Day 12- homework – sakina - GDE

**1. What is ETL in data engineering, and why is it important?**

ETL stands for Extract, Transform, and Load. It’s a fundamental process in data engineering where we first extract data from various sources, then transform it by cleaning, enriching, or reshaping it and finally load it into a data warehouse or destination system. It’s important because raw data from different sources is often inconsistent or unstructured, and ETL ensures that it becomes clean, consistent, and ready for analysis. Efficient ETL pipelines are the backbone of any data-driven decision-making process.

**2. What are the key differences between cloud storage and traditional on-premise storage?**

One of the major differences is scalability and accessibility. Cloud storage, like Google Cloud Storage, is highly scalable and accessible from anywhere with internet access, while on-premise storage is limited by physical infrastructure and local access. Maintenance is another key difference cloud providers manage the hardware and security, whereas on-premise solutions require in-house IT resources. Also, cloud follows a pay-as-you-go model, making it cost-effective, especially for startups and growing businesses.

**3. What is Google Cloud Storage (GCS), and what are some common use cases?**

Google Cloud Storage is an object storage service provided by Google Cloud. It’s designed for storing and retrieving large volumes of unstructured data like images, videos, backups, and log files. Some common use cases I’ve seen include hosting static websites, storing datasets for machine learning, serving as backup storage, and even connecting to BigQuery for external table analysis. GCS is known for its durability, security, and seamless integration with other GCP services.

**4. What is the role of BigQuery in Google Cloud, and how is it different from a traditional database?**

BigQuery is Google’s fully managed, serverless data warehouse optimized for big data analytics. Unlike traditional databases that are optimized for transactions, BigQuery is built for analytical queries across massive datasets. It uses columnar storage and distributed architecture, which makes it extremely fast and scalable. One big advantage is that you don’t need to manage any infrastructure—just load your data and run your queries using SQL. It's ideal for use cases like real-time analytics, dashboards, and reporting.

**5. What are the main types of cloud service models (IaaS, PaaS, SaaS), and which Google Cloud services fall under each?**

The three primary cloud service models are IaaS, PaaS, and SaaS. IaaS, or Infrastructure as a Service, gives you the raw building blocks like virtual machines and storage. In GCP, Compute Engine and Cloud Storage fall under this. PaaS, or Platform as a Service, provides a ready-to-use environment for app development and deployment App Engine and BigQuery are great examples. SaaS, or Software as a Service, is about ready-made applications accessible via the web Google Workspace tools like Gmail and Docs are classic examples. Understanding these models helps in choosing the right tool for the right task.