Databases and Big Data

What are databases?

Databases are structured ways of storing information to be manipulated or moved to other databases later. Their structure consists of a bunch of tables related to each other that have columns and rows for organization.

What are Data Base Management System?

A [DBMS](https://builtin.com/data-science/database-management-system) is a software that we use to create and manage databases.

Types of databases?

There are many types of databases. Below are some of the more relevant ones:

1. Hierarchical

The earliest form of databases. You can think of them as a family tree.

1. Relational

The most popular type of database. Relational databases connect their tables using keys

1. Non-Relational or NoSql

They were invented more recently than relational databases. They are any database that doesn’t use a relational model.

1. Cloud

They refer to information that’s accessible in a hybrid or cloud environment. All users need is internet to get their files to manipulate them.

What is big data?

It is larger and more complex data sets that traditional data processing software can’t manage them, But they can be used to address business problems you wouldn’t have been able to tackle before.

Big Data Processing and Distribution Software?

 They are designed for large-scale installations and assists businesses in organizing massive amounts of data.

Types of big data?

1. Structured

The standardized format with a well-defined structure. Structured data is organized in a table with relationships between the columns and rows. Each field may be accessed independently or in conjunction with information from other areas.

1. Semi-Structured

Data that lacks a precise functional structure but has some structural qualities. Their information is grouped by topic. The advantage is that it’s widely available and can be used to generate in-depth insights.

1. Unstructured

Data that has no predetermined structure. Every day, businesses receive massive volumes of it. However, generating insights from it is difficult and requires huge computational power.

Differences between big data and databases?

|  |  |
| --- | --- |
| Dynamic schema | Fixed Schema |
| Deals with Structured, Semi, and Unstructured Data | Deals with Structured Data |
| Exceeds the capacity of traditional databases | Can be easily accessed, managed, and updated unlike big data |
| Has a data lake | Has a database model |
| Large Datasets (Gigabytes to Terabytes) | Small – Medium Data sets (Petabytes to Exabytes) |
| Analysis Techniques are often inadequate | Analysis Techniques involve using statistical methods and visualizations |
| Provide valuable insights into customer behavior, market trends, and business-critical info | Has a lower potential value due to its limited scope and size |
| Generated per hour, day, or more | Generated per seconds |
| ERP and CRM transaction, financial, and organizational data sources | Social media, device and sensor data, video, images , and audio data sources |

Differences between big data’s and databases’s Software ?