Переделка CMD в формат ISO (YYYY-MM-DD)

Меняем CMD, оставляя одну строку. Меняем --delete-target-dir вместо --hive-import --hive-database s\_dwh --hive-table cardtrnoptype\_sdim.

--hive-overwrite в конце тоже удаляем:

было:

$PMRootDir/Scripts/quotes\_insert.pl 'ssh etladmin@bda11node06 ^sqoop import --connect \^jdbc:oracle:thin:@(DESCRIPTION=(ADDRESS\_LIST=(ADDRESS=(PROTOCOL=TCP)(HOST=exa1-scan)(PORT=1521))(ADDRESS=(PROTOCOL=TCP)(HOST=exa2-scan)(PORT=1521)))(CONNECT\_DATA=(SERVICE\_NAME=DWSTPROD\_TAF)))\^ --username $DWH\_USER --password $DWH\_PASS --query \^select UK, CCODE, NAME, CARDTRNOPTYPEGROUP\_UK, DELETED\_FLAG, DEFAULT\_FLAG, JOB\_INSERT, JOB\_UPDATE, to\_char(AS\_OF\_DAY, ~dd.mm.yyyy~) as AS\_OF\_DAY, DWSCMIX from DWH.CARDTRNOPTYPE\_SDIM where deleted\_flag=~N~ and \\\$CONDITIONS\^ -m 1 --compression-codec GZIP --as-parquetfile **--hive-import --hive-database s\_dwh --hive-table cardtrnoptype\_sdim** --target-dir /storage/s\_dwh/db/cardtrnoptype\_sdim/ **--hive-overwrite**^'

стало:

$PMRootDir/Scripts/quotes\_insert.pl 'ssh etladmin@bda11node06 ^sqoop import --connect \^jdbc:oracle:thin:@(DESCRIPTION=(ADDRESS\_LIST=(ADDRESS=(PROTOCOL=TCP)(HOST=exa1-scan)(PORT=1521))(ADDRESS=(PROTOCOL=TCP)(HOST=exa2-scan)(PORT=1521)))(CONNECT\_DATA=(SERVICE\_NAME=DWSTPROD\_TAF)))\^ --username $DWH\_USER --password $DWH\_PASS --query \^select UK, CCODE, NAME, CARDTRNOPTYPEGROUP\_UK, DELETED\_FLAG, DEFAULT\_FLAG, JOB\_INSERT, JOB\_UPDATE, to\_char(AS\_OF\_DAY, ~YYYY-MM-DD~) as AS\_OF\_DAY, DWSCMIX from DWH.CARDTRNOPTYPE\_SDIM where deleted\_flag=~N~ and \\\$CONDITIONS\^ -m 1 --compression-codec GZIP --as-parquetfile **--delete-target-dir** --target-dir /storage/s\_dwh/db/cardtrnoptype\_sdim/^'

Удаляем строки :

Change group permissions:

hdfs dfs -chown etladmin:s\_dwh\_w /storage/s\_dwh/db/cardtrnoptype\_sdim/\*

Add ACL read permission:

hdfs dfs -setfacl -R --set user::rwx,group::rwx,group:s\_dwh\_r:r-x,other::--- /storage/s\_dwh/db/cardtrnoptype\_sdim

Переделываем запрос из \* в перечисление полей, преобразовывая даты:

*-- to\_char(AS\_OF\_DAY, ~YYYY-MM-DD~) as AS\_OF\_DAY*

*-- to\_char(START\_TIME, ~YYYY-MM-DD HH24:MI:SS~) as START\_TIME*

select \* from DWH.CHANNEL\_HDIM;

select UK, CCODE, ADDITIONAL\_CCODE, NAME\_RUS, NAME\_ENG, NAME\_SHORT, to\_char(START\_DATE, ~dd.mm.yyyy~) as START\_DATE, to\_char(END\_DATE, ~dd.mm.yyyy~) as END\_DATE, USER\_UK, DELETED\_FLAG, DEFAULT\_FLAG, to\_char(VALIDFROM, ~dd.mm.yyyy~) as VALIDFROM, to\_char(VALIDTO, ~dd.mm.yyyy~) as VALIDTO, JOB\_INSERT, JOB\_UPDATE, to\_char(AS\_OF\_DAY, ~yyyy-mm-dd HH24:MI:SS~) as AS\_OF\_DAY, DWSCMIX, CHANNELKIND\_GK, SRC\_CCODE, MODULE\_UK from DWH.CHANNEL\_HDIM

MANAGED\_TABLE

Удаляем старую таблицу:

drop table `s\_dwh.channel\_hdim`

Создаем новую:

CREATE TABLE `s\_dwh.CHANNEL\_HDIM`(

`UK` string,

`CCODE` string,

`ADDITIONAL\_CCODE` string,

`NAME\_RUS` string,

`NAME\_ENG` string,

`NAME\_SHORT` string,

`START\_DATE` string,

`END\_DATE` string,

`USER\_UK` string,

`DELETED\_FLAG` string,

`DEFAULT\_FLAG` string,

`VALIDFROM` string,

`VALIDTO` string,

`JOB\_INSERT` string,

`JOB\_UPDATE` string,

`AS\_OF\_DAY` string,

`DWSCMIX` string,

`CHANNELKIND\_GK` string,

`SRC\_CCODE` string,

`MODULE\_UK` string)

ROW FORMAT SERDE

'org.apache.hadoop.hive.ql.io.parquet.serde.ParquetHiveSerDe'

STORED AS INPUTFORMAT

'org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat'

OUTPUTFORMAT

'org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat'

LOCATION

'hdfs://bda11/storage/s\_dwh/db/CHANNEL\_HDIM'

TBLPROPERTIES (

'COLUMN\_STATS\_ACCURATE'='false',

'avro.schema.url'='hdfs://bda11/storage/s\_dwh/db/CHANNEL\_HDIM/.metadata/schemas/1.avsc',

'kite.compression.type'='snappy',

'numFiles'='0',

'numRows'='-1',

'rawDataSize'='-1',

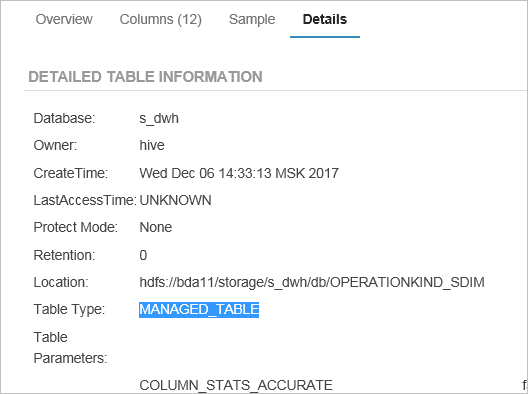
'totalSize'='0',

'transient\_lastDdlTime'='1527848676')

Запускаем CMD

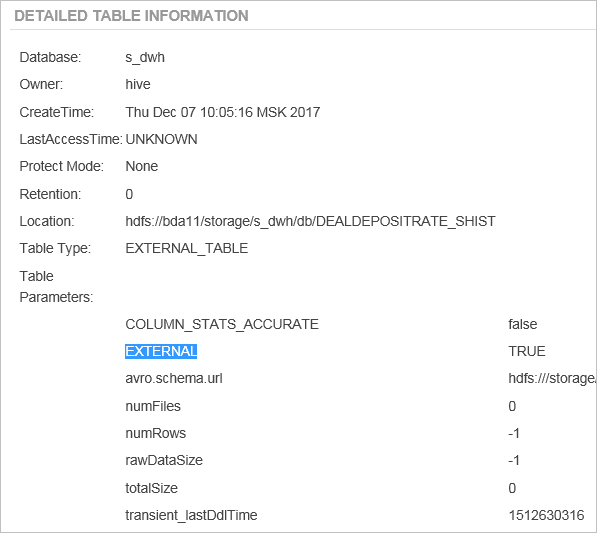
**ПРИМЕЧАНИЕ:**

Если при создании таблицы пишем просто CREATE TABLE то создается MANAGED таблица:



и при ее удалении удаляется папка HDFS с файлами этой таблицы

Если при создании указывается CREATE **EXTERNAL** TABLE то при удалении папка с файлами не удалиться.



MANAGED\_TABLE MANAGED\_TABLE MANAGED\_TABLE MANAGED\_TABLE