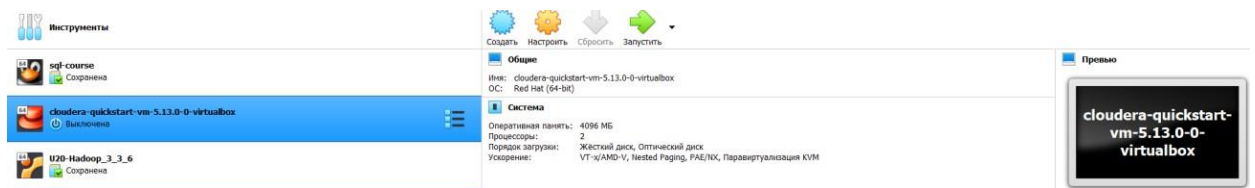
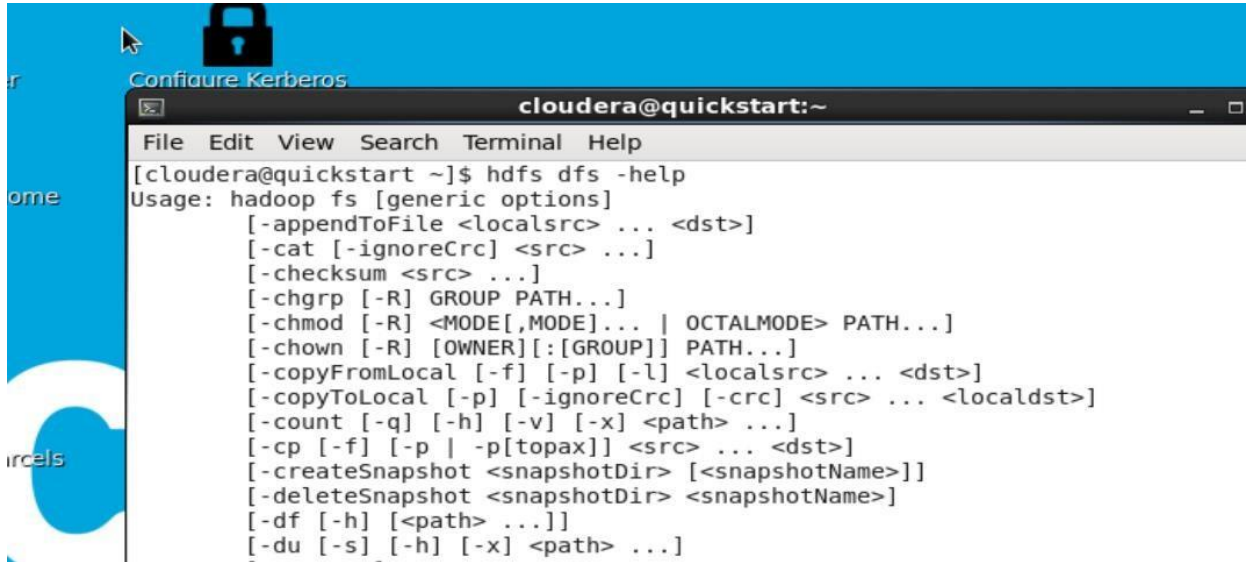


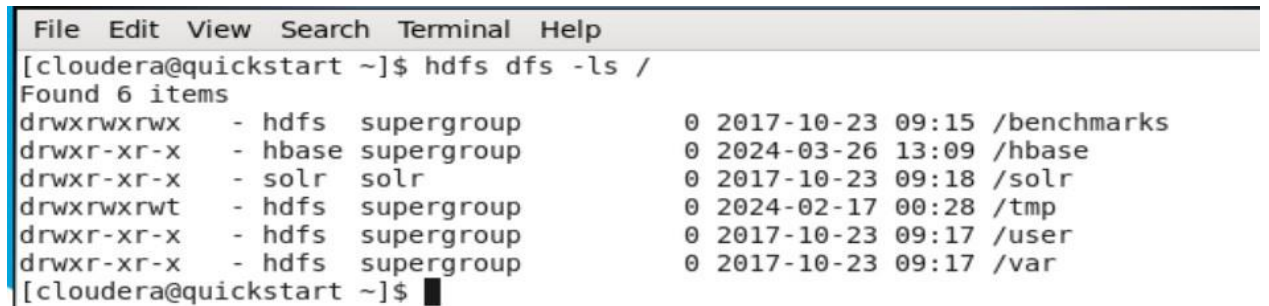
3.1.1. Развернуть виртуальное окружение.+



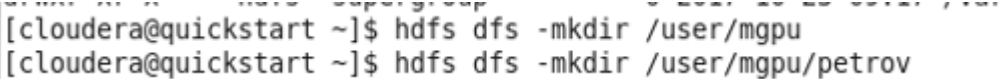
3.1.2. Вывести с помощью команды help описание основных команды shell-клиента.



3.1.3. Просмотреть корневую директорию HDFS.

