Extracting Data using SQL

Agenda:

In today's session, we'll cover essential topics, including:-

- Problem Statement
- Farmer's Database
- Relationships
- ER Diagram
- SQL Commands
- SELECT Query
- ORDER BY
- LIMIT & OFFSET
- Inline Calculation
- Alias (AS)
- ROUND() function

Summary of Previous Lecture:

Problem Statement:
- You're a Data Analyst at Amazon Fresh tasked with growing revenue through farmer's market stores.
- You have access to a MySQL relational database and its schema.

What is a Database?
- A database is a collection of interrelated tables.
- Database Management System (DBMS) is a set of programs to access and manipulate data.
- Operations include CRUD (Create, Read, Update, Delete), search, insert, etc.
- Relational DBMS (RDBMS) stores data in tables (e.g., MySQL, PostgreSQL).

Why Use Databases Over Excel?

- Scalability: Databases handle large datasets effectively.
- Performance: Databases are optimized for fast data retrieval.
- Data Integrity: Databases enforce data accuracy.
- Concurrent Access: Multiple users can access data simultaneously.
- Security: Databases provide user access management.

DB Schema (Entity-Relationship Diagram):
- A schema represents data organization and table relationships.
- Tables are represented as boxes, relationships as lines.
- Data types define the kind of data stored in each column.

Common Data Types
- String: Char(), Varchar()
- Numeric: Int64, Float64
- Date & Time: Date, Time, Datetime, Timestamp

Concept of Keys
- Keys ensure unique identification of records in a table.
- Primary Key: Uniquely identifies rows, can't be updated.
- Unique Key: Unique, non-updatable, may have NULL.
- Foreign Key: Links to the primary key in another table.
- Candidate Key: Any column or set of columns that can act as a primary key. Every table must have at least one candidate key.

Database vs. Data Warehouse
- Data warehouse is optimized for analytics (OLAP).
- Database is for transactions (OLTP).
- Cloud service providers offer data warehousing services (e.g., AWS, GCP, Azure).

Working with BigQuery on GCP:
- BigQuery is a cloud-based data warehouse.
- Set up a project, create a dataset, and upload tables.
- BigQuery is used for analytics and querying large datasets.

BigQuery setup doc - link
BigQuery setup video - link