Databases

Introduction

Databases are collections of data, usually organized under a schema, and stored in a format that is efficient for storing and retrieving the data. When people talk about databases they tend to talk about the underlying **Database Management System (DBMS)**. These are programs like MySQL or PostgreSQL which are designed to complete the tasks of storing, retrieving, updating, caching, deleting, and other data manipulation.

Databases use tables for managing data. Using tables we can handle [big data](https://lms.clarusway.com/mod/lesson/view.php?id=996) storage, build a relationship between data and give priorities to some data on the table. Thus we can reach specific information asking questions to tables. These questions called **query**.

| **https://docs.google.com/uc?id=1mV-LTDPYEzYDrFqW9YLvOZolZAKG7sNZ** |
| --- |
| *Database and tables* |

Q: Explain Database and Database Management System.  
A: 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). Some examples of popular database software or DBMSs include MySQL, MongoDB, PostgreSQL, Microsoft SQL Server

## Databases

### Database Structure

Imagine an education program that stores student information, lesson and class information, teacher information, a student enrolled course information etc.

1. All of these data relate to one another:
   * All students can enroll more than one courses (students, courses, etc)
   * All courses have to have a teacher (teacher, courses)
   * Courses have a status (close, open for enrolling, etc)
2. Imagine this data is stored in [files](https://lms.clarusway.com/mod/lesson/view.php?id=1052) on a hard drive.  
   Is it stored by the student's name? If so, how can we select only students that passed "Computer Essentials and Network" Course?

A database system’s fundamental goal is to provide consistent views of structured data, just like the relationships we’ve laid out between students and courses.

**Structure**

SQL databases are based on around relational algebra

* Tables are the way we look at our relevant data.
* Columns are fields in the table.
* Rows define a relation between fields.
* A Primary key is a set of columns that uniquely identify rows in a table.
* A Foreign key is a column that matches the primary key of another table.

**Table-1 Students**

| **Student-ID (Primary key)** | **Student Name (Column-2)** | **Student Mail Address (Column-3)** |
| --- | --- | --- |
| 0001 (Row-1) | Albert Einstein | einstein@clarusway.com |
| 0002 (Row-2) | Nikola Tesla | tesla@clarusway.com |
| ... | ... | ... |

**Table-2 Courses**

| **Course Code (Primary key)** | **Course Name** | **Enrolled Student (Foreign key to Table 1)** |
| --- | --- | --- |
| CESN501 | Computer Essentials and Networking | 0002 |

Databases

Types of Databases

There are two broad types of databases. These are SQL and NoSQL.

1. SQL: Stores data in tables organized by column and field.
2. NoSQL: Stores data differently than an SQL database.

SQL databases are primarily called Relational Databases (RDBMS); whereas the NoSQL database is primarily called a non-relational or distributed database.

**SQL**

* SQL databases are classic databases and are what we default to talking about when we teach databases.
* SQL databases define and manipulate data based on structured query language (SQL) yapılandırılmış sorgu diline (SQL)
* SQL databases are table-based
* Great support is available for all SQL databases from their vendors. Also, a lot of independent consultations are there who can help you with SQL database for a very large scale deployments

Examples:

* MySQL/MariaDB
* PostgreSQL
* SQLite

**NoSQL**

* A NoSQL database has a dynamic schema for unstructured data.
* NoSQL databases are either key-value pairs, document-based, graph databases or wide-column stores.
* NoSQL databases can also offer more flexibility in storage options, allowing one to spread data across many machines more easily than SQL databases tend to do.
* In NoSQL database you have to rely on community support and only limited outside experts are available for setting up and deploying your large scale NoSQL deployments.

Examples:

* MongoDB
* Apache Casandra
* Elasticsearch

Q: What is SQL and have you heard about NoSQL?  
A: SQL is a programming language used by nearly all **relational databases** to query, manipulate, define data and to provide access control. A NoSQL, or **nonrelational database**, allows unstructured and semistructured data to be stored and manipulated (in contrast to a relational database)

Databases

Database and Query Example

Assume that Apple's HR department wants to hire a person who knows Java programming language and age is under 25. Let's look and make a query to find where is the lucky guy from Clarusway Database?

**Students** (Table)

| **ID** | **Student Name** | **Enrolled Courses** | **Age** |
| --- | --- | --- | --- |
| 0001 | Albertoo Einstein | SQL | 24 |
| 0002 | Nikolass Tesla | Introduction to Testing | 32 |
| 0003 | Steve Jobies | Java | 28 |
| 0004 | Tim Cookie | Java | 21 |

Write Query will be like that:



1

2

SELECT Student Name FROM Students

WHERE Enrolled Courses= Java AND Age < 25 ;

## Big Data

What is Big Data?

Big data is data that contains greater **variety** arriving in increasing **volumes** and with ever-higher **velocity**. This is known as the three Vs.

Put simply, big data is larger, more complex data sets, especially from new data sources. These data sets are so voluminous that traditional data processing software just can’t manage them. But these massive volumes of data can be used to address business problems you wouldn’t have been able to tackle before.

Big data gives you new insights that open up new opportunities and business models. Getting started involves three key actions:

1. **Integrate**: Big data brings together data from many different sources and applications.
2. **Manage**: Big data requires storage. Your storage solution can be in the cloud, on-premises, or both. You can store your data in any form you want and bring your desired processing requirements and necessary process engines to those data sets on an on-demand basis.
3. **Analyze**: Your investment in big data pays off when you analyze and act on your data. Get new clarity with a visual analysis of your varied data sets. Explore the data further to make new discoveries. Share your findings with others. Build data models with [machine learning](https://lms.clarusway.com/mod/lesson/view.php?id=1053) and artificial intelligence. Put your data to work.

| **https://docs.google.com/uc?id=1_xE8K8J3Z-7Yc9ag6zsaVTuuj9dNJsjl** |
| --- |
| *Big Data Sources* |

Q: What are the **three V's** of big data?  
A: **Variety**: Refers to the different data types i.e. various data formats like text, audios, videos, etc.  
**Velocity** is the rate at which data grows. Social media contributes a major role in the velocity of growing data.  
**Volume** represents the volume i.e. amount of data that is growing at a high rate i.e. data volume in Petabytes(1 Petabytes = 1024 Terabytes).

## Big Data

### Creation of Big Data

Data collection plays the most important role in the Big Data cycle. The Internet provides almost unlimited sources of data for a variety of topics. The importance of this area depends on the type of business, but traditional industries can acquire a diverse source of external data and combine those with their transactional data.

Your locations, your conversations from smartphones, weather conditions, bank account activities, stock market values, news values are some examples of big data sources.

## Big Data

### Challenges of Big Data

While big data holds a lot of promise, it is not without its challenges.

Some of these challenges:

* **Quick Data Growth**: Data growing at such a quick rate is making it a challenge to find insights from it. There is more and more data generated every second from which the data that is actually relevant and useful has to be picked up for further analysis.
* **Storage**: Such a large amount of data is difficult to store and manage by organizations without appropriate tools and technologies.
* **Syncing Across Data Sources**: This implies that when organizations import data from different sources the data from one source might not be up to date as compared to the data from another source.
* **Security**: A huge amount of data in organizations can easily become a target for advanced persistent threats, so here lies another challenge for organizations to keep their data secure by proper authentication, data encryption, etc.
* **Unreliable Data**: We can’t deny the fact that big data can’t be 100 percent accurate. It might contain redundant or incomplete data, along with contradictions.

## ``Big Data

### Big Data Use Cases

Big data can help you address a range of business activities, from customer experience to analytics. Here are just a few:

* **Product Development**: Companies like Netflix and P&G use big data to anticipate customer demand.
* **Predictive Maintenance**: Factors that can predict mechanical failures may be deeply buried in structured data, such as the year, make, and model of equipment, as well as in unstructured data that covers millions of log entries, sensor data, error messages, and engine temperature.
* **Customer Experience**: Big data enables you to gather data from social media, web visits, call logs, and other sources to improve the interaction experience and maximize the value delivered.
* **Fraud and Compliance**: Big data helps you identify patterns in data that indicate fraud and aggregate large volumes of information to make regulatory reporting much faster.
* [**Machine Learning**](https://lms.clarusway.com/mod/lesson/view.php?id=1053): You should have big data to train your [machine learning](https://lms.clarusway.com/mod/lesson/view.php?id=1053) models.
* **Operational Efficiency**: With big data, you can analyze and assess production, customer feedback and returns, and other factors to reduce outages and anticipate future demands.
* **Drive Innovation**: Big data can help you innovate by studying interdependencies among humans, institutions, entities, and processes and then determining new ways to use those insights.

| **https://docs.google.com/uc?id=1fK5f9v5fH6-dFmGOske8i1lABSLi7CL-** |
| --- |
| *Big Data application areas* |