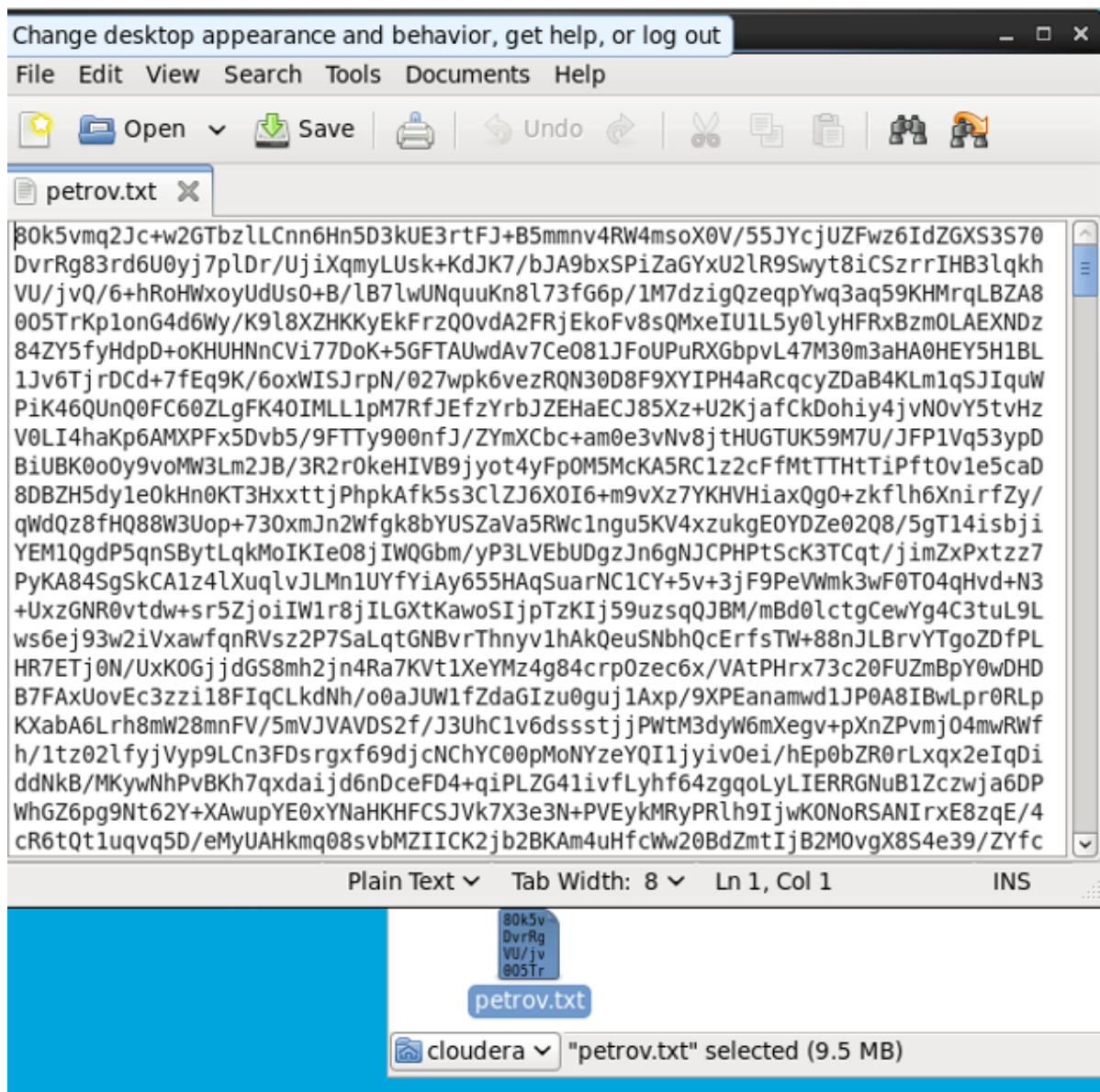


3.1.4. Создать в HDFS в директории /user/mgпу поддиректорию ваше_фио.

