Data Engineer FellowShip with Bytewise Limited.

Meer Danish 14/March/2023

horizontal line

Task # 1

# What is Big Data?

## Big Data:

Big Data is a term used to describe extremely large data, complex

And diverse data sets that cannot be processed using traditional

Data processing techniques.

## Types Of Big Data:

There are typically three types of big data on their structure.

* Structured Data : This Type Of Data IS Easy To Search For Specific Data.
* Unstructured Data : This Is Complicated Data.Any Type Of Data, Example: Emails,Video,Audio, Text, Messages, Transaction And Etc.
* Semi Structured Data : This is Semi Structured Data Means Include Xml And Json Files.

## Characteristics Of Big Data:

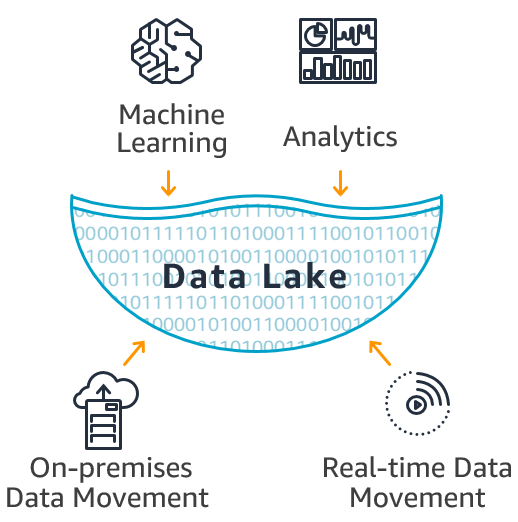
* Volume : Big Data IS Very Large Means Terabytes to Petabytes in size.
* Velocity:It generates Highly Velocity and in real time making it challenging to process and analyze.
* Variety : Big Data Is Often Heterogeneous and Cons from Variety of source and in a variety of formats.

## 

## What is a data lake?

## Data Lake:

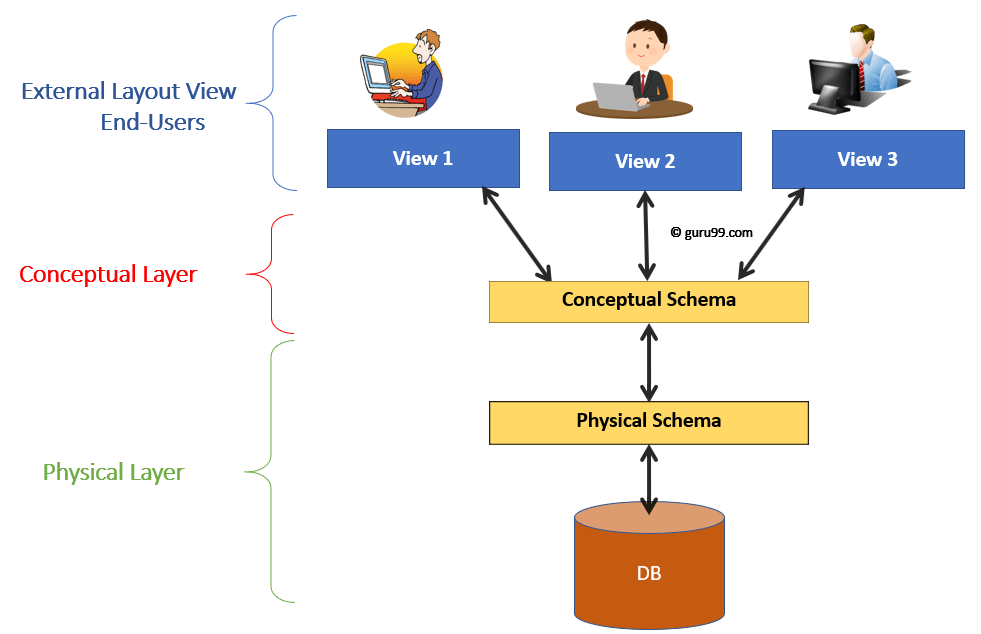
A data lake is a centralized repository that allows you to store all your structured and unstructured data at any scale. You can store your data as-is, without having to first structure the data, and run different types of analytics—from dashboards and visualizations to big data processing, real-time analytics, and machine learning to guide better decisions.



# What is a DataBase?

## DataBase:

A database is an organized collection of structured information, or data, typically stored electronically in a computer system. A database is usually controlled by a [database management system (DBMS)](https://www.oracle.com/in/database/what-is-database/#WhatIsDBMS).



# What is a Data Warehouse?

## Data Warehouse:

Data Warehouse is a subject oriented ,integrated,Non-Volatile and time variant, collection of data in support of management decisions.

