

Государственное автономное образовательное
учреждение высшего образования «Московский городской педагогический
университет»

Институт цифрового образования
Департамент информатики, управления и технологий

Отчет по лабораторной работе № 3.1

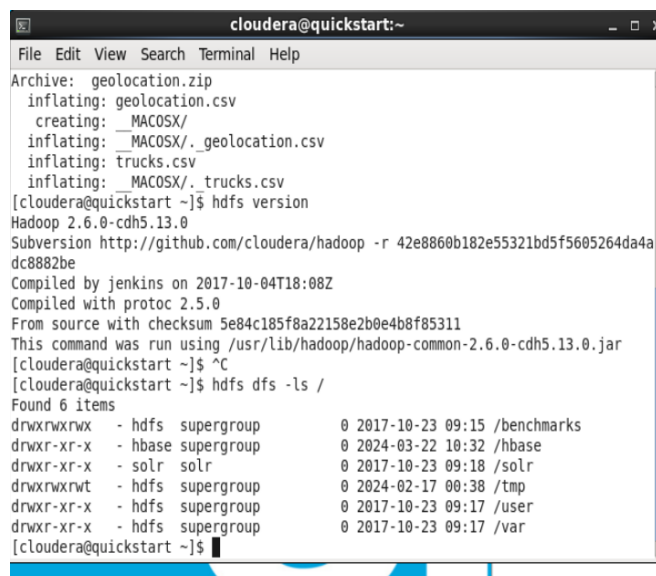
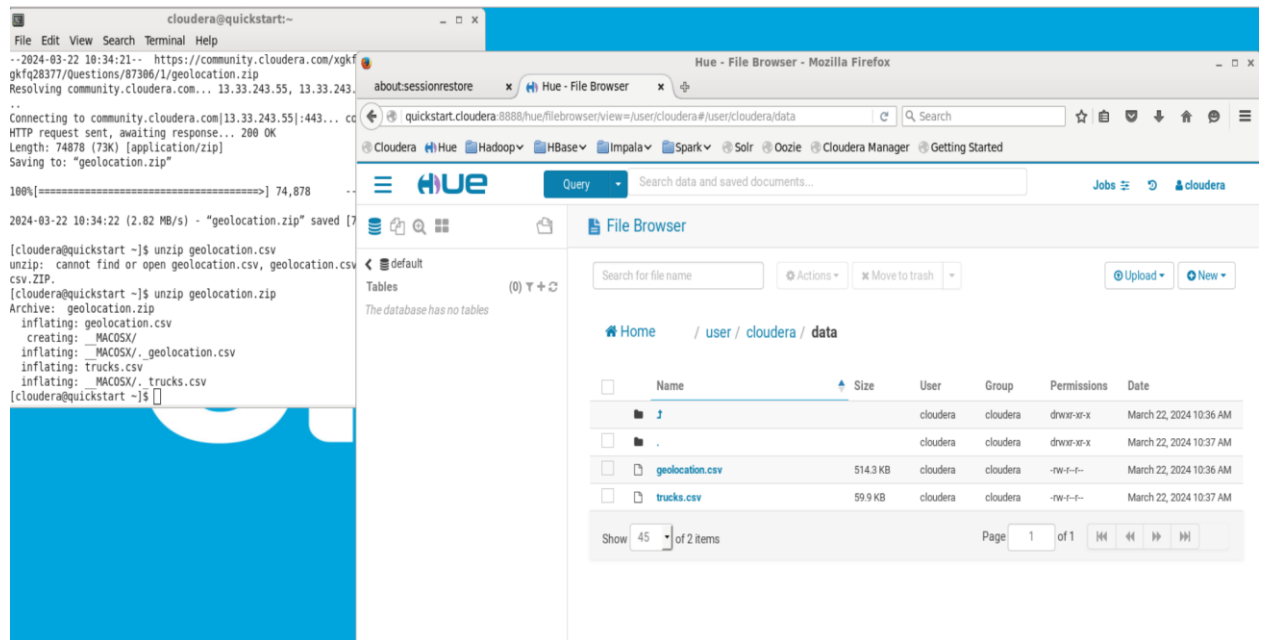
Выполнил: студент БД-231
Кряжева К.С.

