Airflow Project :

Task 1: dummy operator to start task

Task 2: run spark job by reading csv file from /root/airflow/inputfiles/timberland\_stock.csv and send output file to /root/airflow/outputfiles/

Task 3: dummy operator to end task

Dependency:

Task 3 is dependent on task 2 and task 2 is dependent on task 1.

Dag Name: Timberland stock analysis

Timberland\_stock.csv:

**What day had the Peak High in Price?**

#### What is the mean of the Close column?

#### What is the max and min of the Volume column?

#### How many days was the Close lower than 60 dollars?

#### What percentage of the time was the High greater than 80 dollars ?

#### What is the Pearson correlation between High and Volume?

#### What is the max High per year?

#### What is the average Close for each Calendar Month?