅、转储与查询功能。日志服务是阿里云提供的一整套非常优秀的日志管理平台，在您直接使用日志服务进行ECS日志管理的情况下，实时计算可以直接对接日志服务的Log Service存储，避免您进行数据搬运。

注册

1. [登录实时计算控制台](#)。
2. 单击头部导航栏开发页签，进入数据开发页面。
3. 单击左侧菜单栏的数据存储页签。
4. 选择 **Log Service数据存储**后单击右键，在弹出的菜单中单击**注册数据存储**，将Log Service的Project注册到阿里云实时计算。

配置说明

配置	说明
网络探测模式	对于已支持注册存储的数据源，将自动进行网络探测；对于暂不支持注册存储的数据源，您可以使用网络探测检测它们的连通性。
数据存储类型	默认选择Log Service 数据存储。
Endpoint	填写Log Service的Endpoint，不同的地域下Log Service有不同的Endpoint。 <div>❓ 说明 有关专有云Log Service的Endpoint填写，请联系专有云系统管理员。</div>
Project	填写Log Service的Project名称。 <div>❓ 说明 跨一级组织账号的数据存储不能注册。例如A部门用户拥有Log Service的Project A，但B部门用户希望在实时计算使用Project A，目前实时计算暂不支持这类使用情况。</div>
AccessKey ID	填写当前账户的AccessKey ID。
AccessKey Secret	填写当前账户的AccessKey Secret，以便实时计算可以访问Log Service的Project。

使用场景

由于日志服务本身是流数据存储，实时计算只能将其作为流式数据输入和输出，无法作为维表引用。


常见问题

- Q：为什么注册失败？
A：实时计算的数据存储页面仅协助您完成数据管理，其本身就是使用相关存储SDK代为访问各类存储。因此很多情况下可能是您注册过程出现疏忽导致，请排查如下原因：

- 请确认是否已经创建并拥有日志服务的Project。请登录日志服务控制台，您可以访问日志服务控制台看您是否有权访问您的Project。
- 请确认您是日志服务 Project 的属主，特别注意，跨属主的数据存储不能注册。例如A用户拥有日志服务的ProjectA，但B用户希望在实时计算使用ProjectA，目前实时计算暂不支持这类使用情况。
- 请确认您填写的日志服务的Endpoint和Project完全正确。

 **说明** 日志服务Endpoint必须以http开头，且不能以/结尾，例如 `http://cn-hangzhou.log.aliyuncs.com` 是正确的，但 `http://cn-hangzhou.log.aliyuncs.com/` 是错误的。

- 不要重复注册，实时计算提供注册检测机制，避免您重复注册。
- Q：为什么数据抽样仅仅针对时间抽样，不支持其他字段抽样吗？
A：日志服务定位是流数据存储，对外提供的接口也仅仅只有时间参数，因此实时计算也只能提供基于时间的抽样。

 **说明** 如果希望使用日志服务的检索功能，请登录日志服务控制台使用检索。

4.6. 注册表格存储（Tablestore）

表格存储（Tablestore）是构建在阿里云飞天分布式系统之上的NoSQL数据存储服务，提供海量结构化数据的存储和实时访问。Tablestore具备海量存储和低延迟响应的特点，适合给实时计算作为维表和结果表。

注册

1. [登录实时计算控制台](#)。
2. 单击头部导航栏开发页签，进入数据开发页面。
3. 单击左侧菜单栏的数据存储页签。
4. 选择Tablestore数据存储后单击右键，在弹出的菜单中单击注册数据存储，将Tablestore的Instance注册到阿里云实时计算。

配置说明

配置	说明
网络探测模式	对于已支持注册存储的数据源，将自动进行网络探测。对于暂不支持注册存储的数据源，您可以使用网络探测检测它们的连通性。
数据存储类型	默认选择Tablestore数据存储。
Endpoint	填写Tablestore instance的Endpoint，请进入Tablestore控制台查看Tablestore instance的Endpoint信息（需要填写Tablestore内网地址）。
实例名称	填写Tablestore的实例名称。
AccessKey ID	填写当前账户的AccessKey ID。
AccessKey Secret	填写当前账户的AccessKey Secret，以便实时计算可以访问Tablestore的实例。


4.7. 注册云数据库（RDS）

本小节主要介绍RDS在阿里云实时计算中的注册和使用。

RDS简介

云数据库RDS（ApsaraDB for RDS，简称RDS）是一种稳定可靠、可弹性伸缩的在线数据库服务。基于飞天分布式系统和SSD盘高性能存储，支持MySQL、PostgreSQL和PPAS（高度兼容Oracle）引擎。目前实时计算支持的RDS引擎包括MySQL和PostgreSQL。

通常受限于关系型模型，RDS对于大容量高并发请求的支持不如Tablestore。因此实时计算和RDS搭配使用时更多将RDS作为结果表，但对于小批量低并发的数据情况，实时计算仍然可以使用RDS作为维表。

 **说明** 在实时计算中，下游数据库使用云数据库MySQL版等关系数据库（对应的connector为DRDS和RDS），当实时计算频繁写某个DRDS或RDS表时，存在死锁风险。例如在高QPS/TPS或高并发写入情况场景下，不建议使用DRDS或者RDS作为Blinkjob的结果表，建议使用Table Store作为结果表来解决死锁的问题。

注册RDS

1. [登录实时计算控制台](#)。
2. 单击顶部导航栏开发页签，进入数据开发页面。
3. 单击左侧菜单栏的数据存储页签。
4. 选择RDS数据存储后单击右键，在弹出的菜单中单击注册数据存储，将RDS的Instance注册到阿里云实时计算。

