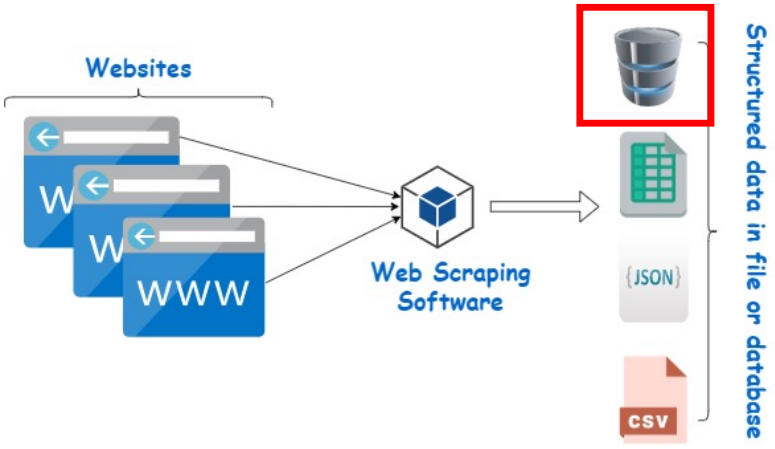
Introduction to Database

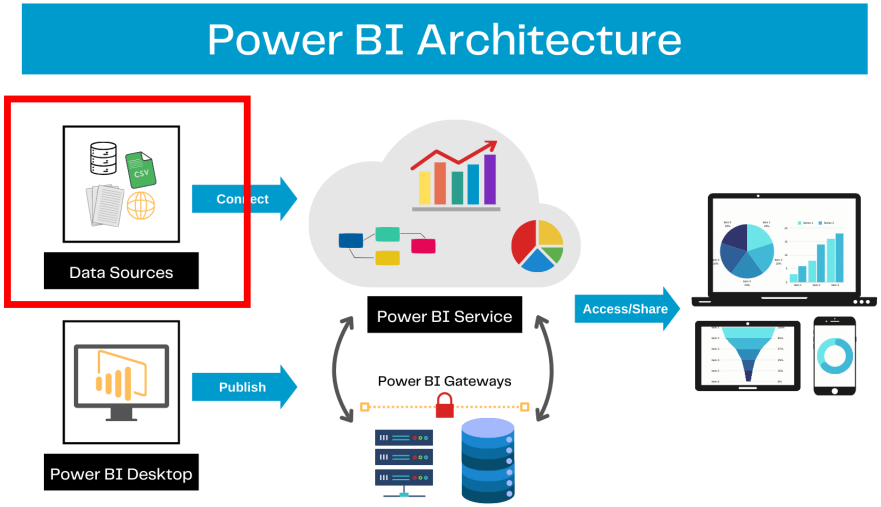
Outlines

- Introduction to DB
- SQL commands

Introduction: Movitation

- Where do you save the scraped data?
- Where do you source the data for your BI dashboard?
- Where do you query to uncover business insights?
- Answer: Database (DB)





Introduction: What is database?

- Database – collection of persistent data
- Database Management System (DBMS) – software system that supports creation, population, and querying of a database
  - MySQL
  - SQLite
  - Microsoft SQL Server
  - Oracle
  - PostgreSQL
  - Etc.

Relational Database: Structured DB

- Relational Database Management System (RDBMS)
  - Consists of a number of *tables* and single *schema* (definition of tables and attributes)
  - Students (sid, name, login, age, gpa)
    - **Students** identifies the table
    - **sid, name, login, age, gpa** identify attributes
    - **sid** is primary key (PK)

sid	name	login	age	gpa
50000	Dave	dave@cs	19	3.3
53666	Jones	jones@cs	18	3.4
53688	Smith	smith@ee	18	3.2
53650	Smith	smith@math	19	3.8
53831	Madayan	madayan@music	11	1.8
53832	Guldu	guldu@music	12	2.0

Why are RDBMS useful?

- Data independence – provides abstract view of the data, without details of storage
- Efficient data access – uses techniques to store and retrieve data efficiently → SQL commands
- Reduced application development time – many important functions already supported
  - Centralized data administration
  - Data Integrity and Security
  - Concurrency control and recovery

Another example: Courses

- Courses (cid, instructor, quarter, dept)

cid	instructor	quarter	dept
Carnatic101	Jane	Fall 06	Music
Reggae203	Bob	Summer 06	Music
Topology101	Mary	Spring 06	Math
History105	Alice	Fall 06	History

Keys

- Primary key – minimal subset of fields that is unique identifier for a tuple
  - sid is primary key for Students
  - cid is primary key for Courses
- Foreign key – connections between tables
  - Courses (cid, instructor, quarter, dept)
  - Students (sid, name, login, age, gpa)
  - How do we express which students take each course?
    - Enrolled(**sid, cid**, grade)

Enrolled

Foreign key

cid	grade	**sid**
Carnatic101	C	53831
Reggae203	B	53832
Topology112	A	53650
History 105	B	53666

