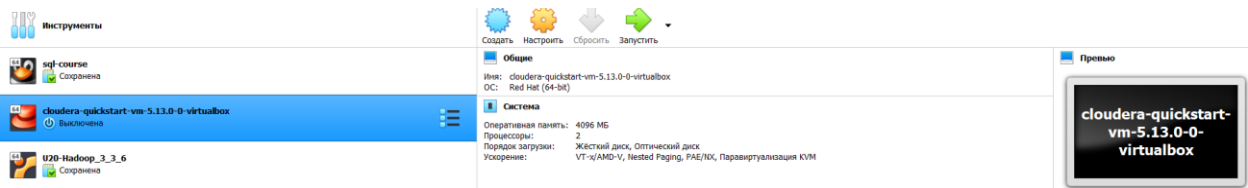
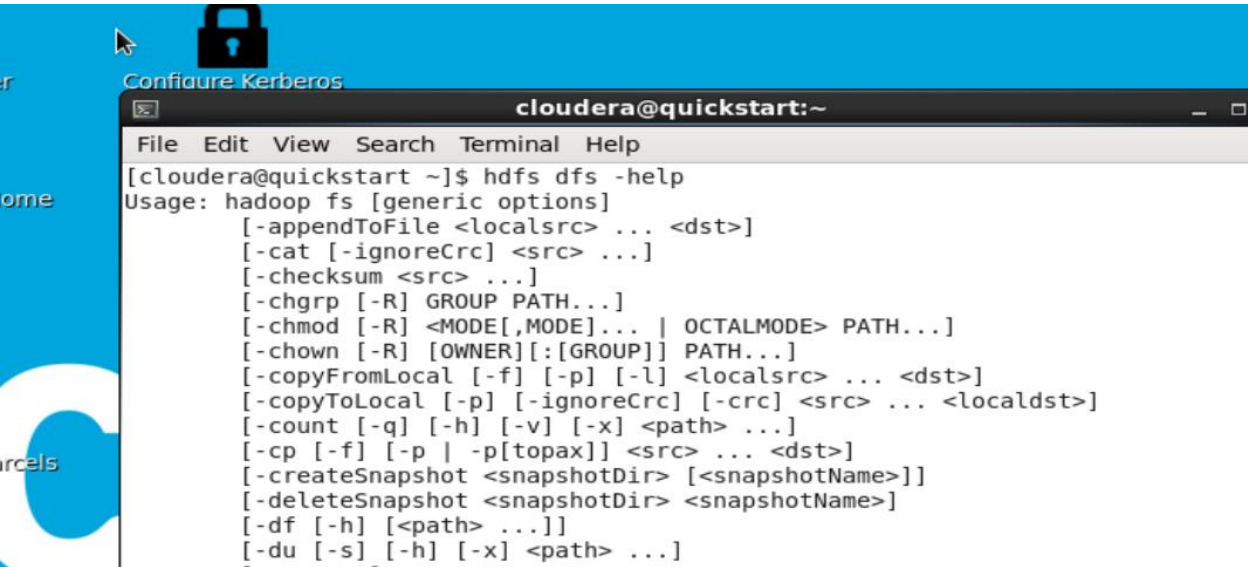


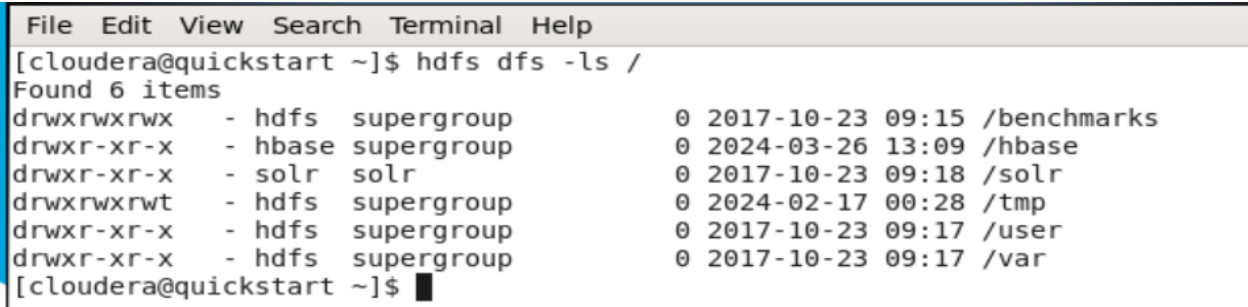
3.1.1. Развернуть виртуальное окружение.+



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

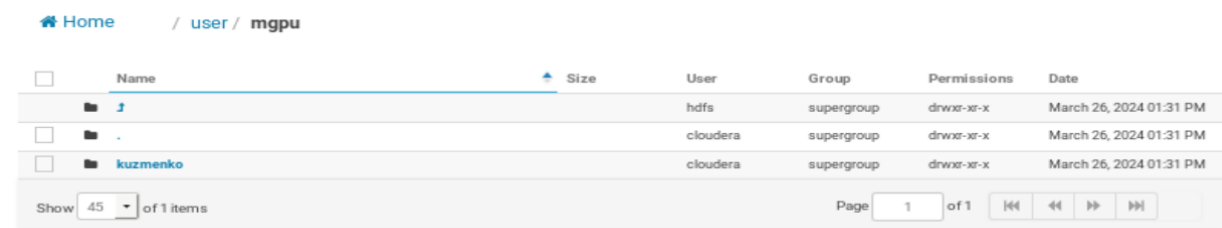


3.1.3. Просмотреть корневую директорию HDFS.