配置说明

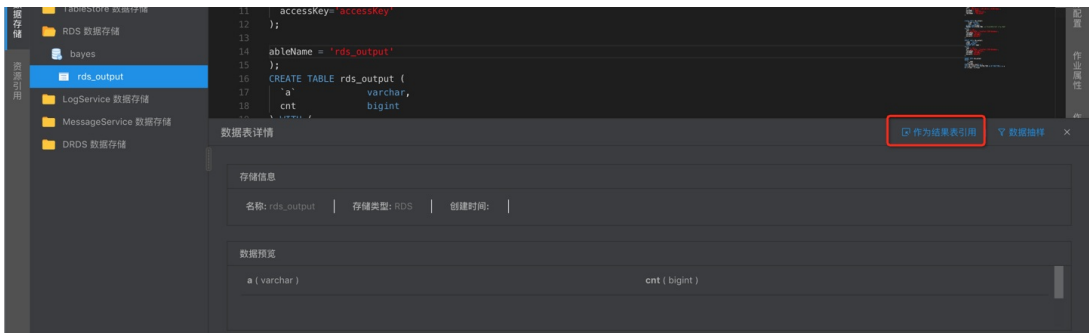
配置	说明
网络探测模式	对于已支持注册存储的数据源，将自动进行网络探测。对于暂不支持注册存储的数据源，您可以使用网络探测检测它们的连通性。
数据存储类型	默认选择RDS数据存储。
URL连接	填写数据库的连接URL。
DBName	<p>填写连接的数据库名称。</p> <div> 说明 DBName是RDS的数据库名称，非Instance的名称。</div> <p>目前RDS使用白名单进行安全保证，需要将实时计算控制台和执行节点的IP地址加入RDS的白名单，否则实时计算可能无法连接RDS，详情请参见设置白名单。</p>
User Name	数据库登录账号。
Password	数据库登录密码。

配置	说明
引擎选择	RDS数据库的类型。支持三种数据库类型： <ul style="list-style-type: none">mysqlpostgresqlsqlserver

引用为结果表

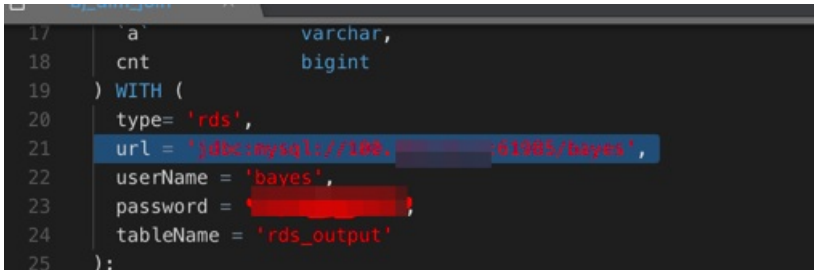
RDS注册成功后，双击该RDS名称，然后选择要作为结果表的表，单击作为结果表引用，将其引用为结果表。

作为结果表引用



引用后，阿里云实时计算会在当前界面自动生成上述DDL信息。

最后结果



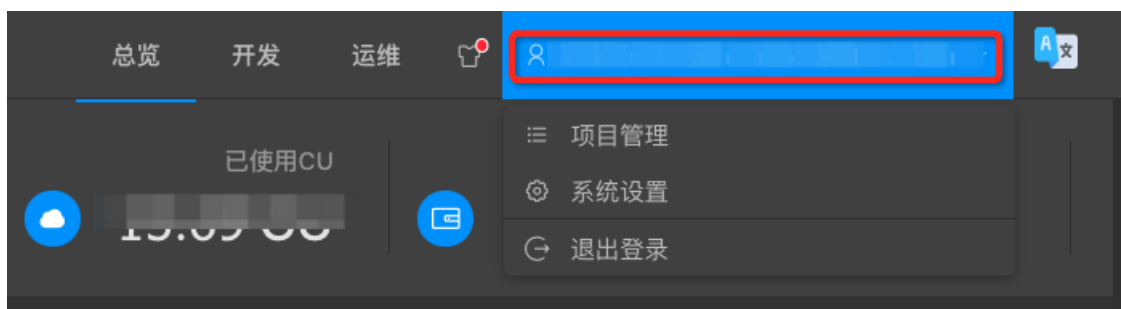
如果使用数据存储的方式发生以下报错情况。

异常信息图



引起该报错的原因是您在创建RDS实例时选择的不是经典网络，而是专有网络（VPC模式），对于这种实例请进行授权操作，步骤如下：

1. 将鼠标悬停在管理员图标处。



2. 单击系统设置。
3. 在左侧导航栏中单击VPC访问授权。
4. 单击新增授权，进入授权流计算访问VPC页面。

授权

授权流计算访问VPC

* 名称:

请输入VPC名称

* 地域:

* VPC ID:

请输入VPC ID

* Instance ID:

请输入Instance ID

* Instance Port:

请输入Instance Port

授权

取消

配置说明

配置

说明

名称

输入VPC的名称。

地域

RDS所在的地域。

VPC ID

输入VPC的ID。

Instance ID

数据库实例的ID。请登录到RDS控制台查看RDS的Instance ID地址。

实例信息

实例ID (实例名称)

部门

项目

区域

实例类型

数据库类型

网络类型

IP地址

最大可用空间 (GB)

最大使用内存 (GB)

CPU

状态

创建时间

操作

实例ID (实例名称)

