Загрузка справочников без --hive-import --hive-database s\_dwh --hive-table

Из-за проблем на 3-м кластере со **scoop import** переделали загрузку справочников, убрав

**--hive-import --hive-database s\_dwh --hive-table**

И вместо этого применив

**--delete-target-dir**

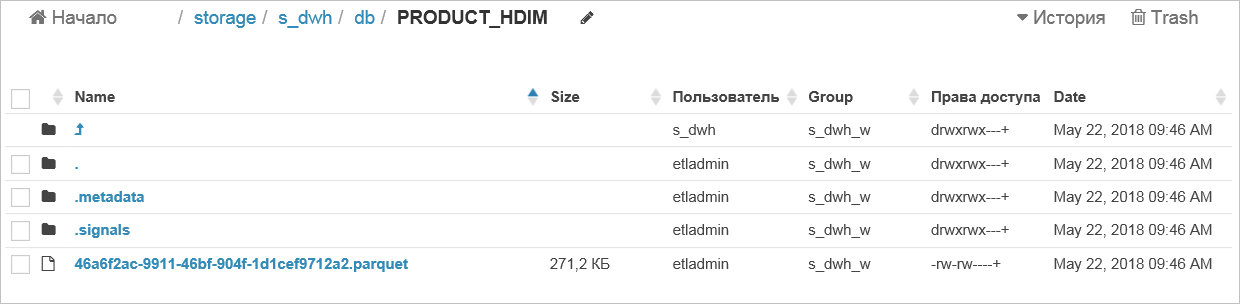
Было так:

$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 \* from CLIENT\_HDIM where validto=to\_date(~31.12.5999~,~dd.mm.yyyy~) and deleted\_flag=~N~ and \\\$CONDITIONS\^ -m 1 --compression-codec GZIP --as-parquetfile **--hive-import --hive-database s\_dwh --hive-table** CLIENT\_HDIM --target-dir /storage/s\_dwh/db/CLIENT\_HDIM/ --hive-overwrite^'

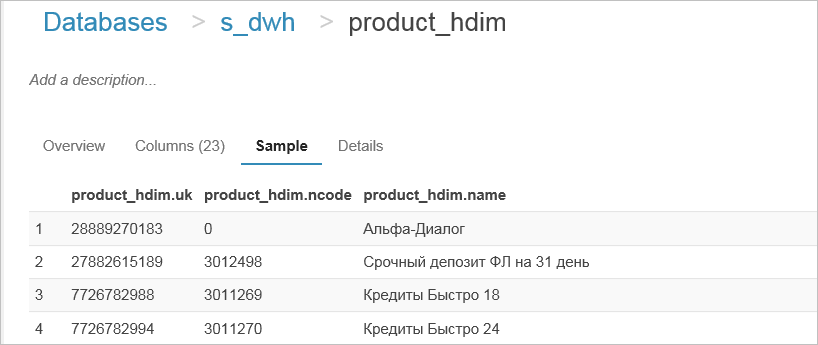
Стало так:

$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 \* from PRODUCT\_HDIM where validto=to\_date(~31.12.5999~,~dd.mm.yyyy~) and deleted\_flag=~N~ and \\\$CONDITIONS\^ -m 1 --compression-codec GZIP --as-parquetfile **--delete-target-dir** --target-dir /storage/s\_dwh/db/PRODUCT\_HDIM/^'

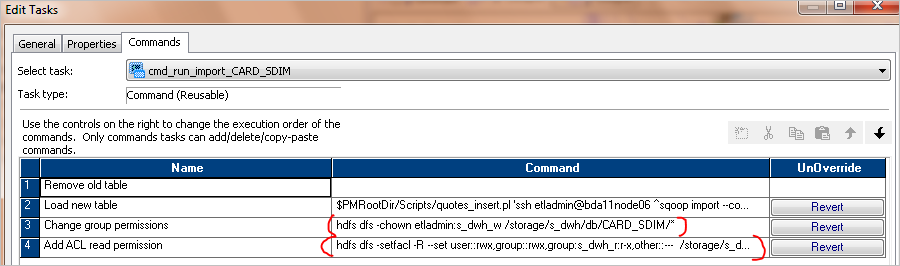
В этом случае уже не hive (хайв) создает таблицы, а мы заранее ее создаем и она автоматически смотрит на файлы HDFS. И если есть паркет файл:



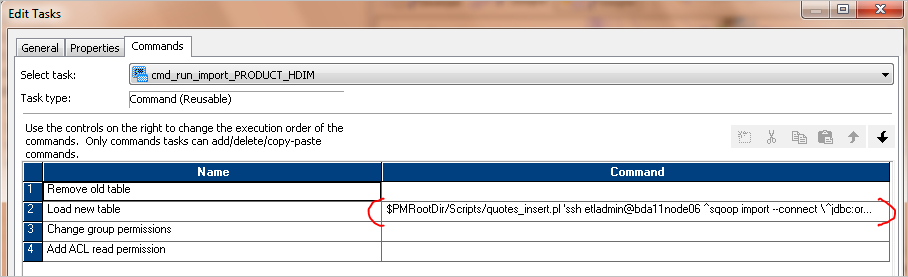
То в таблице сразу есть данные:



И также в этом случае не надо раздавать права на таблицу и выдавать группы как мы это делали:



И теперь достаточно одной строки:



Так как сразу присутствует группа **s\_dwh\_w** а не **hive**:

