云数据仓库套件 Sparkling

操作指南

产品文档

腾讯云

文档目录

操作指南
集群管控
查看集群信息

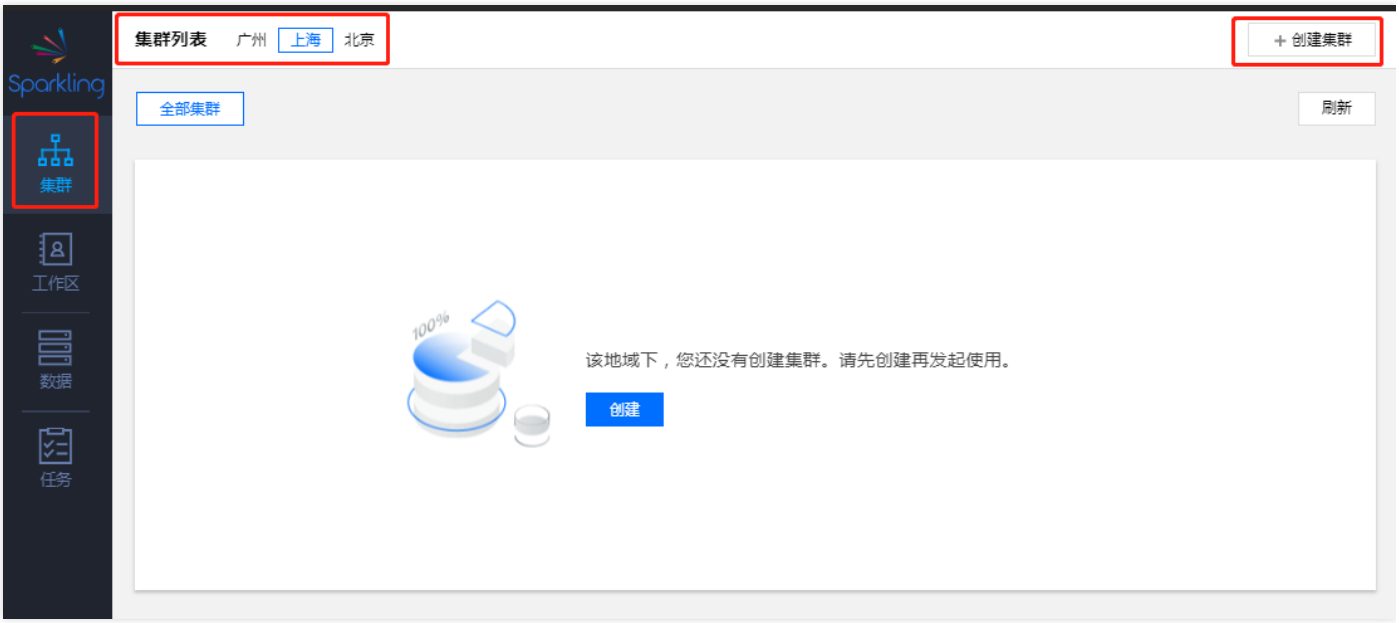
最近更新时间：2019-05-24 09:06:40

操作场景

用户可以通过集群列表页面查看集群创建时间、集群节点机型、节点个数等基本信息。

操作步骤

1. 登录 Sparkling 控制台，在左侧导航单击【集群】进入集群管理页面。



2. 单击左上角地域，可查看当前地域下的集群信息。首次进入时，需要单击右上角【创建集群】，以创建该地域集群，详细操作可参见 创建集群。

> **说明：**
>
> 每个地域下仅支持创建一个集群。

3. 当所选地域下存在集群时，可以在当前页面中查看该集群基本信息。其中节点的机型信息可参见 腾讯云 CVM 实例类型。

4. 单击【SparkUI/logs】可以查看集群 Spark 运行日志信息。同时也可以在该页面进行 集群监控、扩缩容集群、销毁集群 等操作。

集群监控

最近更新时间：2019-05-24 09:06:44

操作场景

用户可在集群监控页面查看当前集群的资源使用、任务个数等实时及历史数据信息，以监控集群资源的使用情况，从而为集群规模的调整提供参考。

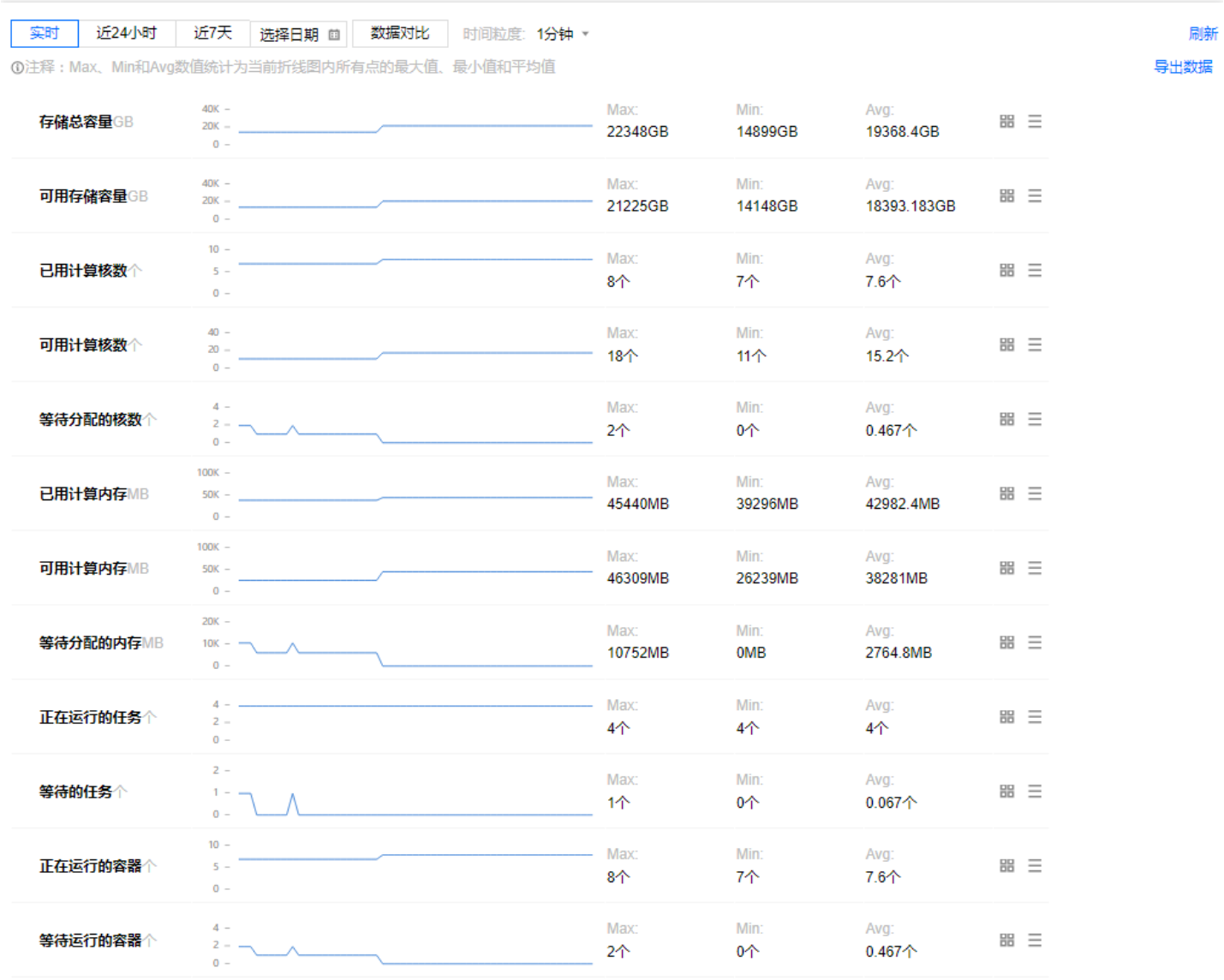
操作步骤

进入集群监控页面

登录 Sparkling 控制台 ，在左侧导航单击【集群】进入集群管理页面。在右侧集群信息栏中单击【查看监控】进入集群监控页面。

查看监控信息

在集群监控页面中可以查看集群实时、近24小时、近7天的数据，该数据为每分钟抓取的集群监控数据。单击右侧【刷新】可以获取最新数据。



单击【选择日期】可以显示该日期下的数据，时间粒度表示以该时间间隔进行数据的展示。

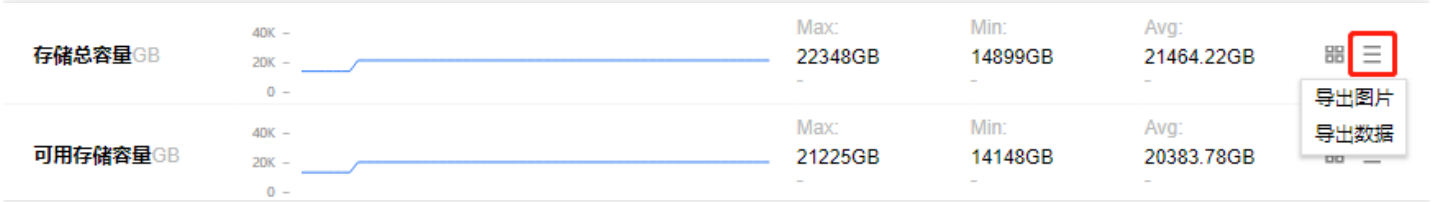
单击【数据对比】可以比较两段时期下的数据走势情况。以一深一浅表示两个时间段的数据走势，鼠标移动到该走

势图上可以查看该点两段时间的数据对比。

导出监控数据

单击右上角【导出数据】，在弹出对话框中，选择时间范围及时间粒度，选择要导出的指标后单击【导出】，即可将数据批量导出为 csv 文件。

单击 ☰ 可以选择单个导出某一指标的图片或数据。



扩缩容集群

最近更新时间：2019-05-24 09:06:51

用户需要调整计算资源以满足业务需要时，可以在集群管理扩缩容已有集群。

操作步骤

1. 登录 Sparkling 控制台，在左侧导航单击【集群】进入集群管理页面。
2. 在集群信息的【操作】栏下，单击【扩缩容】弹出集群信息框，修改当前节点数量，**目前版本不支持同时进行扩容和缩容操作**，确认信息后单击【执行】。

提交申请后，集群的【状态】会显示为【扩容中】或【缩容中】，扩缩容需要时间请耐心等待，成功后【状态】会变成【可用】。

销毁集群

最近更新时间：2019-05-24 09:06:55

用户不再需要使用某个集群时，可以参考如下操作删除集群。**删除的集群无法恢复，同时集群中的数据也会自动删除且无法再访问。**

操作步骤

1. 登录 Sparkling 控制台，在左侧导航单击【集群】进入集群管理页面。
2. 在集群信息的【操作】栏下，单击【销毁】，在弹出的信息框下单击【确认】即可完成集群销毁操作。

数据集成
操作场景

最近更新时间：2019-07-01 09:56:00

云数据仓库套件 Sparkling 支持多样化的数据接入方式和数据源管理。

数据接入方式包括：

- 关系型数据库（RDBMS）接入：可以通过 RDBMS 数据接入方式将云数据库 MySQL、分布式数据库 TDSQL 中的数据接入到 Sparkling 中。更多信息，请参考 MySQL 数据接入、TDSQL 数据接入。
- 腾讯云对象存储（COS）数据接入：可以通过生成账户密钥，建立存储桶（bucket）的方式进行 COS 数据接入。
- 腾讯云消息队列（CKafka）数据接入：可以通过 Kafka 数据接入 方式将腾讯云 CKafka 中的数据接入 Sparkling 中。

数据源管理 提供了查看、筛选、创建和删除数据源，以及测试数据源连通性等功能。

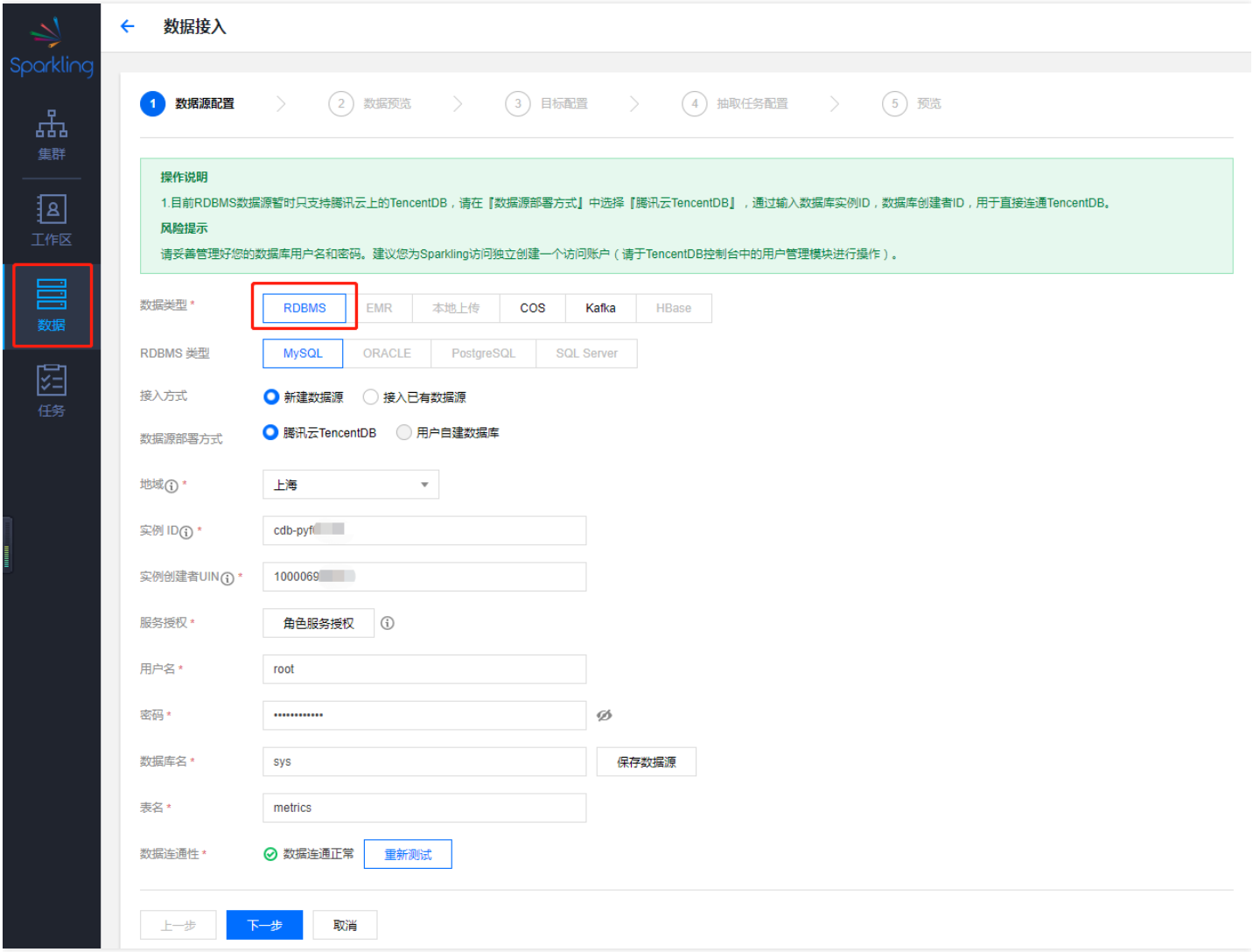
RDBMS 数据接入
MySQL 数据接入

最近更新时间：2019-07-01 09:55:33

本节为您介绍云数据库 MySQL 数据接入方法。通过 RDBMS 数据接入功能可以实现将云数据库 MySQL 中的数据以全量或增量的方式导入 Sparkling 中。用户可以通过时间戳增量追加方式实现增量同步，通过自定义调度周期实现定时数据接入功能。
更多关于云数据库 MySQL 的信息请参见 云数据库 MySQL 产品介绍。

操作步骤

登录 Sparkling 控制台，在左侧导航单击【数据】进入数据接入页面，按以下操作步骤完成 RDBMS 数据接入：



1. 数据源配置

选择【RDBMS】数据类型，支持【新建数据源】和【接入已有数据源】两种数据源接入方式。

- **新建数据源方式**

  a. 接入方式选择【新建数据源】。

  b. 选择数据库所在的地域，对于跨地域数据库访问其稳定性和速率将会受到地域制约。

  c. 填写云数据库实例 ID（可在 云数据库控制台 获取数据库实例 ID）及数据库购买者 UIN（实例购买者在 账号信息 中的账号 ID）。

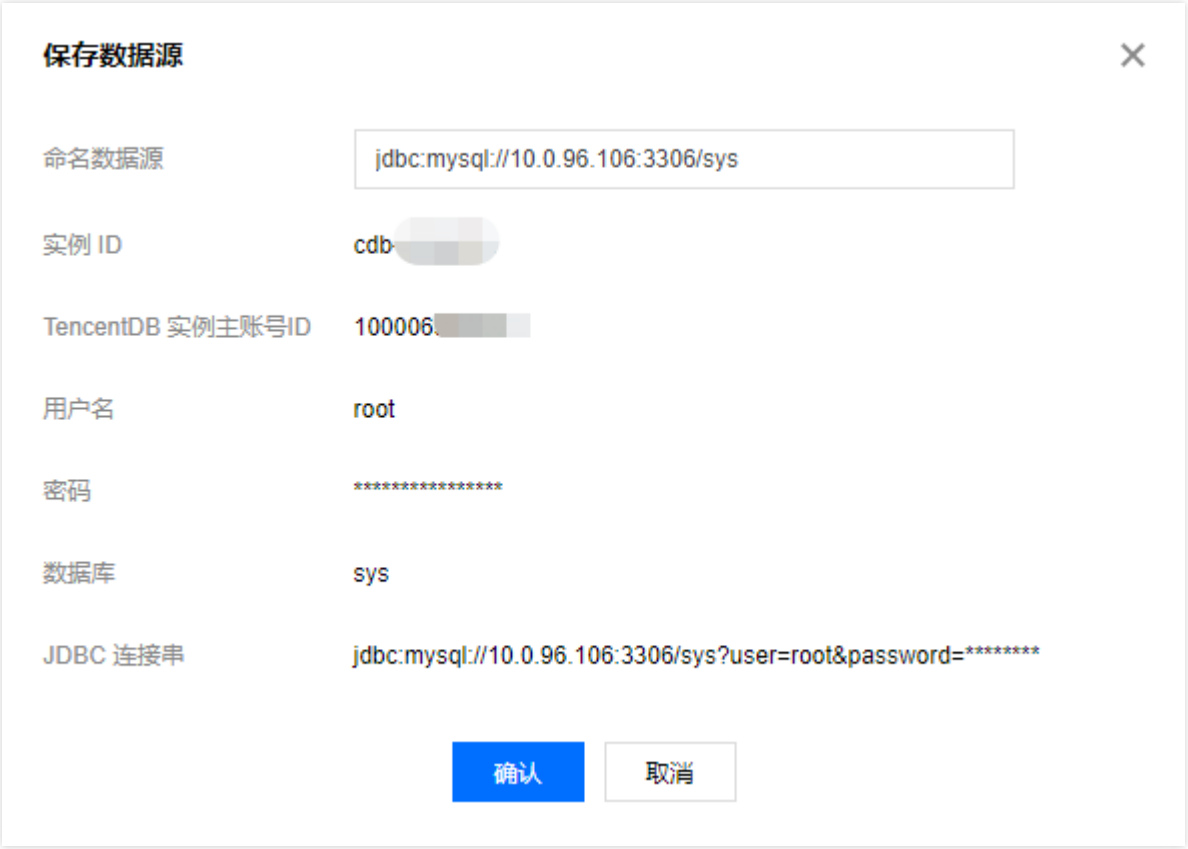
  d. 单击【角色服务授权】，授权 Sparkling 服务访问其他相应服务。

