Data Science

# Databases & SQL

## What is Data

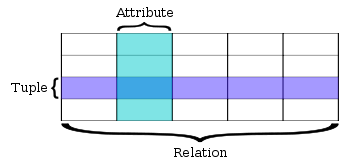
* Data is simply a collection or set of values in the form of text, words, numbers, pictures, audio or video.
* It is most important asset for any organization.

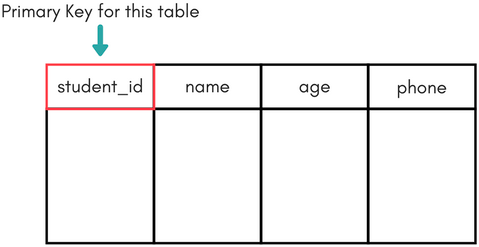
# How to store data ?

* Traditional way - on papers
* File system
* Databases
* Drawbacks -
  + Data redundancy
  + Inconsistency
  + No concurrent access
  + No relationship between multiple files
  + No backup and recovery
  + And many others
* 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 = 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
* Hierarchical DBMS
* Network DBMS
* Relational DBMS
* Object oriented DBMS

# Relational Databases

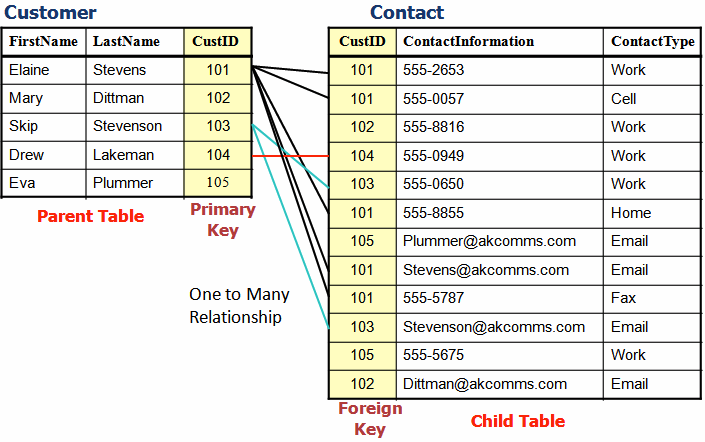
* A relational database is a specific type of database that stores everything in relations or tables.
* Tables have different row and columns



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



# 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.
* Examples of popular RDBMSs -
  + SQLite
  + MySQL
  + PostgreSQL
  + Oracle DB
  + SQL Server

# SQL - 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*
* 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
* 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

* 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);*
* 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

* 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
* Dataset source : <http://www.mysqltutorial.org/mysql-sample-database.aspx>
* Load data from file
  + *source ‘file\_name.sql’*
* 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*