2024 г

3.1.1. Развернуть виртуальное окружение.

3.1.2. Вывести с помощью команды `help` описание основных команды shell-клиента.

3.1.3. Просмотреть корневую директорию `HDFS`.



3.1.4. Создать в HDFS в директории `/user/mgrpu` поддиректорию `ваше_фio`.

3.1.5. Создать в локальной файловой системе случайный текстовый файл размером 10 Mb с именем, образованным вашими инициалами base64 `/dev/urandom | head -c 10000000 > file.txt`.

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
inflating: __MACOSX/.geolocation.csv  
inflating: trucks.csv  
inflating: __MACOSX/.trucks.csv  
[cloudera@quickstart ~]$ hdfs version  
Hadoop 2.6.0-cdh5.13.0  
Subversion http://github.com/cloudera/hadoop -r 42e8860b182e55321bd5f5605264da4a  
dc8882be  
Compiled by jenkins on 2017-10-04T18:08Z  
Compiled with protoc 2.5.0  
From source with checksum 5e84c185f8a22158e2b0e4b8f85311  
This command was run using /usr/lib/hadoop/hadoop-common-2.6.0-cdh5.13.0.jar  
[cloudera@quickstart ~]$ ^C  
[cloudera@quickstart ~]$ hdfs dfs -ls /  
Found 6 items  
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks  
drwxr-xr-x - hbase supergroup 0 2024-03-22 10:32 /hbase  
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr  
drwxrwxrwt - hdfs supergroup 0 2024-02-17 00:38 /tmp  
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user  
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var  
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/mgpu  
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/mgpu/KryazhevaKS  
[cloudera@quickstart ~]$ base64 /dev/urandom | head -c 1000000 > Kryazheva.txt  
[cloudera@quickstart ~]$
```

3.1.6. Заархивировать созданный текстовый файл gzip -c file.txt > file.gz .

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
rwxr-xr-x - hbase supergroup 0 2024-03-22 10:32 /hbase  
rwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr  
rwxrwxrwt - hdfs supergroup 0 2024-02-17 00:38 /tmp  
rwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user  
rwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var  
cloudera@quickstart ~]$ hdfs dfs -mkdir /user/mgpu  
cloudera@quickstart ~]$ hdfs dfs -mkdir /user/mgpu/KryazhevaKS  
cloudera@quickstart ~]$ base64 /dev/urandom | head -c 1000000 > Kryazheva.txt  
cloudera@quickstart ~]$ ls  
cloudera-manager enterprise-deployment.json lib Templates  
m_api.py express-deployment.json __MACOSX trucks.csv  
esktop geolocation.csv Music Videos  
ocuments geolocation.zip parcels workspace  
ownloads kerberos Pictures  
clipse Kryazheva.txt Public  
cloudera@quickstart ~]$ gzip -c Kryazheva.txt > Kryazheva.gz  
cloudera@quickstart ~]$ ls  
cloudera-manager enterprise-deployment.json Kryazheva.txt Public  
m_api.py express-deployment.json lib Templates  
esktop geolocation.csv __MACOSX trucks.csv  
ocuments geolocation.zip Music Videos  
ownloads kerberos parcels workspace  
clipse Kryazheva.gz Pictures  
cloudera@quickstart ~]$
```

3.1.7. Скопировать текстовый файл и архив в директорию /user/mgpu/fio HDFS виртуальной машины.

3.1.8. Просмотреть файл и архив с помощью утилит cat, text в комбинации с каналами и утилитами head, tail -- привести не менее 3 вариантов команд и просмотра файла.

3.1.9. Создать копию файла file.txt вида date_file.txt, где в начале имени файла-копии указана текущая дата. Вывести листинг.

3.1.10. Вывести статистику по директории /user/mgpu/fio виртуальной машины.

```
cloudera@quickstart ~]$ hdfs dfs -cat /user/mgpu/KryazhevaKS/Kryazheva.txt | he
id -n 3
J5A0GBCYrLIDZTNB0+vYX1+n0CLkSgCrvFE0Bvuf+8tcwgUb00Xt00ZufSmrGufXSdnob4XED6j
wMSeJFwXwHdgCEgp/oTGnI1V6BD1I9JtVTx4kzgpBYGVXWZegXyaqULKutBii10PAJ8kw0Vp8
:maBG6z2jeT0TGTk1+GWLJ7nQrzK8LWJv99FWfoVHa4hwL6PLUicrBQbBe7iMTQj7fpm1YONYZCp
:at: Unable to write to output stream.
cloudera@quickstart ~]$ hdfs dfs -cat /user/mgpu/KryazhevaKS/Kryazheva.txt | tail -n 3

31871+WpX4uMu0Bx2P9CEMvyzDjgAf6405X1iFVt0CWrah7LHe6IaG5fErNLWnx9yggj8Q+t9j
XDF0Mef155h0eomQUfIHLMicRwHiQmnqDeROACEELGdPUkuqArU3BwselemXYEqEZUtrlqDVt0
ItkJGwCaBn[cloudera@quickstart ~]$
cloudera@quickstart ~]$ hdfs dfs -cat /user/mgpu/KryazhevaKS/Kryazheva.gz | gunzip | head -n 3

J5A0GBCYrLIDZTNB0+vYX1+n0CLkSgCrvFE0Bvuf+8tcwgUb00Xt00ZufSmrGufXSdnob4XED6j
wMSeJFwXwHdgCEgp/oTGnI1V6BD1I9JtVTx4kzgpBYGVXWZegXyaqULKutBii10PAJ8kw0Vp8
:maBG6z2jeT0TGTk1+GWLJ7nQrzK8LWJv99FWfoVHa4hwL6PLUicrBQbBe7iMTQj7fpm1YONYZCp
:at: Unable to write to output stream.
cloudera@quickstart ~]$
cloudera@quickstart ~]$
```

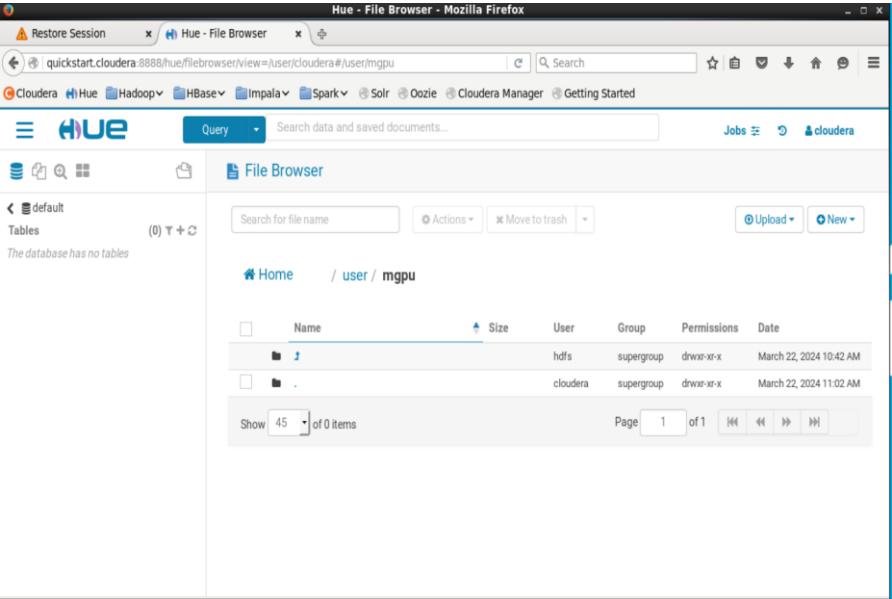
The screenshot shows a terminal window on the left and a Hue File Browser interface on the right. The terminal displays the execution of hdfs commands and the resulting file listing in Hue. The Hue interface shows a file browser view for the path /user/mgpu/KryazhevaKS, displaying a list of files including 22march.txt, Kryazheva.gz, and Kryazheva.txt.

