|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**W1-task 3 :**

Can a database be used as DWH?

Yes, a database can be used as a data warehouse, although it may not be as efficient or scalable as a dedicated data warehouse solution. A data warehouse typically involves a large amount of data that needs to be analyzed and processed for business intelligence purposes. While traditional databases can handle data storage and retrieval, they may not be optimized for complex analytical queries or large-scale data processing. A dedicated data warehouse solution, on the other hand, is specifically designed to handle these tasks efficiently and at scale.

Major differences between structured and unstructured data.

Structured data refers to data that is organized in a specific format, such as a table with rows and columns. This type of data is easily searchable, analyzable, and can be processed by traditional data management tools. Examples of structured data include financial data, customer data, and inventory data.

Unstructured data, on the other hand, refers to data that does not have a predefined structure or format. This type of data is typically generated by humans or machines and includes text, images, videos, and audio files. Unstructured data is more difficult to search, analyze, and process than structured data, and often requires specialized tools and techniques. Examples of unstructured data include social media posts, customer feedback, and sensor data.

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

The duties of a data engineer typically include:

Designing, building, and maintaining data pipelines and ETL processes

Ensuring data quality, integrity, and security

Creating and managing data architectures, data models, and data warehouses

Collaborating with data analysts, data scientists, and business stakeholders to understand their data needs and requirements

Developing and implementing data governance policies and procedures

Monitoring and optimizing data storage and retrieval systems for performance and scalability

Staying up-to-date with the latest data engineering technologies and best practices.

Top of Form