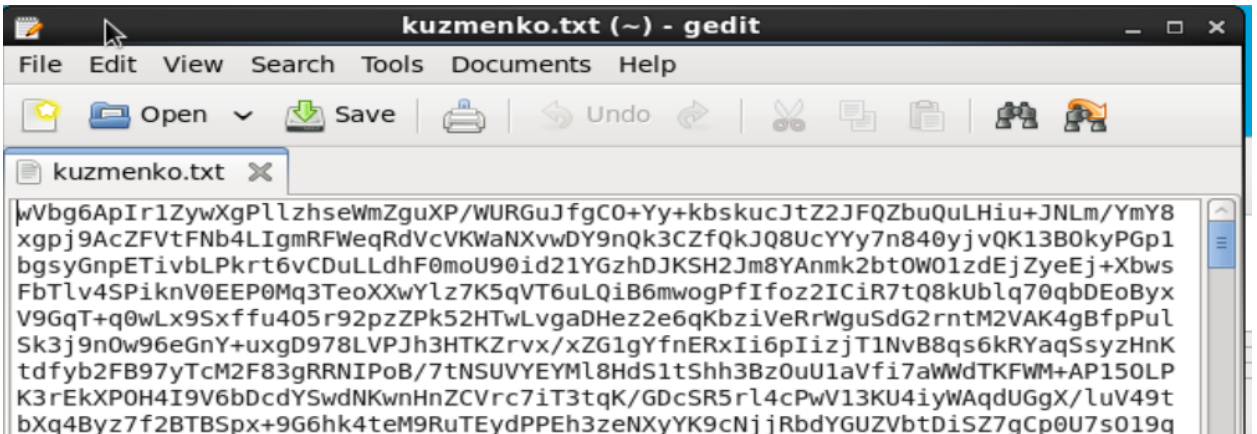
3.1.4. Создать в HDFS в директории /user/mgpu поддиректорию ваше\_фио.

```
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/mgpu
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/mgpu/kuzmenko
[cloudera@quickstart ~]$
```

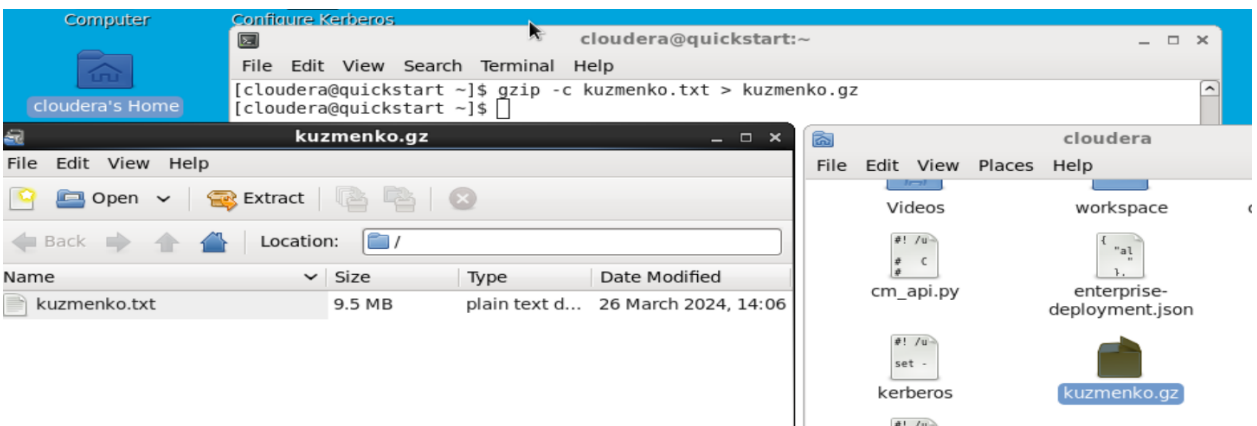


3.1.5. Создать в локальной файловой системе случайный текстовый файл размером 10 Mb с именем, образованным вашими инициалами base64 /dev/urandom | head -c 10000000 > file.txt .

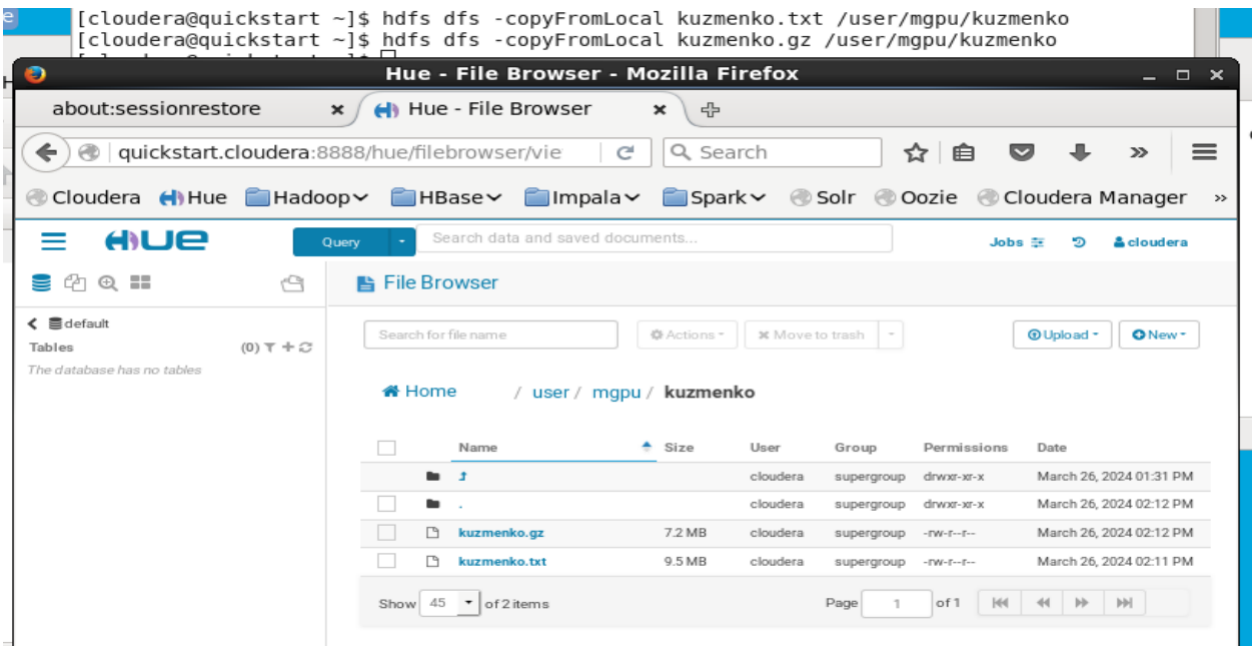
```
[cloudera@quickstart ~]$ base64 /dev/urandom | head -c 10000000 > kuzmenko.txt
```



