

XT00Y01 数据同步系统使用手册V1.0

一、概述

二、功能支持

三、操作步骤

3.1 任务注册

3.1.1 选择源实例

3.1.2 选择目标实例

3.1.3 输入目标表信息

3.1.4 输入源表信息

3.1.5 查看注册结果

3.2 任务调度配置

3.2.1 查询注册的任务

3.2.2 查看定时任务

3.2.3 新增定时任务

3.2.3.1 定时任务类型☆☆☆

3.2.3.2 配置增量定时任务

一、概述

在数仓项目中，一直有数据同步的需求，在使用过SAP Data Service, Apache Sqoop, Apache Spark, Apache Flink等工具后，总结出了一些数据同步的方法，现在初步用一个低代码系统实现了其中一部分功能和需求。

二、功能支持

以下是离线同步：

- 1、SAP HANA 和 Apache Doris之间的双向同步
- 2、MySQL 到 Apache Doris的同步

三、操作步骤

3.1 任务注册

本系统中，注册任务时，源表和目标表均需要提前按照约定好的格式创建完成。

本小节的操作均在【XT离线任务配置】页面中完成。

XTUNNEL数据同步系统

任务注册

XT离线任务配置

XT任务定义查询

任务调度配置

任务日志查询

源库实例编号

源库名

源表名

源表增量字段

任务描述

目标库实例编号

目标库名

目标表名

1、查询时，仅支持输入右侧目标表信息

2、仅支持离线任务

3、注册任务时，左侧均是必填项

查询任务编号

注册任务编号

Q 搜索

过滤 下载

0条记录

第 1 页，共0页

任务定义编号	任务名	源实例ID	源库名	源表名	增量字段	目标表

3.1.1 选择源实例

SQL

1 PRD_DORIS_DW-数据采集代理账号

源库实例编号

源库名

请选择

Q doris

PRD_DORIS_DW-数据采集代理账号

3.1.2 选择目标实例

1 BI_HANA生产机(内网)_写账号

目标库实例编号

请选择

hana

目标库名

BI_HANA生产机(内网)_写账号

3.1.3 输入目标表信息

填入【目标库名】【目标表名】，然后点击【查询任务编号】，查看是否已经注册过。

注：

请写完整的表名，不是ADSO名字

源库实例编号PRD_DORIS_DW-数据采集代... x v

源库名

源表名

源表增量字段

任务描述

目标库实例编号BI_HANA生产机(内网)_写账号 x v

目标库名SAPABAP1

目标表名/BIC/AZRTF2B0172

1 填写目标信息

2 查询是否注册过

查询任务编号

注册任务编号

0 条记录

第 1 页, 共 0 页

任务定义编号	任务名	源实例ID	源库名	源表名	增量字段	目标实

3.1.4 输入源表信息

填入【源库名】 【源表名】 【源表增量字段】 【任务描述】，然后点击【注册任务编号】

注：

【源表增量字段】：是源表中具有增量特性的字段，一般是修改时间，在doris中使用doris__op__stamp即可；若是MySQL的源表，需要选择有独立B+索引的时间字段，若没有索引，需要推动源库去加索引。

【任务描述】：是任务名，一般写表名即可

源库实例编号PRD_DORIS_DW-数据采集代... X

目标库实例编号BI_HANA生产机(内网)_写账号 X

源库名ods_t_f2b_user

源表名ods_t_f2b_user_enterprise_business_li

源表增量字段doris__op__stamp

任务描述企业营业执照资料

1 填写源表信息

1、查询时，仅支持输入右侧目标表信息

2、仅支持离线任务

3、注册任务时，左侧均是必填项

2 点击注册任务

查询任务编号

注册任务编号

Q 搜索...

