Can a database be used as DWH?

1. Yes, a database can be used as a data warehouse (DWH) if it is designed to handle large volumes of data and optimized for reporting and analysis. However, a traditional database may not be the most efficient solution for a DWH, as it may not be able to handle the complexity of data transformations and queries required for a comprehensive data analytics solution. Therefore, specialized DWH solutions such as Amazon Redshift, Snowflake, or Google BigQuery are more commonly used.

Major differences between structured and unstructured data.

1. Structured data refers to data that is organized and formatted in a specific way, usually in tables with clearly defined rows and columns. This data is easy to query and analyze using traditional data analysis tools, such as SQL. Examples of structured data include transactional data, financial data, and customer data. Unstructured data, on the other hand, refers to data that is not organized in a specific format, making it difficult to analyze with traditional data analysis tools. Examples of unstructured data include social media posts, videos, images, and audio recordings. Unstructured data is often analyzed using natural language processing (NLP), machine learning (ML), and other advanced data analysis techniques.

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

1. The duties of a data engineer can vary depending on the organization and the specific role, but some common high-level duties of a data engineer include:

* Designing, building, and maintaining data pipelines that collect, process, and store data from various sources.
* Developing and maintaining data infrastructure, including databases, data warehouses, and data lakes.
* Ensuring data quality and accuracy by developing and implementing data validation and testing procedures.
* Collaborating with data scientists, analysts, and other stakeholders to develop and implement data solutions that meet business requirements.
* Implementing and maintaining data security and compliance measures.
* Staying up-to-date with the latest data engineering technologies and tools, and continuously improving data engineering processes and best practices.