

In-class Lab 2 – Apache Hive Commands

1. Load the large stocks dataset (400 MB) into HDFS and use the dataset to create a Hive table using the location attribute.

```
goodnewsagbadu@lab2-m:~$ wget https://www.dropbox.com/s/ia779edcjcftd84/stocks
--2022-10-29 03:20:15-- https://www.dropbox.com/s/ia779edcjcftd84/stocks
Resolving www.dropbox.com (www.dropbox.com)... 162.125.3.18, 2620:100:601b:18::a27d:812
Connecting to www.dropbox.com (www.dropbox.com) [162.125.3.18]:443... connected.
HTTP request sent, awaiting response... 302 Found
Location: /s/raw/ia779edcjcftd84/stocks [following]
--2022-10-29 03:20:16-- https://www.dropbox.com/s/raw/ia779edcjcftd84/stocks
Reusing existing connection to www.dropbox.com:443.
HTTP request sent, awaiting response... 302 Found
Location: https://ucffb12c23acf3e69710f18a5f19.d1.dropboxusercontent.com/cd/0/inline/BvthXxV6vmxWQw3FswV81S4cA9CZ8G85xo0bLTVWsdpJnh2KiKE31G9baC26PhERya7oarsp2hMXwSsJd1Z6Y6dS_X4dQ1Bb3ZrUG1mAH9FBE01ryW1EO44fXb9u2Db8o2x1_Za2usPuEtJ_2dU9uNXzbTJ-8AZ0-MclS6QyQnCA/file# [following]
--2022-10-29 03:20:16-- https://ucffb12c23acf3e69710f18a5f19.d1.dropboxusercontent.com/cd/0/inline/BvthXxV6vmxWQw3FswV81S4cA9CZ8G85xo0bLTVWsdpJnh2KiKE31G9baC26PhERya7oarsp2hMXwSsJd1Z6Y6dS_X4dQ1Bb3ZrUG1mAH9FBE01ryW1EO44fXb9u2Db8o2x1_Za2usPuEtJ_2dU9uNXzbTJ-8AZ0-MclS6QyQnCA/file
Resolving ucffb12c23acf3e69710f18a5f19.d1.dropboxusercontent.com (ucffb12c23acf3e69710f18a5f19.d1.dropboxusercontent.com)... 162.125.3.15, 2620:100:601b:15::a27d:80f
Connecting to ucffb12c23acf3e69710f18a5f19.d1.dropboxusercontent.com (ucffb12c23acf3e69710f18a5f19.d1.dropboxusercontent.com) [162.125.3.15]:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 428223209 (408M) [text/plain]
Saving to: 'stocks'

stocks                               100%[=====] 408.38M   121MB/s   in 3.5s

2022-10-29 03:20:20 (118 MB/s) - 'stocks' saved [428223209/428223209]

goodnewsagbadu@lab2-m:~$ head stocks
ABCSE,B7J,2010-02-08,8.63,8.70,8.57,8.64,78900,8.64
ABCSE,B7J,2010-02-05,8.63,8.71,8.31,8.58,218700,8.58
ABCSE,B7J,2010-02-04,8.88,8.88,8.59,8.66,89900,8.66
ABCSE,B7J,2010-02-03,8.83,8.92,8.80,8.89,119000,8.89
ABCSE,B7J,2010-02-02,8.77,8.90,8.73,8.87,51900,8.87
ABCSE,B7J,2010-02-01,8.69,8.77,8.66,8.75,38600,8.75
ABCSE,B7J,2010-01-29,8.81,8.81,8.56,8.57,91700,8.57
ABCSE,B7J,2010-01-28,8.90,8.90,8.60,8.69,92100,8.69
ABCSE,B7J,2010-01-27,8.87,8.87,8.68,8.79,82400,8.79
ABCSE,B7J,2010-01-26,8.83,8.92,8.71,8.82,106000,8.82
goodnewsagbadu@lab2-m:~$
```

```
goodnewsagbadu@lab2-m:~$ hadoop fs -mkdir /Lab2
goodnewsagbadu@lab2-m:~$ hadoop fs -copyFromLocal /home/goodnewsagbadu/stocks /Lab2/.
goodnewsagbadu@lab2-m:~$ hadoop fs -ls /Lab2
Found 1 items
-rw-r--r-- 1 goodnewsagbadu hadoop 428223209 2022-10-29 03:39 /Lab2/stocks
goodnewsagbadu@lab2-m:~$
```

```
hive> CREATE EXTERNAL TABLE IF NOT EXISTS stocks (
>   exch STRING,
>   symbol STRING,
>   ymd STRING,
>   price_open FLOAT,
>   price_high FLOAT,
>   price_low FLOAT,
>   price_close FLOAT,
>   volume INT,
>   price_adj_close FLOAT)
> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
> LOCATION '/Lab2/';

OK
Time taken: 0.738 seconds
hive> select * from stocks limit 5;
Query ID = goodnewsagbadu_20221029035549_e4947548-849a-479d-ae16-633380643e00
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1667012789420_0002)
```

	VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	5	5	0	0	0	0
VERTICES: 01/01 [=====] 100% ELAPSED TIME: 9.98 s									
OK									
	ABCSE	QIJ	2007-10-12	92.82	93.56	92.69	93.22	123700	93.22
	ABCSE	QIJ	2007-10-11	94.87	95.47	92.51	93.31	113600	93.31
	ABCSE	QIJ	2007-10-10	93.44	94.74	93.08	94.3	84600	94.3
	ABCSE	QIJ	2007-10-09	93.66	94.23	93.41	94.23	75100	94.23
	ABCSE	QIJ	2007-10-08	94.65	94.65	93.53	93.82	71700	93.82
Time taken: 19.913 seconds, Fetched: 5 row(s)									
hive>									

LAB 2 GOODNEWS AGBADU

2. Write a SQL command to find the stocks with average daily volume larger than 1 million shares

```
hive> select symbol,AVG(volume) as avg_vol from stocks group by symbol having avg_vol > 1000000 limit 10;
Query ID = goodnewsagbadu_20221103170552_92b66082-60f7-4c95-9c23-5a06494bd056
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1667493536235_0003)
```

	VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	5	5	0	0	0	0
Reducer 2	container	SUCCEEDED	17	17	0	0	0	0

```
VERTICES: 02/02 [=====>>>] 100% ELAPSED TIME: 17.27 s
```

OK	
B7J	3908197.3942499785
BBX	1806045.768621236
BVB	1161540.341160833
GG7	1825173.7040358745
GLJ	1162483.5794447726
II7	1825173.7040358745
K9J	1655103.6090225563
K9K	4564766.409365754
KE	1262947.1479891352
OOO	1971458.5358624777

```
Time taken: 25.658 seconds, Fetched: 10 row(s)
hive>
```

3. Write a Hive query to find the top 3 stocks by volume for the year 2004.

```
hive> select symbol, volume from stocks where ymd like '2004%' order by volume DESC limit 3;
Query ID = goodnewsagbadu_20221029043003_315c1160-11a6-4015-b777-f8ea2a28ced8
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1667012789420_0004)
```

	VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	5	5	0	0	0	0
Reducer 2	container	SUCCEEDED	1	1	0	0	0	0

```
VERTICES: 02/02 [=====>>>] 100% ELAPSED TIME: 12.40 s
```

OK	
IRC	145015500
GRC	145015500
BRC	145015500

```
Time taken: 13.155 seconds, Fetched: 3 row(s)
hive>
```

4. Write a Hive query to find the top 3 stocks by volume and whose symbol start with the first letter of your name (example for Saber, it is symbols starting with “S”).

```
hive> select symbol, volume from stocks where symbol like 'G%' order by volume DESC limit 3;
Query ID = goodnewsagbadu_20221029212200_acaf6f32-46a0-426e-8c21-047a1d814ce7
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1667012789420_0016)
```

	VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	5	5	0	0	0	0
Reducer 2	container	SUCCEEDED	1	1	0	0	0	0

```
VERTICES: 02/02 [=====>>] 100% ELAPSED TIME: 14.11 s
```

```
OK
GK      329786100
GK      321561400
GK      206852900
Time taken: 22.662 seconds, Fetched: 3 row(s)
hive>
```

5. Write a Hive query to find all the stocks symbols whose closing price is larger than your age.

```
hive> select symbol,price_close from stocks where price_close > 29 limit 5;
Query ID = goodnewsagbadu_20221029192028_5a5599ad-db20-4b76-8752-43a5f3527da7
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1667012789420_0009)
```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	5	5	0	0	0	0

```
VERTICES: 01/01 [=====>>] 100% ELAPSED TIME: 10.08 s
OK
QIJ      93.22
QIJ      93.31
QIJ      94.3
QIJ      94.23
QIJ      93.82
Time taken: 10.833 seconds, Fetched: 5 row(s)
hive>
```

6. Write a Hive query to find the top 10 stocks with largest intraday price change (difference between high and low price during a trading day) and also display the amount of the change.

```
hive> select symbol, price_high - price_low AS change from stocks order by change DESC LIMIT 10;
Query ID = goodnewsagbadu_20221103173632_4a52e1c3-0e12-4b36-b158-4755a4a05b93
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1667493536235_0004)
```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	5	5	0	0	0	0
Reducer 2	container	SUCCEEDED	1	1	0	0	0	0

```
VERTICES: 02/02 [=====>>] 100% ELAPSED TIME: 17.36 s
OK
ZBR      108.34
HBR      108.34
BBR      108.34
QBR      108.34
WBR      108.34
BCL      99.899994
ICL      99.899994
GCL      99.899994
OCL      99.899994
QBR      82.34
Time taken: 26.217 seconds, Fetched: 10 row(s)
hive>
```