  e. 填写要接入的数据表所在的【用户名】、【密码】、【数据库名】、【表名】。

  f. 选择【数据连通性】>【测试连通性】确认是否可以连接到要接入数据表所在的数据库。

  g. 待显示【数据连通正常】后，单击【下一步】完成数据源配置操作，同时可以选择【保存数据源】选项将该数

据源保存为已有数据源，待下次接入该数据源时可采用【接入已有数据源】方式。



- **接入已有数据源方式**

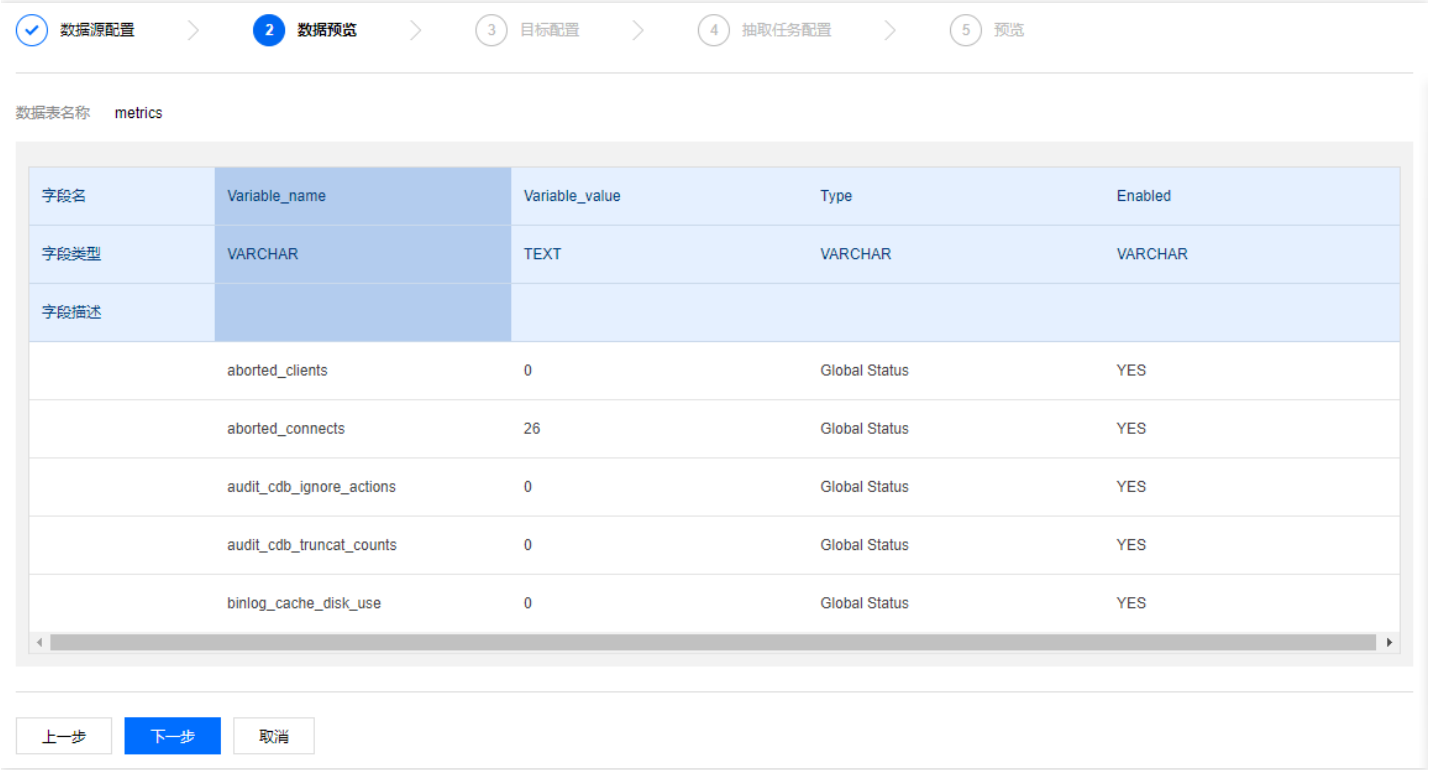
  a. 接入方式选择【接入已有数据源】。

  b. 选择历史保存的已有数据源。

c. 输入表名后，单击【下一步】完成数据源配置操作。

2. 数据预览

在【数据预览】页可以预览数据表中的字段信息，默认抽取五行数据进行预览，预览无误后单击【下一步】。



3. 目标配置

支持【新建表】和【导入已有目标表】两种方式。

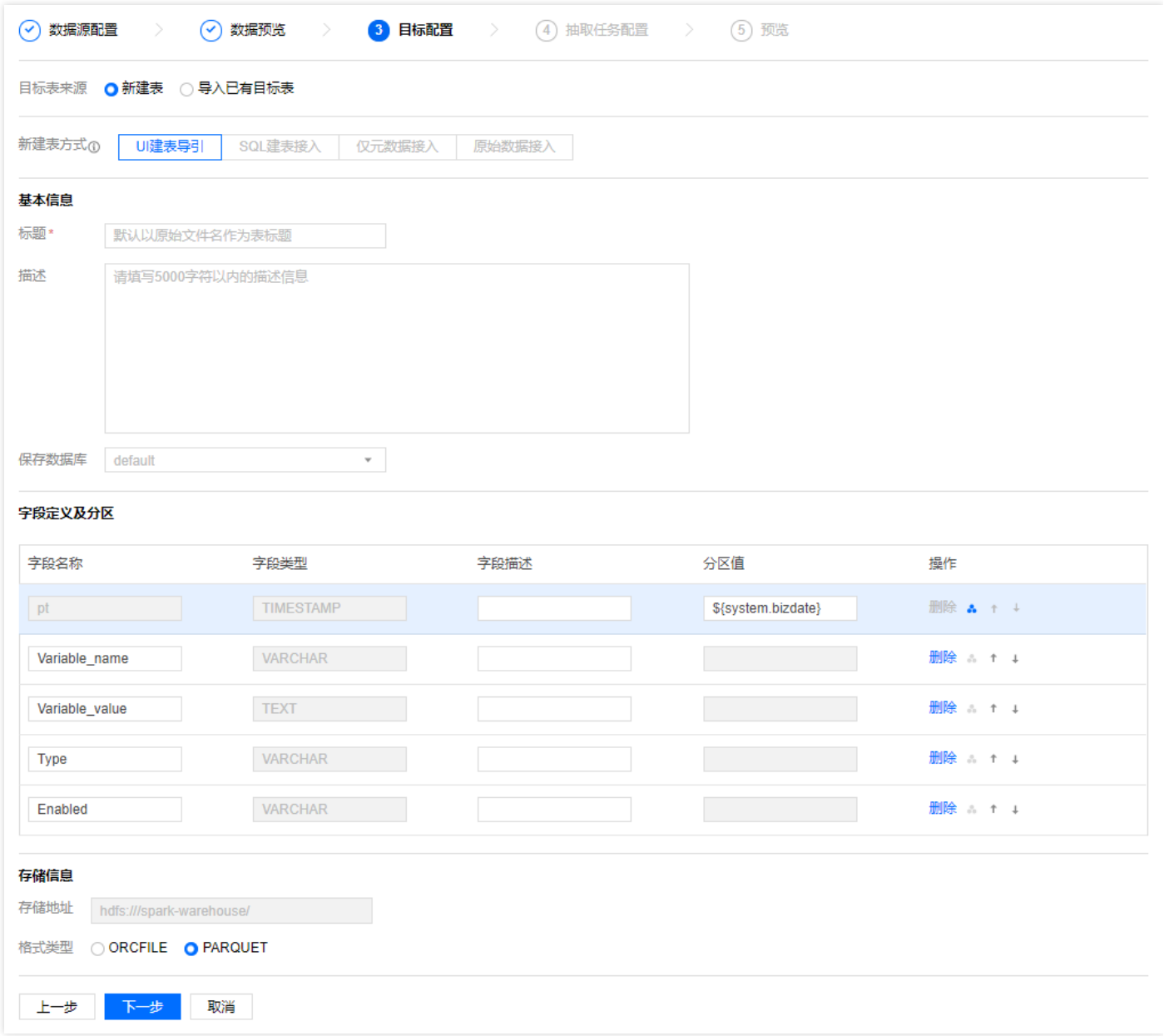
- **新建表方式**

  a. 选择【新建表】并填写【标题】和【描述】。

  b. 字段定义及分区部分可以设置字段名、描述及字段顺序，支持字段裁剪（可通过删除字段实现仅导入部分字段功能）。其中第一列“pt”列为默认分区字段，不支持删除操作，默认以该时间戳作为分区值。

  c. 选择格式类型，支持【ORCFILE】和【PARQUET】两种格式。

d. 确认无误后，单击【下一步】完成新建表方式目标配置操作。



- **导入已有目标表方式**

  导入已有目标表方式需要集群中已经包含该目标表，其中目标表包括数据导入时创建的表及在工作区中自建的表。

  a. 选择【导入已有目标表】后选择目标表名。

  b. 设置目标分区。选择【依据例行任务导入动态分区】将会根据您的例行任务配置信息写入对应的分区文件；选择【自定义分区】请按照分区字段明确分区命名。

  c. 单击【字段映射】进行字段匹配。若【下一步】处于置灰状态，说明字段映射未成功，请确认目标表与待导入数据源表字段可以匹配。

  d. 确认无误后，单击【下一步】完成导入已有目标表方式目标配置操作。

4. 调度任务配置

支持【单次】和【例行】两种调度方式。

- **单次方式**
  a. 调度周期选择【单次】。
  b. 数据加载方式可选择【时间戳增量追加】或【整表全量导入】。选择【时间戳增量追加】后需增加过滤条件，该方式只导入符合该过滤条件所包含的时间范围的数据，具体可参看语法说明；选择【整表全量导入】后可选择【写入前清理已有数据(Insert Overwrite)】或【写入前保留已有数据(Insert)】。

> **说明：**
> 时间戳增量追加方式仅适用于导入数据中包含时间戳的情形，若数据中不存在时间戳形式的列，该选项将无法选中。

c. 确认无误后，单击【下一步】完成单次抽取任务配置操作。

- **例行方式**

  a. 调度周期选择【例行】。

  b. 设置间隔周期，支持每天、每周、每月、每小时，例如选择：每天0时0分，即每日0点0分自动开始执行任务。

  c. 设置例行时长，即任务持续时间，支持一周、一个月、三个月、一年、永久。

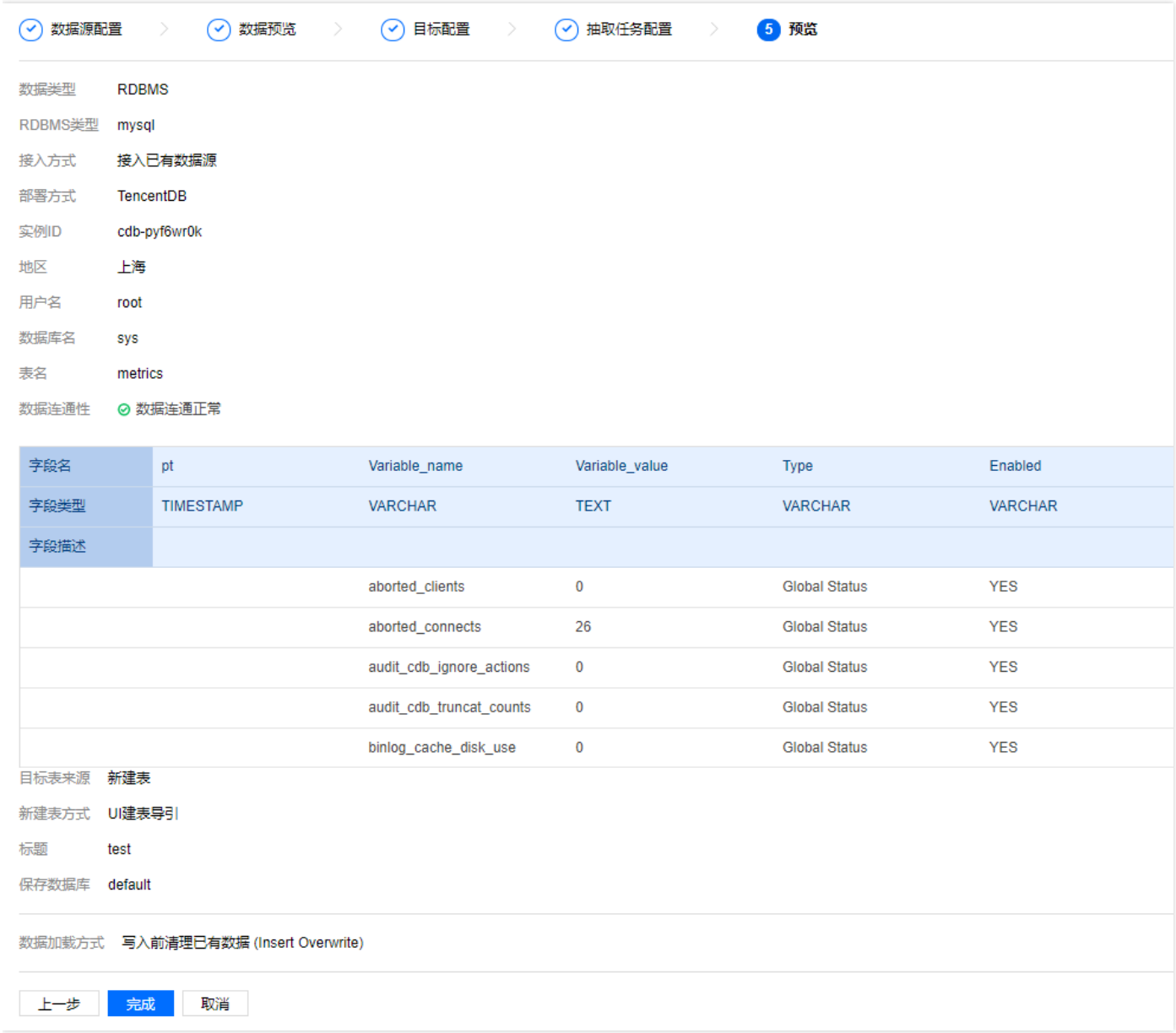
  d. 数据加载方式可选择【时间戳增量追加】或【整表全量导入】。选择【时间戳增量追加】后需增加过滤条件，该方式只导入符合该过滤条件所包含的时间范围的数据，具体可参看语法说明；选择【整表全量导入】后可选择【写入前清理已有数据(Insert Overwrite)】或【写入前保留已有数据(Insert)】。

  e. 确认无误后，单击【下一步】完成例行抽取任务配置操作。

> **说明：**
>
> 例行调度方式仅适用于目标配置为【导入已有目标表方式】，若选择【新建表方式】则该选项将无法选中。

5. 预览

在【预览】页可以查看当前设置的数据源信息、数据预览信息、目标表信息及任务调度信息。确认无误后单击【完成】即可完成 MySQL 数据接入任务设置。


数据源配置
数据预览
目标配置
抽取任务配置
5 预览
数据类型 RDBMS
RDBMS类型 mysql
接入方式 接入已有数据源
部署方式 TencentDB
实例ID cdb-pyf6wr0k
地区 上海
用户名 root
数据库名 sys
表名 metrics
数据连通性 数据连通正常
字段名 pt Variable_name Variable_value Type Enabled
字段类型 TIMESTAMP VARCHAR TEXT VARCHAR VARCHAR
字段描述
aborted_clients 0 Global Status YES
aborted_connects 26 Global Status YES
audit_cdb_ignore_actions 0 Global Status YES
audit_cdb_truncat_counts 0 Global Status YES
binlog_cache_disk_use 0 Global Status YES
目标表来源 新建表
新建表方式 UI建表导引
标题 test
保存数据库 default
数据加载方式 写入前清理已有数据 (Insert Overwrite)
上一步
完成
取消

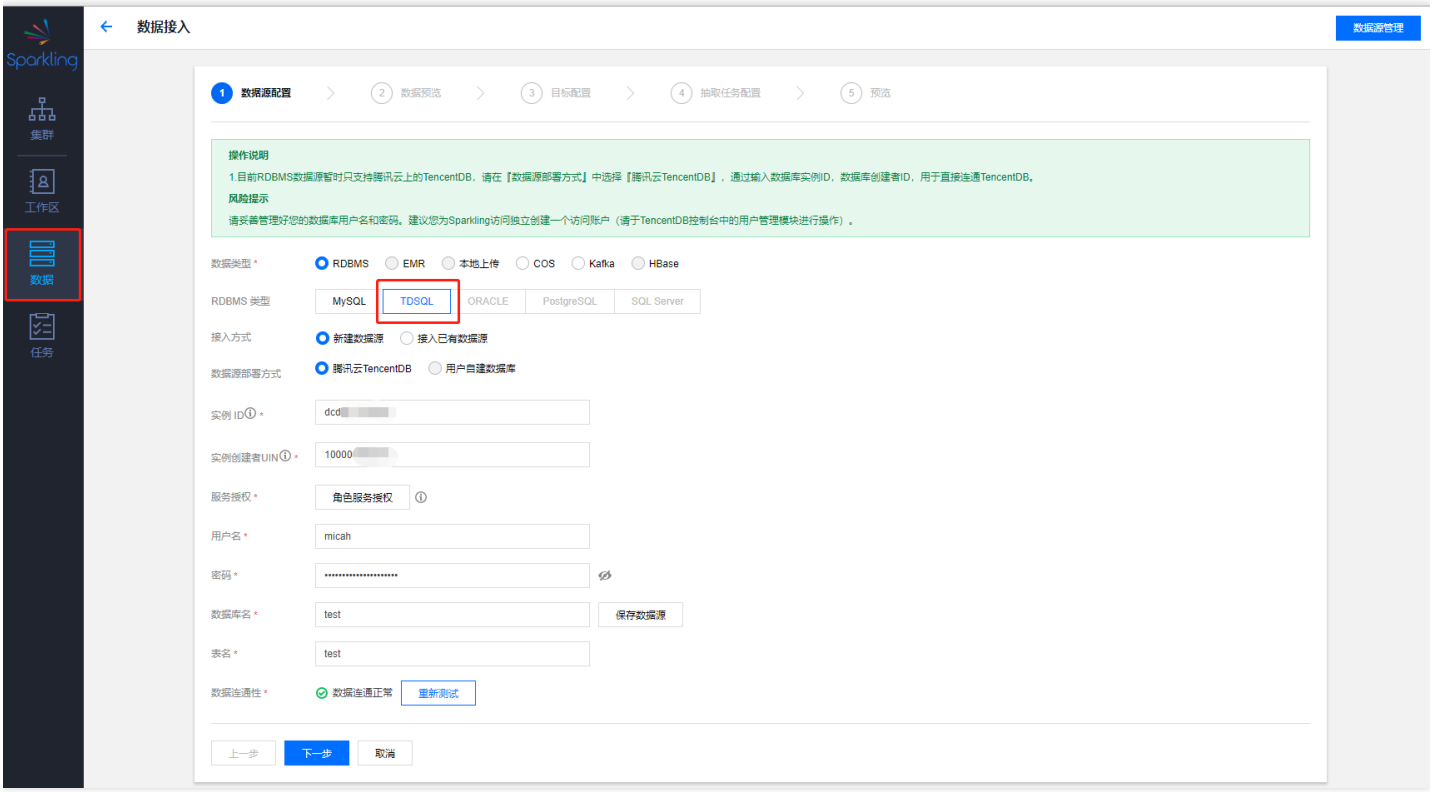
TDSQL 数据接入

最近更新时间：2019-07-01 09:55:45

本节为您介绍分布式数据库 TDSQL 数据接入方法。通过 RDBMS 数据接入功能可以实现将分布式数据库 TDSQL 中的数据以全量或增量的方式导入 Sparkling 中。用户可以通过时间戳增量追加方式实现增量同步，通过自定义调度周期实现定时数据接入功能。
更多关于云数据库 TDSQL 的信息请参见 分布式数据库 TDSQL 产品介绍。

