Департамент образования и науки города Москвы

Государственное автономное образовательное учреждение

высшего образования города Москвы

«Московский городской педагогический университет»

Институт цифрового образования

Департамент информатики, управления и технологий

ДИСЦИПЛИНА:

Инструменты для хранения и обработки больших данных

Практическая работа 6

Тема:

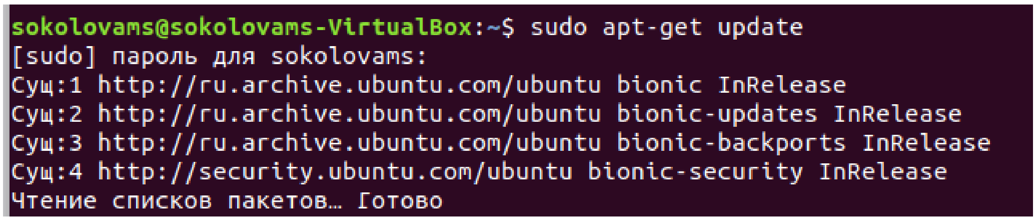
Introduction to YARN + Hive

Выполнила: Соколова М. С., группа: АДЭУ-201

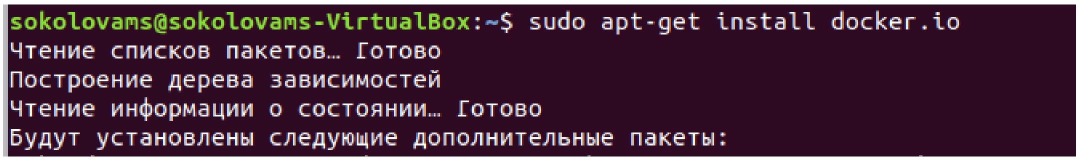
Преподаватель: Босенко Т. М.

Москва

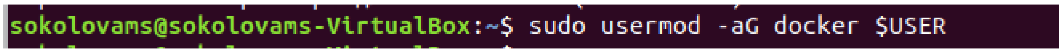
2023



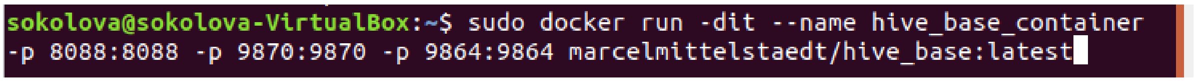
```
sokolovams@sokolovams-VirtualBox:~$ sudo apt-get update
[sudo] пароль для sokolovams:
Сущ:1 http://ru.archive.ubuntu.com/ubuntu bionic InRelease
Сущ:2 http://ru.archive.ubuntu.com/ubuntu bionic-updates InRelease
Сущ:3 http://ru.archive.ubuntu.com/ubuntu bionic-backports InRelease
Сущ:4 http://security.ubuntu.com/ubuntu bionic-security InRelease
Чтение списков пакетов… Готово
```



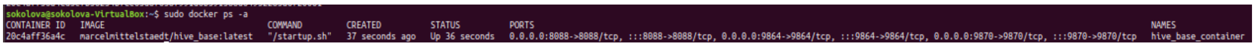
```
sokolovams@sokolovams-VirtualBox:~$ sudo apt-get install docker.io
Чтение списков пакетов… Готово
Построение дерева зависимостей
Чтение информации о состоянии… Готово
Будут установлены следующие дополнительные пакеты:
```



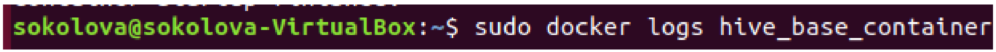
```
sokolovams@sokolovams-VirtualBox:~$ sudo usermod -aG docker $USER
```



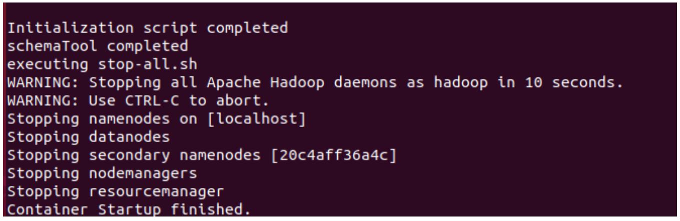
```
sokolova@sokolova-VirtualBox:~$ sudo docker run -dit --name hive_base_container
-p 8088:8088 -p 9870:9870 -p 9864:9864 marcelmittelstaedt/hive_base:latest
```



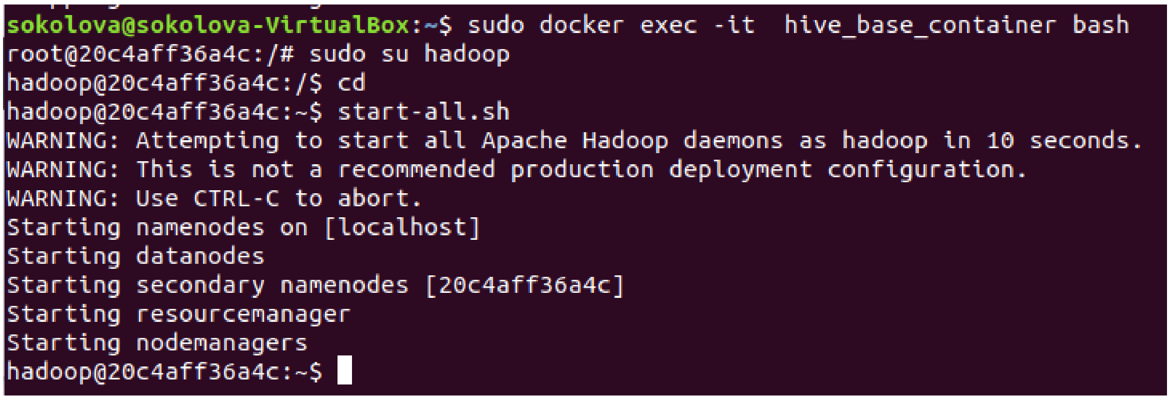
```
sokolova@sokolova-VirtualBox:~$ sudo docker ps -a
CONTAINER ID   IMAGE                                 COMMAND          CREATED          STATUS          PORTS                                                                                                                                    NAMES
20c4aff36a4c   marcelmittelstaedt/hive_base:latest   "/startup.sh"    37 seconds ago   Up 36 seconds   0.0.0.0:8088->8088/tcp, :::8088->8088/tcp, 0.0.0.0:9864->9864/tcp, :::9864->9864/tcp, 0.0.0.0:9870->9870/tcp, :::9870->9870/tcp   hive_base_container
```



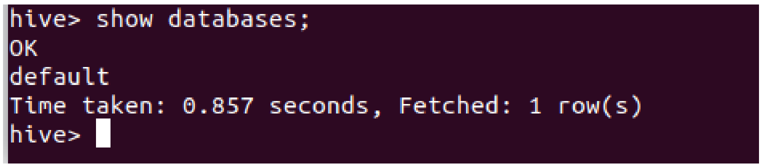
```
sokolova@sokolova-VirtualBox:~$ sudo docker logs hive_base_container
```



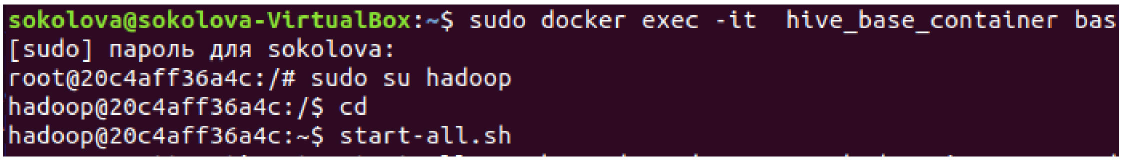
```
Initialization script completed
schemaTool completed
executing stop-all.sh
WARNING: Stopping all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [20c4aff36a4c]
Stopping nodemanagers
Stopping resourcemanager
Container Startup finished.
```



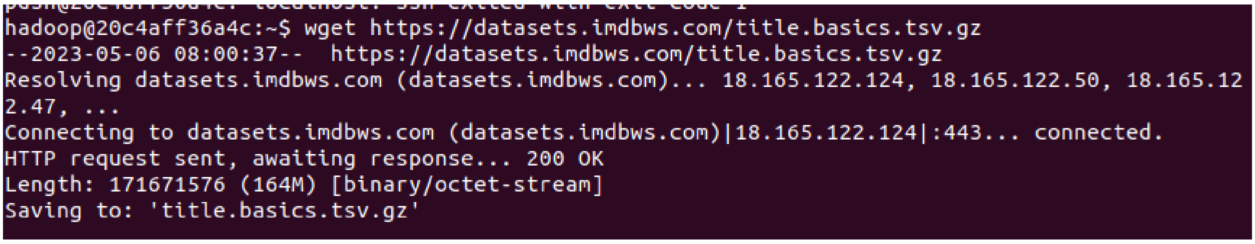
```
sokolova@sokolova-VirtualBox:~$ sudo docker exec -it  hive_base_container bash
root@20c4aff36a4c:/# sudo su hadoop
hadoop@20c4aff36a4c:/$ cd
hadoop@20c4aff36a4c:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [20c4aff36a4c]
Starting resourcemanager
Starting nodemanagers
hadoop@20c4aff36a4c:~$
```



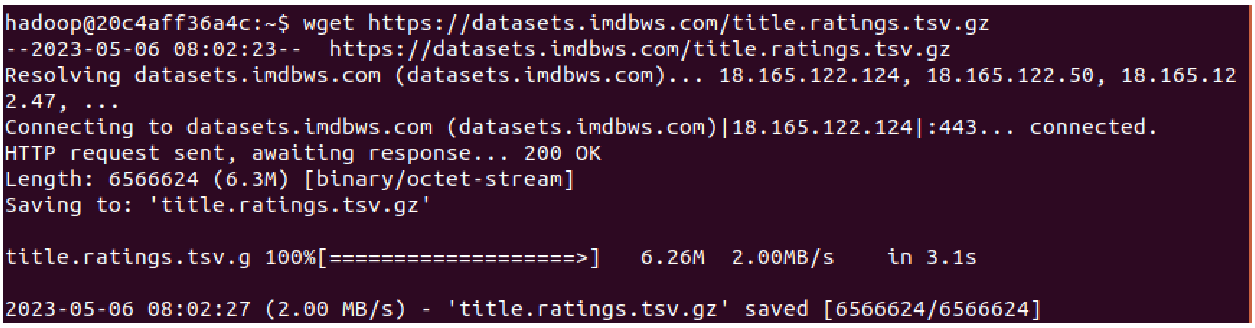
```
hive> show databases;
OK
default
Time taken: 0.857 seconds, Fetched: 1 row(s)
hive>
```



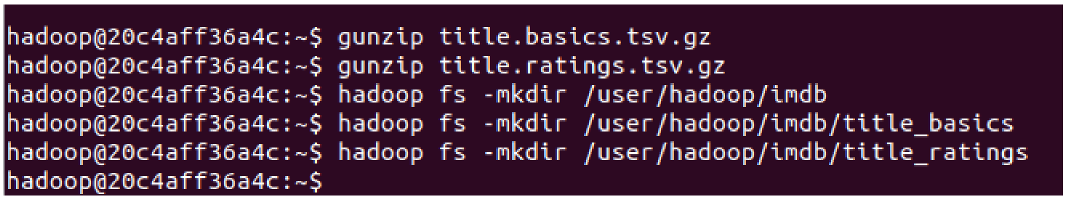
```
sokolova@sokolova-VirtualBox:~$ sudo docker exec -it  hive_base_container bas
[sudo] пароль для sokolova:
root@20c4aff36a4c:/# sudo su hadoop
hadoop@20c4aff36a4c:/$ cd
hadoop@20c4aff36a4c:~$ start-all.sh
```



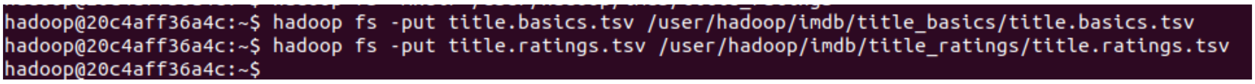
```
hadoop@20c4aff36a4c:~$ wget https://datasets.imdbws.com/title.basics.tsv.gz
--2023-05-06 08:00:37--  https://datasets.imdbws.com/title.basics.tsv.gz
Resolving datasets.imdbws.com (datasets.imdbws.com)... 18.165.122.124, 18.165.122.50, 18.165.12
2.47, ...
Connecting to datasets.imdbws.com (datasets.imdbws.com)|18.165.122.124|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 171671576 (164M) [binary/octet-stream]
Saving to: 'title.basics.tsv.gz'
```



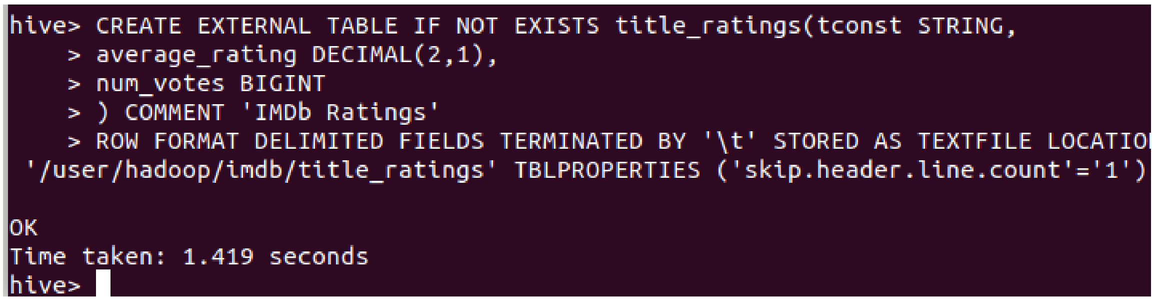
```
hadoop@20c4aff36a4c:~$ wget https://datasets.imdbws.com/title.ratings.tsv.gz
--2023-05-06 08:02:23--  https://datasets.imdbws.com/title.ratings.tsv.gz
Resolving datasets.imdbws.com (datasets.imdbws.com)... 18.165.122.124, 18.165.122.50, 18.165.12
2.47, ...
Connecting to datasets.imdbws.com (datasets.imdbws.com)|18.165.122.124|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 6566624 (6.3M) [binary/octet-stream]
Saving to: 'title.ratings.tsv.gz'

title.ratings.tsv.g 100%[===================>]   6.26M  2.00MB/s    in 3.1s

2023-05-06 08:02:27 (2.00 MB/s) - 'title.ratings.tsv.gz' saved [6566624/6566624]
```



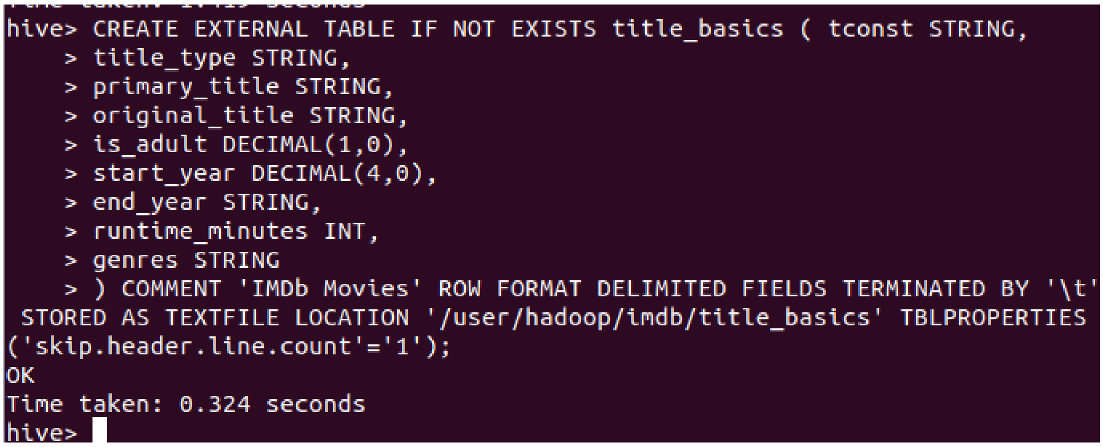
```
hadoop@20c4aff36a4c:~$ gunzip title.basics.tsv.gz
hadoop@20c4aff36a4c:~$ gunzip title.ratings.tsv.gz
hadoop@20c4aff36a4c:~$ hadoop fs -mkdir /user/hadoop/imdb
hadoop@20c4aff36a4c:~$ hadoop fs -mkdir /user/hadoop/imdb/title_basics
hadoop@20c4aff36a4c:~$ hadoop fs -mkdir /user/hadoop/imdb/title_ratings
hadoop@20c4aff36a4c:~$
```



