Assignment-1

APIs - (Part 1 of 2 parts)

Sri Krishnamurthy, analyticsneu@gmail.com

Due: April 2nd 2021

Summary:

Datalytics Inc. is interested in monetizing it's data and making it's data available as an API. They have invited your team to put together an application using Fast API to illustrate how it works. Datalytics Inc. reviewed multiple APIs available and were inspired by https://developer.moodysanalytics.com/.

Task 1:

- Review https://ma.moodys.com/2020_9_AMER_EBU_PortalIntroductionandAPILibraryOverview_WebTel_MAU12329_ThankYou.html
- Use the APIWebinar code and review how the APIs are designed.
- Use https://github.com/mingrammer/diagrams to document your understanding of how this application works.

Task 2: Data Ingestion

We discussed different methods of ingesting data into the Snowflake database. Use SnowFlake as a staging environment and import the data into the Snowflake database. Discuss the design options you considered and how you came up with this architecture. Use Airflow to orchestrate the best design as an ingestion pipeline.

Task 3: Design the Fast API

Review the API design of the Moody's API and build out an API that can be used to query different aspects of the dataset.

- Review https://fastapi.tiangolo.com/tutorial/ for an intro to API
- Use **Pytest** to write unit tests for the API

Task 4: Enabling API key authentication

https://medium.com/data-rebels/fastapi-authentication-revisited-enabling-api-key-authentication-122dc5975680 Review this to enable API key authentication

Task 5: Test API

Build a Jupyter notebook to implement all the test cases your API supports.

Use **Locust** to load test the API

Data sets:

Team 1
https://www.kaggle.com/borismarjanovic/price-volume-data-for-all-us-stocks-etfs

Team 2 https://www.kaggle.com/hadiyad/lendingclub-data-sets

Team 3 https://www.kaggle.com/shasun/tool-wear-detection-in-cnc-mill

Team 4 https://www.kaggle.com/behrad3d/nasa-cmaps

Team 5 https://www.kaggle.com/tamilsel/healthcare-providers-data

Team 6
https://www.kaggle.com/inIT-OWL/production-plant-data-for-condition-monitoring