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

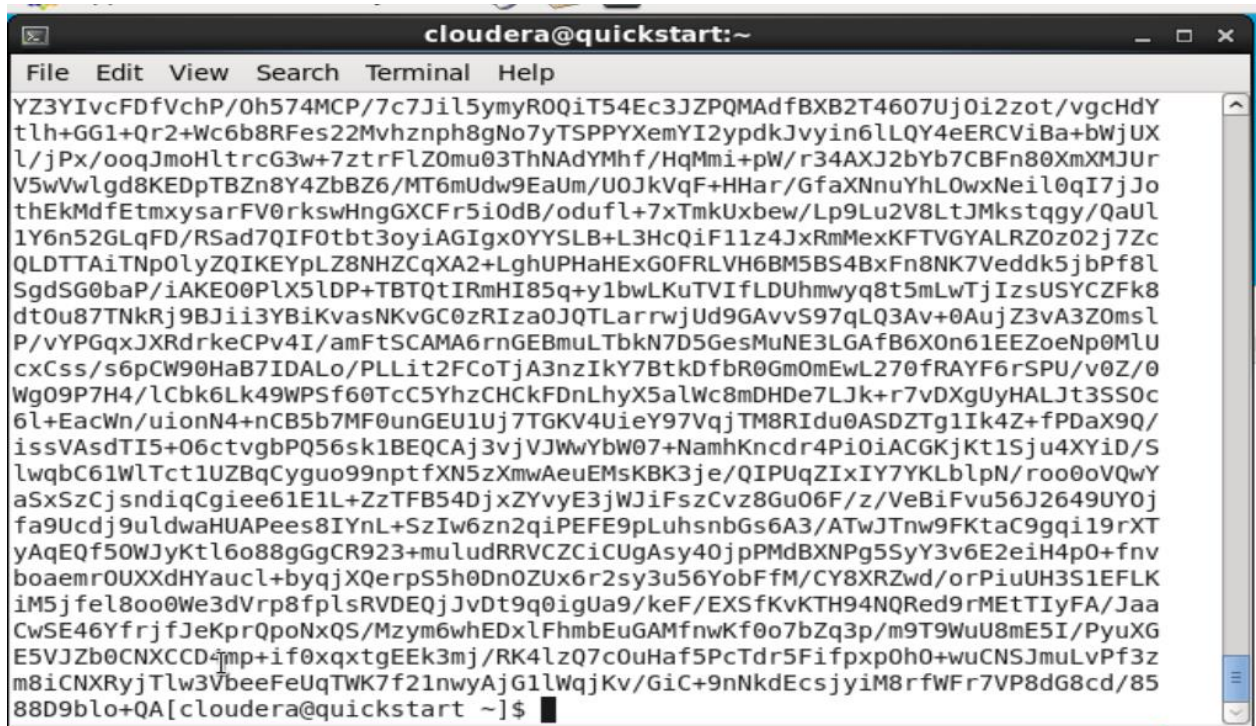


3.1.7. Скопировать текстовый файл и архив в директорию /user/mgpu/fio HDFS виртуальной машины.