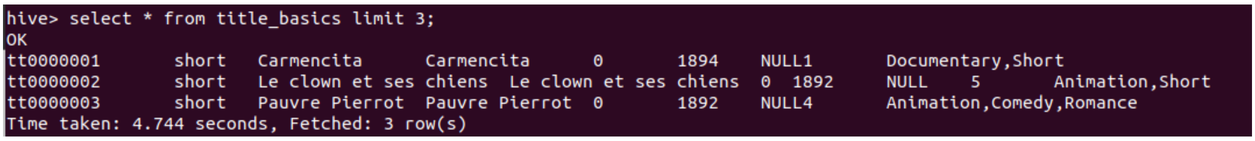
```
hadoop@20c4aff36a4c:~$ hadoop fs -put title.basics.tsv /user/hadoop/imdb/title_basics/title.basics.tsv
hadoop@20c4aff36a4c:~$ hadoop fs -put title.ratings.tsv /user/hadoop/imdb/title_ratings/title.ratings.tsv
hadoop@20c4aff36a4c:~$
```



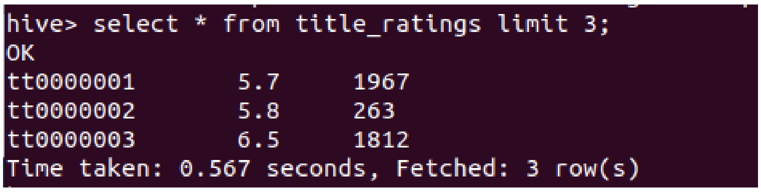
```
hive> CREATE EXTERNAL TABLE IF NOT EXISTS title_ratings(tconst STRING,
    > average_rating DECIMAL(2,1),
    > num_votes BIGINT
    > ) COMMENT 'IMDb Ratings'
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE LOCATIO
 '/user/hadoop/imdb/title_ratings' TBLPROPERTIES ('skip.header.line.count'='1')

OK
Time taken: 1.419 seconds
hive>
```



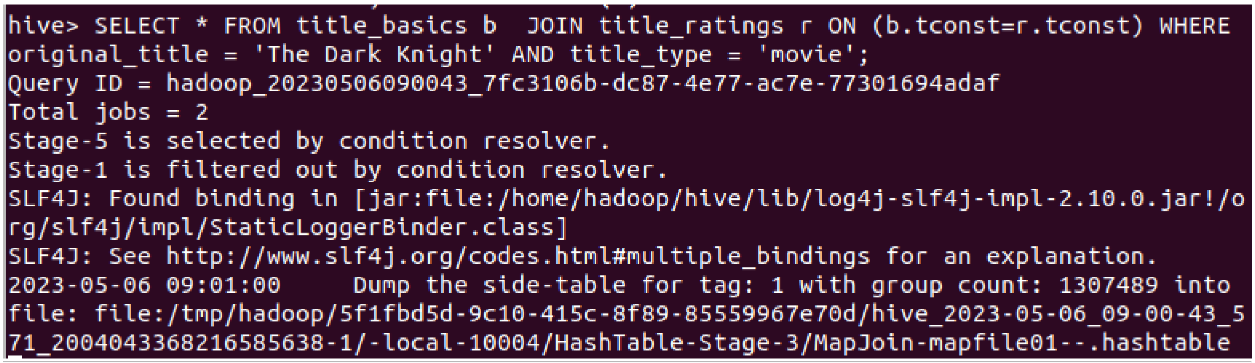
```
hive> CREATE EXTERNAL TABLE IF NOT EXISTS title_basics ( tconst STRING,
    > title_type STRING,
    > primary_title STRING,
    > original_title STRING,
    > is_adult DECIMAL(1,0),
    > start_year DECIMAL(4,0),
    > end_year STRING,
    > runtime_minutes INT,
    > genres STRING
    > ) COMMENT 'IMDb Movies' ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'
 STORED AS TEXTFILE LOCATION '/user/hadoop/imdb/title_basics' TBLPROPERTIES
('skip.header.line.count'='1');
OK
Time taken: 0.324 seconds
hive>
```



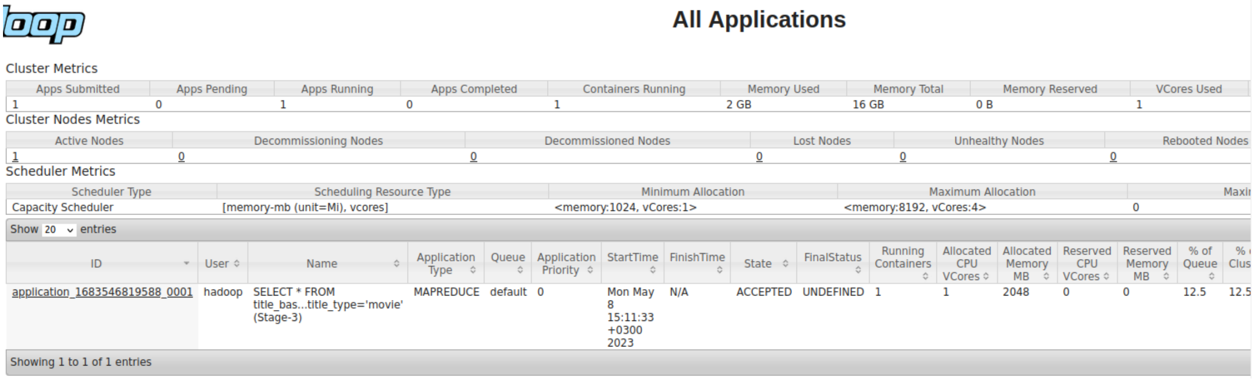
```
hive> select * from title_basics limit 3;
OK
tt0000001       short   Carmencita      Carmencita      0       1894    NULL1       Documentary,Short
tt0000002       short   Le clown et ses chiens  Le clown et ses chiens  0  1892     NULL    5       Animation,Short
tt0000003       short   Pauvre Pierrot  Pauvre Pierrot  0       1892    NULL4       Animation,Comedy,Romance
Time taken: 4.744 seconds, Fetched: 3 row(s)
```



```
hive> select * from title_ratings limit 3;
OK
tt0000001       5.7     1967
tt0000002       5.8     263
tt0000003       6.5     1812
Time taken: 0.567 seconds, Fetched: 3 row(s)
```



```
hive> SELECT * FROM title_basics b  JOIN title_ratings r ON (b.tconst=r.tconst) WHERE
original_title = 'The Dark Knight' AND title_type = 'movie';
Query ID = hadoop_20230506090043_7fc3106b-dc87-4e77-ac7e-77301694adaf
Total jobs = 2
Stage-5 is selected by condition resolver.
Stage-1 is filtered out by condition resolver.
SLF4J: Found binding in [jar:file:/home/hadoop/hive/lib/log4j-slf4j-impl-2.10.0.jar!/o
rg/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2023-05-06 09:01:00     Dump the side-table for tag: 1 with group count: 1307489 into
file: file:/tmp/hadoop/5f1fbd5d-9c10-415c-8f89-85559967e70d/hive_2023-05-06_09-00-43_5
71_2004043368216585638-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile01--.hashtable
```



All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used
1	0	1	0	1	2 GB	16 GB	0 B	1

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
1	0	0	0	0	0

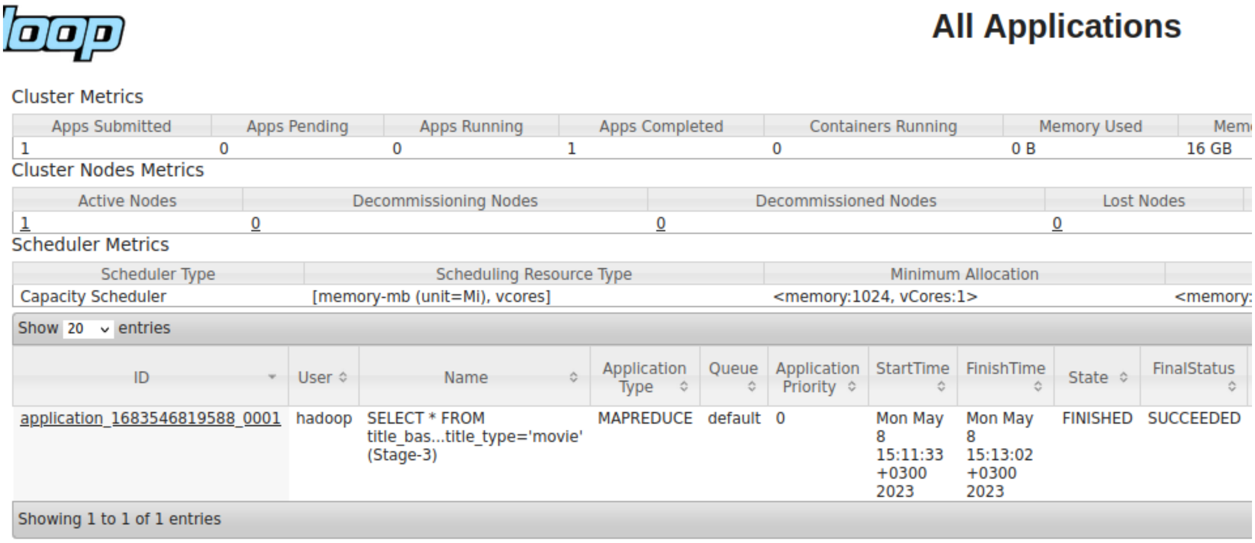
Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maxi
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	0

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Reserved CPU VCores	Reserved Memory MB	% of Queue	% Clus
application_1683546819588_0001	hadoop	SELECT * FROM title_bas...title_type='movie' (Stage-3)	MAPREDUCE	default	0	Mon May 8 15:11:33 +0300 2023	N/A	ACCEPTED	UNDEFINED	1	1	2048	0	0	12.5	12.5

Showing 1 to 1 of 1 entries



All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Mem
1	0	0	1	0	0 B	16 GB

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes
1	0	0	0

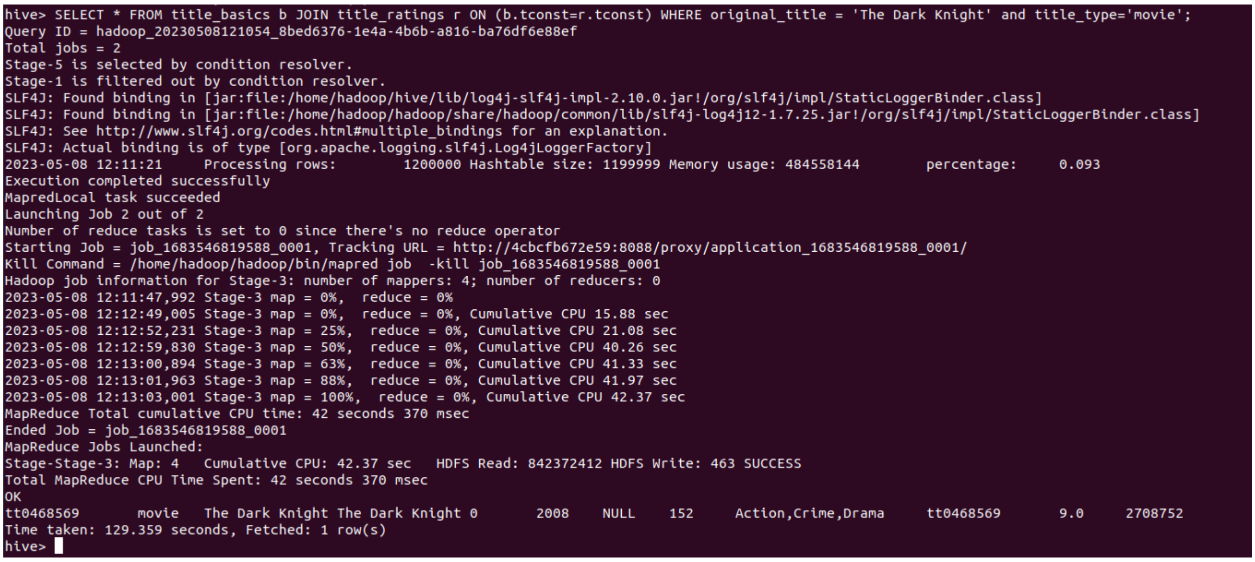
Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:

Show 20 entries

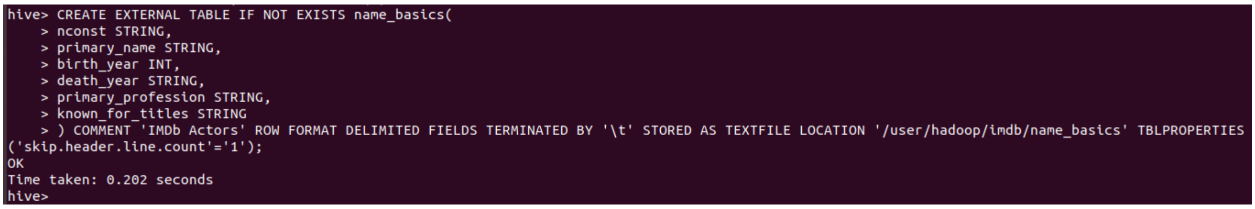
ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus
application_1683546819588_0001	hadoop	SELECT * FROM title_bas...title_type='movie' (Stage-3)	MAPREDUCE	default	0	Mon May 8 15:11:33 +0300 2023	Mon May 8 15:13:02 +0300 2023	FINISHED	SUCCEEDED

Showing 1 to 1 of 1 entries



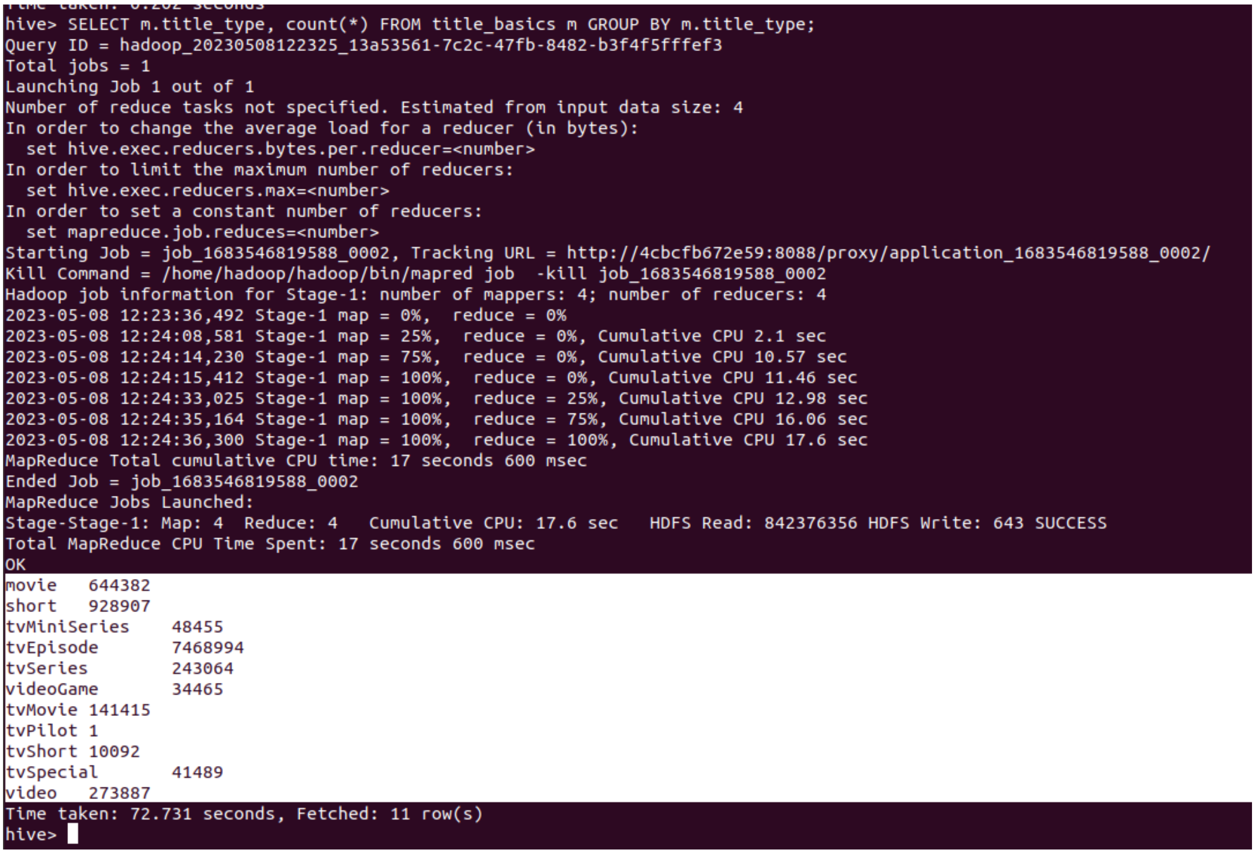
```
hive> SELECT * FROM title_basics b JOIN title_ratings r ON (b.tconst=r.tconst) WHERE original_title = 'The Dark Knight' and title_type='movie';
Query ID = hadoop_20230508121054_8bed6376-1e4a-4b6b-a816-ba76df6e88ef
Total jobs = 2
Stage-5 is selected by condition resolver.
Stage-1 is filtered out by condition resolver.
SLF4J: Found binding in [jar:file:/home/hadoop/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/hadoop/hadoop/share/hadoop/common/lib/slf4j-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]
2023-05-08 12:11:21     Processing rows:        1200000 Hashtable size: 1199999 Memory usage: 484558144        percentage:     0.093
Execution completed successfully
MapredLocal task succeeded
Launching Job 2 out of 2
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1683546819588_0001, Tracking URL = http://4cbcfb672e59:8088/proxy/application_1683546819588_0001/
Kill Command = /home/hadoop/hadoop/bin/mapred job  -kill job_1683546819588_0001
Hadoop job information for Stage-3: number of mappers: 4; number of reducers: 0
2023-05-08 12:11:47,992 Stage-3 map = 0%,  reduce = 0%
2023-05-08 12:12:49,005 Stage-3 map = 0%,  reduce = 0%, Cumulative CPU 15.88 sec
2023-05-08 12:12:52,231 Stage-3 map = 25%,  reduce = 0%, Cumulative CPU 21.08 sec
2023-05-08 12:12:59,830 Stage-3 map = 50%,  reduce = 0%, Cumulative CPU 40.26 sec
2023-05-08 12:13:00,894 Stage-3 map = 63%,  reduce = 0%, Cumulative CPU 41.33 sec
2023-05-08 12:13:01,963 Stage-3 map = 88%,  reduce = 0%, Cumulative CPU 41.97 sec
2023-05-08 12:13:03,001 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 42.37 sec
MapReduce Total cumulative CPU time: 42 seconds 370 msec
Ended Job = job_1683546819588_0001
MapReduce Jobs Launched:
Stage-Stage-3: Map: 4   Cumulative CPU: 42.37 sec   HDFS Read: 842372412 HDFS Write: 463 SUCCESS
Total MapReduce CPU Time Spent: 42 seconds 370 msec
OK
tt0468569       movie   The Dark Knight The Dark Knight 0       2008    NULL    152     Action,Crime,Drama      tt0468569       9.0     2708752
Time taken: 129.359 seconds, Fetched: 1 row(s)
hive> █
```

Create Hive Table name_basics

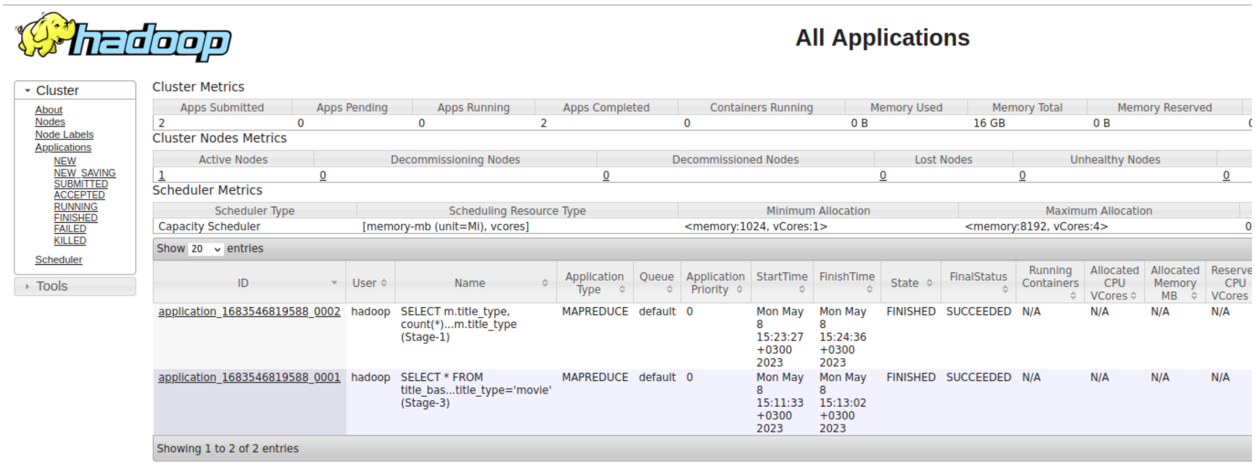


```
hive> CREATE EXTERNAL TABLE IF NOT EXISTS name_basics(
    > nconst STRING,
    > primary_name STRING,
    > birth_year INT,
    > death_year STRING,
    > primary_profession STRING,
    > known_for_titles STRING
    > ) COMMENT 'IMDb Actors' ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE LOCATION '/user/hadoop/imdb/name_basics' TBLPROPERTIES
('skip.header.line.count'='1');
OK
Time taken: 0.202 seconds
hive>
```

Сколько фильмов и сколько сериалов содержится в наборе данных IMDB?



```
Time taken: 0.202 seconds
hive> SELECT m.title_type, count(*) FROM title_basics m GROUP BY m.title_type;
Query ID = hadoop_20230508122325_13a53561-7c2c-47fb-8482-b3f4f5fffef3
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 4
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1683546819588_0002, Tracking URL = http://4cbcfb672e59:8088/proxy/application_1683546819588_0002/
Kill Command = /home/hadoop/hadoop/bin/mapred job  -kill job_1683546819588_0002
Hadoop job information for Stage-1: number of mappers: 4; number of reducers: 4
2023-05-08 12:23:36,492 Stage-1 map = 0%,  reduce = 0%
2023-05-08 12:24:08,581 Stage-1 map = 25%,  reduce = 0%, Cumulative CPU 2.1 sec
2023-05-08 12:24:14,230 Stage-1 map = 75%,  reduce = 0%, Cumulative CPU 10.57 sec
2023-05-08 12:24:15,412 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 11.46 sec
2023-05-08 12:24:33,025 Stage-1 map = 100%,  reduce = 25%, Cumulative CPU 12.98 sec
2023-05-08 12:24:35,164 Stage-1 map = 100%,  reduce = 75%, Cumulative CPU 16.06 sec
2023-05-08 12:24:36,300 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 17.6 sec
MapReduce Total cumulative CPU time: 17 seconds 600 msec
Ended Job = job_1683546819588_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 4  Reduce: 4   Cumulative CPU: 17.6 sec   HDFS Read: 842376356 HDFS Write: 643 SUCCESS
Total MapReduce CPU Time Spent: 17 seconds 600 msec
OK
movie   644382
short   928907
tvMiniSeries    48455
tvEpisode       7468994
tvSeries        243064
videoGame       34465
tvMovie 141415
tvPilot 1
tvShort 10092
tvSpecial       41489
video   273887
Time taken: 72.731 seconds, Fetched: 11 row(s)
hive> █
```



All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved
2	0	0	2	0	0 B	16 GB	0 B

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	
1	0	0	0	0	0

Scheduler Metrics

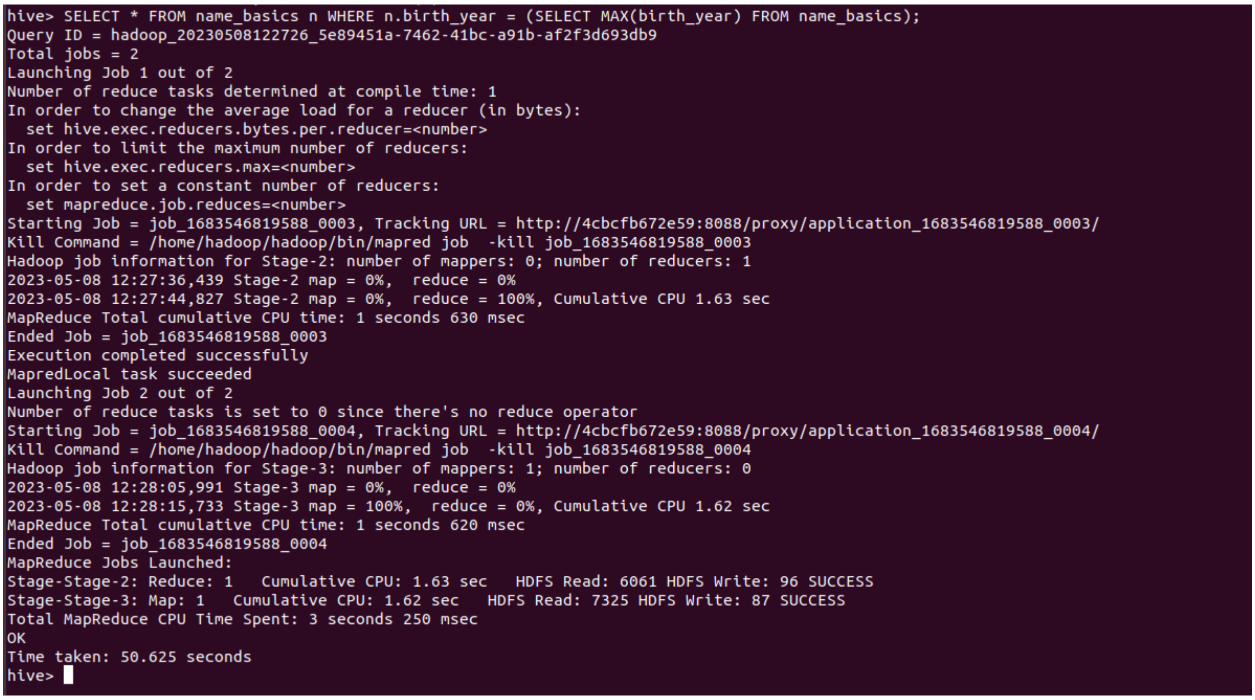
Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	0

Show 20 entries

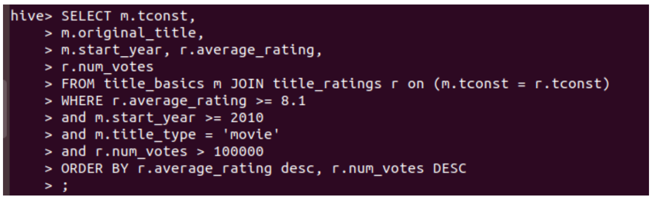
ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Reserve CPU VCores
application_1683546819588_0002	hadoop	SELECT m.title_type, count(*)...m.title_type (Stage-1)	MAPREDUCE	default	0	Mon May 8 15:23:27 +0300 2023	Mon May 8 15:24:36 +0300 2023	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A
application_1683546819588_0001	hadoop	SELECT * FROM title_bas...title_type='movie' (Stage-3)	MAPREDUCE	default	0	Mon May 8 15:11:33 +0300 2023	Mon May 8 15:13:02 +0300 2023	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A

Showing 1 to 2 of 2 entries

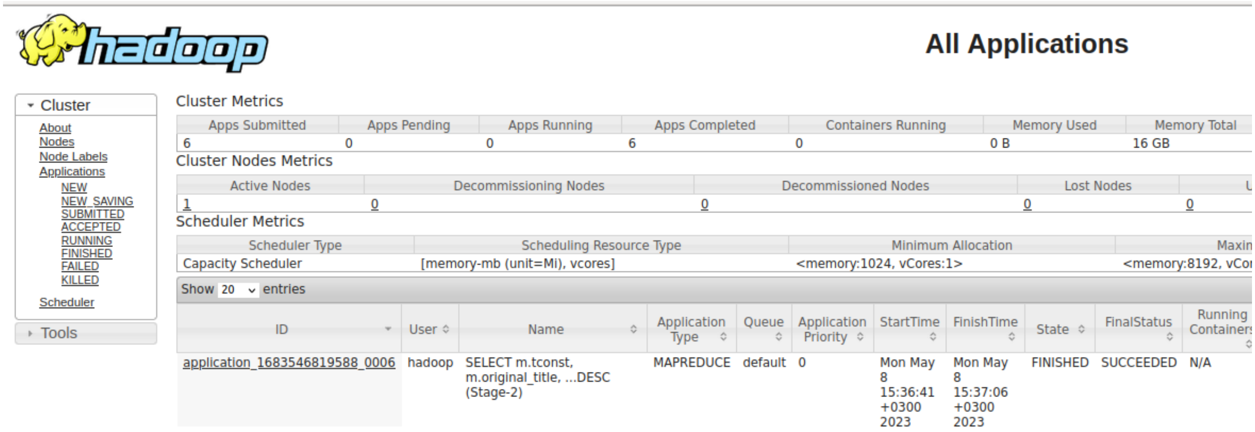
Кто самый молодой актер/сценарист/… в наборе данных?



```
hive> SELECT * FROM name_basics n WHERE n.birth_year = (SELECT MAX(birth_year) FROM name_basics);
Query ID = hadoop_20230508122726_5e89451a-7462-41bc-a91b-af2f3d693db9
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1683546819588_0003, Tracking URL = http://4cbcfb672e59:8088/proxy/application_1683546819588_0003/
Kill Command = /home/hadoop/hadoop/bin/mapred job  -kill job_1683546819588_0003
Hadoop job information for Stage-2: number of mappers: 0; number of reducers: 1
2023-05-08 12:27:36,439 Stage-2 map = 0%,  reduce = 0%
2023-05-08 12:27:44,827 Stage-2 map = 0%,  reduce = 100%, Cumulative CPU 1.63 sec
MapReduce Total cumulative CPU time: 1 seconds 630 msec
Ended Job = job_1683546819588_0003
Execution completed successfully
MapredLocal task succeeded
Launching Job 2 out of 2
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1683546819588_0004, Tracking URL = http://4cbcfb672e59:8088/proxy/application_1683546819588_0004/
Kill Command = /home/hadoop/hadoop/bin/mapred job  -kill job_1683546819588_0004
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2023-05-08 12:28:05,991 Stage-3 map = 0%,  reduce = 0%
2023-05-08 12:28:15,733 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 1.62 sec
MapReduce Total cumulative CPU time: 1 seconds 620 msec
Ended Job = job_1683546819588_0004
MapReduce Jobs Launched:
Stage-Stage-2: Reduce: 1   Cumulative CPU: 1.63 sec   HDFS Read: 6061 HDFS Write: 96 SUCCESS
Stage-Stage-3: Map: 1   Cumulative CPU: 1.62 sec   HDFS Read: 7325 HDFS Write: 87 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 250 msec
OK
Time taken: 50.625 seconds
hive> 
```



```
hive> SELECT m.tconst,
    > m.original_title,
    > m.start_year, r.average_rating,
    > r.num_votes
    > FROM title_basics m JOIN title_ratings r on (m.tconst = r.tconst)
    > WHERE r.average_rating >= 8.1
    > and m.start_year >= 2010
    > and m.title_type = 'movie'
    > and r.num_votes > 100000
    > ORDER BY r.average_rating desc, r.num_votes DESC
    > ;
```



