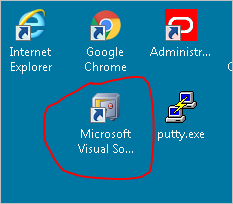
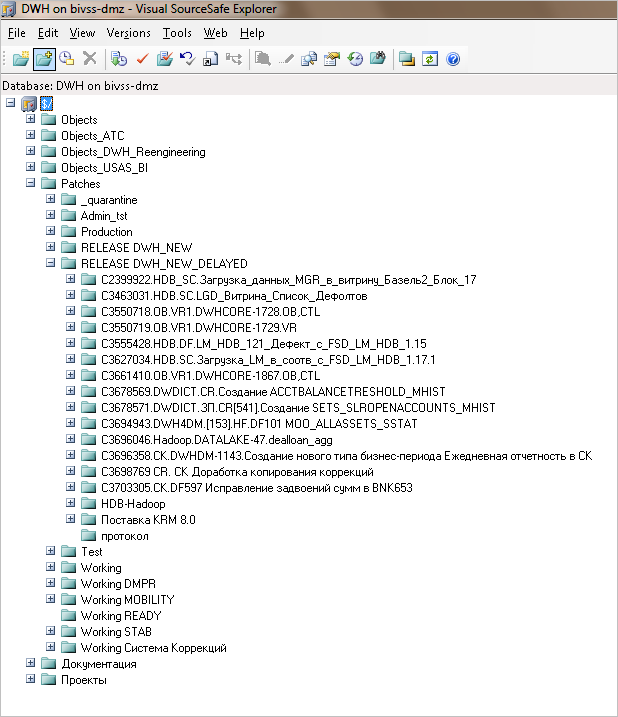
**Установка патча в Hadoop**

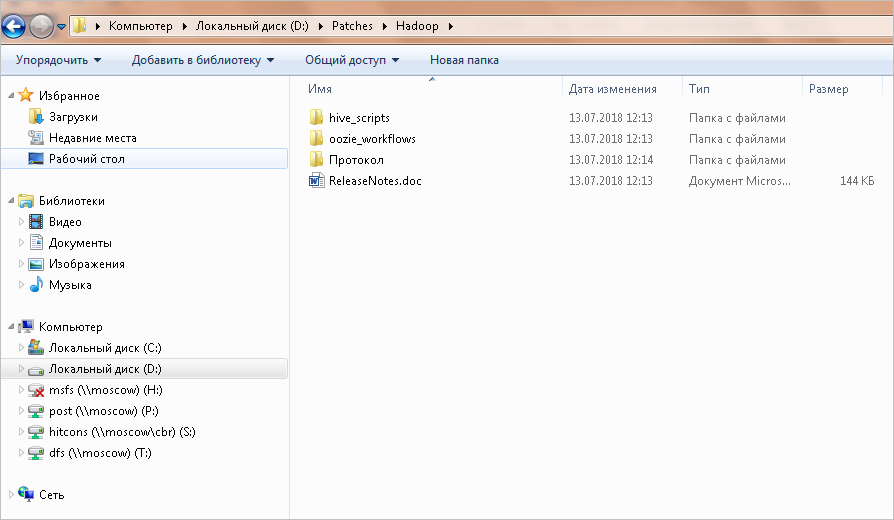
**VSS**



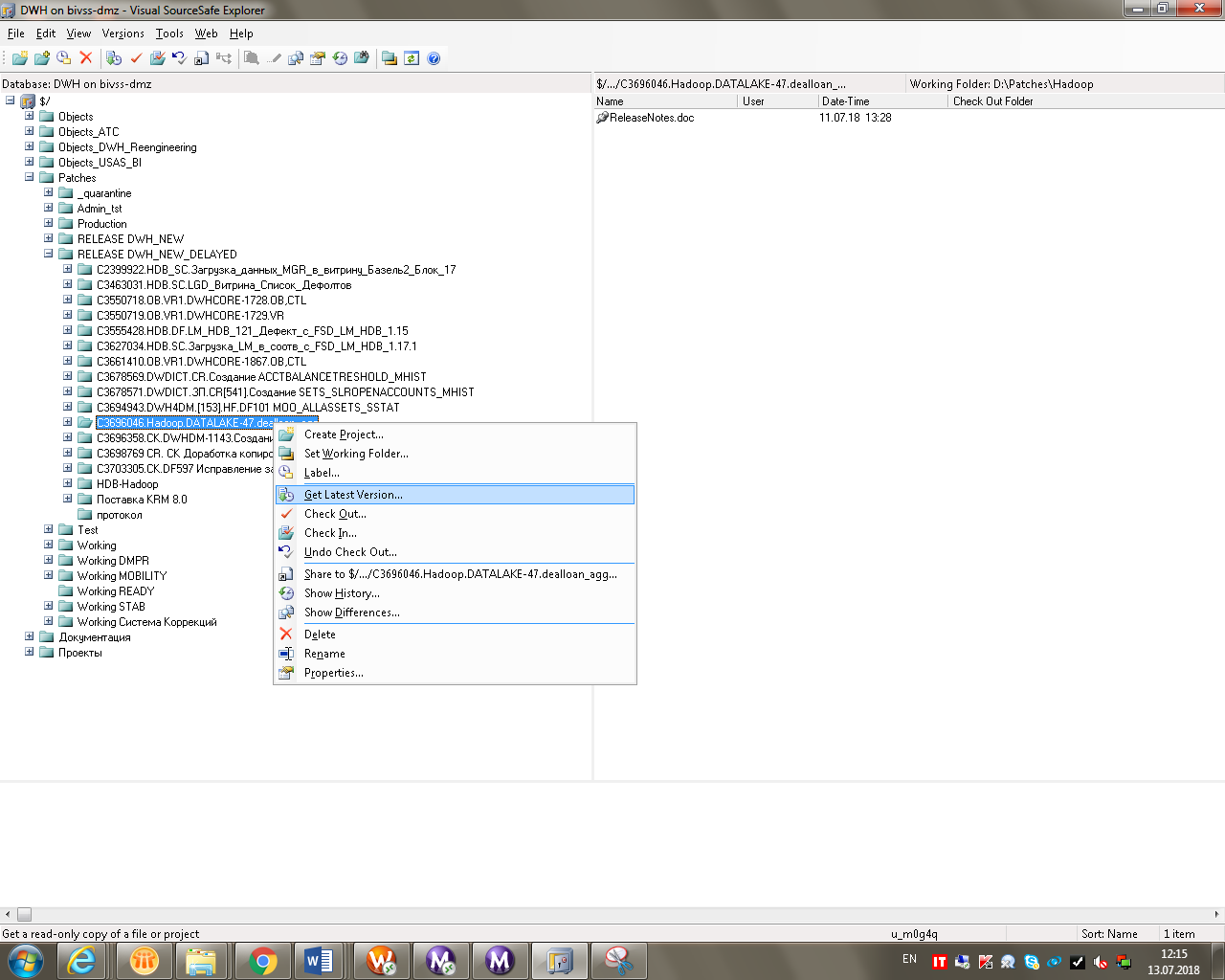
логин/пароль: **u\_m0g4q** / **u\_m0g4q**



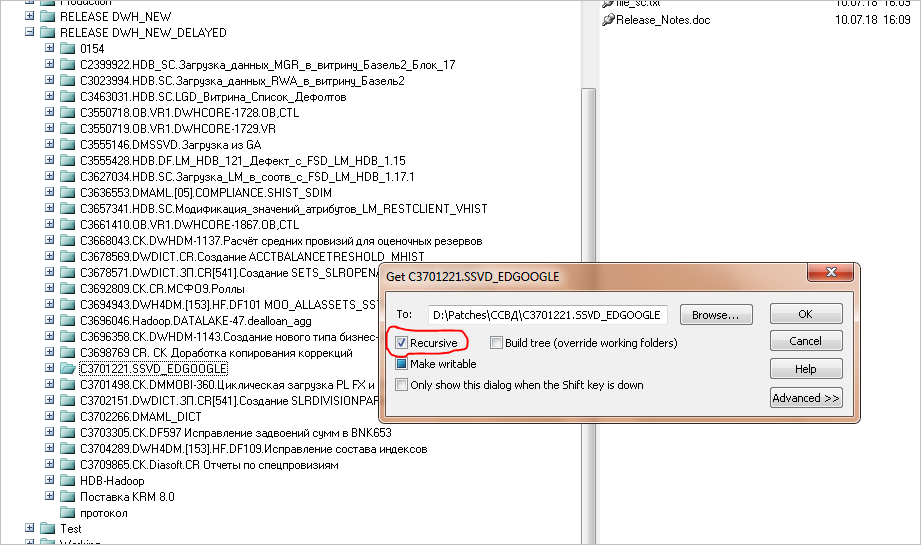
Скачиваем патч в папку на своем компе



ПКМ на патче:



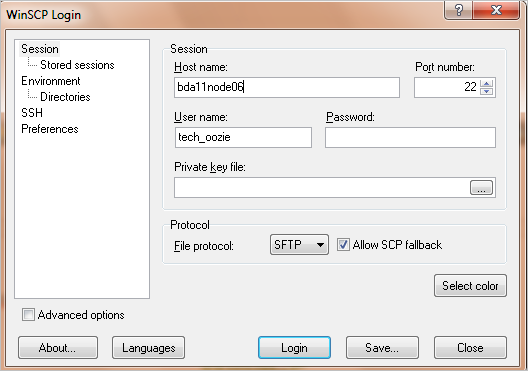
Ставим галочку **recursive:**



Открываем ReleaseNotes.doc

**WinSCP**

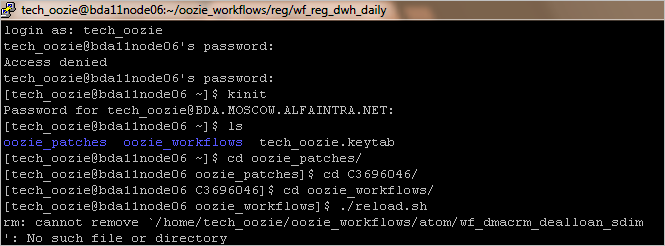
Сервер **bda11node06** пользователь **tech\_oozie // tech\_oozie**

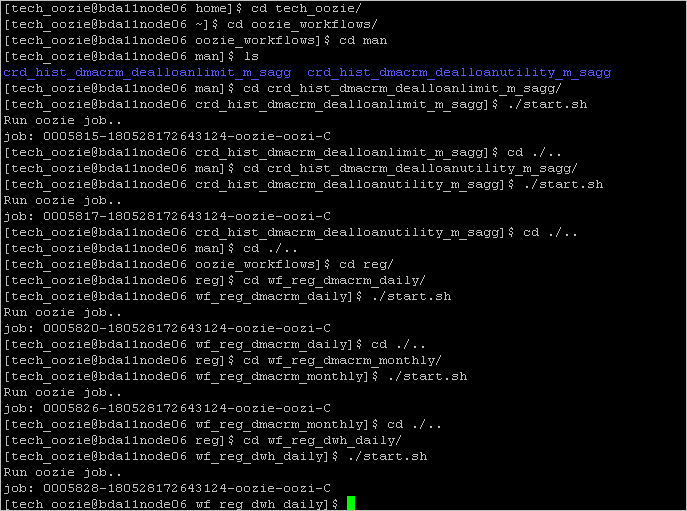


**Putty**

Заходим на сервер **bda11node06**

юзер/пароль: tech\_oozie // tech\_oozie





Команды:

**kinit** сразу после авторизации и еще раз пароль – устанавливаем надежное соединение, чтобы Karberos соединение не прерывалось

**ls** – просмотр каталога в котором находимся

**cd** – переход на каталог ниже

**cd ..** – поднимаемся на уровень выше по директориям

**cd** **/** – выход в директорию **root**

**cd** **~** – выход в директорию **home**

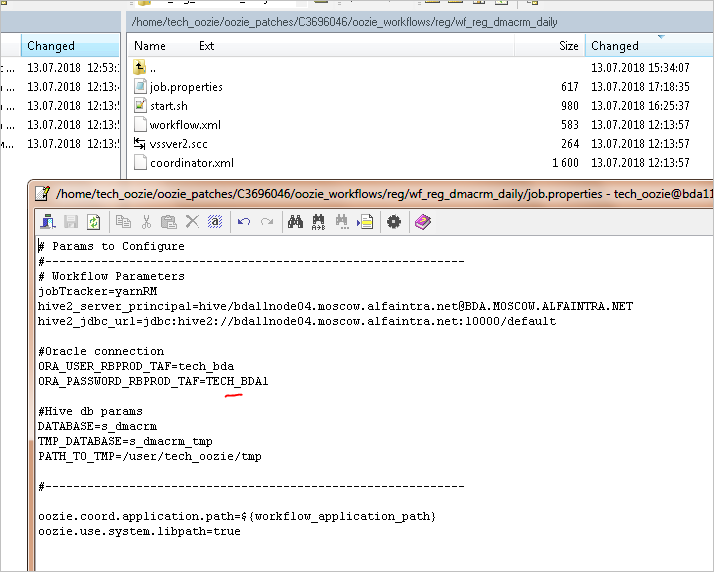
**pwd –** выводит полный путь до текущей папки:



**./reload.sh** – запуск reload.sh

**./start.sh** – запуск start.sh

Если ошиблись в параметрах потоков:



Исправляем, выполняем ./reload.sh

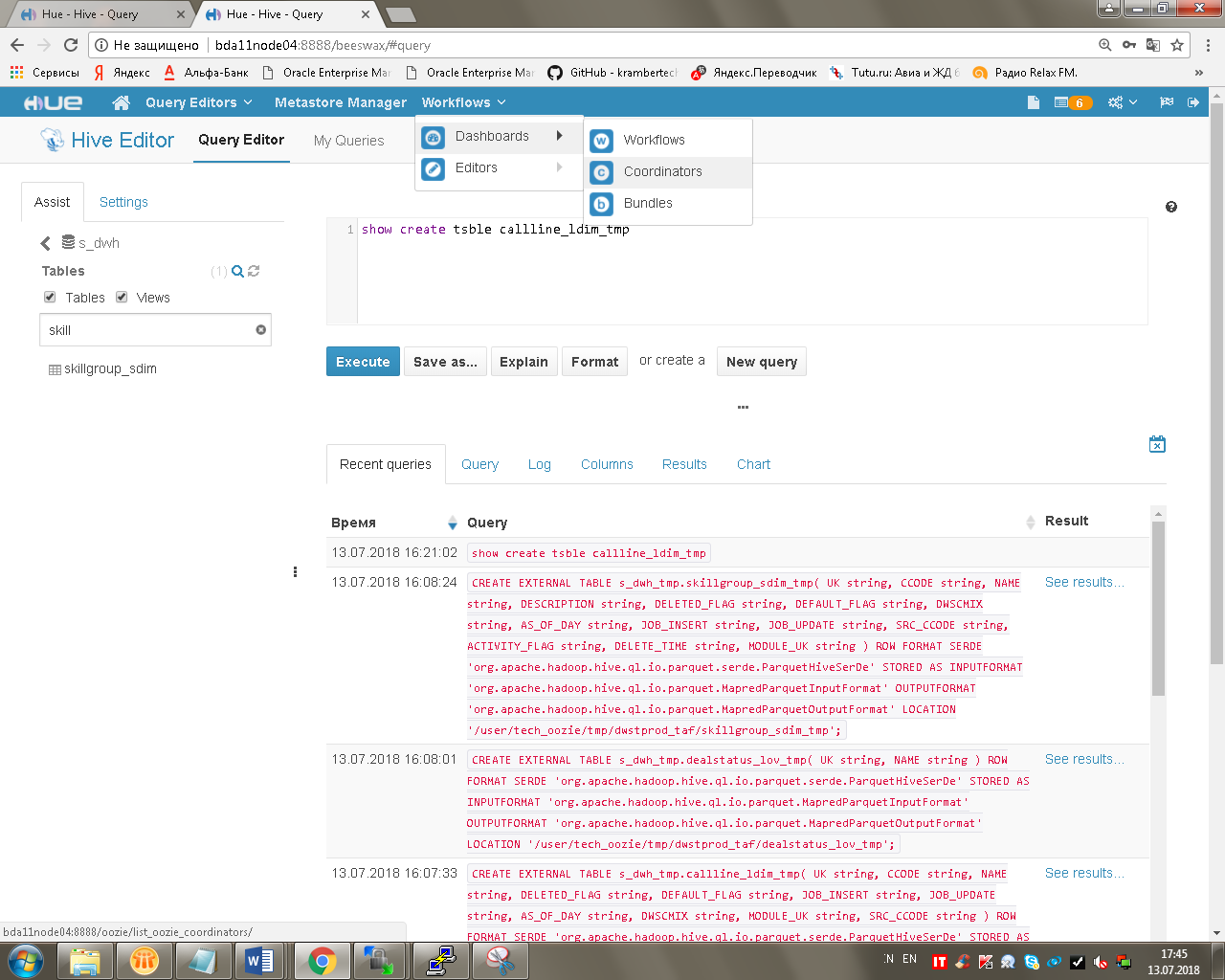
/home/<TECH\_OOZIE>/oozie\_patches/C3696046/oozie\_workflows/reload.sh.