NAS_DVCL

NAS_DVCL

cn-qdandshu-qd1

主实例

MySQL 5.6

经典网络

192.168.1.100

250

16,384

2

运行中

2018/7/4 7:56:35:34

Instance Port

数据库实例的访问端口。请登录到RDS控制台，然后单击目标实例名称，在实例基本信息部分获取内网端口号。

5. 注册RDS数据仓储，填写注册信息。

（如果使用的不是当前一级部门账号的存储资源）跨属主的数据存储不能注册。例如A部门用户拥有RDS的实例A，但B部门用户希望在实时计算使用实例A，目前这种情况实时计算暂不支持数据存储的方式注册，需要使用明文的方式将数据库注入进作业中。

说明 如果使用一级部门账号下的存储资源，则不建议使用明文的方式使用RDS。

B用户需要将作业with参数中的url、userName、password、tableName参数按照实例A的信息填写。

配置信息

```
91  
92 CREATE TABLE datahub_output (  
93     id          varchar  
94 ) WITH (  
95     type= 'rds',  
96     url = 'jdbc:mysql://  
97     userName = 'root@  
98     password = '8%gty2203',  
99     tableName = 'datahub_output'  
100 );  
101
```

使用明文的方式需要用户进行白名单设置。

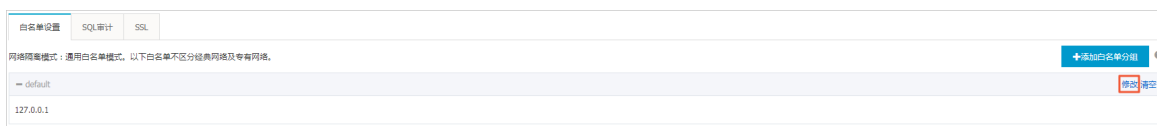
设置白名单

部分数据存储为了自身的安全需求，提供白名单机制，即仅允许用户指定的白名单IP地址访问RDS，不在白名单中的IP将无法访问RDS，包括阿里云产品。但同样机制也会拦截其他阿里云产品写入数据存储的需求。以RDS为例，新建的RDS数据库是完全拒绝任何外部连接的，此时必须要添加访问白名单，即加入访问者的IP才能够访问RDS。

RDS支持访问内网地址和外网地址。针对RDS，您只需要补充实时计算白名单网段地址即可。

操作步骤如下：

1. 登录RDS管理控制台。
2. 在实例列表页面，单击实例名称栏中目标实例的ID。
3. 在左侧导航栏中，单击数据安全性。
4. 在白名单设置页签中，单击default白名单分组中的修改。



说明

- 若需要ECS实例通过内网地址连接到RDS，请确保两者处于同一地域内且网络类型相同，否则设置了白名单也无法连接成功。
- 您也可以单击添加白名单分组新建自定义分组。

5. 在修改白名单分组对话框中，填写需要访问该实例的IP地址或IP段，然后单击确定。

修改白名单分组

*分组名称:

default

*组内白名单:

192.168.0.44

加载ECS内网IP

还可添加999个白名单

指定IP地址：192.168.0.1 允许192.168.0.1的IP地址访问RDS
指定IP段：192.168.0.0/24 允许从192.168.0.1到192.168.0.255的IP地址访问RDS
多个IP设置，用英文逗号隔开，如192.168.0.1,192.168.0.0/24

新白名单将于1分钟后生效

确定

取消

- 填写IP段，例如10.10.10.0/24，则表示10.10.10.X的IP地址都可以访问该RDS实例。
- 添加多个IP地址或IP段，请用英文逗号隔开（逗号前后都不能有空格），例如192.168.0.1,172.16.213.9。
- 单击加载ECS内网IP后，将显示您当前阿里云账号下所有ECS实例的IP地址，可快速添加ECS内网IP地址到白名单中。

 说明 当您在default分组中添加新的IP地址或IP段后，默认地址127.0.0.1会被自动删除。

常见问题

- 问题描述

运行中报错异常栈，如图所示。

```
com.alibaba.blink.streaming.connectors.common.exception.BlinkRuntimeException: 821102

系统错误-内部错误：无法连接到RDS
    at com.alibaba.blink.connectors.rds.RdsExceptionUtil.getBlinkException(RdsExceptionUtil.java:33)
    at com.alibaba.blink.connectors.rds.RdsOutputFormat.open(RdsOutputFormat.java:248)
    at com.alibaba.blink.streaming.connectors.common.output.TupleOutputFormatAdapterSink.open(TupleOutputFormatAdapterSink.java:50)
    at org.apache.flink.api.common.functions.util.FunctionUtils.openFunction(FunctionUtils.java:36)
    at org.apache.flink.streaming.api.operators.AbstractUdfStreamOperator.open(AbstractUdfStreamOperator.java:156)
    at org.apache.flink.streaming.runtime.tasks.StreamTask.openAllOperators(StreamTask.java:445)
    at org.apache.flink.streaming.runtime.tasks.StreamTask.invoke(StreamTask.java:297)
    at org.apache.flink.runtime.taskmanager.Task.run(Task.java:758)
    at java.lang.Thread.run(Thread.java:834)
Caused by: com.mysql.jdbc.exceptions.jdbc4.CommunicationsException: Communications link failure

The last packet sent successfully to the server was 0 milliseconds ago. The driver has not received any packets from the server.
    at sun.reflect.NativeConstructorAccessorImpl.newInstance0(Native Method)
    at sun.reflect.NativeConstructorAccessorImpl.newInstance(NativeConstructorAccessorImpl.java:62)
    at sun.reflect.DelegatingConstructorAccessorImpl.newInstance(DelegatingConstructorAccessorImpl.java:45)
    at java.lang.reflect.Constructor.newInstance(Constructor.java:423)
    at com.mysql.jdbc.Util.handleNewInstance(Util.java:425)
    at com.mysql.jdbc.SQLException.createCommunicationsException(SQLException.java:989)
    at com.mysql.jdbc.MysqlIO.<init>(MysqlIO.java:341)
    at com.mysql.jdbc.ConnectionImpl.coreConnect(ConnectionImpl.java:2251)
    at com.mysql.jdbc.ConnectionImpl.connectOneTryOnly(ConnectionImpl.java:2284)
    at com.mysql.jdbc.ConnectionImpl.createNewIO(ConnectionImpl.java:2083)
    at com.mysql.jdbc.ConnectionImpl.<init>(ConnectionImpl.java:806)
    at com.mysql.jdbc.JDBC4Connection.<init>(JDBC4Connection.java:47)
    at sun.reflect.NativeConstructorAccessorImpl.newInstance0(Native Method)
    at sun.reflect.NativeConstructorAccessorImpl.newInstance(NativeConstructorAccessorImpl.java:62)
    at sun.reflect.DelegatingConstructorAccessorImpl.newInstance(DelegatingConstructorAccessorImpl.java:45)
    at java.lang.reflect.Constructor.newInstance(Constructor.java:423)
```

