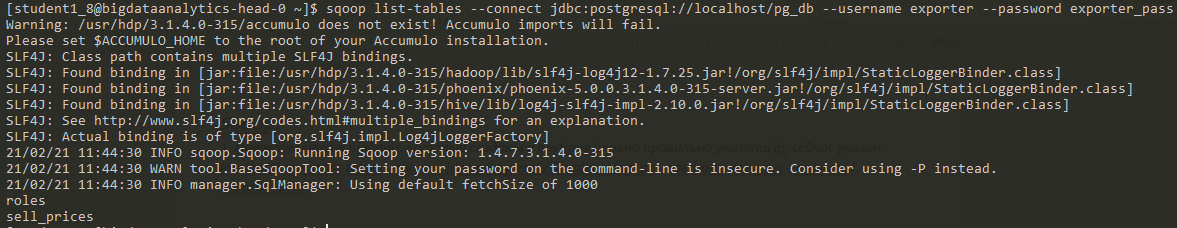
***Sqoop:***

***1. Получаем список таблиц, которые созданы в postgres (нужно правильно указать ip, сейчас указан localhost)***

sqoop list-tables --connect jdbc:postgresql://localhost/pg\_db --username exporter --password exporter\_pass



***2. Создаем таблицу в hive, куда собираемся выгружать данные CREATE TABLE lesson6\_9.character( charid string, charname string, abbrev string, description string, speechcount int)***

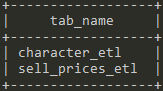
0: jdbc:hive2://bigdataanalytics-worker-0.nov> create database l6\_etl;

0: jdbc:hive2://bigdataanalytics-worker-0.nov> use database l6\_etl;

0: jdbc:hive2://bigdataanalytics-worker-0.nov> CREATE c TABLE l6\_etl.character\_ETL(charid string, charname string, abbrev string, description string, speechcount int);

0: jdbc:hive2://bigdataanalytics-worker-0.nov> CREATE external TABLE `sell\_prices\_etl`( `store\_id` string, `item\_id` string, `wm\_yr\_wk` string, `sell\_price` string);

0: jdbc:hive2://bigdataanalytics-worker-0.nov> show tables;



0: jdbc:hive2://bigdataanalytics-worker-0.nov> select \* from character\_etl;



***3. Запускаем импорт: sqoop import --connect jdbc:postgresql://node3.novalocal/pg\_db --username exporter --password exporter\_pass --table character --hive-import --hive-database lesson6\_9 --hive-table character***

[student1\_8@bigdataanalytics-head-0 ~]$ sqoop import --connect jdbc:postgresql://localhost/pg\_db --username exporter --password exporter\_pass --table sell\_prices\_etl - -hive-import --hive-database l6\_etl --hive-table sell\_prices\_etl -m 1

21/02/21 16:55:18 ERROR manager.SqlManager: Error executing statement: org.postgresql.util.PSQLException: ERROR: relation "sell\_prices\_etl" does not exist

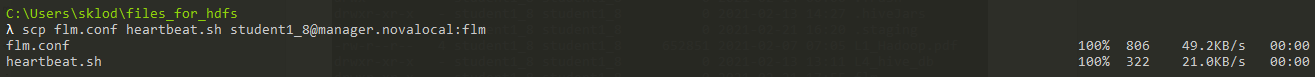
org.postgresql.util.PSQLException: ERROR: relation "sell\_prices\_etl" does not exist

[student1\_8@bigdataanalytics-head-0 ~]$ sqoop import --connect jdbc:postgresql://localhost/pg\_db --username exporter --password exporter\_pass --table sell\_prices --hiv e-import --hive-database l6\_etl.db --hive-table sell\_prices\_etl -m 1

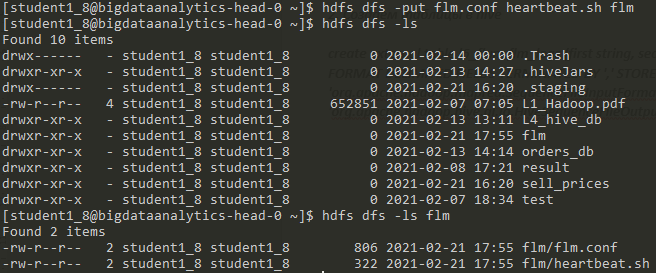
21/02/21 16:54:30 ERROR tool.ImportTool: Import failed: org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://bigdataanalytics-head-0.novalocal:8020/user/student1\_8/sell\_prices already exists

*Я так и не смогла добиться импорта и перестала понимать, что вообще происходит. Меняла localhost на 10.0.0.8 – результат такой же*

***Flume:***



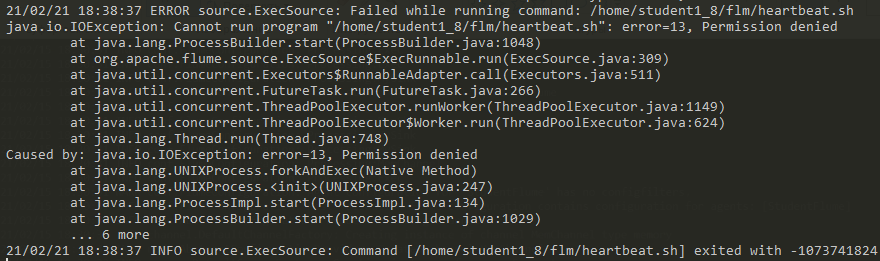




***1. Запускаем flume (нужно правильно указать пути до папки и файла) /opt/apache-flume/bin/flume-ng agent --conf /home/admin/flm/ --conf-file /home/admin/flm/flm.conf --name StudentFlume -Dflume.root.logger=INFO,console***

/opt/apache-flume/bin/flume-ng agent --conf /home/student1\_8/flm/ --conf-file /home/student1\_8/flm/flm.conf --name StudentFlume -Dflume.root.logger=INFO,console

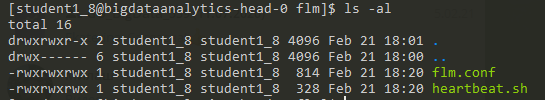
*После этого скрипта на меня вывалилось много текста, но суть в том, что код не сработал. Я уверена, что причина в этом:*



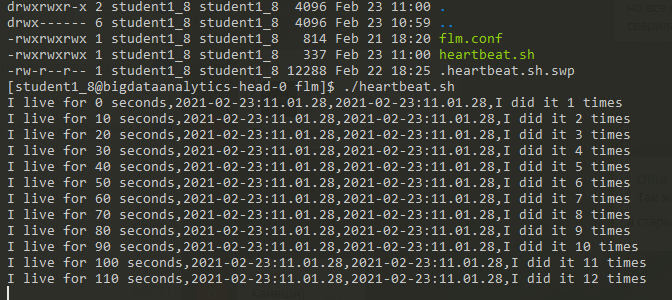
*Я так и не разобралась, почему heartbeat.sh не запустился. Я скопировала скрипт с методички и сравнила с тем, который был на уроке – все верно. Вероятно, проблема в каких-то доступах или разрешениях.*

*Соответственно да п.2 я не дошла – мне нечего складывать в табличку((((*

*На уроке разобрались с правами, получили новый файлик heartbeat.sh*



*На следующий день перегрузила файл еще раз и он заработал:*



***2. Создаем таблицы в hive***

create external table l6\_flum.flm\_logs (first string, second string, third string, fourth string, fifth string) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS INPUTFORMAT 'org.apache.hadoop.mapred.SequenceFileInputFormat' OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveSequenceFileOutputFormat' location '/user/student1\_8/flm'

Логи пишутся в папку и в таблицу!!