操作步骤

登录 Sparkling 控制台，在左侧导航单击【数据】进入数据接入页面，按以下操作步骤完成 RDBMS 数据接入：



1. 数据源配置

选择【RDBMS】数据类型，支持【新建数据源】和【接入已有数据源】两种数据源接入方式。

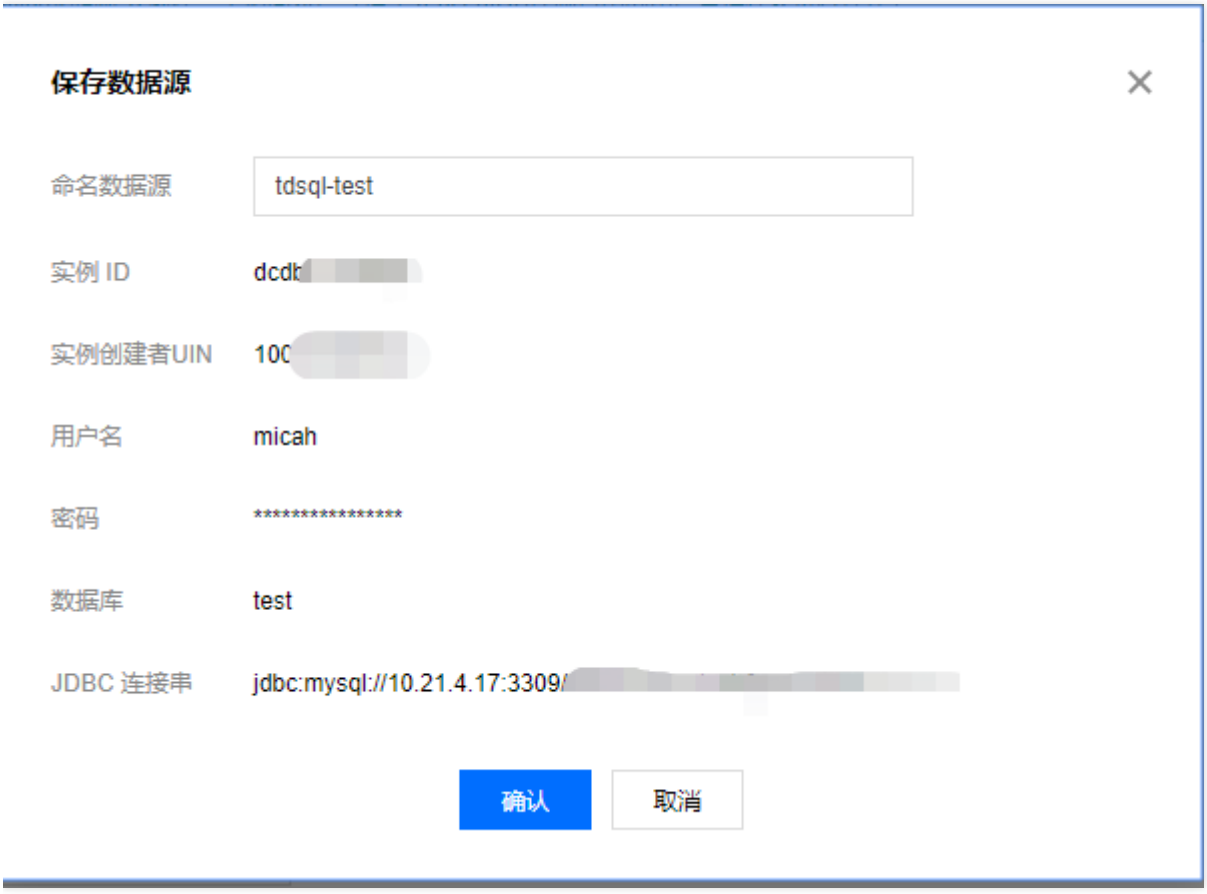
- **新建数据源方式**
  a. 接入方式选择【新建数据源】。
  b. 填写分布式数据库实例 ID（可在 分布式数据库控制台 获取数据库实例 ID）及数据库购买者 UIN（实例购买者在 账号信息 中的账号 ID）。

c. 单击【角色服务授权】，授权 Sparkling 服务访问其他相应服务。

d. 填写要接入的数据表所在的【用户名】、【密码】、【数据库名】、【表名】。

e. 选择【数据连通性】>【测试连通性】确认是否可以连接到要接入数据表所在的数据库。

f. 待显示【数据连通正常】后，单击【下一步】完成数据源配置操作，同时可以选择【保存数据源】选项将该数据源保存为已有数据源，待下次接入该数据源时可采用【接入已有数据源】方式。

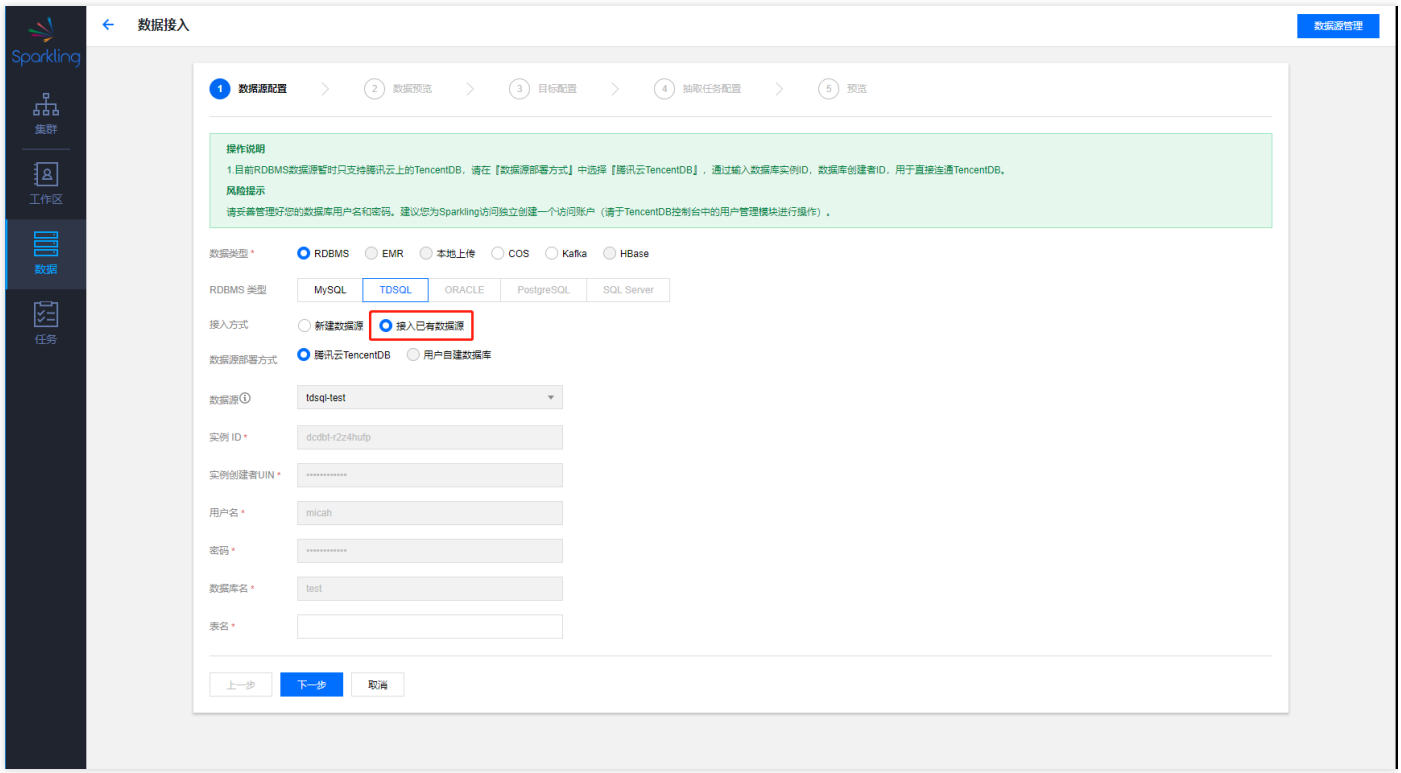


- **接入已有数据源方式**

  a. 接入方式选择【接入已有数据源】。

  b. 选择历史保存的已有数据源。

c. 输入表名后，单击【下一步】完成数据源配置操作。



2. 数据预览

在【数据预览】页可以预览数据表中的字段信息，默认抽取五行数据进行预览，预览无误后单击【下一步】。

3. 目标配置

支持【新建表】和【导入已有目标表】两种方式。

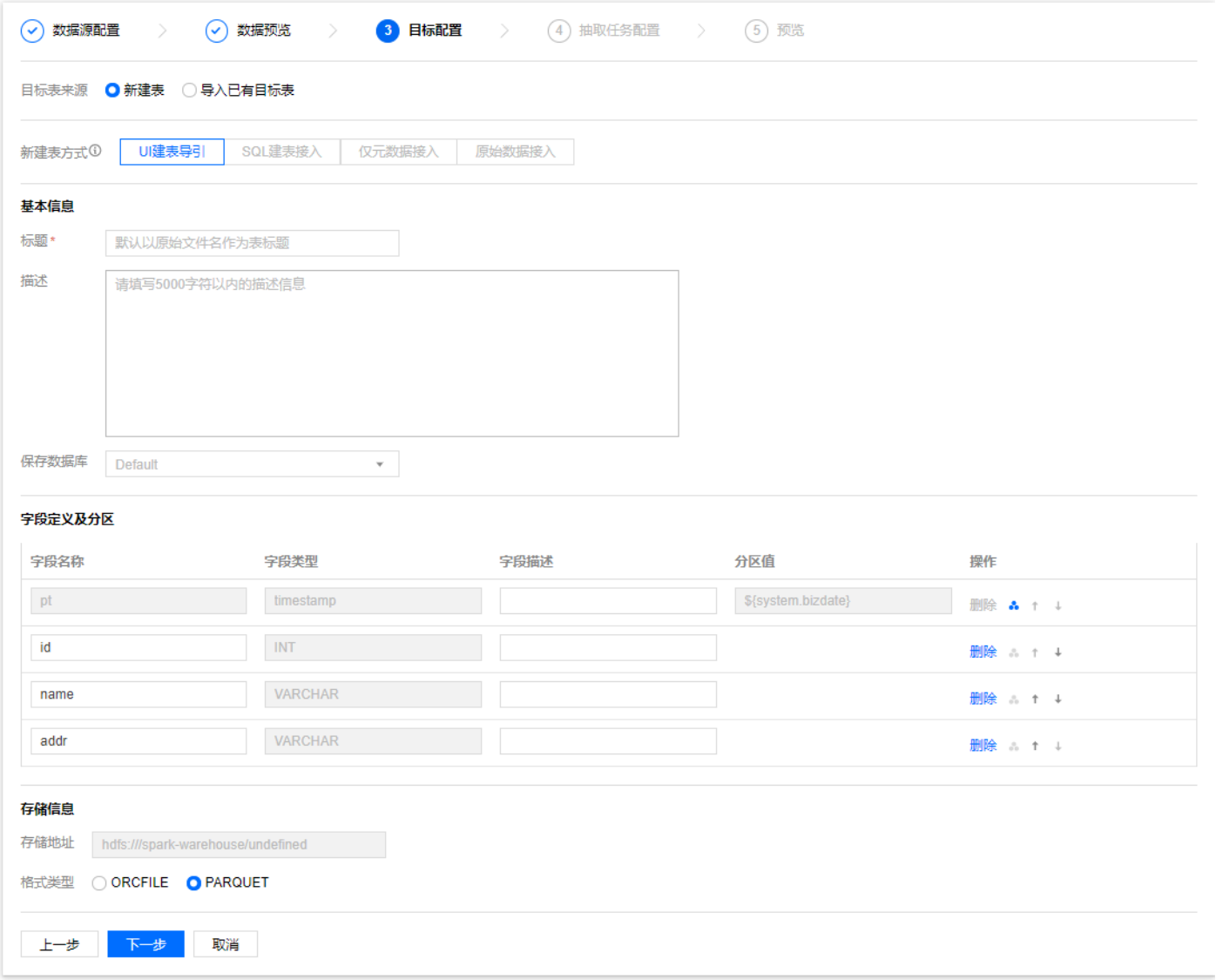
- **新建表方式**

  a. 选择【新建表】并填写【标题】和【描述】。

  b. 字段定义及分区部分可以设置字段名、描述及字段顺序，支持字段裁剪（可通过删除字段实现仅导入部分字段功能）。其中第一列“pt”列为默认分区字段，不支持删除操作，默认以该时间戳作为分区值。

  c. 选择格式类型，支持【ORCFILE】和【PARQUET】两种格式。

d. 确认无误后，单击【下一步】完成新建表方式目标配置操作。



- **导入已有目标表方式**

  导入已有目标表方式需要集群中已经包含该目标表，其中目标表包括数据导入时创建的表及在工作区中自建的表。

  a. 选择【导入已有目标表】后选择目标表名。

  b. 设置目标分区。选择【依据例行任务导入动态分区】将会根据您的例行任务配置信息写入对应的分区文件；选择【自定义分区】请按照分区字段明确分区命名。

  c. 单击【字段映射】进行字段匹配。若【下一步】处于置灰状态，说明字段映射未成功，请确认目标表与待导入数据源表字段可以匹配。

  d. 确认无误后，单击【下一步】完成导入已有目标表方式目标配置操作。

> **说明：**

使用导入已有目标表方式的前提是已经建立了可以与新导入数据实现字段映射的数据表。

4. 抽取任务配置

目前支持【单次】和【例行】调度方式。

- **单次方式**

  a. 调度周期选择【单次】。

  b. 数据加载方式可选择【时间戳增量追加】或【整表全量导入】。选择【时间戳增量追加】后需增加过滤条件，该方式只导入符合该过滤条件所包含的时间范围的数据，具体可参看语法说明；选择【整表全量导入】后可选择【写入前清理已有数据(Insert Overwrite)】或【写入前保留已有数据(Insert)】。

> **说明：**
>
> 时间戳增量追加方式仅适用于导入数据中包含时间戳的情形，若数据中不存在时间戳形式的列，该选项将无法选中。

c. 确认无误后，单击【下一步】完成单次抽取任务配置操作。

- **例行方式**

  a. 调度周期选择【例行】。

  b. 设置间隔周期，支持每天、每周、每月、每小时，例如选择：每天0时0分，即每日0点0分自动开始执行任务。

  c. 设置例行时长，即任务持续时间，支持一周、一个月、三个月、一年、永久。

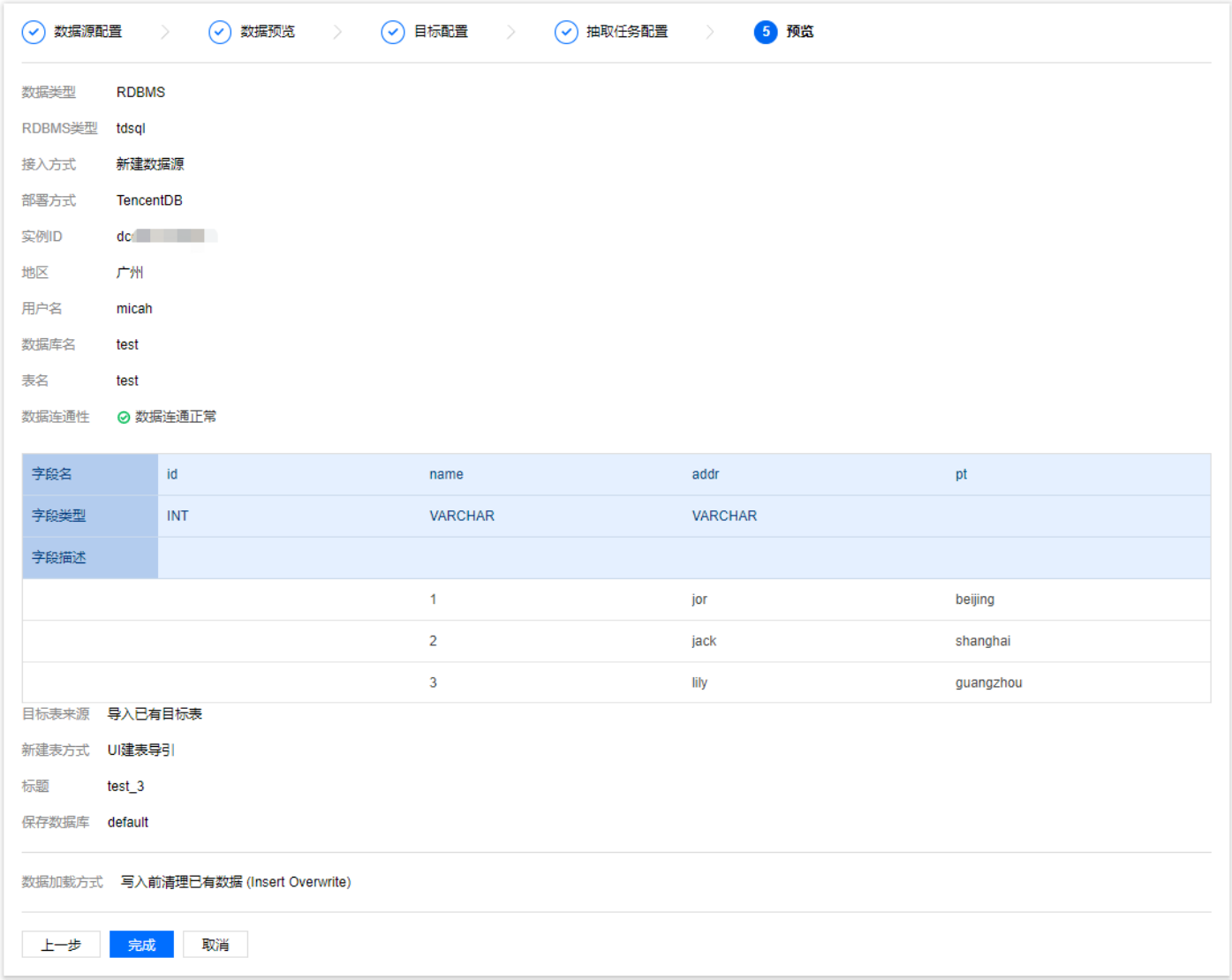
  d. 数据加载方式可选择【时间戳增量追加】或【整表全量导入】。选择【时间戳增量追加】后需增加过滤条件，该方式只导入符合该过滤条件所包含的时间范围的数据，具体可参看语法说明；选择【整表全量导入】后可选择【写入前清理已有数据(Insert Overwrite)】或【写入前保留已有数据(Insert)】。

> **说明：**
>
> 时间戳增量追加方式仅适用于导入数据中包含时间戳的情形，若数据中不存在时间戳形式的列，该选项将无法选中。

e. 确认无误后，单击【下一步】完成例行抽取任务配置操作。

5. 预览

在【预览】页可以查看当前设置的数据源信息、数据预览信息、目标表信息及任务调度信息。确认无误后单击【完成】即可完成 TDSQL 数据接入任务设置。


数据源配置
数据预览
目标配置
抽取任务配置
5 预览
数据类型 RDBMS
RDBMS类型 tdsql
接入方式 新建数据源
部署方式 TencentDB
实例ID dc
地区 广州
用户名 micah
数据库名 test
表名 test
数据连通性 数据连通正常
字段名 id name addr pt
字段类型 INT VARCHAR VARCHAR
字段描述
1 jor beijing
2 jack shanghai
3 lily guangzhou
目标表来源 导入已有目标表
新建表方式 UI建表导引
标题 test_3
保存数据库 default
数据加载方式 写入前清理已有数据 (Insert Overwrite)
上一步
完成
取消

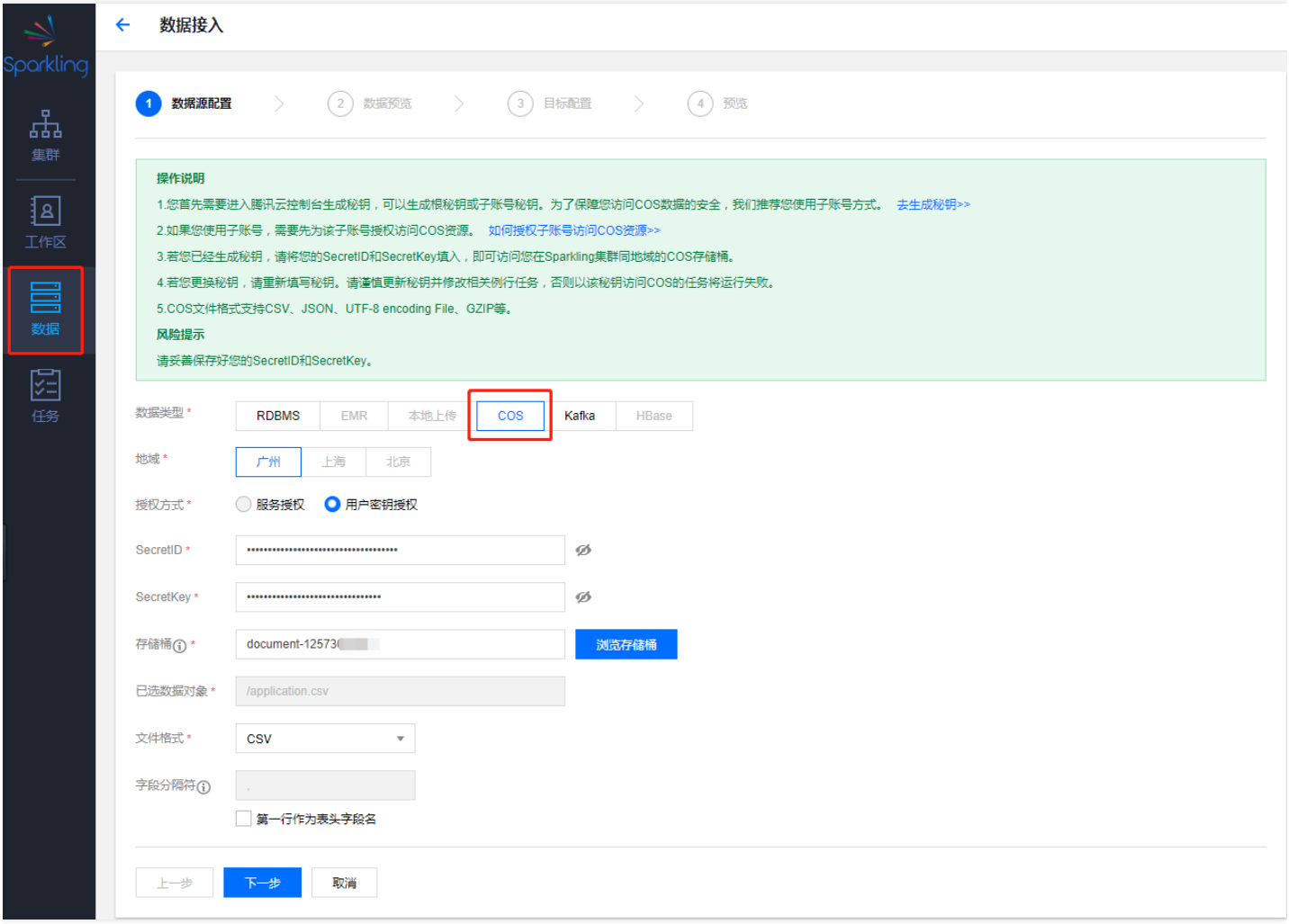
COS 数据接入

最近更新时间：2020-01-15 15:07:52

本节将为您介绍 COS 数据接入方法。更多关于 COS 的信息请参见 COS 产品介绍。

操作步骤

登录 Sparkling 控制台，在左侧导航单击【数据】进入数据接入页面，按以下操作步骤完成 COS 数据接入：



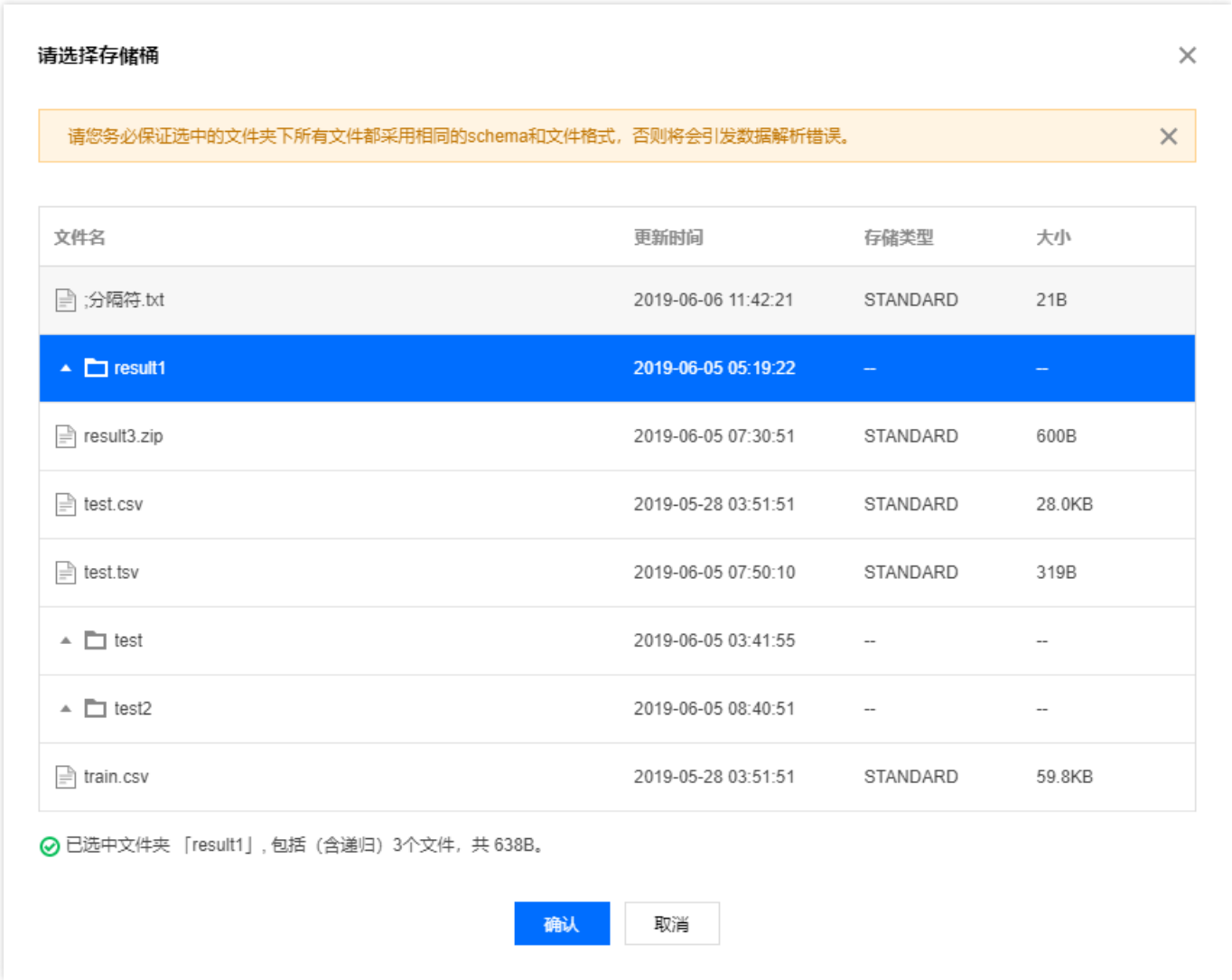
1. 数据源配置

- 数据类型：选择【COS】数据类型。
- 地域：选择 COS 存储桶所在地域。
- 授权方式：选择用户密钥授权。
- SecretID/SecretKey：填写您已生成的密钥，可在 API 密钥管理 中生成并查看。

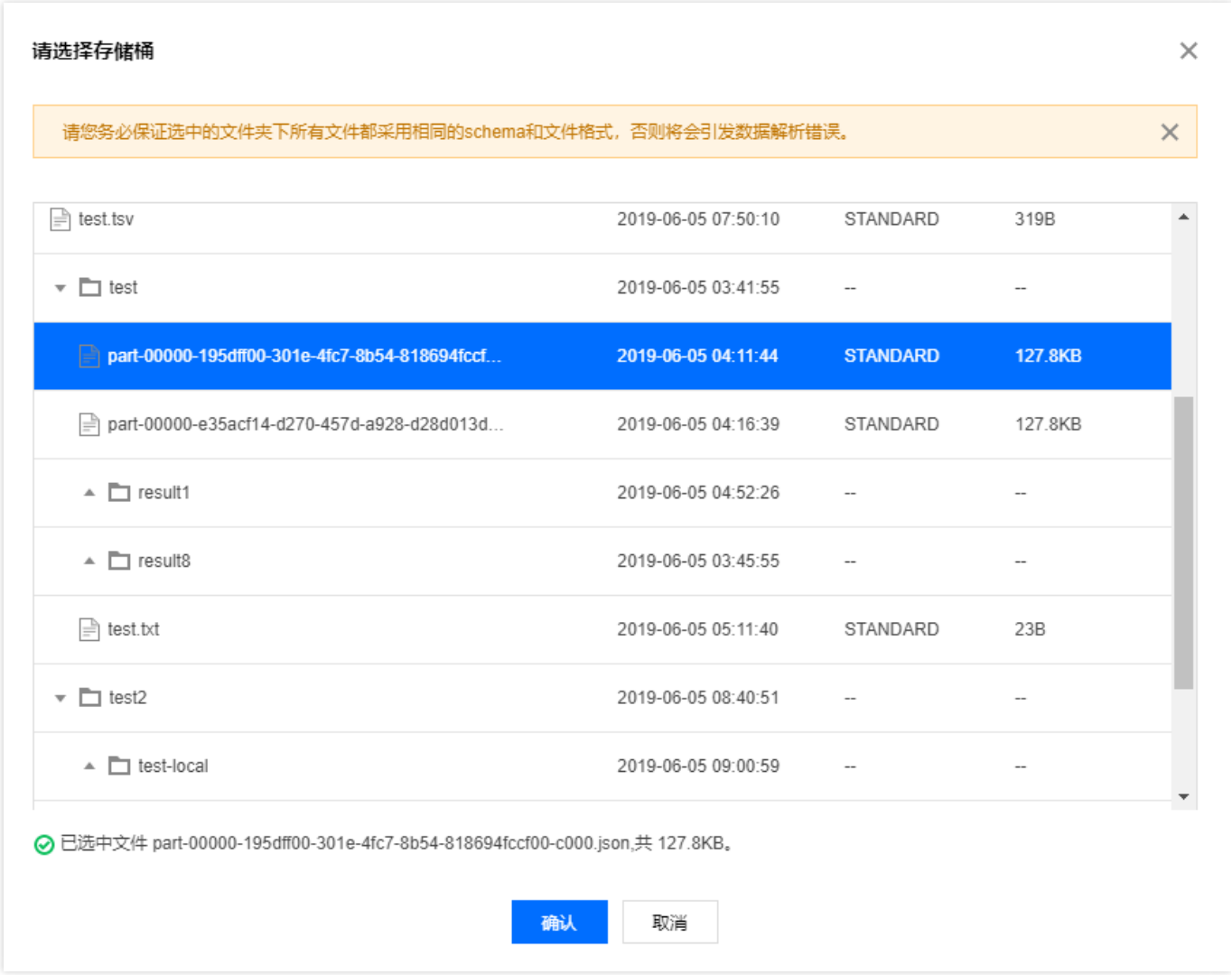
- 存储桶：填写您在 COS 中已生成的存储桶名称和您的 APPID，单击【浏览存储桶】查看当前存储桶下的数据并选择要导入的数据文件。数据文件导入方式支持【文件夹导入】和【文件导入】两种方式。

> 说明：
> 存储桶名称需要按<目标存储桶名称-APPID>格式填写，例如：sparkling-12334513，桶名和 APPID 可在账户信息中查看。

- 导入方式：
  - 文件夹导入方式：单击所选文件夹，左下角显示所选文件夹所包含文件个数及大小，确认信息无误后单击【确认】。



- 文件导入方式：单击所选文件，左下角显示所选文件名称及大小，确认信息无误后单击【确认】。



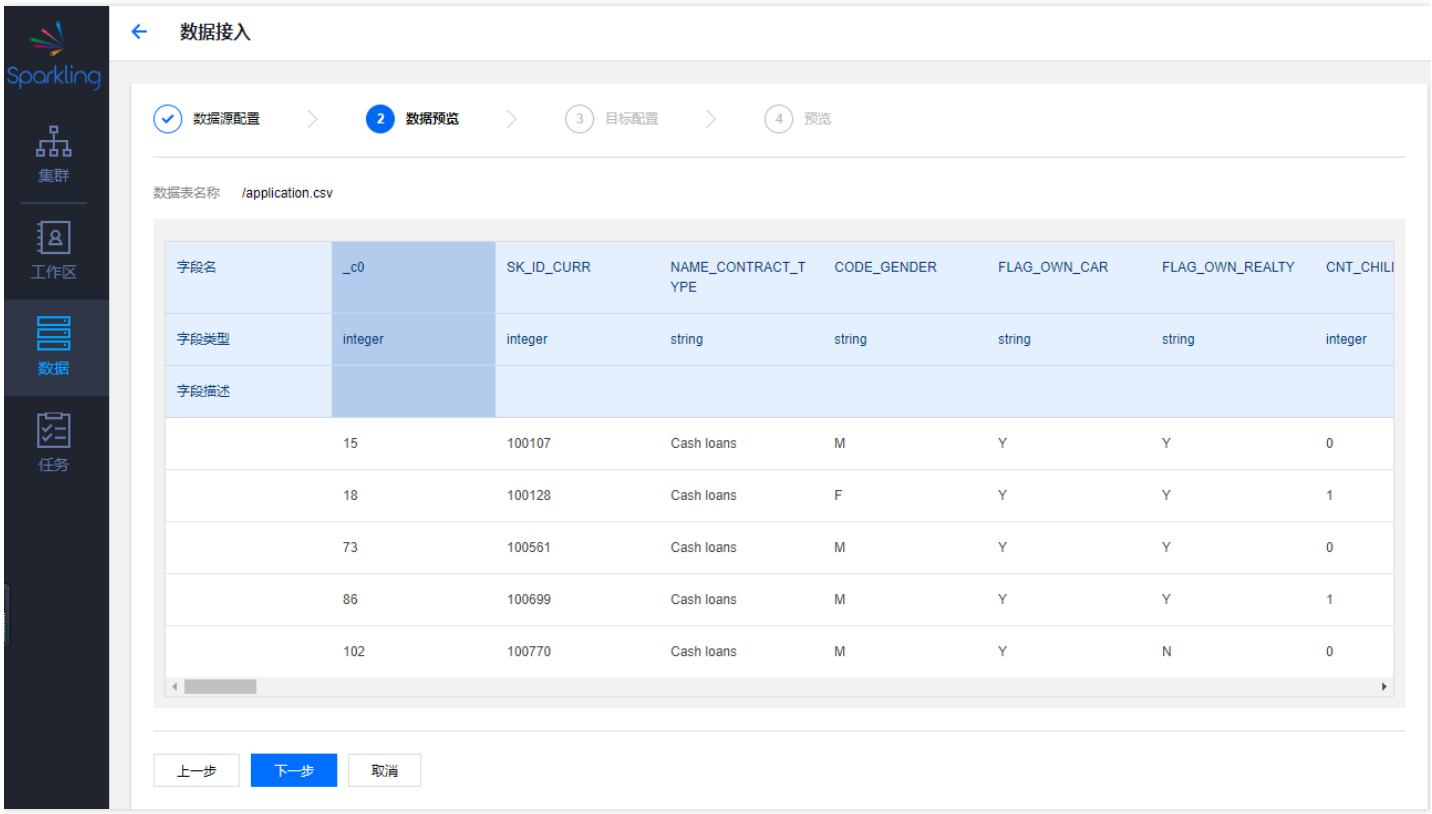
- 文件格式：支持 CSV、TSV、PARQUET、ORC、AVRO、JSON 及其他自定义分隔符日志的文件。

> 说明：
> COS 导入 JSON 文件时要求将 JSON 文件的每条记录必须用换行符分割。

- 字段分隔符：选择是否将第一行作为表头字段名。

2. 数据预览

确认信息无误后，单击【下一步】进行数据预览，本页默认显示前五行数据。



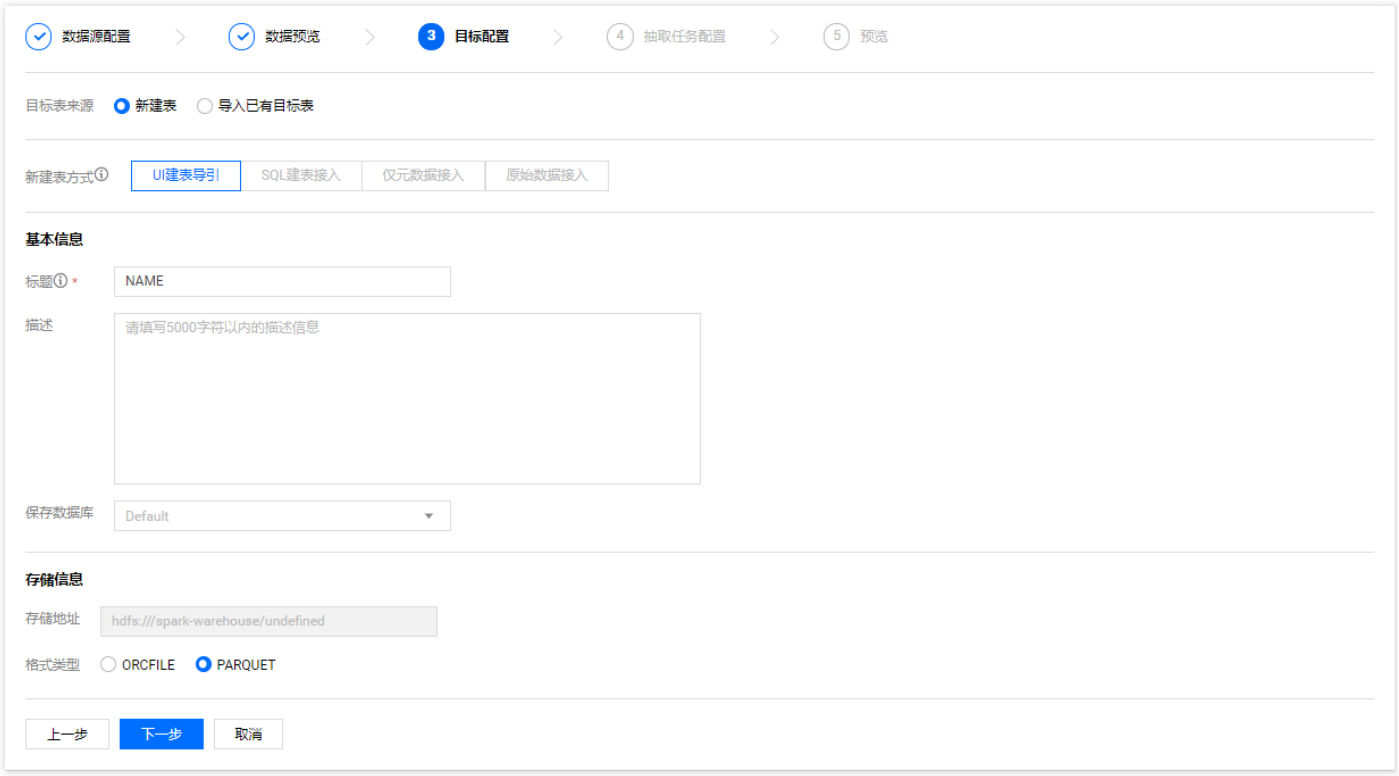
3. 目标配置

支持【新建表】和【导入已有目标表】两种方式。

新建表方式

1. 选择【新建表】并填写【标题】和【描述】。
2. 选择格式类型，支持【ORCFILE】和【PARQUET】两种格式。

3. 确认无误后，单击【下一步】完成新建表方式目标配置操作。



导入已有目标表方式

导入已有目标表方式需要集群中已经包含该目标表，其中目标表包括数据导入时创建的表及在工作区中自建的表。

1. 选择【导入已有目标表】后选择目标表名。
2. 设置目标分区。选择【依据例行任务导入动态分区】将会根据您的例行任务配置信息写入对应的分区文件；选择【自定义分区】请按照分区字段明确分区命名。
3. 单击【字段映射】进行字段匹配。若【下一步】处于置灰状态，说明字段映射未成功，请确认目标表与待导入数据源表字段可以匹配。
4. 确认无误后，单击【下一步】完成导入已有目标表方式目标配置操作。

> 说明：
> 使用导入已有目标表方式的前提是已经建立了可以与新导入数据实现字段映射的数据表。

4. 抽取任务配置

支持【单次】和【例行】两种调度方式。

单次方式

COS 单次调度任务支持【整表全量导入/覆盖】方式。

例行方式

目前支持【增量追加】和【整表全量导入/覆盖】两种方式。

- 增量追加
  a. 调度周期选择【例行】。
  b. 数据加载方式选择【增量追加】。
  c. 设置例行时间长，即任务持续时间，支持一周、一个月、三个月、一年、永久。

d. 确认无误后，单击【下一步】完成单次抽取任务配置操作。

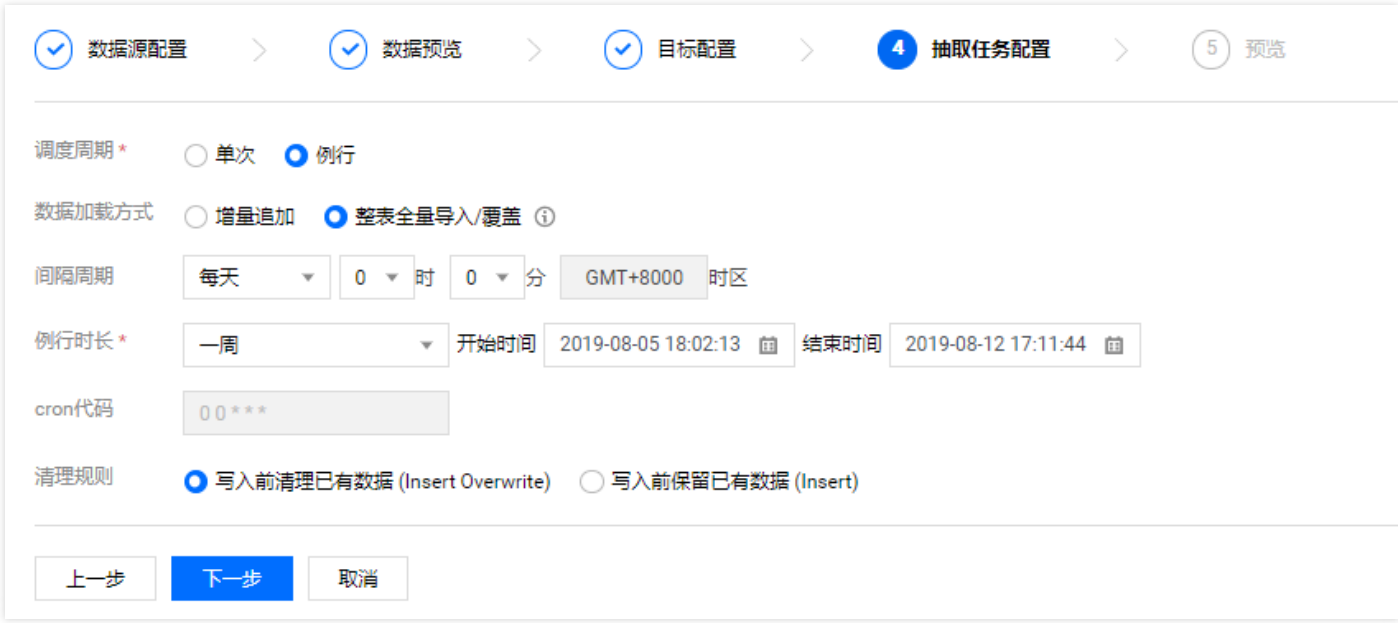
- 整表全量导入/覆盖

a. 调度周期选择【例行】。

b. 设置间隔周期，支持每天、每周、每月、每小时，例如选择：每天0时0分，即每日0点0分自动开始执行任务。

c. 设置例行时长，即任务持续时间，支持一周、一个月、三个月、一年、永久。

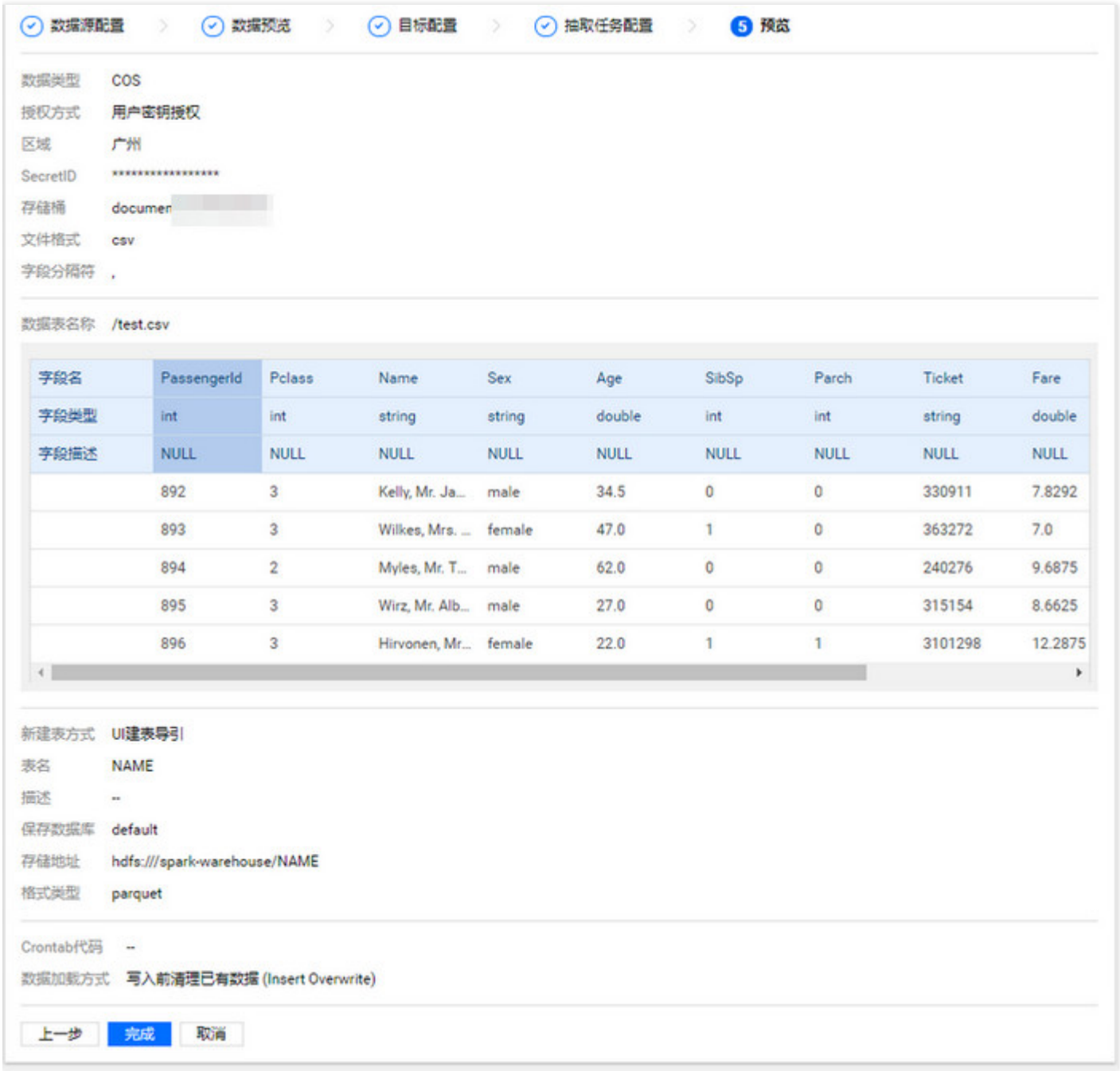
d. 数据清理规则可选择【写入前清理已有数据(Insert Overwrite)】或【写入前保留已有数据(Insert)】。



e. 确认无误后，单击【下一步】完成例行抽取任务配置操作。

5. 预览

任务预览无误后单击【完成】即可。



Kafka 数据接入

最近更新时间：2020-01-15 15:31:25

本节将介绍 CKafka 数据接入方法。更多关于 CKafka 的信息请参见 CKafka 产品介绍。

操作步骤

登录 Sparkling 控制台，在左侧导航单击【数据】进入数据接入页面，按以下操作步骤完成 Kafka 数据接入：

1. 数据源配置

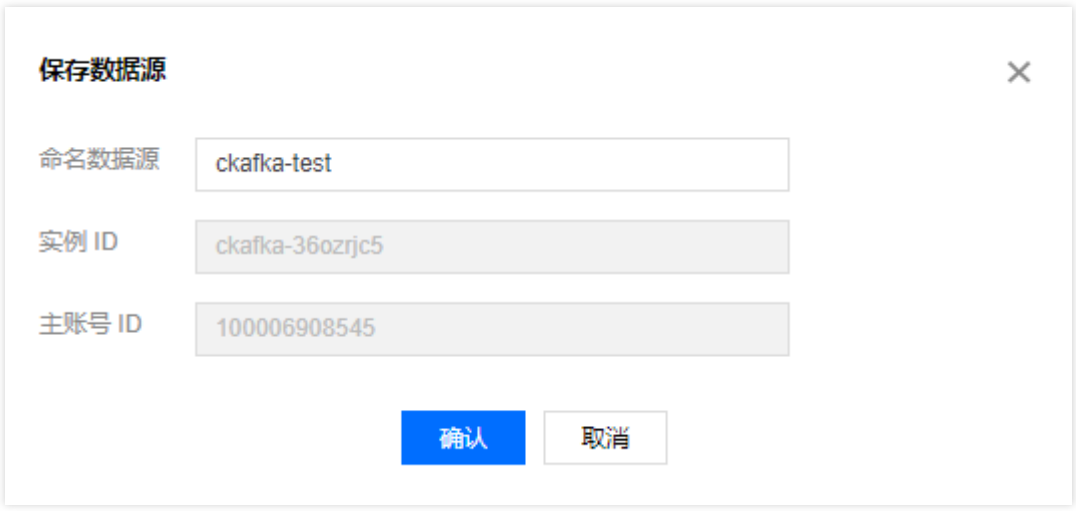
选择【Kafka】数据类型，支持【新建数据源】和【接入已有数据源】两种数据源接入方式。

新建数据源方式

1. 接入方式选择【新建数据源】。
2. 数据源部署方式选择腾讯云 CKafka。
3. 填写腾讯云 CKafka 实例 ID（可在 CKafka 控制台 获取）及 CKafka 创建者 UIN（实例创建者在 账号信息 中的账号 ID）。

4. 单击角色服务授权，授权 Sparkling 服务访问其他相应服务。

5. 填写要接入的 Topic 。
6. 单击【测试连通性】确认是否可以连接到要接入的 Topic 所在的 CKafka。
7. 待显示【数据连通正常】后，可以选择【保存数据源】选项将该数据源保存为已有数据源，待下次接入该数据源时可采用【接入已有数据源】方式。



8. 数据源高级配置。
   - 选择当前数据编码格式：目前 Sparkling 支持 AVRO、JSON、CSV 三种编码格式。CSV 编码格式下可以设置字段分隔符：目前可以使用“,”、“\t”、空格、“|”、“;”等，不支持“\n”、“*”等分隔符。
   - 配置读取开始 OFFset，可以根据需求选择数据订阅是否忽略历史消息或从最开始订阅数据。如果数据量比较大，从最开始订阅数据时会耗用比较长时间。

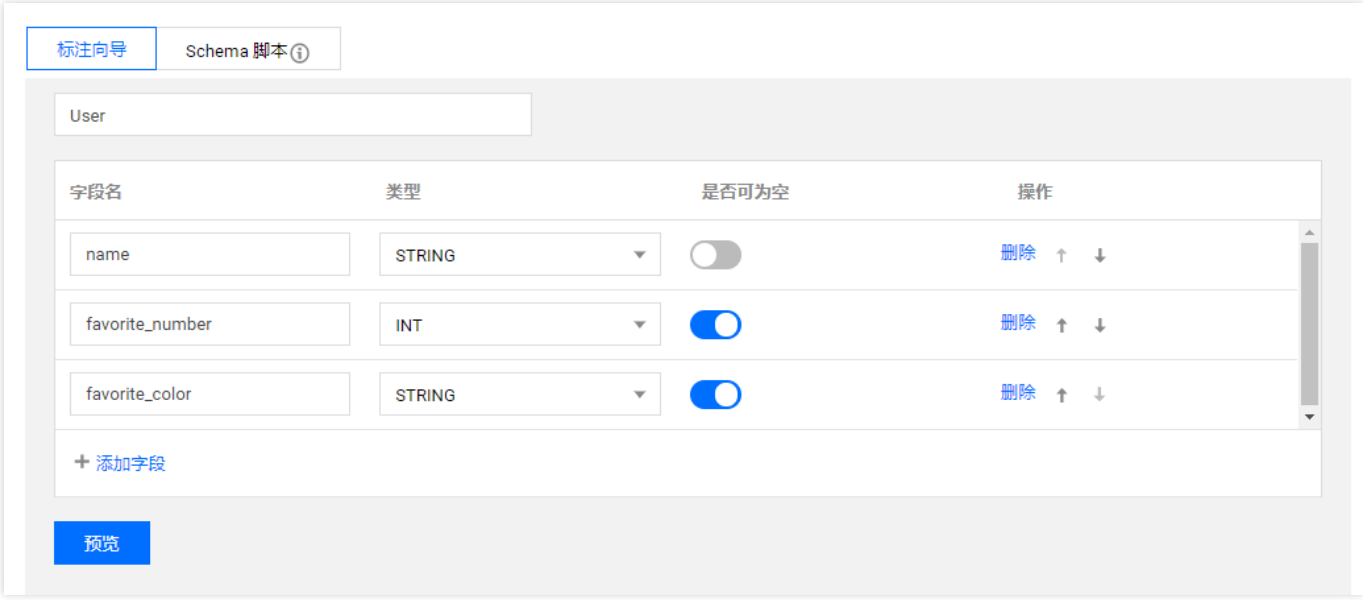
- 选择 Rcord Key：目前 Sparkling 只采用用户自定义 Schema 解析 Value 部分。

