

Time Series Data Migration and Analysis

Description

You have a large dataset containing time series data from a sensor network, which includes timestamps and sensor readings from multiple sensors deployed in different locations. The goal is to migrate this time series data to a relational database, perform analysis to derive meaningful insights, and enrich the data with additional information using lookup operations.

Dataset

The dataset consists of the following columns:

- Timestamp (in UTC format)
- SensorID (unique identifier for each sensor)
- LocationID (unique identifier for each location)
- ReadingValue (sensor reading at the given timestamp)

Tasks

Generate Parquet Files

1. Prepare sample data in the specified format: Timestamp (UTC format), SensorID, LocationID, ReadingValue.
2. Generate Parquet files from the prepared data for analysis. Ensure appropriate encoding and schema are applied.

Data Migration

1. Design a relational database schema to store the time series data. You can create tables for sensors, locations, readings, etc., and establish appropriate relationships between them.
2. Migrate the time series data from its current format (Parquet files) to the relational database.

Data Normalization and Initial Analysis

1. Normalize the data to eliminate redundancy and ensure data integrity. Create separate tables for sensors, locations, and readings, and establish foreign key relationships.
2. Write SQL queries to find the average, minimum, and maximum sensor readings for a specific sensor within a given time period.

3. Write SQL queries to find the sensors that experienced readings above a certain threshold value.
4. Calculate the hourly, daily, and monthly averages for sensor readings and store the results in appropriate tables.
5. Write SQL queries to identify trends in sensor readings over time, such as identifying sensors with increasing or decreasing readings.
6. Show how to achieve the same using Pandas working with the Parquet files.

Data Enrichment with Location Information

1. Enrich the sensor data with location information using a lookup.
2. Write SQL queries and Pandas code to perform the lookup and create a new dataset with location names added.

ETL vs ELT Evaluation

1. Discuss the differences between ETL (Extract, Transform, Load) and ELT (Extract, Load, Transform) processes.
2. Analyze the advantages and use cases for each approach in the context of this dataset.
3. Provide a recommendation on whether to use ETL or ELT for this specific dataset based on the given requirements and constraints.