

Code or Steps for Data Ingest

gyo202, sh4023, hk2874

Link to the Dataset: <https://tradingeconomics.com/>

We extracted Taiwan and Sweden's historical consumer_confidence, gdp, consumer spending data between the year 1980 and 2020, downloaded on our Desktop.

historical_country_Taiwan_indicator_Consumer_Confidence.csv

historical_country_Taiwan_indicator_GDP.csv

historical_country_Taiwan_indicator_Consumer_Spending.csv

historical_country_Sweden_indicator_GDP.csv

historical_country_Sweden_indicator_Consumer_Confidence.csv

historical_country_Sweden_indicator_Consumer_Spending.csv

- 1) We first created a directory for data sets we have found.

`hdfs dfs -mkdir project1`

```
[[hk2874@login-1-1 ~]$ hdfs dfs -ls
Found 10 items
drwxrwx---+ - hk2874 users          0 2020-10-12 14:00 .Trash
drwx-----+ - hk2874 users          0 2020-10-11 13:30 .staging
drwxr-xr-x+ - hk2874 users          0 2020-09-16 05:35 class1
drwxr-xr-x+ - hk2874 users          0 2020-10-27 01:00 hiveInput
drwxr-xr-x+ - hk2874 users          0 2020-09-21 12:27 hw3
drwxr-xr-x+ - hk2874 users          0 2020-09-21 12:45 hw3_noMapReduce
drwxr-xr-x+ - hk2874 users          0 2020-10-11 13:30 hw4a
drwxr-xr-x+ - hk2874 users          0 2020-10-11 11:59 hw4aa
drwxr-xr-x+ - hk2874 users          0 2020-10-27 00:37 impalaInput
drwxrwxr-x+ - hk2874 users          0 2020-11-08 03:02 project1
```

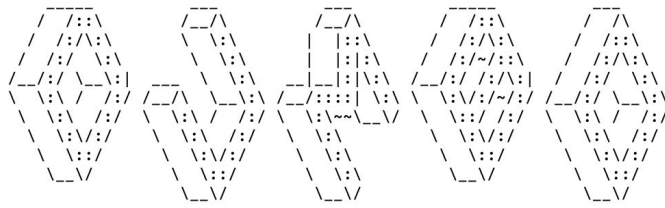
`> /user/hk2874/project 1`

- 2) From our desktop, using scp, sent all the files like below to home directory

`scp historical_country_Sweden_indicator_GDP.csv`

`hk2874@dumbo.es.its.nyu.edu:/home/hk2874`

- 3) For six files in dumbo cluster, moved to hdfs project1 directory



```
*****
https://wikis.nyu.edu/display/NYUHPC/Clusters+-+Dumbo
*****
```

```
hk2874@login-1-1 ~]$ ls
IB_NYC_2019.csv
hadoop-3.1.4-src
hadoop-3.1.4-src.tar.gz
historical_country_Sweden_indicator_Consumer_Confidence.csv
historical_country_Sweden_indicator_Consumer_Spending.csv
historical_country_Sweden_indicator_GDP.csv
historical_country_Taiwan_indicator_Consumer_Confidence.csv
historical_country_Taiwan_indicator_Consumer_Spending.csv
historical_country_Taiwan_indicator_GDP.csv
```

```
[hk2874@login-1-1 ~]$ hdfs dfs -put historical_country_Sweden_indicator_GDP.csv project1
[hk2874@login-1-1 ~]$ hdfs dfs -put historical_country_Sweden_indicator_Consumer_Spending.csv
project1
[hk2874@login-1-1 ~]$ hdfs dfs -put historical_country_Sweden_indicator_Consumer_Confidence.c
sv project1
[hk2874@login-1-1 ~]$ hdfs dfs -put historical_country_Taiwan_indicator_GDP.csv project1
[hk2874@login-1-1 ~]$ hdfs dfs -put historical_country_Taiwan_indicator_Consumer_Spending.csv
project1
[hk2874@login-1-1 ~]$ hdfs dfs -put historical_country_Taiwan_indicator_Consumer_Confidence.c
sv project1
```

4) Shared HDFS files with team members using hdfs dfs -setfacl command

```
[hk2874@login-1-1 ~]$ hdfs dfs -setfacl -R -m user:gyo202:rwX /user/hk2874/project1 ]
[hk2874@login-1-1 ~]$ hdfs dfs -setfacl -R -m default:user:sh2023:rwX /user/hk2874/project1 ]
[hk2874@login-1-1 ~]$ hdfs dfs -setfacl -R -m default:user:sh4023:rwX /user/hk2874/project1 ]
[hk2874@login-1-1 ~]$ hdfs dfs -setfacl -R -m user:sh4023:rwX /user/hk2874/project1 ]
[hk2874@login-1-1 ~]$ hdfs dfs -getfacl /user/hk2874/project1 ]
# file: /user/hk2874/project1
# owner: hk2874
# group: users
user::rwX
user:gyo202:rwX
user:hive:rwX
user:impala:rwX
user:sh4023:rwX
group::r-x
mask::rwX
other::r-x
default:user::rwX
default:user:gyo202:rwX
default:user:hive:rwX
default:user:hk2874:rwX
default:user:impala:rwX
default:user:sh2023:rwX
default:user:sh4023:rwX
default:group::r-x
default:mask::rwX
default:other::r-x
```

5) Teammates confirmed the access to data and hdfs file like this.

```
[sh4023@login-2-1 ~]$ hdfs dfs -ls /user/hk2874/project1
Found 6 items
-rw-rw-r--+ 3 hk2874 users 29564 2020-11-08 03:27
/user/hk2874/project1/historical_country_Sweden_indicator_Consumer_Confidence.csv
-rw-rw-r--+ 3 hk2874 users 15243 2020-11-08 03:26
/user/hk2874/project1/historical_country_Sweden_indicator_Consumer_Spending.csv
-rw-rw-r--+ 3 hk2874 users 4550 2020-11-08 03:26
/user/hk2874/project1/historical_country_Sweden_indicator_GDP.csv
-rw-rw-r--+ 3 hk2874 users 24158 2020-11-08 03:28
/user/hk2874/project1/historical_country_Taiwan_indicator_Consumer_Confidence.csv
-rw-rw-r--+ 3 hk2874 users 15352 2020-11-08 03:27
/user/hk2874/project1/historical_country_Taiwan_indicator_Consumer_Spending.csv
-rw-rw-r--+ 3 hk2874 users 3100 2020-11-08 03:27
/user/hk2874/project1/historical_country_Taiwan_indicator_GDP.csv
```