

# Analysis 2 - Spark SQL (Hive Metastore)

## Settings

- Review active interpreters: md (needs to be on the top), spark2, sh, shUser
- Review correct file to analyse in %sh and %spark2.pyspark

total 1.8G

```
-rw-rw-r-- 1 bd01 bd01 119M Feb  5 19:02 prep_household_data_15min_singleindex.csv
-rw-rw-r-- 1 bd01 bd01 1.6G Feb  5 19:00 prep_household_data_1min_singleindex.csv
-rw-rw-r-- 1 bd01 bd01  34M Feb  5 18:56 prep_household_data_60min_singleindex.csv
```

----hdfs----

Found 6 items

```
drwx-----  - zeppelin hdfs          0 2021-02-06 06:00 /user/zeppelin/.Trash
drwxr-xr-x  - zeppelin hdfs          0 2021-02-06 13:38 /user/zeppelin/conf
drwxr-xr-x  - zeppelin hdfs          0 2021-02-06 12:57 /user/zeppelin/notebook
-rw-r--r--  1 zeppelin hdfs 124520554 2021-02-05 20:36 /user/zeppelin/prep_household_data_15min_singleindex.csv
-rw-r--r--  1 zeppelin hdfs 1712423587 2021-02-06 11:26 /user/zeppelin/prep_household_data_1min_singleindex.csv
-rw-r--r--  1 zeppelin hdfs  34626321 2021-02-05 20:35 /user/zeppelin/prep_household_data_60min_singleindex.csv
```

```
+-----+-----+-----+
|      col_name|data_type|comment|
+-----+-----+-----+
|      _c0|      int|  null|
| utc_timestamp|timestamp|  null|
|cet_cest_timestamp|  string|  null|
|   interpolated|  string|  null|
|         unit|  string|  null|
|   dishwasher|  double|  null|
|     freezer|  double|  null|
|   grid_import|  double|  null|
|   heat_pump|  double|  null|
|         pv|  double|  null|
| washing_machine|  double|  null|
| circulation_pump|  double|  null|
|   grid_export|  double|  null|
| refrigerator|  double|  null|
|         ...|  double|  null|
```

