REFRESH таблиц Impala

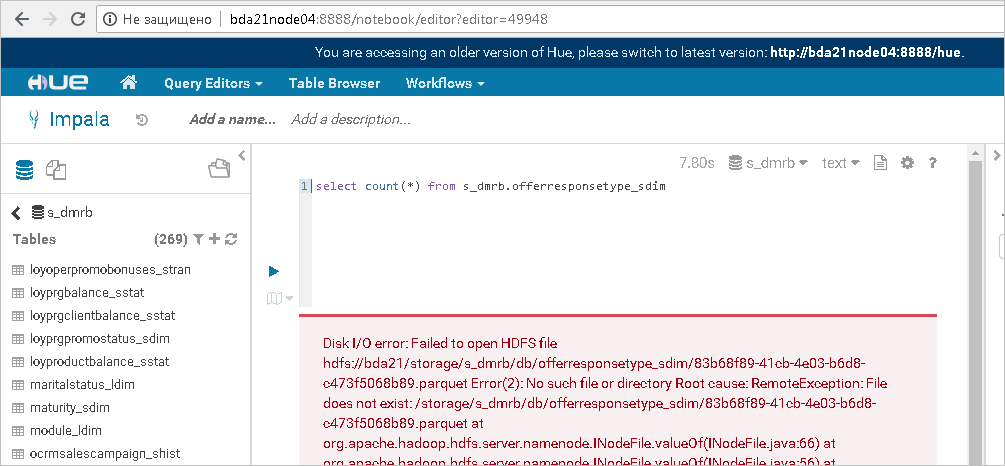
Письмо пользователю:

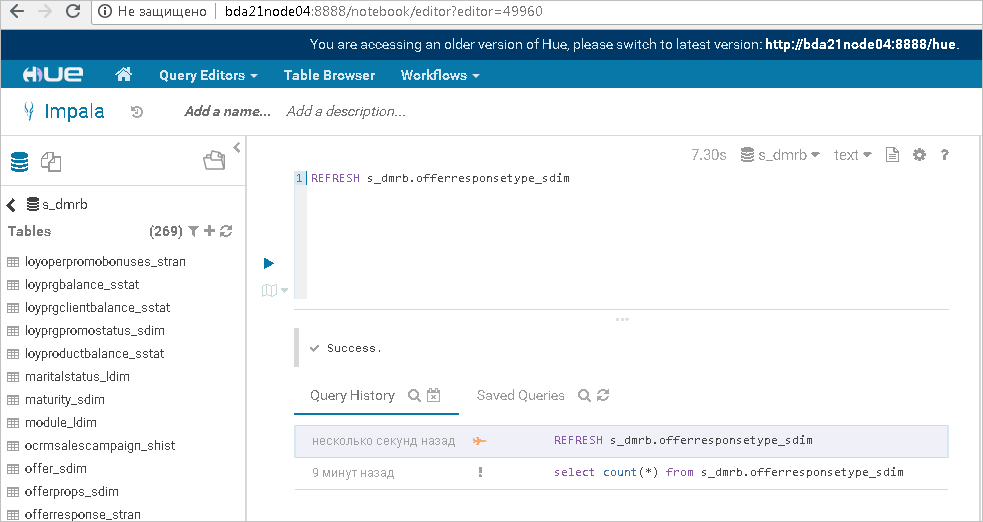
Добрый день.

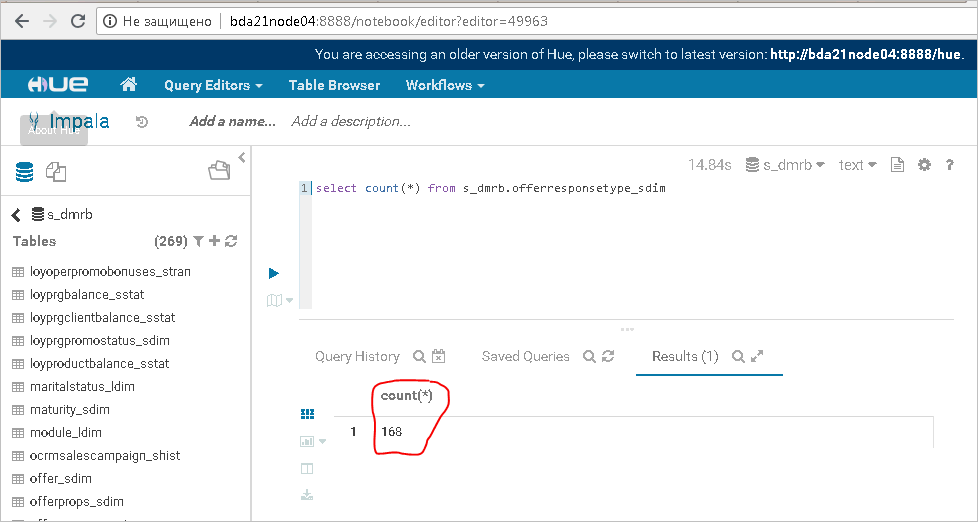
Особенностью Impala является необходимость обновления метаданных, если таблицы менялись в другом приложении, например, в Hive.

Для этого можно использовать команду REFRESH: REFRESH s\_dmrb.offerresponsetype\_sdim

Более подробно можно почить здесь: http://confluence.moscow.alfaintra.net/pages/viewpage.action?pageId=213446170







Вадим,

загрузка ETL происходит через hive по разным таблицам в разное время (а таких объектов у нас ежедневно несколько тысяч обновляется) . Запросы в hive не требуют обновления метаданных, а вот в Impala требуется или выполнять refresh для каждой таблицы (причем это занимает доп.время в загрузках, а загрузки у нас идут по 20 часов из 24 в день), или обновлять по всем базам в определенное время. У каждого подхода есть свои плюсы и минусы. Причем это потеряет всю необходимость при запросах в Impala на других кластерах, т.к. там свой сервис Impala и свои метаданные.

На данный момент встраивать refresh в ETL нет возможности.

Эта проблема сейчас находится в стадии проработки на стороне Развития Hadoop\_Datalake. Комментарии может дать Евгений.

*С уважением,*

*Полюхова Евгения,*

*УАТиС Дирекция BI*

*Центр по управлению и монетизации данных*

*АО "Альфа-Банк"  
т.:+7 /495/ 620-91-91, добавочный 011-4007*

*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\**

Добрый день.

Если это Impala, то нужно делать refresh метаданных, после того, как файлы с данными таблицы изменились.

Для этого можно использовать команду REFRESH: REFRESH s\_dmrb.client\_sdim

Инструкция по Impala : http://confluence.moscow.alfaintra.net/pages/viewpage.action?pageId=213446170