Подключение по SH к Impala на ЛД

|  |
| --- |
| impala-shell -k -i hue.lab.df.sbrf.ru:21000 -q "show databases" |

**Обновление метаданных Impala на ЛД**

|  |
| --- |
| В Impala необходимо перед обращением к данным обновлять метаинформацию    invalidate metadata database.table;  или  refresh database.table;    также в рамках сессии советую выполнять команду:  SET PARQUET\_FALLBACK\_SCHEMA\_RESOLUTION=1; |

**Переопределение фактора репликации директории HDFS**

|  |
| --- |
| export HADOOP\_CLIENT\_OPTS="-Xmx32G"; hdfs dfs -setrep -R -w 2 /data/core/internal/COD/  hadoop fs -Ddfs.replication=5 -cp hadoop-test/test.csv hadoop-test/test\_with\_rep5.csv |

Распределение узлов Impala по инженерам

|  |
| --- |
| Род С  94, 72, 63, 75, 76,84, 66,21,32, 20  Серов И  93,82, 29, 61,36,67, 69,22, 52, 85  Пилипейко Г  49, 62,64,65,79, 88,59,68, 44, 31  Сальников А  50, 77, 80,60, 95, 37, 38,70, 30, 35  Формат команды будет соответственно такой  impala-shell -k -i februs{DD}.lab.df.sbrf.ru:21000 -q "show databases" |

**Создание PARQUETной таблицы в Impala**

|  |
| --- |
| use core\_internal\_crm\_kb;  drop table if exists CX\_CG\_ACCOUNT;  create external table CX\_CG\_ACCOUNT LIKE PARQUET '/data/ods/internal/CRM\_KB/src/DT\_039\_000000/CX\_CG\_ACCOUNT/load20190403/6d8329f5-ff8d-4439-a200-ffb746bf77ac.parquet'  STORED AS PARQUET LOCATION '/data/ods/internal/CRM\_KB/src/DT\_039\_000000/CX\_CG\_ACCOUNT/load20190403'  TBLPROPERTIES ('Дата загрузки'='03.04.2019','ЗНО'='ЗНО0070730880','Задача в JIRA'='BIBS-3754','Откуда загружено'='ОДС'); |

**Из bigint корректно в date Impala**

|  |
| --- |
| --Запрос дает разный результат в Hive и Impala  --Дело в том, что Impala в отличии от Hive не преобразует автоматически (без настройки) GMT в GMT+3, делаем это сами.  select 1469998800000,  from\_unixtime(cast(1469998800000/1000 as bigint),'yyyy-MM-dd HH:mm:ss'),  from\_utc\_timestamp(from\_unixtime(cast(1469998800000/1000 as bigint),'yyyy-MM-dd HH:mm:ss'), 'EEST') --из GMT0 в GMT+3 |