9. 指定数据 Schema 格式。

   AVRO 格式目前支持 Web 标注向导或上传 Schema 脚本方式指定，两种方式可以相互切换。

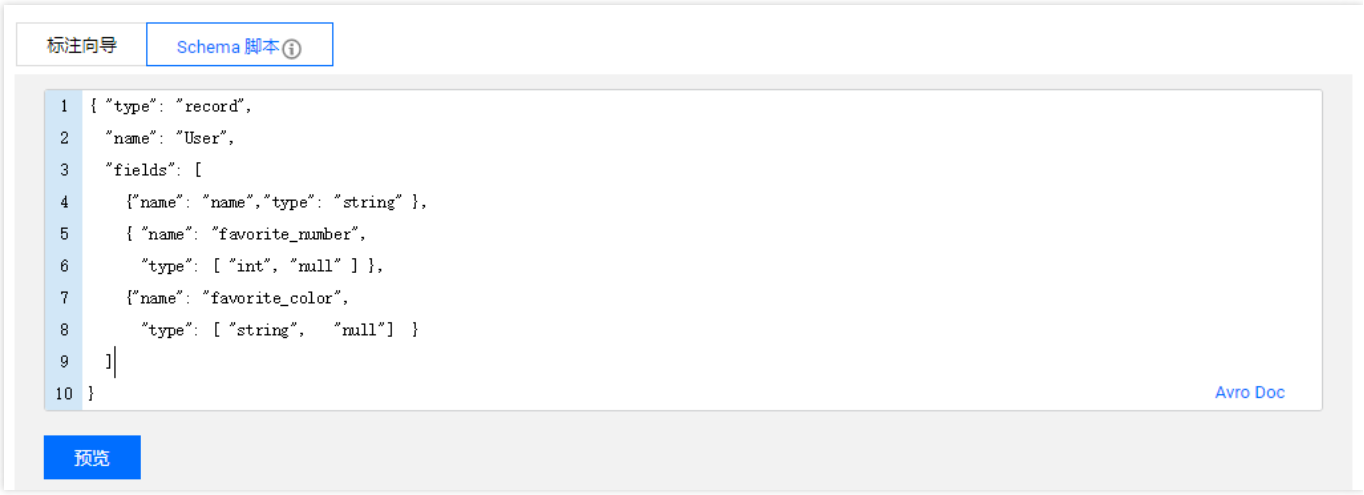
   - 标注向导

     用户需要输入Schema 名称和添加字段。字段添加完成后会自动生成 Schema 脚本，可单击【Schema 脚本】查看。



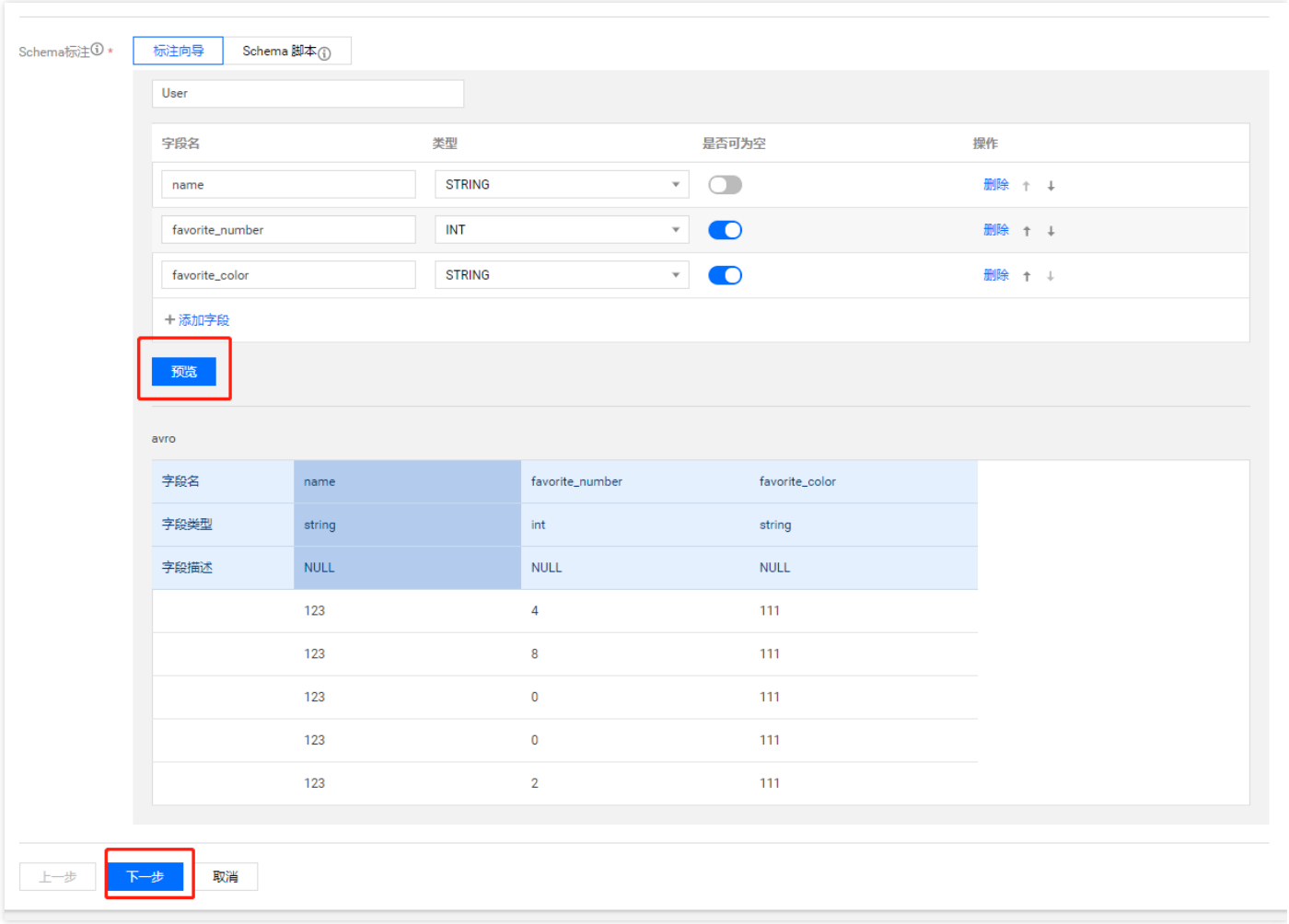
   - Schema 脚本

     用户可手动输入Schema 脚本。手动输入完成后，会自动在【标注向导】生成对应字段。



说明：
JSON 和 CSV 格式目前只支持用户手动添加字段。CSV 格式下可以在第四列输入该字段所处序列（排序默认从0开始）。

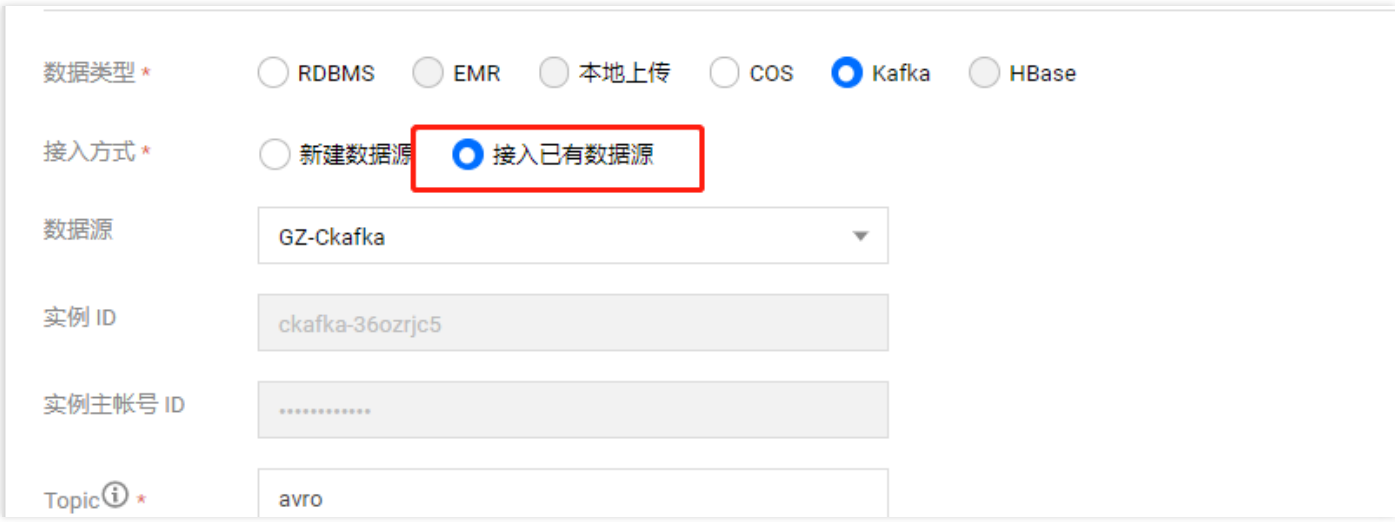
0. 填写完毕后单击【预览】可以查看数据预览，预览无误后单击【下一步】完成数据源配置操作。



接入已有数据源方式

1. 接入方式选择【接入已有数据源】。
2. 选择历史保存的已有数据源。

3. 输入 Topic 后的操作同“新建数据源方式”的第8 - 10步。

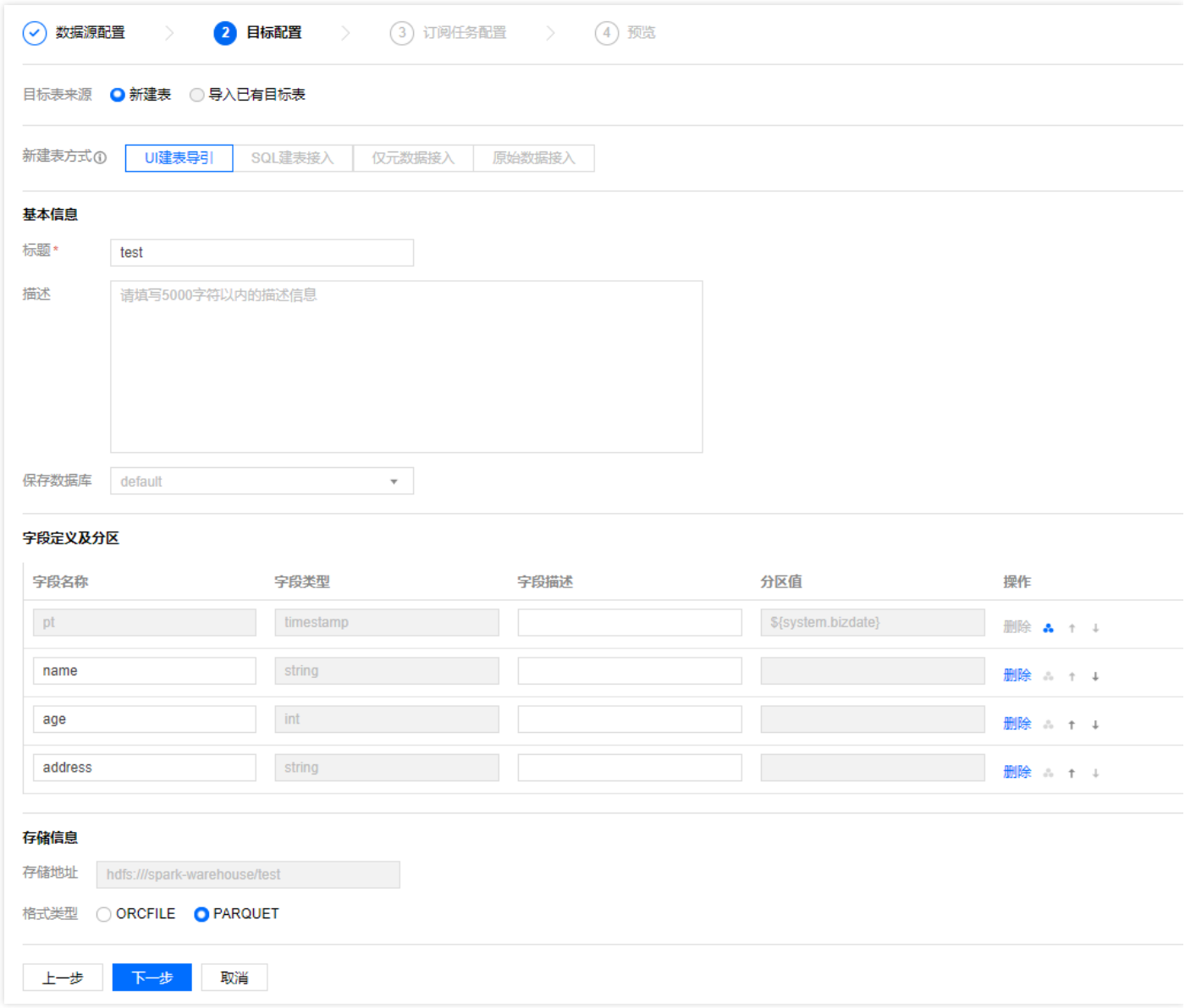


2. 目标配置

目前支持【新建表】方式目标配置。

- 选择【新建表】并填写【标题】和【描述】。
- 字段定义及分区部分可以设置字段名、描述及字段顺序，支持字段裁剪（可通过删除字段实现仅导入部分字段功能）。其中第一列“pt”列为默认分区字段，不支持删除操作，默认以该时间戳作为分区值。
- 选择格式类型，支持【ORCFILE】和【PARQUET】两种格式。

- 确认无误后，单击【下一步】完成新建表方式目标配置操作。



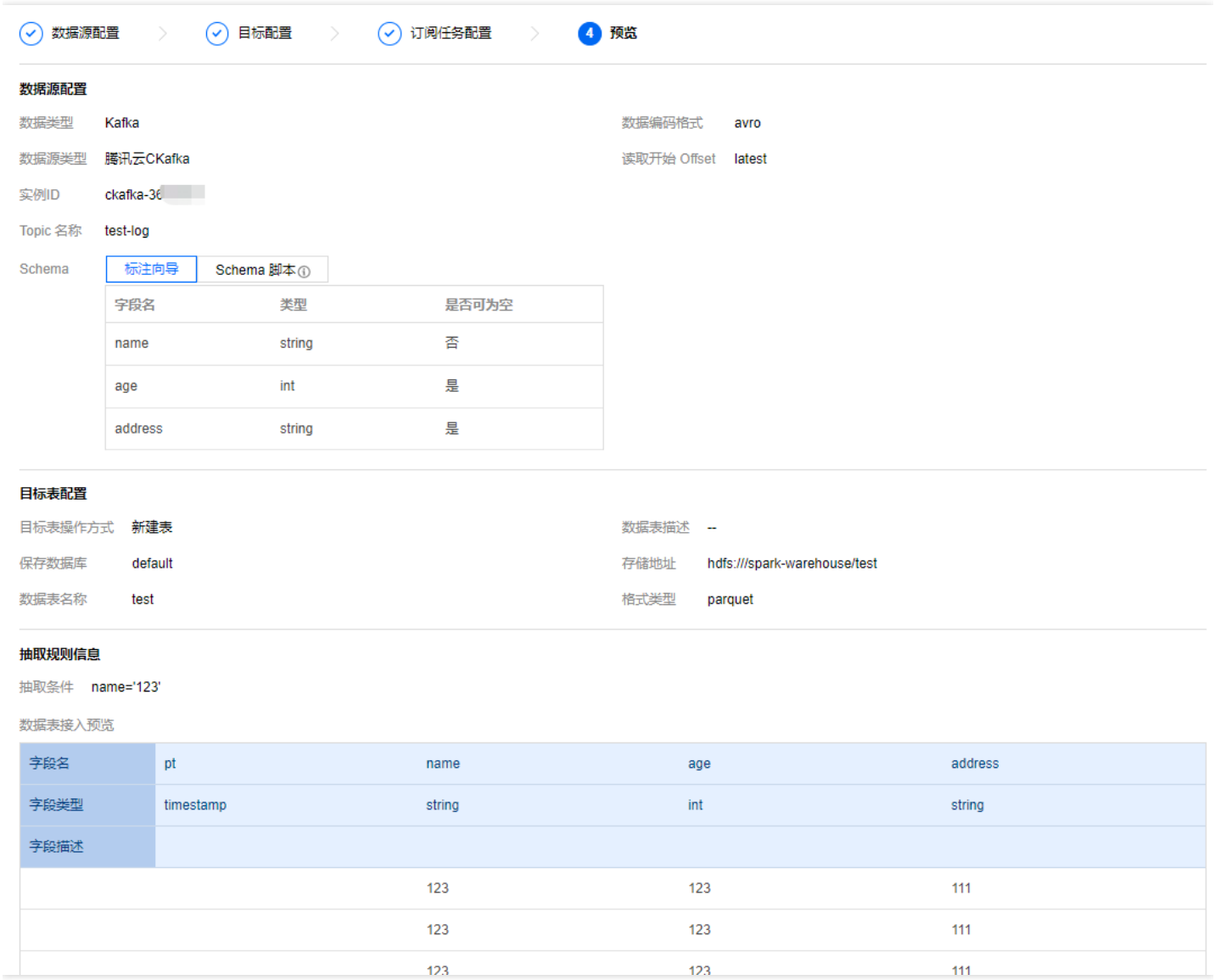
3. 订阅任务配置

- 填写抽取条件。参考 SQL 语法填写 WHERE 后的过滤语句，不包括“WHERE”字段，如图所示筛选 name 为123的数据。
- 设置起始时刻。将从该时刻开始进行数据导入任务调度。
- 设置执行时长。设置该任务持续时间。

- 设置单次数据条目。设置任务单次导入的数据条目，选择不限制表示单次导入当前阶段的所有最新数据。

4. 预览

在【预览】页可以查看当前设置的数据源配置、目标表配置、抽取规则、订阅任务配置等信息。



确认无误后单击【完成】即可完成 Kafka 数据接入任务设置。

数据源管理

最近更新时间：2020-01-15 15:58:18

本节为您介绍通过 Sparkling 控制台管理数据源的相关操作。

操作步骤

登录 Sparkling 控制台，在左侧导航单击【数据】进入数据接入页面，并单击数据接入页面右上角的【数据源管理】进入数据源管理界面。

查看数据源和筛选数据源类型

单击【全部数据源】可查看当前集群下保存的所有数据源信息，在右侧下拉框中可筛选要查看的数据类型。例如选择 RDBMS，数据源管理列表中展示所有 RDBMS 数据源信息。

查看数据源名称、创建时间和数据源类型

可在第一、二、三栏中查看【数据源名称】、【创建时间】和【数据源类型】。

查看数据源详情

选择第四栏【链接信息】>【详情】，可查看【数据源名称】、【实例 ID】及数据库购买者【Uin】（实例购买者在账号信息 中的账号 ID）。

查看数据源连通状态

选择第五栏【连通状态】>【成功】，可查看当前数据源的连通性验证时间。连通状态包括【成功】、【失败】和【未知】三种，展现当前数据源所处连通性验证是否成功。

测试联通性和删除数据源

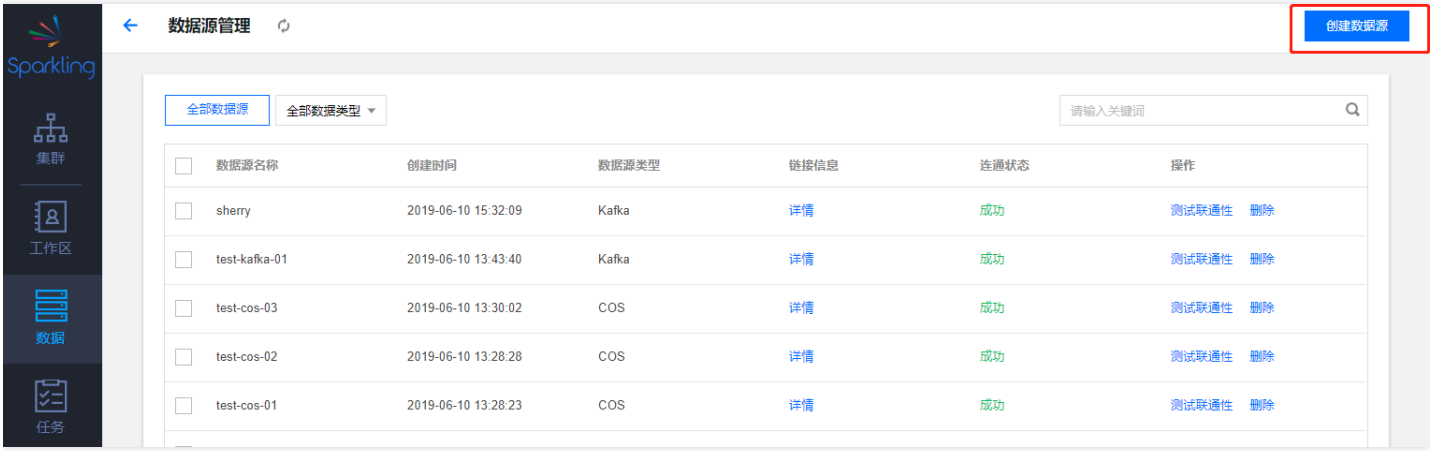
选择最后一栏【操作】>【测试联通性】可重新测试数据源联通性，单击【删除】可删除该单个数据源。

搜索数据源

页面右上角支持数据源全文搜索（除数据源状态）。通过输入包括数据源名称、创建时间、数据源类型、实例 ID、数据库购买者【Uin】进行相关数据源搜索。

创建数据源

单击页面右上角【创建数据源】，以进入创建数据源页面。



- **创建 RDBMS 数据源**

  a. 填写数据源名称，并选择数据类型为【RDBMS】。

  b. 选择 RDBMS 类型。RDBMS 数据类型包括腾讯云云数据库 MySQL 和腾讯云分布式数据库 TDSQL。

  c. 选择数据库所在的地域，对于跨地域数据库访问其稳定性和速率将会受到地域制约。

  d. 填写实例 ID（可在 云数据库控制台 获取云数据库实例 ID；可在 分布式数据库控制台 获取分布式数据库实例 ID）及数据库购买者 UIN（实例购买者在 账号信息 中的账号 ID）。

