Data Engineering Assignment (pySpark)

The file `stock_prices.csv` contains the daily closing price of a few stocks on the NYSE/NASDAQ.

Load up `stock_prices.csv` as a *DataFrame*, *Dataset,* or *RDD* and complete the assignment objectives.

Assumptions

- Use only the closing price to determine returns.
- If a price is missing on a given date, you can compute returns from the closest available date
- Return can be trivially computed as the % difference of two prices

- Each result of a question should save as a separate file.

Objectives

1. **Compute the average daily return of all stocks for every date.**

   Print the results to the screen

date	average_return
yyyy-MM-dd	return of all stocks on that date

2. **Which stock was traded most frequently - as measured by closing price * volume - on average?**

ticker	frequency

3. **Which stock was the most volatile as measured by the annualized standard deviation of daily returns?**

ticker	standard deviation

4. **What were the top three 30-day return dates (% increase in closing price as compared to the closing price 30 days prior) — present ticker and date combinations**

AWS Deployment

In Vi, we take the approach of managing and provisioning cloud infrastructure through definition files (IAC Infrastructure as Code).

(Bonus: Please use AWS Cloudformation to deploy the code you created and add the following)

If your not familiar with Cloudformation Use the aws UI for the following

1. Create a glue job.
2. The code should save the results into S3 Bucket.
3. Create Glue Catalog Table mapping for each result file.

4. Results should be queryable from Athena.

AWS- credentials:

Console URL	https://249751718460.signin.aws.amazon.com/console
User	guest
Password	1q2w3e4r5t6y!
Access key	AKIATUJS4ZI6DMB7KRGA
Secret Access key	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

Remarks

1. If there are permission issues with reading/writing from/to S3 buckets, it is recommended to add the following prefix to the bucket's name: "aws-glue". For example name the bucket: "aws-glue-home-assignment"
2. To see the logs, please use the "legacy" job UI

Git Deployment

After completing the home assignment, please share your Github repo with the project's code.