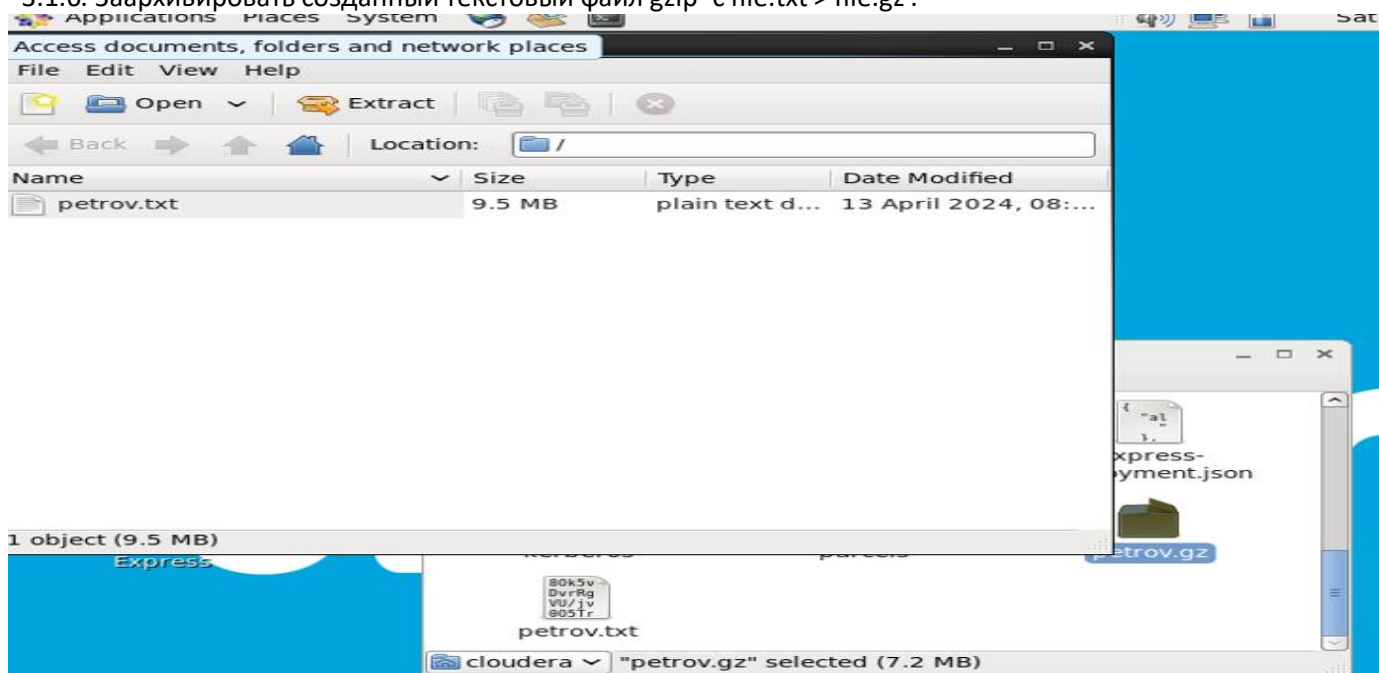


3.1.5. Создать в локальной файловой системе случайный текстовый файл размером 10 Mb с именем, образованным вашими инициалами base64 /dev/urandom | head -c 10000000 > file.txt .



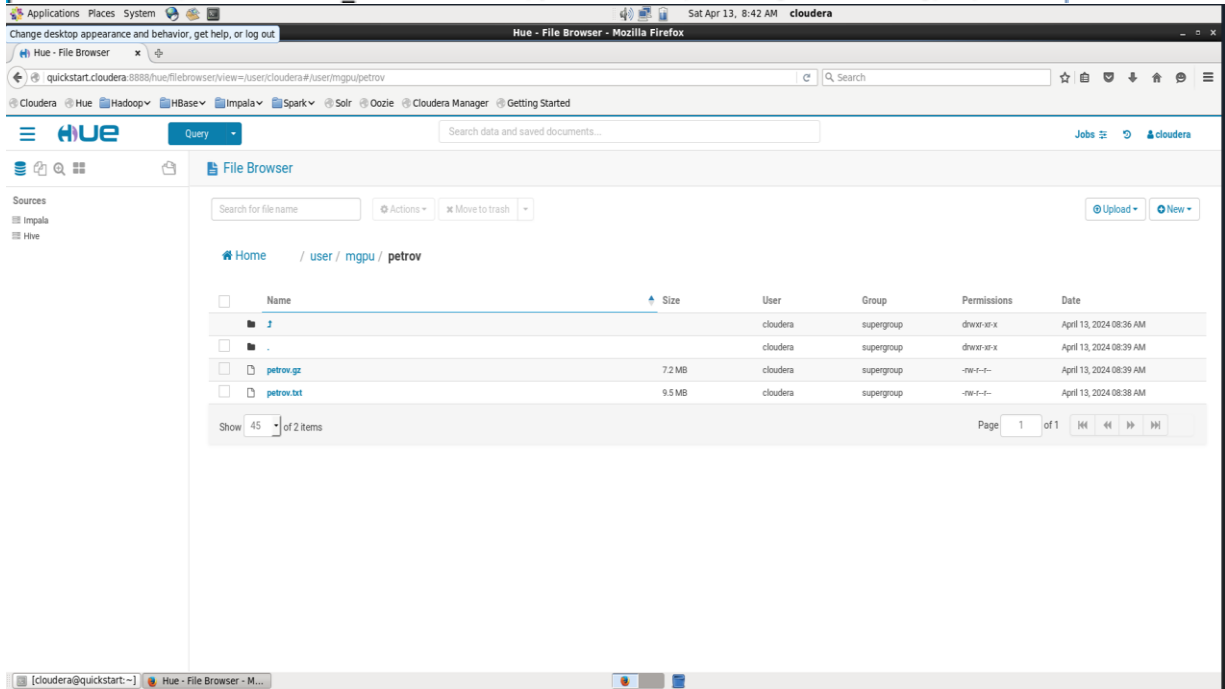


3.1.6. Заархивировать созданный текстовый файл gzip -c file.txt > file.gz .



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

```
[cloudera@quickstart ~]$ hdfs dfs -copyFromLocal petrov.txt /user/mgpu/petrov  
^[[cloudera@quickstart ~]$ hdfs dfs -copyFromLocal petrov.gz /user/mgpu/petrov
```



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 petrov.txt
```

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

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