  e. 选择【创建者UIN】>【服务角色授权】，授权 Sparkling 服务访问其他相应服务。

  f. 填写要接入的数据表所在的【用户名】、【密码】、【数据库名】。

  g. 选择【连通测试】>【测试连通性】确认是否可以连接到所填写的数据库。

h. 待显示【数据连通正常】后，单击【保存数据源】完成创建 RDBMS 数据源。

- **创建 COS 数据源**

  a. 填写数据源名称，并选择【COS】数据类型。

  b. 地域：选择 COS 存储桶所在地域。

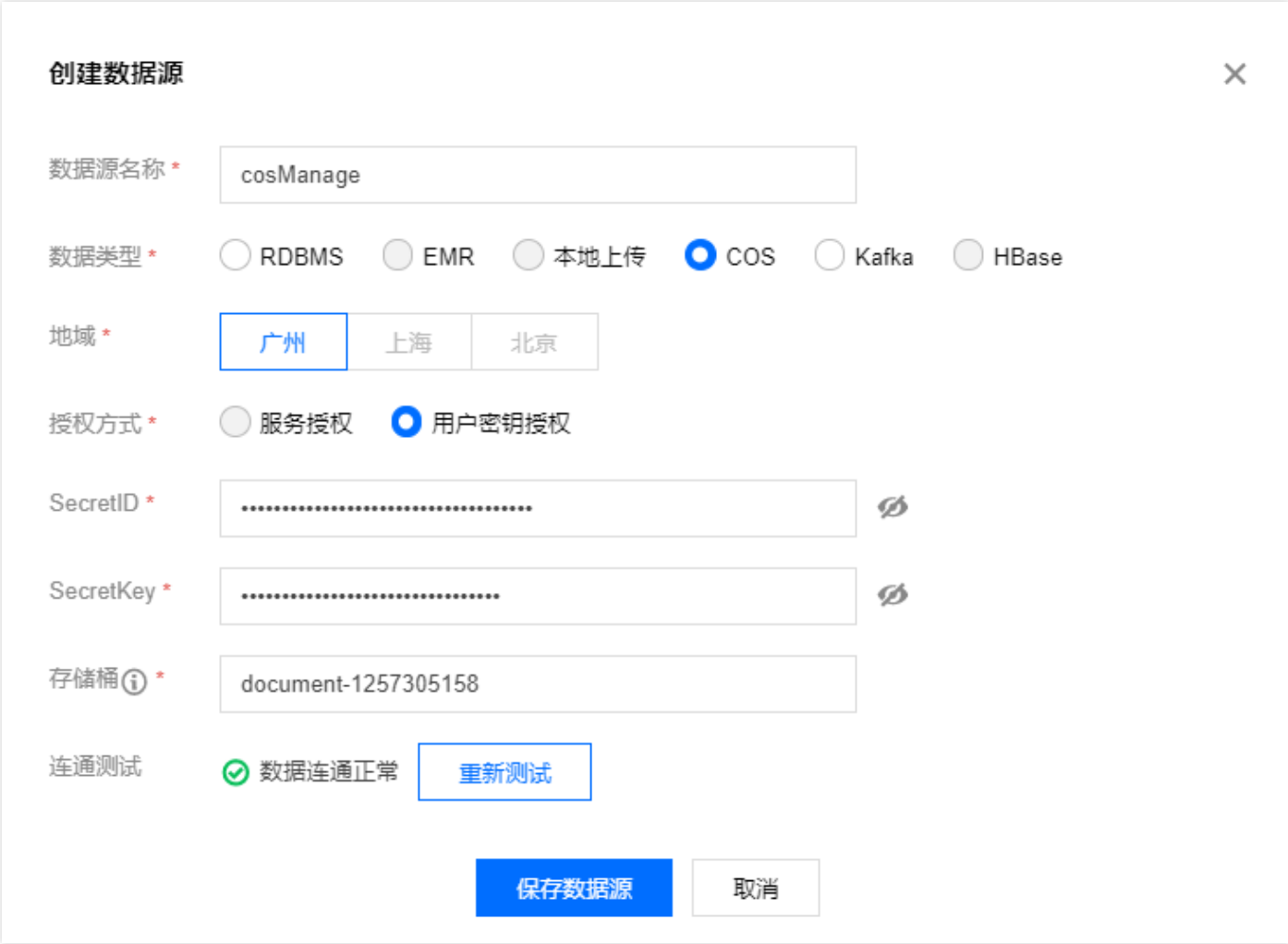
  c. 授权方式：选择用户密钥授权。

  d. SecretID/SecretKey：填写您已生成的密钥，可在 API 密钥管理 中生成并查看。

  e. 存储桶：填写您在 COS 中已生成的存储桶名称和您的 APPID。

  说明：
  存储桶名称需要按<目标存储桶名称-APPID>格式填写，例如 sparkling-12334513，桶名和 APPID 可在账户信息中查看。

f. 选择【连通测试】>【测试连通性】确认是否可以连接到所填写的数据库。
g. 待显示【数据连通正常】后，单击【保存数据源】完成创建 COS 数据源。

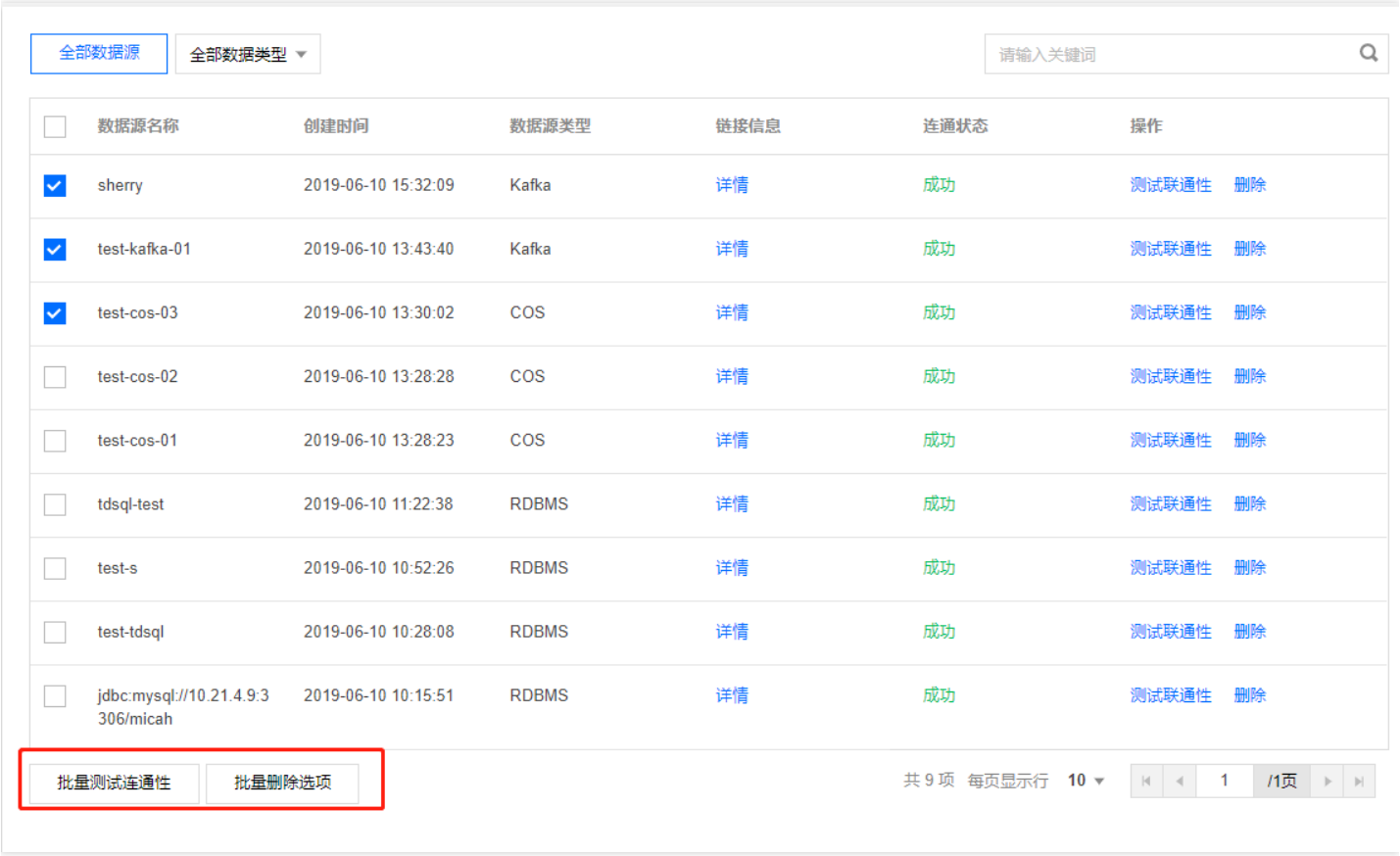


- **创建 Kafka 数据源**
  a. 填写数据源名称，并选择【Kafka】数据类型。
  b.填写腾讯云 CKafka 实例 ID（可在 CKafka 控制台 获取）及 CKafka 创建者 UIN（实例创建者在 账号信息 中的账号 ID）。
  c. 选择【创建者UIN】>【服务角色授权】，授权 Sparkling 服务访问其他相应服务。
  d. 选择【连通测试】>【测试连通性】确认是否可以连接到所填写的数据库。

e. 待显示【数据连通正常】后，单击【保存数据源】完成创建 Kafka 数据源。

批量测试连通性和批量删除

当同时选中多个数据源时，页面底部可以进行数据源【批量测试连通性】和【批量删除选项】操作。



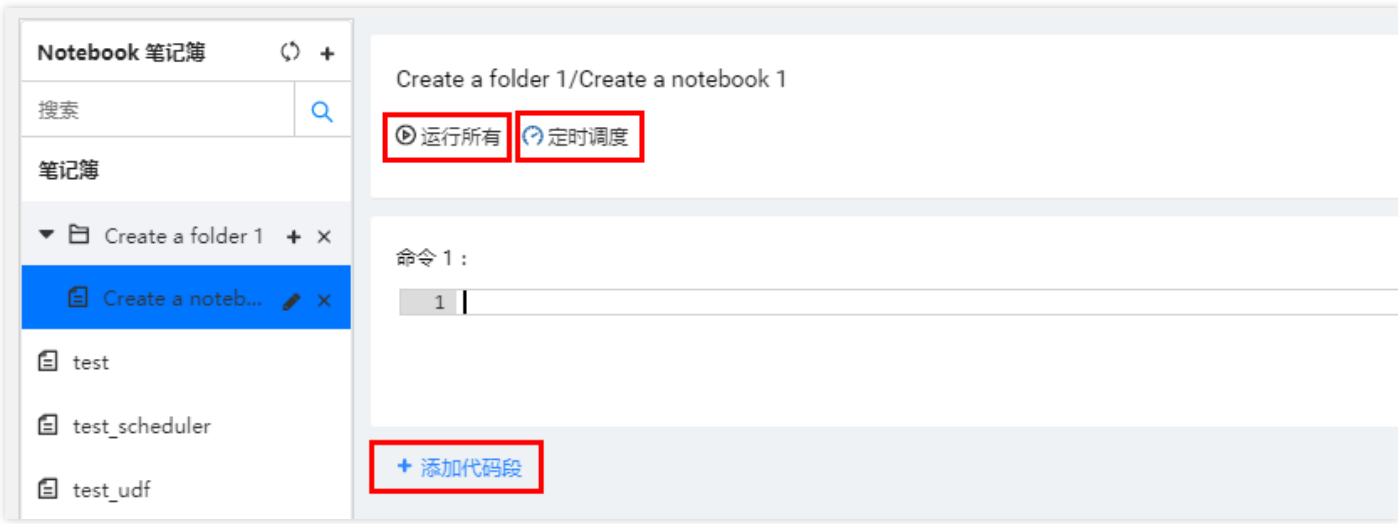
数据开发

最近更新时间：2020-01-15 16:20:13

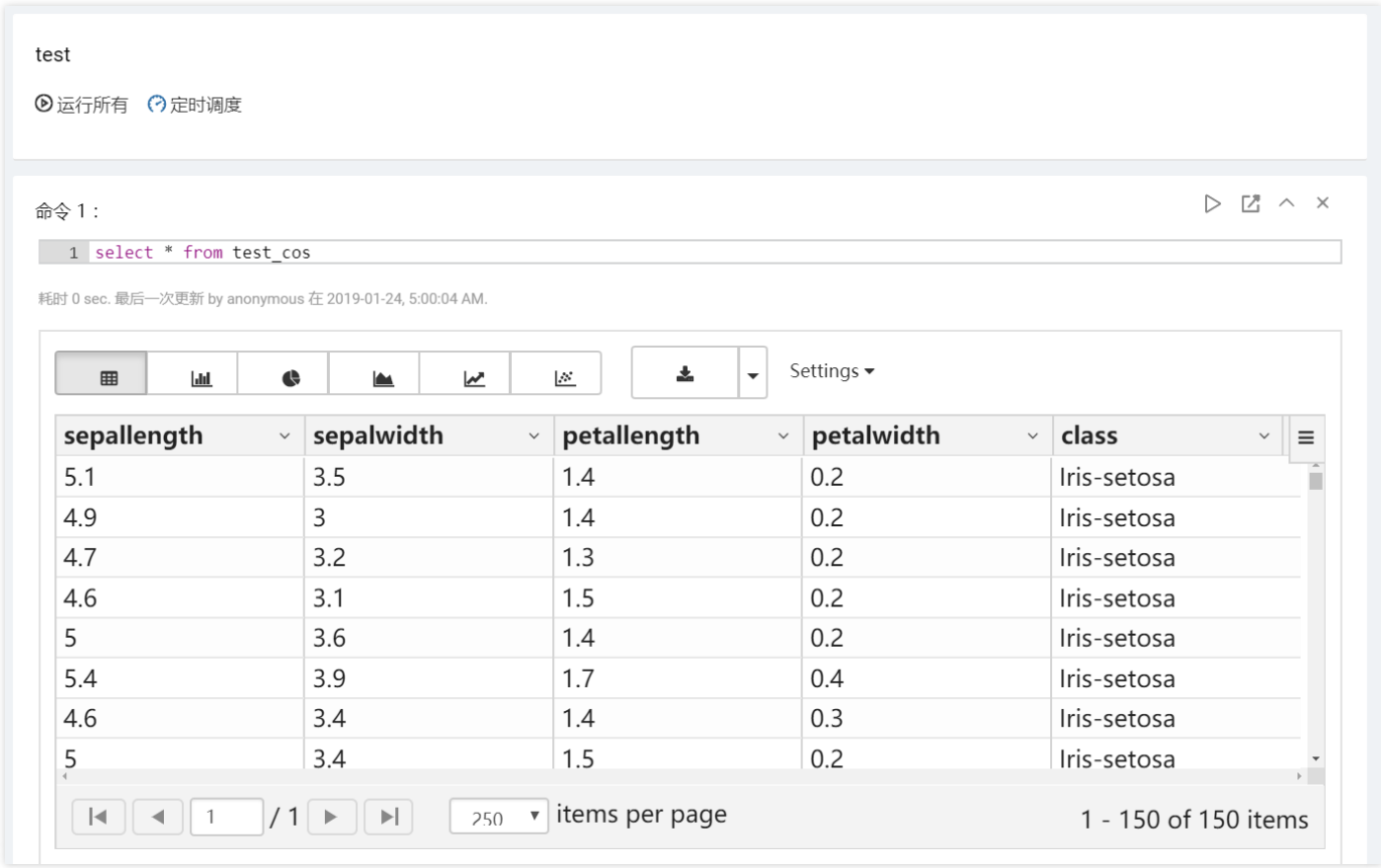
云数据仓库套件 Sparkling 提供数据开发 IDE 供用户对数据进行 ETL 、清洗加工和计算等操作，同时支持任务定时调度功能。笔记簿（Notebook）是数据开发 IDE 的核心功能组件，通过使用笔记簿可方便用户进行任务管理。

操作步骤

1. 进入 集群管理 页面，在左侧导航单击【工作区】进入数据开发页面。
2. 管理笔记簿，在左侧栏可对笔记簿进行新建、刷新、重命名和删除操作。



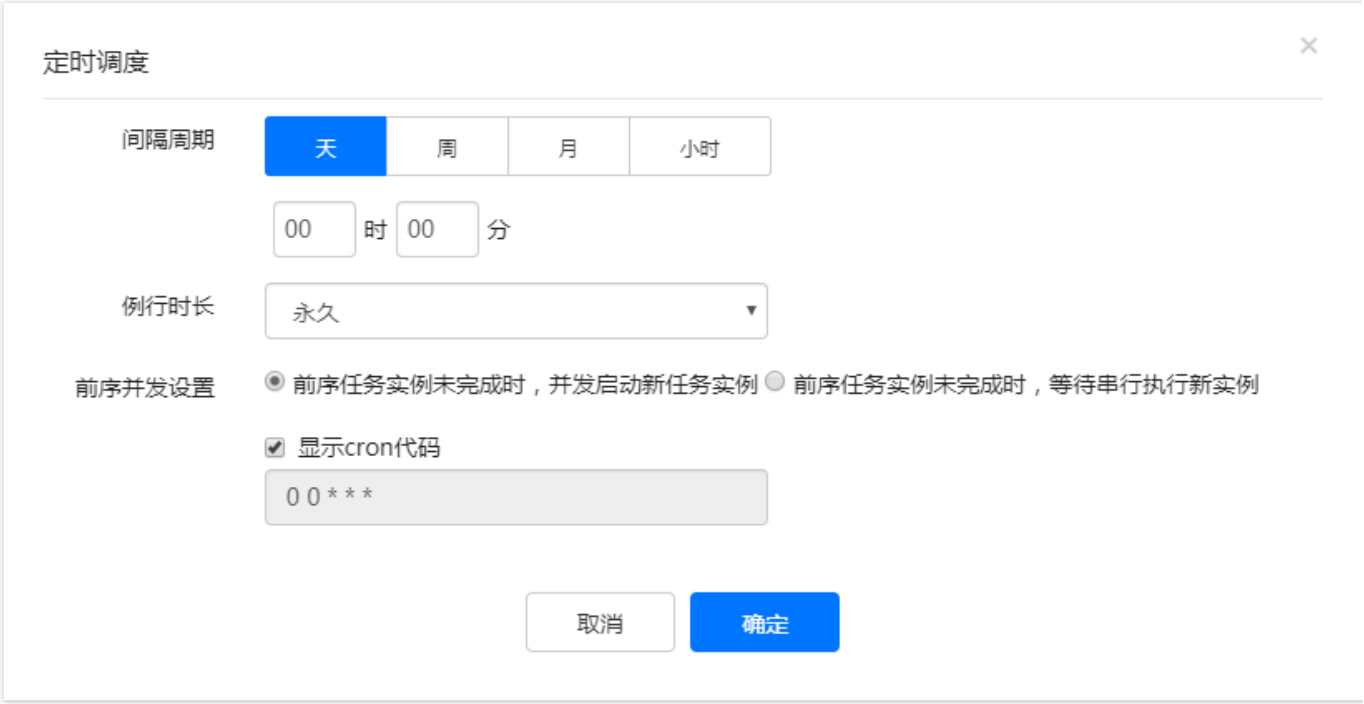
3. 编辑执行命令。



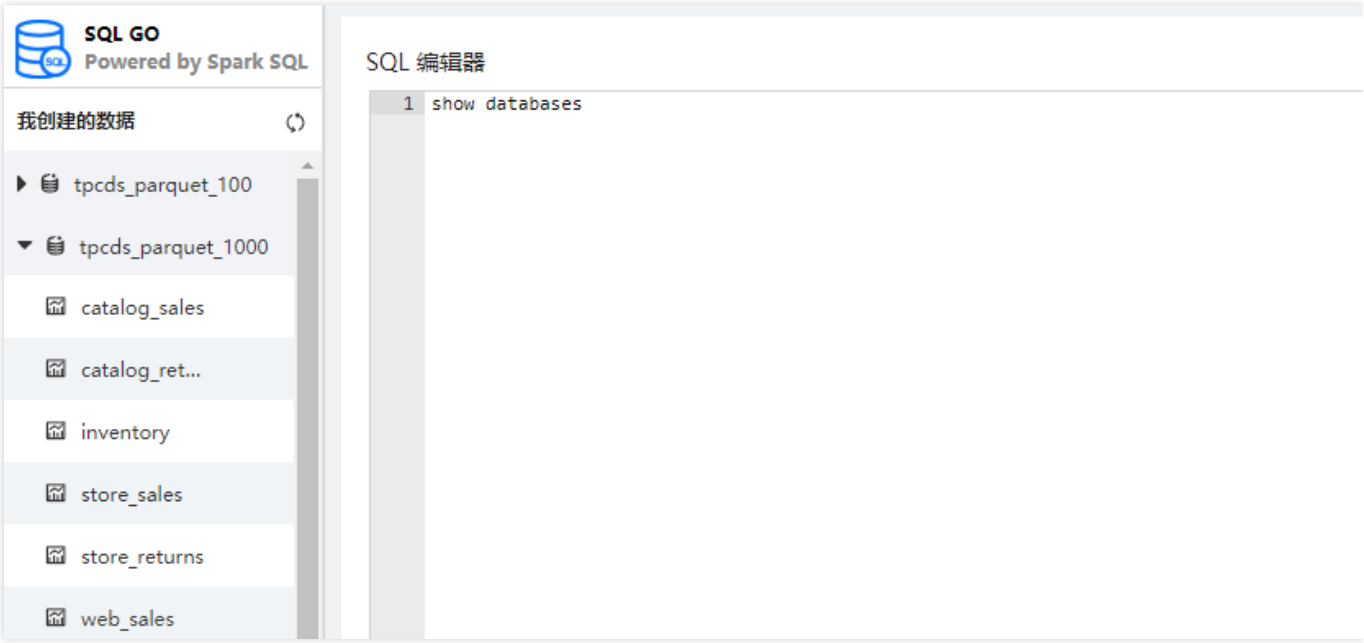
- 运行并查看执行结果。在命令栏中输入 SQL 语句后，在右上角运行或使用快捷键 Shift+Enter 查看当前语句执行结果。
- 可视化查询结果，可通过单击以下的图标对查询结果进行可视化操作，如绘制柱状图、饼图等。

