

mysql2udw: 把 Mysql 表格和数据导入到 UDW

一 mysql2udw 简介

mysql2udw 是参考 <https://github.com/philipsoutham/py-mysql2pgsql> 实现的把 mysql 的表格和数据导入到 udw 的工具。

二 mysql2udw 的安装

2.1 下载地址

<http://mysql2udw.ufile.ucloud.cn/mysql2udw.tar.gz>

2.2 安装

mysql2udw 安装依赖

- python (2.6 或者 2.7)
- setuptools
- mysql-python
- psycopg2
- pyyaml
- termcolor
- argparse
- pytz

首先我们需要安装好 python2.6 或者 python2.7 的环境。然后解压下载的 mysql2udw.tar.gz 进入 mysql2udw 目录, tools 里面包含了 mysql2udw 依赖工具。

如果是 centos 系统, 并且已经有 python 环境了, 可以执行 mysql2udw 目录的下的 install.sh 脚本可以帮助我们自动完成 mysql2udw 的安装。

如果是其他的 linux 系统

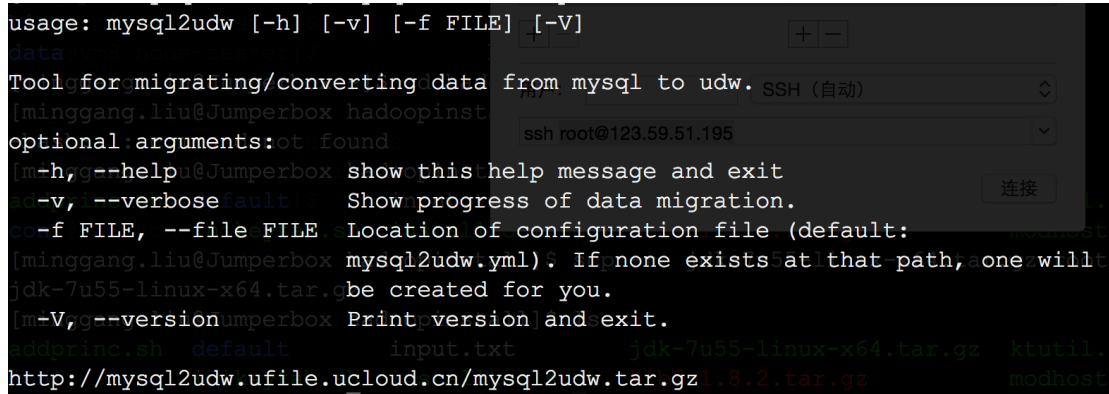
1. 安装 python
2. 请把 tools 的依赖手动安装,
3. 安装完成之后进入 mysql2udw/mysql-to-udw 目录执行 `python setup.py install`。

三 mysql2udw 的使用

3.1 mysql2udw 使用简介

mysql2udw 是通过执行配置文件来实现 mysql 到 udw 的具体数据迁移操作。在 mysql2udw 的解压目录下的 examples 目录有几个场景的配置模板。

输入 `mysql2udw --help` 我们也可以看到 mysql2udw 工具的使用说明



```
usage: mysql2udw [-h] [-v] [-f FILE] [-V]
Tool for migrating/convertng data from mysql to udw. SSH (自动)
[minggang.liu@Jumperbox hadoopinst ssh root@123.59.51.195]
optional arguments:
  -h, --help           show this help message and exit
  -v, --verbose        Show progress of data migration.
  -f FILE, --file FILE Location of configuration file (default:
                        mysql2udw.yml). If none exists at that path, one will
                        be created for you.
  -V, --version        Print version and exit.
http://mysql2udw.ufile.ucloud.cn/mysql2udw.tar.gz 1.0.2.tar.gz modhost
```

-v 可以展示数据迁移过程

-V mysql2udw 版本

-f 执行 mysql 到 udw 迁移的配置文件

例如: `mysql2udw_sample.yml` 配置文件如下所示, `mysql2udw_sample.yml` (注意把 mysql 和 udw 的相关配置修改自己 mysql 和 udw 地址) 配置完成之后执行 `mysql2udw -v -f mysql2udw_sample.yml` 就可以执行 mysql 到 udw 的迁移了。

mysql:

hostname: localhost

port: 3306

socket: /tmp/mysql.sock

username: mysql2psql

password:

database: mysql2psql_test

compress: false

destination:

file 如果配置了, 结果会输出到文件中, 否则直接输出到 udw, 一般直接写入 udw 效率
更高

file:

```

udw:
  hostname: localhost
  port: 5432
  username: mysql2psql
  password:
  database: mysql2psql_test

# 如何下面的 only_tables 配置，只会导入下面 table1 和 table2 的数据或者结构
#only_tables:
#- table1
#- table2

# 如果配置了 exclude_tables 除了下面的 table 之外的 table 都会被处理.
#exclude_tables:
#- table3

# tables_sql 支持自定义 sql，如果没有指定，则会导入整个表的所有字段的所有数据
#tables_sql:
# table1: select * from table1 limit 10

#如果 suppress_data 为 true，只会导入表结构到 udw，数据不会导入
suppress_data: false

# 如果 suppress_ddl 为 true 只导入数据到 udw，schema 不会导入
suppress_ddl: false

# 如果 force_truncate 为 true，导入表格之前会执行 truncate 操作
force_truncate: false

```

配置说明

mysql	连接 mysql 数据库相关的配置
file	mysql 数据导入到指定的文件，如果 file 不配置，数据直接写入 udw
udw	连接 udw 相关的配置
only_tables	指定需要导入的数据表
exclude_tables	排除要导入的表格
tables_sql	用户自定义的 sql
suppress_data	如果为 true，只导入表结构
suppress_ddl	如果为 true，只导入数据
force_truncate	为 true，导入表格之前会执行 truncate 操作

3.2 mysql2udw 使用说明

supress_data: true 只迁表结构

supress_ddl: true 只迁迁移数据

上述两个同时为 false 的时候，同时迁移表结构和数据

3.2.1 迁移整个数据库

参考 mysql2udw 安装文件下面 examples/all_tables.yml 配置

1.迁移一个数据库的所有表结构和数据

```
#only_tables:
```

```
#exclude_tables:
```

```
#tables_sql:
```

```
supress_data: false
```

```
supress_ddl: false
```

上面前面三个相关的配置注销，下面两个配置 false。

2. 只迁移一个数据库中所有表结构

```
supress_data: true, 其他配置和上述 1 保持一致
```

3. 只迁移一个数据库中所有表的数据

```
supress_ddl: true
```

3.2.2 过滤单个或者多个表格

参考 mysql2udw 安装文件下面 examples/exclude_tables.yml 配置

```
exclude_tables:
```

```
- table3
```

```
- table4
```

3.2.3 迁移一个或者多个表格

参考 mysql2udw 安装文件下面 examples/olny_tables.yml 配置

```
only_tables:
```

```
- table3
```

```
- table4
```

3.2.4 自定义 sql 迁移一个或者多个表格的数据

参考 mysql2udw 安装文件下面 examples/user_sql.yml 配置

mysql2udw 自定义 sql 为了方面进行增量的数据迁移

tables_sql 配置项提供了用户自定义 sql 迁移表格数据, 要实现自定的 sql 迁移表格的数据需要注意以下事项。

1. mysql 和 udw 必须已经创建好 tables_sql 下面的表格, 要求表格和自定义 sql 中有相同的字段。

2. 需要配置 only_tables 和 tables_sql

3. supress_ddl 配置为 true

4. tables_sql 可以是单表查询、也可以是多表的联合查询, 可以是数据库表的所有字段也可以是部分字段

only_tables:

- table3

- table4

- table5

tables_sql:

table3: select id,time,product from table1 limit 10

table4: select * from table1 left join table2 on table1.id = table2.id

table5: select * from table3 where start_time>=1400000000 and start_time<1500000000

supress_ddl: true

3.2.5 自定义 sql 实现高级特性

1. mysql 数据的增量导入到 udw

only_tables:

- table1

tables_sql:

table1: select * from table3 where start_time>=1400000000 and start_time<1500000000

supress_ddl: true

如上所示每次可以通过一个或者多个字段范围控制导入 udw 的数据

2. mysql 中数据分析结果导入到 udw

场景：把 mysql 中 table1 和 table2 联合查询的结果导入 udw。

1) 在 mysql 和 udw 分别创建一个表格 t_result

2) 配置配置文件

```
only_tables:
```

```
- t_result
```

```
tables_sql:
```

```
  t_result: select * from table1 left join table2 on table1.id = table2.id
```

```
supress_ddl: true
```

3) 执行 `mysql2udw -v -f test.yml`