DataBases Introduction
DataBase Foundations

Author: Eng. Carlos Andrés Sierra, M.Sc.

cavirguezs@udistrital.edu.co

Lecturer
Computer Engineer
School of Engineering
Universidad Distrital Francisco José de Caldas

2024-III

Outline

Outline

Modular Software Components

- **Software Components** are the building blocks of software systems.

- Modular Software is a software design technique that emphasizes separating the functionality of a program into independent, interchangeable modules.

  ↳ use more technologies

- Software Applications are the final product of software development.

- Software Development is the process of creating software applications.

Modular Software Components

- **Software Components** are the building blocks of software systems.
- **Modular Software** is a software design technique that emphasizes separating the functionality of a program into independent, interchangeable **modules**.
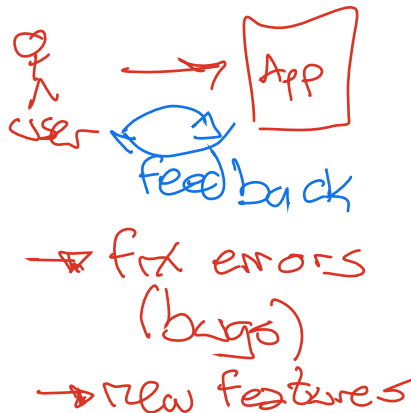- Software Applications are the final product of software development.
- Software Development is the process of creating software applications.

Modular Software Components

- **Software Components** are the building blocks of software systems.
- **Modular Software** is a software `design` `technique` that emphasizes separating the functionality of a program into independent, interchangeable **modules**..
- **Software Applications** are the final product of software development.
- Software Development is the process of creating software applications.

Modular Software Components

- **Software Components** are the building blocks of software systems.

- **Modular Software** is a software design technique that emphasizes separating the functionality of a program into independent, interchangeable **modules**.

- **Software Applications** are the final product of software development.

- **Software Development** is the process of creating software applications.

History of Data

Ubidots → IoT / 7 Petabytes/week | IoT

HISTORY OF DATA

| 19,000 BC | 1600s | 1800s | 1900s | 1990s |

The Ishango bone holds the first evidence of data collection and storage.

John Graunt introduces the concept of data analysis in 1663.

Herman Hollerith designs a machine that helped complete the US census in 1890.

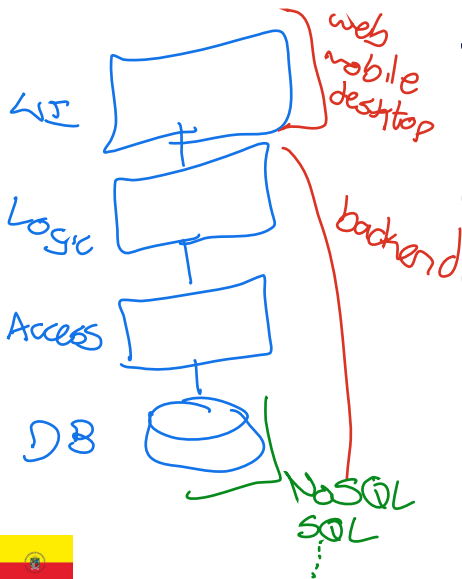Fritz Pfleumer invents the magnetic tape which later inspired the invention of floppy disks and hard disk drives.
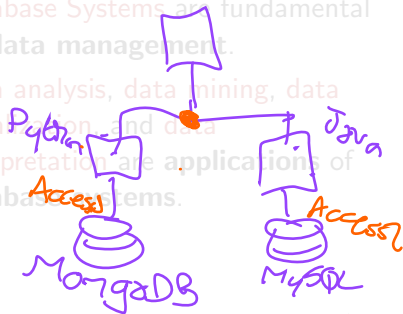
Sir Tim Berners Lee invents the World Wide Web.

Applications

- Software based on **layers** of abstraction and modularity lets implement different database strategies.

- Database Systems are fundamental for **data management**.

- Data analysis, data mining, data visualization, and data interpretation are **applications** of **database systems**.

