1) Start Hadoop ecosystem

- Start Hadoop at the Terminal
  > start-all.sh

- Verfify Hadoop distribution is running – go to localhost:8088 or type jps
  > jps

```
hadoopuser@lainy-virtualmachine:~$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [localhost]
localhost: starting namenode, logging to /usr/local/hadoop-2.7.7/logs/hadoop-hadoopuser-namenode-lai
ny-virtualmachine.out
localhost: starting datanode, logging to /usr/local/hadoop-2.7.7/logs/hadoop-hadoopuser-datanode-lai
ny-virtualmachine.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop-2.7.7/logs/hadoop-hadoopuser-secon
darynamenode-lainy-virtualmachine.out
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hadoopuser-resourcemanager-lainy-vi
rtualmachine.out
localhost: starting nodemanager, logging to /usr/local/hadoop-2.7.7/logs/yarn-hadoopuser-nodemanager
-lainy-virtualmachine.out
hadoopuser@lainy-virtualmachine:~$ jps
4768 SecondaryNameNode
5412 Jps
4932 ResourceManager
5092 NodeManager
4360 NameNode
4523 DataNode
hadoopuser@lainy-virtualmachine:~$
```

2) Store data into hive data warehouse

- Start Hive at the Terminal
  > hive

- Select database 'crudeoil_predict'
  > show databases;
  > use crudeoil_predict;

- Show available tables
  > show tables;

```
hive> show tables;
OK
news_klse
news_raw
oilprices
testing
Time taken: 0.066 seconds, Fetched: 4 row(s)
hive>
```

== 1st Table ==

- Create table 'oilprices' in 'crudeoil_predict' database
  > CREATE TABLE oilprices IF NOT EXISTS (DateOil string, Open_Price double, High_Price double, Low_Price double, Close_Price double, Adjusted_Price double, Volume int)

> COMMENT 'crude oil prices from year 2000'

> ROW FORMAT DELIMITED

> FIELDS TERMINATED BY ','

> LINES TERMINATED BY '\n'

> TBLPROPERTIES ("skip.header.line.count"="1");

```
hive> CREATE TABLE oilprices (DateOil string, Open_Price double, High_Price double, Low_Price double
, Close_Price double, Adjusted_Price double, Volume int)
    > COMMENT 'crude oil prices from year 2000'
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ','
    > LINES TERMINATED BY '\n'
    > TBLPROPERTIES ("skip.header.line.count"="1");
OK
Time taken: 0.476 seconds
hive>
```

- Load data CrudeoilPricesdata.csv from local drive to hive table 'oilprices'

  > LOAD DATA LOCAL INPATH '/home/hadoopuser/

  CrudeoilPricesdata.csv'

  > OVERWRITE INTO TABLE oilprices;

```
hive> LOAD DATA LOCAL INPATH '/home/hadoopuser/CrudeoilPricesdata.csv'
    > OVERWRITE INTO TABLE oilprices;
Loading data to table oilprice_predict.oilprices
Table oilprice_predict.oilprices stats: [numFiles=1, numRows=0, totalSize=332037, rawDataSize=0]
OK
Time taken: 2.004 seconds
hive>
```

- View attribute in oilprices table

  > DESCRIBE oilprices;

```
hive> describe oilprices;
OK
dateoil                 string
open_price              double
high_price              double
low_price               double
close_price             double
adjusted_price          double
volume                  int
Time taken: 0.336 seconds, Fetched: 7 row(s)
hive>
```

- To check availability of data, we do simple data query. Query 5 rows
  from the table oilprices

  > SELECT * FROM oilprices

  > LIMIT 5;

```
hive> SELECT * FROM oilprices
    > LIMIT 5;
OK
2000-03-22      27.65   28.25   27.25   27.459999       27.459999       92302
2000-03-23      27.65   27.780001       27.16   27.309999       27.309999       79373
2000-03-24      27.85   28.15   27.549999       27.98   27.98   55693
2000-03-26      NULL    NULL    NULL    NULL    NULL    NULL
2000-03-27      27.65   28.02   27.41   27.73   27.73   59199
Time taken: 0.258 seconds, Fetched: 5 row(s)
hive>
```

== 2nd Table ==

- Create table 'news_raw' in 'crudeoil_predict' database
  > CREATE TABLE news_raw IF NOT EXISTS (DateNews string, Year string, Time string, news_headline string)
  > COMMENT 'news article from market insider'
  > ROW FORMAT DELIMITED
  > FIELDS TERMINATED BY ','
  > LINES TERMINATED BY '\n'
  > TBLPROPERTIES ("skip.header.line.count"="1");

- Load data news_raw.csv from local drive to hive table 'news_raw'
  > LOAD DATA LOCAL INPATH '/home/hadoopuser/news_raw.csv'
  > OVERWRITE INTO TABLE news_raw;

- View attribute in news_raw table
  > DESCRIBE news_raw;

```
hive> describe news_raw;
OK
datenews                string
news_headline           string
Time taken: 0.249 seconds, Fetched: 2 row(s)
```

- To check availability of the table, we do simple data query. Query 5 rows from the table news_raw
  > SELECT * FROM news_raw
  > LIMIT 5;

```
hive> SELECT * FROM news_raw
    > LIMIT 5;
OK
3/21/2019       The Price Of Oil And The Price Of Oil Stocks: New Lows
3/20/2020       Canada oil sands work camp reports coronavirus infection
3/20/2020       Physical gold squeezed further; Royal Canadian Mint shuts down production for two we
eks
3/20/2020       Formula milk sells for triple usual price online
3/20/2020       Qatar eases exit restrictions on migrant employees in the oil and gas industry
Time taken: 0.557 seconds, Fetched: 5 row(s)
hive>
```

== 3rd Table ==

- Create table 'news_klse' in 'crudeoil_predict' database
  > CREATE TABLE news_klse IF NOT EXISTS (DateNews string, Year string, Time string, news_headline string)
  > COMMENT 'news from KLSE'
  > ROW FORMAT DELIMITED
  > FIELDS TERMINATED BY ','
  > LINES TERMINATED BY '\n'
  > TBLPROPERTIES ("skip.header.line.count"="1");

- Load data news_klse.csv from local drive to hive table 'news_klse'
  > LOAD DATA LOCAL INPATH '/home/hadoopuser/ news_klse.csv'
  > OVERWRITE INTO TABLE news_klse;

- View attribute in news_klse table
  > DESCRIBE news_klse

```
hive> describe news_klse;
OK
datenews            string
year                string
time                string
news_headline       string
Time taken: 0.263 seconds, Fetched: 4 row(s)
hive>
```

- To check availability of the table, we can do simple data query. Query 5 rows from the table news_klse
  > SELECT * FROM news_klse
  > LIMIT 5;

```
hive> SELECT * FROM news_klse
    > LIMIT 5;
OK
0       "TheEdge16 Mar    2020 08:52am - 1 hour ago"     "CGS-CIMB Research cuts KLCI end-2020 target
 to 1
1       "TheEdge16 Mar    2020 08:52am - 1 hour ago"     "CGS-CIMB Research cuts KLCI end-2020 target
 to 1
2       "TheEdge13 Mar    2020 23:05pm - 2 days ago"     "Malls
3       "TheEdge13 Mar    2020 09:10am - 3 days ago"     Govt to maintain RM21.6b budget for NFCP
4       "TheStar11 Mar    2020 09:23am - 5 days ago"     "Bursa ekes out slight gains on Sime Plantat
ion
Time taken: 0.218 seconds, Fetched: 5 row(s)
hive>
```

=== THE END ===