**Task 03**

**Abdul Razaque**

**Data Engineering**

* **Can a database be used as DWH?**

Yes, a database can be used as a Data Warehouse (DWH) if it is specifically designed and optimized for analytical queries, and it meets certain requirements such as high performance, scalability, and reliability.

A DWH requires a different design approach than a traditional transactional database. A DWH usually involves large volumes of historical data that is organized and stored in a way that enables efficient querying and analysis. It may also include data from multiple sources and integrate them into a single source of truth.

A database that can serve as a DWH needs to support features such as:

* Column-oriented storage to optimize query performance
* Data compression to save storage space and improve performance
* Partitioning to enable scalability and manageability of the data
* Support for complex data models, such as star or snowflake schema
* Advanced indexing and query optimization techniques
* High availability and fault tolerance to ensure data reliability

Examples of databases that can be used as a DWH include Amazon Redshift, Google BigQuery, Microsoft Azure SQL Data Warehouse, etc.

* **Major differences between structured and Unstructured data:**
* **Structured data:**
* Structured data is standardized, clearly defined, and searchable data.
* Structured data is organized into predefined formats such as tables with rows and columns.
* Structured data is usually stored in databases that are designed to store and manage structured data.
* Structured data can be easily analyzed using traditional data analysis techniques such as SQL queries and data visualization tools.
* Structured data tends to be more organized and compact.
* Structured data tends to be simpler and easier to understand.
* **Unstructured data:**
* Structured data is quantitative, while unstructured data is qualitative.
* Unstructured data can take many different forms and is not organized in a specific way.
* Unstructured data, on the other hand, can be stored in a variety of ways such as file systems, content management systems, and NoSQL databases.
* Unstructured data requires more advanced analysis techniques such as natural language processing, machine learning, and deep learning algorithms.
* Unstructured data can be much larger and more difficult to manage due to its diverse and unpredictable nature.
* Unstructured data can be much more complex and challenging to make sense of due to its lack of structure and diverse content.
* **What are the duties of a data engineer? (high-level):**
* Analyze and organize raw data.
* Build data systems and pipelines.
* Evaluate business needs and objectives.
* Interpret trends and patterns.
* Conduct complex data analysis and report on results.
* Prepare data for prescriptive and predictive modeling.
* Build algorithms and prototypes.