All Applications

- Cluster
  - About
  - Nodes
  - Node Labels
  - Applications
    - NEW
    - NEW_SAVING
    - SUBMITTED
    - ACCEPTED
    - RUNNING
    - FINISHED
    - FAILED
    - KILLED
  - Scheduler
- Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total
6	0	0	6	0	0 B	16 GB

Cluster Nodes Metrics

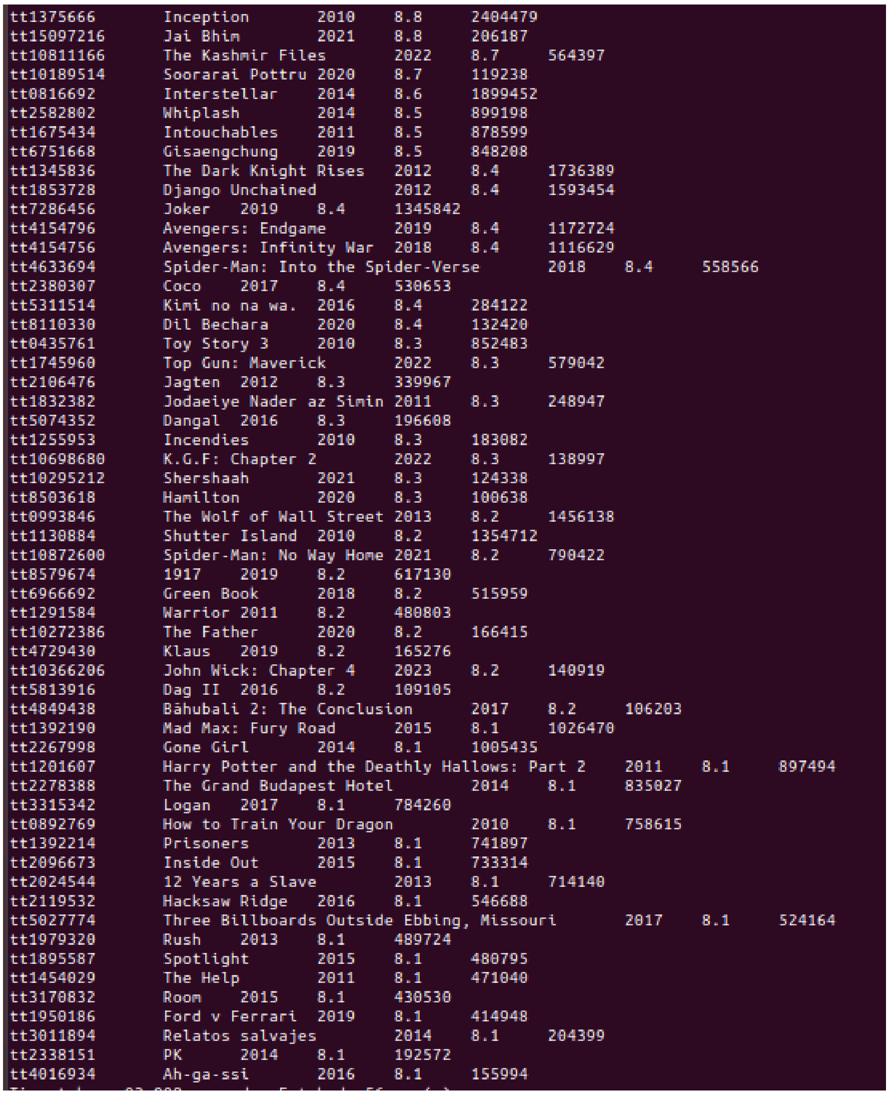
Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	U
1	0	0	0	0

Scheduler Metrics

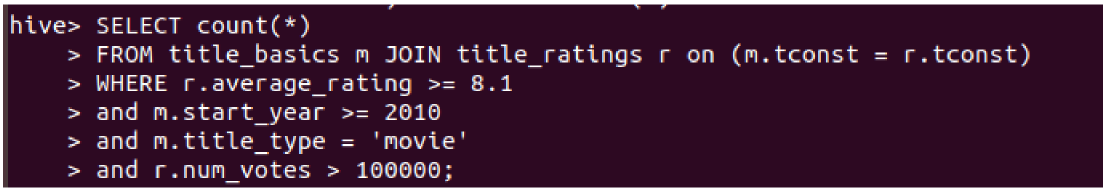
Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maxin
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCor

Show 20 entries

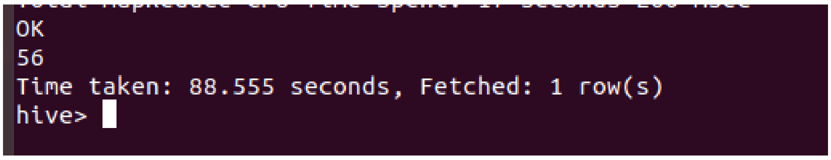
ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Container
application_1683546819588_0006	hadoop	SELECT m.tconst, m.original_title, ...DESC (Stage-2)	MAPREDUCE	default	0	Mon May 8 15:36:41 +0300 2023	Mon May 8 15:37:06 +0300 2023	FINISHED	SUCCEEDED	N/A



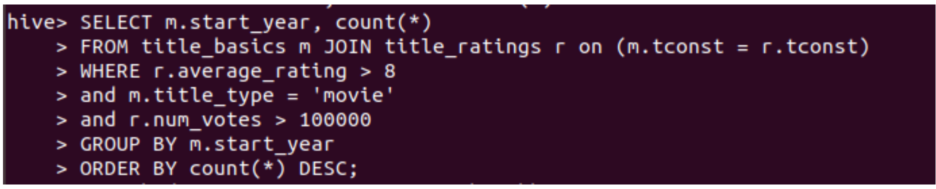
```
tt1375666       Inception       2010    8.8     2404479
tt15097216      Jai Bhim        2021    8.8     206187
tt10811166      The Kashmir Files       2022    8.7     564397
tt10189514      Soorarai Pottru 2020    8.7     119238
tt0816692       Interstellar    2014    8.6     1899452
tt2582802       Whiplash        2014    8.5     899198
tt1675434       Intouchables    2011    8.5     878599
tt6751668       Gisaengchung    2019    8.5     848208
tt1345836       The Dark Knight Rises   2012    8.4     1736389
tt1853728       Django Unchained        2012    8.4     1593454
tt7286456       Joker   2019    8.4     1345842
tt4154796       Avengers: Endgame       2019    8.4     1172724
tt4154756       Avengers: Infinity War  2018    8.4     1116629
tt4633694       Spider-Man: Into the Spider-Verse       2018    8.4     558566
tt2380307       Coco    2017    8.4     530653
tt5311514       Kimi no na wa.  2016    8.4     284122
tt8110330       Dil Bechara     2020    8.4     132420
tt0435761       Toy Story 3     2010    8.3     852483
tt1745960       Top Gun: Maverick       2022    8.3     579042
tt2106476       Jagten  2012    8.3     339967
tt1832382       Jodaeiye Nader az Simin 2011    8.3     248947
tt5074352       Dangal  2016    8.3     196608
tt1255953       Incendies       2010    8.3     183082
tt10698680      K.G.F: Chapter 2        2022    8.3     138997
tt10295212      Shershaah       2021    8.3     124338
tt8503618       Hamilton        2020    8.3     100638
tt0993846       The Wolf of Wall Street 2013    8.2     1456138
tt1130884       Shutter Island  2010    8.2     1354712
tt10872600      Spider-Man: No Way Home 2021    8.2     790422
tt8579674       1917    2019    8.2     617130
tt6966692       Green Book      2018    8.2     515959
tt1291584       Warrior 2011    8.2     480803
tt10272386      The Father      2020    8.2     166415
tt4729430       Klaus   2019    8.2     165276
tt10366206      John Wick: Chapter 4    2023    8.2     140919
tt5813916       Dag II  2016    8.2     109105
tt4849438       Bāhubali 2: The Conclusion      2017    8.2     106203
tt1392190       Mad Max: Fury Road      2015    8.1     1026470
tt2267998       Gone Girl       2014    8.1     1005435
tt1201607       Harry Potter and the Deathly Hallows: Part 2    2011    8.1     897494
tt2278388       The Grand Budapest Hotel        2014    8.1     835027
tt3315342       Logan   2017    8.1     784260
tt0892769       How to Train Your Dragon        2010    8.1     758615
tt1392214       Prisoners       2013    8.1     741897
tt2096673       Inside Out      2015    8.1     733314
tt2024544       12 Years a Slave        2013    8.1     714140
tt2119532       Hacksaw Ridge   2016    8.1     546688
tt5027774       Three Billboards Outside Ebbing, Missouri       2017    8.1     524164
tt1979320       Rush    2013    8.1     489724
tt1895587       Spotlight       2015    8.1     480795
tt1454029       The Help        2011    8.1     471040
tt3170832       Room    2015    8.1     430530
tt1950186       Ford v Ferrari  2019    8.1     414948
tt3011894       Relatos salvajes        2014    8.1     204399
tt2338151       PK      2014    8.1     192572
tt4016934       Ah-ga-ssi       2016    8.1     155994
```



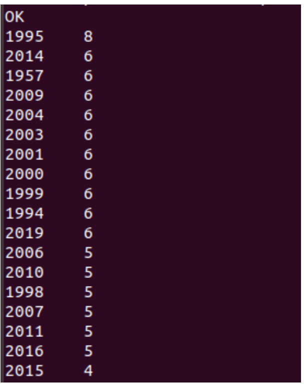
```
hive> SELECT count(*)
    > FROM title_basics m JOIN title_ratings r on (m.tconst = r.tconst)
    > WHERE r.average_rating >= 8.1
    > and m.start_year >= 2010
    > and m.title_type = 'movie'
    > and r.num_votes > 100000;
```



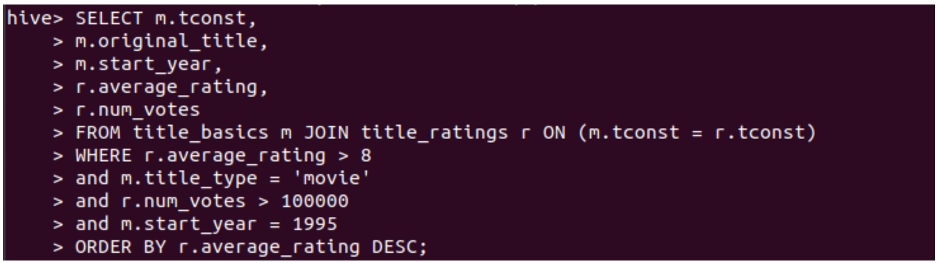
```
OK
56
Time taken: 88.555 seconds, Fetched: 1 row(s)
hive>
```



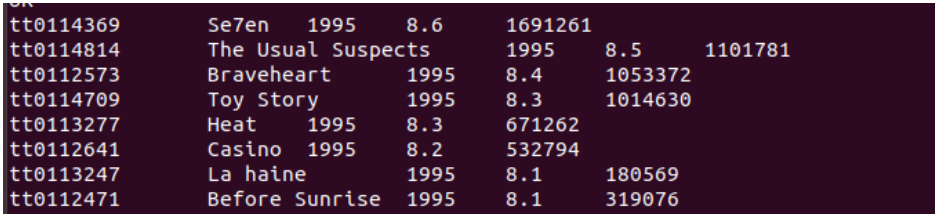
```
hive> SELECT m.start_year, count(*)
    > FROM title_basics m JOIN title_ratings r on (m.tconst = r.tconst)
    > WHERE r.average_rating > 8
    > and m.title_type = 'movie'
    > and r.num_votes > 100000
    > GROUP BY m.start_year
    > ORDER BY count(*) DESC;
```



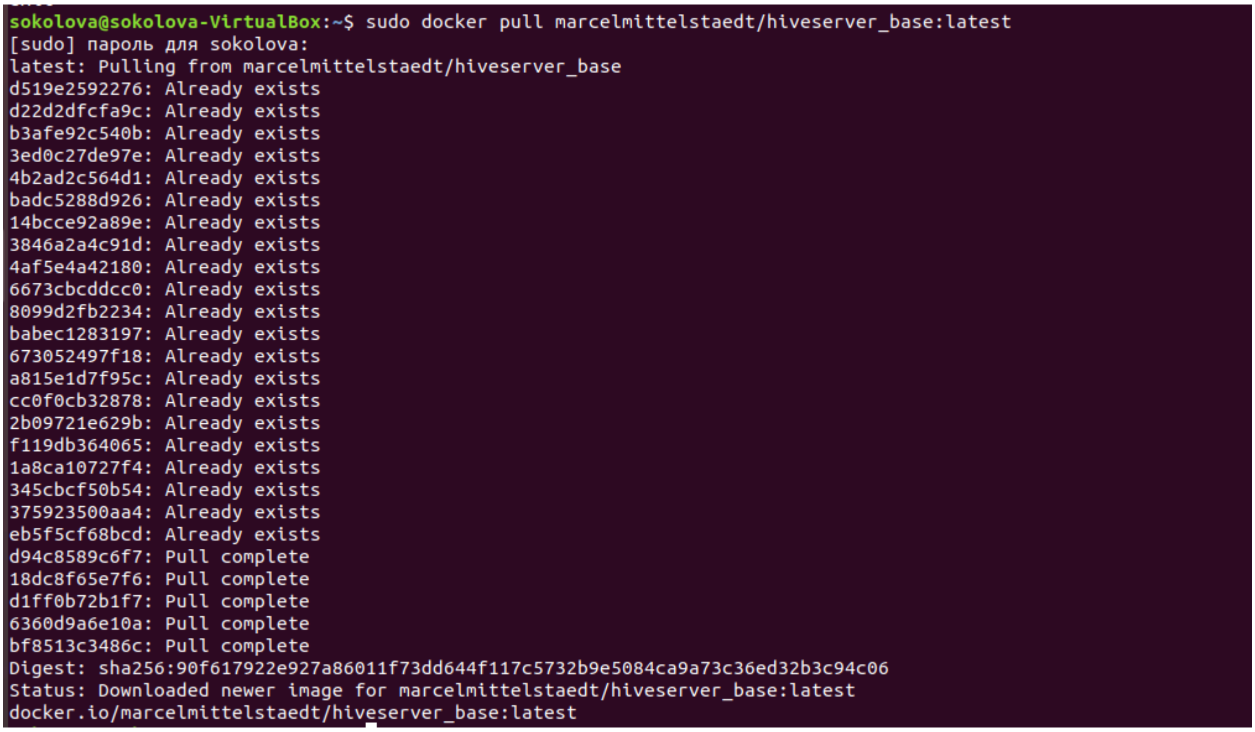
```
OK
1995    8
2014    6
1957    6
2009    6
2004    6
2003    6
2001    6
2000    6
1999    6
1994    6
2019    6
2006    5
2010    5
1998    5
2007    5
2011    5
2016    5
2015    4
```