```
cloudera@quickstart:~
File Edit View Search Terminal Help
YZ3YIvcFDfVchP/0h574MCP/7c7Jil5ymyROqIT54Ec3JZPQMAdfBxB2T4607Uj0i2zot/vgcHdY
tlh+GG1+Qr2+Wc6b8RFes22Mvhznph8gNo7yTSPPYXemyYI2ypdkJvyin6LLQY4eERCvIBa+bWjUX
l/jPx/ooqJmoHltrcG3w+7ztrFLZ0mu03ThNAdYMhf/HqMmi+pW/r34AXJ2bYb7CBFn80XmXmJUr
V5wVwlgd8KEDpTBzn8Y4ZbBZ6/MT6mUdw9EaUm/UOJkVqF+HHar/GfaXNnuYhL0wxNeil0qI7jJo
thEkMdfEtmxysarFV0rkswHngGXCfR5i0dB/odufl+7xTmkUxbew/Lp9Lu2V8LtlJmKstqgy/QaUl
1Y6n52GLqFD/RSad7QIF0tbt3oyiAGIgx0YYSLB+L3HcQiF11z4JxRmMexKFTVGYALRZ0Z02j7Zc
QLDTTAiTNp0lyZQIKEYpLZ8NHZCqXA2+LghUPHahHexG0FRLVH6BMSB54BxFn8NK7Veddk5jbPf8l
SgdSG0baP/iAKE00PLX5LDP+TBTQtIRmHI85q+ylbwLKuTVIflDUhmwyq8t5mLwTjIzsUSYCZFk8
dt0u87TNkRj9BJii3YBiKvasNKvGC0zRIza0JQTLarrwJUD9GAvvS97qLQ3Av+0AuJZ3vA3Z0msl
P/vYPGqxJXRdrkeCpV4I/amFtSCAMA6rnGEBmulTbkN7D5GesMuNE3LGAfB6X0n61EEZoeNp0MLU
cxCss/s6pCW90HaB7IDALo/PLLit2FCotJ3znzIkY7BtkDfBR0GmOmEwL270fRAYF6rSPU/v0Z/0
Wg09P7H4/LCbk6Lk49WPSf60TcC5YhzCHCKfDnLhyX5alWc8mDHDe7LJk+r7vDXgUYHALJt3SS0c
6l+EacWn/uionN4+nCB5b7MF0unGEU1Uj7TGKV4UieY97VqjTM8RIdu0ASDZTg1Ik4Z+fPdaX9Q/
issVAsdTl5+06ctvgbPQ56sk1BEQCAj3vjVJWwYbW07+NamhKncdr4Pi0iACGKjKt1Sju4XY
id/S
lwqbC61WLTct1UZBqCyguo99nptfXN5zXmwAeuEMsKBK3je/QIPUqZIXIY7YKLbLpN/roo0oVQwY
aSxSzcjsndiqCgiee61E1L+ZzTFB54DjxZYvyE3jWJiFszCvz8Gu06F/z/VeBiFvu56J2649UY0j
fa9Ucdj9uldwaHUApees8IYnL+SzIw6zn2qiPEFE9pLuhsnbGs6A3/ATwJTnw9FKtaC9gqi19rXT
yAqEQf50WJyKtl6o88gGgCR923+muludRRVCZCiCugAsy40jppMdBXNpg5SyY3v6E2eiH4p0+fnv
boaemrOUXXdHYaucl+byqjXQerpS5h0Dn0ZUx6r2sy3u56YobFfM/CY8XRZwd/orPiuUH3S1EFLK
iM5jfel80o0We3dVrp8fplsRVDEqJvDt9q0igUa9/keF/EXSfKvKTH94NQRed9rMetTIyFA/Jaa
CwSE46YfrjJfJeKprQpoNxQS/Mzym6whEDxLFhmbEuGAMfnwKf0o7bZq3p/m9T9WuU8mE5I/P
yuXG
E5VJZb0CNXCCD4mp+if0xqxtgEEk3mj/RK4lzQ7c0uHaf5PcTdr5Fifpxp0h0+wuCNSJmuLv
Pf3z
m8iCNXRyJtlw3VbeeFeUqTWK7f21nwyAjG1lWqjKv/GiC+9nNkdEcsjyiM8rfWFr7VP8dG8cd/85
88D9blo+QA[cloudera@quickstart ~]$
```

```
550c
129860 6l+EacWn/uionN4+nCB5b7MF0unGEU1Uj7TGKV4UieY97VqjTM8RIdu0ASDZTg1Ik4Z+fPda
X9Q/
129861 issVAsdTl5+06ctvgbPQ56sk1BEQCAj3vjVJWwYbW07+NamhKncdr4Pi0iACGKjKt1Sju4XY
id/S
129862 lwqbC61WLTct1UZBqCyguo99nptfXN5zXmwAeuEMsKBK3je/QIPUqZIXIY7YKLbLpN/roo0o
VQwY
129863 aSxSzcjsndiqCgiee61E1L+ZzTFB54DjxZYvyE3jWJiFszCvz8Gu06F/z/VeBiFvu56J2649
UY0j
129864 fa9Ucdj9uldwaHUApees8IYnL+SzIw6zn2qiPEFE9pLuhsnbGs6A3/ATwJTnw9FKtaC9gqi1
9rXT
129865 yAqEQf50WJyKtl6o88gGgCR923+muludRRVCZCiCugAsy40jppMdBXNpg5SyY3v6E2eiH4p0
+fnv
129866 boaemrOUXXdHYaucl+byqjXQerpS5h0Dn0ZUx6r2sy3u56YobFfM/CY8XRZwd/orPiuUH3S1
EFLK
129867 iM5jfel80o0We3dVrp8fplsRVDEqJvDt9q0igUa9/keF/EXSfKvKTH94NQRed9rMetTIyFA
/Jaa
129868 CwSE46YfrjJfJeKprQpoNxQS/Mzym6whEDxLFhmbEuGAMfnwKf0o7bZq3p/m9T9WuU8mE5I/P
yuXG
129869 E5VJZb0CNXCCD4mp+if0xqxtgEEk3mj/RK4lzQ7c0uHaf5PcTdr5Fifpxp0h0+wuCNSJmuLv
Pf3z
129870 m8iCNXRyJtlw3VbeeFeUqTWK7f21nwyAjG1lWqjKv/GiC+9nNkdEcsjyiM8rfWFr7VP8dG8c
d/85
```



