

hdfs导出DB

功能说明

hdfs数据导出到DB ,支持mysql,postgre

其他说明

任务设置

1. 基本信息

参考 [基本信息设置](#)

2. 调度

参考 [调度设置](#)

3. 参数

3.1 源服务器

待导入数据所在的hdfs server

更多信息参考 [服务器配置](#)

3.2 目标服务器

存储最终结果所在的HDFS server

更多信息参考 [服务器配置](#)

3.3 源目录

存放待导入DB数据所在hdfs 存放目录

3.4 分隔符

待导入DB数据在hdfs 字段分隔符

3.5 数据入库模式

支持两种方式，第一种是append，一种是truncate。

如果是append模式导入，重跑任务可能会出现重复数据。truncate 模式每次导入之前都会清空数据。

如果数据量比较大，推荐用户使用分区表，按小时，或者按天分区，并使用truncate 模式。这样大数据量的情况下，既不会出现重复删除大量数据，造成资源浪费，也不会影响用户数据的正常使用。

3.6 允许误差

整数值，表示允许失败的百分比。如果不填，表示不允许失败。

3.7 目标表

数据导出的目标表名，如果带有db名称，则使用db.table

3.8 目标表列名

数据导出的目标表表列名，改列名并不一定使用跟目标表列数一致，而是需要与hdfs 数据，通过分隔符切分的数据字段个数一致。多个列名，通过 逗号 , 分隔

3.9 是否分区

数据导出的目标表是否是分区表。

3.10 分区字段

指定数据导出的目标表的分区字段，只支持时间格式的分区，如按小时，天，月等。

3.11 读并发度

指定读取db 数据时使用的读并发线程数。

3.12 写并发度

指定写入hdfs 时使用的写并发线程数。

demo

demo资源