Student

sid	name	login
50000	Dave	dave@cs
53666	Jones	jones@cs
53688	Smith	smith@ee
53650	Smith	smith@math
53831	Madayan	madayan@music
53832	Guldu	guldu@music

Example Schema (simplified)

- Courses (cid, instructor, quarter, dept)
- Students (sid, name, gpa)
- Enrolled (cid, sid, grade)

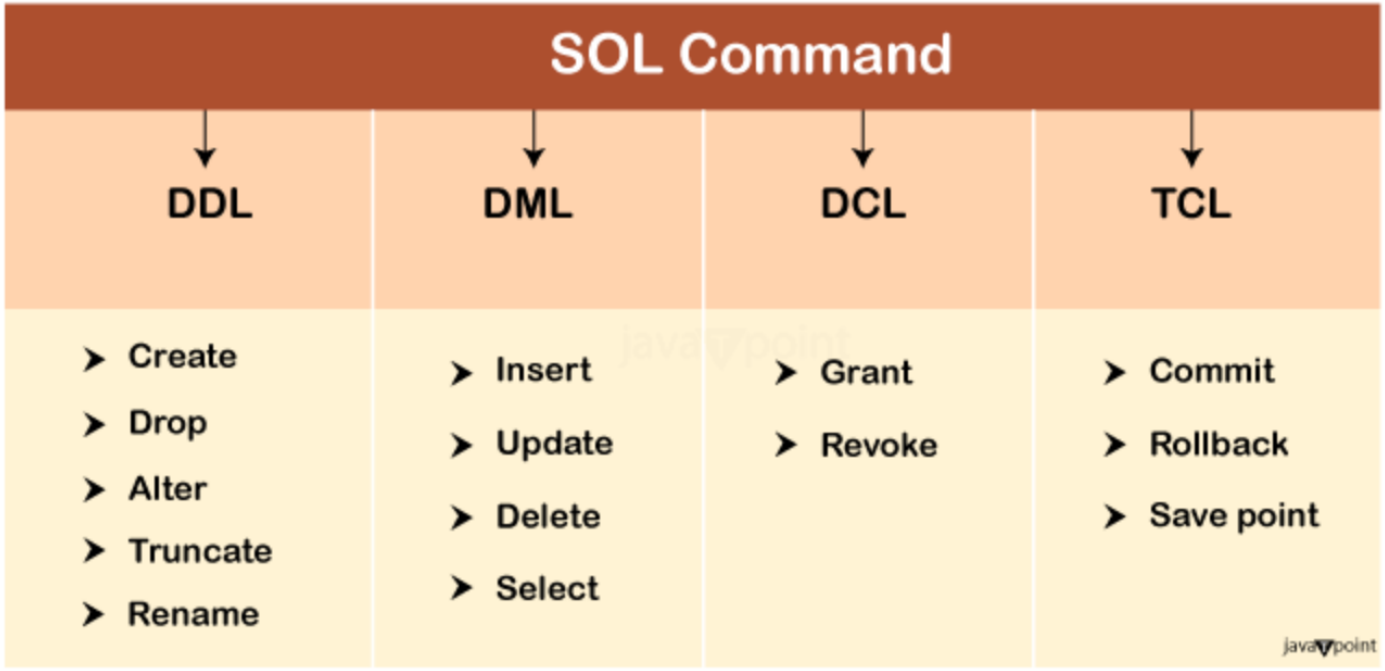
SQL commands

- SQL commands are instructions. It is used to communicate with the database. It is also used to perform specific tasks, functions, and queries of data.
- SQL can perform various tasks like create a table, add data to tables, drop the table, modify the table, set permission for users.
- CRUD: Create, read (select), update and delete

https://www.javatpoint.com/dbms-sql-command

Types of SQL commands

- Data Definition Language (DDL)
- Data Manipulation Language (DML)
- Data Control Language (DCL)
- Transaction Control Language (TCL)



Intro to SQL

- CREATE TABLE
  - Create a new table, e.g., students, courses, enrolled
- SELECT-FROM-WHERE
  - List all CS courses
- INSERT
  - Add a new student, course, or enroll a student in a course

Create Table

- CREATE TABLE Enrolled
  (sid CHAR(20),
  cid CHAR(20),
  grade CHAR(20),
  PRIMARY KEY (sid, cid),
  FOREIGN KEY (sid) references Students)

Select-From-Where query

- “Find all students who are under 18”

```
SELECT *
FROM   Students S
WHERE  S.age < 18
```

Queries across multiple tables (joins)

- “Print the student name and course ID where the student received an ‘A’ in the course”

```
SELECT S.name, E.cid
FROM    Students S, Enrolled E
WHERE  S.sid = E.sid AND E.grade = ‘A’
```

Other SQL features

- MIN, MAX, AVG
  - Find highest grade in fall database course
- COUNT, DISTINCT
  - How many students enrolled in CS courses in the fall?
- ORDER BY, GROUP BY
  - Rank students by their grade in fall database course