Applications

- Software based on **layers** of abstraction and modularity lets implement different database strategies.
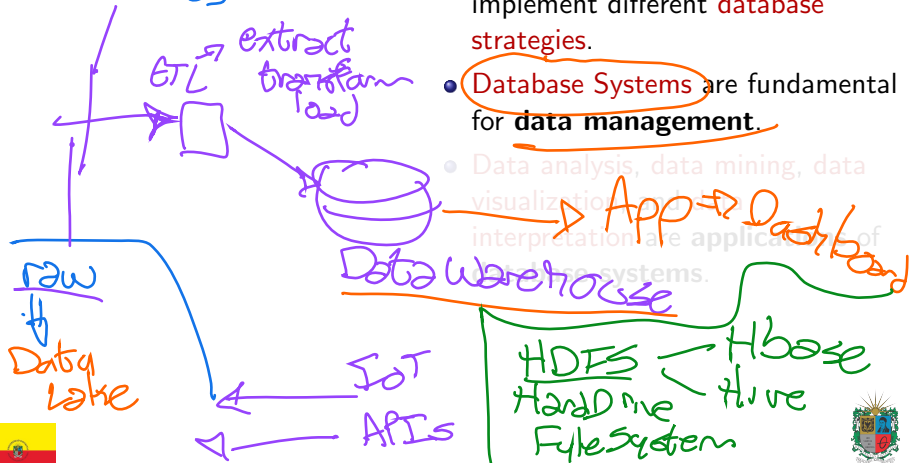- Database Systems are fundamental for **data management**.
- Data analysis, data mining, data visualization, and data interpretation are **applications** of warehouse systems.

Applications
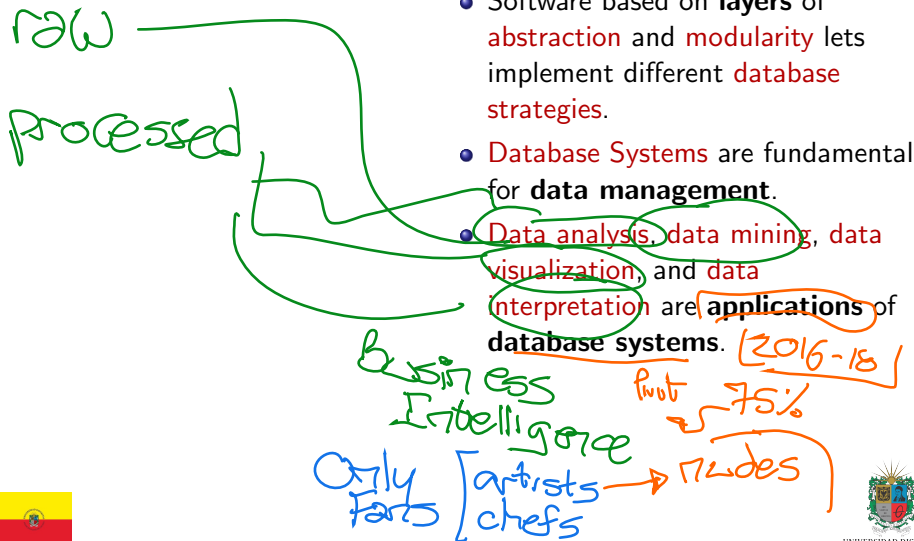
- Software based on **layers** of abstraction and modularity lets implement different database strategies.

- Database Systems are fundamental for **data management**.

