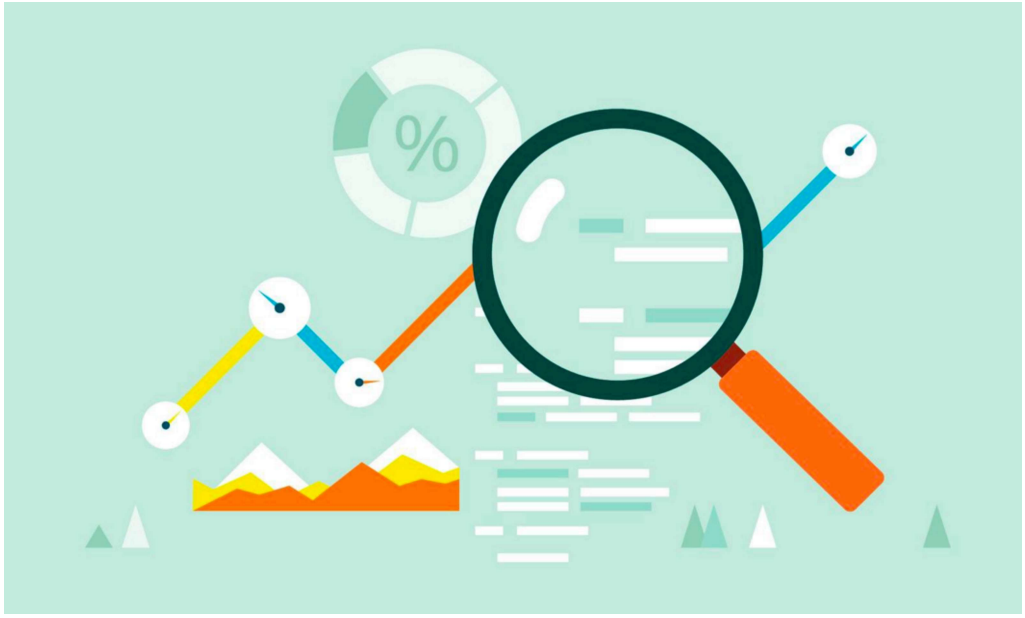
Data Science



Databases & SQL

What is Data

- Data is simply a collection or set of values in the form of text, words, numbers, pictures, audio or video.
- It is most important asset for any organization.

How to store data ?

Data Storage

- Traditional way - on papers
- File system
- Databases

File system

- Drawbacks -
  - Data redundancy
  - Inconsistency
  - No concurrent access
  - No relationship between multiple files
  - No backup and recovery
  - And many others

Database

- It is an organized collection or repository of our data
- It is a collection of interrelated data which helps in efficient retrieval, insertion and deletion of data.
- Users can perform different queries in order to perform actions based on their requirement

DBMS

- DBMS = Database Management System
- DBMS is an application which is used to maintain or manage databases
- Eg. Oracle, sybase, Microsoft SQL server, PostgreSQL and many others are there

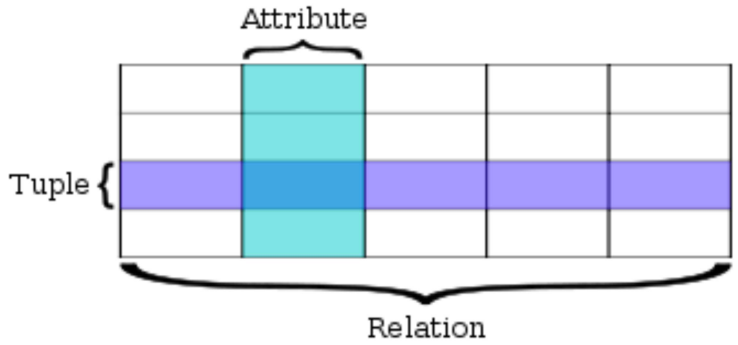
Types of DBMS

- Hierarchical DBMS
- Network DBMS
- Relational DBMS
- Object oriented DBMS

Relational Databases

Relational Database

- A relational database is a specific type of database that stores everything in relations or tables.
- Tables have different row and columns

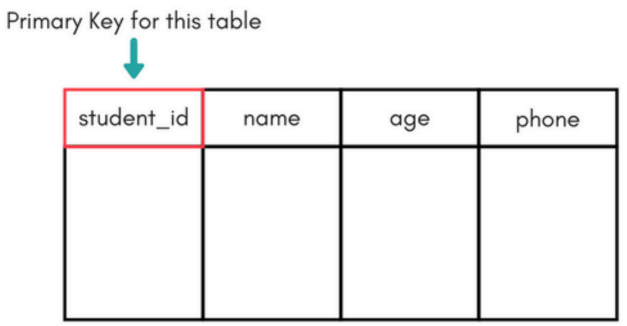


RDBMS

- A software system which is used to maintain relational databases is termed as relational database management system (RDBMS).

Primary Key

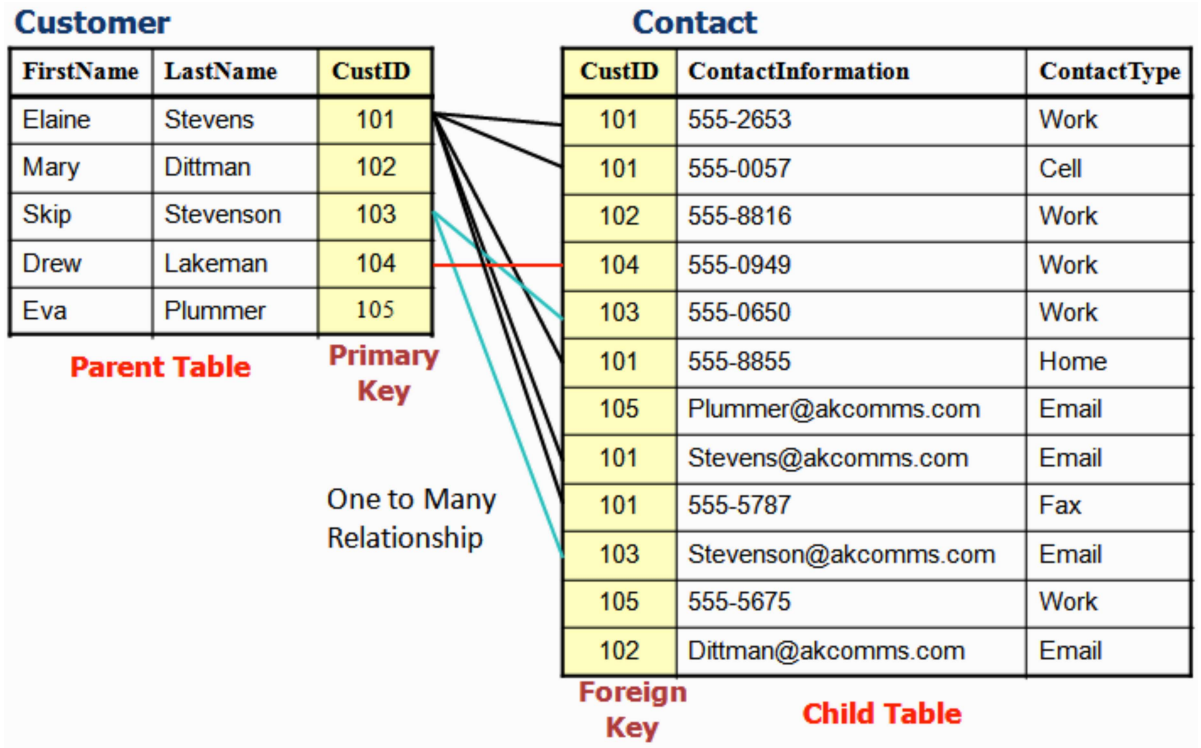
- Primary key is a minimal set of columns or attributes that uniquely identifies row in a table.
- A primary key's main features are -
  - It must contain a unique value for each row of data.
  - It cannot contain null values.



Foreign Key

- Foreign keys are columns that point to or that matches the primary key columns in other tables.
- The foreign key can be used to cross-reference tables. Foreign keys do not need to have unique values in the referencing relation.

Foreign Key


Customer
Contact
FirstName
LastName
CustID
Elaine
Stevens
101
Mary
Dittman
102
Skip
Stevenson
103
Drew
Lakeman
104
Eva
Plummer
105
Parent Table
Primary Key
One to Many Relationship
CustID
ContactInformation
ContactType
101
555-2653
Work
101
555-0057
Cell
102
555-8816
Work
104
555-0949
Work
103
555-0650
Work
101
555-8855
Home
105
Plummer@akcomms.com
Email
101
Stevens@akcomms.com
Email
101
555-5787
Fax
103
Stevenson@akcomms.com
Email
105
555-5675
Work
102
Dittman@akcomms.com
Email
Foreign Key
Child Table

SQL

SQL

SQL

- Structured query language
- Language used to interact with relational database
- An SQL query is how you access the data. Using an SQL query, you can create and delete, or modify tables, as well as select, insert, and delete data from existing tables.

Example - RDBMS

- Examples of popular RDBMSs -
  - SQLite
  - MySQL
  - PostgreSQL
  - Oracle DB
  - SQL Server

SQL - Basic Commands

Basic commands

- Show all databases
  - *Show databases;*
- Create database
  - *create database database_name*
- Use database
  - *use database_name*
- List all tables inside a database
  - *show tables*

Basic commands

- Create a new table
  - *create table table-name (*
    column_name1 datatype1,
    column_name2 datatype2,
    column_name3 datatype3,
    column_name4 datatype4
    );
  - Create table with different constraints
    - Primary Key
    - Not null
    - With some default value

Basic commands

- Describe a table
  - *describe table_name* or
    desc table_name
- Insert data in a table
  - *insert into table-name values(data1,data2,..);*
  - *insert INTO table_name (column1, column2, column3, ...)*
    VALUES (value1, value2, value3, ...);

Update Table

Alter - Add column

- Add one column to an existing table
  - *alter table table_name add(column_name datatype);*
- Add multiple columns
  - *alter table table_name add(column_name1 datatype1, column_name2 datatype2, column_name3 datatype3);*
- Add column with default value
  - *alter table table_name add(column_name1 datatype1 default data);*

Alter - Modify column

- Change data type of an existing column
  - *alter table table_name modify column_name datatype;*
- Rename a column
  - *alter table table_name change old_column_name new_column_name datatype;*
- Delete column
  - *alter table table_name drop column column_name;*
  - *alter table table_name drop column column_name1, drop column column_name1, ...;*

Retrieve Data

Select

- This command allows us to retrieve the specific information as per our requirement from a relational database. It returns a result set of records from one or more tables.
- Different variations
  - Select *
  - Select one column
  - Select multiple columns

Load data

- Dataset source : http://www.mysqltutorial.org/mysql-sample-database.aspx
- Load data from file
  - *source 'file_name.sql'*

Restrict result set

- Limit
- Distinct

Filter Result set

where

- Using this WHERE clause, we can specify a selection criteria to select the required records from a table.
- The WHERE clause works like an if condition in any programming language.
- You can specify any condition using different operators -
  - Relational operators
    - >, <, =, <=, >=
  - Logical operators
    - AND, OR
  - Is Null and is not Null

Aggregate Functions

Aggregate Functions

- Aggregate functions perform a calculation on a set of values and return a single value.
- Different functions
  - COUNT
    - count(*)
    - count(column_name)
    - count(distinct column_name)
  - AVG
  - MAX
  - MIN
  - SUM

Update & Delete

Update and Delete

- Update row of a table
  - *UPDATE table_name set column_name = 'column_value';*
  - *UPDATE table-name set column_name = 'column_value' where condition;*
- Delete
  - *DELETE from table_name where condition;*
- Truncate
  - *TRUNCATE table_name*
- Drop
  - *DROP table table_name*