```
SS0c
129860 6l+EacWn/uionN4+nCB5b7MF0unGEU1Uj 7TGKV4UieY97Vqj TM8RIdu0ASDZTg1Ik4
X90/
129861 issVAsdTI5+06ctvgbPQ56sk1BEQCAj3vjVJWwYbW07+NamhKncdr4Pi0iACGKjKt1
iD/S
129862 lwqbC61WlTct1UzBqCyguo99nptfXN5zXmwAeuEMsKbK3je/QIPUqZIXIY7YKLblpM
VQwY
129863 aSxSzCjsndiqCgiee61E1L+ZzTFB54DjxZYvyE3jWJiFszCvz8Gu06F/z/VeBiFvu5
UY0j
129864 fa9Ucdj9uIdwaHUAPees8IYnL+SzIw6zn2qiPEFE9pLuhsnbGs6A3/ATwJTnw9FKt2
9rXT
129865 yAqEQf50WJyKtl6o88gGgCR923+muludRRVCZCiCUgAsy40jpPmDBXNPg5SyY3v6E2
+fnv
129866 boaemr0UXXdHYaucL+byqjXQerpS5h0Dn0ZUx6r2sy3u56YobFFm/CY8XRZwd/orPi
EFLK
129867 iM5jfel8oo0We3dVrp8fplsRVDEQjJvDt9q0igUa9/keF/EXSfKvKTH94NQRed9rME
/Jaa
129868 CwSE46YfrfjJJeKprQpoNxQ5/Mzym6whEDxLFhmbEuGAMfnwKf0o7bZq3p/m9T9WuU8
yuXG
129869 E5VJZb0CNXCCD4mp+if0xqxtgEEk3mj/RK4lzQ7c0uHaf5PcTdr5Fifpxp0h0+wuCF
Pf3z
129870 m8iCNXRyjTlw3VbeeFeUqTWK7f21nwyAjG1lWqjKv/GiC+9nNkdEcsjyiM8rfWFr7A
d/85
```

```
[cloudera@quickstart ~]$ head petrov.txt
```

```
30k5vmq2Jc+w2GTbzLLCnn6Hn5D3kUE3rtFJ+B5mmnv4RW4msoX0V/55JYcjUZFWz6IdZGXS3S70
>vrRg83rd6U0yj7plDr/UjiXqmyLUsk+KdJK7/bJA9bxSPiZaGYxU2lR9Swyt8iCSzrrIHB3lqkh
/U/jvQ/6+hRoHwXoyUdUs0+B/LB7lwUNquuKn8l73fG6p/1M7dzigQzeqpYwq3aq59KHMqrLBZA8
>05TrKplonG4d6Wy/K9l8XZHKKyEkFrzQ0vdA2FRjEkoFv8sQMxeIU1L5y0lyHFRxBzm0LAEXNDz
34ZY5fyHdpD+oKHUHNnCVi77DoK+5GFTAUwdAv7Ce081JFoUPuRXGbpvL47M30m3aHA0HEY5H1BL
LJv6TjrDCd+7fEq9K/6oxWISJrpN/027wpk6vezRQN30D8F9XYIPH4aRcqcyZDaB4KLm1qSJIQuW
>ik46QunQ0FC60ZLgFK40IMLL1pM7RfJEfzYrbJZEHaECJ85Xz+U2KjafCkDohiy4jvN0vY5tvHz
/0LI4haKp6AMXPFx5Dvb5/9FTTy900nfJ/ZYmXCbc+am0e3vNv8jtHUGTUK59M7U/JFP1Vq53ypD
3iUBK0o0y9voMW3Lm2JB/3R2r0keHIVB9jyot4yFp0M5McKA5RC1z2cFfMTTHTiPft0vle5caD
3DBZH5dyle0kHn0KT3HxxttjPhpkAfk5s3CLZJ6X0I6+m9vXz7YKHVHiaxQg0+zkflh6XnirfZy/
```

```
[cloudera@quickstart ~]$ head -c 7 petrov.txt
```

```
30k5vmq[cloudera@quickstart ~]$ head -n 7 petrov.txt
```

```
30k5vmq2Jc+w2GTbzLLCnn6Hn5D3kUE3rtFJ+B5mmnv4RW4msoX0V/55JYcjUZFWz6IdZGXS3S70
>vrRg83rd6U0yj7plDr/UjiXqmyLUsk+KdJK7/bJA9bxSPiZaGYxU2lR9Swyt8iCSzrrIHB3lqkh
/U/jvQ/6+hRoHwXoyUdUs0+B/LB7lwUNquuKn8l73fG6p/1M7dzigQzeqpYwq3aq59KHMqrLBZA8
>05TrKplonG4d6Wy/K9l8XZHKKyEkFrzQ0vdA2FRjEkoFv8sQMxeIU1L5y0lyHFRxBzm0LAEXNDz
34ZY5fyHdpD+oKHUHNnCVi77DoK+5GFTAUwdAv7Ce081JFoUPuRXGbpvL47M30m3aHA0HEY5H1BL
LJv6TjrDCd+7fEq9K/6oxWISJrpN/027wpk6vezRQN30D8F9XYIPH4aRcqcyZDaB4KLm1qSJIQuW
>ik46QunQ0FC60ZLgFK40IMLL1pM7RfJEfzYrbJZEHaECJ85Xz+U2KjafCkDohiy4jvN0vY5tvHz
```

