**Bytewise DE Task**

**BY: Mohammad Yaqoob**

**Date: March/15/2023**

**Task # 3:**

* Can a database be used as DWH?

A database can be used as a data warehouse (DWH) if it is designed and configured properly to support the needs of a data warehouse. A data warehouse typically involves large volumes of data that need to be transformed, integrated, and analyzed in order to support business intelligence (BI) and decision-making. To use a database as a data warehouse, you would need to design your schema, data loading, and querying strategies to support the requirements of your data warehouse. This might involve techniques such as partitioning, indexing, and query optimization to ensure that your queries can be executed efficiently over large data sets.

* Major differences between structured and Unstructured data.

Structured data and unstructured data are two different types of data with different characteristics and properties. Here are the major differences between them:

1. Definition: Structured data is organized and formatted in a specific way, with defined fields, values, and relationships between them. Unstructured data, on the other hand, is data that does not have a specific structure or format, making it more difficult to organize and analyze.
2. Format: Structured data is usually stored in databases, spreadsheets, or tables, where the data is arranged in rows and columns. Unstructured data, on the other hand, can be stored in various formats such as text, images, and audio, video, and social media posts.
3. Processing: Structured data is easier to process and analyze, as the format is consistent and predictable. Unstructured data, however, requires more processing and analysis, as the format and content can vary greatly.
4. Volume: Structured data is usually smaller in volume, as it is organized and stored in a specific format. Unstructured data, on the other hand, can be very large in volume, as it includes various types of data such as text, images, and videos.
5. Analysis: Structured data is easier to analyze using tools such as SQL queries and statistical analysis. Unstructured data, however, requires more advanced analysis techniques such as natural language processing and machine learning to extract insights and patterns.
6. Source: Structured data is typically generated by transactional systems such as point-of-sale systems, financial systems, and customer relationship management systems. Unstructured data, on the other hand, is generated by a wide range of sources including social media, email, sensors, and website logs.

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

The duties of a data engineer can vary depending on the organization and industry they work in, but generally include the following high-level responsibilities:

1. Data Infrastructure Design: A data engineer designs and builds data infrastructure that supports data processing, storage, and retrieval.
2. Data Pipeline Development: A data engineer develops and maintains data pipelines that extract, transform, and load (ETL) data from various sources into the data warehouse or other data storage systems.
3. Data Quality Management: A data engineer ensures that the data is accurate, complete, and consistent by developing and implementing data quality checks and cleansing procedures.
4. Data Warehousing: A data engineer is responsible for designing and maintaining data warehouse systems, including data modeling, database design, and optimization.
5. Data Security: A data engineer ensures the security and privacy of the data by implementing access controls, data encryption, and other security measures.
6. Performance Optimization: A data engineer optimizes the performance of data systems by monitoring system performance, identifying bottlenecks, and implementing performance improvements.
7. Data Analysis Support: A data engineer provides support to data analysts and data scientists by designing and implementing data structures that support data analysis and machine learning.
8. Documentation: A data engineer documents the design, implementation, and maintenance of data infrastructure and systems.