- 解决方案

将自己所在区域的IP添加至白名单，具体步骤请参见[设置白名单](#)。

5. 数据开发

5.1. 创建作业

本节介绍如何创建实时计算作业。

操作步骤

1. 登录实时计算控制台。
2. 单击顶部菜单栏开发页签，进入数据开发页面。
3. 在顶部菜单栏中单击新建作业。
4. 在新建作业对话框中配置相关参数。

配置	说明
文件名称	输入新作业的名称，需以字母开头，只能包含小写英文字母、数字、下划线（_），长度限制为3~64个字符。
作业类型	可以创建2种作业类型，FLINK_STREAM/SQL或FLINK_STREAM/DATASTREAM。
存储位置	在文件夹目录中，指定该作业的代码文件所属的文件夹。您还可以单击现有文件夹右侧的图标，新建子文件夹。

5. 单击确定完成作业配置。

5.2. 开发阶段

5.2.1. SQL辅助

数据开发提供了一套完整的在线SQL IDE工具，支持如下功能辅助您进行Flink SQL开发。

- Flink SQL语法检查

您在修改IDE文本后即可进行自动保存，保存操作可以触发SQL语法检查功能。语法校验出错误后，将在IDE界面提示出错行数、列数以及错误原因。

- Flink SQL智能提示

您在输入Flink SQL过程中，IDE提供包括关键字、内置函数、表/字段智能记忆等提示功能。

- Flink SQL语法高亮

针对Flink SQL关键字，提供不同颜色的语法高亮功能，以区分Flink SQL不同结构。

5.2.2. SQL版本管理

数据开发涵盖了日常开发工作的关键领域，包括代码辅助、代码版本，数据开发提供了一个代码版本管理功能。您每次提交即可生成一个代码版本，代码版本为追踪修改以及日后回滚所用。

- 版本管理

数据开发为您提供代码版本管理功能。每提交一次作业即可生成一个代码版本。代码版本用于版本追踪、版本修改以及后期版本回滚。

在开发页面右侧的版本信息页面，单击操作 > 更多，可以选择相应的版本管理功能：

- 对比：查看最新代码和指定版本的差异。
- 回滚：使用回滚功能回滚到指定版本。
- 删除：删除旧版本作业。
- 锁定：锁定当前作业版本。

❓ 说明 解锁前无法提交新版本。

● 版本清理

您每次提交一个作业，发布线上，实时计算均会生成一份代码快照用于日后的代码追踪使用。实时计算为您设定了版本最大上限值，专有云默认是20个版本（其他环境请咨询实时计算系统管理员）。如果生成的版本超过最大值，则系统将不允许提交，报错提示您需要删除部分旧版本作业。

在作业开发页面右侧导航栏，单击版本信息 > 操作 > 更多 > 删除，删除过期且业务不需要的版本后，即可再次进行上线作业操作。

代码对比



5.2.3. 数据存储管理

开发页面提供了一整套数据存储管理的便捷工具，您通过开发页面注册数据源，即可享受到多种便利的数据存储服务，如下所示。

● 数据预览

数据开发页面中，为各类数据存储类型提供数据预览功能。使用数据预览可以有效辅助用户洞察上下游数据特征，识别关键业务逻辑，快速完成业务开发工作。

● DDL辅助生成

实时计算支持自动生成DDL语句引用外部数据存储。实时计算为您提供辅助生成DDL的功能，减少您编写流式任务的复杂度，有效降低编写SQL的错误率，并最终提高业务产出效率。

5.3. 调试阶段

作业开发模块为您提供了一套模拟的运行环境，您可以在调试环境中自定义上传数据，模拟运行，检查输出结果。

当您写完所有的业务逻辑，接下来的操作步骤如下所示。

1. [登录实时计算控制台](#)，进入阿里实时计算产品首页。
2. 单击头部导航栏开发页签，进入数据开发页面。
3. 在左侧导航栏选择作业开发。
4. 在作业开发区域，双击文件夹或作业名称，打开目标作业。
5. 在顶部菜单栏中单击语法检查。

 **说明** 语法检查功能可以检测SQL中是否存在语法错误并显示相应的错误提示。

6. 在顶部菜单栏中单击**调试**进入**调试作业**页面进行作业调试。

为方便您调试作业，实时计算支持以下2种调试模式。

- o 本地上传方式
 - a. 单击**下载模板**。
 - b. 根据模板，准备数据。
 - c. **数据上传**。完成后可在**数据预览**界面查看已上传数据。
- o 线上抽样方式
 - a. 单击**随机抽样线上数据**或**顺序抽样线上数据**。
 - b. 抽样成功后可在**数据预览**界面查看抽样数据。

7. 单击**确定**，启动调试。

8. 在弹出的调试结果输出窗口，查看调试输出结果。

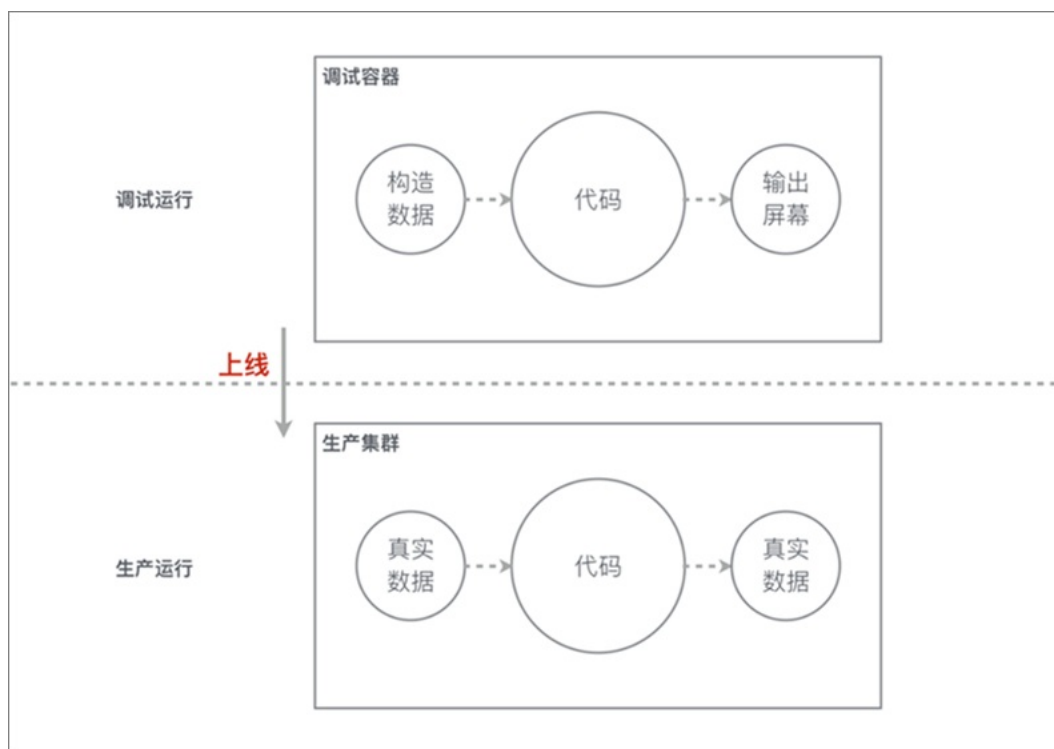
在该环境运行Flink SQL可以实现以下功能。

- 生产完全隔离

调试环境下，所有的Flink SQL运行将在独立的调试容器运行，且所有的输出将被直接改写为调试结果屏幕，不会对线上生产实时计算作业、线上生产的数据存储系统造成任何影响，让您可以放心大胆运行任务。

数据调试实际上不会真正写入到外部数据源，而是被实时计算拦截输出到屏幕，因此在实时计算调试完成的代码是在调试容器中完成，真正线上运行过程中可能由于对目标数据源写入格式导致运行失败。这类错误调试阶段无法完全规避，只能到线上运行才能发现。假设您的结果数据输出到RDS系统，其中某些字段输出字符串数据长度大于RDS建表最大值，在Debug环境下我们无法测试出该类问题，但实际生产运行过程中会有引发异常。后续，实时计算将提供针对本地调试运行也支持写出到真实数据源的功能，届时可以有效辅助用户缩短调试和生产的差距，尽可能在调试阶段解决问题。

生产调试



- 支持构造测试数据

调试环境下，所有的Flink SQL运行均不会从源头数据存储系统读取数据，包括DataHub的流式输入、RDS等维表输入，调试作业均不会读取。调试环境要求您必须进行自己构建测试数据集，并将测试数据上传到数据开发。