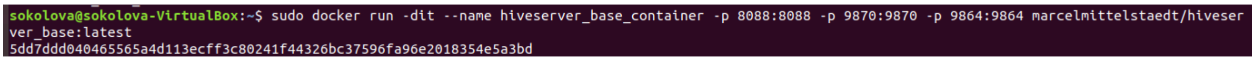
```
hive> SELECT m.tconst,
    > m.original_title,
    > m.start_year,
    > r.average_rating,
    > r.num_votes
    > FROM title_basics m JOIN title_ratings r ON (m.tconst = r.tconst)
    > WHERE r.average_rating > 8
    > and m.title_type = 'movie'
    > and r.num_votes > 100000
    > and m.start_year = 1995
    > ORDER BY r.average_rating DESC;
```



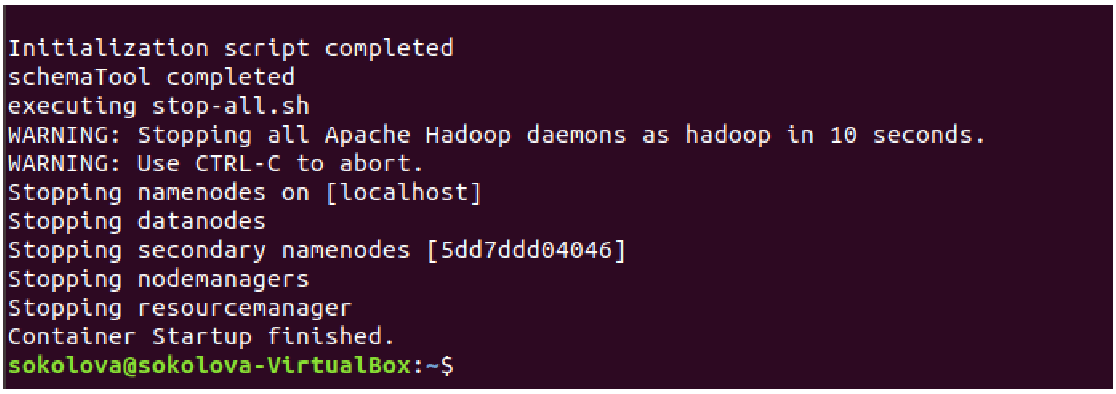
```
tt0114369       Se7en   1995    8.6     1691261
tt0114814       The Usual Suspects      1995    8.5     1101781
tt0112573       Braveheart      1995    8.4     1053372
tt0114709       Toy Story       1995    8.3     1014630
tt0113277       Heat    1995    8.3     671262
tt0112641       Casino  1995    8.2     532794
tt0113247       La haine        1995    8.1     180569
tt0112471       Before Sunrise  1995    8.1     319076
```



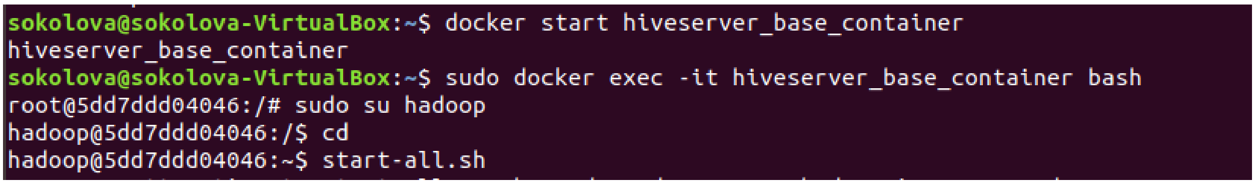
```
sokolova@sokolova-VirtualBox:~$ sudo docker pull marcelmittelstaedt/hiveserver_base:latest
[sudo] пароль для sokolova:
latest: Pulling from marcelmittelstaedt/hiveserver_base
d519e2592276: Already exists
d22d2dfcfa9c: Already exists
b3afe92c540b: Already exists
3ed0c27de97e: Already exists
4b2ad2c564d1: Already exists
badc5288d926: Already exists
14bcce92a89e: Already exists
3846a2a4c91d: Already exists
4af5e4a42180: Already exists
6673cbcddcc0: Already exists
8099d2fb2234: Already exists
babec1283197: Already exists
673052497f18: Already exists
a815e1d7f95c: Already exists
cc0f0cb32878: Already exists
2b09721e629b: Already exists
f119db364065: Already exists
1a8ca10727f4: Already exists
345cbcf50b54: Already exists
375923500aa4: Already exists
eb5f5cf68bcd: Already exists
d94c8589c6f7: Pull complete
18dc8f65e7f6: Pull complete
d1ff0b72b1f7: Pull complete
6360d9a6e10a: Pull complete
bf8513c3486c: Pull complete
Digest: sha256:90f617922e927a86011f73dd644f117c5732b9e5084ca9a73c36ed32b3c94c06
Status: Downloaded newer image for marcelmittelstaedt/hiveserver_base:latest
docker.io/marcelmittelstaedt/hiveserver_base:latest
```



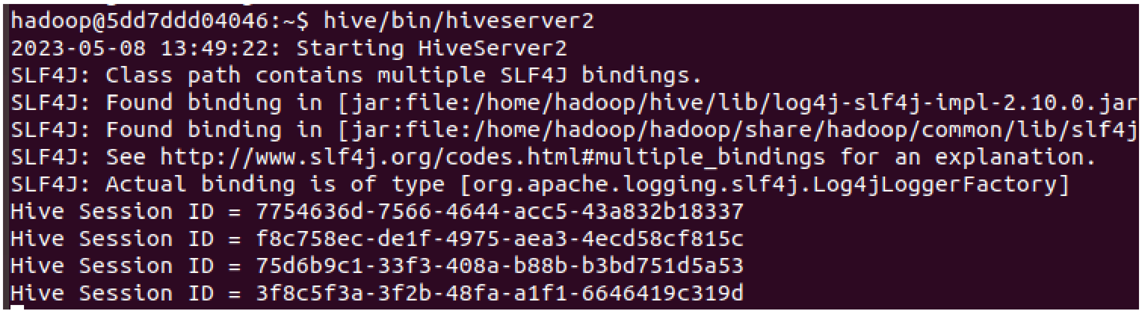
```
sokolova@sokolova-VirtualBox:~$ sudo docker run -dit --name hiveserver_base_container -p 8088:8088 -p 9870:9870 -p 9864:9864 marcelmittelstaedt/hiveser
ver_base:latest
5dd7ddd040465565a4d113ecff3c80241f44326bc37596fa96e2018354e5a3bd
```



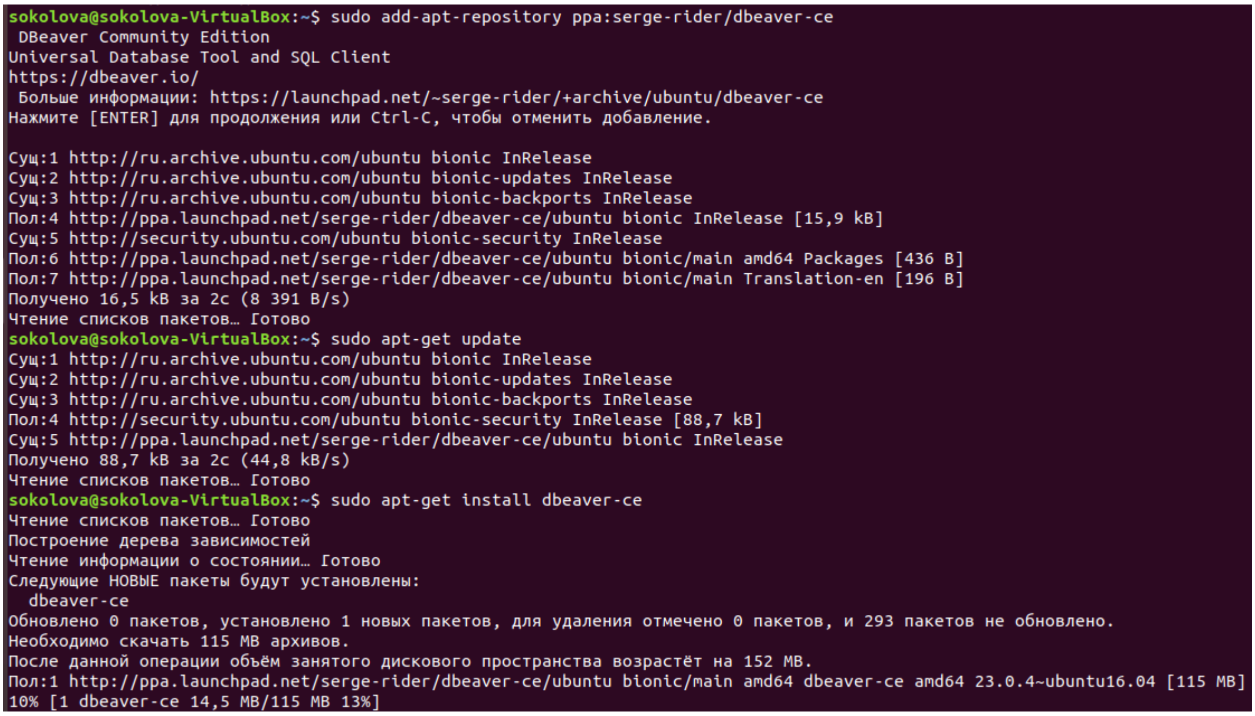
```
Initialization script completed
schemaTool completed
executing stop-all.sh
WARNING: Stopping all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [5dd7ddd04046]
Stopping nodemanagers
Stopping resourcemanager
Container Startup finished.
sokolova@sokolova-VirtualBox:~$
```



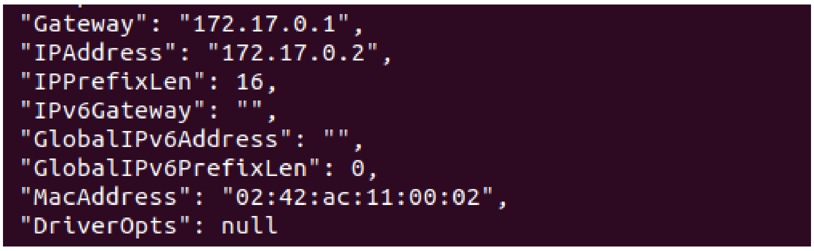
```
sokolova@sokolova-VirtualBox:~$ docker start hiveserver_base_container
hiveserver_base_container
sokolova@sokolova-VirtualBox:~$ sudo docker exec -it hiveserver_base_container bash
root@5dd7ddd04046:/# sudo su hadoop
hadoop@5dd7ddd04046:/$ cd
hadoop@5dd7ddd04046:~$ start-all.sh
```



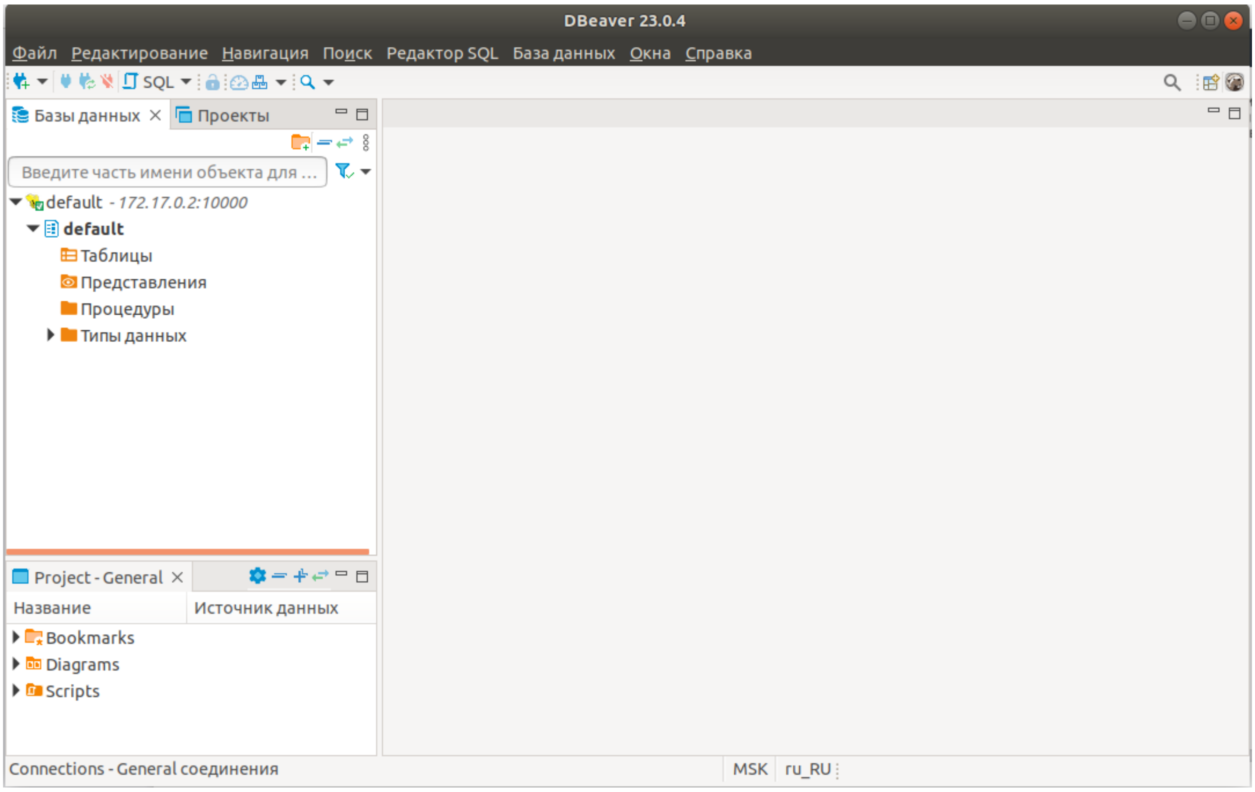
```
hadoop@5dd7ddd04046:~$ hive/bin/hiveserver2
2023-05-08 13:49:22: Starting HiveServer2
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoop/hive/lib/log4j-slf4j-impl-2.10.0.jar
SLF4J: Found binding in [jar:file:/home/hadoop/hadoop/share/hadoop/common/lib/slf4j
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Hive Session ID = 7754636d-7566-4644-acc5-43a832b18337
Hive Session ID = f8c758ec-de1f-4975-aea3-4ecd58cf815c
Hive Session ID = 75d6b9c1-33f3-408a-b88b-b3bd751d5a53
Hive Session ID = 3f8c5f3a-3f2b-48fa-a1f1-6646419c319d
```



