

Database Week 9

Khaul Molapo

1.

If I had to set up a data warehouse for an e-commerce system, I'd make it store all the business data from different sources like sales, customers, and products in one big place. The data would come from systems like the app, website, and payment platform, then be cleaned, sorted, and stored for reporting. I'd use ETL tools to pull the data from each source, transform it so it fits nicely, and then load it into the warehouse. The warehouse would have separate areas for raw data and processed data. This helps me run reports fast without messing up the live systems. For example, managers can check sales trends, top products, or customer behavior easily. The main goal is to have all the data in one place, easy to access, analyze, and make decisions from, without stressing the main system that runs the store.

2.

A star schema has one big main table in the middle called the fact table, and smaller tables around it called dimension tables, like customers, products, and dates. It's simple and fast for reading data. A snowflake schema is similar but the dimension tables are split into smaller related tables, making it more organized but a bit slower. The star schema is easier to understand and good for quick reports. The snowflake schema uses less space and handles complex data better. I'd use star schema for simple analytics and snowflake when the data is big and more detailed.

3.

In Python, I can simulate an ETL process by writing a small script that extracts data from a CSV file, cleans it, and then loads it into a database. For example, I'd read a sales file, fix errors like missing prices, change column names, and then insert everything into a SQL table. This shows how ETL helps move data from one place to another while cleaning it up. It's useful for data warehouses because you get neat, ready-to-use data instead of messy raw data. It saves time later when doing reports or analysis since everything's already prepared.

4.

If I were designing a data warehouse, I'd make it have layers — one for raw data, one for cleaned data, and one for reports. Data from systems like sales or customer service would first be collected, then cleaned and stored properly. The warehouse would use a star schema so reports load faster. This setup helps with analyzing sales trends, customer habits, and inventory without slowing down the main system. The goal is simple: make decision-

making easier by putting all data in one clean, organized space. It saves time, reduces confusion, and makes reports way easier to generate.