Technical Take-Home Assessment for Senior Data Engineer

Objective:

Develop a data pipeline that extracts weather data from the OpenWeatherMap API and loads it into a PostgreSQL database, applying necessary transformations.

Scenario:

You are provided with access to the OpenWeatherMap API, which returns weather forecast data in JSON format. Your task is to design and implement a pipeline that efficiently extracts, transforms, and loads this data into a PostgreSQL database.

API Details:

- **API**: OpenWeatherMap API
- **Documentation**: OpenWeatherMap API Documentation
- **API Key**: You will need to sign up for a free API key to access the OpenWeatherMap API.

Requirements:

1. API Interaction:

- Write a python script to interact with the OpenWeatherMap API.
- Implement error handling and comply with the API's rate limits.

2. Data Extraction:

- Extract weather forecast data at regular intervals.
- Handle potential connectivity issues or API downtimes.
- Note: a few (10-20) api calls is fine here and be sure to stay well within the free tier of the api key.

3. Data Transformation:

- Transform the JSON data into a format suitable for database storage.
- Include necessary data cleaning and normalization steps.

4. Database Design:

- Design a schema for the PostgreSQL database to store the weather data.
- Consider appropriate data types, indexes, and constraints for the schema.

5. ETL Process:

- Outline an ETL process that integrates extraction, transformation, and loading steps.
- How would you orchestrate the workflow? Please outline the tech stack you would use.
- Ensure the process is scalable and maintainable.

6. Logging and Monitoring:

- Outline logging for tracking the pipeline's operations.
- Set up basic monitoring and alerts for the pipeline.

7. Testing:

- What unit tests for critical components of the pipeline would you consider?
- Note any testing strategies or frameworks utilized.

Deliverables:

- Source code for the data pipeline.
- Database schema (ERD).
- Code can be in either notebook or script format.
- Outlines and documentation should be in a short presentation.

Evaluation Criteria:

- Code quality and clarity.
- Efficiency and scalability of the solution.
- Adherence to best practices in data engineering.