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.

```
[cloudera@quickstart ~]$ cat kuzmenko.txt
```



```
cloudera@quickstart:~
File Edit View Search Terminal Help
YZ3YIvcFDfVchP/Oh574MCP/7c7Jil5ymyR0QiT54Ec3JZPQMAdfBxB2T4607Uj0i2zot/vgcHdY
tlh+GG1+Qr2+Wc6b8RFes22Mvhznph8gNo7yTSPPYXemYI2ypdkJvyin6LLQY4eERCViBa+bWjUX
l/jPx/ooqJmoHltrcG3w+7ztrFLZ0mu03ThNAdYMHf/HqMmi+pW/r34AXJ2bYb7CBFn80XmXMJUr
V5wVwlgd8KEDpTBZn8Y4ZbBZ6/MT6mUdw9EaUm/U0JkVqF+HHar/GfaXNnuYhL0wxNeil0qI7jJo
thEkMdfEtmxysarFV0rksWngGXCFr5i0dB/odufL+7xTmkUxbew/Lp9Lu2V8LtJMkstqgy/QaUl
1Y6n52GLqFD/RSad7QIF0tbt3oyiAGIgx0YYSLB+L3HcQiF11z4JxRmMexKFTVGyALRZ0z02j7Zc
QLDTTAiTNp0lyZQIKEYpLZ8NHZCqXA2+LghUPHaHEXG0FRLVH6BM5BS4BxFn8NK7Veddk5jbPf8l
SgdSG0baP/iAKE00PLX5LDP+TBTQtIRmHI85q+y1bwLKuTViFLDUhmwyq8t5mLwTjIzsUSYCZFk8
dt0u87TNkrj9BJii3YBiKvasNKvGC0zRiza0JQTLarrwjUd9GAvvS97qLQ3Av+0AujZ3vA3Z0msl
P/vYPGqxJXRdrkeCPv4I/amFtSCAMA6rnGEBmuLTbKN7D5GesMuNE3LGAfB6X0n61EEZoeNp0MLU
cxCss/s6pCW90HaB7IDALo/PLLit2FCotJA3nzIky7BtkDfbR0Gm0mEwL270fRAYF6rSPU/v0Z/0
Wg09P7H4/lCbK6Lk49WPSf60TcC5YhzCHCKFDnLhyX5aWc8mDHDe7LJk+r7vDXgUyHALJt3SS0c
6l+EacWn/uionN4+nCB5b7MF0unGEU1Uj7TGKV4UieY97VqjTM8RIdu0ASDZTg1Ik4Z+fPDax9Q/
issVAsdTII5+06ctvgbPQ56sk1BEQCAj3vjVJWwYbW07+NamhKncdr4Pi0iACGKjKt1Sju4XYiD/S
lwqbC61WLTct1UZBqCyguo99nptfXN5zXmwAeuEMsKBK3je/QIPUqZIXiY7YKLbLPn/roo0oVQwY
aSxSzCjsndiqCgiee61E1L+ZzTFB54DjxZYvyE3jWJiFsZCvz8Gu06F/z/VeBiFvu56J2649UY0j
fa9Ucdj9uldwaHUAPees8IYNL+SzIw6zn2qiPEFE9pLuhsnbGs6A3/ATwJTnw9FKtaC9gqi19rXT
yAqEQf50WJyKtl6o88gGgCR923+muludRRVCZCiUGAsy40jpPmdBXNPg5SyY3v6E2eiH4p0+fnnv
boaemrOUXXdHYaucl+byqjXQerpS5h0Dn0ZUx6r2sy3u56YobFfm/CY8XRZwd/orPiuUH3S1EFLK
iM5jfel80o0We3dVrp8fplsRVDEqJvDt9q0igUa9/keF/EXSfKvKTH94NQRed9rMetTiyFA/Jaa
CwSE46YfrjfJekprQp0NqXS/Mzym6whEDxlFhmbEuGAMfnwKf0o7bZq3p/m9T9WuU8mE5I/P
yuXG
E5VJZb0CNXCCD4mp+if0xqxtgEEk3mj/RK4lz7c0uHaf5PcTdr5Fifpxp0h0+wuCNSJmuLvPf3z
m8iCNXRyJtlw3VbeeFeUqTWK7f21nwyAjG1lWqjKv/GiC+9nNkdEcsjyiM8rfWFr7VP8dG8cd/85
88D9blo+QA[cloudera@quickstart ~]$
```

```
[cloudera@quickstart ~]$ cat -n kuzmenko.txt
```



```
cloudera@quickstart:~
129860 6l+EacWn/uionN4+nCB5b7MF0unGEU1Uj7TGKV4UieY97VqjTM8RIdu0ASDZTg1Ik4Z+fPDax9Q/
129861 issVAsdTII5+06ctvgbPQ56sk1BEQCAj3vjVJWwYbW07+NamhKncdr4Pi0iACGKjKt1Sju4XYiD/S
129862 lwqbC61WLTct1UZBqCyguo99nptfXN5zXmwAeuEMsKBK3je/QIPUqZIXiY7YKLbLPn/roo0oVQwY
129863 aSxSzCjsndiqCgiee61E1L+ZzTFB54DjxZYvyE3jWJiFsZCvz8Gu06F/z/VeBiFvu56J2649UY0j
129864 fa9Ucdj9uldwaHUAPees8IYNL+SzIw6zn2qiPEFE9pLuhsnbGs6A3/ATwJTnw9FKtaC9gqi19rXT
129865 yAqEQf50WJyKtl6o88gGgCR923+muludRRVCZCiUGAsy40jpPmdBXNPg5SyY3v6E2eiH4p0+fnnv
129866 boaemrOUXXdHYaucl+byqjXQerpS5h0Dn0ZUx6r2sy3u56YobFfm/CY8XRZwd/orPiuUH3S1EFLK
129867 iM5jfel80o0We3dVrp8fplsRVDEqJvDt9q0igUa9/keF/EXSfKvKTH94NQRed9rMetTiyFA/Jaa
129868 CwSE46YfrjfJekprQp0NqXS/Mzym6whEDxlFhmbEuGAMfnwKf0o7bZq3p/m9T9WuU8mE5I/P
129869 E5VJZb0CNXCCD4mp+if0xqxtgEEk3mj/RK4lz7c0uHaf5PcTdr5Fifpxp0h0+wuCNSJmuLvPf3z
129870 m8iCNXRyJtlw3VbeeFeUqTWK7f21nwyAjG1lWqjKv/GiC+9nNkdEcsjyiM8rfWFr7VP8dG8cd/85
d/85
```

```
[cloudera@quickstart ~]$ cat -b kuzmenko.txt
```



```
SS0c
129860 6l+EacWn/uionN4+nCB5b7MF0unGEU1Uj7TGKV4UieY97VqjTM8RIdu0ASDZTg1Ik4
X9Q/
129861 issVAsdTI5+06ctvgbPQ56sk1BEQCAj3vjVJWwYbW07+NamhKncdr4Pi0iACGKjKt1
iD/S
129862 lwqbC61WlTct1UzBqCyguo99nptfXN5zXmwAeuEMsKBK3je/QIPUqZIXIY7YKLblpN
VQwY
129863 aSxSzCjsndiqCgiee61E1L+ZzTFB54DjxZYvyE3jWJiFszCvz8Gu06F/z/VeBiFvu5
UY0j
129864 fa9Ucdj9uldwaHUApees8IYnL+SzIw6zn2qiPEFE9pLuhsnbGs6A3/ATwJTnw9FKta
9rXT
129865 yAqEQf50WJyKtl6o88gGgCR923+muludRRVCZCiCuGAsy40jpPmDBXNPg5SyY3v6E2
+fnv
129866 boaemr0UXXdHYaucl+byqjXQerpS5h0Dn0ZUx6r2sy3u56YobFfM/CY8XRZwd/orPi
EFLK
129867 iM5jfel8oo0We3dVrp8fplsRVDEQjJvDt9q0igUa9/keF/EXSfKvKTH94NQRed9rME
/Jaa
129868 CwSE46YfrjfJekPrQpoNxQS/Mzym6whEDxlFhmbEuGAMfnwKf0o7bZq3p/m9T9WuU8
yuXG
129869 E5VJZb0CNXCCD4mp+if0xqxtgEEk3mj/RK4lzQ7c0uHaf5PcTdr5Fifpxp0h0+wuCf
Pf3z
129870 m8iCNXRyjTlw3VbeeFeUqTWK7f21nwyAjG1lWqjKv/GiC+9nNkdEcsjyiM8rfWFr7l
d/85
```

