

ALIBABA CLOUD

# 阿里云

实时计算（流计算）  
快速入门

文档版本：20220126

 阿里云

## 法律声明

阿里云提醒您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 <b>设置</b> > <b>网络</b> > <b>设置网络类型</b> 。
<b>粗体</b>	表示按键、菜单、页面名称等UI元素。	在 <b>结果确认</b> 页面，单击 <b>确定</b> 。
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> <i>Instance_ID</i>
[ ] 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.概述	05
2.搭建上下游存储	07
2.1. 搭建数据源表存储	07
2.2. 搭建数据维表存储	08
2.3. 搭建数据结果表存储	09
3.数据开发	10
4.作业上线	12
5.生产运维	13

# 1.概述

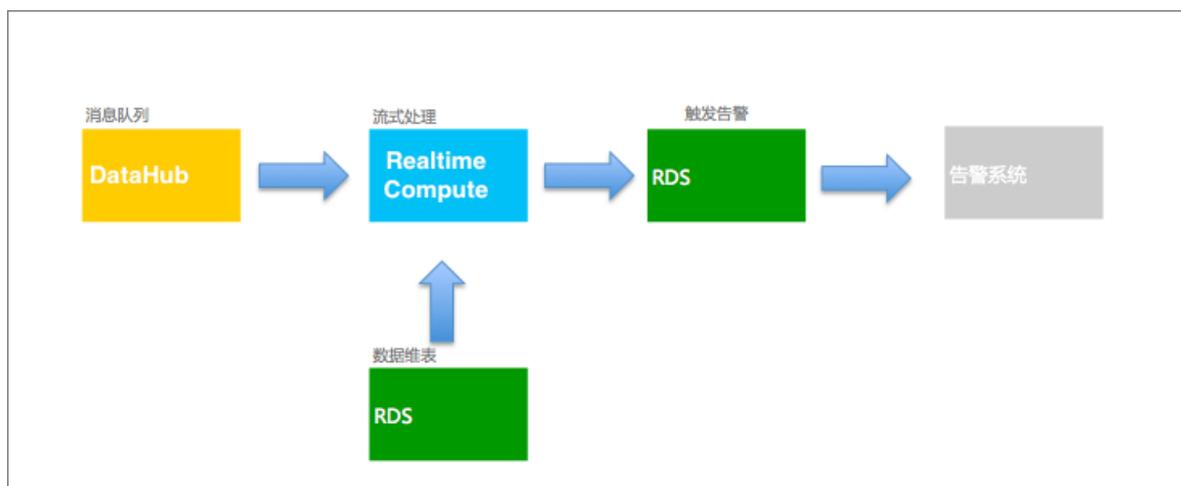
本文通过安全监控作业案例，帮助您了解实时计算Flink版作业开发流程。

## 实时安全监控案例介绍

实时监控报警作为保障数据安全的重要环节，受到越来越多的关注。为了保证同时满足数据的正确性和实效性，实时计算Flink版将流式数据源表和存放白名单的维度表进行实时JOIN，识别源表中数据对应的地址是否存在于白名单中，采取过滤等措施。

? 说明 本案例适用于实时计算Flink版共享模式（已停购）和独享模式。

## 安全监控业务架构图



- DataHub源表结构

字段名	类型	注释
name	varchar	名称
place	varchar	地址

- RDS维表结构

字段名	类型	注释
place	varchar	白名单地址

- RDS结果表结构

字段名	类型	注释
name	varchar	过滤后的名称
place	varchar	地址

## 作业操作流程

- 准备工作

- 开通服务和创建项目

详情请参见[开通服务和创建项目](#)。

 **说明** 您可以根据需要选择购买共享模式和独享模式，请参见[概述](#)。

- 搭建上下游数据存储

- [搭建数据源表存储](#)
- [搭建数据维表存储](#)
- [搭建数据结果表存储](#)