过滤 下载

0 条记录 第 1 页, 共0页

任务定义编号	任务名	源实例ID	源库名	源表名	增量字段	目标实例ID

3 再次查询任务注册结果

4

任务注册完成

源库实例编号

PRD_DORIS_DW-数据采集代... X

源库名

ods_t_f2b_user

源表名

ods_t_f2b_user_enterprise_business_li

源表增量字段

doris___op___stamp

任务描述

企业营业执照资料

目标库实例编号

BI_HANA生产机(内网)_写账号 X

目标库名

SAPABAP1

目标表名

/BIC/AZRTF2B0172

1、查询时，仅支持输入右侧目标表信息

2、仅支持离线任务

3、注册任务时，左侧均是必填项

一些不合法的注册会提示出来

查询任务编号

注册任务编号

Q 搜索...

过滤

下载

0条记录

第 1 页，共0页

任务定义编号	任务名	源实例ID	源库名	源表名	增量字段	目标实例

3.1.5 查看注册结果

再次点击【查询任务编号】，记住此时出现的【任务定义编号】，在后边的步骤中，有这个编号会更方便一些。

源库实例编号

PRD_DORIS_DW-数据采集代... X

源库名

ods_t_f2b_user

源表名

ods_t_f2b_user_enterprise_business_li

源表增量字段

doris____op____stamp

任务描述

企业营业执照资料

目标库实例编号

BI_HANA生产机(内网)_写账号 X

目标库名

SAPABAP1

目标表名

/BIC/AZRTF2B0172

1、查询时，仅支持输入右侧目标表信息

2、仅支持离线任务

3、注册任务时，左侧均是必填项

查询任务编号

注册任务编号

搜索

过滤

下载

1条记录

第 1 页, 共1页

再次查询注册结果，

需要记住这个【任务定义编号】

后面需要用到

3.2 任务调度配置

本小节均在【XT任务调度配置】页面

数据采集

你好, 张英杰

任务注册

任务调度配置

XT任务调度配置

任务日志查询

查询类型

PRD_DORIS

查询

搜索

过滤

下载

0条记录

第 1 页, 共0页

查看定时任务	新增定时任务	任务定义编号	任务名	源库名	源表名	目标库名	目标表名	有效状态

3.2.1 查询注册的任务

先查询出要配置调度的注册任务，支持以下五种方式查询注册的任务。

- 【源库表名】：勾选源库表名+输入框中写入正确的表名
- 【目标库表名】：勾选目标库表名+输入框中写入正确的表名
- 【任务定义编号】：勾选任务定义编号+输入框中写入任务ID（上一步骤中的任务定义ID）
- 【源实例ID】：勾选源实例ID，查询该实例下所有的注册任务
- 【目标实例ID】：勾选目标实例ID，查询该实例下所有的注册任务

以上五种方式任选其一。



3.2.2 查看定时任务

当出现我们要配置的注册任务时，先点击【查看】查看一下，是否已经在调度中的实例任务，是否有满足自己条件的任务，避免过度调度。



[illegible]

3.2.3 新增定时任务

当我们需要新增一些任务时，可以点击【新增】

调度任务类型分为三类：

【全量初始化】：需要选择初始条件，选择一个调度时间

【一次性】： 需要选择初始条件，选择一个调度时间

【增量定时】：需要选择初始条件，可以选择**【间隔定时】**或**【cron定时】**两种

另外还有一些配置需要做：

【小切片开关】 源头是mysql类型的，建议开启【制作】小切片；源头是Doris的目前还没有实现，暂时选择【不制作】

【切片有效分钟数】对于一些短时任务，有效期应该略大于调度周期，比如，20分钟运行一次的调度任务，切片有效期建议在20分钟左右，避免需要扎堆调度的问题。

其中，【全量初始化】和【一次性】都是一次性的任务，不会多次调度。



查看定时任务	新增定时任务	任务定义编号	任务名	源库名	源表名	目标库名	目标表名	有效状态
查看	新增	696	企业营业执照资...	ods_tf2b_u...	ods_tf2b_user_en...	SAPABAP1	/BIC/AZRTF2B0172	1