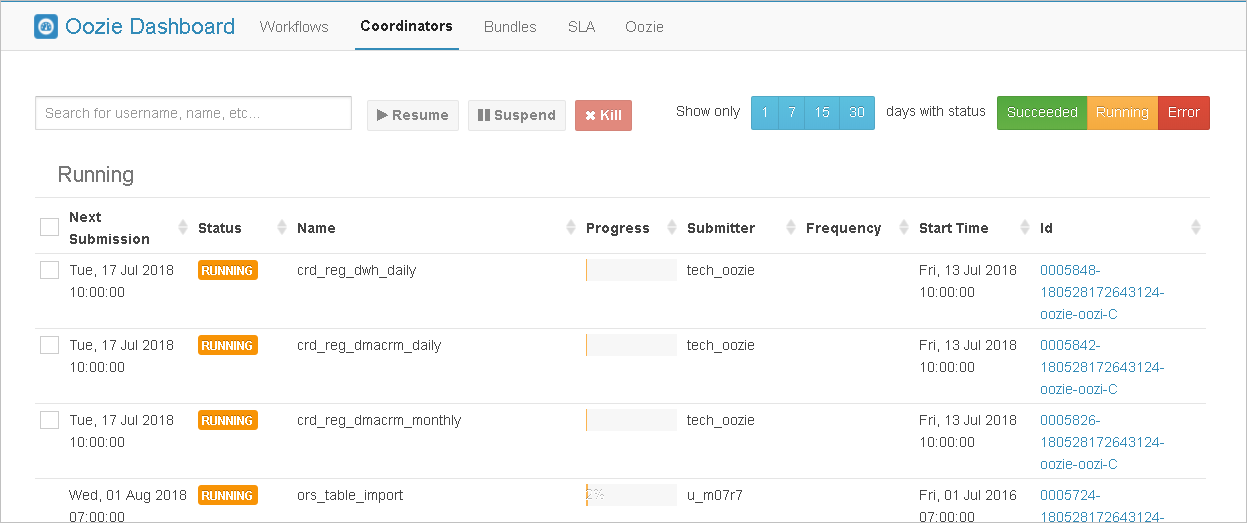
И еще раз запускаем ./start.sh

/home/<TECH\_OOZIE>/oozie\_workflows/reg/wf\_reg\_dwh\_daily/start.sh

Тех потоков которые исправляли

Просмотр работы потоков:

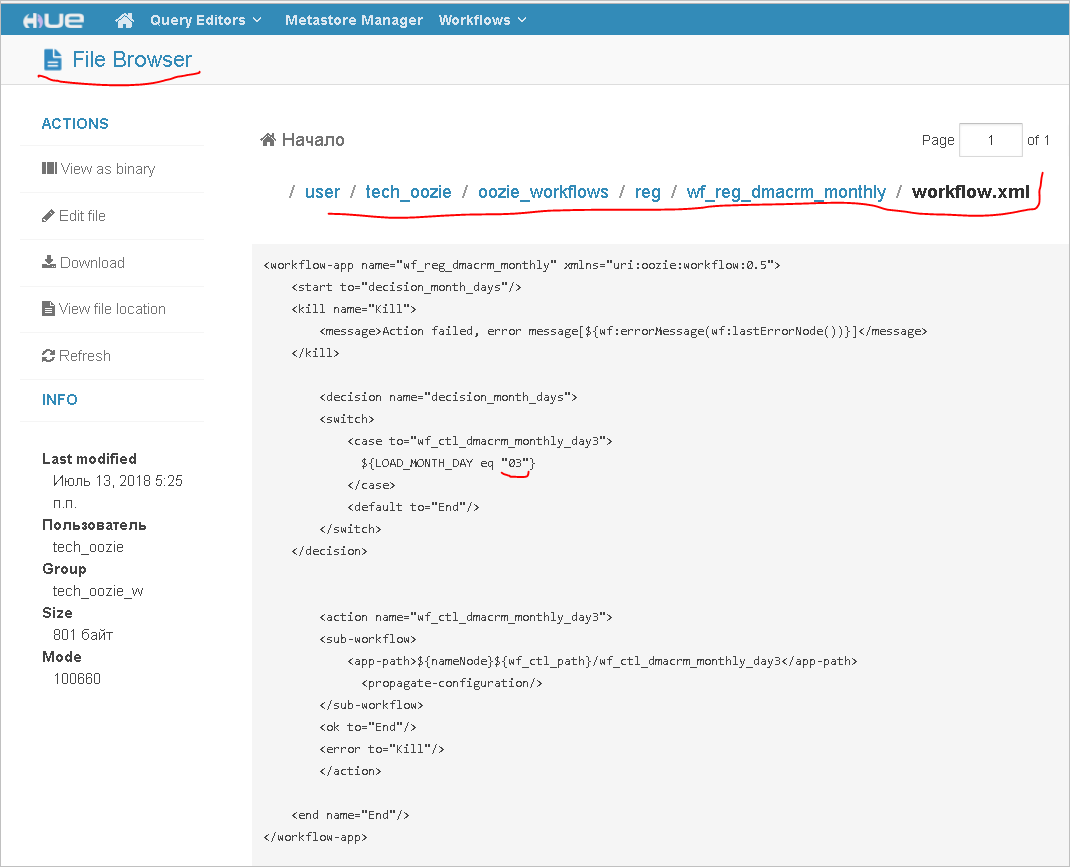




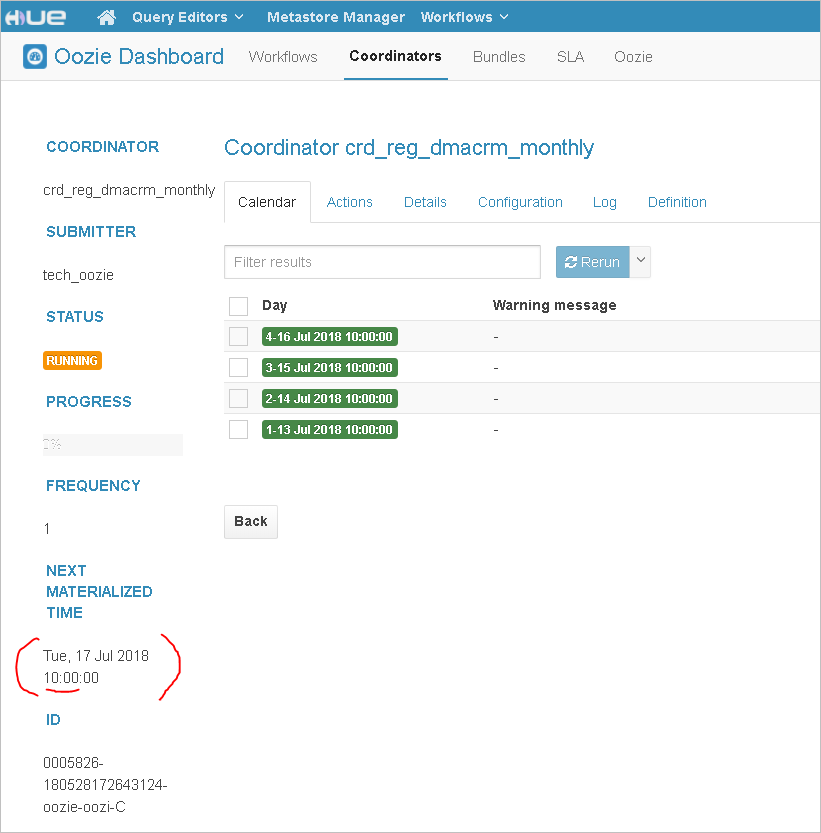
Посмотреть когда запуститься поток можно в File Browser:

Параметр **LOAD\_MONTH\_DAY**.

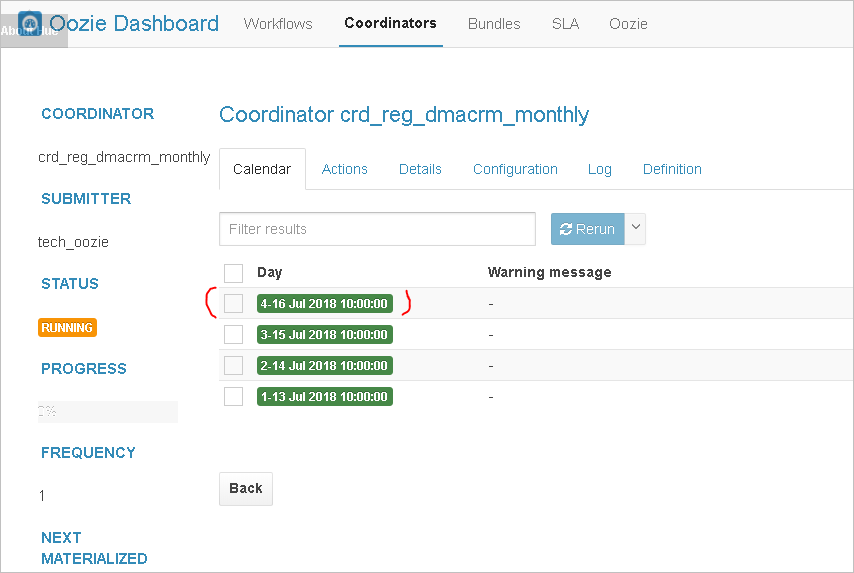
Поток запускается каждый день и каждый раз параметр LOAD\_MONTH\_DAY сравнивается с Т-1. Когда Т-1 будет равен указанному значению (в данном случае 3-м) поток запустится.



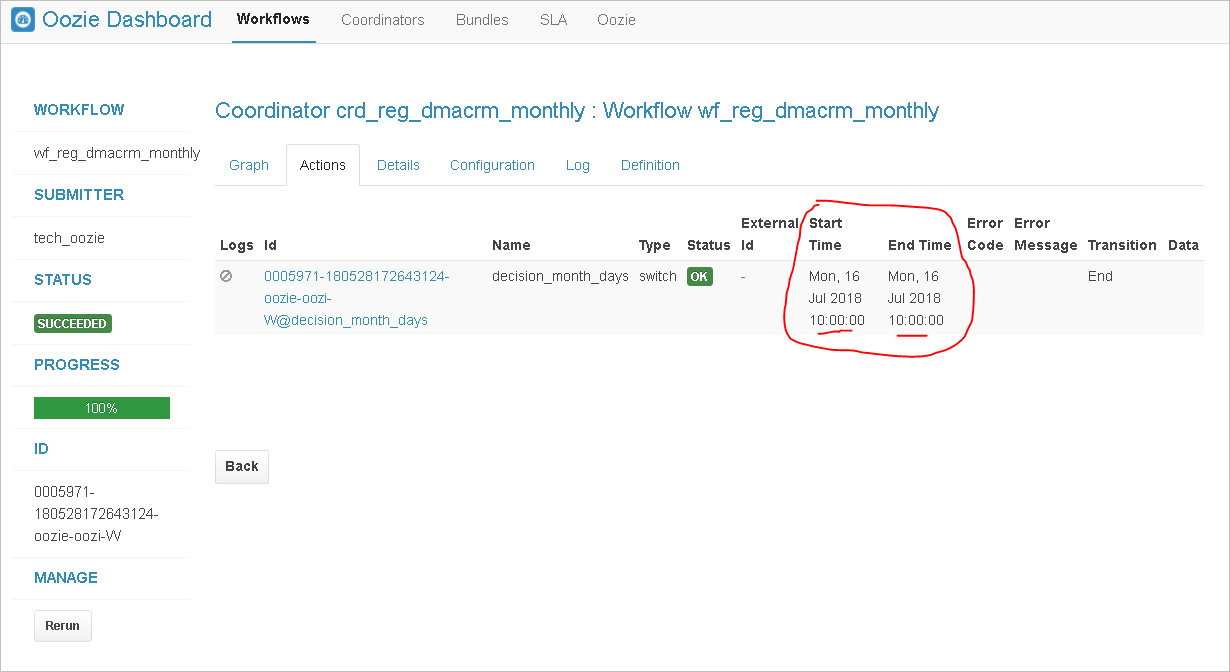
Т.е. поток запускается каждый день в 10.00, но если параметр не совпадает, работает в холостую.