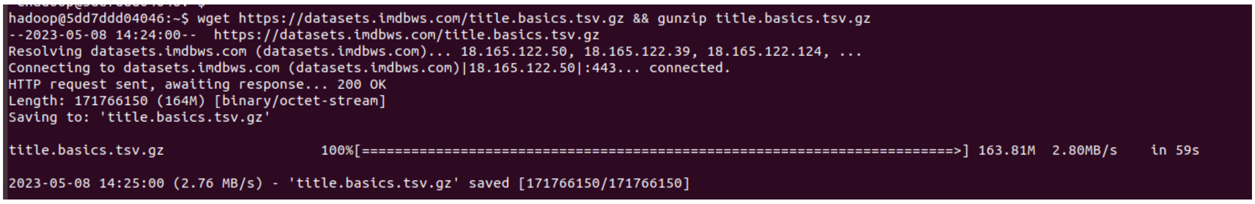
```
sokolova@sokolova-VirtualBox:~$ sudo add-apt-repository ppa:serge-rider/dbeaver-ce
 DBeaver Community Edition
Universal Database Tool and SQL Client
https://dbeaver.io/
 Больше информации: https://launchpad.net/~serge-rider/+archive/ubuntu/dbeaver-ce
Нажмите [ENTER] для продолжения или Ctrl-C, чтобы отменить добавление.

Сущ:1 http://ru.archive.ubuntu.com/ubuntu bionic InRelease
Сущ:2 http://ru.archive.ubuntu.com/ubuntu bionic-updates InRelease
Сущ:3 http://ru.archive.ubuntu.com/ubuntu bionic-backports InRelease
Пол:4 http://ppa.launchpad.net/serge-rider/dbeaver-ce/ubuntu bionic InRelease [15,9 kB]
Сущ:5 http://security.ubuntu.com/ubuntu bionic-security InRelease
Пол:6 http://ppa.launchpad.net/serge-rider/dbeaver-ce/ubuntu bionic/main amd64 Packages [436 B]
Пол:7 http://ppa.launchpad.net/serge-rider/dbeaver-ce/ubuntu bionic/main Translation-en [196 B]
Получено 16,5 kB за 2с (8 391 B/s)
Чтение списков пакетов… Готово
sokolova@sokolova-VirtualBox:~$ sudo apt-get update
Сущ:1 http://ru.archive.ubuntu.com/ubuntu bionic InRelease
Сущ:2 http://ru.archive.ubuntu.com/ubuntu bionic-updates InRelease
Сущ:3 http://ru.archive.ubuntu.com/ubuntu bionic-backports InRelease
Пол:4 http://security.ubuntu.com/ubuntu bionic-security InRelease [88,7 kB]
Сущ:5 http://ppa.launchpad.net/serge-rider/dbeaver-ce/ubuntu bionic InRelease
Получено 88,7 kB за 2с (44,8 kB/s)
Чтение списков пакетов… Готово
sokolova@sokolova-VirtualBox:~$ sudo apt-get install dbeaver-ce
Чтение списков пакетов… Готово
Построение дерева зависимостей
Чтение информации о состоянии… Готово
Следующие НОВЫЕ пакеты будут установлены:
  dbeaver-ce
Обновлено 0 пакетов, установлено 1 новых пакетов, для удаления отмечено 0 пакетов, и 293 пакетов не обновлено.
Необходимо скачать 115 MB архивов.
После данной операции объём занятого дискового пространства возрастёт на 152 MB.
Пол:1 http://ppa.launchpad.net/serge-rider/dbeaver-ce/ubuntu bionic/main amd64 dbeaver-ce 23.0.4~ubuntu16.04 [115 MB]
10% [1 dbeaver-ce 14,5 MB/115 MB 13%]
```



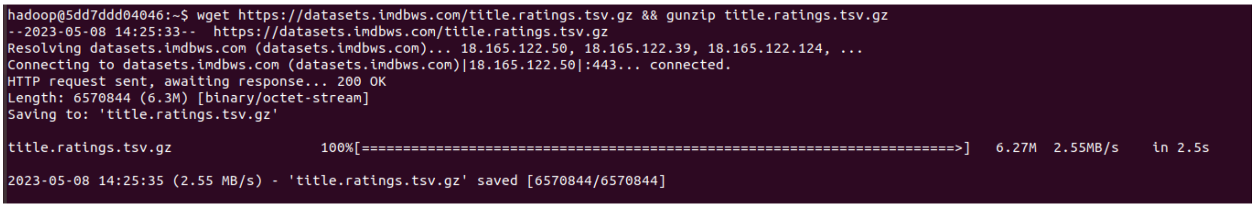
```
 "Gateway": "172.17.0.1",
 "IPAddress": "172.17.0.2",
 "IPPrefixLen": 16,
 "IPv6Gateway": "",
 "GlobalIPv6Address": "",
 "GlobalIPv6PrefixLen": 0,
 "MacAddress": "02:42:ac:11:00:02",
 "DriverOpts": null
```



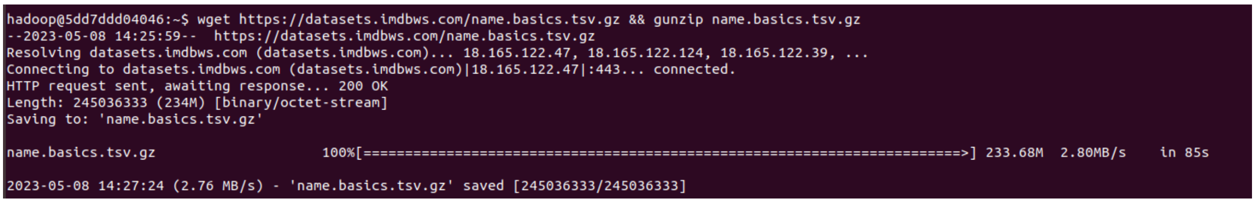
DBeaver 23.0.4
Файл Редактирование Навигация Поиск Редактор SQL База данных Окна Справка
SQL
Базы данных
Проекты
Введите часть имени объекта для ...
default - 172.17.0.2:10000
default
Таблицы
Представления
Процедуры
Типы данных
Project - General
Название
Источник данных
Bookmarks
Diagrams
Scripts
Connections - General соединения
MSK
ru_RU



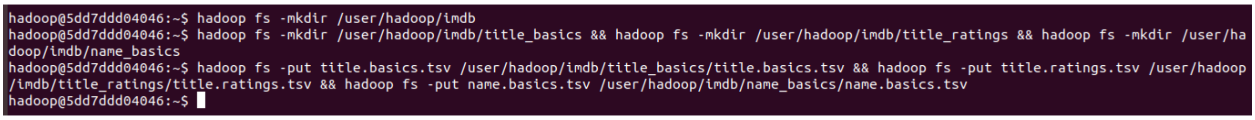
```
hadoop@5dd7ddd04046:~$ wget https://datasets.imdbws.com/title.basics.tsv.gz && gunzip title.basics.tsv.gz
--2023-05-08 14:24:00--  https://datasets.imdbws.com/title.basics.tsv.gz
Resolving datasets.imdbws.com (datasets.imdbws.com)... 18.165.122.50, 18.165.122.39, 18.165.122.124, ...
Connecting to datasets.imdbws.com (datasets.imdbws.com)|18.165.122.50|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 171766150 (164M) [binary/octet-stream]
Saving to: 'title.basics.tsv.gz'

title.basics.tsv.gz             100%[=========================================================================>] 163.81M  2.80MB/s    in 59s

2023-05-08 14:25:00 (2.76 MB/s) - 'title.basics.tsv.gz' saved [171766150/171766150]
```



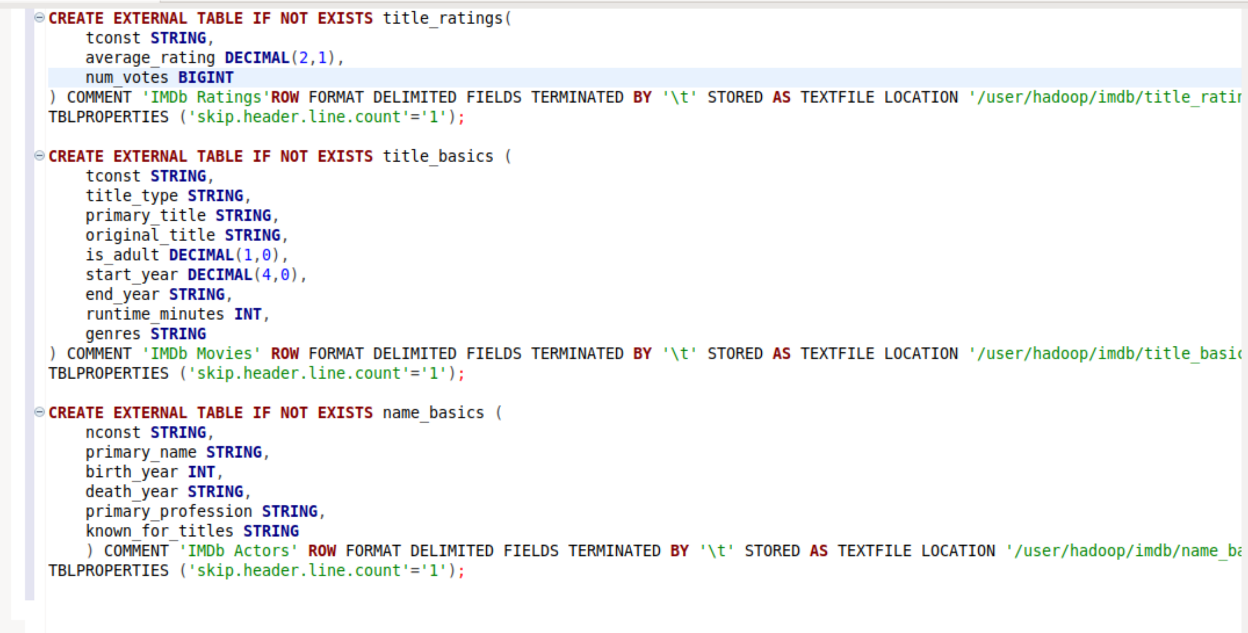
```
hadoop@5dd7ddd04046:~$ wget https://datasets.imdbws.com/title.ratings.tsv.gz && gunzip title.ratings.tsv.gz
--2023-05-08 14:25:33--  https://datasets.imdbws.com/title.ratings.tsv.gz
Resolving datasets.imdbws.com (datasets.imdbws.com)... 18.165.122.50, 18.165.122.39, 18.165.122.124, ...
Connecting to datasets.imdbws.com (datasets.imdbws.com)|18.165.122.50|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 6570844 (6.3M) [binary/octet-stream]
Saving to: 'title.ratings.tsv.gz'

title.ratings.tsv.gz            100%[=========================================================================>]   6.27M  2.55MB/s    in 2.5s

2023-05-08 14:25:35 (2.55 MB/s) - 'title.ratings.tsv.gz' saved [6570844/6570844]
```



```
hadoop@5dd7ddd04046:~$ wget https://datasets.imdbws.com/name.basics.tsv.gz && gunzip name.basics.tsv.gz
--2023-05-08 14:25:59--  https://datasets.imdbws.com/name.basics.tsv.gz
Resolving datasets.imdbws.com (datasets.imdbws.com)... 18.165.122.47, 18.165.122.124, 18.165.122.39, ...
Connecting to datasets.imdbws.com (datasets.imdbws.com)|18.165.122.47|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 245036333 (234M) [binary/octet-stream]
Saving to: 'name.basics.tsv.gz'

name.basics.tsv.gz              100%[=========================================================================>] 233.68M  2.80MB/s    in 85s

2023-05-08 14:27:24 (2.76 MB/s) - 'name.basics.tsv.gz' saved [245036333/245036333]
```



```
hadoop@5dd7ddd04046:~$ hadoop fs -mkdir /user/hadoop/imdb
hadoop@5dd7ddd04046:~$ hadoop fs -mkdir /user/hadoop/imdb/title_basics && hadoop fs -mkdir /user/hadoop/imdb/title_ratings && hadoop fs -mkdir /user/ha
doop/imdb/name_basics
hadoop@5dd7ddd04046:~$ hadoop fs -put title.basics.tsv /user/hadoop/imdb/title_basics/title.basics.tsv && hadoop fs -put title.ratings.tsv /user/hadoop
/imdb/title_ratings/title.ratings.tsv && hadoop fs -put name.basics.tsv /user/hadoop/imdb/name_basics/name.basics.tsv
hadoop@5dd7ddd04046:~$ 
```



```sql
CREATE EXTERNAL TABLE IF NOT EXISTS title_ratings(
    tconst STRING,
    average_rating DECIMAL(2,1),
    num_votes BIGINT
) COMMENT 'IMDb Ratings'ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE LOCATION '/user/hadoop/imdb/title_ratir
TBLPROPERTIES ('skip.header.line.count'='1');

CREATE EXTERNAL TABLE IF NOT EXISTS title_basics (
    tconst STRING,
    title_type STRING,
    primary_title STRING,
    original_title STRING,
    is_adult DECIMAL(1,0),
    start_year DECIMAL(4,0),
    end_year STRING,
    runtime_minutes INT,
    genres STRING
) COMMENT 'IMDb Movies' ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE LOCATION '/user/hadoop/imdb/title_basic
TBLPROPERTIES ('skip.header.line.count'='1');

CREATE EXTERNAL TABLE IF NOT EXISTS name_basics (
    nconst STRING,
    primary_name STRING,
    birth_year INT,
    death_year STRING,
    primary_profession STRING,
    known_for_titles STRING
    ) COMMENT 'IMDb Actors' ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE LOCATION '/user/hadoop/imdb/name_ba
TBLPROPERTIES ('skip.header.line.count'='1');
```

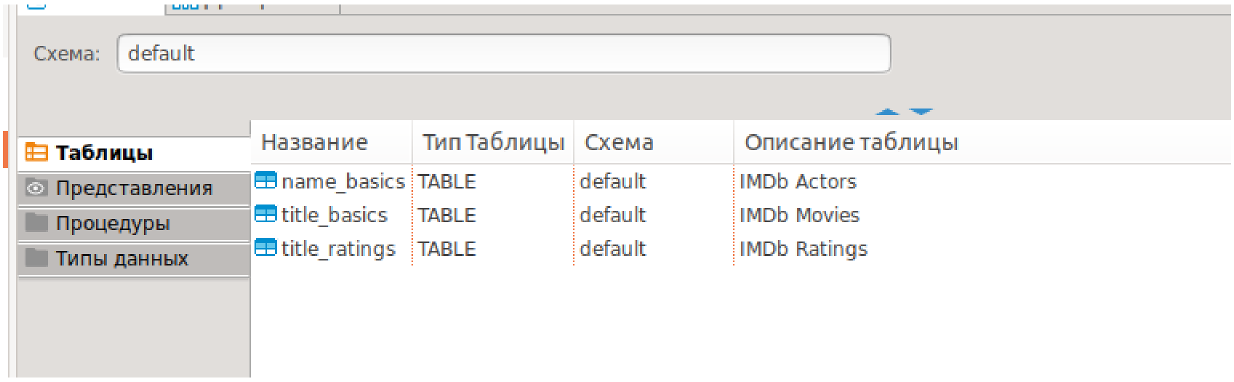
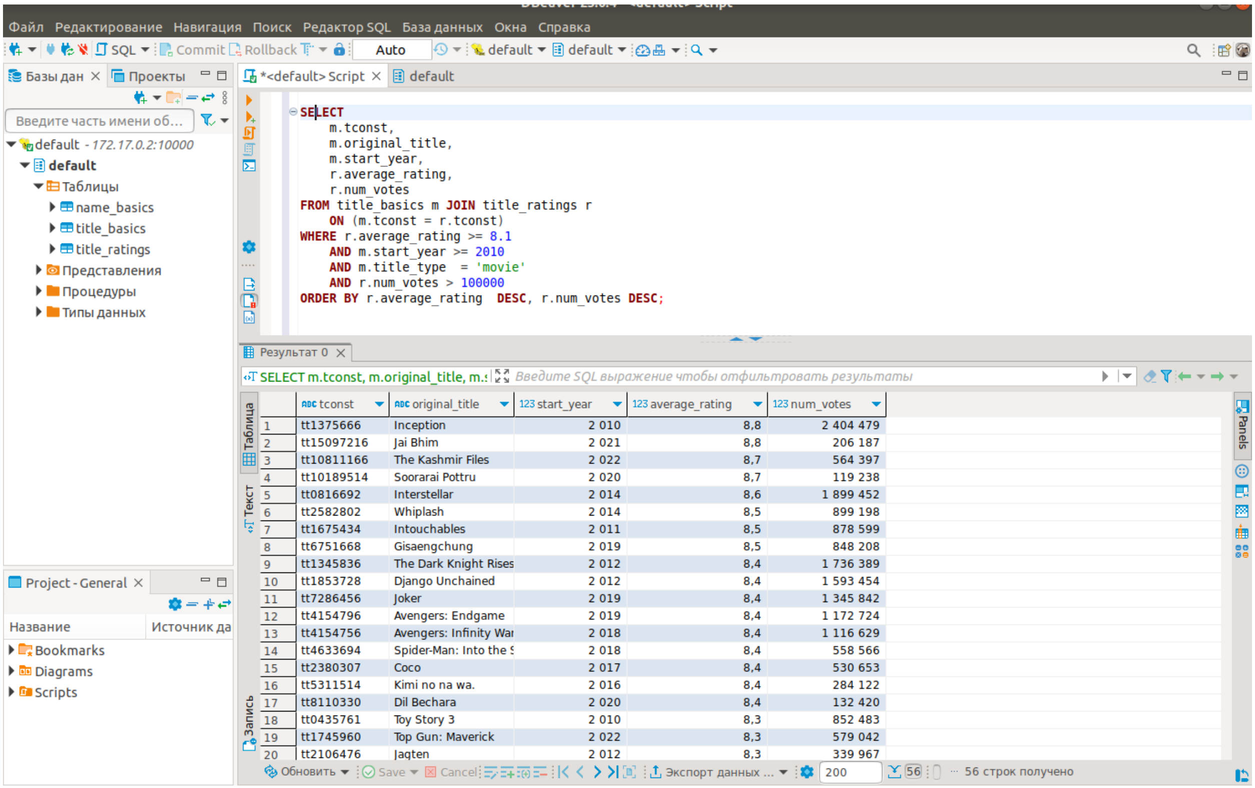


Схема: default

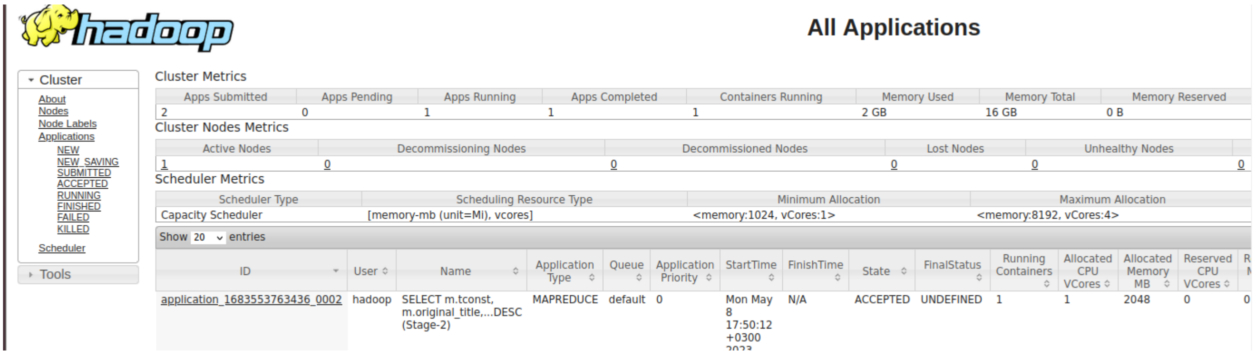
Название	Тип Таблицы	Схема	Описание таблицы
name_basics	TABLE	default	IMDb Actors
title_basics	TABLE	default	IMDb Movies
title_ratings	TABLE	default	IMDb Ratings



```sql
SELECT
    m.tconst,
    m.original_title,
    m.start_year,
    r.average_rating,
    r.num_votes
FROM title_basics m JOIN title_ratings r
    ON (m.tconst = r.tconst)
WHERE r.average_rating >= 8.1
    AND m.start_year >= 2010
    AND m.title_type  = 'movie'
    AND r.num_votes > 100000
ORDER BY r.average_rating  DESC, r.num_votes DESC;
```

SELECT m.tconst, m.original_title, m.s

	tconst	original_title	start_year	average_rating	num_votes
1	tt1375666	Inception	2 010	8,8	2 404 479
2	tt15097216	Jai Bhim	2 021	8,8	206 187
3	tt10811166	The Kashmir Files	2 022	8,7	564 397
4	tt10189514	Soorarai Pottru	2 020	8,7	119 238
5	tt0816692	Interstellar	2 014	8,6	1 899 452
6	tt2582802	Whiplash	2 014	8,5	899 198
7	tt1675434	Intouchables	2 011	8,5	878 599
8	tt6751668	Gisaengchung	2 019	8,5	848 208
9	tt1345836	The Dark Knight Rises	2 012	8,4	1 736 389
10	tt1853728	Django Unchained	2 012	8,4	1 593 454
11	tt7286456	Joker	2 019	8,4	1 345 842
12	tt4154796	Avengers: Endgame	2 019	8,4	1 172 724
13	tt4154756	Avengers: Infinity War	2 018	8,4	1 116 629
14	tt4633694	Spider-Man: Into the S	2 018	8,4	558 566
15	tt2380307	Coco	2 017	8,4	530 653
16	tt5311514	Kimi no na wa.	2 016	8,4	284 122
17	tt8110330	Dil Bechara	2 020	8,4	132 420
18	tt0435761	Toy Story 3	2 010	8,3	852 483
19	tt1745960	Top Gun: Maverick	2 022	8,3	579 042
20	tt2106476	Jagten	2 012	8,3	339 967



All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved
2	0	1	1	1	2 GB	16 GB	0 B

Cluster Nodes Metrics

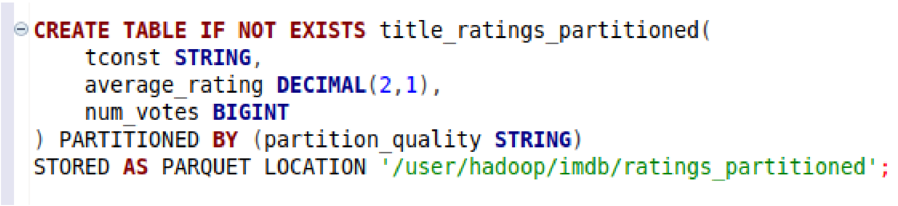
Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	
1	0	0	0	0	0

Scheduler Metrics

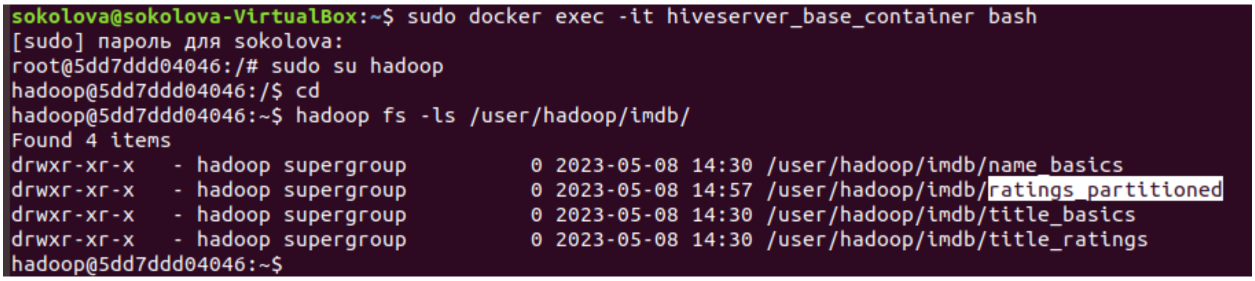
Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>

Show 20 entries

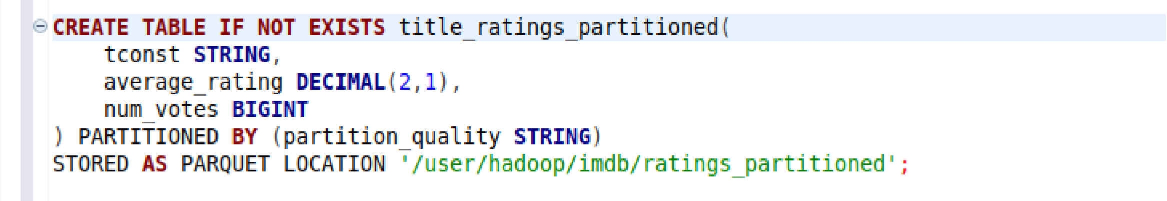
ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Reserved CPU VCores	R M
application_1683553763436_0002	hadoop	SELECT m.tconst, m.original_title,...DESC (Stage-2)	MAPREDUCE	default	0	Mon May 8 17:50:12 +0300 2023	N/A	ACCEPTED	UNDEFINED	1	1	2048	0	0



```sql
CREATE TABLE IF NOT EXISTS title_ratings_partitioned(
    tconst STRING,
    average_rating DECIMAL(2,1),
    num_votes BIGINT
) PARTITIONED BY (partition_quality STRING)
STORED AS PARQUET LOCATION '/user/hadoop/imdb/ratings_partitioned';
```



```
sokolova@sokolova-VirtualBox:~$ sudo docker exec -it hiveserver_base_container bash
[sudo] пароль для sokolova:
root@5dd7ddd04046:/# sudo su hadoop
hadoop@5dd7ddd04046:/$ cd
hadoop@5dd7ddd04046:~$ hadoop fs -ls /user/hadoop/imdb/
Found 4 items
drwxr-xr-x   - hadoop supergroup          0 2023-05-08 14:30 /user/hadoop/imdb/name_basics
drwxr-xr-x   - hadoop supergroup          0 2023-05-08 14:57 /user/hadoop/imdb/ratings_partitioned
drwxr-xr-x   - hadoop supergroup          0 2023-05-08 14:30 /user/hadoop/imdb/title_basics
drwxr-xr-x   - hadoop supergroup          0 2023-05-08 14:30 /user/hadoop/imdb/title_ratings
hadoop@5dd7ddd04046:~$
```



```sql
CREATE TABLE IF NOT EXISTS title_ratings_partitioned(
    tconst STRING,
    average_rating DECIMAL(2,1),
    num_votes BIGINT
) PARTITIONED BY (partition_quality STRING)
STORED AS PARQUET LOCATION '/user/hadoop/imdb/ratings_partitioned';
```

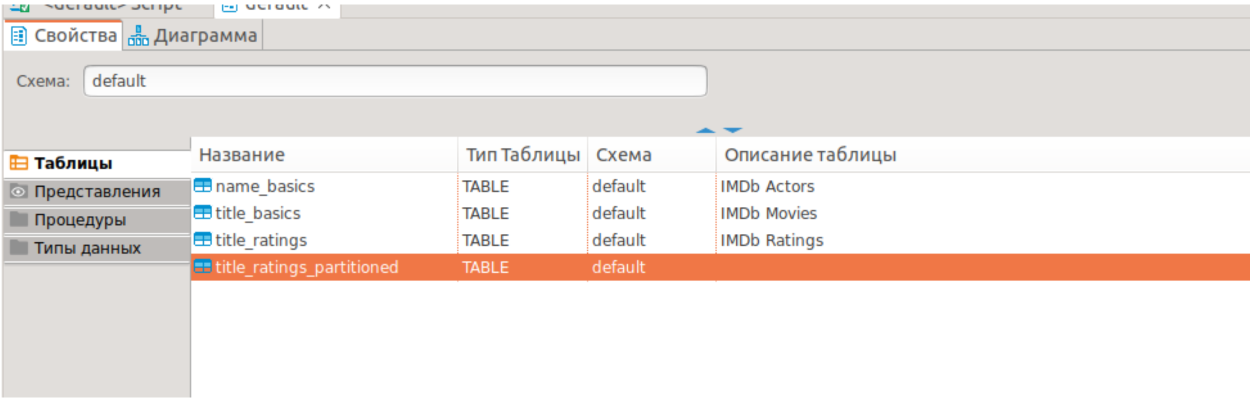
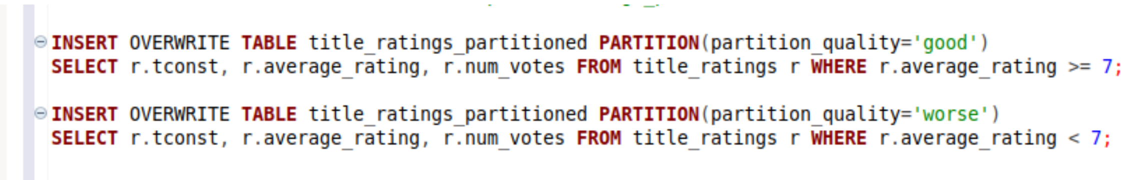
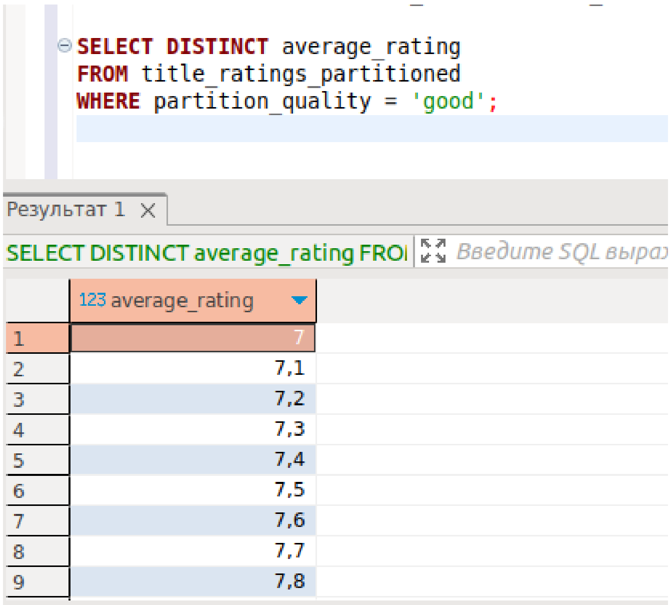


Схема: default

Название	Тип Таблицы	Схема	Описание таблицы
name_basics	TABLE	default	IMDb Actors
title_basics	TABLE	default	IMDb Movies
title_ratings	TABLE	default	IMDb Ratings
title_ratings_partitioned	TABLE	default	

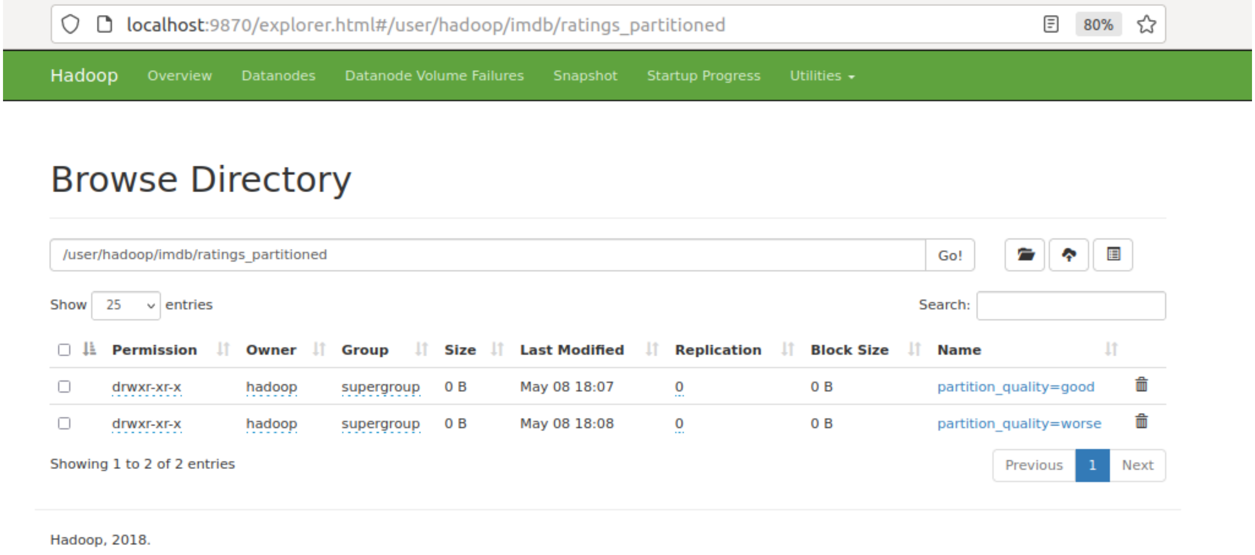


```sql
INSERT OVERWRITE TABLE title_ratings_partitioned PARTITION(partition_quality='good')
SELECT r.tconst, r.average_rating, r.num_votes FROM title_ratings r WHERE r.average_rating >= 7;

INSERT OVERWRITE TABLE title_ratings_partitioned PARTITION(partition_quality='worse')
SELECT r.tconst, r.average_rating, r.num_votes FROM title_ratings r WHERE r.average_rating < 7;
```



```sql
SELECT DISTINCT average_rating
FROM title_ratings_partitioned
WHERE partition_quality = 'good';
```

	123 average_rating
1	7
2	7,1
3	7,2
4	7,3
5	7,4
6	7,5
7	7,6
8	7,7
9	7,8



Browse Directory

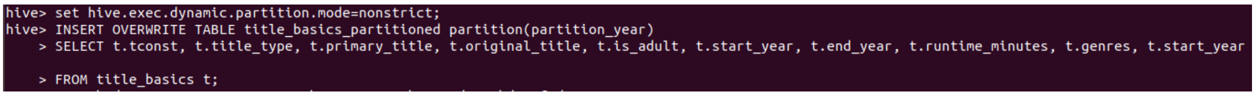
/user/hadoop/imdb/ratings_partitioned Go!