3.2.3.1 定时任务类型☆☆☆

本系统支持的定时类型有三种：

date：一次性，选择一个时间点，在执行后，本调度任务就此完结

cron：选择指定的时间点周期性地执行任务。比如选择cron hour='7,14,18'，cron minute='5'，就是在每天的7点5分，14点5分，18点5分分别启动一次任务。

interval：间隔一定时间周期性执行。比如选择interval type='hours',interval num='2'，就是每2小时执行一次。目前仅开放['hours','minutes']两种。

后台实现采用的是apscheduler.schedulers.background.BackgroundScheduler

有兴趣的可以去进一步了解

3.2.3.2 配置增量定时任务

配置任务时需要注意的事项

配置定时任务类型

×

DORIS 写到HANA的暂不支持切片

有效期一般匹配间隔周期，避免任务扎堆争抢

小切片开关

☐ 制作

☒ 不制作

切片有效分钟数

20

× v

数据同步类型 ?

☐ 全量初始化

☒ 增量定时

☐ 一次性

设置同步条件

doris___op___stamp >= now() - INTERVAL 3 DAY

调度条件需要改为30 minute

定时类型

☐ cron定时

☒ 间隔定时

间隔类型

分钟数

× v

间隔数量

20

× v

一般选择20分钟调度一次

关闭

提交

修改同步条件时，可以点击左侧的提示

▼

SQL |

```
1  【注】
2
3  以下语法仅在【doris】【mysql】进行过验证
4
5  - INTERVAL 1 MINUTE 表示前移1分钟
6
7  - INTERVAL 1 HOUR 表示前移1小时
8
9  - INTERVAL 1 DAY 表示前移1天
10
11 - INTERVAL 1 MONTH 表示前移1月
12
13 - INTERVAL 1 YEAR 表示前移1年
```

配置定时任务类型

【注】以下语法仅在【doris】【mysql】进行过验证
- INTERVAL 1 MINUTE 表示前移1分钟 - INTERVAL
1 HOUR 表示前移1小时 - INTERVAL 1 DAY 表示前
移1天 - INTERVAL 1 MONTH 表示前移1月 -
INTERVAL 1 YEAR 表示前移1年

不制作

切片有效分钟数20

数据同步类型

增量定时

一次性

设置同步条件

doris___op___stamp >= now() - INTERVAL 30 minute

定时类型

cron定时

间隔定时

间隔类型

分钟数

间隔数量

20

任务已加入任务池

配置定时任务类型

小切片开关

制作

不制作

切片有效分钟数20

数据同步类型

增量定时

一次性

设置同步条件

doris___op___stamp >= now() - INTERVAL 30 minute

定时类型

cron定时

间隔定时

注意，提交相同条件的任务会提示重复

间隔类型

分钟数

间隔数量

20

点击【提交】

关闭

提交

3.2.3.3 查看刚才的定时任务

查看任务时，有一些内容需要释义

11

定时任务实例

[illegible]

刷新日志

关闭

定时任务实例



cron定时使用的					间隔定时使用的		一次性定时使用的	
调度状态	定时类型	任务类型	cron小时数	cron分钟数	interval类型	interval数	预约时间	
DOING	interval	incremental	*	*	minutes	20		

【定时类型】是本调度任务选择的触发方式
cron是cron定时
interval是间隔定时
date是一次性

刷新日志

关闭

Q 搜索...

过滤 下载

1条记录 < 第 1 页, 共1页 >

切片制作开始时间	切片制作结束时间	制作秒数	制作状态	切片数量	同步开始时间	同步完成时间	同步分钟	同步状态	同步记录数
2024-01-20 18:07:59	2024-01-20 18:07:59	0	DONE	1	2024-01-20 18:08:09	2024-01-20 18:08:41	0.53	DONE	0
分为两段									
1、制作切片									
2、数据同步									