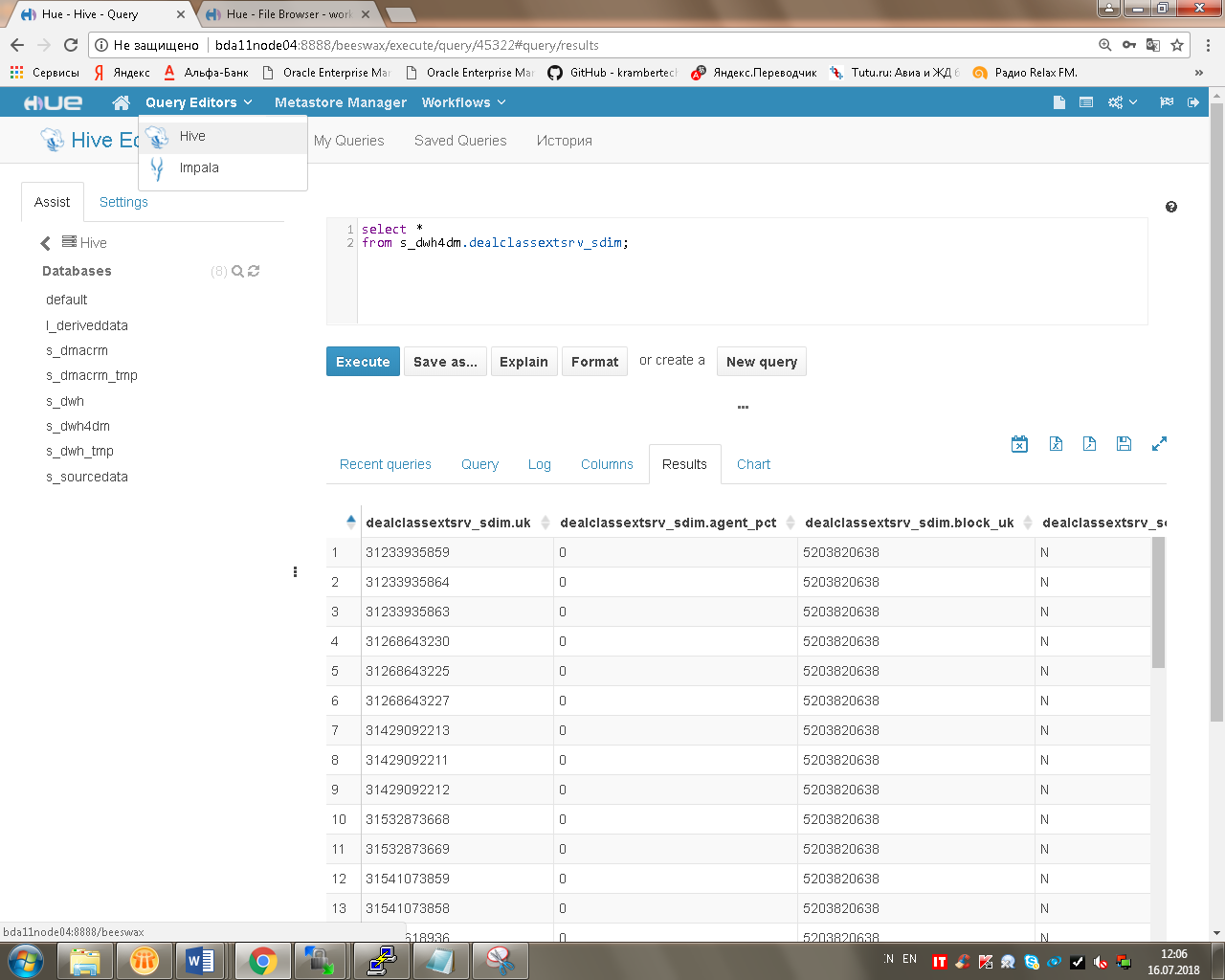


Сколько по времени работал поток можно посмотреть, если нажать на последний запуск :



и посмотреть на Start Time и End Time во вкладке **Action:**



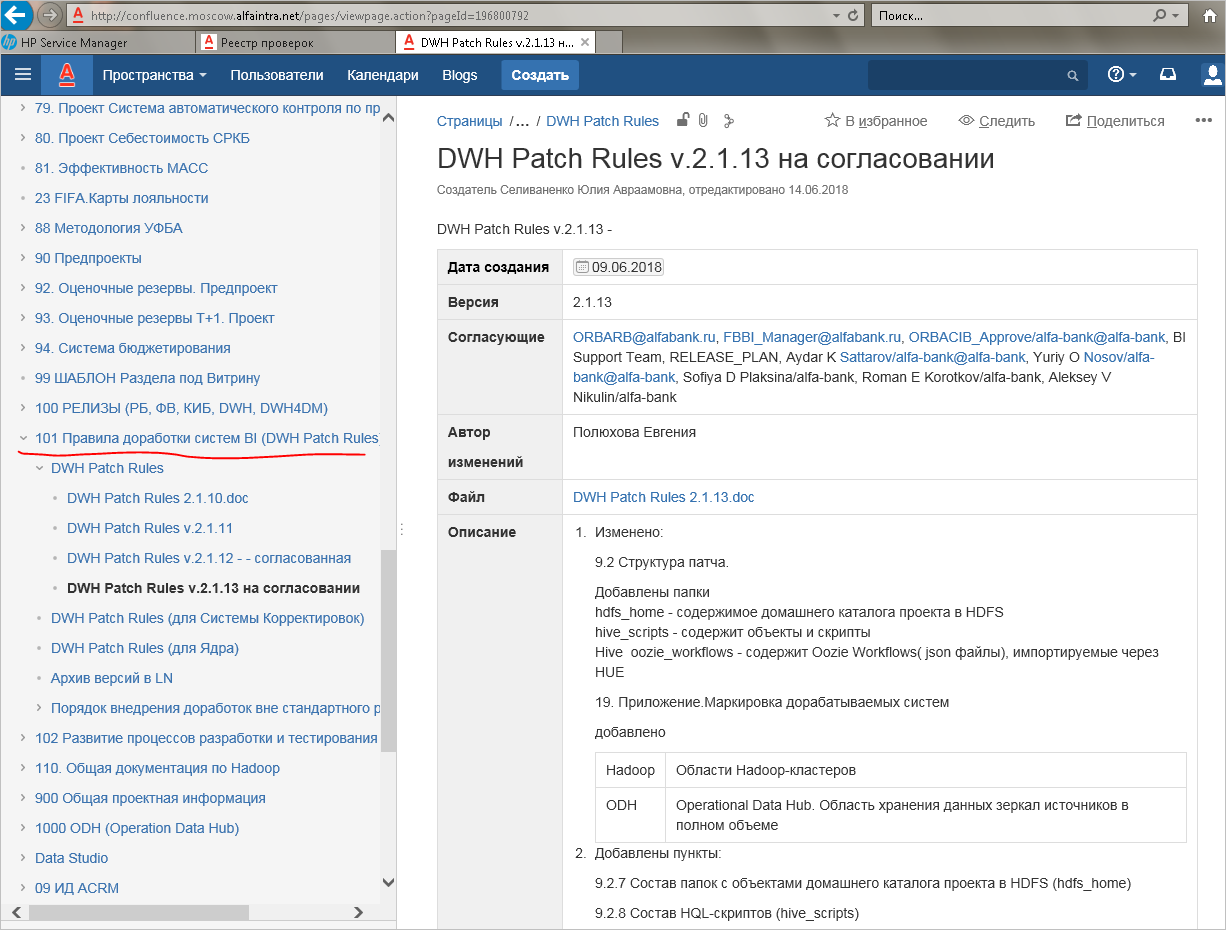


[**Patch Rules**](http://confluence.moscow.alfaintra.net/pages/viewpage.action?pageId=196800792) находится здесь:

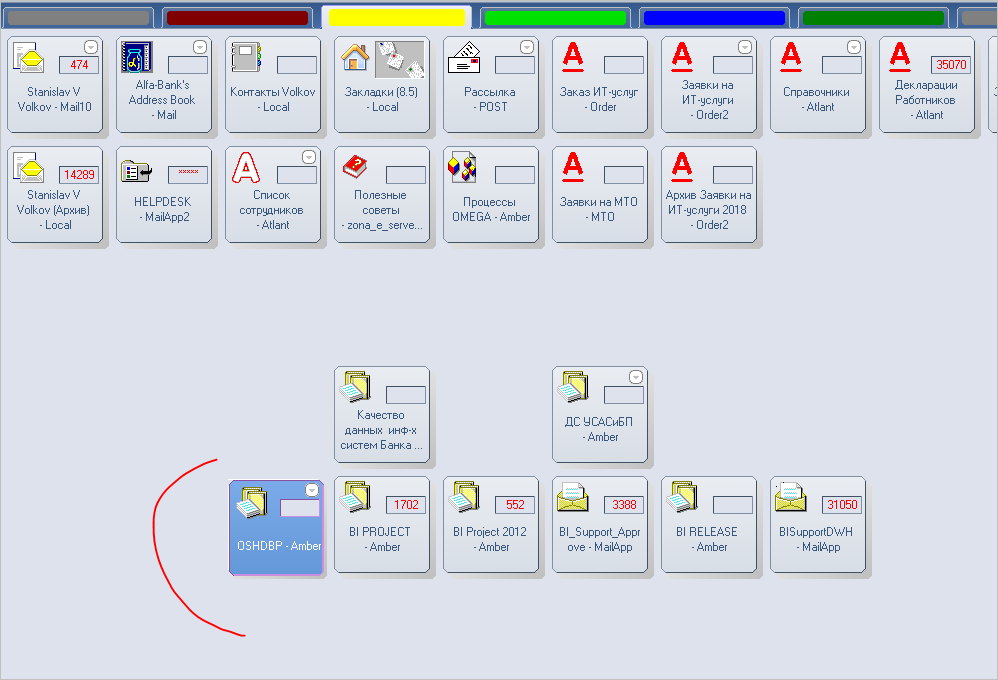
<http://confluence.moscow.alfaintra.net/pages/viewpage.action?pageId=196800792>

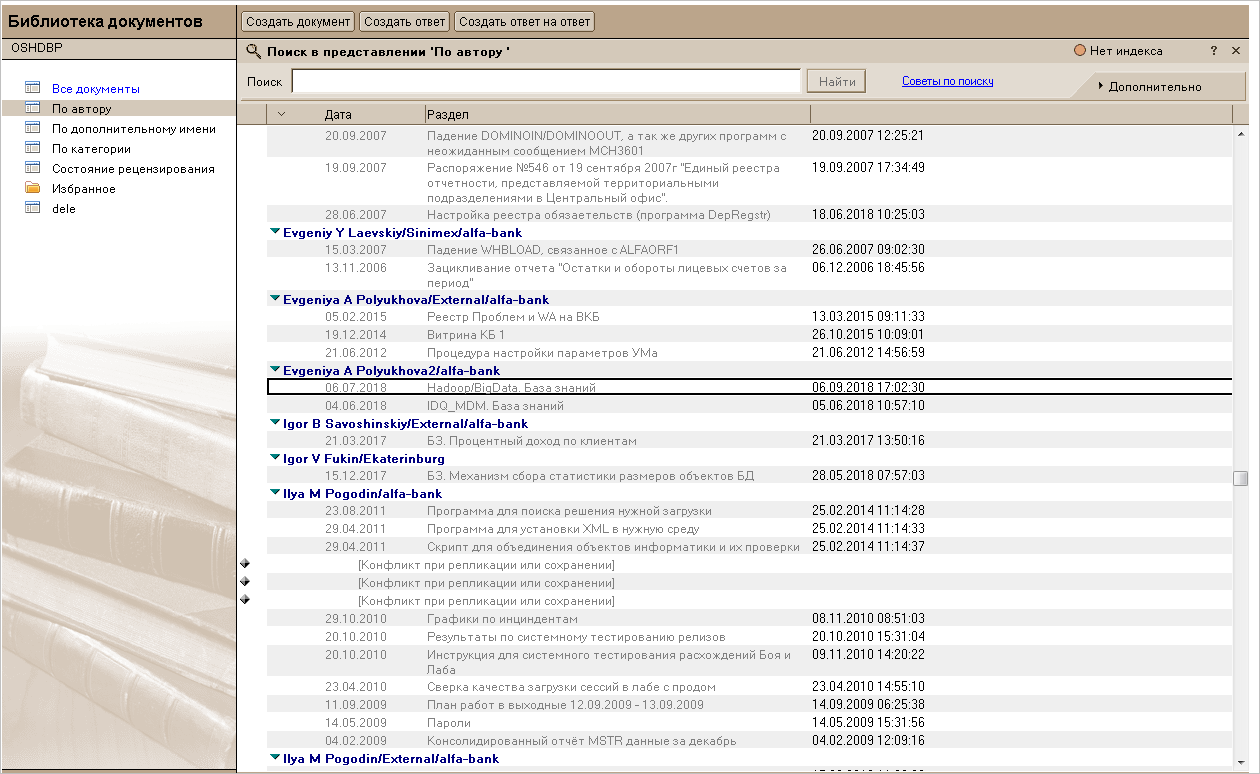
может сразу не загружаться, тогда попробовать **http://confluence.moscow.alfaintra.net/pages**

Пункт 101:



База Знаний в LN



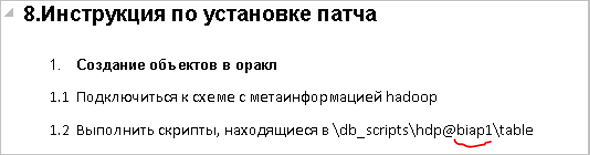


Установка патча в ТЕСТ

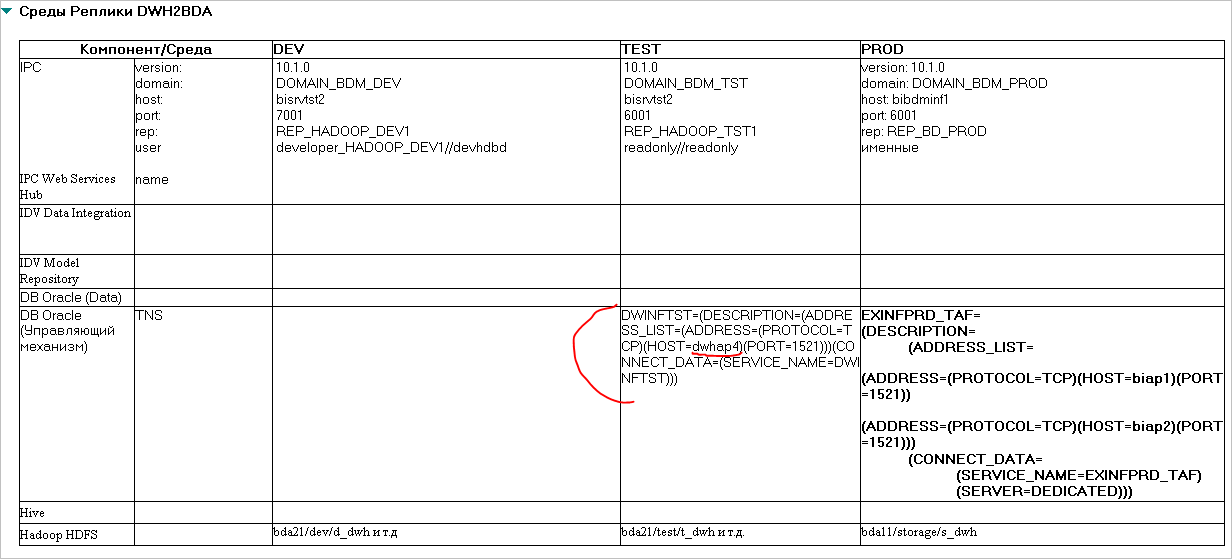
ReleaseNotes пишется для установки в БОЙ и поэтому, когда ставим в тест, подставляем реквизиты для теста, которые смотрим в описании сред.

