Task 3

- Can a database be used as DWH?

A data warehouse is a type of database the integrates copies of transaction data from disparate source systems and provisions them for analytical use.

- Major differences between structured and Un-structured data.

| **Basis of Differentiation** | **Structured Data** | **Unstructured Data** |
| --- | --- | --- |
| Analysis | Quantitative | Qualitative |
| Schema Creation | Schema-on-write | Schema-on-read |
| Searching | Easy using SQL-based methods | May need special tools |
| Format | Predefined, using alphanumeric characters | Typically non-character-oriented digital representations |
| Storage | May require more storage to accommodate defined data structures | Some forms require less storage; others have large file formats, requiring more storage |
| |  |  |  | | --- | --- | --- | | Storage Format | Relational database management systems, data warehouses | Applications, NoSQL databases, data lakes | | | |

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

**Data Engineer Duties**

1. Work on Data Architecture

They use a systematic approach to plan, create, and maintain [data architectures](https://www.simplilearn.com/what-is-data-architecture-article) while also keeping it aligned with business requirements.

2. Collect Data

Before initiating any work on the database, they have to obtain data from the right sources. After formulating a set of dataset processes, data engineers store optimized data.

3. Conduct Research

Data engineers conduct research in the industry to address any issues that can arise while tackling a business problem.

4. Improve Skills

Data engineers don’t rely on theoretical database concepts alone. They must have the knowledge and prowess to work in any development environment regardless of their [programming language](https://www.simplilearn.com/best-programming-languages-start-learning-today-article). Similarly, they must keep themselves up-to-date with machine learning and its algorithms like the [random forest](https://www.simplilearn.com/tutorials/machine-learning-tutorial/random-forest-algorithm), [decision tree](https://www.simplilearn.com/the-power-of-decision-trees-in-machine-learning-article), [k-means](https://www.simplilearn.com/tutorials/machine-learning-tutorial/k-means-clustering-algorithm), and others.

They are proficient in analytics tools like [Tableau](https://www.simplilearn.com/learn-tableau-tips-to-start-article), Knime, and [Apache Spark.](https://www.simplilearn.com/apache-spark-guide-for-newbies-article) They use these tools to generate valuable business insights for all types of industries. For instance, data engineers can make a difference in the health industry and identify patterns in patient behavior to improve diagnosis and treatment. Similarly, law enforcement engineers can observe changes in crime rates.

5. Create Models and Identify Patterns

Data engineers use a descriptive data model for data aggregation to extract historical insights. They also make[predictive models](https://www.simplilearn.com/what-is-predictive-analytics-article) where they apply forecasting techniques to learn about the future with actionable insights. Likewise, they utilize a prescriptive model, allowing users to take advantage of recommendations for different outcomes. A considerable chunk of a data engineer’s time is spent on identifying hidden patterns from stored data.

6. Automate Tasks

Data engineers dive into data and pinpoint tasks where manual participation can be eliminated with automation.