```
[cloudera@quickstart ~]$ █
```

```
[cloudera@quickstart ~]$ tail petrov.txt
```

```
UV+ae03lCwlxqhWy3humTrs9YIkphpb+9Hq0g0ZBN39QQGL9mENQNl5PE+sXs1vR4e/45TUXSDC8
i0QfDVTBhthXAC1gGLa0pX6nGfhkYf881UIbSr+mjDsuSbY2teRjBD2SAA1IIYAFgP9XslvoEH00
etNo8DZl2TnAm4cdgsaUe/HhHVDfYm8ISdWHnhXk3tKIGUKlQl/gHJsfxz0zVNjsAvftPVHYUR7A
V6K3aMecYN25UhnMjGfB0k2q1Bv0eq3JkTfenr4TG79HU10CmYoL5DQYQinSxfnUkAysWQmmpL3F
/9FKDrc/zepWq55kxVzKK4mhwqkbzjgwl7Z1EMwonR1aEwDsvjQItA7ia3G/7dXwUq1ISpGAE/S1
EZsUX0dctdC8iJg3e/YnXurFlQ494vNgFaArBVSEzYMeEXpkYt1Nq5Z2VSEUACpYQ4Mt7WlPd2T5
fx77iTsy0grs3bTpx5+w2KzDlYaxgizGa0qC8yJFCIFkwacwnJVWNIRsDdLfMEmzl4AB87EkFa3m
MWu6oWFnzQB+mdMR8Wr5ttIl23PgTocegVKHmj8lyonLqySEgbhpqe1qA8e3XmWiagXnI9qGM4bp
HadyaNY0fUHONKkseIkSv3dmdbkYy2lX8RJ+bTbm+02D+Xc1EfYrELepuAhrkw/7DQjogCByc3I6
+/Tx7CUB9l[cloudera@quickstart ~]$
```

```
[cloudera@quickstart ~]$ tail -c 7 petrov.txt
```

```
x7CUB9l[cloudera@quickstart ~]$ tail -n 7 petrov.txt
```

```
V6K3aMecYN25UhnMjGfB0k2q1Bv0eq3JkTfenr4TG79HU10CmYoL5DQYQinSxfnUkAysWQmmpL3F
/9FKDrc/zepWq55kxVzKK4mhwqkbzjgwl7Z1EMwonR1aEwDsvjQItA7ia3G/7dXwUq1ISpGAE/S1
EZsUX0dctdC8iJg3e/YnXurFlQ494vNgFaArBVSEzYMeEXpkYt1Nq5Z2VSEUACpYQ4Mt7WlPd2T5
fx77iTsy0grs3bTpx5+w2KzDlYaxgizGa0qC8yJFCIFkwacwnJVWNIRsDdLfMEmzl4AB87EkFa3m
MWu6oWFnzQB+mdMR8Wr5ttIl23PgTocegVKHmj8lyonLqySEgbhpqe1qA8e3XmWiagXnI9qGM4bp
HadyaNY0fUHONKkseIkSv3dmdbkYy2lX8RJ+bTbm+02D+Xc1EfYrELepuAhrkw/7DQjogCByc3I6
+/Tx7CUB9l[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 petrov.txt 05.03.2024_petrov.txt
[cloudera@quickstart ~]$ ls
05.03.2024_petrov.txt  Desktop  eclipse  kerberos  parcels  Pictures  Videos
cloudera-manager      Documents  enterprise-deployment.json  lib  petrov.gz  Public  workspace
cm_api.py             Downloads  express-deployment.json    Music  petrov.txt  Templates
[cloudera@quickstart ~]$ hdfs dfs -cp /user/mgpu/petrov/petrov.txt /user/mgpu/petrov/ 05.03.2024_petrov.txt
cp: `05.03.2024_petrov.txt': No such file or directory
[cloudera@quickstart ~]$ hdfs dfs -ls /user/mgpu/petrov
Found 2 items
-rw-r--r-- 1 cloudera supergroup 7599439 2024-04-13 08:39 /user/mgpu/petrov/petrov.gz
-rw-r--r-- 1 cloudera supergroup 10000000 2024-04-13 08:38 /user/mgpu/petrov/petrov.txt
[cloudera@quickstart ~]$ hdfs dfs -cp /user/mgpu/petrov/petrov.txt /user/mgpu/petrov/05.03.2024_petrov.txt
[cloudera@quickstart ~]$ hdfs dfs -ls /user/mgpu/petrov
Found 3 items
-rw-r--r-- 1 cloudera supergroup 10000000 2024-04-13 08:54 /user/mgpu/petrov/05.03.2024_petrov.txt
-rw-r--r-- 1 cloudera supergroup 7599439 2024-04-13 08:39 /user/mgpu/petrov/petrov.gz
-rw-r--r-- 1 cloudera supergroup 10000000 2024-04-13 08:38 /user/mgpu/petrov/petrov.txt
[cloudera@quickstart ~]$
```

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