**Пункт 1**

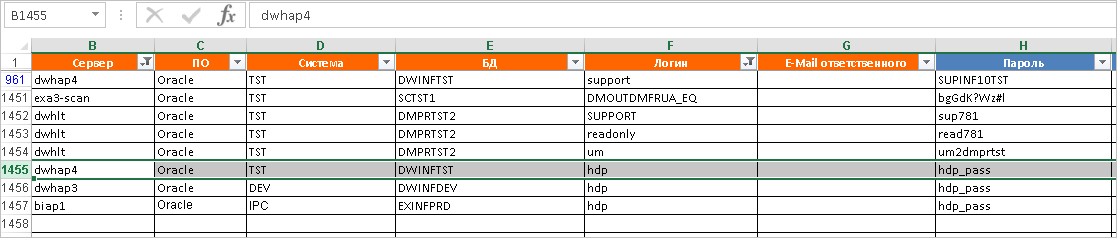
Написано выполнить скрипты в **biap1**:



это для БОЯ. Смотрим тнс для теста и подключаемся к **dwhap4**:

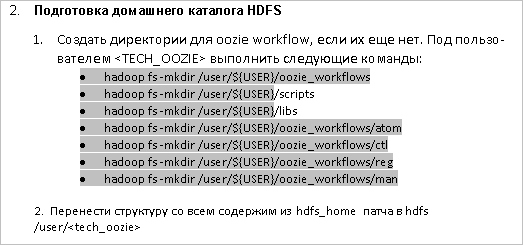


Пароль для учетки **hdp** смотрим в файле паролей:

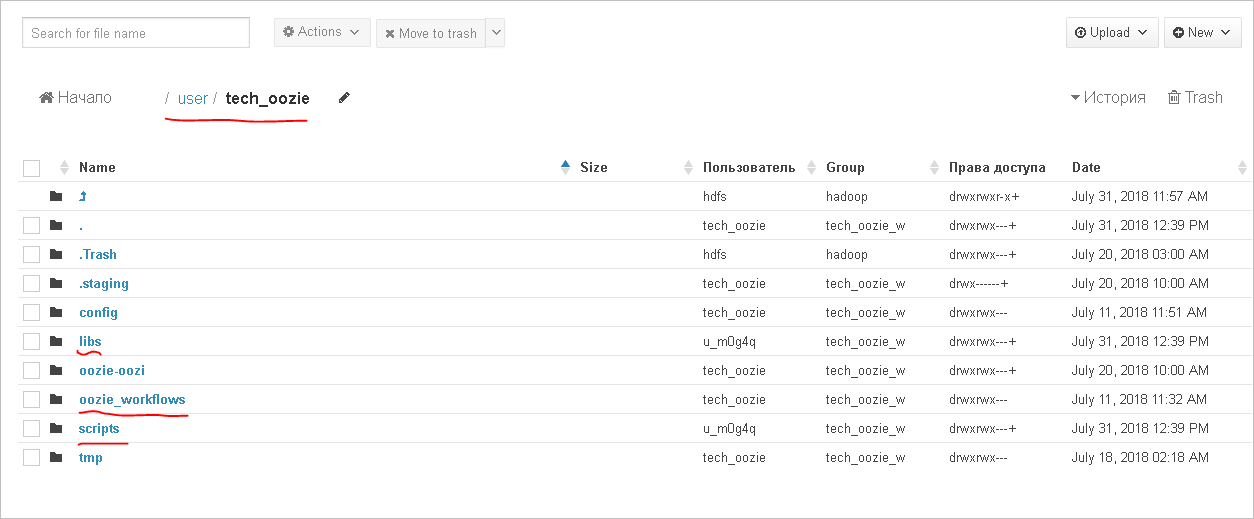


**Пункт 2**

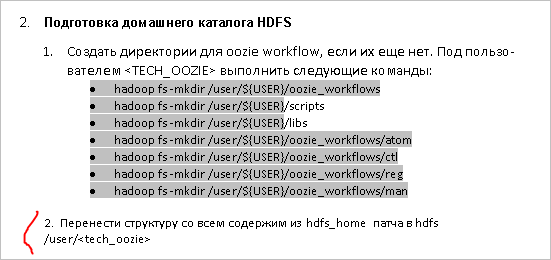
Можно создавать папки через командную строку как написано:



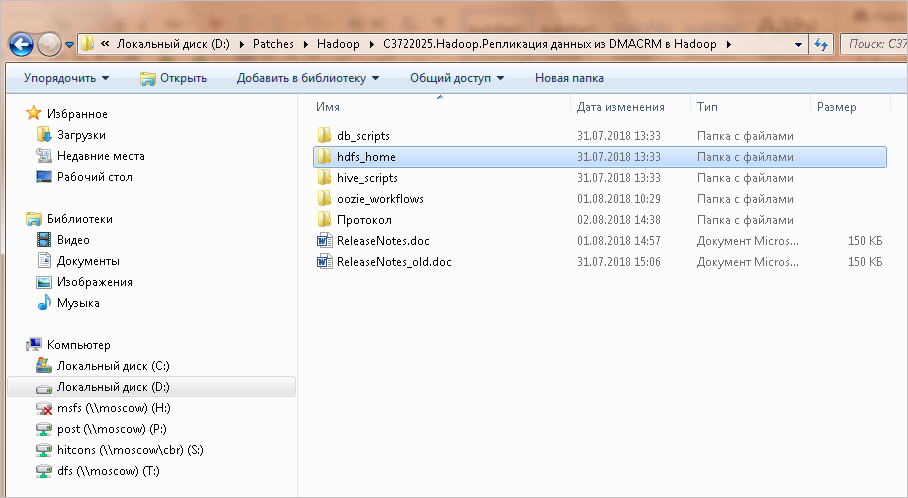
можно через браузер:



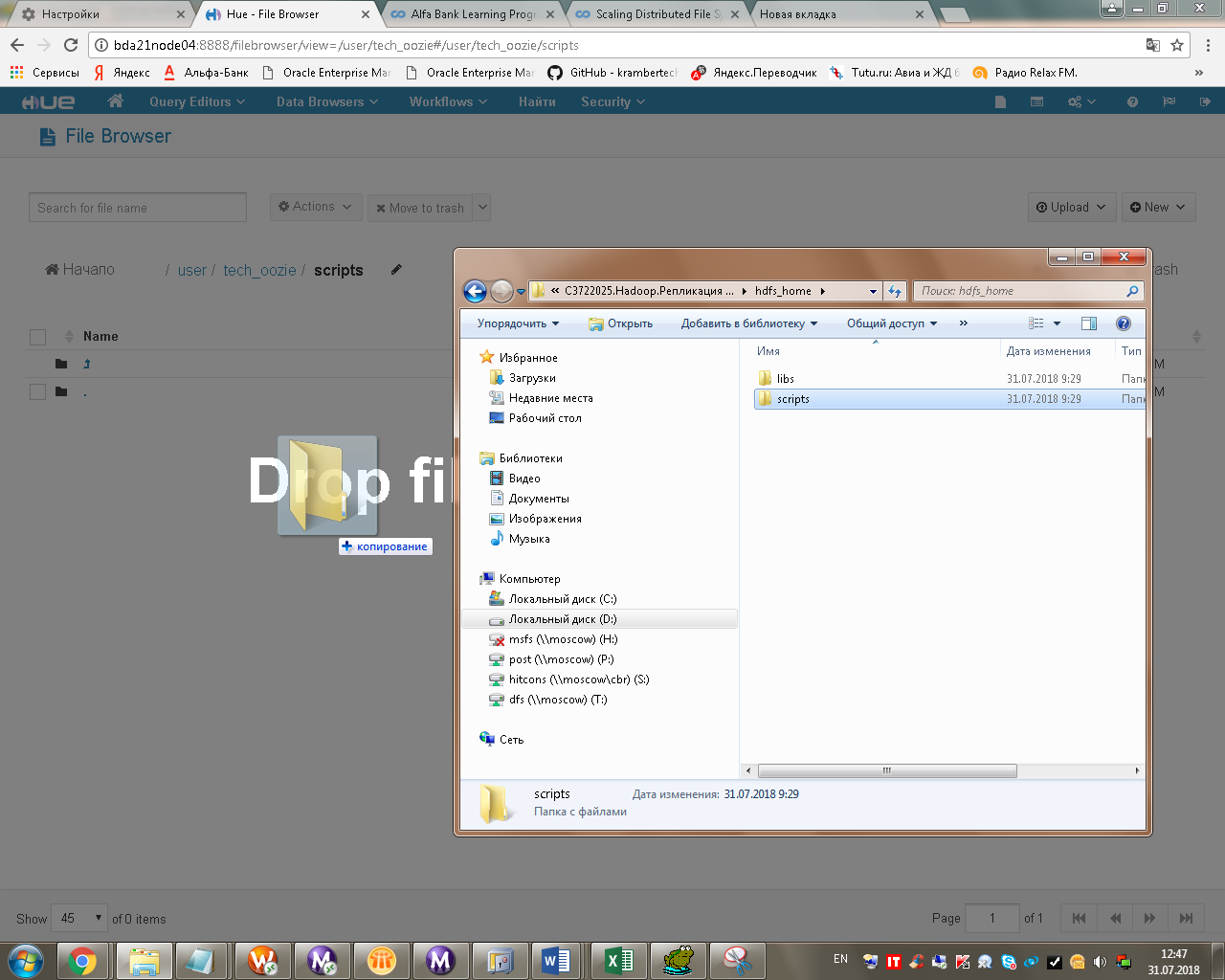
дальше пункт 2, перенос структуры:



Имеется в виду все содержимое папки **hdfs\_home:**

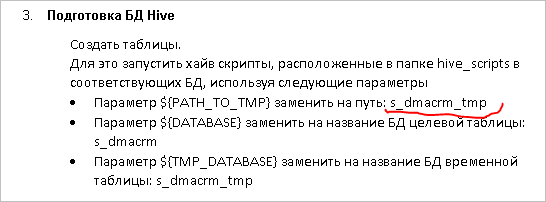


делаем перетаскиванием папок из проводника windows в hdfs:

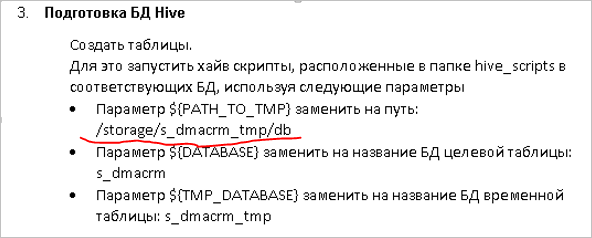


**Пункт 3**

В описании написано Параметр ${PATH\_TO\_TMP} заменить на путь: **s\_dmacrm\_tmp** – это ошибка,

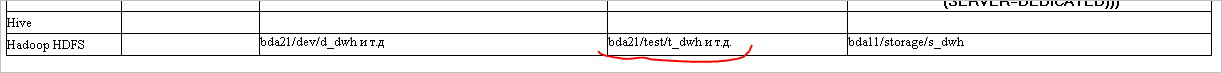


Путь нужен такой - **/storage/s\_dmacrm\_tmp/db:**



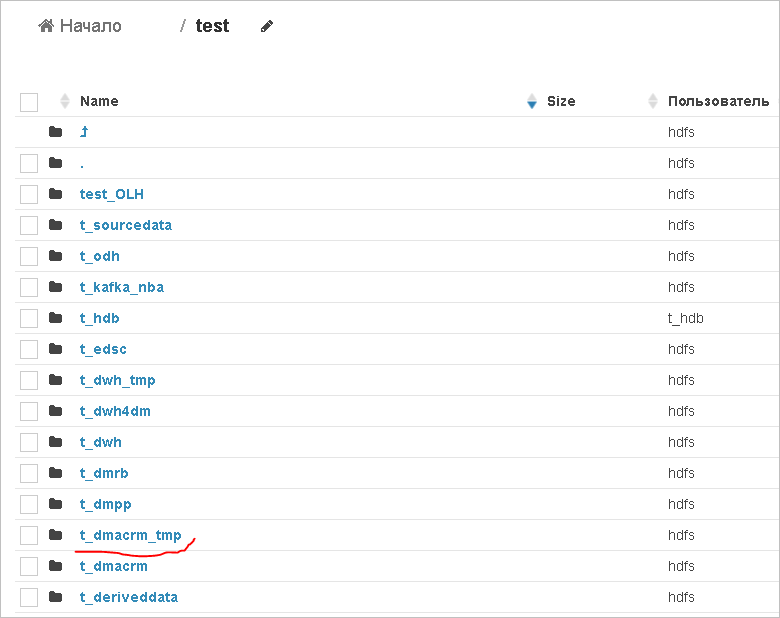
для ТЕСТА смотрим сами - **/test/t\_dmacrm\_tmp/db**

**Путь установки в HDFS** для теста, тоже смотрим в описании сред:

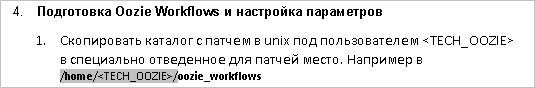


Т.е. в ReleaseNotes.doc описана установка в бой и устанавливать надо не в область **s\_dmacrm\_tmp**

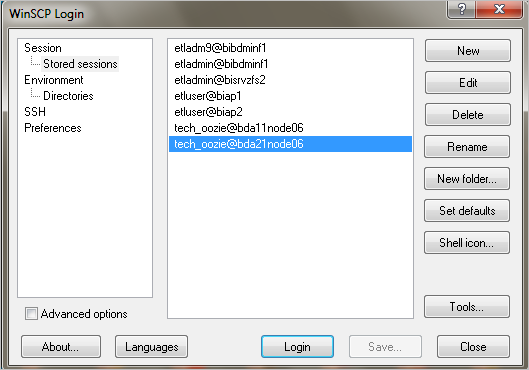
а в **/test/t\_dmacrm\_tmp**:



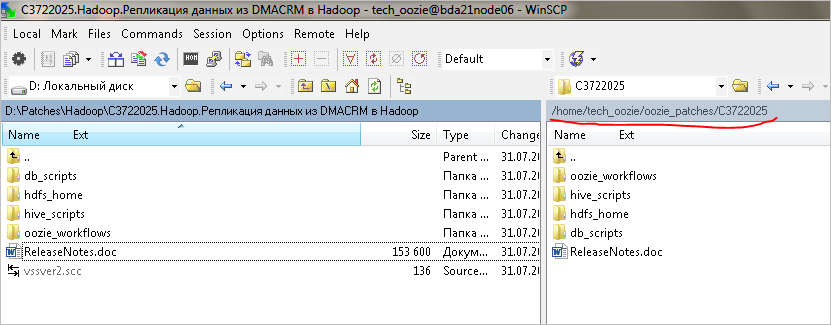
Пункт 4



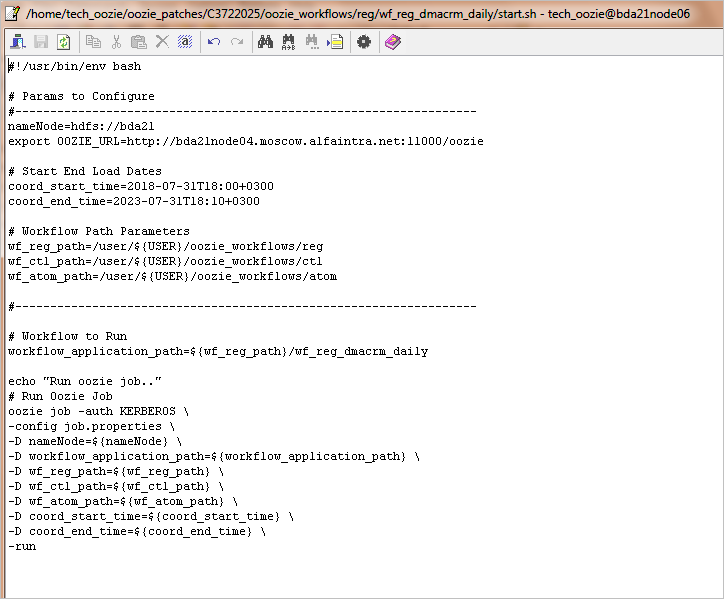
Заходим под tech\_oozie в WinSCP:



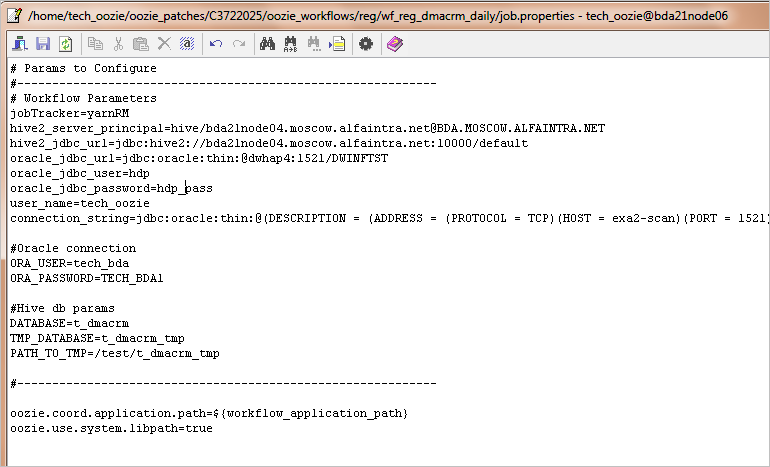
Копируем патч целикомсо своего ПК в **/home/tech\_oozie/oozie\_patches/C3722025**:



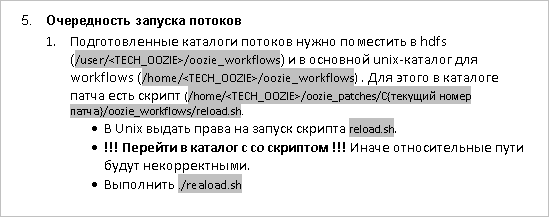
Меняем файлы **start.sh**:



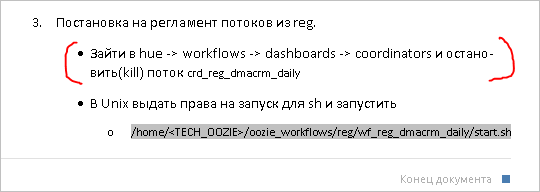
и **job.properties:**



Пункт 5



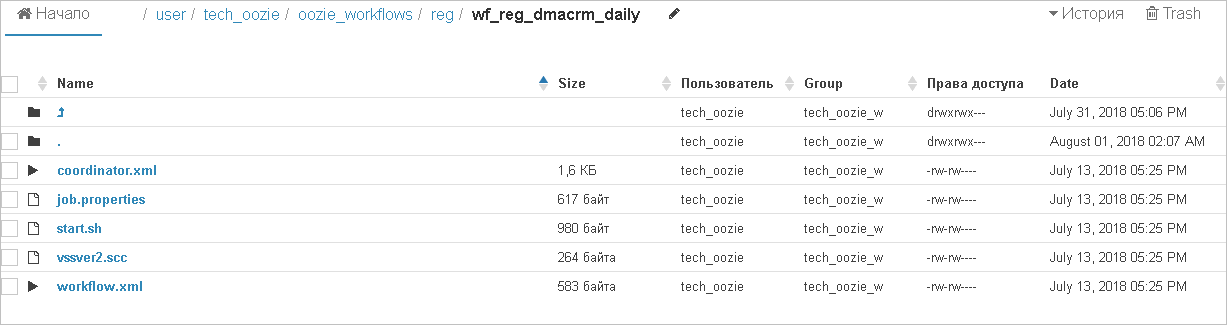
Перед тем, как выполнить **./reload.sh** надо выполнить подпункт 3 – остановить поток **crd\_reg\_dmacrm\_daily**



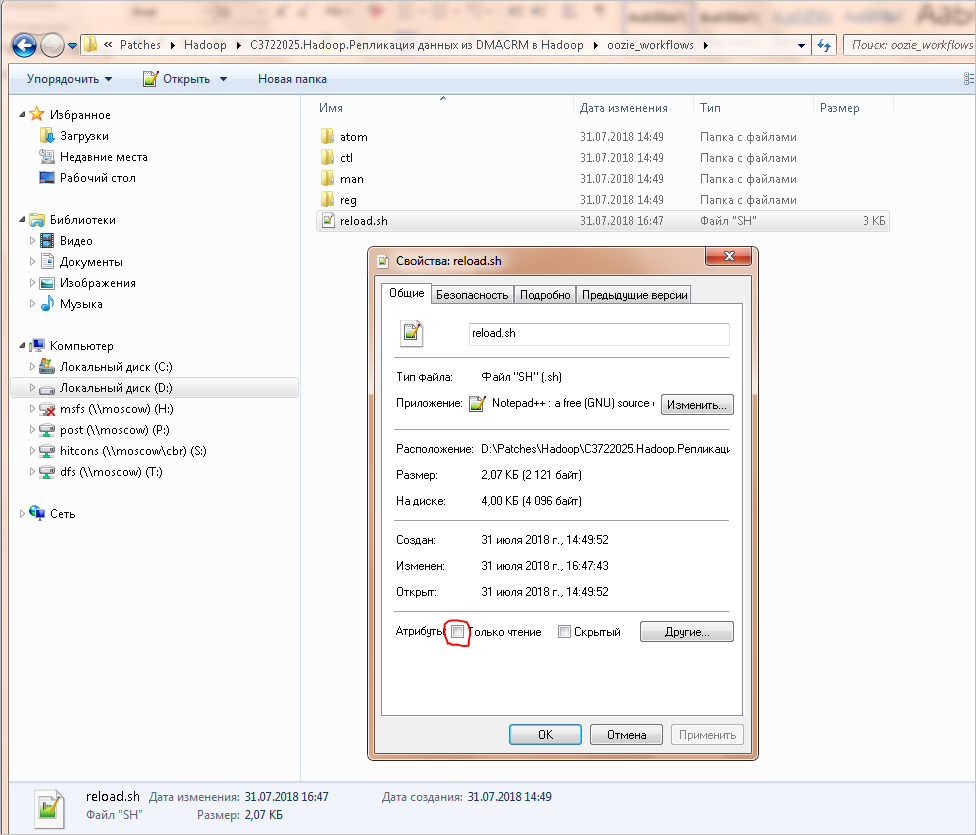
Если этого не сделать, то не сработает строка:



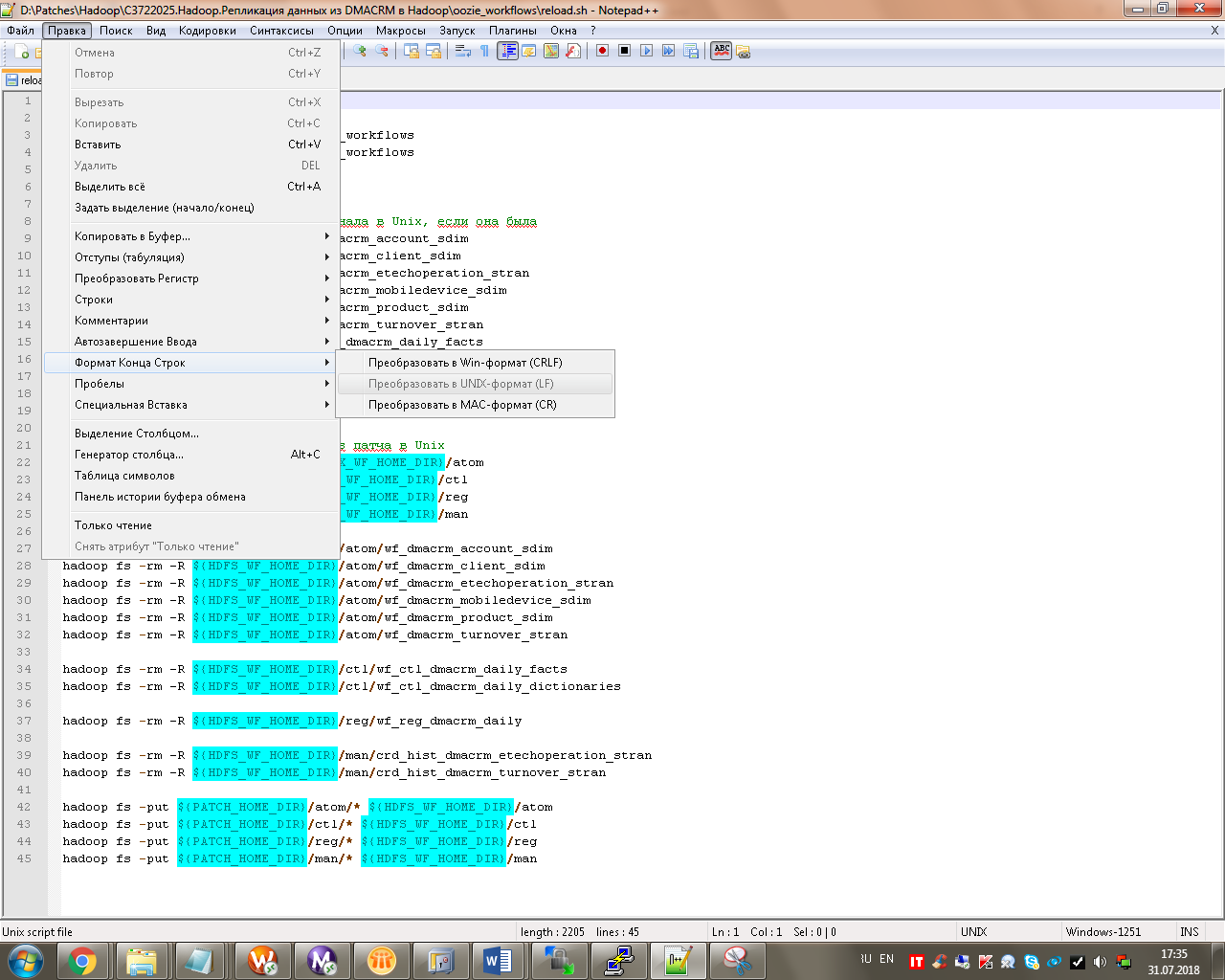
в reload.sh и поток **wf\_reg\_dmacrm\_daily** в hdfs не перепишется (файлы останутся за старую дату):



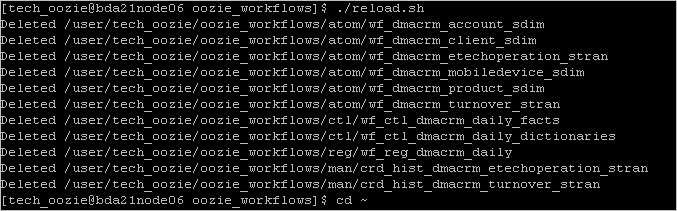
Перед тем, как выполнить **./reload.sh** надо обязательно выполнить преобразование формата конца строк в UNIX-формат. Для этого у себя на ПК у файла убираем галку «Только чтение» и открываем его в Notepad++.



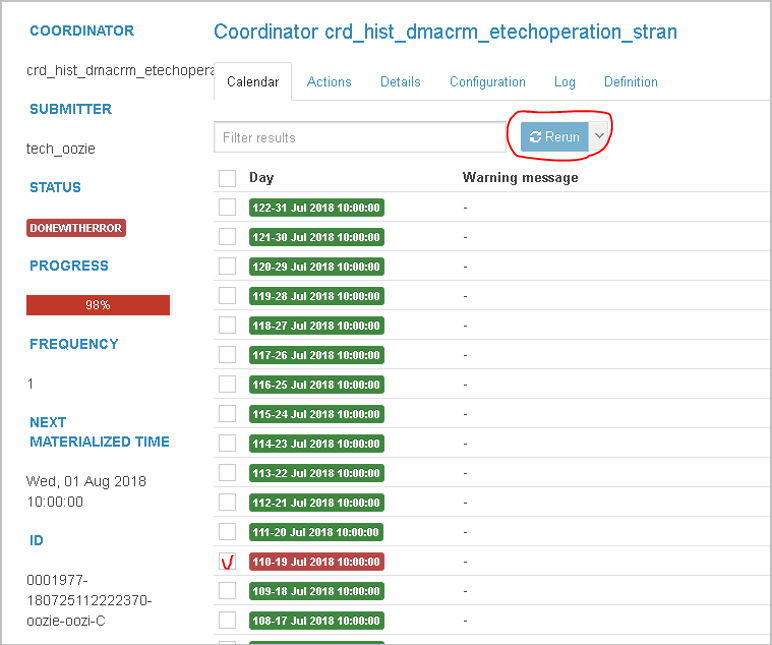
Выбираем **Правка –> Формат Конца Строк --> Преобразовать в UNIX-формат** , сохраняем:



Результат работы **./reload.sh** выглядит так:

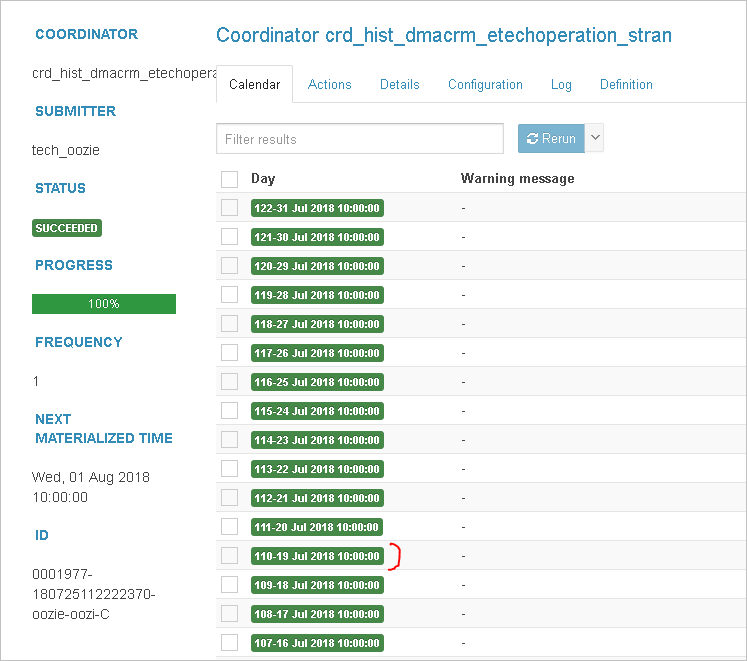


Если при загрузке истории какой-то день упал, его можно перезагрузить, отметив галкой и нажав **Rerun**:



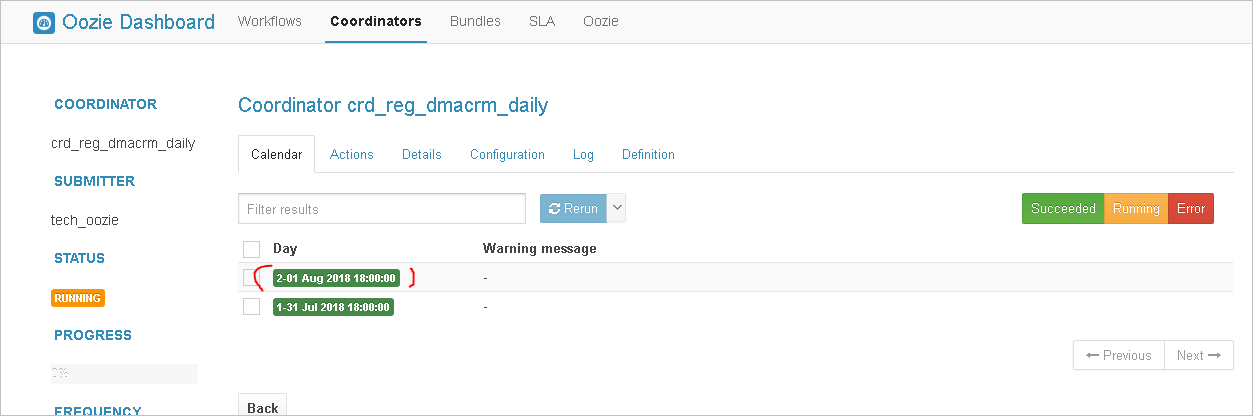
Ничего не менять в окне выбора параметров и нажать **Submit**

