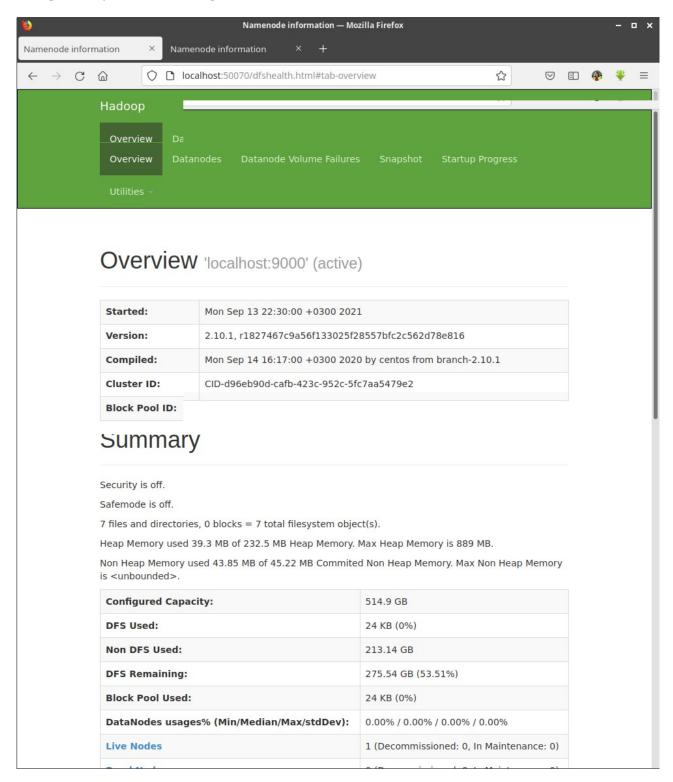
Скачиваем файлы образа и собераем его, выполнив команду, находясь внутри директории:

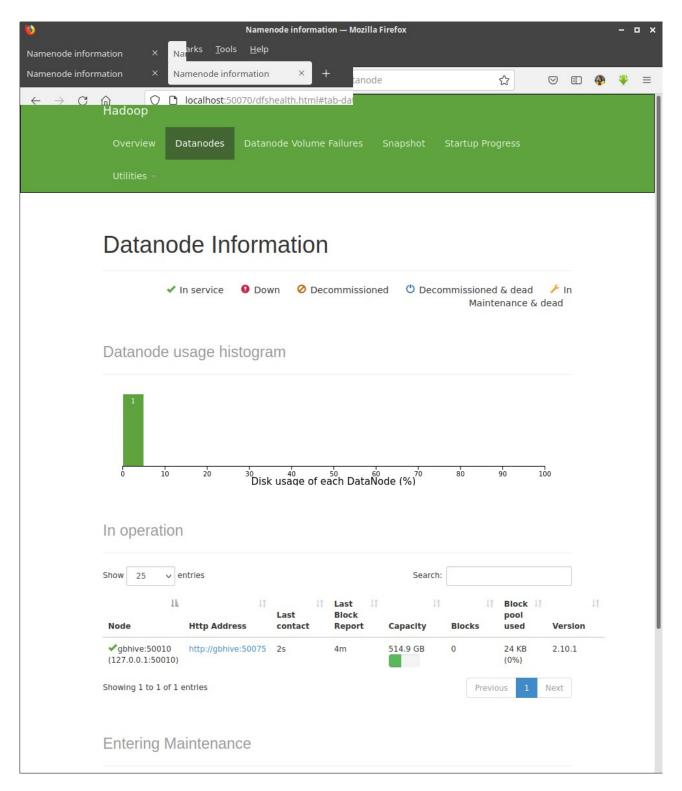
docker build -t hive-hadoop.

ПОДНИМАЕМ ЛОКАЛЬНЫЙ КЛАСТЕР

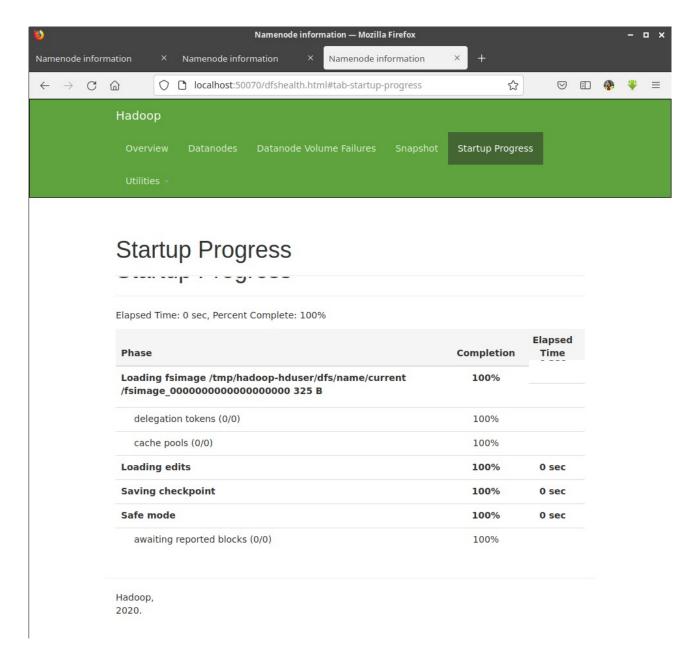
Смотрим статус NameNode: http://localhost:50070



