

Databases and Information Systems

CS303

Overview and Introduction
02-08-2023

- Course Instructor: Anantha Padmanabha
- Email Id : ananthap@iitdh.ac.in
- Office : 2A7

- TA :
 - Vivek Shahare 212011006@iitdh.ac.in
 - Bonthu Vyuhita 212011001@iitdh.ac.in

Course Overview (CS303)

- Relational Databases and SQL

- SQL queries, Entity-Relationship diagram, Normalization, Transactions, Security, Query Optimization,

- NoSQL Databases

- Overview and properties
- Graph Databases: Queries in Cypher (Neo4J)
- Key-Value Pair Databases, Document Databases, Column Databases

- Decision support, Data mining, Data Warehousing, Information retrieval

- Special Databases

- Spatial, Temporal, Ontology based, RDF, Knowledge Base

References

- Database System Concepts by Silberschatz, Korth and Sudarshan
- NoSQL for mere mortals by Dan Sullivan
- Slides and in-class discussions

Evaluation

- Mid-semester Exam : 30%
- End semester Exam : 50%
- Quiz (2) : 10%
- Assignment : 10%

Introduction

What is Database Management?

- Provide way to **store and retrieve data** in a **convenient and efficient** manner.
- Computerized management of high volumes of data
- **Application:**
 - **Marketing and Enterprises**
 - Sales, Accounting, Human resources, Manufacturing ...
 - **Banking and Finance**
 - Customer information, Transactions, Finance, Employees ...
 - **Universities**
 - Student information, Course registration, Grades, Faculty, Departments ...
 - **Airlines**
 - Reservation system, Flight plans ...
 - **Telecommunications, Hospitals**

Before Database Systems

- Before DBMS, Operating systems used to store data as system files.
(Same as saying store everything in a **text files**)
- **Disadvantages:**
 - **Data redundancy and Inconsistency :** Duplication of data and disagreement over same data
 - **Difficult to access :** Searching some information from the data
 - **Data Isolation :** Different file formats of different files
 - **Integrity problems :** Check some conditions (like balance of customer > 0)
 - **Atomicity problems :** Restoring consistency on failure
 - **Concurrent access problems :** Multiple people accessing the same data
 - **Security problems :** Control of access to data

Database System

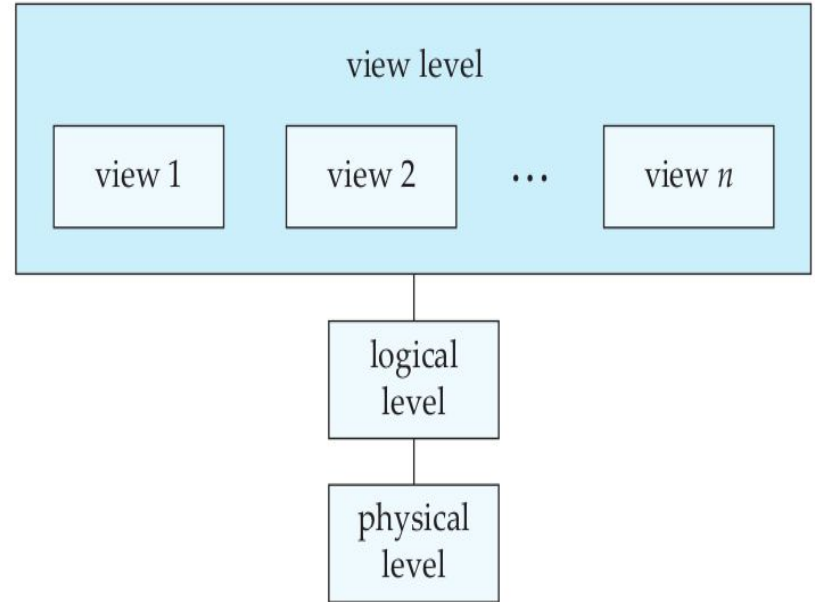
- Collection of interrelated data and programs that allow users to access and modify the data
- Functions:
 - Store and Retrieve data
 - Modify data
 - Ensure consistency
 - Security
 - Transaction Management
 - Failure handling

Data Abstraction

- Efficient retrieval needs complex data structures for storage
(Needs to be hidden from users)
- Not all data should be accessible to everyone
- These requirements lead to layered approach where the implementations are hidden from each other (Data Abstraction)

Data Abstraction

- Hide the complexity of the system through various levels to simplify user's interaction with the system.
- Three levels of Data abstraction in Database systems:
 - **Physical Level** : How data is actually stored. Describes low level data structures
 - **Logical Level** : What data is stored and what is the relationship between them. User of Logical Level need not know the details of Physical Level
(Physical data independence)
 - **View Level** : For end users. Describes part of the database.



Instances and Schemas

- Databases change over time
- Collection of information stored in the database at a particular moment is called an **Instance**
- The overall design of the database is called the Database Schema
(Does not change under usual circumstances)
 - **Physical Schema** : Describes database design at physical level
 - **Logical Schema** : Describes database design at the logical level
 - Many schemas at View level called **subschemas**

Database Languages

- **Data Definition Language** : Specifies database schema
 - Domain constraints (DC)
 - Referential Integrity constraints (RIC)
 - Assertions (constraints that cannot be specified as DC or RIC)
 - Authorization
- **Data Manipulation Language** : Express database queries and updates
 - Retrieve information from database
 - Add new data / Delete data / Modify existing data
 - Two type of DML
 - **Procedural** : User describes what data is needed and how to get that data
 - **Declarative** : User specifies what data is needed without specifying how to get it

Query

- Statement requesting retrieval of information.
- Common to assume Query and Data Manipulation Language are same
- Data Abstraction is applied at every level:
 - At Physical level : Define algorithm that allows efficient access to data
 - Higher levels : Ease of Use for humans to interact with the system.
- Query processor translates Data Manipulation Language queries into sequence of actions at the physical level of the database system

How to store data?

- Text
- Tables (Relational Database Architecture)
- Graphs
-

Reference:

Database System Concepts by Silberschatz, Korth and Sudarshan
Chapter 1