Name	Size	User	Group	Permissions	Date
22march.txt	9.5 MB	cloudera	supergroup	-rw-r--r--	March 22, 2024 10:57 AM
Kryazheva.gz	7.2 MB	cloudera	supergroup	-rw-r--r--	March 22, 2024 10:50 AM
Kryazheva.txt	9.5 MB	cloudera	supergroup	-rw-r--r--	March 22, 2024 10:50 AM

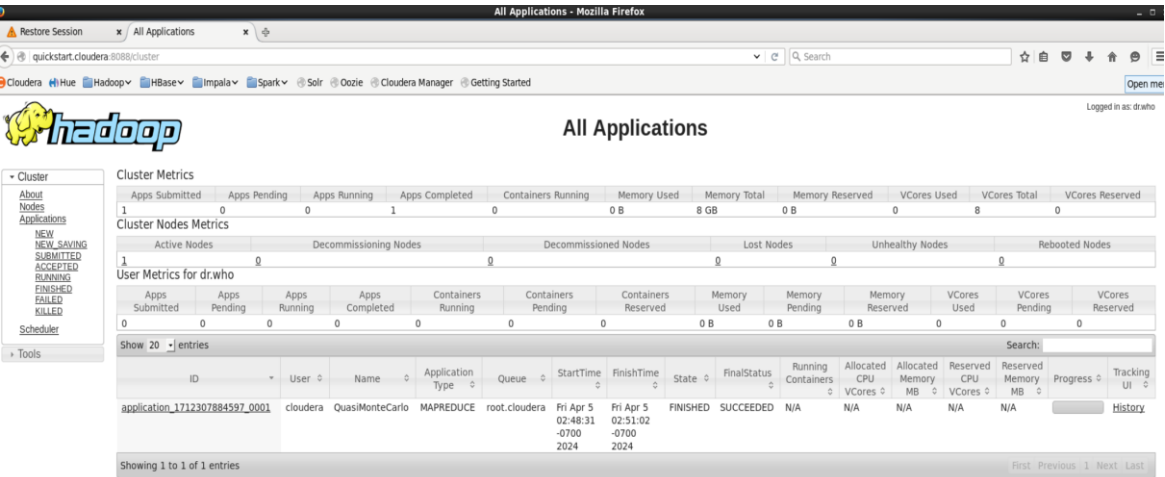
The screenshot shows the Hue File Browser interface for the path /user/mgpu/KryazhevaKS. The interface displays a table of files with columns for Name, Size, User, Group, Permissions, and Date.

Name	Size	User	Group	Permissions	Date
22march.txt	9.5 MB	cloudera	supergroup	-rw-r--r--	March 22, 2024 10:57 AM
Kryazheva.gz	7.2 MB	cloudera	supergroup	-rw-r--r--	March 22, 2024 10:50 AM
Kryazheva.txt	9.5 MB	cloudera	supergroup	-rw-r--r--	March 22, 2024 10:50 AM

3.1.11. Удалить поддиректорию **/foo** со всем содержимым.



3.1.12. Подсчитать количество слов в файле внутри HDFS с помощью методологии **Map Reduce** (размер файла не менее 128 Мб).



default

ables

(0)

he database has no tables

Search for file name

Actions

Move to trash

Uplo

Home

/ user / cloudera / geoloc

	Name	Size	User	Group	Permissions	Date
			cloudera	cloudera	drwxr-xr-x	April 05, 2024 02:57 AM
	.		cloudera	cloudera	drwxr-xr-x	March 22, 2024 10:37 AM
	geolocation.csv	514.3 KB	cloudera	cloudera	-rw-r--r--	March 22, 2024 10:36 AM
	trucks.csv	59.9 KB	cloudera	cloudera	-rw-r--r--	March 22, 2024 10:37 AM