- 作业操作步骤

- i. [数据开发](#)
- ii. [作业上线](#)
- iii. [生产运维](#)

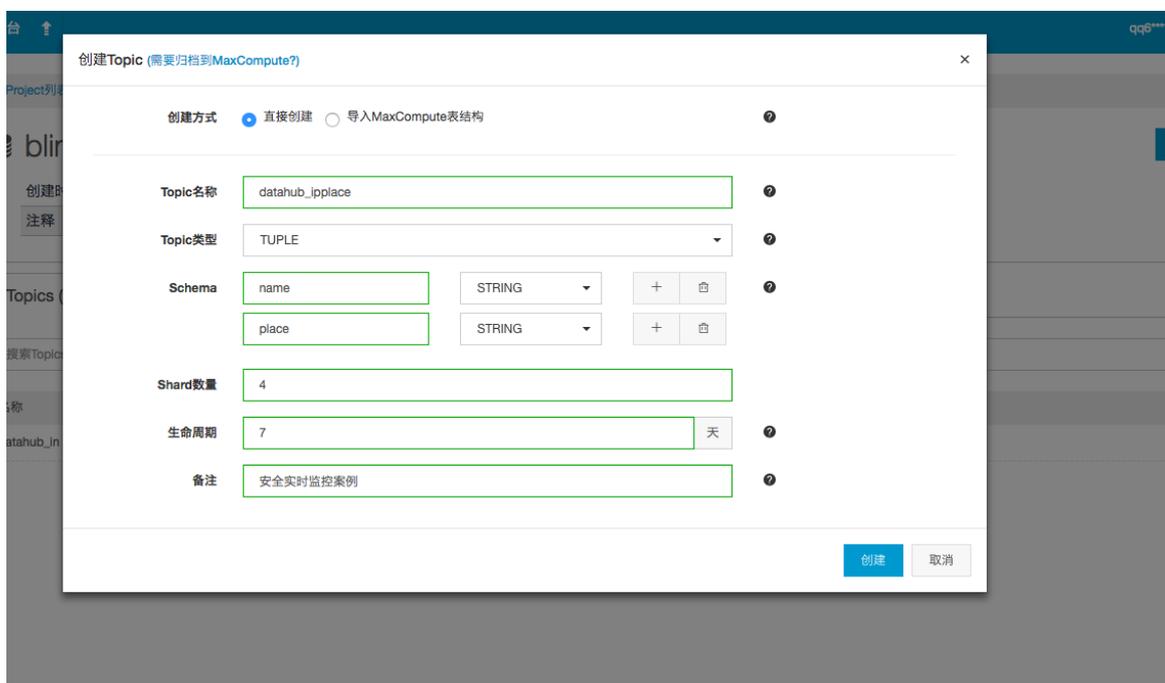
## 2. 搭建上下游存储

### 2.1. 搭建数据源表存储

本文以DataHub数据源表存储为例，为您介绍如何准备实时计算数据源表存储。

#### 创建数据源表存储

1. 登录DataHub控制台。
2. 在项目管理页面的顶部，选择项目区域。
3. 单击页面右上角的创建Project。
4. 在创建DataHub的Project窗口中，输入项目（Project）参数。
5. 在项目列表，目标项目的操作列下，单击查看。
6. 单击右上角的创建Topic。
7. 填写Topic参数信息。



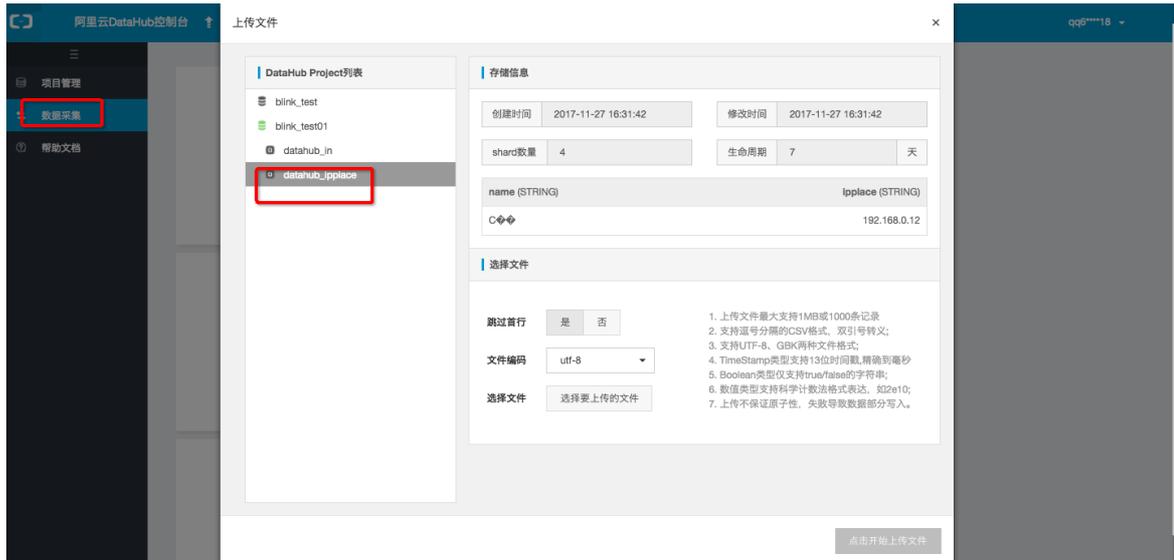
**说明** DataHub中的Topic类似于数据库中的表。

8. 单击确认。

#### 输入数据源表测试数据

**说明** 您可以使用测试数据进行测试。

1. 单击左侧导航栏中的数据采集。
2. 单击文件上传。
3. 在DataHub Project列表中，选择目标Topic，本示例为datahub\_ipplace。



4. 单击选择文件区域中的选择要上传的文件，选择测试数据。
5. 单击点击开始上传文件。

## 2.2. 搭建数据维表存储

本文以云数据库RDS MySQL版维表存储为例，为您介绍如何搭建实时计算Flink版数据维表存储。

### 创建数据维表存储

**注意** 本示例仅支持RDS经典网络。

1. 创建云数据库RDS MySQL版实例，详情请参见 [创建RDS MySQL实例](#)。
2. 创建云数据库RDS MySQL版账号和数据库，详情请参见 [创建数据库和账号](#)。
3. 设置白名单，详情请参见 [数据存储白名单配置](#)。
4. 登录数据库，详情请参见 [通过DMS登录RDS数据库](#)。
5. 在左侧已登录实例中，双击您已创建的数据库名称。
6. 在您创建的数据库下，创建 `rds_dim` 维表。
  - i. 在表页签下，单击[这里](#)开始建表。

**说明** 您也可以在表页签下的空白区域，鼠标右键，单击创建表。

- ii. 在基本信息 > 表中输入 `rds_dim`。
- iii. 在列信息中，单击+新增后，输入如下信息。

列名	类型	长度	主键
place	varchar	32	是

**说明** 选中主键字段下的方格，设置 `place` 为主键。

- iv. 单击保存变更 > 直接执行。

## 输入维表测试数据

1. 在页面左侧表页签下，双击rds\_dim。
2. 在SQL Console中输入 `INSERT INTO `rds_dim`(`place` ) values ('beijing');`。
3. 在SQL Console左上角，单击执行。

## 2.3. 搭建数据结果表存储

本文以RDS MySQL版数据结果表存储为例，为您介绍如何准备实时计算数据结果表存储。

### 创建数据结果表存储

 **说明** 本示例仅支持RDS经典网络。

1. 创建RDS MySQL实例，步骤请参见[创建RDS MySQL实例](#)。
2. 创建RDS MySQL账号和数据库，步骤请参见[创建数据库和账号](#)。
3. 在您创建的数据库下，创建 `rds_ipplace` 结果表，表格信息如下：
  - i. 鼠标悬停至页面顶部的创建，单击下拉菜单中的表。
  - ii. 在基本信息 > 表名 中输入 `rds_dim`。
  - iii. 在列信息中输入如下信息。

列名	类型	长度	主键
name	varchar	32	是
place	varchar	32	否

 **说明** 单击主键字段下的方格，设置 `name` 为主键。

- iv. 单击保存 > 确定。

## 3.数据开发

完成准备工作后，您可以进行实时计算数据开发。

### 创建作业

1. 登录[实时计算控制台](#)。
2. 在页面顶部，单击开发。
3. 单击新建作业。
4. 在新建作业页面，输入作业配置信息。

作业参数	说明
文件名称	作业的名称。 <div style="border: 1px solid #ccc; background-color: #e6f2ff; padding: 5px; margin-top: 5px;"> <span style="font-size: 1.2em; color: #007bff;">?</span> 说明 作业名称在当前项目中需保持唯一。         </div>
作业类型	<ul style="list-style-type: none"> <li>◦ 共享模式：仅支持 FLINK_STREAM/SQL作业类型。</li> <li>◦ 独享模式：支持 FLINK_STREAM/DATASTREAM和FLINK_STREAM/SQL作业类型。</li> </ul>
存储位置	指定该作业的代码文件所属的文件夹。 您还可以在现有文件夹右侧，单击  图标，新建子文件夹。

5. 单击确定。

### 引用数据存储

完成作业创建后，您可以通过注册数据存储的方式，关联已搭建的上下游存储。

- 注册引用源表数据
  - i. 注册您的DataHub Project，步骤参见[注册数据总线DataHub](#)。
  - ii. 在开发页面的数据存储页签，双击DataHub 数据存储。
  - iii. 双击目标数据库名称。
  - iv. 双击datahub\_ipplace 。
  - v. 在数据表详情区域，单击作为输入表引用。
- 注册引用维表数据
  - i. 参考[注册云数据库RDS版](#)，完成RDS的注册。
  - ii. 在开发页面的数据存储页签，双击RDS数据存储。
  - iii. 双击目标数据库名称。
  - iv. 双击rds\_dim 。
  - v. 在数据表详情区域，单击作为维表引用。
- 注册引用结果表数据
  - i. 参考[注册云数据库RDS版](#)，完成RDS的注册。
  - ii. 在开发页面的数据存储页签，双击RDS数据存储。
  - iii. 双击目标数据库名称。

- iv. 双击rds\_ipplace。
- v. 在数据表详情区域，单击作为结果表引用。

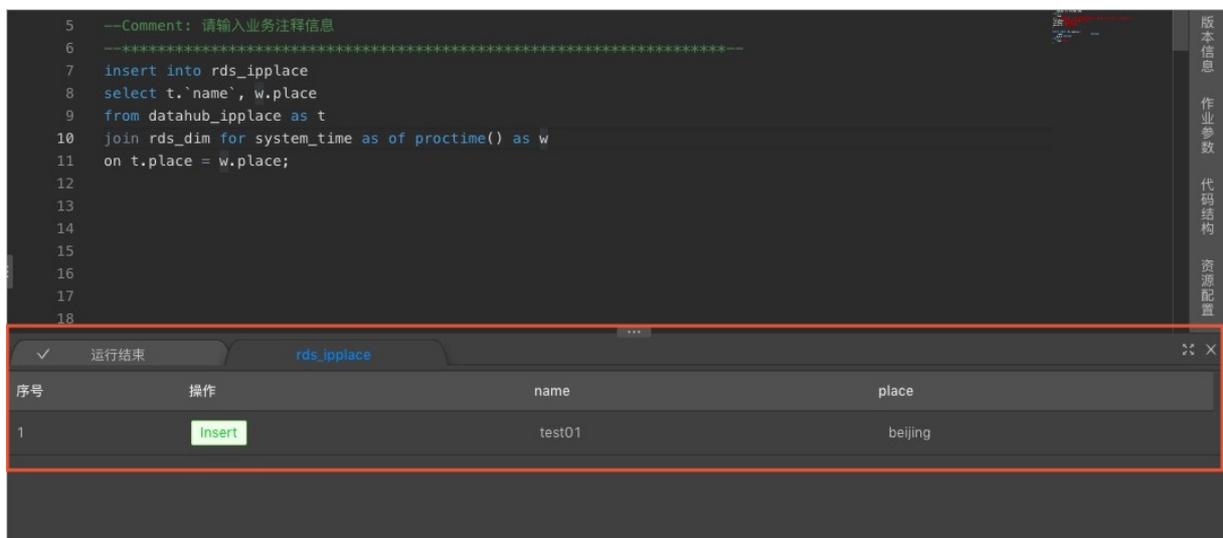
### 编写测试SQL

完成数据存储引用后，在作业编辑页面输入如下SQL代码，完成作业的开发。

```
insert into rds_ipplace
select t.`name`, w.place
from datahub_ipplace as t
join rds_dim for system_time as of proctime() as w
on t.place = w.place;
```

### 作业调试

完成作业开发后，可进行本地调试。作业调试详情，参见 [本地调试](#)。本地调试结果示例图如下。



源表和维表的测试数据如下：

- [DataHub源表测试数据](#)
- [RDS维表测试数据](#)

## 4.作业上线

作业调试完成，经验证逻辑无误后，可以在开发页面中完成作业上线。

上线作业操作将您的改动提交到数据运维中。作业上线步骤如下：

1. 登录[实时计算控制台](#)。
2. 单击页面顶部的[开发](#)。
3. 在开发页面左侧的作业开发页签中，双击目标作业。
4. 单击作业编辑页面顶部的[上线](#)。
5. 在上线新版本向导中，完成上线操作。

### i. 资源配置

选择对应的资源配置方式。第1次启动建议使用系统默认配置。

 **说明** 实时计算支持手动资源配置和自动资源配置2种资源配置方式：

- 手动资源配置方法参见[手动配置调优](#)。
- 自动资源配置方法根据实时计算版本，分为以下2种方式：
  - 实时计算3.0及以上版本：AutoScale自动配置，详情请参见[AutoScale自动配置调优](#)。
  - 实时计算3.0以下版本：AutoConf自动配置，详情请参见[AutoConf自动配置调优](#)。

