BTTF Logistics Data Engineer Case Study

Overview

Back-to-the-Future Logistics, Inc. (BTTF Logistics) is a leading transportation company operating in over 20 countries across Asia. They deliver more than a million full truck loads yearly with a fleet of over 8,000 trucks and trailers.

Current Architecture and Usage Patterns

BTTF Logistics wants to understand how they can compete and prevail in this competitive market. Their fleet is already equipped with multiple sensors for reading GPS location, truck load, maintenance parameters of the truck and the trailer, as well as other telemetry data. BTTF uses a VIA Mobile360 D700 CAN bus enabled dash cam to collect the sensor data in-vehicle, however, they are not processing that information as of today.

BTTF is using a CRM system to manage their clients and the logistics for truck fleet scheduling, as well as legacy databases that contain additional information for contextualization purposes. That data is stored in SQL database, and is approximately 5 TB in size. That number is expected to grow to 15TB over the next 3 years.

Today, around 500 users are interacting with this data. This number is expected to grow to 1000 users over the next 2 years. Apart from a handful of BI analysts that use Tableau to directly access data in the source systems, there is no unified interface to access and explore the data.

Therefore, BTTF Logistics is looking for a partner to build a scalable, cost-efficient Cloud Data Platform that can integrate with current and future data sources, as well as democratize data access to enable new use cases. Both Amazon Web Services and Microsoft Azure are valid options for BTTF Logistics in terms of hyperscaler choice.

Key Challenges

BTTF Logistics has identified few key challenges:

- Lack of a Data Platform means data is spread across multiple data sources and it is hard to integrate new sources
- Need to store data in a centralized repository with consideration for storage costs and scalability
- Need to integrate weather data as a new data source to analyze how outside temperature affects fuel consumption

Your Task

As a Data Engineer, your task is to:

1. Design a solution architecture for a data platform that addresses BTTF's challenges
2. Develop proof-of-concept code for key components of the platform
3. Document the implementation approach in sufficient detail

Your solution should focus on these key areas:

1. Solution Architecture

Design a comprehensive solution architecture that includes:

- Data ingestion mechanisms for different data sources
- Data storage approach (Data Lake structure)
- Data processing and transformation components
- Analytics and reporting layer

2. Weather Data Integration Implementation

- Select a weather API and design a data collection approach
- Write a working script in your preferred programming language that demonstrates:
    - Weather data collection for specific locations
    - Data transformation and cleaning
    - Proper error handling and logging
    - Output to an appropriate file format

3. Data Model Design

- Design a data model that combines weather data with shipment data
- Create schema definitions for your proposed tables/datasets
- Document how this model supports the analytical requirements

4. Processing Pipeline Design

- Design a processing pipeline to calculate fuel consumption metrics
- Write working code that demonstrates:
    - Loading and joining relevant datasets
    - Computing average fuel consumption per temperature range
    - Creating aggregated metrics for reporting

5. Visualization Approach

- Explain what tools would be appropriate for implementing these visualizations

Available Datasets

Sample dump of the shipment and cities table will be provided along with this case study. You can download the dump from this S3 bucket. This dump contains two tables that have the following schema:

Cities Table

Column	Description
id	City ID
name	City name (joins with shipment table)
country_code	Country code
country_name	Country name
latitude	Latitude (for weather API)
longitude	Longitude (for weather API)

Shipment Table

Column	Description
id	Shipment ID
truck	Truck ID
driver	Driver ID
shipment_start_timestamp	Start time, used for extracting temperature
shipment_end_timestamp	End time
start_location	Start city, used for extracting temperature
end_location	End city, used for extracting temperature
shipment_distance	Distance in km (geodistance +20%)
consumed_fuel	Fuel used in liters

You can use the average temperature of the start_location and end_location at shipment_start_timestamp and shipment_end_timestamp as the temperature for a trip.

Deliverables

Please prepare:

1. Solution Architecture Document (3-5 pages)
    - High-level architecture diagram
    - Component descriptions and interactions
    - Technology selections with justifications
    - Scaling considerations
2. Implementation Examples

    - Working script for weather data collection in your preferred programming language (must be executable locally)
    - Data processing code (in any appropriate programming language)
    - Sample data model schemas (SQL DDL or equivalent)
3. Presentation & Code Walk through (45 minutes)
    - Overview of your solution architecture
    - Walk-through of your implementation approach
    - Demonstration of working code components
    - Discussion of how your solution addresses BTTF's challenges

Guidance

- Focus on a clean, well-documented solution rather than completeness
- Make reasonable assumptions where requirements are ambiguous, but document them

Good luck!