实时计算针对不同任务提供测试数据模板，您完全可以下载数据模板开始直接填写构造数据，不必担心测试数据构造困难。

说明 强烈建议您使用下载的数据模板构造数据，以避免报错。

- 调试分隔符

默认情况下，调试文件使用逗号作为分隔符，例如您构造了如下的测试文件。

```
id,name,age
1,alicloud,13
2,stream,1
```

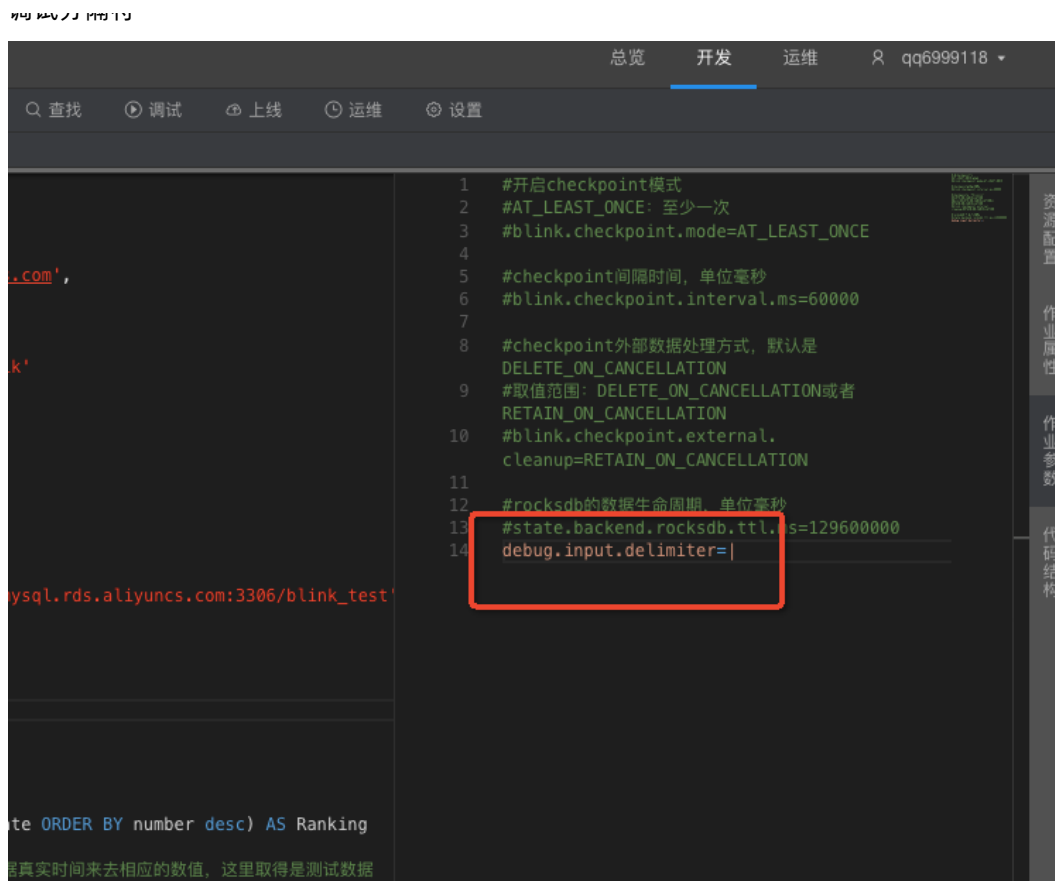
在不指定调试分隔符情况下，默认使用了逗号进行分割。但一旦您需要使用JSON作为字段内容，字段内容即包含了逗号，此时您需要指定分割符为其他字符。

说明 实时计算仅支持指定单个英文字符为分隔符，不允许字符串，例如不允许aaa作为分隔符。

```
id|name|age
1|alicloud|13
2|stream|1
```

此时您需要针对该数据存储的作业参数设置 `debug.input.delimiter=|`。

调试分隔符



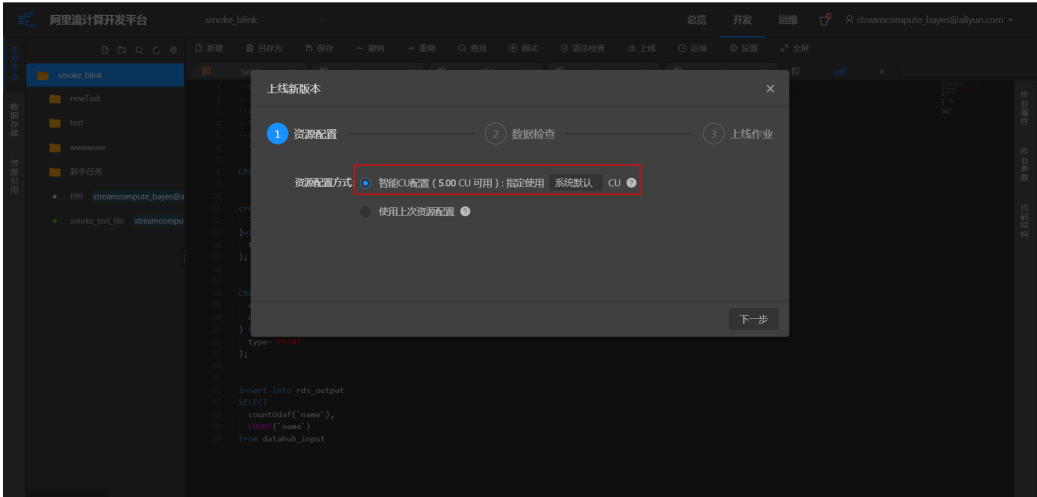
5.4. 上线阶段

当您完成开发、调试，验证Flink SQL正确无误后，就可以将其上线，使作业在生产环境中运行。

操作步骤

1. [登录实时计算控制台](#)，进入阿里云实时计算产品首页。
2. 单击头部导航栏开发页签，进入作业开发页面。
3. 在菜单栏中单击上线。
4. 选择智能CU配置，第一次不需要指定CU数量，直接用系统默认配置即可。单击下一步。