День перегружается и поток переходит в состояние **Succeeded:**

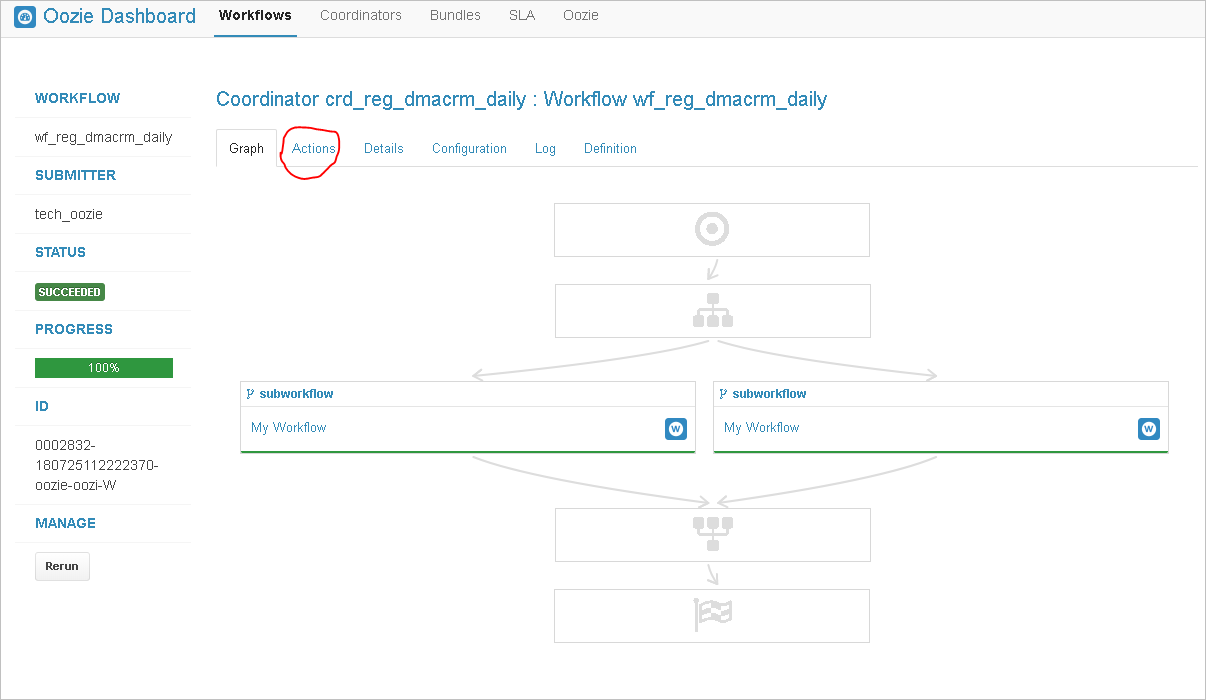


**Просмотр ошибок работы потока**

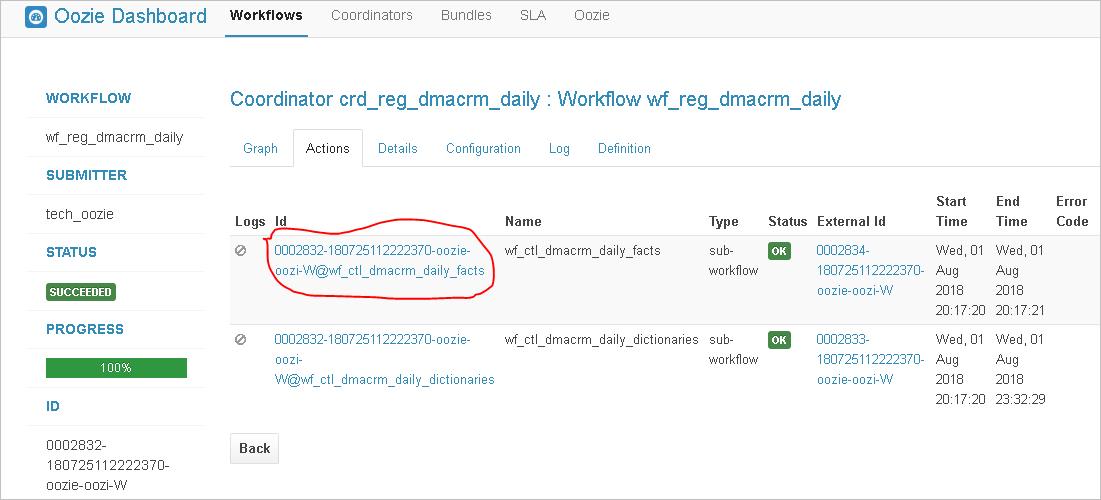
Нажимаем интересующий запуск:



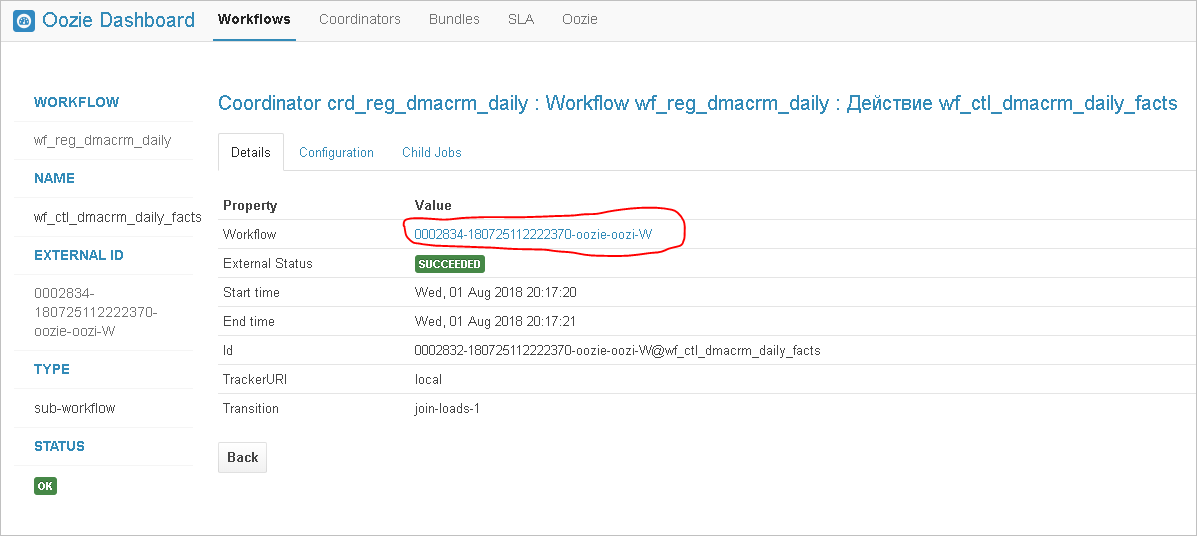
Нажимаем **Action**:



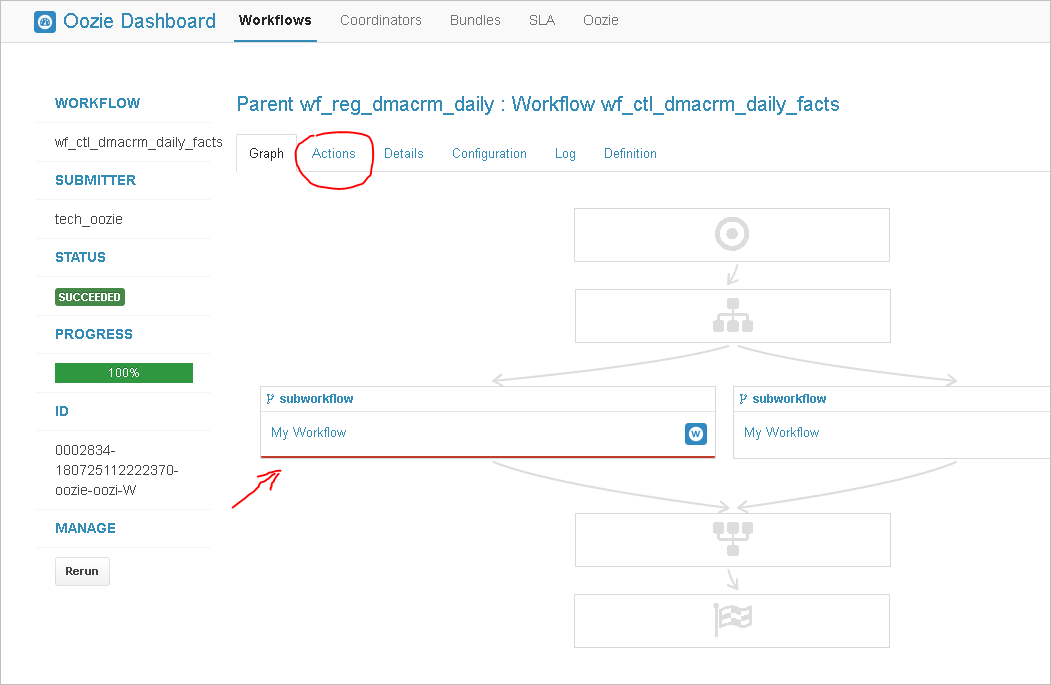
Нажимаем **ссылку**:



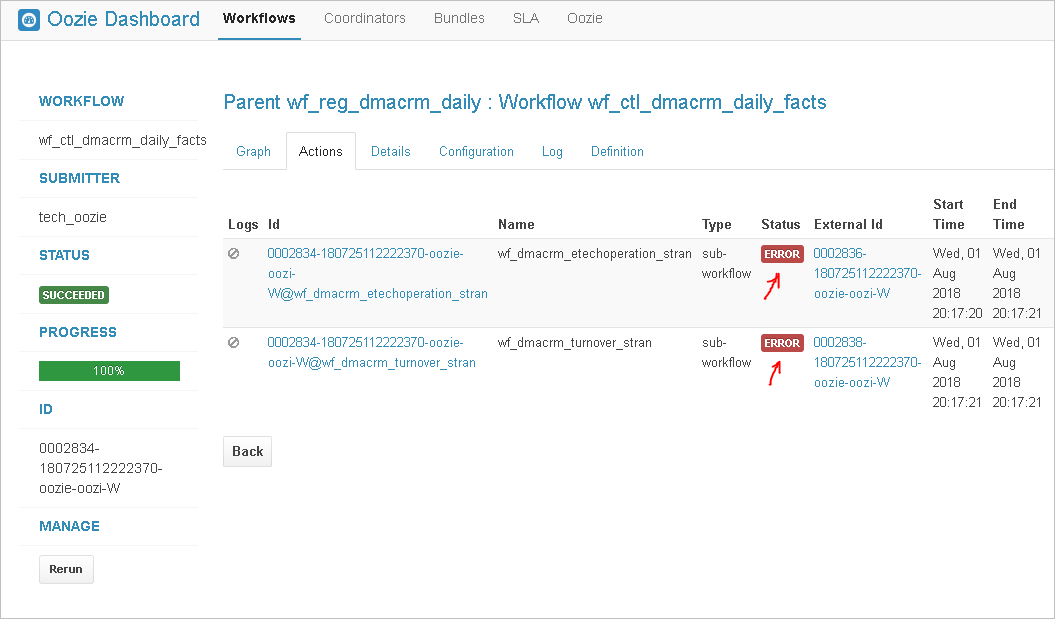
Нажимаем **воркфлоу**:



Нажимаем **Action** (красный цвет полосы говорит о том, что есть ошибки):

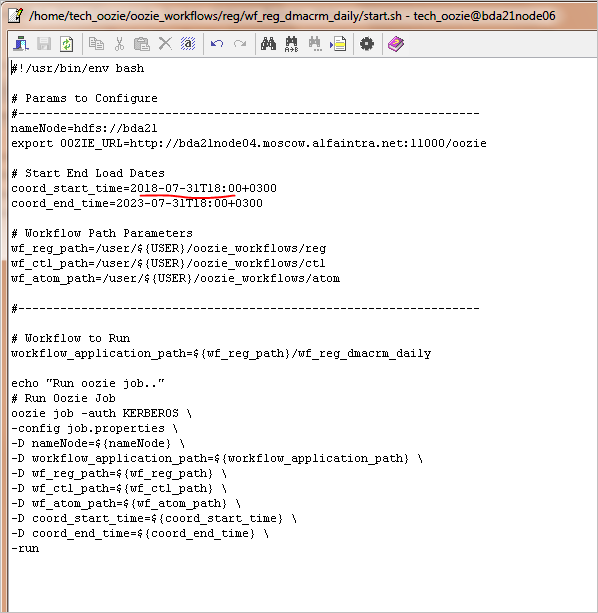


И видим ошибки инкрементальной загрузки двух таблиц (это загрузки в потоке **wf\_reg\_dmacrm\_daily**), для которых мы грузили историю:



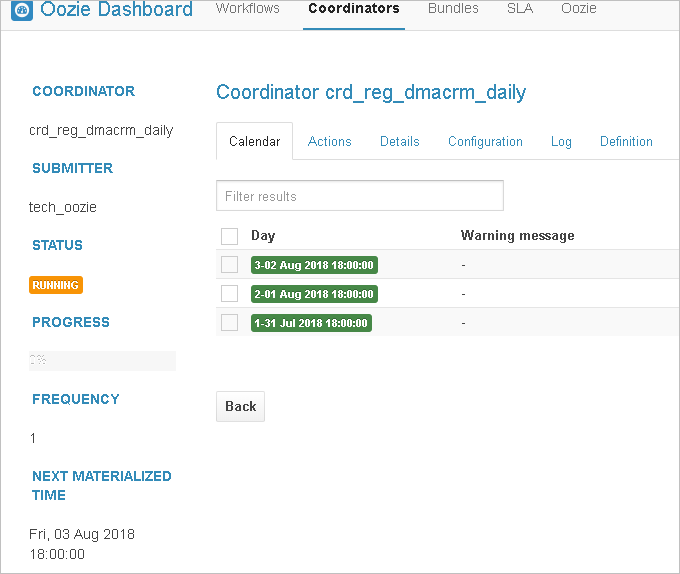
Запуск потока с датой меньшей текущей

Если поток стартуется запуском ./start.sh и в файле указана дата старта меньше текущей:

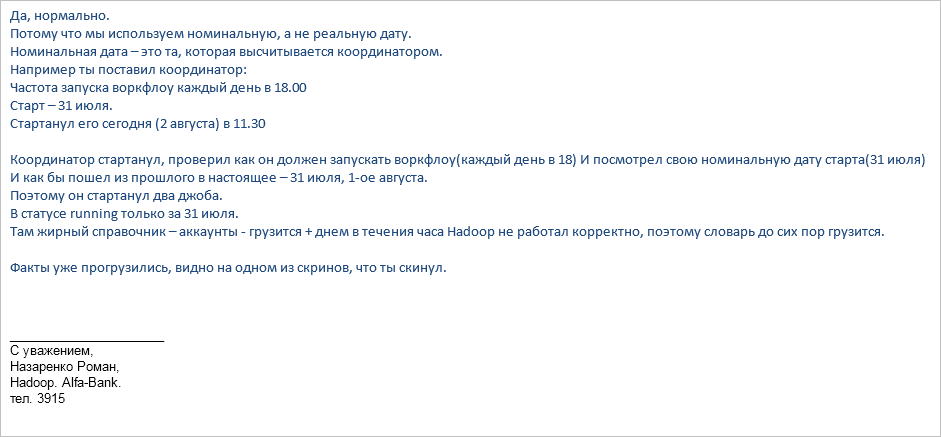


То поток сразу начинает отрабатывать за прошедшие даты, не дожидаясь указанного времени (18.00).

Т.е. в данном случае при старте 2 августа, начинает работать за 31-ое и потом за 1-ое число

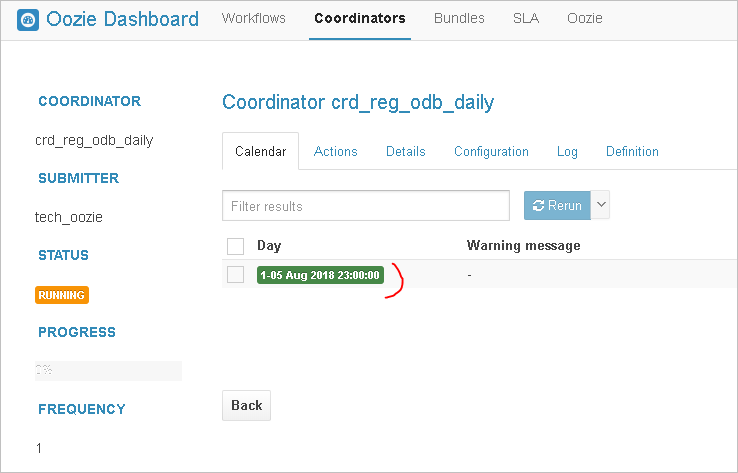


Вот письмо Ромы на эту тему:



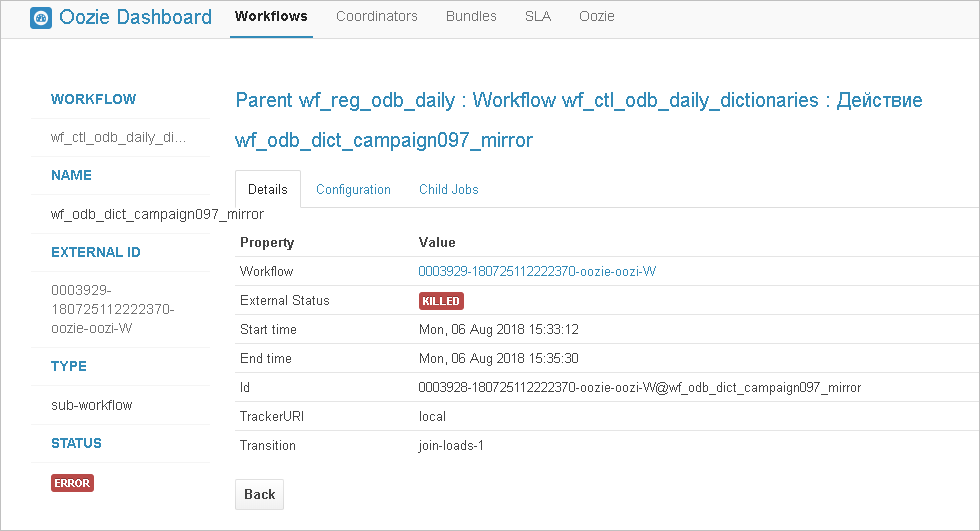
Как посмотреть ошибки

Надо зайти в запуск и провалиться до конца:

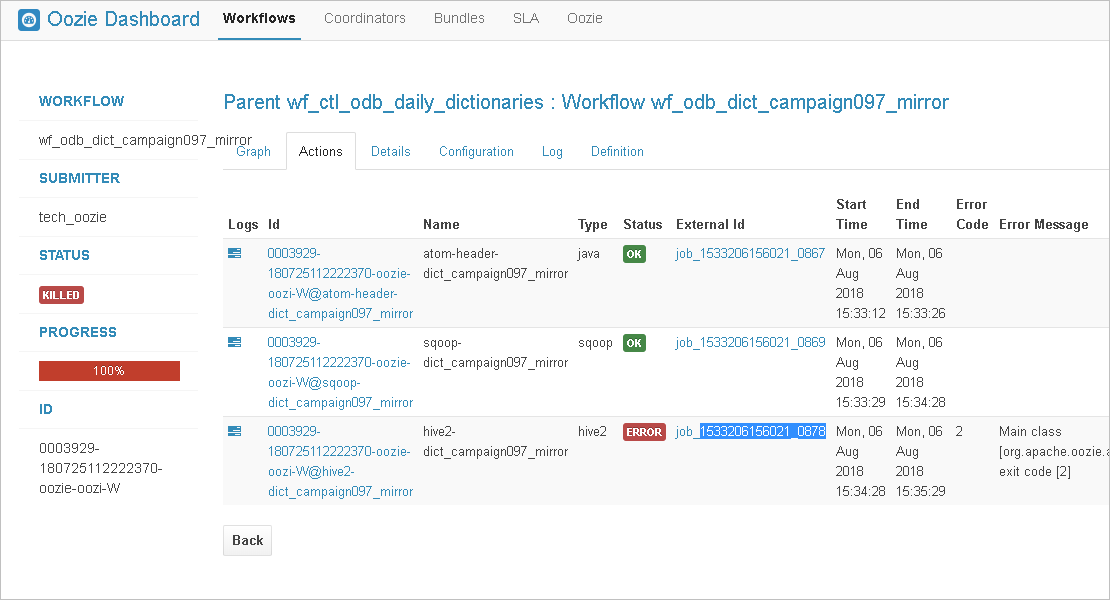








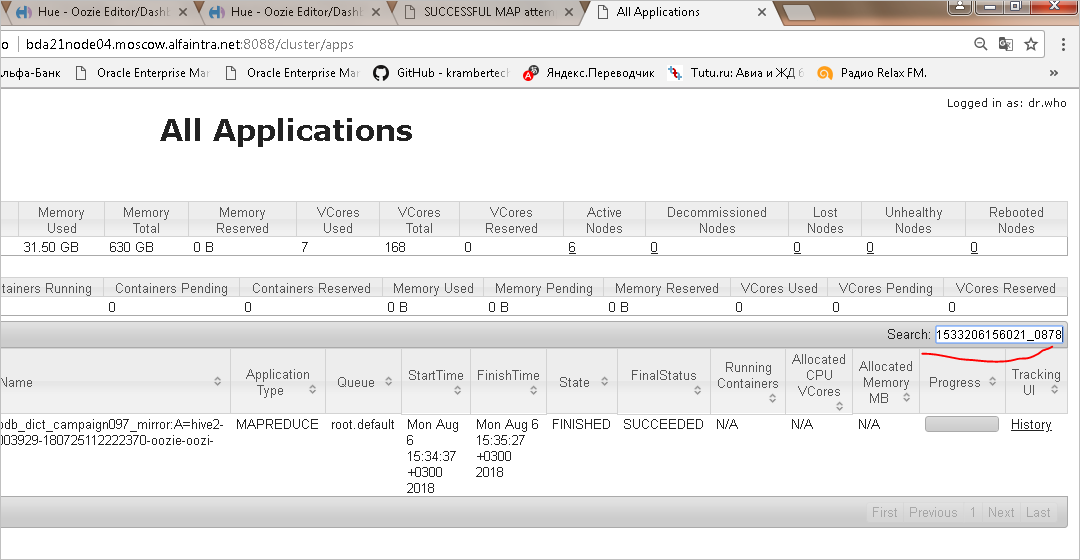
До последнего, не знаю как назвать, “потока” и скопировать **job\_id**



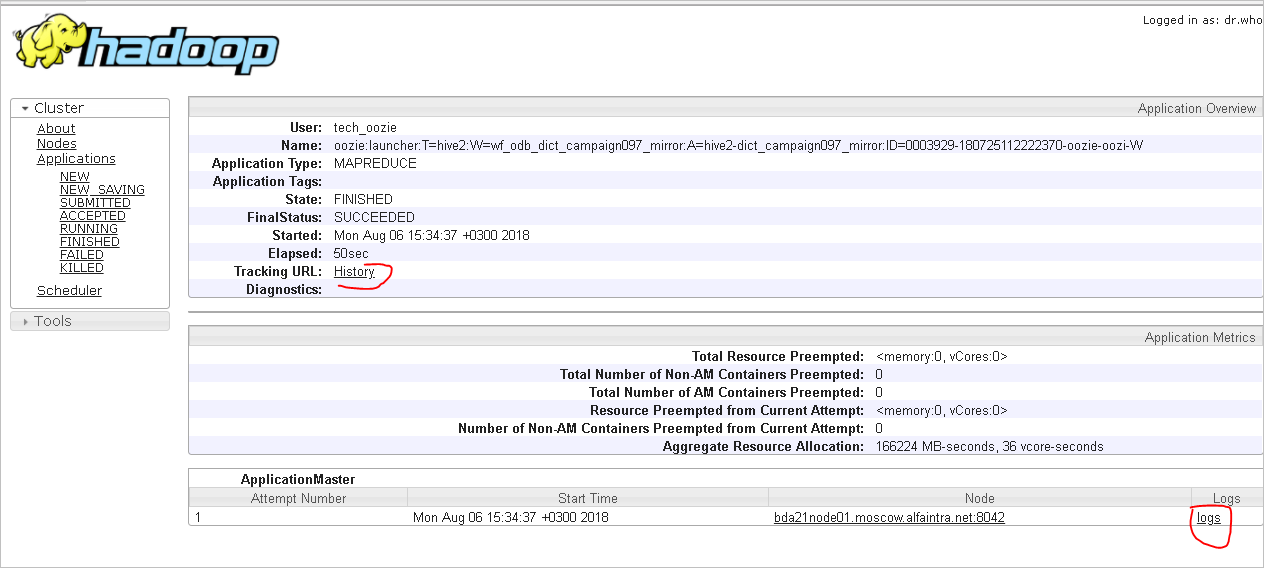
Вставить в <Hadoop YARN> –

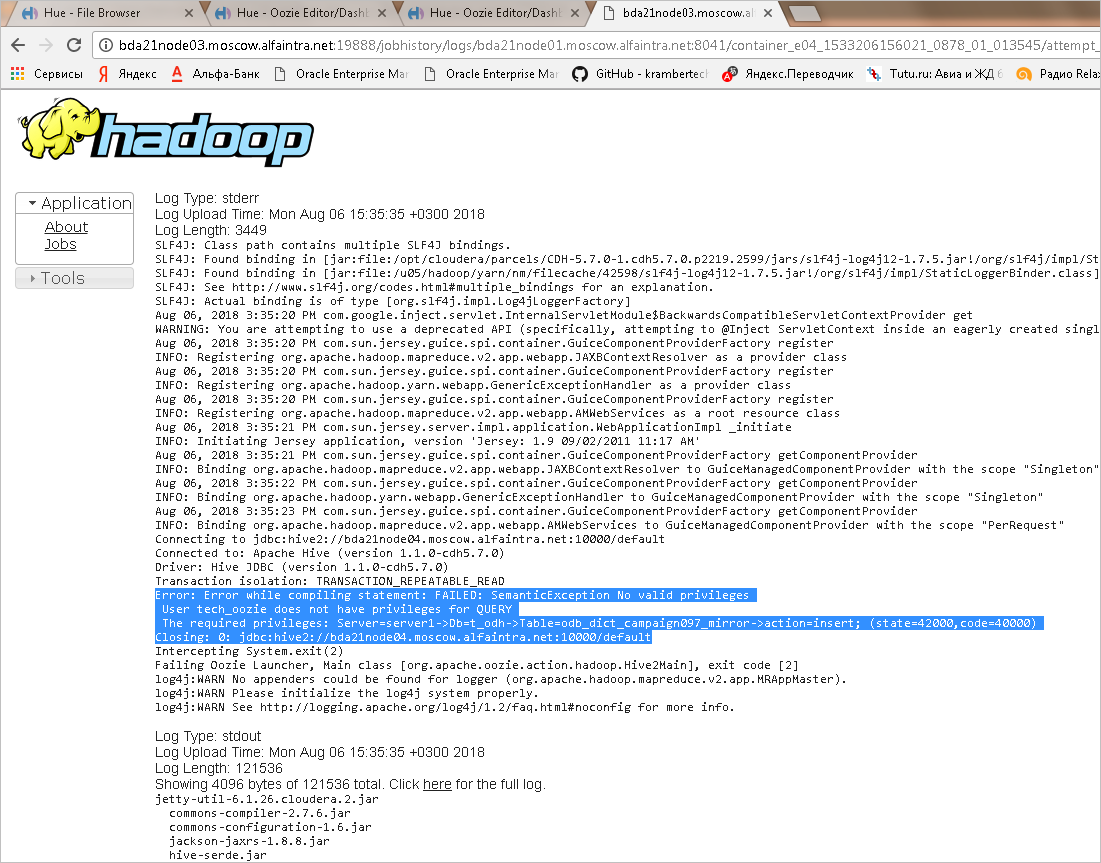
http://bda**11**node03.moscow.alfaintra.net:8088/cluster/apps

[http://bda**21**node04.moscow.alfaintra.net:8088/cluster/apps](http://bda21node04.moscow.alfaintra.net:8088/cluster/apps)



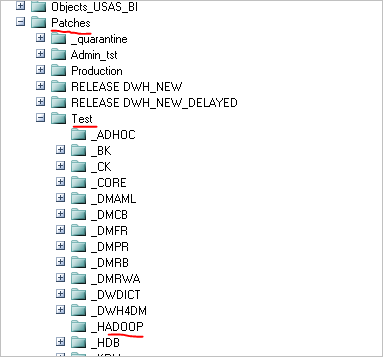
И искать ошибки или в **Logs** или **History**



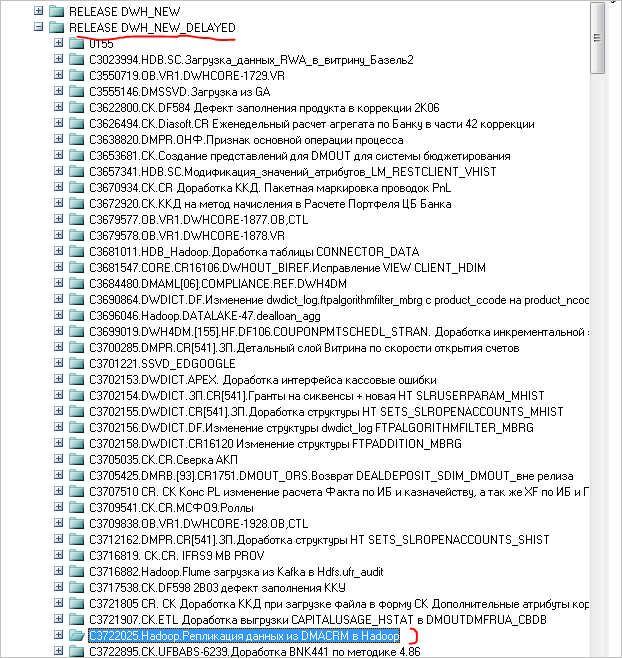


Движение патчей в VSS

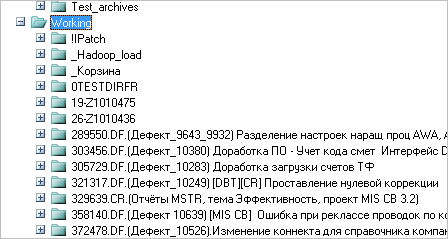
Патчи которые ставятся в ТЕСТ находятся в папке Test:



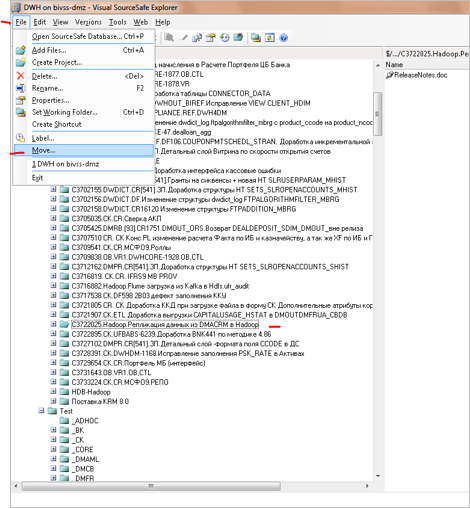
Когда, патч переходит на установку в БОЙ он находится в папке **RELESE DWH\_NEW\_DELAYED**



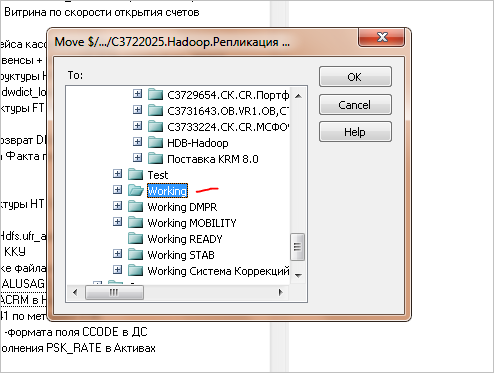
Если в патче надо исправить ошибки, то для того, чтобы с ним могли работать разработчики, его надо спустить в папку **Working**:



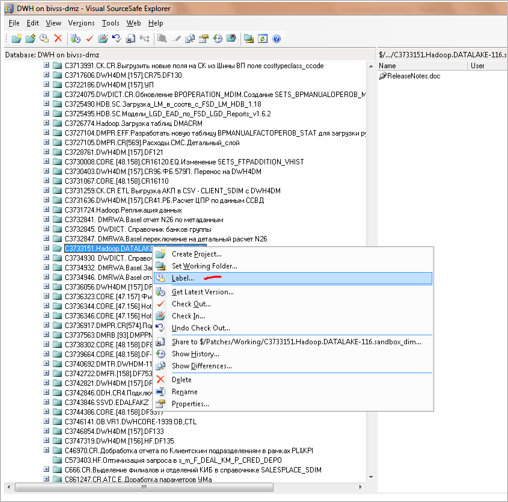
Для этого выбираем патч в списке и верхнем меню **File** -> **Move**



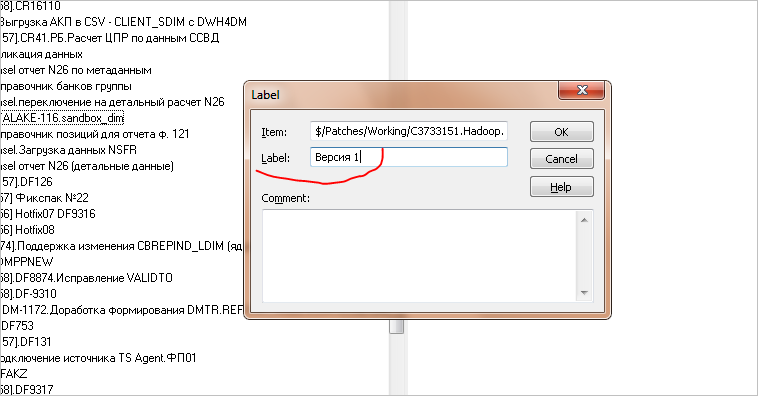
Выбираем **Working** и ОК



После того, как патч исправят, надо его поднять обратно в ТЕСТ и при этом поставить тот Label



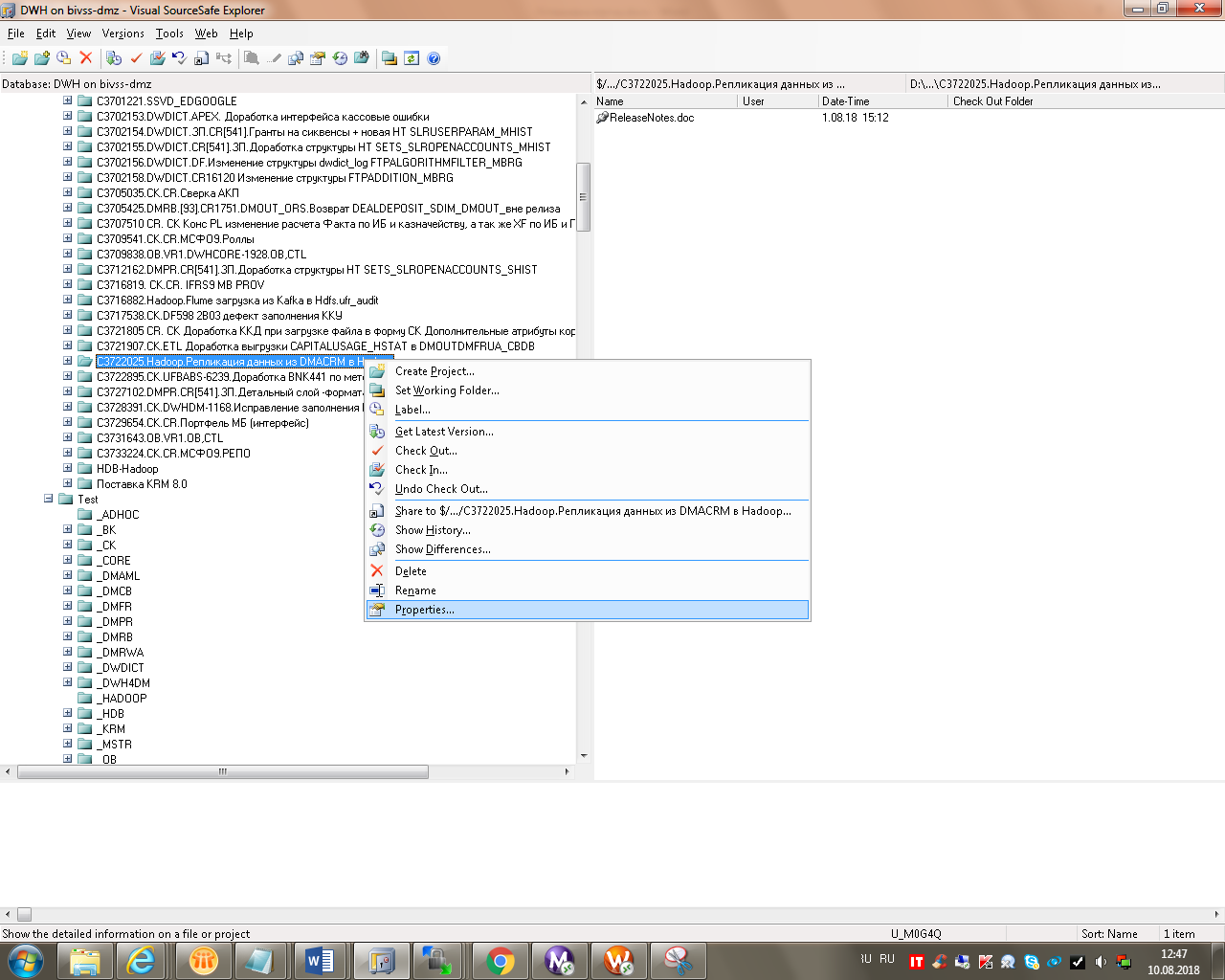
который у него был:

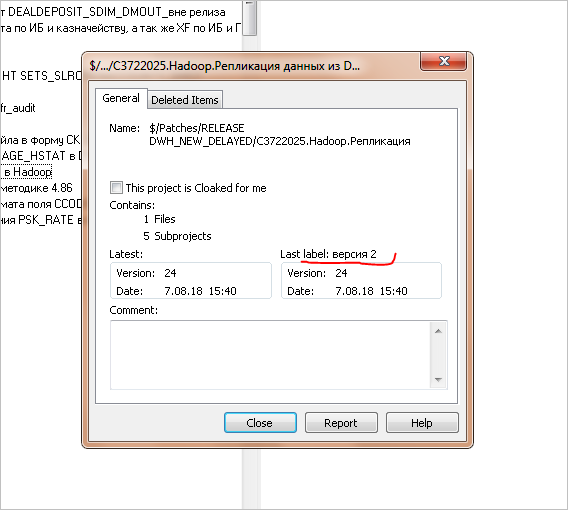


Т.е. если патч исправляется без закрытия задачи и создания новой, то версия сохраняется.

Если патч отклоняется на доработку с закрытием задачи, то Версия становится +1

Эти Версии прописывает Юля Селиваненко и посмотреть какая текущая можно в свойствах патча:





**Когда патч установлен** пишем Перминову Александру пиьмо чтобы перевести его в релиз

Александр, добрый день.

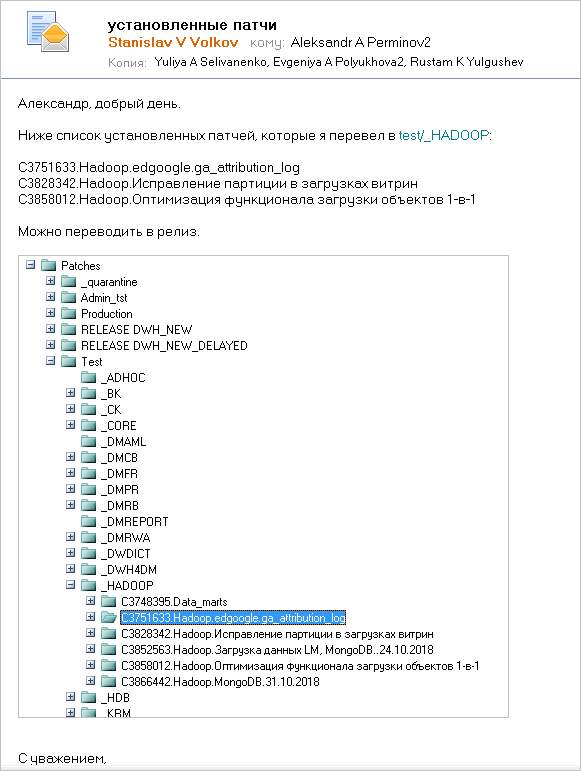
Ниже список установленных патчей в test/\_HADOOP:

C3751633.Hadoop.edgoogle.ga\_attribution\_log

C3828342.Hadoop.Исправление партиции в загрузках витрин

C3858012.Hadoop.Оптимизация функционала загрузки объектов 1-в-1

Можно переводить в релиз.



**Решение для задачи на СТ:**

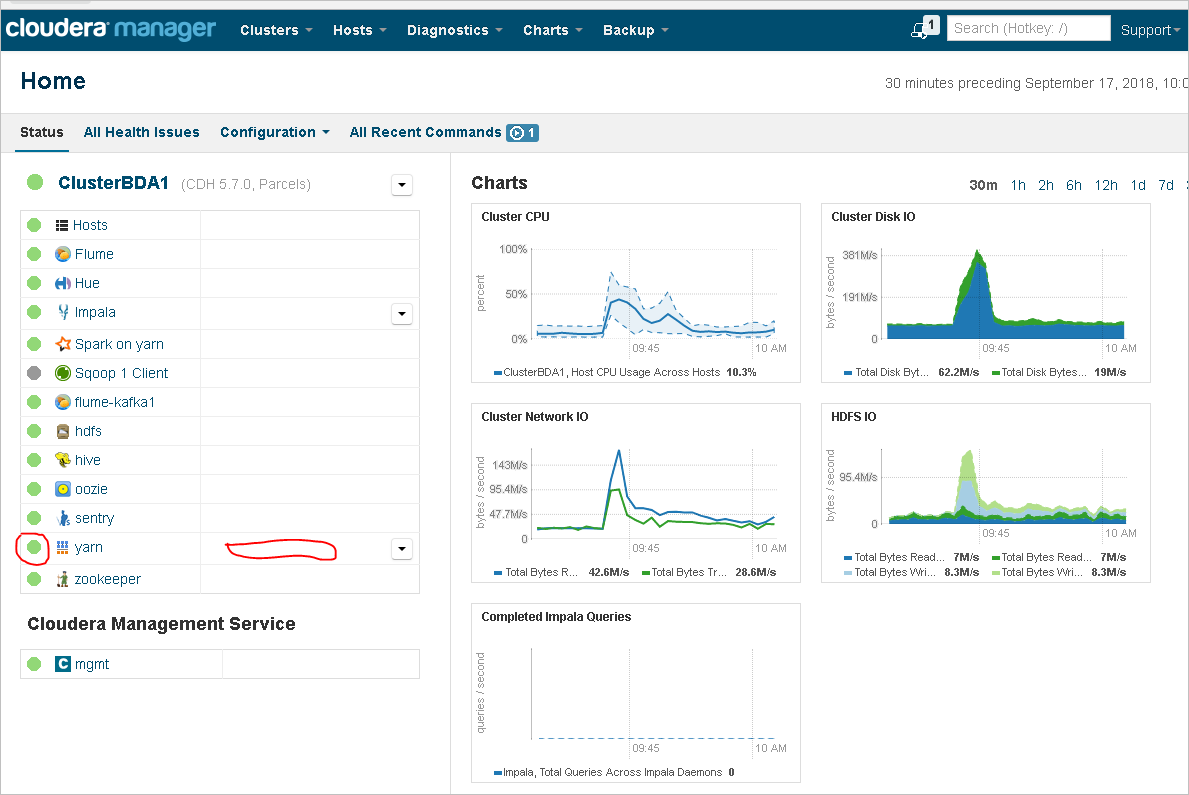
СТ выполнено без ошибок. Патч можно внедрять в БОЙ

Посмотреть состояние кластера можно в Cloudera Manager

http://bda11node03.moscow.alfaintra.net:7180/cmf/home

readonly // readonly

Состояние компонентов не должно быть красным:



Если надо запустить переливку reload.sh для отдельных папок,

можно для этого отредактировать файл таким образом:

#!/usr/bin/env bash

HDFS\_WF\_HOME\_DIR=/user/${USER}/oozie\_workflows

UNIX\_WF\_HOME\_DIR=/home/${USER}/oozie\_workflows

PATCH\_HOME\_DIR=${PWD}

# Перенос новой версии функционала из патча в Unix

cp -Rf ${PATCH\_HOME\_DIR}/atom/wf\_deriveddata\_salary\_income\* ${UNIX\_WF\_HOME\_DIR}/atom

cp -Rf ${PATCH\_HOME\_DIR}/ctl/wf\_ctl\_nba\_deriveddata\_daily\* ${UNIX\_WF\_HOME\_DIR}/ctl

#cp -Rf ${PATCH\_HOME\_DIR}/reg/\* ${UNIX\_WF\_HOME\_DIR}/reg

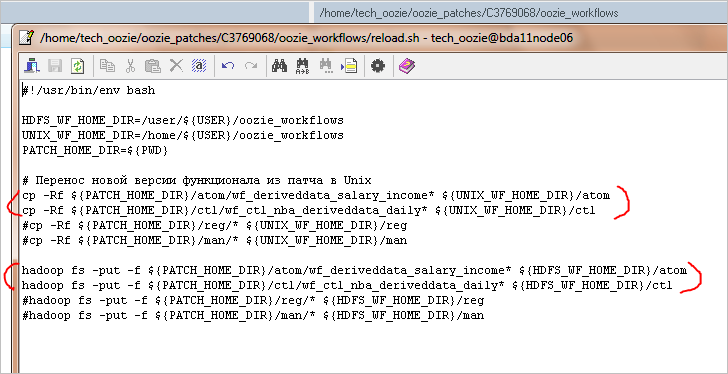
#cp -Rf ${PATCH\_HOME\_DIR}/man/\* ${UNIX\_WF\_HOME\_DIR}/man

hadoop fs -put -f ${PATCH\_HOME\_DIR}/atom/wf\_deriveddata\_salary\_income\* ${HDFS\_WF\_HOME\_DIR}/atom

hadoop fs -put -f ${PATCH\_HOME\_DIR}/ctl/wf\_ctl\_nba\_deriveddata\_daily\* ${HDFS\_WF\_HOME\_DIR}/ctl

#hadoop fs -put -f ${PATCH\_HOME\_DIR}/reg/\* ${HDFS\_WF\_HOME\_DIR}/reg

#hadoop fs -put -f ${PATCH\_HOME\_DIR}/man/\* ${HDFS\_WF\_HOME\_DIR}/man



Команда для запуска **run\_hive\_scripts\_user.sh**

с добавкой для записи логов в отдельный файл log.txt вместо вывода на экран

sh ./run\_hive\_scripts.sh  >  log.txt  2>&1

Параметр отвечающий за то в каком пуле будут запускаться лаунчеры находится здесь:

(если поставить default будут запускаться в tech\_oozie)



Задача выполнена, дефектов не зафиксировано