Hello and thank you for dedicating your time to this home assignment, it is much appreciated.

The assignment contains several tasks that are related to your future day-to-day job. These tasks involve several skills that we, at FireArc, see crucial for succeeding in the Data Engineer role.

We are looking forward to seeing how you express yourself through the assignment.

Time given to complete the assignment completely: 72 hours.

If you have any questions, feel free to contact us at: BI-Team@fire-arc.com

Good luck!

Scenario

You have been hired as a data engineer by FireArc Inc. Your task is to design and implement a data pipeline that collects data from NYC Open Data's Fire Incident Dispatch Data API, stores it in a data warehouse (any DB you favor) and performs some basic data transformations. The data pipeline should be implemented using Python and SQL.

Requirements

Collect data from the Fire Incident Dispatch Data API provided by NYC Open Data at the following URL:
https://data.cityofnewyork.us/Public-Safety/Fire-Incident-Dispatch-Data/8m42-w767

Parse the data and transform it into a format suitable for loading into a data warehouse.

Load the transformed data into a data warehouse using Python and the appropriate client library (e.g. sqlalchemy, google-cloud-bigquery, psycopg2 etc.)

Design a data model (Kimball) that enables efficient querying of the data. Define and implement appropriate indexes and partitioning strategies in the data warehouse to optimize data access, storage, query performance.
Write SQL queries to demonstrate the effectiveness of your indexing and partitioning strategies, including aggregations, filtering, and joining of multiple tables.
If needed, define and describe any maintenance tasks.

Describe the ETL pipeline for collection, transformation, and loading of the data into the data warehouse.

Deliverables

- Code for the data pipeline that includes scripts for collecting data from the API, transforming the data, and loading it into the data warehouse.
- SQL queries & processes
- Data model document that describes the structure of the data and the relationships between the tables, including partitioning and indexing strategies.
- Description of the ETL pipeline.

Submission Guidelines

- Create a GitHub repository that will contain your assignment.
- Commit your code, SQL queries, data model & ETL document to master.
- Share the repository with: bi-team@fire-arc.com (firearc-bi)
- You have 72 hours to complete the assignment.

Good luck!