Show 45 of 2 itemsPage 1 of 1

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
lrwxrwxrwt - hdfs supergroup 0 2024-02-17 00:38 /tmp  
lrwxr-xr-x - hdfs supergroup 0 2024-03-22 10:42 /user  
lrwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var  
[cloudera@quickstart ~]$ yarn jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar wordcount geoloc/geolocation.csv  
/ output  
24/04/05 03:01:10 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032  
24/04/05 03:01:11 INFO input.FileInputFormat: Total input paths to process : 1  
24/04/05 03:01:11 INFO mapreduce.JobSubmitter: number of splits:1  
24/04/05 03:01:11 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1712307884597_0003  
24/04/05 03:01:12 INFO impl.YarnClientImpl: Submitted application application_1712307884597_0003  
24/04/05 03:01:12 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1712307884597_0003/  
24/04/05 03:01:12 INFO mapreduce.Job: Running job: job_1712307884597_0003  
24/04/05 03:01:19 INFO mapreduce.Job: Job job_1712307884597_0003 running in uber mode : false  
24/04/05 03:01:19 INFO mapreduce.Job: map 0% reduce 0%  
24/04/05 03:01:25 INFO mapreduce.Job: map 100% reduce 0%  
24/04/05 03:02:21 INFO mapreduce.Job: map 100% reduce 100%  
24/04/05 03:02:24 INFO mapreduce.Job: Job job_1712307884597_0003 completed successfully  
24/04/05 03:02:24 INFO mapreduce.Job: Counters: 49  
File System Counters  
FILE: Number of bytes read=331421  
FILE: Number of bytes written=950105  
FILE: Number of read operations=0  
FILE: Number of large read operations=0
```

All Applications - Mozilla Firefox

quickstart.cloudera:8088/cluster

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total
2	0	0	2	0	0 B	8 GB	0 B	0	8

Cluster Nodes Metrics

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

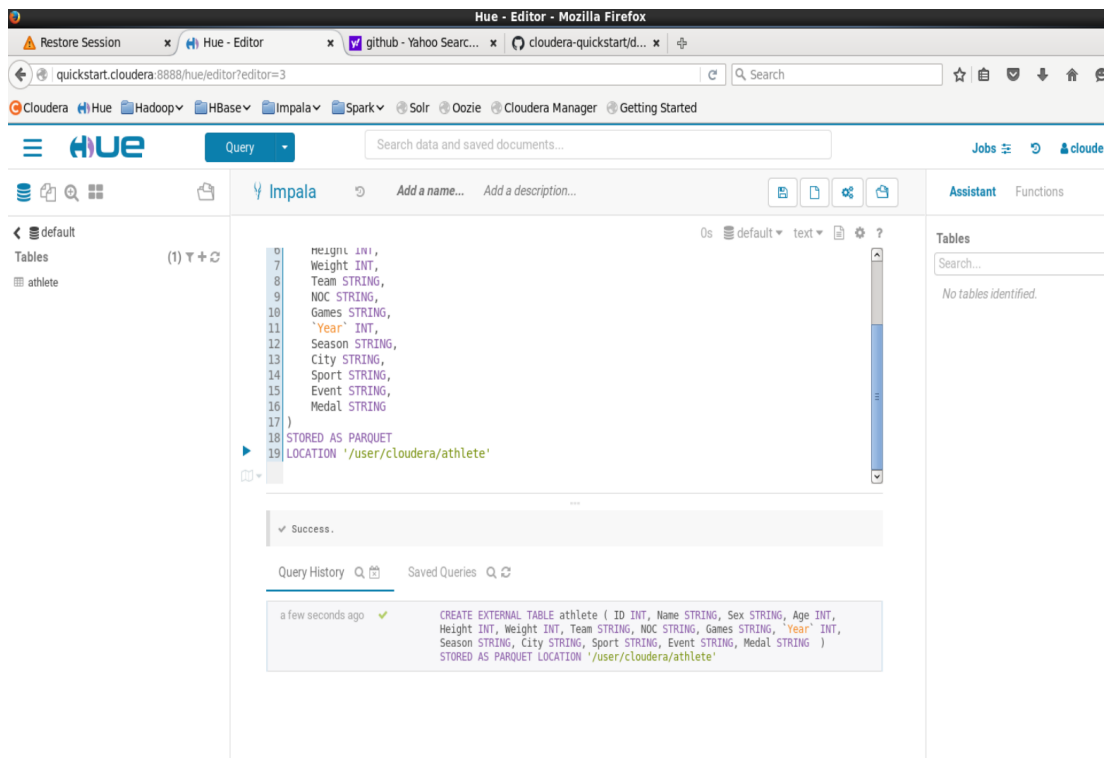
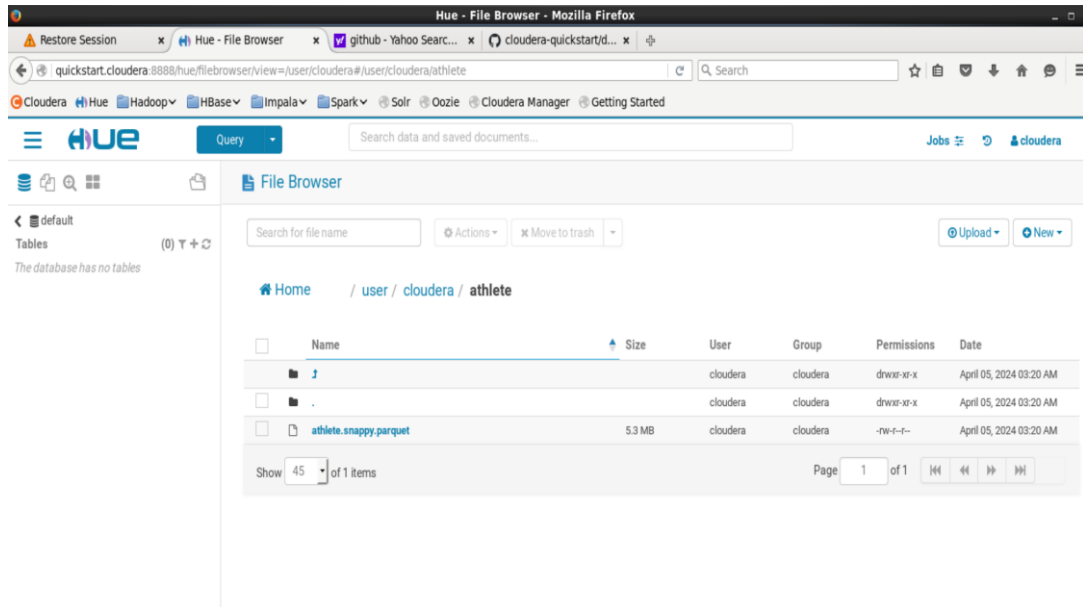
User Metrics for dr.who

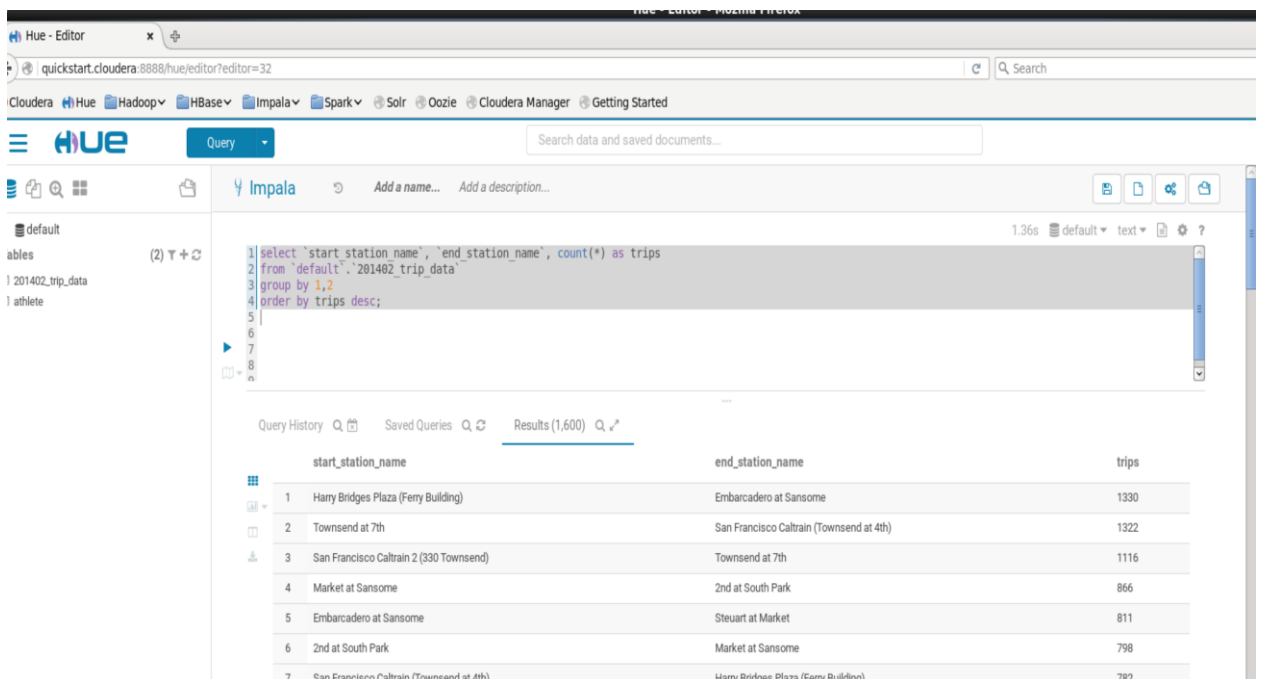
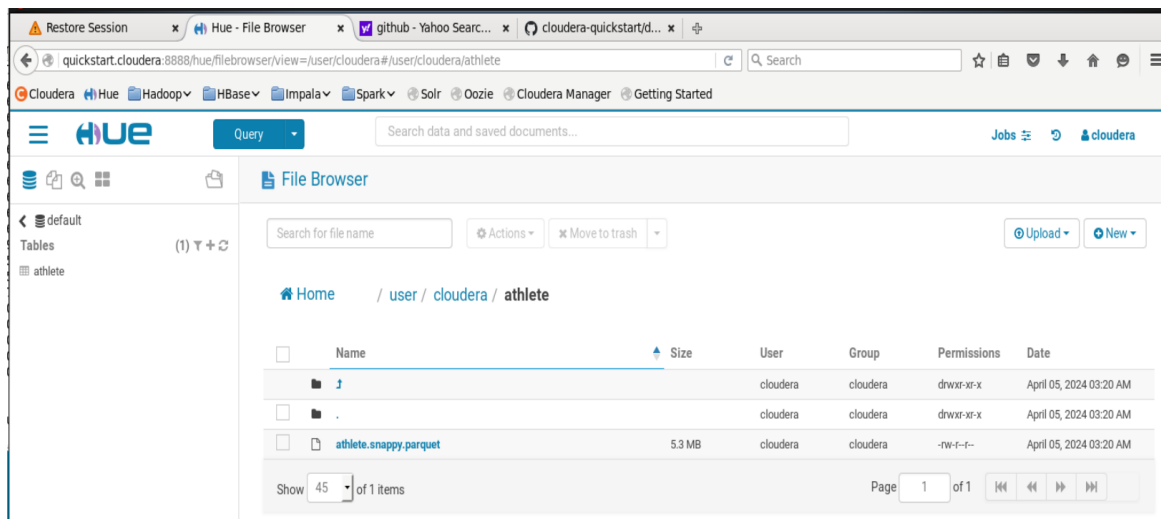
Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Containers Pending	Containers Reserved	Memory Used	Memory Pending	Memory Reserved	VCores Used	VCores Total
0	0	0	0	0	0	0 B	0 B	0 B	0 B	0	0

Show 20 entries

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCo	Allocated Memory MB	Reserved CPU VCo	Rese Mem MB
application_1712307884597_0003	cloudera	word count	MAPREDUCE	root.cloudera	Fri Apr 5 03:01:12 -0700 2024	Fri Apr 5 03:02:22 -0700 2024	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A
application_1712307884597_0001	cloudera	QuasiMonteCarlo	MAPREDUCE	root.cloudera	Fri Apr 5 02:48:31 -0700 2024	Fri Apr 5 02:51:02 -0700 2024	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A

Showing 1 to 2 of 2 entries





```
select `start_station_name`, `end_station_name`, count(*) as trips
from `default`.`201402_trip_data`
group by 1,2
order by trips desc;
```

