

HIVE automation (sqoop.jar)

The purpose of the program is to sqoop tables from **DM to hive** and/or **execute .HQL files** on hive **daily and in parallel**. The schedule should be specified in the **schedule.txt** file:

Located at lynxkite@lynxkite01. The file looks like this:

```
ZSHHOURLYGGSN ggsn.ggsn_daily_tmp ggsn.hql 14 00
ZSHHOURLYGGSN ggsn.ggsn_daily_tmp - 14 56
ZSHUSGVASCS5 - vas.hql 16 44
//example usage.example_final_tmp example.hql 18 10
```

Every line is responsible for uploading a table or executing an HQL file every day at a given time. The lines should look like (left to right)

- basename* of the source table (*example*)
- actual name of the target table including it's schema (*usage.example_final_tmp*)
- name of HQL file (*example.hql*)
- uploading/executing time should be specified (*18 10* = 18:10).

The prepared files on DM should look like this: *example20160526*. The program will always find the newest table with the specified basename.

Line comment: //

The lines are executed in **PARALLEL**, so multiple lines can have the same time.

So each line does two things **SQOOPING** and **EXECUTING**. Either one of these steps can be skipped by replacing the tablename or .hql file's name with a '-'.

A casual use is to sqoop a table to a _tmp table and then inserting that into the final table. The insertion should be written in the called .hql file.

Only executing but in a smart way:

You might only want to schedule an HQL command. So you can write a line like this:

```
- - example.hql 12 23
```

But usually you will need the table's date that you are working with. If you write a table name instead of the first dash the program will use the newest date-extension for that table in the .hql file as a *DAY_ID* parameter.

For this feature the .hql file's **first line should not be used** for hive commands. The reason for this is that the program replaces the first line with something like this: SET DAY_ID = '20160814';. But any other placeholder is good for the first line. In this case every \${hiveconf:DAY_ID} will be replaced with 20160814.