**TASK NO: 3**

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

Yes, Data Warehouse is also a special or giant database.

Data warehouses and databases both act as data storage and management tools. However, there are a few key differences to acknowledge. First, data warehouses have analytical capabilities. They enable companies to make analytical queries that track and record certain variables for business intelligence. In contrast, a database is a simple collection of data in one place. Databases’ main purpose is to store data securely and allow users to access it easily.

## **Differences between structured and unstructured data?**

While structured (quantitative) data gives a “birds-eye view” of customers, unstructured (qualitative) data provides a deeper understanding of customer behavior and intent. Let’s explore some of the key areas of difference and their implications:

**Sources:** Structured data is sourced from GPS sensors, online forms, network logs, web server logs, [OLTP systems](https://www.ibm.com/cloud/learn/oltp), etc., whereas unstructured data sources include email messages, word-processing documents, PDF files, etc.

**Forms:**Structured data consists of numbers and values, whereas unstructured data consists of sensors, text files, audio and video files, etc.

**Models:** Structured data has a predefined data model and is formatted to a set data structure before being placed in data storage (e.g., schema-on-write), whereas unstructured data is stored in its native format and not processed until it is used (e.g., schema-on-read).

**Storage:**Structured data is stored in tabular formats (e.g., excel sheets or SQL databases) that require less storage space. It can be stored in data warehouses, which makes it highly scalable. Unstructured data, on the other hand, is stored as media files or No SQL databases, which require more space. It can be stored in data lakes which makes it difficult to scale.

**Uses:**Structured data is used in machine learning (ML) and drives its algorithms, whereas unstructured data is used in [natural language processing](https://www.ibm.com/cloud/learn/natural-language-processing) (NLP) and text mining.

**DUTIES OF DATA ENGINEER:**

Data engineer involves working with large collections of data, recording and analyzing it, then converting it into easy-to-read reports or presentations.

They may also be responsible for creating organized databases with detailed programs to help businesses collect and interpret data.

They do this by developing, maintaining, and testing infrastructures for data generation.

Data engineers work closely with data scientists and are largely in charge of architecting solutions for data scientists that enable them to do their jobs