```
[cloudera@quickstart ~]$ hdfs dfs -ls /user/mgpu/petrov
Found 3 items
-rw-r--r-- 1 cloudera supergroup 10000000 2024-04-13 08:54 /user/mgpu/petrov/05.03.2024_petrov.txt
-rw-r--r-- 1 cloudera supergroup 7599439 2024-04-13 08:39 /user/mgpu/petrov/petrov.gz
-rw-r--r-- 1 cloudera supergroup 10000000 2024-04-13 08:38 /user/mgpu/petrov/petrov.txt
[cloudera@quickstart ~]$
```

3.1.11. Удалить поддиректорию /fio со всем содержимым.

```
[cloudera@quickstart ~]$ hdfs dfs -rm -R /user/mgpu/petrov
Deleted /user/mgpu/petrov
[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_petrov_SV.txt
[cloudera@quickstart ~]$ hdfs dfs -copyFromLocal 3.1.12_petrov_SV.txt /user/mgpu/petrov
[cloudera@quickstart ~]$ yarn jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar wordcount /user/mgpu/petrov/3.1.12_petrov_SV.txt /user/mgpu/petrov/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_1711484710610_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
```



3.2. Создание таблицы в Hive

1. Скачать [датасет](#) или [тут](#)

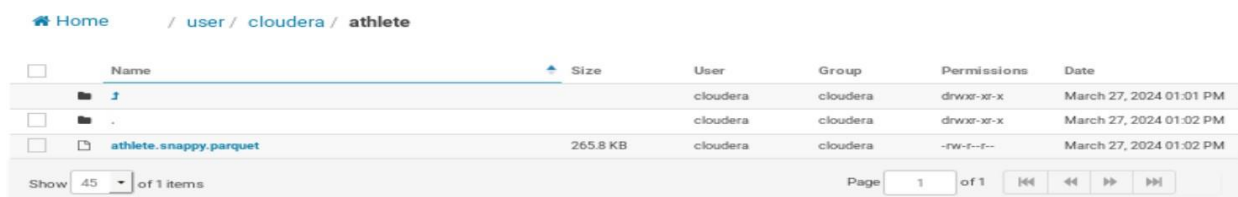
```
[cloudera@quickstart ~]$ wget https://github.com/BosenkoTM/cloudera-quickstart/blob/main/data/athlete.snappy.parquet
--2024-04-13 09:25:17-- https://github.com/BosenkoTM/cloudera-quickstart/blob/main/data/athlete.snappy.parquet
Resolving github.com... 140.82.121.4
Connecting to github.com|140.82.121.4|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: unspecified [text/html]
Saving to: "athlete.snappy.parquet"
```

```
[ <=> ] 148,633 --.-K/s in 0.1s
```

```
2024-04-13 09:25:18 (1.31 MB/s) - "athlete.snappy.parquet" saved [148633]
```

```
[cloudera@quickstart ~]$ ls
05.03.2024_petrov.txt  cm_api.py  eclipse  lib  petrov.txt  Videos
3.1.12_petrov_SV.txt  Desktop    enterprise-deployment.json  Music  Pictures  workspace
athlete.snappy.parquet Documents  express-deployment.json    parcels  Public
cloudera-manager      Downloads  kerberos  petrov.gz  Templates
```

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**.



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

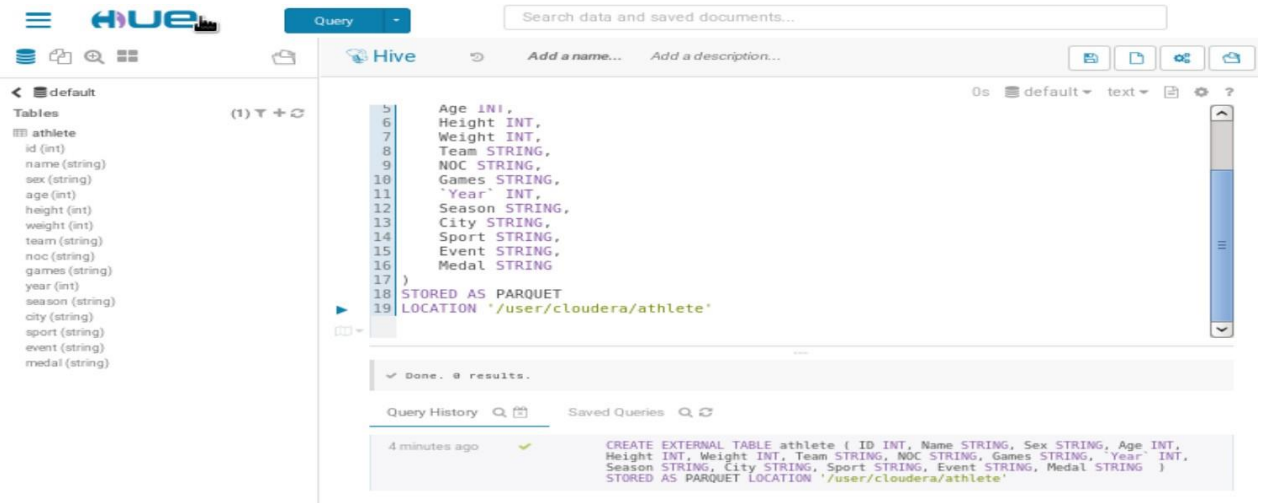
```
CREATE EXTERNAL TABLE athlete (
  ID INT,
  Name STRING,
  Sex STRING,
  Age INT,
  Height INT,
  Weight INT,
```



```

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.

| <input type="checkbox"/> | Name | Size | User | Group | Permissions | Date |
|--------------------------|--------------------------------------|---------|----------|----------|-------------|-------------------------|
| <input type="checkbox"/> | ↑ | | 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"/> | 201402_trip_data.csv | 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.

```

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'

```

✓ Success.