Show 25 entries Search:

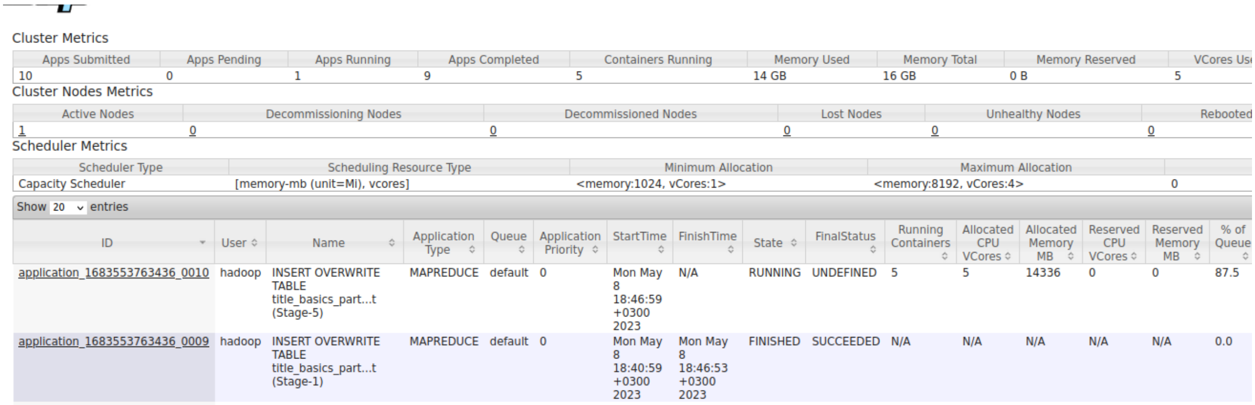
☐	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
☐	drwxr-xr-x	hadoop	supergroup	0 B	May 08 18:07	0	0 B	partition_quality=good	🗑
☐	drwxr-xr-x	hadoop	supergroup	0 B	May 08 18:08	0	0 B	partition_quality=worse	🗑

Showing 1 to 2 of 2 entries

Hadoop, 2018.



```
hive> set hive.exec.dynamic.partition.mode=nonstrict;
hive> INSERT OVERWRITE TABLE title_basics_partitioned partition(partition_year)
    > SELECT t.tconst, t.title_type, t.primary_title, t.original_title, t.is_adult, t.start_year, t.end_year, t.runtime_minutes, t.genres, t.start_year

    > FROM title_basics t;
```



Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Use
10	0	1	9	5	14 GB	16 GB	0 B	5

Cluster Nodes Metrics

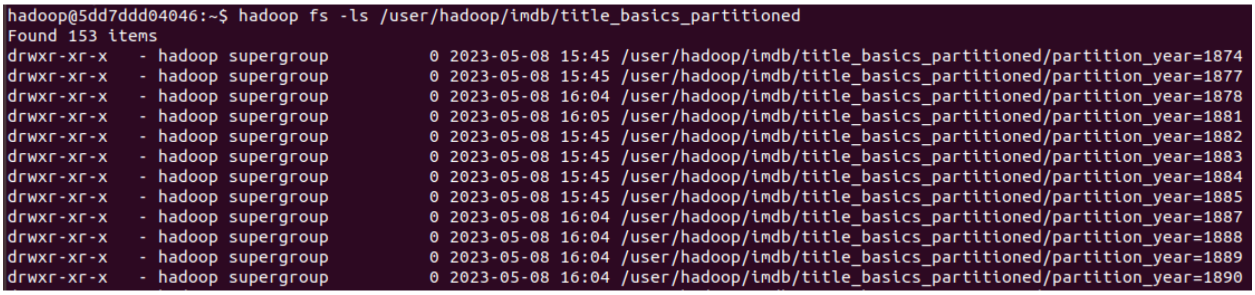
Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted
1	0	0	0	0	0

Scheduler Metrics

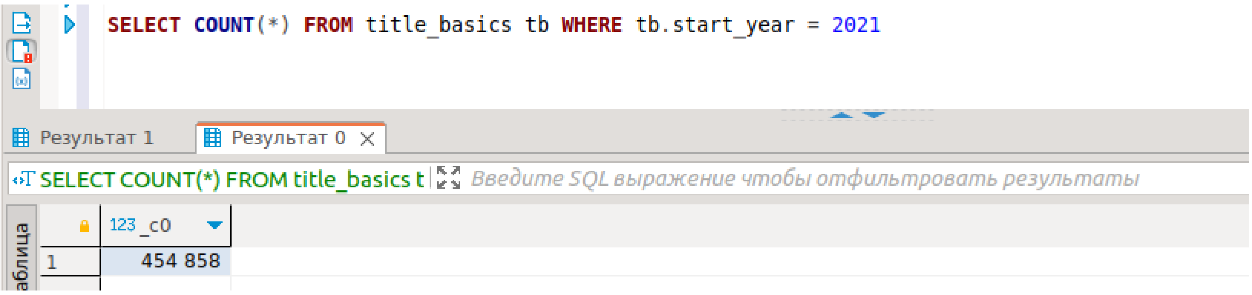
Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	0

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Reserved CPU VCores	Reserved Memory MB	% of Queue
application_1683553763436_0010	hadoop	INSERT OVERWRITE TABLE title_basics_part...t (Stage-5)	MAPREDUCE	default	0	Mon May 8 18:46:59 +0300 2023	N/A	RUNNING	UNDEFINED	5	5	14336	0	0	87.5
application_1683553763436_0009	hadoop	INSERT OVERWRITE TABLE title_basics_part...t (Stage-1)	MAPREDUCE	default	0	Mon May 8 18:40:59 +0300 2023	Mon May 8 18:46:53 +0300 2023	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A	0.0



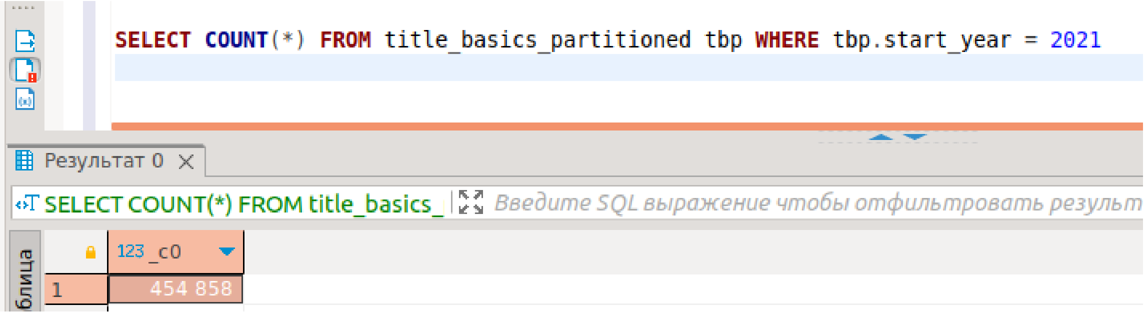
```
hadoop@5dd7ddd04046:~$ hadoop fs -ls /user/hadoop/imdb/title_basics_partitioned
Found 153 items
drwxr-xr-x   - hadoop supergroup          0 2023-05-08 15:45 /user/hadoop/imdb/title_basics_partitioned/partition_year=1874
drwxr-xr-x   - hadoop supergroup          0 2023-05-08 15:45 /user/hadoop/imdb/title_basics_partitioned/partition_year=1877
drwxr-xr-x   - hadoop supergroup          0 2023-05-08 16:04 /user/hadoop/imdb/title_basics_partitioned/partition_year=1878
drwxr-xr-x   - hadoop supergroup          0 2023-05-08 16:05 /user/hadoop/imdb/title_basics_partitioned/partition_year=1881
drwxr-xr-x   - hadoop supergroup          0 2023-05-08 15:45 /user/hadoop/imdb/title_basics_partitioned/partition_year=1882
drwxr-xr-x   - hadoop supergroup          0 2023-05-08 15:45 /user/hadoop/imdb/title_basics_partitioned/partition_year=1883
drwxr-xr-x   - hadoop supergroup          0 2023-05-08 15:45 /user/hadoop/imdb/title_basics_partitioned/partition_year=1884
drwxr-xr-x   - hadoop supergroup          0 2023-05-08 15:45 /user/hadoop/imdb/title_basics_partitioned/partition_year=1885
drwxr-xr-x   - hadoop supergroup          0 2023-05-08 16:04 /user/hadoop/imdb/title_basics_partitioned/partition_year=1887
drwxr-xr-x   - hadoop supergroup          0 2023-05-08 16:04 /user/hadoop/imdb/title_basics_partitioned/partition_year=1888
drwxr-xr-x   - hadoop supergroup          0 2023-05-08 16:04 /user/hadoop/imdb/title_basics_partitioned/partition_year=1889
drwxr-xr-x   - hadoop supergroup          0 2023-05-08 16:04 /user/hadoop/imdb/title_basics_partitioned/partition_year=1890
```



```sql
SELECT COUNT(*) FROM title_basics tb WHERE tb.start_year = 2021
```

Результат 1
Результат 0
SELECT COUNT(*) FROM title_basics t
Введите SQL выражение чтобы отфильтровать результаты
аблица

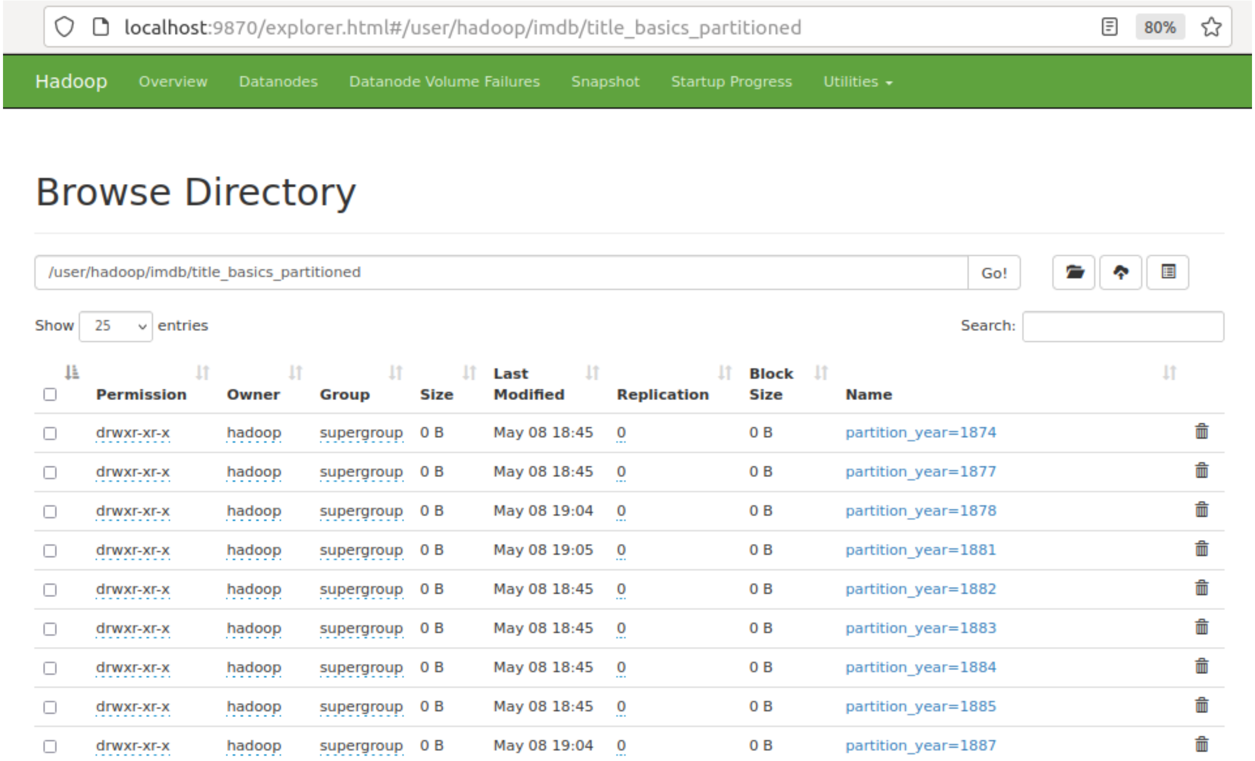
	_c0
1	454 858



```sql
SELECT COUNT(*) FROM title_basics_partitioned tbp WHERE tbp.start_year = 2021
```

Результат 0
SELECT COUNT(*) FROM title_basics_
Введите SQL выражение чтобы отфильтровать результ
блица

	_c0
1	454 858



Browse Directory

/user/hadoop/imdb/title_basics_partitioned Go!

Show 25 entries Search:

□	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
□	drwxr-xr-x	hadoop	supergroup	0 B	May 08 18:45	0	0 B	partition_year=1874
□	drwxr-xr-x	hadoop	supergroup	0 B	May 08 18:45	0	0 B	partition_year=1877
□	drwxr-xr-x	hadoop	supergroup	0 B	May 08 19:04	0	0 B	partition_year=1878
□	drwxr-xr-x	hadoop	supergroup	0 B	May 08 19:05	0	0 B	partition_year=1881
□	drwxr-xr-x	hadoop	supergroup	0 B	May 08 18:45	0	0 B	partition_year=1882
□	drwxr-xr-x	hadoop	supergroup	0 B	May 08 18:45	0	0 B	partition_year=1883
□	drwxr-xr-x	hadoop	supergroup	0 B	May 08 18:45	0	0 B	partition_year=1884
□	drwxr-xr-x	hadoop	supergroup	0 B	May 08 18:45	0	0 B	partition_year=1885
□	drwxr-xr-x	hadoop	supergroup	0 B	May 08 19:04	0	0 B	partition_year=1887