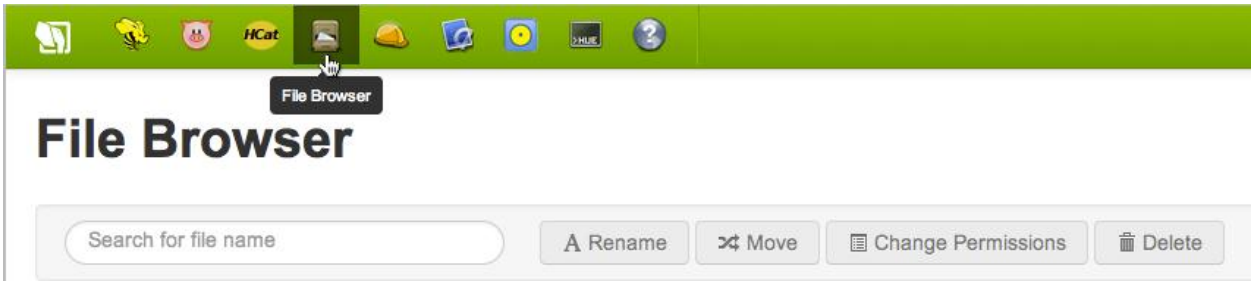
PROJECT REPORT

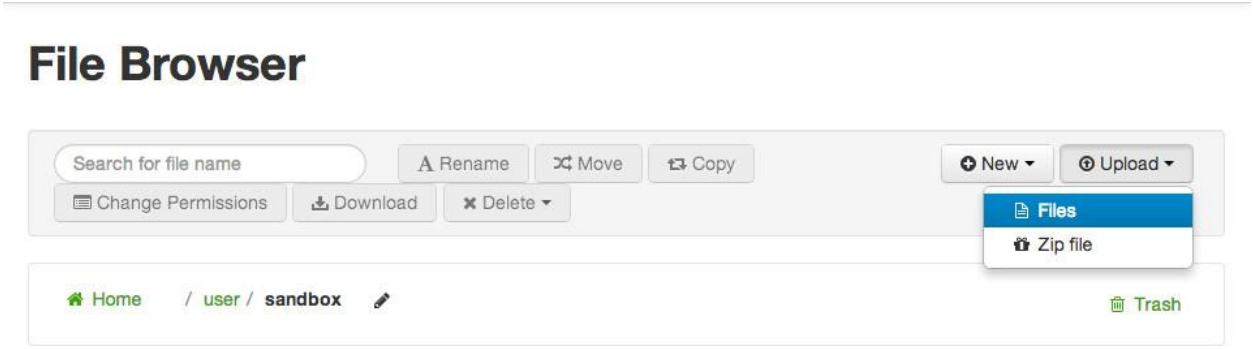
PIG DATA ANALYSIS

Data analysis to be performed after uploading the .csv file provided.

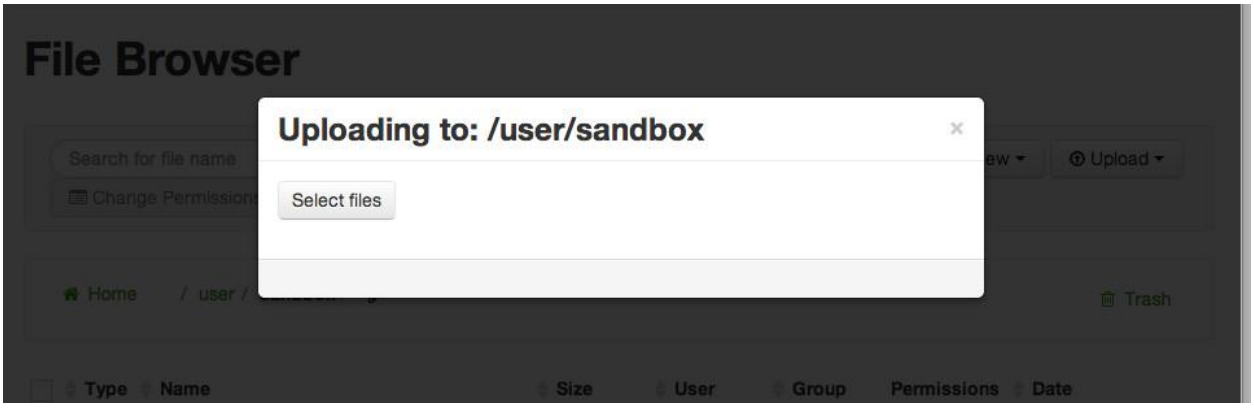
Upload the data files



Click **Upload** and select **Files** to load data to HDFS:



This action brings up the following dialog box:



Viewing a set of data

```
STOCK_A  =  LOAD '<hdfs_path>/<filename>.csv' using PigStorage(',')
AS (exchange:chararray, stock_symbol:chararray, date:chararray, stock_price_open:float,
stock_price_high:float, stock_price_low:float, stock_price_close:float, stock_volume:int,
stock_price_adj_close:float);
B = LIMIT STOCK_A 10;
Dump B;
```

Select specific Columns

```
STOCK_A  =  LOAD '<hdfs_path>/<filename>.csv' using PigStorage(',')
AS (exchange:chararray, stock_symbol:chararray, date:chararray, stock_price_open:float,
stock_price_high:float, stock_price_low:float, stock_price_close:float, stock_volume:int,
stock_price_adj_close:float);
B = LIMIT STOCK_A 10;
C = FOREACH B GENERATE stock_symbol, date, stock_price_adj_close;
DUMP C;
```

Perform Join

```
STOCK_A  =  LOAD '<hdfs_path>/<filename>.csv' using PigStorage(',')
AS (exchange:chararray, stock_symbol:chararray, date:chararray, stock_price_open:float,
stock_price_high:float, stock_price_low:float, stock_price_close:float, stock_volume:int,
stock_price_adj_close:float);
DIV_A = LOAD '<hdfs_path>/<filename>.csv' using PigStorage(',')
AS (exchange:chararray, stock_symbol:chararray, date:chararray, dividends:float);
C = JOIN STOCK_A BY (stock_symbol, date),  DIV_A BY (stock_symbol, date);
DUMP C;
```

Sort data using Order By

```
DIV_A = LOAD '<hdfs_path>/<filename>.csv' using PigStorage(',')
AS (exchange:chararray, stock_symbol:chararray, date:chararray, dividends:float);
B = ORDER DIV_A BY stock_symbol, date asc;
DUMP B;
```