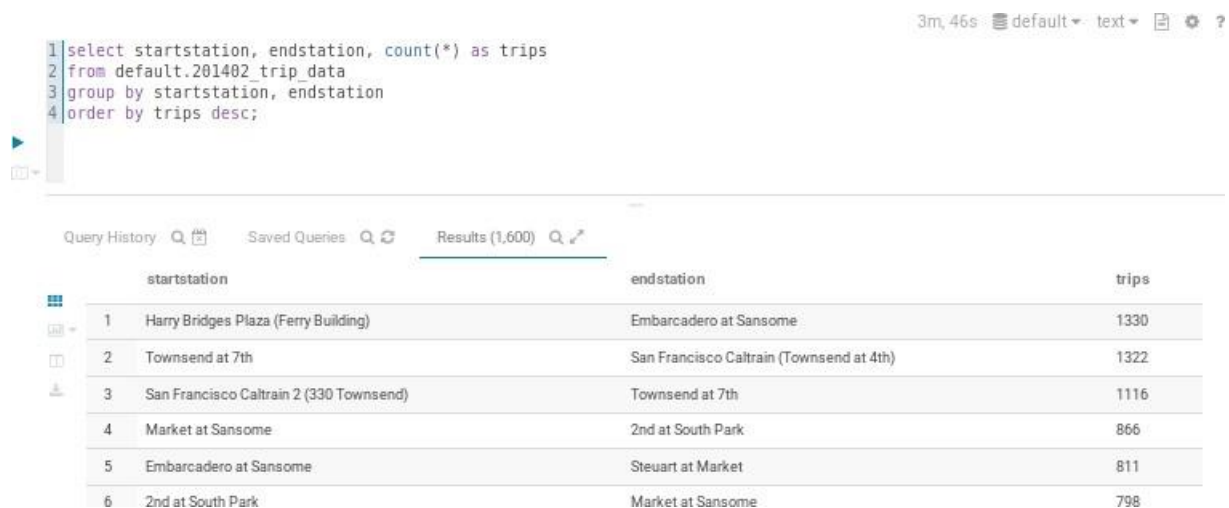
выполнить запрос

```

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

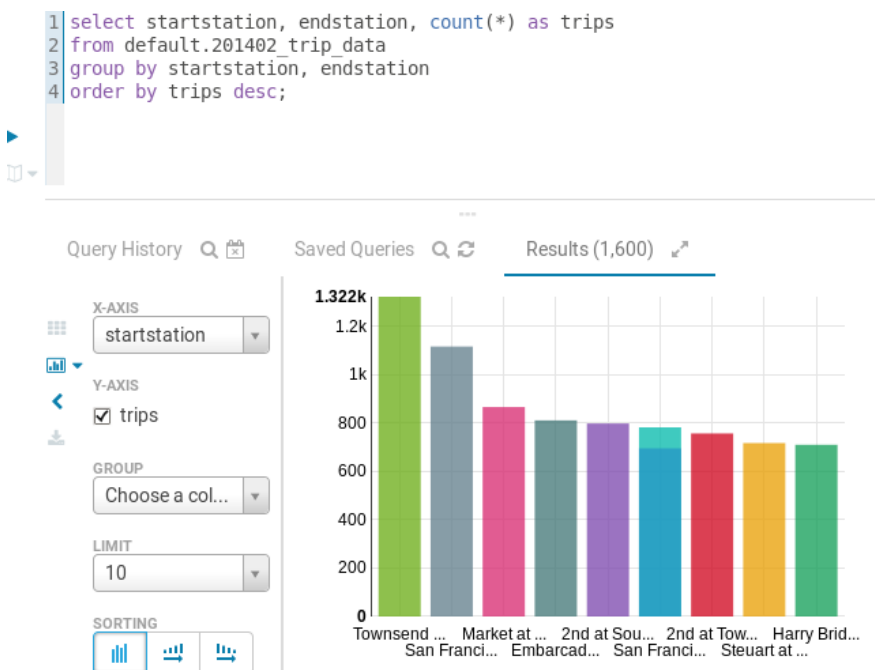
```

| 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 | | | | | 0 |
| 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 | Memory Reserved | | | |
| 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 VCo | Reserve | |
| application_1711484710610_0007 | cloudera | select startstation, endstation, count...desc(Stage-2) | MAPREDUCE | root.cloudera | Thu Mar 28 13:23:41 -0700 2024 | Thu Mar 28 13:24:42 -0700 2024 | FINISHED | SUCCEEDED | N/A | N/A | | N |



Создать гистограмму, щелкнув значок «Hue Bar»:

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



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 |
| 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