- 下载查询结果，支持 CSV 和 TSV 两种格式的数据。
- 添加代码段，单击左下角【添加代码段】。
- 运行所有命令，单击左上角【运行所有】，可依次执行当前笔记簿下所有命令栏中的语句。
- 定时调度，单击左上角【定时调度】，弹出定时调度设置框，可选择间隔周期并定义调度时间，例如“天”为00时00分，即每日0点0分自动开始执行任务。选择例行时长及前序并发设置项后单击【确定】，即可完成定时调度设置。可在页面左侧【任务】栏查看该 notebook 任务具体细节。

> 说明：
> 此处调度周期设置完成后，任务第一次执行是开始时间 + 一个调度周期。

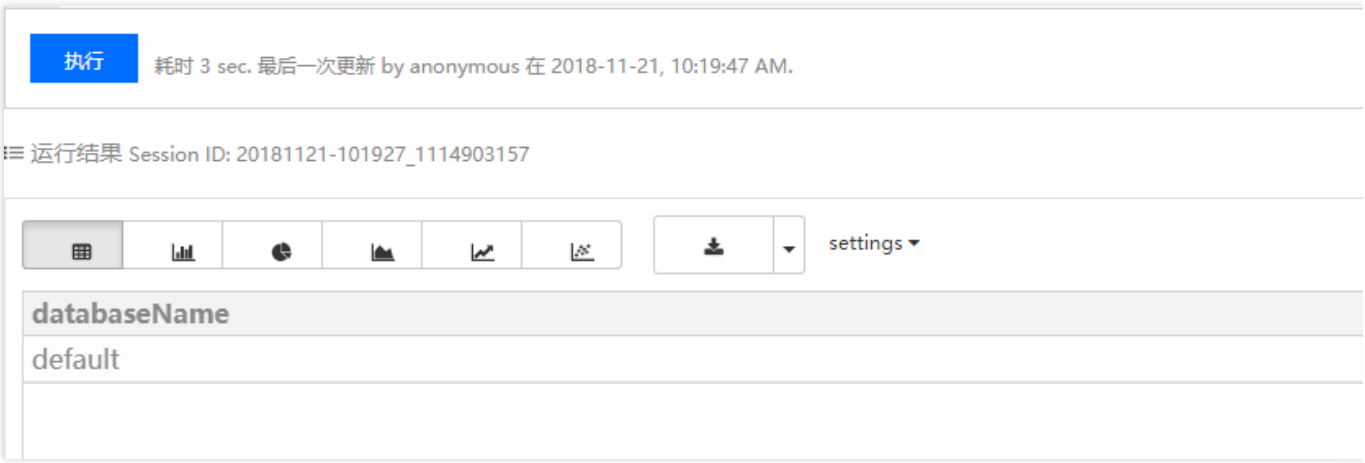


4. 使用 SQL IDE。通过 SQL IDE 可以查看数据表信息，完成数据的管理、查询、下载等操作。
   - 单击右上角 进入 SQL IDE 界面。
   - 查看数据目录。在 SQL IDE 界面左侧导航栏可以查看当前集群下所有创建和授权的数据。



   - 编辑 SQL 代码。在右侧 SQL 编辑器中可编辑要执行的代码。目前 SQL 支持 DDL 和 DML 语法规范，并完全兼容 ANSI SQL 2003。编辑完成后单击【执行】，可查看耗时、更新时间和运行结果，完成数据可视化和下载

等操作。


执行
耗时 3 sec. 最后一次更新 by anonymous 在 2018-11-21, 10:19:47 AM.
运行结果 Session ID: 20181121-101927_1114903157
settings
databaseName
default

任务管理

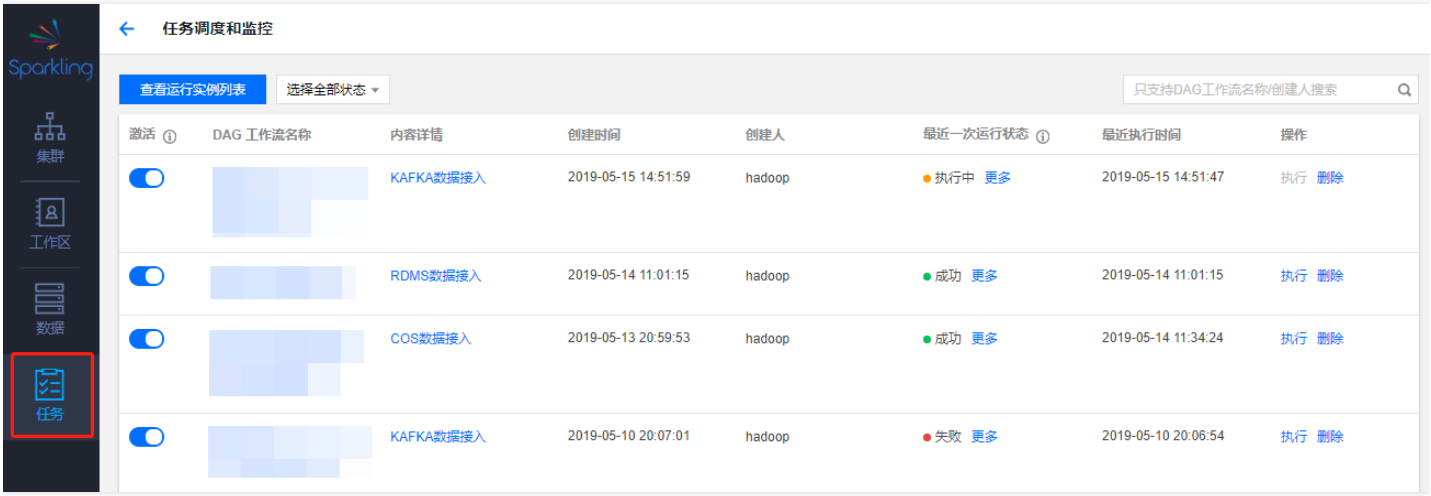
最近更新时间：2019-09-26 10:59:40

操作场景

云数据仓库套件 Sparkling 允许用户管理和调度任务，任务类型包括笔记簿开发任务、数据集成接入任务，用户可以在任务管理界面选择是否开启任务、是否立即执行该任务、查看任务运行时间及任务运行细节信息等。本节为您介绍通过 Sparkling 控制台管理执行任务的相关操作。

操作步骤

登录 Sparkling 控制台，在左侧导航单击【任务】进入任务管理页面。其中笔记簿开发任务以“notebook_”起始，数据集成接入任务以“datasync_”起始。



- **开启\关闭任务**

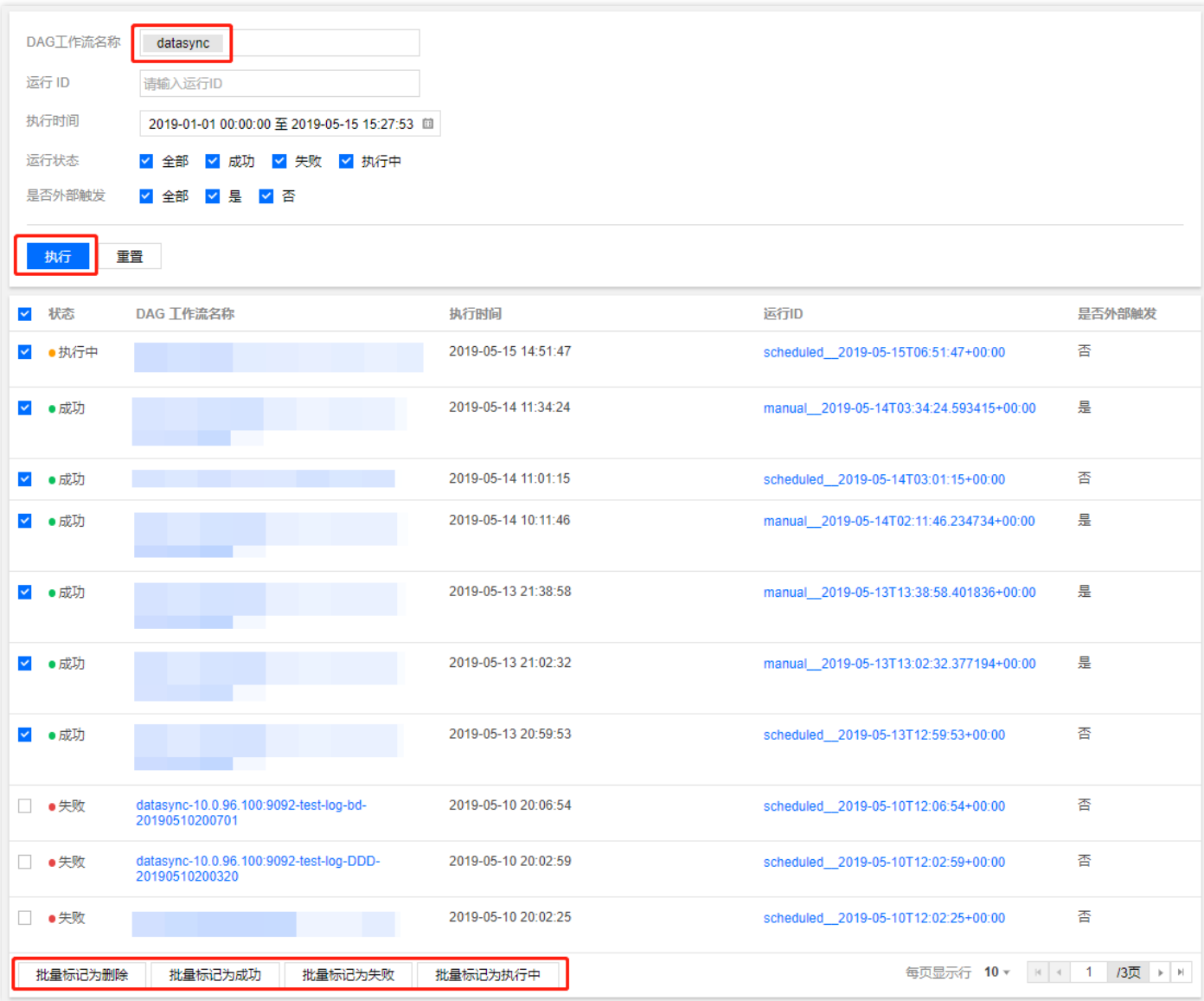
  单击页面第一栏激活按钮可切换任务状态，  表示任务处于开启状态，  表示任务处于关闭状态。当任务处于关闭状态时，RDBMS、COS 数据导入任务的当前执行中任务不受影响，下一次执行将会被暂停，而 Kafka 数据导入任务将会被立即暂停。
- **立即执行\删除任务**

  单击最后一栏的【执行】可以手动触发立即执行该任务，单击【删除】可以删除该任务，执行中的任何任务都将被立即终止。
- **查看任务详情**

  单击第二栏 DAG 任务工作流名称，可查看任务 DAG 详情。

单击第三栏内容详情可以查看任务细节，其中 Datasync 任务可以查看数据源信息、目标表信息、抽取规则信息、定时任务信息等，Notebook 任务可以跳转到该 Notebook 页面。

- **查看任务创建时间、创建人、最近执行时间**
  通过第四、五、七栏可以查看任务创建时间、创建人、最近执行时间。
- **查看任务状态**
  通过第六栏可以查看任务最近一次运行状态。
- **搜索任务**
  按字段搜索：页面右上角支持任务字段搜索，通过输入 DAG 工作流名称或创建人进行相关任务搜索，支持搜索部分字段。
  按条件搜索：单击左上角【查看运行实例列表】，输入 DAG 工作流名称（支持关键字联想）或运行 ID，执行时间、运行状态、是否外部触发等筛选条件，单击【执行】进行条件搜索。



页面底部可以批量标记任务状态。您可以勾选要修改的任务后单击批量标记，从而实现批量修改任务状态的功能。您也可以单击【批量标记为删除】批量删除任务记录。

个人中心

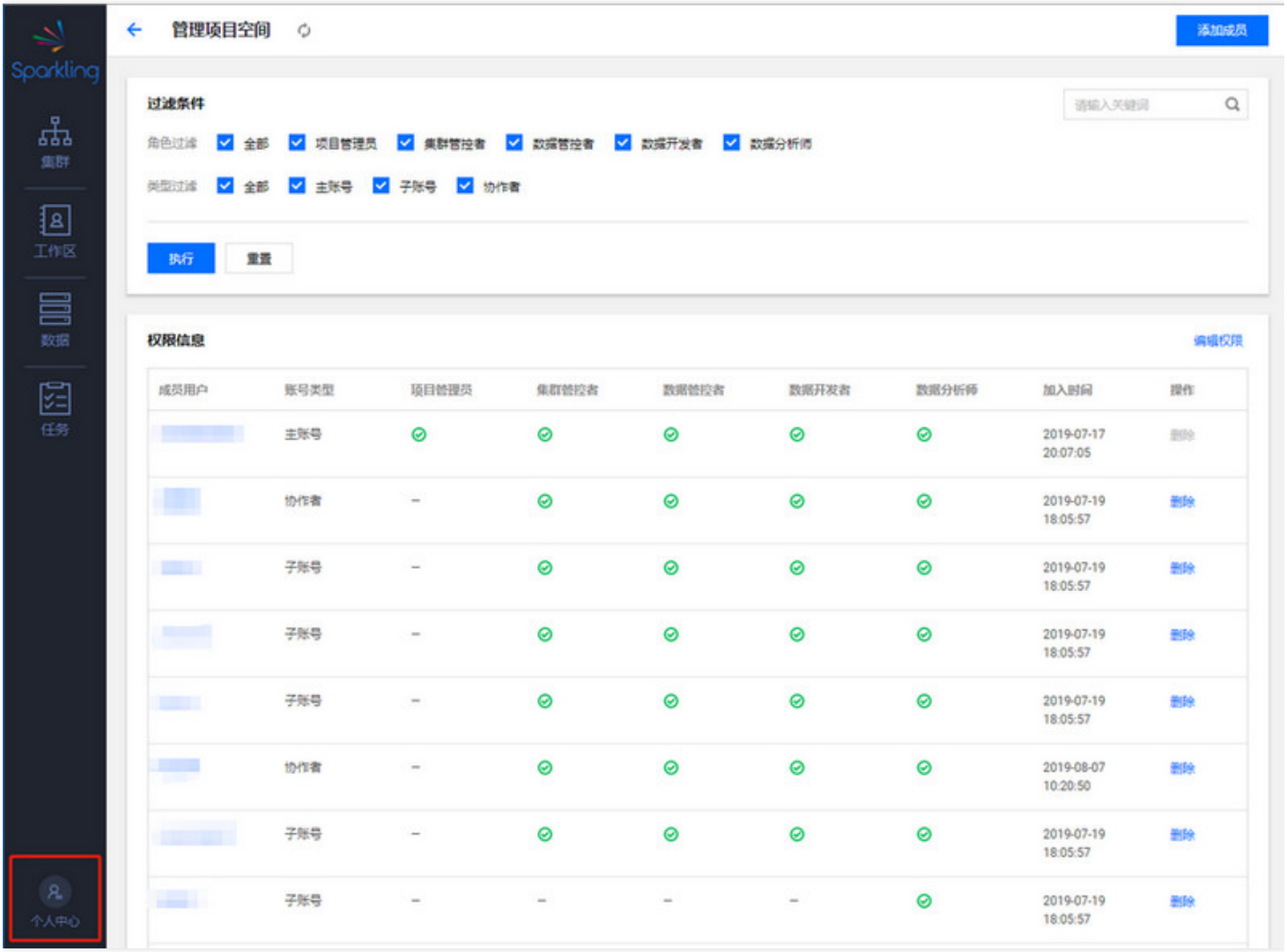
最近更新时间：2020-01-15 16:42:48

操作场景

云数据仓库套件 Sparkling 个人中心支持项目管理员进行项目空间管理，主要包括成员管理和权限管理两部分。本节将为您介绍通过成员及权限管理项目空间的方法，其中涉及的 Sparkling 账号登录、权限及角色详情，请参见 Sparkling 多租户策略。

操作步骤

登录 Sparkling 控制台，在左侧导航单击【个人中心】进入项目空间管理页面，按以下操作步骤完成项目空间管理：



添加成员

添加成员功能包括添加已有子账号（包括子用户和协作者）或新建子账号后添加至项目空间。

- **添加已有子账号**

  a. 单击项目空间管理页面【添加成员】。

  b. 选择待添加账号。

c. 单击　　将所选账号移动到【已添加账号】。

d. 选择账号权限。

e. 单击【确定】完成添加已有子账号。

- **新建子账号并添加至项目空间**

a. 单击【CAM 控制台】进入 CAM 控制台。

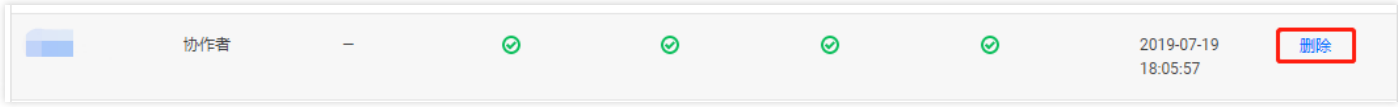
b. 详细控制台添加子账号操作请参见 CAM 访问管理。

c. 返回 Sparkling，单击刷新同步账号信息。

d. 后续步骤同上“添加已有子账号”的步骤 b - e。

删除成员

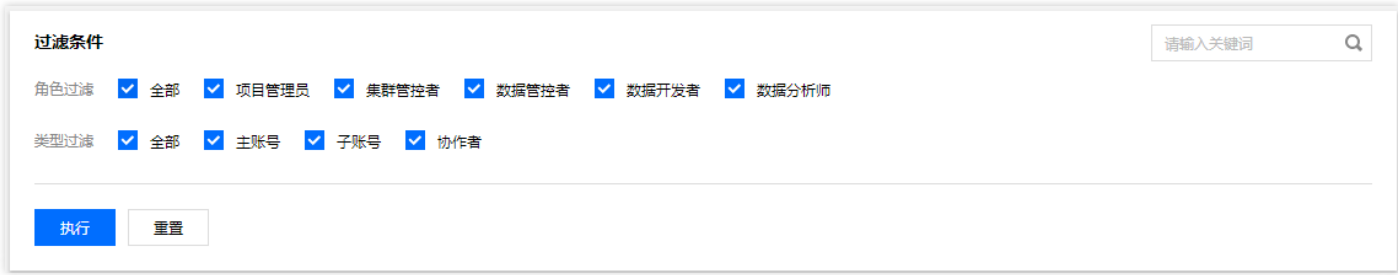
- 在权限信息栏找到相关账号用户，确认用户信息。确认无误后单击操作栏中的【删除】。



- 在删除用户页面，单击【删除】完成删除成员操作。

筛选成员

- 通过设置【角色过滤】和【类型过滤】条件，设置完成后单击【执行】查看筛选结果。
- 在右上角搜索框中通过成员用户关键词查询。



权限管理

Sparkling 将权限附加到角色，使角色拥有对 Sparkling 操作和资源的权限。Sparkling 中支持的用户角色包括项目管理员、集群管控者、数据管控者、数据开发者、数据分析师。Sparkling 角色与权限详情，请参见 Sparkling 多租户策略。

Sparkling 权限管理操作步骤如下：

1. 在项目空间管理页面单击【编辑权限】。

2. 选择要修改权限的账户。单击每列角色名称下对应的单选框，进行用户权限的设置和取消。 表示成员用户无此权限， 表示成员用户拥有此权限。

说明：
Sparkling 不支持编辑主账号权限。

3. 单击右上角【保存】，完成编辑权限操作。