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

Theoretically, a database can be used as a Datawarehouse but it will not be an optimized solution since databases are geared towards OLTP but datawarehouses are more focused on OLAP.

**Q. Major differences between structured and Un-structured data.**

Here are some of the differences in structured and unstructured data:

|  |  |
| --- | --- |
| **Structured data** | **Unstructured data** |
| It is highly organized | It doesn’t follow any predefined pattern |
| Can be accessed via SQL queries | Can only be accessed with the help of advanced tools |
| Stored in relational databases like MySQL | Stored in NoSQL databases like MongoDB |
| It is generally smaller in size | It can be in huge size and volume |
| It is easier to process | It requires advanced algorithms and NLP techniques to extract meaningful insights from it. |
| It is generally more accurate due to its predefined structure. | It can be less accurate due to inconsistencies and errors. |

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

Following are the general responsibilities of a data engineer:

* A typical data engineering project starts from collecting data and ends at a decision
* The first phase is collecting data via SFTP or API calls
* Now the data engineer starts to ‘prepare’ data for the project i.e. cleaning the data (adding/removing columns, changing data types etc)
* After this the data engineer writes the core business logic to extract insights from the data using libraries such as python, spark etc
* Now the insights are used to create reports to be further used by Business intelligence engineer/ Data scientists.
* A data engineer also creates and maintains ETL pipelines from various sources to store data in data warehouses or data lakes