- Data analysis, data mining, data visualization, and data interpretation are **applications** of **database systems**.
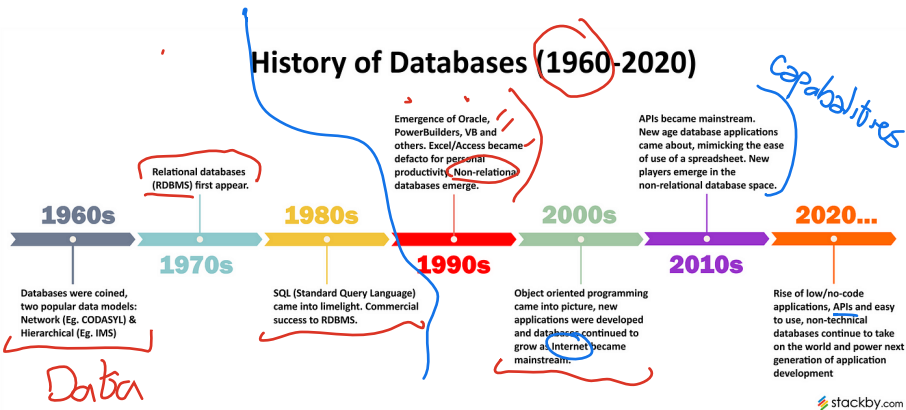
(handwritten annotations on the slide:)

raw

processed

Business Intelligence

Only Facts

artists
chefs

nodes

[2016-18]

Pivot ~ 75%

History of DataBases

History of Databases (1960-2020)

Case of Study: DataBase System

Outline

From Data to Information

- **Data**: is a set of **values** of **qualitative** or **quantitative** variables.

- Data Management: is the process of collecting, storing, **processing**, and **analyzing** data

- Data Analysis: is the process of **inspecting**, **cleansing**, **transforming**, and **modeling** data with the goal of **discovering** useful **information**, informing **conclusions**, and supporting **decision-making**

From Data to Information

- Data: is a set of **values** of **qualitative** or **quantitative** variables.

- Data Management: is the process of **collecting**, **storing**, **processing**, and **analyzing** data.

- Data Analysis: is a process of **inspecting**, **cleansing**, **transforming**, and **modeling** data with the goal of **discovering** useful **information**, informing **conclusions**, and supporting **decision-making**.

DBA ⇒ Data Base Administrator

decision making

From Data to Information

- Data: is a set of **values** of **qualitative** or **quantitative** variables.

- Data Management: is the process of **collecting**, **storing**, **processing**, and **analyzing** data.

- Data Analysis: is a process of **inspecting**, **cleansing**, **transforming**, and **modeling** data with the goal of **discovering** useful **information**, informing **conclusions**, and supporting **decision-making**.

Business Intelligence

ETLs

Structured and Unstructured Data

Structured Data *vs* Unstructured Data

SQL

NoSQL

Structured Data	Unstructured Data
Can be displayed in rows, columns and relational databases	Cannot be displayed in rows, columns and relational databases
Numbers, dates and strings	Images, audio, video, word processing files, e-mails, spreadsheets
Estimated 20% of enterprise data *(Gartner)*	Estimated 80% of enterprise data *(Gartner)*
Requires less storage	Requires more storage
Easier to manage and protect with legacy solutions	More difficult to manage and protect with legacy solutions

Tables, Columns and Rows

- **Table** is a collection of related data held in a structured format within a database.
- **Column** is a set of data values of a particular simple type, one for each row of the table.
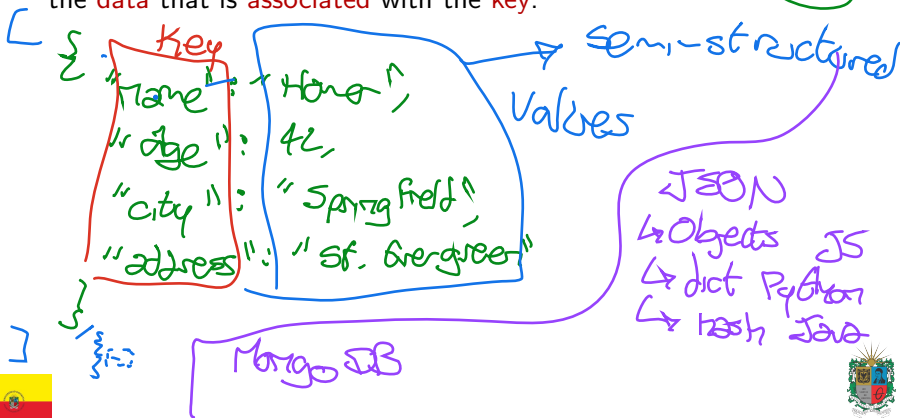- **Row** is a set of data values of a particular relationship, one for each column of the table.

Key-Value Data Structures

- **Key-Value Data Structures** are a type of data structure that can map keys to values.
- **Key** is a unique identifier for a record in a data fragment. **Value** is the data that is associated with the key.

Primary and Foreign Keys

- **Primary Key** is a unique identifier for a record in a data set.
- **Foreign Key** is a column or group of columns in a table that links to a primary key in another table.

CRUD Operations

extract

- CRUD is an acronym for **Create**, **Read**, **Update**, and **Delete**.
- Create is the process of **adding** new **records** to a **data set**.
- Read is the process of **retrieving records** from a **data set**.
- Update is the process of **modifying records** in a **data set**.
- Delete is the process of **removing records** from a **data set**.

(1) new post ⇒ Create

(2) news feed ⇒ read

(3) edit chat ⇒ update

(4) remove posts ⇒ delete

Pk, value

Name

City

Est.

Genre

FK

Outline

DataBase Classification

- **DataBase** is a collection of **data** that is organized so that it can be **easily accessed**, **managed**, and **updated**.

- Relational DataBase is a type of **database** that stores and provides access to **data points** that are **related** to one another.

- NoSQL DataBase is a type of **database** that provides a mechanism for **storage** and **retrieval** of **data** that is **modeled** in **means** other than the **tabular relations** used in **relational databases**.
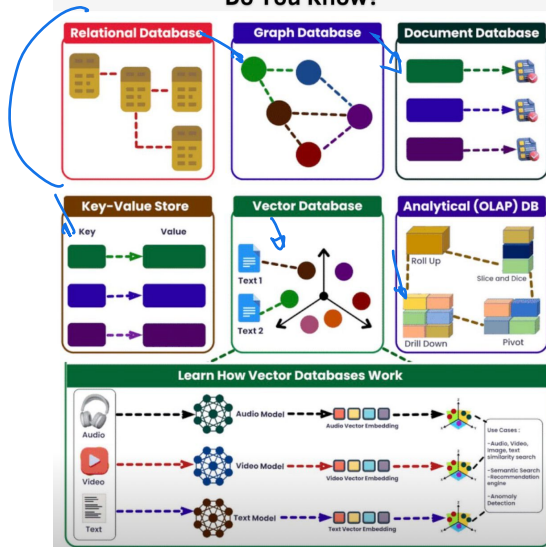
DataBase Classification

- **DataBase** is a collection of **data** that is **organized** so that it can be **easily accessed**, **managed**, and **updated**.
- **Relational DataBase** is a type of **database** that stores and provides access to **data points** that are **related** to one another.
- NoSQL DataBase is a type of **database** that provides a mechanism for **storage** and **retrieval** of **data** that is **modeled** in **means** other than the **tabular relations** used in **relational databases**.

DataBase Classification

- **DataBase** is a collection of **data** that is **organized** so that it can be **easily accessed**, **managed**, and **updated**.
- **Relational DataBase** is a type of **database** that stores and provides access to **data points** that are **related** to one another.
- **NoSQL DataBase** is a type of **database** that provides a mechanism for **storage** and **retrieval** of **data** that is **modeled** in **means** other than the **tabular relations** used in **relational databases**.

Types of Database

Outline

Entity-Relationship Model

Entity-Relationship Model is a **data model** for **describing** the **data** or **information** aspects of a **business domain** or its **processes**.
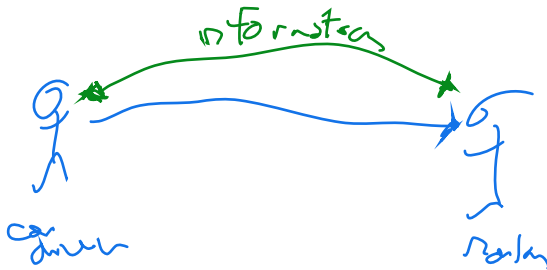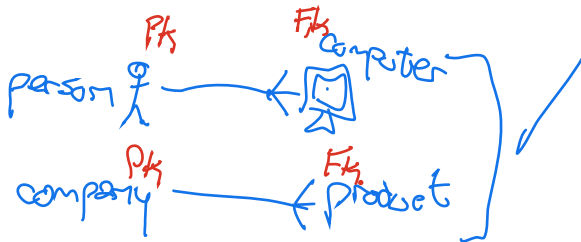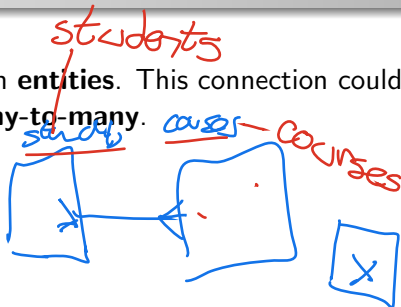
Entity Definition

Entity is a **thing** or **object** in the **real world** that is **distinguishable** from other

$$entity \approx object$$

DB   DAO

class row {
c_1;
c_2;
c_3;
c_4;
...
methods
}

DB Table $c_1 c_2 c_3 c_4$

RaO,

list of objects
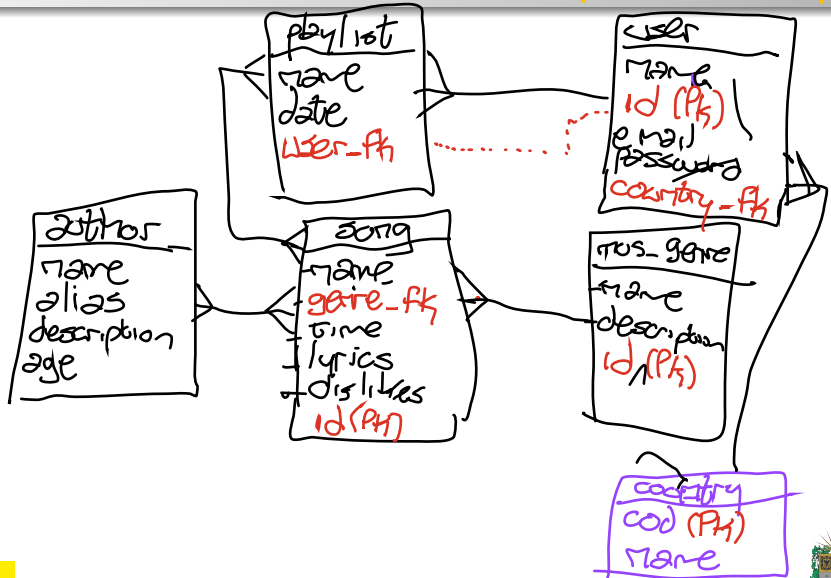
$obj1$
$obj2$
$obj3$

list of rows
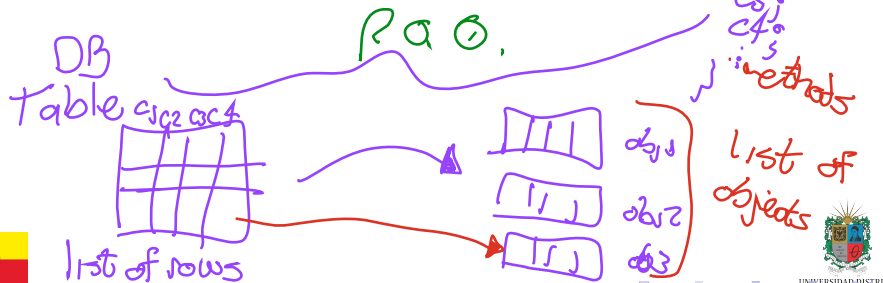
Relationship between Entities

Relationship is a **connection** between **entities**. This connection could be **one-to-one**, **one-to-many**, and **many-to-many**.

Creating our own **Espotifai**

Outline

Thanks!

Questions?

Repo: *https://github.com/EngAndres/ud-public/tree/main/courses/databases-foundations*