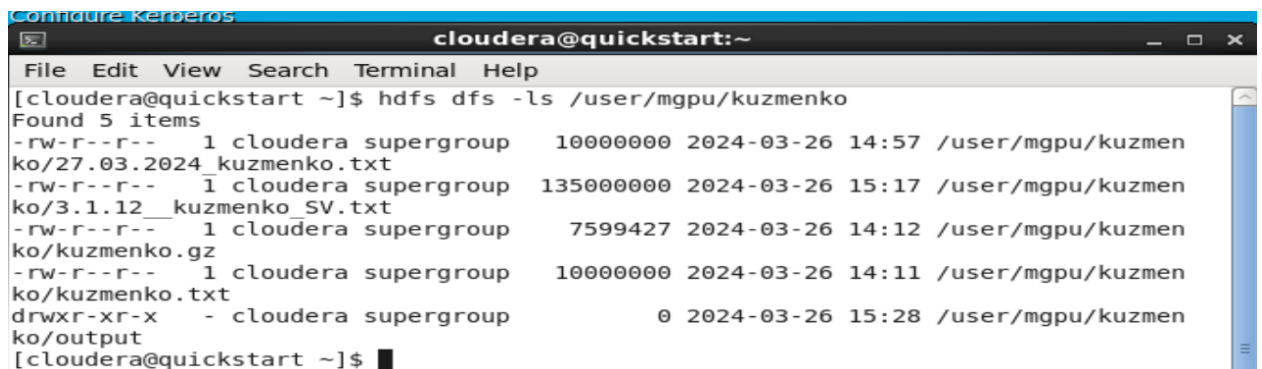
```
[cloudera@quickstart ~]$ head kuzmenko.txt
wVbg6ApIr1ZyWxgPlLzhseWmZguXP/WURGuJfgCO+Yy+kbskucJtZ2JFQZbuQuLHiu+JNLm/YmY8
xgpj9AcZFvtFNb4LIgmRFWeqRdVcVKWaNxvWdY9nQk3CZfQkQJ8UcYyy7n840yjjvQK13B0kyPGp1
bgsyGnpETivbLPkrt6vCDuLLdhF0moU90id21YGzhDJKSH2Jm8YAnmk2bt0W01zdEjZyeEj+Xbws
FbTlv4SPiknV0EEP0Mq3TeoXXwYlZ7K5qVT6uLQiB6mwogPfIfoz2ICiR7tQ8kUblq70qbDEoByx
V9GqT+q0wLx9Sxffu405r92pzZPk52HTwLvgaDHez2e6qKbziVeRrWguSdG2rntM2VAK4gBfpPuL
Sk3j9n0w96eGnY+uxgD978LVPJh3HTKZrvx/xZG1gYfnERxi6pIizjT1NvB8qs6kRYaqSSyzHnK
tdfyb2FB97yTcm2F83gRRNIPoB/7tNSUVYeyML8HdS1tShh3Bz0uU1aVfi7aWwdTKFWM+AP150LP
K3rEkXP0H4I9V6bDcdYSwdNKwnHnZCvrc7iT3tqK/GDcSR5rl4cPwV13KU4iyWAqdUGgX/luV49t
bXq4Byz7f2BTBSpx+9G6hk4teM9RuTEydPPEh3zeNXyYK9cNjjRbdYUGUZVbtDiSZ7gCp0U7s019q
4468Id+JC30mrZgoxeRJKhYJe7lfr+AmrnVdS8J3zihRxtPQ9+HcrpnGotd67U4qK9tkuqdjV/va
[cloudera@quickstart ~]$ head -c 7 kuzmenko.txt
wVbg6Ap[cloudera@quickstart ~]$ head -n 7 kuzmenko.txt
wVbg6ApIr1ZyWxgPlLzhseWmZguXP/WURGuJfgCO+Yy+kbskucJtZ2JFQZbuQuLHiu+JNLm/YmY8
xgpj9AcZFvtFNb4LIgmRFWeqRdVcVKWaNxvWdY9nQk3CZfQkQJ8UcYyy7n840yjjvQK13B0kyPGp1
bgsyGnpETivbLPkrt6vCDuLLdhF0moU90id21YGzhDJKSH2Jm8YAnmk2bt0W01zdEjZyeEj+Xbws
FbTlv4SPiknV0EEP0Mq3TeoXXwYlZ7K5qVT6uLQiB6mwogPfIfoz2ICiR7tQ8kUblq70qbDEoByx
V9GqT+q0wLx9Sxffu405r92pzZPk52HTwLvgaDHez2e6qKbziVeRrWguSdG2rntM2VAK4gBfpPuL
Sk3j9n0w96eGnY+uxgD978LVPJh3HTKZrvx/xZG1gYfnERxi6pIizjT1NvB8qs6kRYaqSSyzHnK
tdfyb2FB97yTcm2F83gRRNIPoB/7tNSUVYeyML8HdS1tShh3Bz0uU1aVfi7aWwdTKFWM+AP150LP
[cloudera@quickstart ~]$
```

```
[cloudera@quickstart ~]$ tail kuzmenko.txt
lwqbC61WlTct1UzBqCyguo99nptfXN5zXmwAeuEMsKBK3je/QIPUqZIXIY7YKLblpN/roo0oVQwY
aSxSzCjsndiqCgiee61E1L+ZzTFB54DjxZYvyE3jWJiFszCvz8Gu06F/z/VeBiFvu56J2649UY0j
fa9Ucdj9uldwaHUApees8IYnL+SzIw6zn2qiPEFE9pLuhsnbGs6A3/ATwJTnw9FKtaC9gqi19rXT
yAqEQf50WJyKtl6o88gGgCR923+muludRRVCZCiCuGAsy40jpPmDBXNPg5SyY3v6E2eiH4p0+fnv
boaemr0UXXdHYaucl+byqjXQerpS5h0Dn0ZUx6r2sy3u56YobFfM/CY8XRZwd/orPiuUH3S1EFLK
iM5jfel8oo0We3dVrp8fplsRVDEQjJvDt9q0igUa9/keF/EXSfKvKTH94NQRed9rMEtTiYFA/Jaa
CwSE46YfrjfJekPrQpoNxQS/Mzym6whEDxlFhmbEuGAMfnwKf0o7bZq3p/m9T9WuU8mE5I/PyuXG
E5VJZb0CNXCCD4mp+if0xqxtgEEk3mj/RK4lzQ7c0uHaf5PcTdr5Fifpxp0h0+wuCNSJmuLvPf3z
m8iCNXRyjTlw3VbeeFeUqTWK7f21nwyAjG1lWqjKv/GiC+9nNkdEcsjyiM8rfWFr7VP8dG8cd/85
88D9blo+QA[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ tail -c 7 kuzmenko.txt
9blo+QA[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ tail -n 7 kuzmenko.txt
yAqEQf50WJyKtl6o88gGgCR923+muludRRVCZCiCuGAsy40jpPmDBXNPg5SyY3v6E2eiH4p0+fnv
boaemr0UXXdHYaucl+byqjXQerpS5h0Dn0ZUx6r2sy3u56YobFfM/CY8XRZwd/orPiuUH3S1EFLK
iM5jfel8oo0We3dVrp8fplsRVDEQjJvDt9q0igUa9/keF/EXSfKvKTH94NQRed9rMEtTiYFA/Jaa
CwSE46YfrjfJekPrQpoNxQS/Mzym6whEDxlFhmbEuGAMfnwKf0o7bZq3p/m9T9WuU8mE5I/PyuXG
E5VJZb0CNXCCD4mp+if0xqxtgEEk3mj/RK4lzQ7c0uHaf5PcTdr5Fifpxp0h0+wuCNSJmuLvPf3z
m8iCNXRyjTlw3VbeeFeUqTWK7f21nwyAjG1lWqjKv/GiC+9nNkdEcsjyiM8rfWFr7VP8dG8cd/85
88D9blo+QA[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
```

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

```
[cloudera@quickstart ~]$ cp kuzmenko.txt 27.03.2024_kuzmenko.txt
[cloudera@quickstart ~]$ ls
27.03.2024_kuzmenko.txt Desktop eclipse kerberos lib Pictures Videos
cloudera-manager Documents enterprise-deployment.json kuzmenko.gz Music Public workspace
cm_api.py Downloads express-deployment.json kuzmenko.txt parcels Templates
[cloudera@quickstart ~]$ hdfs dfs -cat /user/mgpu/kuzmenko/kuzmenko.txt ^C
[cloudera@quickstart ~]$ hdfs dfs -cp /user/mgpu/kuzmenko/kuzmenko.txt /user/mgpu/kuzmenko/27.03.2024_kuzmenko.txt
[cloudera@quickstart ~]$ hdfs dfs -ls /user/mgpu/kuzmenko/
Found 3 items
-rw-r--r-- 1 cloudera supergroup 10000000 2024-03-26 14:57 /user/mgpu/kuzmenko/27.03.2024_kuzmenko.txt
-rw-r--r-- 1 cloudera supergroup 7599427 2024-03-26 14:12 /user/mgpu/kuzmenko/kuzmenko.gz
-rw-r--r-- 1 cloudera supergroup 10000000 2024-03-26 14:11 /user/mgpu/kuzmenko/kuzmenko.txt
[cloudera@quickstart ~]$
```

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



```
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hdfs dfs -ls /user/mgpu/kuzmenko
Found 5 items
-rw-r--r-- 1 cloudera supergroup 10000000 2024-03-26 14:57 /user/mgpu/kuzmenko/27.03.2024_kuzmenko.txt
-rw-r--r-- 1 cloudera supergroup 135000000 2024-03-26 15:17 /user/mgpu/kuzmenko/3.1.12_kuzmenko_SV.txt
-rw-r--r-- 1 cloudera supergroup 7599427 2024-03-26 14:12 /user/mgpu/kuzmenko/kuzmenko.gz
-rw-r--r-- 1 cloudera supergroup 10000000 2024-03-26 14:11 /user/mgpu/kuzmenko/kuzmenko.txt
drwxr-xr-x - cloudera supergroup 0 2024-03-26 15:28 /user/mgpu/kuzmenko/output
[cloudera@quickstart ~]$
```

3.1.11. Удалить поддиректорию /fio со всем содержимым.

```
[cloudera@quickstart ~]$ hdfs dfs -rm -R /user/mgpu/kuzmenko
Deleted /user/mgpu/kuzmenko
[cloudera@quickstart ~]$
```

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

```
[cloudera@quickstart ~]$ base64 /dev/urandom | head -c 135000000 > 3.1.12_kuzmenko_SV.txt
[cloudera@quickstart ~]$ hdfs dfs -copyFromLocal 3.1.12_kuzmenko_SV.txt /user/mgpu/kuzmenko

[cloudera@quickstart ~]$ yarn jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar wordcount /user/mgpu/kuzmenko/3.1.12_kuzmenko_SV.txt /user/mgpu/kuzmenko/output
24/03/26 15:26:59 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
24/03/26 15:27:03 INFO input.FileInputFormat: Total input paths to process : 1
24/03/26 15:27:04 INFO mapreduce.JobSubmitter: Number of splits:1
24/03/26 15:27:05 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1711484710610_0001
24/03/26 15:27:07 INFO impl.YarnClientImpl: Submitted application application_1711484710610_0001
24/03/26 15:27:08 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1711484710_0001/
24/03/26 15:27:08 INFO mapreduce.Job: Running job: job_1711484710610_0001
24/03/26 15:27:44 INFO mapreduce.Job: Job job_1711484710610_0001 running in uber mode : false
24/03/26 15:27:44 INFO mapreduce.Job: map 0% reduce 0%
24/03/26 15:28:10 INFO mapreduce.Job: map 41% reduce 0%
24/03/26 15:28:16 INFO mapreduce.Job: map 67% reduce 0%
24/03/26 15:28:23 INFO mapreduce.Job: map 87% reduce 0%
24/03/26 15:28:25 INFO mapreduce.Job: map 100% reduce 0%
24/03/26 15:28:40 INFO mapreduce.Job: map 100% reduce 100%
24/03/26 15:28:40 INFO mapreduce.Job: Job job_1711484710610_0001 completed successfully
24/03/26 15:28:40 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=291038990
  FILE: Number of bytes written=436845764
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=135000139
  HDFS: Number of bytes written=138506495
  HDFS: Number of read operations=6
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=37809
  Total time spent by all reduces in occupied slots (ms)=12389
  Total time spent by all map tasks (ms)=37809
  Total time spent by all reduce tasks (ms)=12389
  Total vcore-milliseconds taken by all map tasks=37809
```

```
Home / user / mgpu / kuzmenko / output / part-r-00000

+++0eClfWNEItbY1IM55A6HEh3QnmCX8xHJyB79YL/a2op1STKKSbTmcncRjTnxVEW1H1uG6/C+ 1
+++2f2aEKxy+JJB6/9g8Zuc4qImp555GSEIThCp6oY+2U37cLvY1TvgWTmwfY1Jj9xNu8PhUYMI 1
+++E0YQGVq3CDIuszxxA103NGACJt+pl+dhY4N/iF6C+BtsDpwMjP3j0G6EbTUG6xD/7ETy7tU68 1
+++0hxhBtA/+tw0fVnJ4w8i7Tua6pYSMV2u7yNWyJPAyJtpNICxVdsXM7QmJN3xkgWTPIS0TpkZ 1
+++VhYXJSupm3LJdVU2BUFG/VmJ0B+oFkSTuvMgXcBa2wzPeIgy41igy7tzBxaX2wiV28Xpb1tId 1
+++dOMXr1mHb1T9+xlFqO+GXQnw8C6yB6SEh7/PpgN97BK/xVH5KdYhHvayFT92jDqdNfjnJ+nmjg 1
+++fnhtFsF9buUS4HXQsAdmmK0vZm4nQsZxPZRTTMB5T12UuVYiAx2/R1/Gmugana896KIB+Set 1
+++gmKegqP+tO46HTDor9PI0bH4qFeZWPX/QjdKUBkb4NBKQ1Qdhk6UPLw8QakEj1JmG6fkbHtAw 1
+++nkJ0Gwvv/558TZttdmubAESEsR4n6rU7cPJuw2vV76snk8mJCrSwuYWSa++0DFqcMYohrjqkB 1
+++rdPPpMnROJZRJ7btAGQhzw9mru3Y90TsA22zXwjJ+14QQjWRfHuHcWw/ahHqFx9zELoze2vIP 1
++/B6XFjhVGI4IAfe4o+2UbG4YdtW0D5Uts0Xx9LHKu5JScENJqUndXpTL6D/y76w0HL1Nx45Rw 1
++/JKgRe5Wk5mSSKy8CADgEUvD8+o1A/2cNKom3nkW/iOLOZsQb/+oPB1/D3yt616YUWbX9+wg+ 1
++/w0M9y1+aPdexCE0+zQptJNypU310GQ4Yi7DWB8BKMhs1E9S5BIJqdc22Npd9XJmmLb0oS1oK 1
++02mQYT4ay1t5EiRcBmpHv1YJhEpG81TKMA5L6/Gtc+JJ1TAzD8iJrQPrMJoqKjxJ6jjYD7uod 1
++0Bous8Dnfx4E6EamgqUHD/ntpiRC3LWxF7bU7gBd8uoXsFdUxKbhbOxnd9eAWHakMpEn54Ybwr 1
++0w7NB8i+kcPww/U5hV7100k10kOCxf5tZ5xtw0ppL9jzDJ3EY7PEkTVJLRwuaYdHqLG+CV4D0 1
++178qhSEAnrtgISuHxoZ4a+t5eNkK1OTwfdDCDN/e73tYgMZFSN3s10cXqs+6Gskfj9xP372KDS 1
++1E3KB/60bImBYdfkZJ5kP19KbHhwpn0z/Rn4UTGn/MdxXQNAh1JBYPF/4uVzWDanraRE9f3tv/ 1
++1nns70eBnA70M1A0dFnc1161Vf30eYhC7VfK2NvXbUuV7T0DnucVc3Vn+3u+7V46V4K 1
```