资源配置



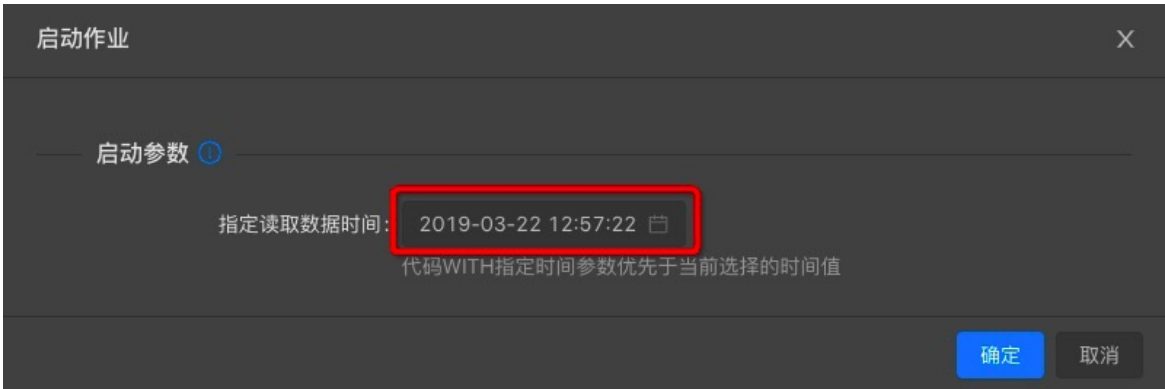
5. 检查数据。完成后，单击下一步。
6. 单击上线完成作业的上线。
7. 进入运维页面，启动作业。
 - i. 单击页面顶部的运维。
 - ii. 在运维，单击目标作业操作列下的启动。

5.5. 启动作业

完成作业开发和作业上线后，您可以在运维页面启动作业。

操作步骤

1. 登录实时计算控制台。
2. 单击页面顶部的运维。
3. 单击目标作业操作列下的启动。
4. 在启动作业页面，单击指定数据读取数据时间（即指定启动位点）文本框。



5. 指定读取数据时间（启动位点），单击确定，完成作业启动。

启动位点表示从数据源表中读取数据的时间点：

 - 选择当前时间：表示从当前时间开始读取数据。
 - 选择历史时间：表示从历史时间点开始读取数据，通常用于回追历史数据。

5.6. 暂停作业

修改资源配置后，可以经过暂停和恢复的步骤使变更生效。本文为您介绍如何暂停作业。


背景信息

注意

- 只能对运行状态为运行的作业进行暂停操作。
- 暂停操作不会清除任务状态，即如果有COUNT操作，作业暂停 > 恢复后，COUNT会从上上次成功Checkpoint的状态开始继续计算。
- 实时计算3.5.0以上版本才可以使用暂停（Checkpoint）功能，否则右上角会出现报错：发生错误系统错误：BLINK版本异常。错误原因：blink version >= blink-3.5 is required, instance blink-3.4.4。

操作步骤

1. 登录实时计算控制台。
2. 单击页面顶部的运维。
3. 在运维页面，单击目标作业操作列下的暂停。

 说明 更多目录的暂停（Checkpoint）在进行暂停作业的同时，会主动触发一次Checkpoint，因此暂停（Checkpoint）所消耗的时间可能会比暂停的时间长。

5.7. 停止作业

更改SQL逻辑、更改作业版本、增加WITH参数或增加作业参数后，经过停止和启动的步骤，才能使变更生效。本文为您介绍如何停止作业。

注意

- 只能对运行状态为运行或启动中的作业进行停止操作。
- 停止操作会清除任务状态，即如果有COUNT操作，作业停止 > 启动后，COUNT从0开始计算。
- 实时计算3.5.0以上版本才可以使用停止（checkpoint）功能，否则右上角会出现报错：发生错误系统错误：BLINK版本异常。错误原因：blink version >= blink-3.5 is required, instance blink-3.4.4。

作业停止操作步骤如下：

1. 登录实时计算控制台。
2. 单击页面顶部的运维。
3. 在运维页面，单击目标作业操作列下的停止。

❓ 说明 更多目录的停止（checkpoint）功能与停止功能的唯一区别是：停止（checkpoint）在进行停止作业的同时，会主动触发一次Checkpoint，因此停止（checkpoint）作业所消耗的时间可能会比停止作业所消耗的时间稍长一些。但作业停止后依然会清除作业的状态，该功能在个别场景下会有其他作用，例如在上游存储为kafka时，系统触发一次Checkpoint会提交一次offset，确保提交到kafka服务端的offset和实际消费的数据量一致。

5.8. 查看日志

您可以通过查看作业运行日志了解作业的运行状况。本文为您介绍如何查看日志。

操作步骤

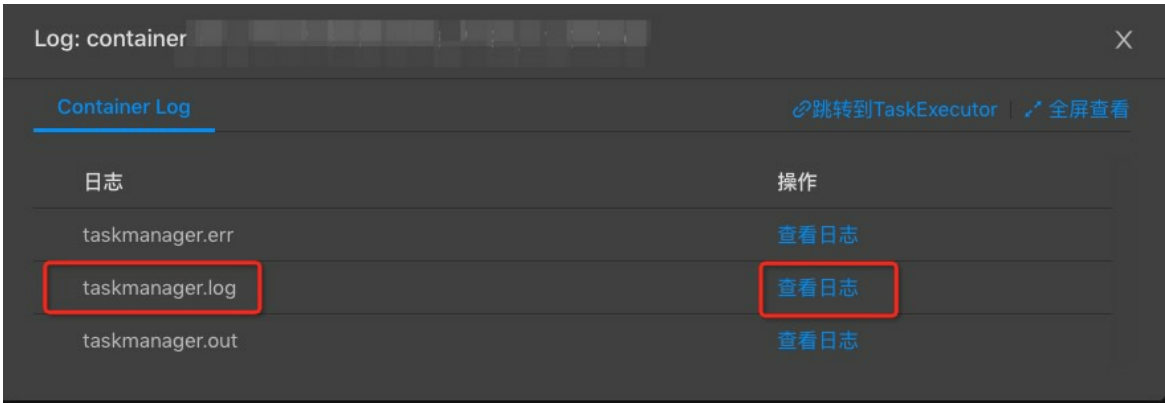
1. 登录作业运维页面。
 - i. 登录实时计算控制台。
 - ii. 单击页面顶部运维。
 - iii. 在作业列表区域，单击作业名称列中的目标作业。
2. 在作业运维 > 运行信息页面底部，单击目标Vertex名称。

The screenshot shows the '运行信息' (Run Information) page. At the top, there are tabs for '运行信息', '数据曲线', 'Failover', 'Checkpoints', 'JobManager', 'TaskExecutor', '血缘关系', and '属性参数'. Below these, there's a 'Task状态' section with counts for '创建', '运行', '失败', '完成', '调度', '取消中', and '已取消'. A table follows with columns for '输入TPS', '输入RPS', '输出RPS', '输入BPS', '消耗CU', and '启动时间'. Below this is a 'Vertex拓扑' section with a table of vertices. A red box highlights the '运行信息' tab and the 'Source: RandomSource -> fro...' entry in the table.

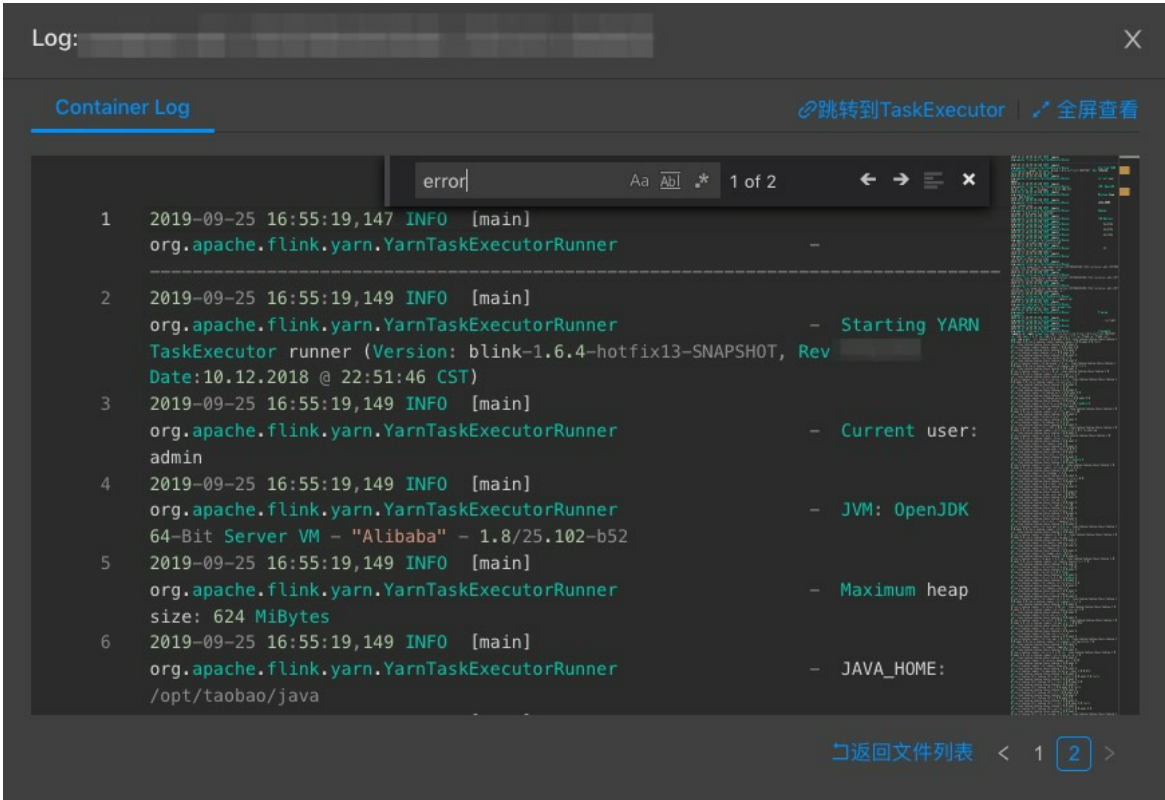
3. 在Execution Vertex > Subtask List，单击目标Subtask操作列下的查看日志。

The screenshot shows the 'Subtask List' page. At the top, there are tabs for 'Vertex Topology', 'Subtask List', 'Metrics Graph', 'Metrics Data', and 'Accumulators'. Below these, there's a table of subtasks with columns for 'ID', 'Status', 'In Queue', 'Out Queue', 'RecCnt', 'SendCnt', 'TPS', 'Retries', 'Duration', 'Host', 'Start Time', and '操作'. A red box highlights the 'Subtask List' tab and the '查看日志' (View Log) button in the '操作' column.

4. 在Log窗口，单击taskmanager.log操作列下的查看日志。



5. 在Container Log页面查看日志详情。



② 说明 可以使用快捷键（Windows系统为Ctrl加F，Mac系统为command加F）的方式，触发日志搜索功能，查看指定的日志内容。建议从最后一页往前查看，日志中的第一个error记录描述了作业报错的Root cause。