Смотрим статус Datanode Information: http://localhost:50070

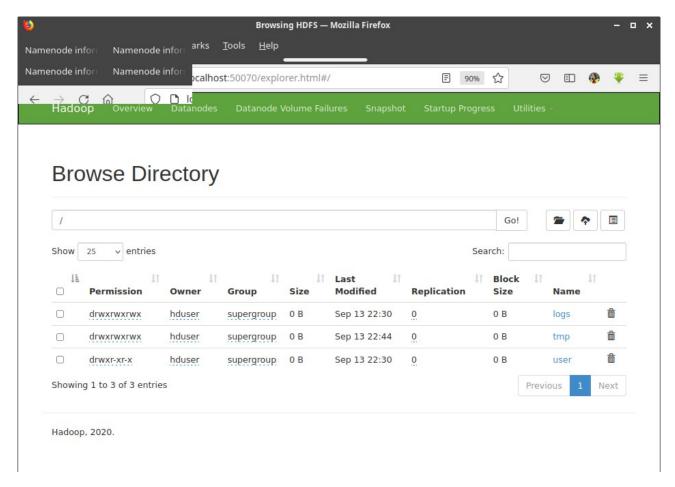


Смотрим статус Startup Progress: http://localhost:50070



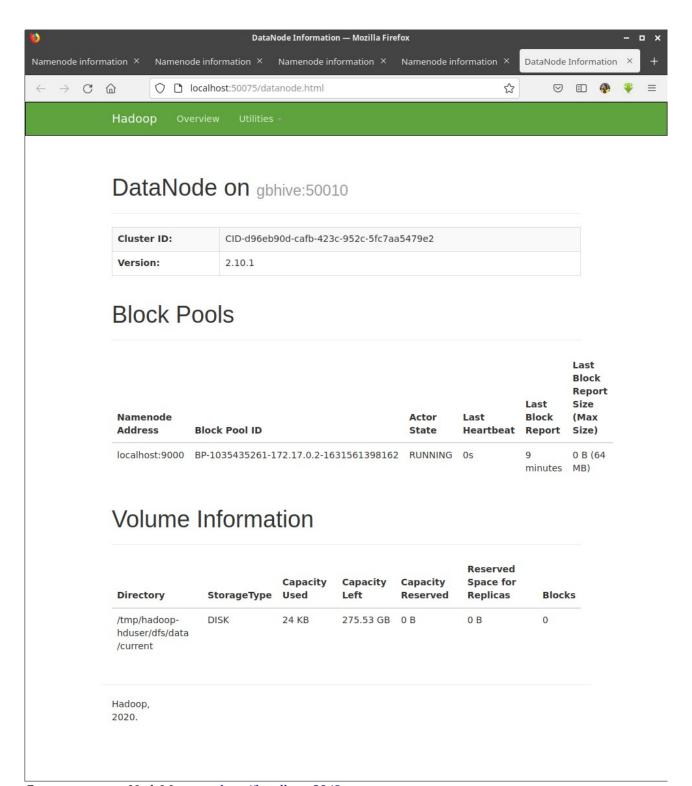
ПОДНИМАЕМ ЛОКАЛЬНЫЙ КЛАСТЕР

Смотрим Browse Directory: http://localhost:50070

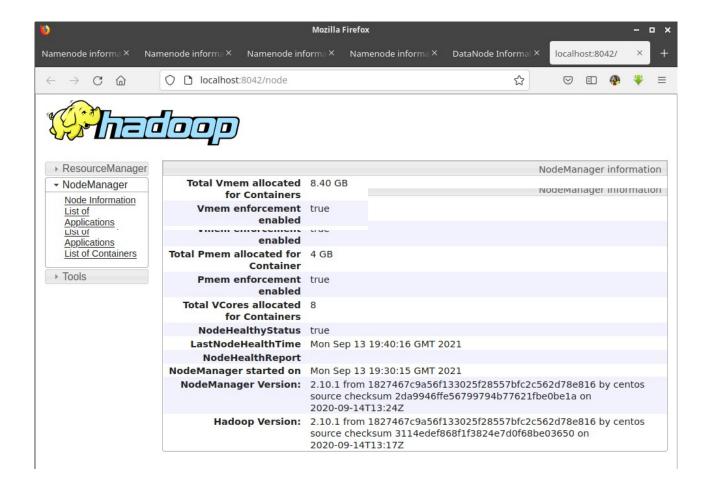


ПОДНИМАЕМ ЛОКАЛЬНЫЙ КЛАСТЕР

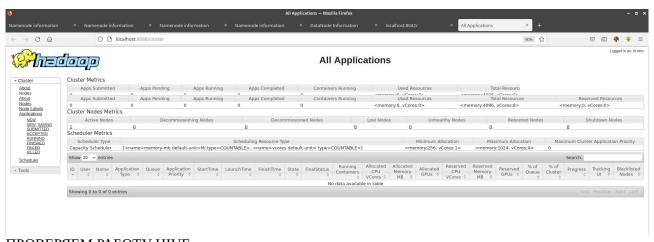
Смотрим статус DataNode: http://localhost:50075



Смотрим статус NodeManager: http://localhost:8042



Смотрим статус ResourceManager: http://localhost:8088



ПРОВЕРЯЕМ РАБОТУ HIVE

Выведем список таблиц в hive:

```
$_
                        Терминал - hduser@gbhive: ~
                                                                           Правка
                          Терминал
 Файл
                   Вид
                                      Вкладки
                                                  Справка
0.0.0.0: Warning: Permanently added '0.0.0.0' (ECDSA) to the list of known hosts
0.0.0.0: starting secondarynamenode, logging to /home/hduser/hadoop/logs/hadoop-
hduser-secondarynamenode-gbhive.out
starting yarn daemons
starting resourcemanager, logging to /home/hduser/hadoop/logs/yarn--resourcemana
localhost: Warning: Permanently added 'localhost' (ECDSA) to the list of known h
localhost: starting nodemanager, logging to /home/hduser/hadoop/logs/yarn-hduser
-nodemanager-gbhive.out
hduser@gbhive:~$ hive -e 'show tables;'
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hduser/hive/lib/log4j-slf4j-impl-2.6.2.j
ar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/hduser/hadoop/share/hadoop/common/lib/sl
f4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Logging initialized using configuration in jar:file:/home/hduser/hive/lib/hive-c
ommon-2.3.9.jar!/hive-log4j2.properties Async: true
Time taken: 20.176 seconds
```

PAGOTA C CRON

Добавляем в систему: sudo apt-get install cron

Включаем:

sudo service cron start

Открываем для редактирования:

crontab -e

Добавляем правила:

* * * * * echo `date` >> /home/hduser/datus.log

0**** hdfs dfs -put /home/hduser/datus.log "/user/hduser/`date +%s`.log"

0 * * * * wget -q mail.ru -O /home/hduser/mail-ru.html

Смотрим существующие правила: crontab -l

```
💶 Терминал - igor@igor-MS-7808: ~/Настройка потоков данных. Apache Airflo 📁 🗖
 Файл
         Правка
                   Вид
                         Терминал
                                      Вкладки
                                                 Справка
 проблемы зависимостей — оставляем не настроенным
При обработке следующих пакетов произошли ошибки:
python3-lib2to3
python3-distutils
 libglib2.0-dev-bin
 libglib2.0-dev:amd64
libharfbuzz-dev:amd64
libicu-le-hb-dev:amd64
libicu-dev
libxml2-dev:amd64
(base) igor@igor-MS-7808:~/Настройка потоков данных. Apache Airflow/1. Планирова
ние задач. Введение Apache AirFlow/hive-hdp-image$ sudo service cron start
(base) igor@igor-MS-7808:~/Настройка потоков данных. Apache Airflow/1. Планирова
ние задач. Введение Apache AirFlow/hive-hdp-image$ crontab -e
no crontab for igor - using an empty one
Select an editor. To change later, run 'select-editor'.

    /bin/nano

                  <---- easiest
 /usr/bin/vim.basic
 /usr/bin/vim.tiny
 4. /usr/bin/emacs25
 /bin/ed
Choose 1-5 [1]:
```

Запуск команды каждые 4 часа:

0 */4 * * * ./run.sh

Задачи определяют «что запускать?», а операторы — «как запустить?»