Task 3

Q1) Can a database be used as DWH?

Yes, a database can be used as a data warehouse (DWH)

Q2) Difference between structure and unstructured data?

Structured data refers to data that is organized in a specific format or structure, such as a table, spreadsheet, or database. Structured data is typically stored in a fixed format and has a predefined schema, which makes it easy to search, analyze, and process. Examples of structured data include customer information, transaction data, inventory data, and financial data.

Unstructured data, on the other hand, refers to data that does not have a specific format or structure, and is often stored in a variety of different formats, such as text, images, audio, or video. Unstructured data is typically more difficult to process and analyze than structured data, because it lacks a predefined schema or organization. Examples of unstructured data include social media posts, emails, documents, images, and videos.

Q3) What are the duties of a data engineer? (high-level)

- Data architecture: Designing and building the architecture for data storage and processing, including databases, data warehouses, data lakes, and data pipelines.
- Data integration: Developing and maintaining processes for extracting data from various sources, transforming it into a usable format, and loading it into a central repository.
- Data quality: Ensuring the accuracy, completeness, and consistency of data by developing and implementing data quality checks, data validation rules, and error handling processes.
- Data security: Implementing security measures to protect sensitive data, including access control, encryption, and data masking.
- Performance optimization: Optimizing data storage and processing performance by tuning database configurations, improving data retrieval times, and minimizing processing latency.