Task: 03– by Bytewise Team

# Can we use the database as a data warehouse?

Technically, it is possible to use a database as a data warehouse, but it may not always be the best solution.

A database is typically designed to handle transactions and perform operations quickly and efficiently, while a data warehouse is designed to store large amounts of data and support complex reporting and analysis.

Using a database as a data warehouse can work for small to medium-sized businesses with relatively simple reporting requirements. However, as the data volume and complexity increase, it can become difficult to manage and maintain, and performance may suffer.

Therefore, it is generally recommended to use a dedicated data warehouse solution for larger and more complex data sets, and use databases for transaction processing and other operational tasks.

# Major differences between structured and Unstructured data?

The main differences are listed down.

* Structure data has a fixed format like in tables or schemas. While others have no fixed format, they can be anything text, images, videos, etc.
* Structure data is organized and easy to read. While other is not organized and need a particular way to deduce meaning.
* Structure data is easy to analyze using traditional data mining techniques and tools as it is organized well and has a fixed format. While unstructured needs special advanced analysis techniques like ML and NLP to make data useful to the organization.
* Structure data has a small volume and is stored in a database while un-structure data is in raw form and have to be stored in files and other specific places made for it due to its large volume.

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

The following are the duties of a **Data Engineer** responsible for:

* **Design, Build and Maintain Data Pipeline:** Data engineers are the hardworking guys, who work in the backend and are responsible for all the amount of data that comes in.
* **Tooling Part:** Data engineers generally use Python, SQL, and ETL tools (Informatica, SAP data services, etc), NoSQL, and Bigdata. All these technologies are very closely inclined toward the data engineer profile.
* **Work in DataOps:** Data engineers generally work in a team known as DataOps also called Data Operations. They do all the dirty work to make sure that data pipelines are intact, the data is coming as well as data is of the highest quality.
* **Manages 5Vs of Big Data:** `Many new roles come in the world of IT/data field when the Big data was introduced. To understand Big data you need to know 5Vs.

**Which are Volume, Variety, Velocity, Veracity, and Value.** The first 4Vs of Big data are tackled or managed by data engineers and the 5V is the Value part, It's the duty of the data analyst and data scientist to deliver the value part of data.

So, folks to end this summary, Data engineers sometimes do the job of Data analysts, and Data scientists might go back and do the job of Data engineers. It all depends on what kind of team you have or the Business model you have.