### ii. 数据检查

通过数据检查后，单击[下一步](#)。

### iii. 上线作业

单击[上线](#)。

 **说明** 作业上线后只是将作业提交至集群，并没有启动作业。启动作业请参见[启动](#)。

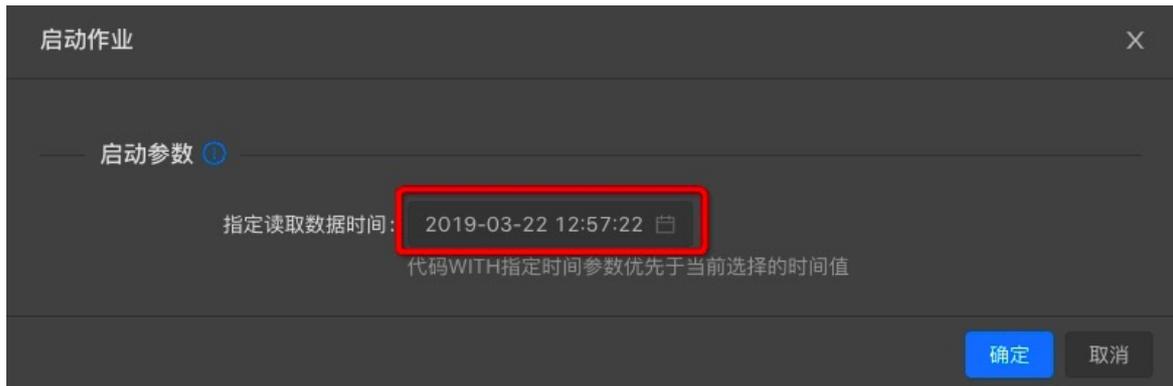
完成作业上线后，您可在生产环境中进行作业启动等生产运维工作。

## 5.生产运维

完成作业上线后，您可在生产环境中进行作业启动等生产运维工作。

### 作业启动步骤

1. 登录**实时计算控制台**。
2. 单击页面顶部的**运维**。
3. 在**运维**，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。

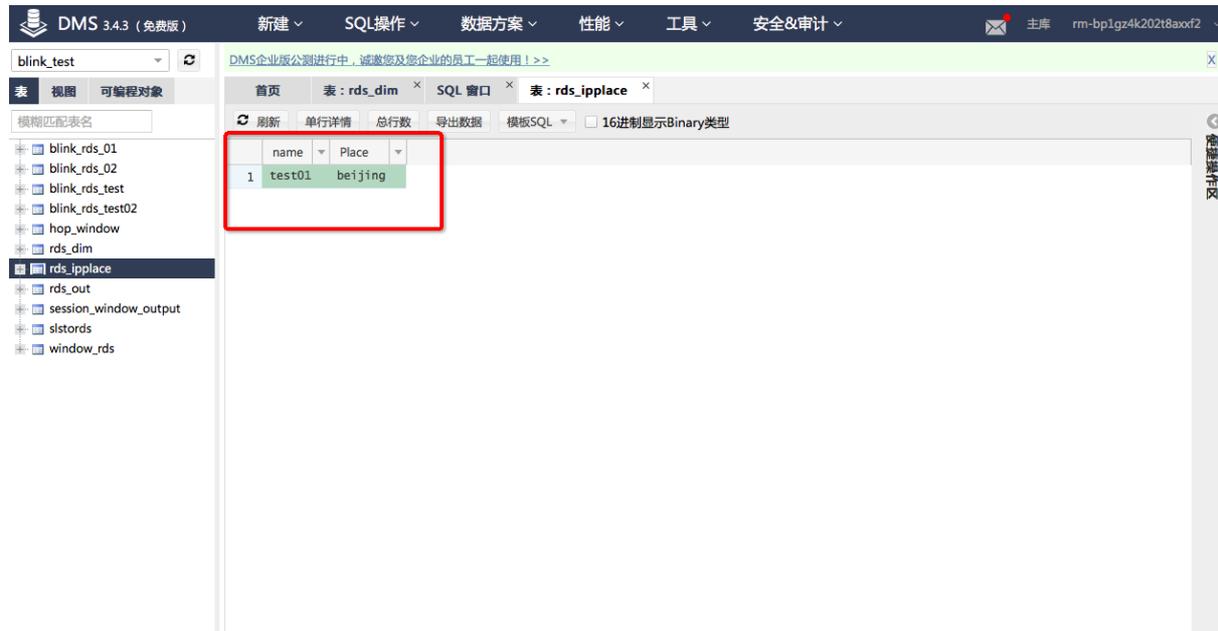


5. 指定读取数据时间（启动位点），单击**确定**，完成作业启动。

- 说明** 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。

### 查看结果

使用测试数据完成作业运行后，可以发现，DataHub源表（datahub\_ipplace）中，仅RDS维表（rds\_dim）输入的 place 为 beijing 的数据，输出到了RDS结果表（rds\_ipplace）中。从而验证了业务代码的正确性，实现了实时监控的需求。



## 业务运维

实时计算提供了丰富的业务运维功能，详情请参见 [运行信息](#)。