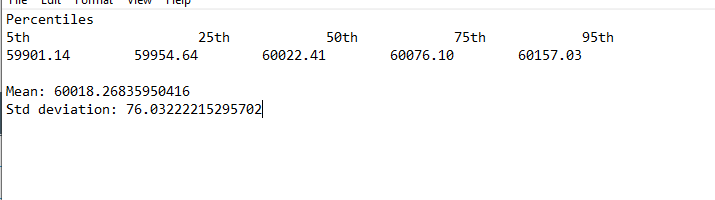
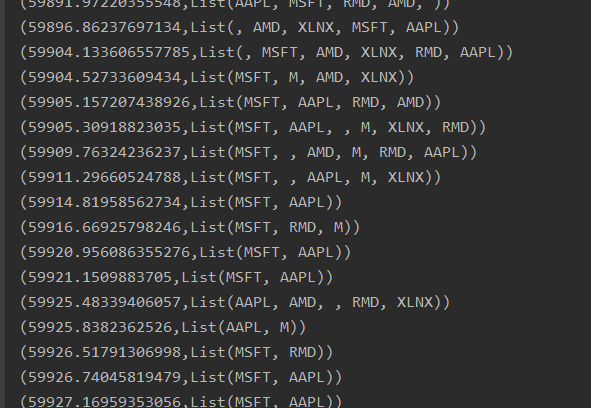
Aws Spark

For this program we created a spark program to run multiple simulations in parallel and collect data. The simulations consisted of buying and selling stocks randomly based on if there was a significant increase in price. We used the RDD structures to support running in parallel as well as to broadcast variables to all the child nodes. The simulations allow customizations on how many days you would like to run the program and how many simulations you would like to run in parallel. After all simulations finish we sort them by the total funds of all the stocks that were in the simulation. With enough simulations this gives us a nice normal curve that we can extract more data from. Then from here we can take percentiles to show you how the bottom 5% and top 5% performed as well as the 50th percentile which shows on average what should happen. There were also built in functions for calculating mean and standard deviation and those were included in the output as well.



Here is a sample of what the simulation did left is the total value of all stocks in the list. The list is the stocks in that simulation.



All of the simulations start with the same stocks and funding and as events occur sell and buy other ones.

The data was collected using a simple python script at the alpha vantage api here is how the data gets downloaded when the script completes.

