

Jiahui Tang

2.3. Stock Summary (10 points)

Write a Spark script to calculate the average daily stock price at close of Alphabet Inc. (GOOG) per year since 2009. The output should be the year and the average price. You can use the `GOOGLE.csv` file from Problem 1.3.

You may need to use **DataFrames** to simplify the processing of the data. A DataFrame is a Dataset organized into named columns. It is conceptually equivalent to a table in a relational database or a data frame in R/Python, but with richer optimizations under the hood.

```
$ spark-submit P23_spark.py
```

Submission

- `P23_spark.py`: Spark script
- `P23.pdf`: The command line that you used to execute the job and any information required to reproduce the execution

Command Used:

```
spark-submit P23_spark.py
```

Output:

```
Year,avg(Close)
2016,743.4867066984126
2012,319.3312970280001
2017,921.7808373107566
2014,558.270557611111
2013,439.26437728571443
2009,218.42353286904762
2018,1109.197705057143
2011,282.6487258015873
2015,601.5505467777779
2010,266.0809473849206
```

In []: