INTRODUCTION TO DATABASES

What will be covered today ?

- History and Origin of SQL.
- What is SQL and what does it stand for ?
- What's a database ?
- Types of Databases.
- What's a relational database ?
- What's MySQL

History of The SQL Language

- In the early 1970's IBM began developing the System R project, which aimed to create a prototype for a relational database management system (RDBMS).
- Donald D. Chamberlin and Raymond F. Boyce, two IBM researchers, developed a language called SEQUEL (Structured English Query Language) as part of the System R project. SEQUEL later evolved into SQL.
- Over the years, SQL underwent several revisions and enhancements to improve functionality and address industry needs.
- SQL gained popularity as the standard language for interacting with relational databases.
- Major database vendors, such as Oracle, IBM, Microsoft, and MySQL, implemented SQL in their database management systems, contributing to its widespread adoption.

Definition

- SQL is a tool for **organizing**, **managing, and retrieving data stored by a computer database**. The name "SQL" is an abbreviation for **Structured Query Language**.
- As the name implies, SQL is a computer language that you use to interact with a **database**.
- A **database** is an organized collection of structured information, or data, typically stored electronically in a computer system.
- To manage and interact with a database you use a **Database Management System (DBMS)**.
- A **DBMS (Database Management System**) is **a software application or system** that allows users to efficiently organize, store, retrieve, and manipulate data in a database.
- It serves as an interface between users and the database, providing a set of tools and functions to manage the data.

TYPES OF DATABASES

- There are two main types of data bases **relational databases** and **Non-Relational Databases.**
- Other types of databases include:
  1) Object-Oriented Databases (OODBMS): Object-oriented databases store data in the form of objects, similar to object-oriented programming concepts. Examples include db4o,ObjectDB , MongoDb Realm
  2) Hierarchical Databases: Hierarchical databases organize data in a tree-like structure, where each record has a parent-child relationship.IBM's Information Management System (IMS) is an example of a hierarchical database.
  3) Network Databases: Network databases are similar to hierarchical databases, but they allow more complex relationships by using a graph-like structure. Eg IDS and IDMS
  4) Columnar Databases: Columnar databases store data in columns rather than rows, optimizing data storage and query performance for analytical workloads. Examples include Vertica, Apache HBase, and Google Bigtable.

- There are several popular **DBMSs (Database Management Systems)** used in various applications and industries. Here are some examples:

1) **Oracle Database:** Oracle Database is a widely used relational database management system developed by Oracle Corporation.
2) **MySQL:** MySQL is an open-source relational database management system. It is known for its speed, scalability, and ease of use. MySQL is widely used in web applications and is the default database system for popular content management systems like WordPress.
3) **Microsoft SQL Server**: Microsoft SQL Server is a relational database management system developed by Microsoft.
4) **PostgreSQL**: PostgreSQL is a powerful open-source object-relational database management system. It emphasizes standards compliance, extensibility, and data integrity. PostgreSQL is known for its advanced features, such as support for complex queries, indexing options, and extensive data types.
5) **MongoDB**: MongoDB is a popular open-source NoSQL document database. It is designed for flexibility, scalability, and ease of development. MongoDB stores data in flexible JSON-like documents and is commonly used in modern web applications and big data environments.
6) I**BM Db2**: IBM Db2 is a family of relational database management systems developed by IBM. It provides robust data management capabilities and is known for its scalability, high availability, and performance. Db2 is used in enterprise-level applications and data warehouses.
7) **SQLite**: SQLite is a lightweight, embedded relational database management system. It is widely used in mobile and embedded applications due to its small footprint, simplicity, and zero-configuration setup.

RELATIONAL DATABASES

- This database uses Structured Query Language (SQL) for performing CRUD operations creating, reading, updating, and deleting data.
- They are the most widely used type of database.
- They organize data into tables with predefined relationships between them.
- RDBMSs(Relational Database Management Systems) use the SQL language for data manipulation and retrieval.
- Examples include Oracle, MySQL, Microsoft SQL Server,MariaDb and PostgreSQL.


ORACLE


PostgreSQL


Microsoft®
SQL Server®


MySQL


SQLite
Database


MariaDB

Non-Relational Databases

- Most databases can be categorized as either relational or non-relational.
- Non-relational databases are sometimes referred to as “NoSQL,” which stands for “**Not Only SQL”**.
- The main difference between them is how they store their information.
- A non-relational database stores data in a non-tabular form, and tends to be more flexible than the traditional, SQL-based, relational database structures.
- It does not follow the relational model provided by traditional relational database management systems but rather it uses other data structures like documents,json files, key/value pairs , Graph-like data structures.

Examples of Non-Relational Databases

- MongoDB: MongoDB is a widely used document database that stores data in **JSON-like documents**.
- Apache Cassandra: Apache Cassandra is a distributed and highly scalable **columnar database**.
- Redis: Redis is an **in-memory key-value store** that provides fast data retrieval and caching. It supports a wide range of data structures, including strings, lists, sets, and hashes.
- Neo4j: Neo4j is a **graph database** that stores data in nodes and relationships (edges). It is designed for handling highly interconnected data and complex relationships.
- Apache HBase: Apache HBase is a **distributed columnar database** that runs on top of Hadoop.
- Amazon DynamoDB: DynamoDB is a fully managed NoSQL database service provided by Amazon Web Services (AWS). **It is a key-value store** with high scalability, automatic scaling, and low latency.

Amazon
DynamoDB

Apache HBase

redis

Apache
CASSANDRA

MySQL

- MySQL is an open-source relational database management system (RDBMS) that is widely used for managing and storing structured data.
- It was developed by a Swedish company called MySQL AB, which was later acquired by Sun Micro systems and subsequently by Oracle Corporation.
- MySQL is known for its performance, scalability, ease of use, and robustness
- MySQL employs a client-server architecture, where multiple client applications can connect to a central MySQL server to access and manipulate data.
- The server handles data storage, retrieval, and management, while clients interact with the server to perform operations on the data.
- MySQL includes various tools and utilities for database administration, configuration, and monitoring. These tools simplify tasks such as database setup, performance tuning, backup and recovery, and managing user accounts.

How do Relational Databases Work ?

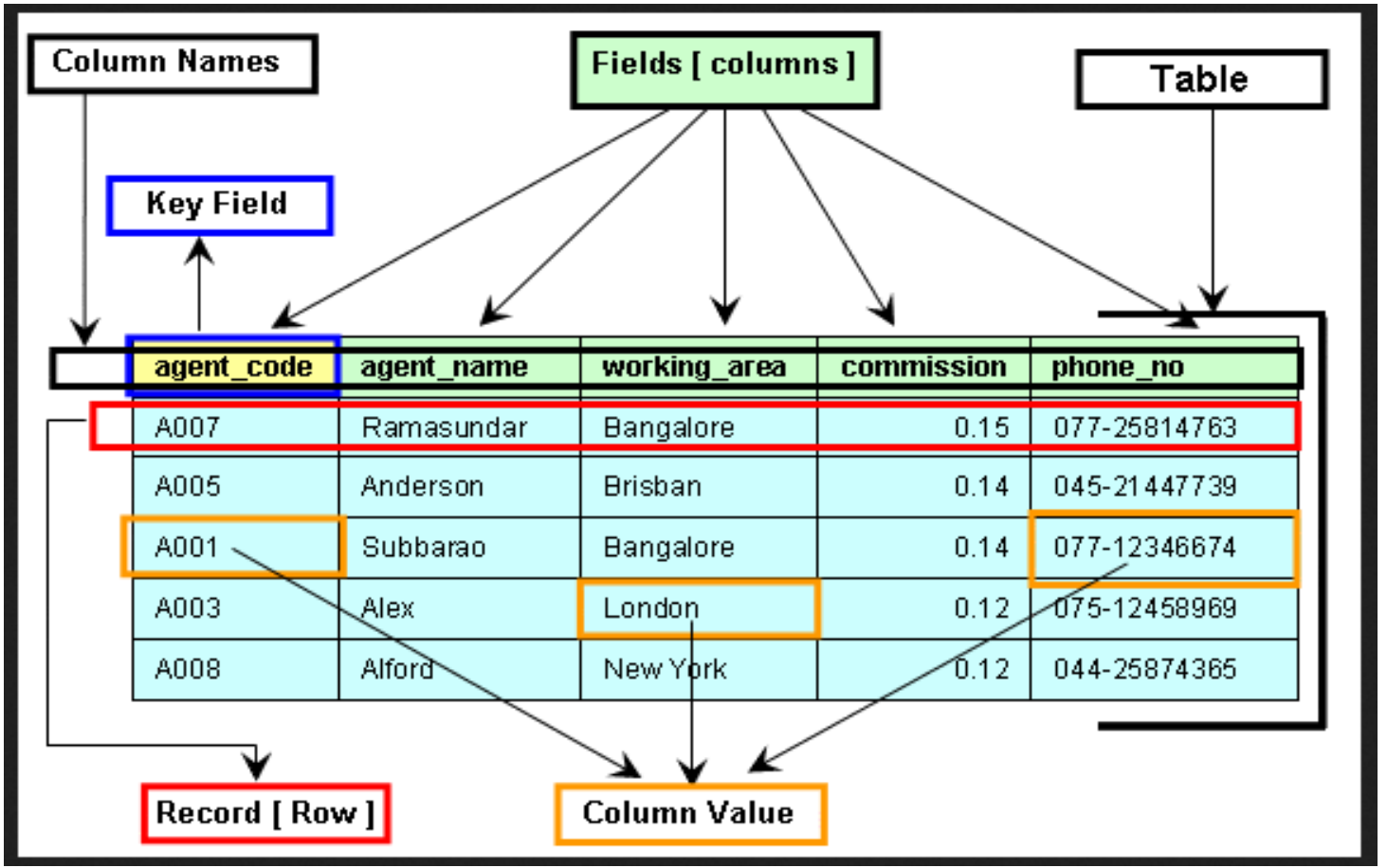
- Relational databases work based on the relational model, which organizes data into many tables and establishes relationships between the tables based on the data in the tables.
- The structure and organization of the database is known as **schema**.
- Schema includes the tables, columns and data types, relationships, and constraints that govern the data stored in the database.
- To do this effectively we use **Queries.**
- A **query** refers to a request or instruction to retrieve specific information from a database.
- Queries are used to interact with a database management system (DBMS) and extract data based on specific criteria or conditions. They allow users to retrieve, manipulate, filter, and analyze data stored within a database.

Tables

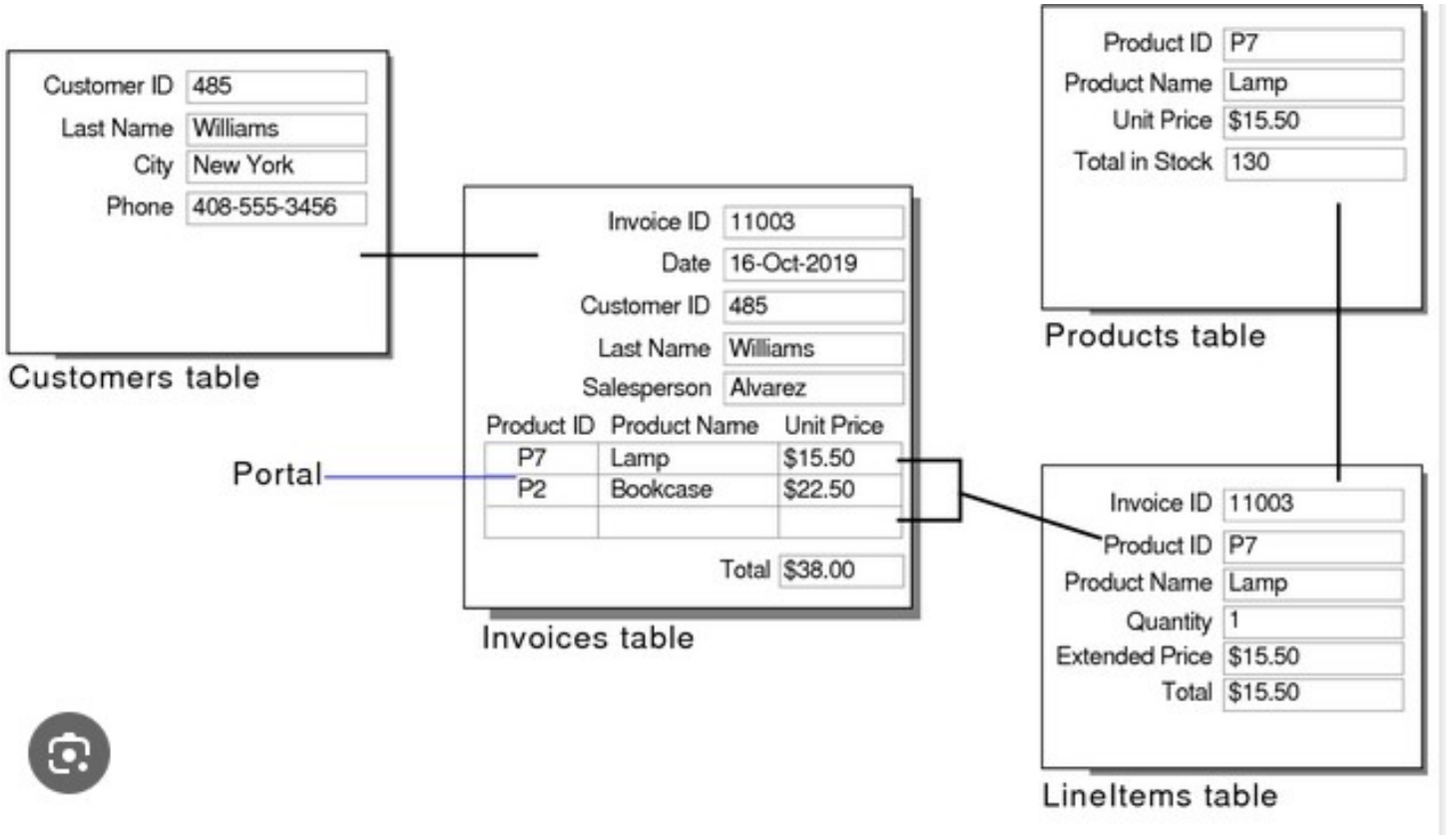
- The organizing principle in a relational database is the table, a rectangular, row/column arrangement of data values.
- Each table in a database has a unique table name that identifies its contents. The **table schema defines the structure** and organization of the table, specifying the columns, data types, constraints, and relationships
- The components of a table in more detail:

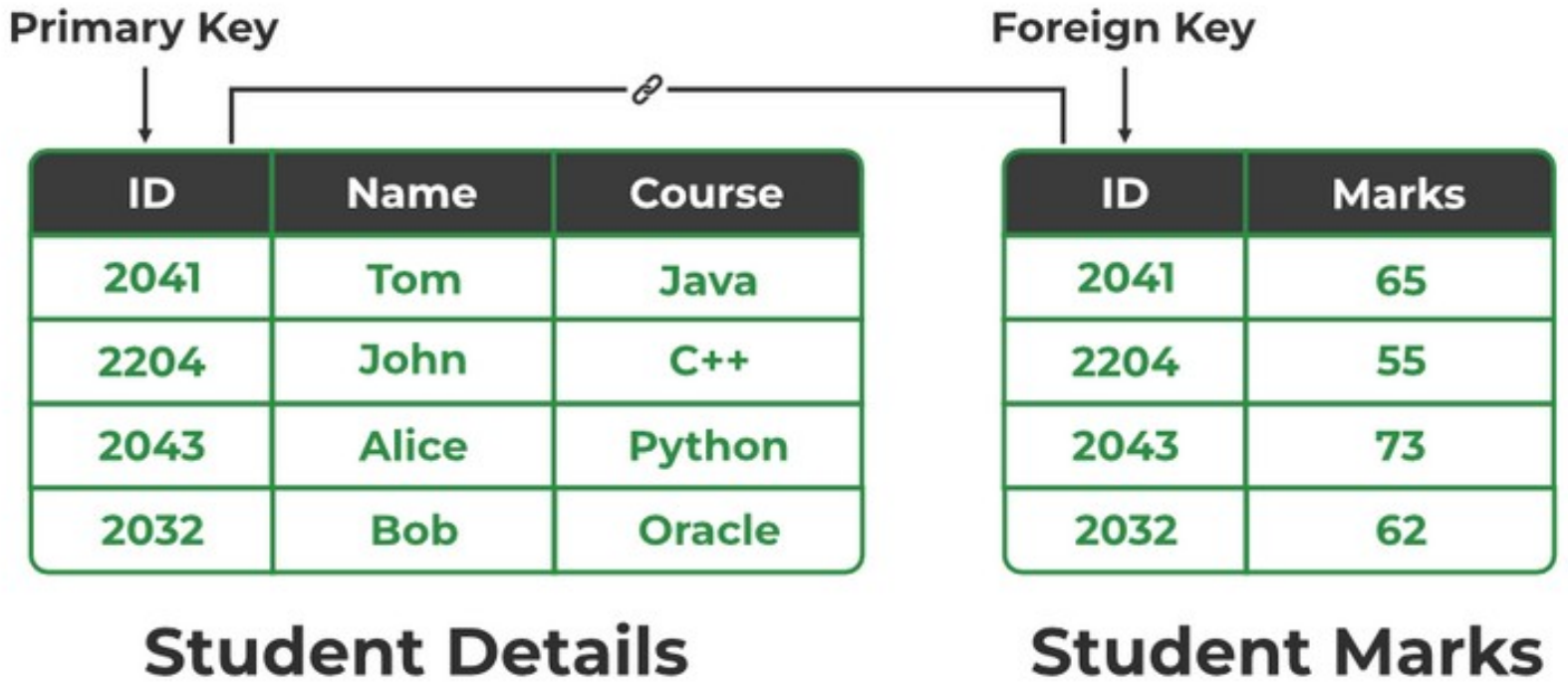
1) Rows (Records):
   A row, also referred to as a record or tuple, represents a single instance of data stored in the table. Each row contains specific data values that correspond to the attributes or columns defined in the table's schema.
2) Columns (Attributes):
   Columns, also known as attributes or fields, represent the individual data elements associated with the entities or concepts represented by the table. Each column has a specific name and data type, defining the kind of data it can store, such as strings, numbers, dates, or binary data.
3) Primary Key:
   A primary key is a column or a combination of columns that uniquely identifies each row in the table.

   It ensures the uniqueness and integrity of data within the table, preventing duplicate entries and serving as a reference for establishing relationships with other tables. Examples of primary keys include unique identifiers like employee IDs, customer IDs, or order IDs.
4) Foreign Key:
   A foreign key is a column or a set of columns in a table that references the primary key of another table. It establishes relationships or connections between tables, allowing data to be linked across tables.
   Foreign keys enable the representation of complex data relationships in a relational database by connecting related entities.
5) Data Types:
   Each column in a table has a specific data type that determines the kind of data it can store.
   Data types can include integers, floating-point numbers, strings, dates, Boolean values, or binary data, among others.


Column Names
Fields [columns]
Table
Key Field
agent_code
agent_name
working_area
commission
phone_no
A007
Ramasundar
Bangalore
0.15
077-25814763
A005
Anderson
Brisban
0.14
045-21447739
A001
Subbarao
Bangalore
0.14
077-12346674
A003
Alex
London
0.12
075-12458969
A008
Alford
New York
0.12
044-25874365
Record [Row]
Column Value

Example of many Tables in a Relational Database




Primary Key
Foreign Key
ID
Name
Course
2041
Tom
Java
2204
John
C++
2043
Alice
Python
2032
Bob
Oracle
ID
Marks
2041
65
2204
55
2043
73
2032
62
Student Details
Student Marks

SQL Syntax

- Most of the actions you need to perform on a database are done with SQL statements.
- SQL statements consists of keywords that are easy to understand.
- SQL keywords are NOT case sensitive: select is the same as SELECT
- Some database systems require a semicolon at the end of each SQL statement.
- Semicolon is the standard way to separate each SQL statement in database systems that allow more than one SQL statement to be executed in the same call to the server.

Some of The Most Important SQL Commands

- SELECT - extracts data from a database
- UPDATE - updates data in a database
- DELETE - deletes data from a database
- INSERT INTO - inserts new data into a database
- CREATE DATABASE - creates a new database
- ALTER DATABASE - modifies a database
- CREATE TABLE - creates a new table
- ALTER TABLE - modifies a table
- DROP TABLE - deletes a table
- CREATE INDEX - creates an index (search key)
- DROP INDEX - deletes an index

There are 5 main types of commands:

- DDL (Data Definition Language) commands
- DML (Data Manipulation Language) commands
- DCL (Data Control Language) commands
- Transaction Control Language(TCL) commands
- Data Query Language(DQL) commands.

Data Definition Language

- DDL is the short name for Data Definition Language, which deals with database schemas and descriptions, of how the data should reside in the database.
  - CREATE: to create a database and its objects like (table, index, views, store procedure, function, and triggers)
  - ALTER: alters the structure of the existing database
  - DROP: delete objects from the database
  - TRUNCATE: remove all records from a table, including all spaces allocated for the records are removed
  - COMMENT: add comments to the data dictionary
  - RENAME: rename an object

Data Manipulation Language

- DML is the short name for Data Manipulation Language which deals with data manipulation and it is used to store, modify, retrieve, delete and update data in a database.
- **Data query language(DQL)** is the subset of “Data Manipulation Language”.
- The most common command of DQL is SELECT statement. SELECT statement help on retrieving the data from the table without changing anything in the table.
  - SELECT: retrieve data from a database
  - INSERT: insert data into a table
  - UPDATE: updates existing data within a table
  - DELETE: Delete all records from a database table
  - MERGE: UPSERT operation (insert or update)
  - CALL: call a PL/SQL or Java subprogram
  - EXPLAIN PLAN: interpretation of the data access path
  - LOCK TABLE: concurrency Control

Data Control Language

- DCL is short for Data Control Language which acts as an access specifier to the database.(basically to grant and revoke permissions to users in the database
- GRANT: grant permissions to the user for running DML(SELECT, INSERT, DELETE,…) commands on the table
- REVOKE: revoke permissions to the user for running DML(SELECT, INSERT, DELETE,…) command on the specified table

Transactional Control Language

- TCL is short for Transactional Control Language which acts as an manager for all types of transactional data and all transactions.
- Some of the commands of TCL are
  - Roll Back: Used to cancel or Undo changes made in the database
  - Commit: It is used to apply or save changes in the database
  - Save Point: It is used to save the data on the temporary basis in the database