### 3.2. Создание таблицы в Hive

1. Скачать [датасет](#) или [тут](#)

```
[cloudera@quickstart ~]$ wget https://github.com/BosenkoTM/cloudera-quickstart/blob
/main/data/athlete.snappy.parquet
--2024-03-27 12:51:44-- https://github.com/BosenkoTM/cloudera-quickstart/blob/main
/data/athlete.snappy.parquet
Resolving github.com... 140.82.121.4
Connecting to github.com|140.82.121.4|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: unspecified [text/html]
Saving to: "athlete.snappy.parquet"

[ <==> ] 272,193 1.52M/s in 0.2s

2024-03-27 12:51:45 (1.52 MB/s) - "athlete.snappy.parquet" saved [272193]

[cloudera@quickstart ~]$ ls
27.03.2024_kuzmenko.txt Documents
3.1.12_kuzmenko_SV.txt Downloads
athlete.snappy.parquet eclipse
cloudera-manager enterprise-deployment.json kuzmenko.kgz Pictures
cm_api.py express-deployment.json kuzmenko.txt Public
Desktop kerberos lib Music Templates
Videos
workspace
```

2. Через [HUE](#) загрузите файл в папку `/user/cloudera/athlete`.
3. В навигационном меню выберите `Files`.
4. Создайте папку.
5. Загрузите файл в `HDFS`, нажав `Upload`.

Home

/ user / cloudera / athlete

<input type="checkbox"/>	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	<a href="#">.</a>		cloudera	cloudera	drwxr-xr-x	March 27, 2024 01:01 PM
<input type="checkbox"/>	<a href="#">.</a>		cloudera	cloudera	drwxr-xr-x	March 27, 2024 01:02 PM
<input type="checkbox"/>	<a href="#">athlete.snappy.parquet</a>	265.8 KB	cloudera	cloudera	-rw-r--r--	March 27, 2024 01:02 PM

Show45of 1 items

Page1of 1

6. Перейдите в "Editor > Hive" и выполните запрос:

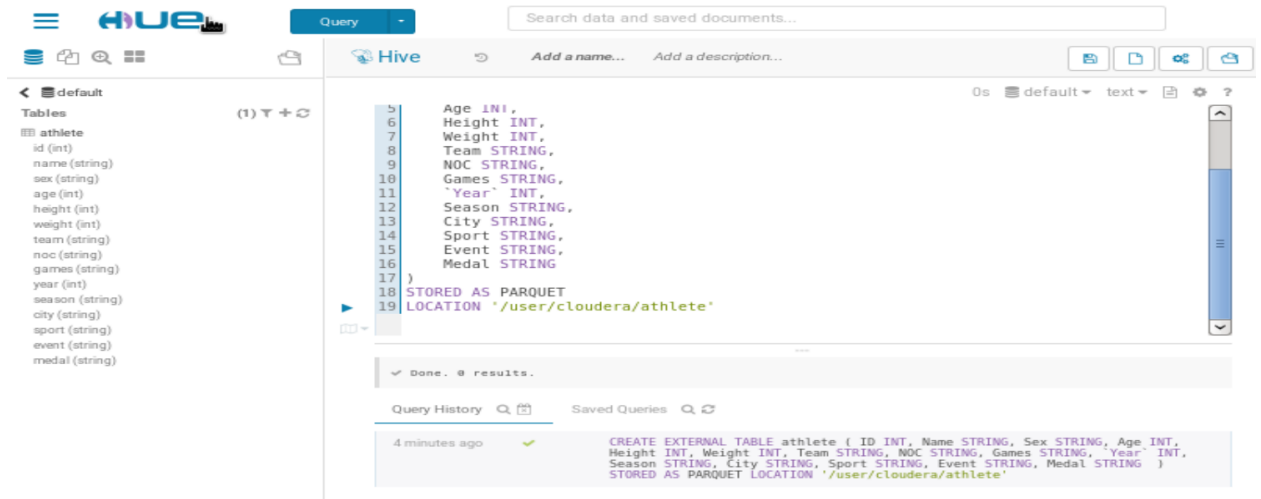
```
CREATE EXTERNAL TABLE athlete (  
  ID INT,  
  Name STRING,  
  Sex STRING,  
  Age INT,  
  Height INT,  
  Weight INT,
```



```

Team STRING,
NOC STRING,
Games STRING,
`Year` INT,
Season STRING,
City STRING,
Sport STRING,
Event STRING,
Medal STRING
)
STORED AS PARQUET
LOCATION '/user/cloudera/athlete'

```



3.3 Проанализировать и визуализировать данные с помощью Impala(высокоскоростной механизм запросов SQL) или Hive.

