**🌐 What is BigQuery?**

**BigQuery** is a **serverless, fully-managed data warehouse** by **Google Cloud Platform (GCP)**.

It's used to:

* Store **large amounts of structured data**
* Run **fast SQL queries** on that data
* Perform **analytics and reporting** in real time

**🧠 Simple Definition:**

**BigQuery is like a super powerful online database where you can upload your data and run fast SQL queries without managing servers.**

**🏗️ Key Concepts in BigQuery**

**1. Project**

* Your GCP workspace where BigQuery resources live.
* Contains datasets and billing info.

**2. Dataset**

* A container for tables.
* Like a folder where you keep related data.

**3. Table**

* Stores structured data (rows and columns).
* Example: sales\_data, customer\_info

**4. Schema**

* Defines the structure of a table (columns: name, type, etc.)

**5. SQL Queries**

* BigQuery uses **standard SQL** to query data.
* Very similar to other SQL systems like MySQL or PostgreSQL.

**⚙️ How BigQuery Works (Behind the Scenes)**

**1. Storage Layer**

* Stores petabytes of structured data.
* You pay for how much data you store.

**2. Compute Layer**

* When you run a SQL query, BigQuery automatically manages compute resources.
* You pay based on the **amount of data scanned**, not compute time.

**3. Serverless Model**

* No need to manage infrastructure (VMs, servers, etc.).
* Google handles performance tuning, scaling, and maintenance.

**🔄 Typical BigQuery Workflow**

STEP 1: Load data

→ Upload CSV, JSON, or connect to other sources like Cloud Storage, Cloud SQL, etc.

STEP 2: Organize

→ Create datasets and tables.

STEP 3: Query

→ Use SQL to analyze your data.

STEP 4: Visualize or Export

→ Send results to Data Studio, Looker, or export to CSV.

**💻 Example SQL Query in BigQuery**

Suppose we have a table sales\_data:

SELECT

customer\_name,

SUM(order\_amount) AS total\_spent

FROM

`my\_project.my\_dataset.sales\_data`

WHERE

order\_date >= '2024-01-01'

GROUP BY

customer\_name

ORDER BY

total\_spent DESC

✅ This SQL query gives you total spending per customer since Jan 2024.

**📥 Supported Data Sources for BigQuery**

You can load data from:

* **CSV / JSON / AVRO / Parquet files**
* **Google Cloud Storage**
* **Google Sheets**
* **Cloud SQL**
* **Firebase / Google Analytics**
* **Streaming data (Pub/Sub)**

**💰 Pricing Model (High Level)**

| **Cost Type** | **Description** |
| --- | --- |
| **Storage** | How much data is stored in BigQuery |
| **Query** | Based on how much data is scanned |
| **Streaming Inserts** | When you insert data row-by-row in real time |

**📊 Real-World Use Cases**

| **Industry** | **How BigQuery Helps** |
| --- | --- |
| E-commerce | Analyze product sales, user behavior, returns |
| Healthcare | Track patient records and treatment outcomes |
| Finance | Detect fraud, generate reports |
| Marketing | Measure ad performance, campaign ROI |

**🛠️ Tools That Work With BigQuery**

* **Google Data Studio / Looker** – Visualization
* **Python / Jupyter Notebooks** – Data analysis
* **Terraform** – Infrastructure automation
* **Cloud Composer (Airflow)** – Workflow orchestration
* **dbt** – Data transformation