Загрузить и разархивировать babs\_open\_data\_year\_1.zip.

```




[cloudera@quickstart ~]$ unzip babs_open_data_year_1.zip
Archive:  babs_open_data_year_1.zip
  creating: 201402_babs_open_data/
  inflating: 201402_babs_open_data/201402_station_data.csv
  inflating: 201402_babs_open_data/201402_status_data.csv
  inflating: 201402_babs_open_data/201402_trip_data.csv
  inflating: 201402_babs_open_data/201402_weather_data.csv
  inflating: 201402_babs_open_data/README.txt
  creating: 201408_babs_open_data/
  inflating: 201408_babs_open_data/201408_station_data.csv
  inflating: 201408_babs_open_data/201408_status_data.csv
  inflating: 201408_babs_open_data/201408_trip_data.csv
  inflating: 201408_babs_open_data/201408_weather_data.csv
  inflating: 201408_babs_open_data/README.txt

```

Перенести данные 201402\_trip\_data.csv в HDFS.

Home

/ user / cloudera / trip\_data

<input type="checkbox"/>	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	 <a href="#">↑</a>		cloudera	cloudera	drwxr-xr-x	March 28, 2024 12:38 PM
<input type="checkbox"/>	 .		cloudera	cloudera	drwxr-xr-x	March 28, 2024 12:38 PM
<input type="checkbox"/>	 <a href="#">201402_trip_data.csv</a>	16.4 MB	cloudera	cloudera	-rw-r--r--	March 28, 2024 11:51 AM

Show

45

of 1 items

Page

1

of 1

⏪

⏴

⏵


⏩


Создать таблицу в Hive с привязкой к внешним данным 201402\_trip\_data.csv.


0s

default

text









```
1 drop table if exists 201402_trip_data;
2 CREATE EXTERNAL TABLE 201402_trip_data (
3   TripID INT,
4   Duration INT,
5   StartDate STRING,
6   startstation STRING,
7   StartTerminal INT,
8   EndDate STRING,
9   endstation STRING,
10  EndTerminal INT,
11  Bike INT,
12  SubscriptionType STRING,
13  ZipCode STRING
14 )
15 ROW FORMAT DELIMITED
16 FIELDS TERMINATED BY ','
17 LOCATION '/user/cloudera/trip_data'
```

2/2





✓ Success.

выполнить запрос

```
select `startstation`, `endstation`, count(*) as trips
from `default`.`201402_trip_data`
group by `startstation`, `endstation`
order by trips desc;
```

VIEWED

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved
7	0	0	7	0	0 B	8 GB	0 B

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unheal
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	Memor Reserve
0	0	0	0	0	0	0	0 B	0 B	0 B

Show 20 entries

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCoers
application_1711484710610_0007	cloudera	select startstation, endstation, coun...desc(Stage-2)	MAPREDUCE	root.cloudera	Thu Mar 28 13:23:41 -0700	Thu Mar 28 13:24:42 -0700	FINISHED	SUCCEEDED	N/A	N/A



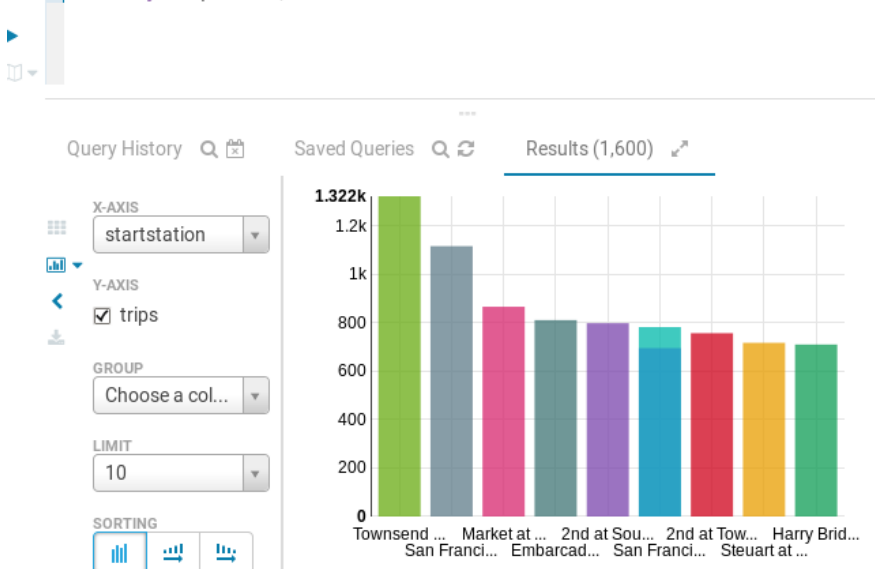
```
1 select startstation, endstation, count(*) as trips
2 from default.201402_trip_data
3 group by startstation, endstation
4 order by trips desc;
```

	startstation	endstation	trips
1	Harry Bridges Plaza (Ferry Building)	Embarcadero at Sansome	1330
2	Townsend at 7th	San Francisco Caltrain (Townsend at 4th)	1322
3	San Francisco Caltrain 2 (330 Townsend)	Townsend at 7th	1116
4	Market at Sansome	2nd at South Park	866
5	Embarcadero at Sansome	Steuart at Market	811
6	2nd at South Park	Market at Sansome	798

Создать гистограмму, щелкнув значок «Hue Bar»:

Установить ось X в качестве начальной станции, а ось Y — в качестве маршрута. Установить лимит 10.

```
1 select startstation, endstation, count(*) as trips
2 from default.201402_trip_data
3 group by startstation, endstation
4 order by trips desc;
```



Выгрузить результаты, выбрав CSV или Excel.

☒ COLUMNS

☒ startstation

☒ endstation

☒ trips

CSV

Excel

Clipboard

Save

	startstation	endstation
1	Harry Bridges Plaza (Ferry Building)	Embarcadero
2	Townsend at 7th	San Francisco
3	San Francisco Caltrain 2 (330 Townsend)	Townsend at
4	Market at Sansome	2nd at South f
5	Embarcadero at Sansome	Steuart at Ma
6	2nd at South Park	Market at Sar
7	San Francisco Caltrain (Townsend at 4th)	Harry Bridges
8	2nd at Townsend	Harry Bridges



query-hive-26.xlsx