Database Management System

B. Tech. II CSE (Sem-4th)

CS204

Credits: 5 (L:3 T:1 P:2)

Covering

- Data
- Need of Database
- Database
  - Collection of related data (recorded facts)
- DBMS
  - Generalized software package for implementing and maintaining a computerized database
  - Provides many services to manage data resources
- Categories of database users
- When Not to Use a DBMS

Data

- A raw fact or unorganized form (such as alphabets, numbers, or symbols) that refers to or represent, condition, ideas or objects
- Limitless and present everywhere in the universe
- e.g. 13, Harsh, 395007

Information

Data which

- Verified to be accurate and timely
- Specific and organized for a purpose
- Represented within a context that gives it meaning and relevance
- Lead to an increase in understanding and decrease in uncertainty is called information
- e.g. Roll No=13, First Name=Harsh, Pincode=395007

Need of DBMS

- Earlier existing traditional **file processing**
  - For a given problem, each user defines and implements the files needed for a specific software application
  - As the application base grows
    - Many shared files
    - Multitude of file structures
    - Need to exchange data among applications

Need of DBMS

- Earlier existing traditional **file processing** Issues
  - Redundancy: multiple copies
  - Inconsistency: independent updates
  - Inaccuracy: concurrent updates
  - Incompatibility: multiple formats
  - Insecurity
  - In-auditability: poor chain of responsibility
  - Inflexibility: changes are difficult to apply

Drawbacks of Traditional File Processing

- Atomicity of updates
  - Failures may leave database in an inconsistent state with partial updates carried out
  - Example: Transfer of funds from one account to another **should either complete or not happen at all**
- Concurrent access by multiple users
  - Concurrent access needed for performance
  - Uncontrolled concurrent accesses can lead to inconsistencies
    - Example: Two people reading a balance (say 100) and updating it by withdrawing money (say 50 each) at the same time
- Security problems
  - Hard to provide user access to some, but not all, data

File Management System

- A collection of programs that manage and store data in files and folders in a computer hard disk
- Manages the way of reading and writing data to the hard disk
- Also known as conventional **file system**
- Actually stores data in the isolated files which have their own physical location on the drive, and users manually go to these locations to access these files
- The easiest way to store the data like text, videos, images, audios, etc. in general files
- Data redundancy is high in file management system, and it cannot be controlled easily
- Data consistency is not met, and the integration of data is hard to achieve
- Operating System such as **Linux** and **Windows** has its own file system
  - e.g. **NTFS** is the Windows file system, and **EXT** is the Linux file system
- Provide less security to these files where they have options such as hide files, locks, and sharing on files

Database Approach

- Single repository maintains data that is defined once and then accessed by various users
- **Offer solutions to problems by Abstraction, Schema, Data Models**
- Addresses a variety of problems
  - No Redundancy: no multiple copies
  - Consistency
  - Accuracy: concurrent updates
  - Compatibility: multiple formats
  - Security: proliferation
  - Auditability: chain of responsibility
  - Flexibility: changes are easy to apply

Characteristics of the Database Approach

- Programs isolated from data through **abstraction**
  - Does not expose details of how (or where) data is stored or how operations are implemented
  - Data sharing through multiple **views**
- Multiuser **transaction** processing
  - Encapsulates sequence of operations to behave atomically
- Data is **self-defining**
  - Database system contains complete definition of structure and constraints as *meta-data*
  - Database catalog used by
    - DBMS software
    - Database users who need information about database structure

Databases/Applications to manage Data

- Traditional database applications
  - Store numeric short textual information
  - Typically for managing enterprises
- Text and multimedia databases
  - Store documents, digital images, audio, and video streams
- Geographic information systems (GIS)
  - Store maps, weather data, and satellite images
  - For route-finding, agriculture, and natural resource management
- Data warehouses and online analytical processing (OLAP) systems
  - Store historical business information
  - For business analytics and decision support
- Real-time and active database technology
  - Store process models, constraints, and key performance indicators
  - Control industrial and manufacturing processes

Course Content

- Introduction to database systems
- Relational data model
- Entity-Relationship (ER) modelSQL
- Extended ER model
- Mapping ER models to relational
- Relational algebra
- Views and view management
- DBMS architecture
- Query processing and optimization
- Transaction management
- Data warehouses

Goal

- Implement a database in DBMS software, define schemas, develop queries and perform manipulations in SQL
- Transform a business description into Entity-Relationship diagrams, which were then used to specify a SQL Schema
- Apply knowledge of data redundancy, security, backups and recovery, data dependencies and other constraints in database design

Subject Assessment Plan

- Theory and Practicals
  - Laboratory assignments for different topics
  - A project
    - The project is an end-to-end design and implementation project to model part of an enterprises database (e.g. product and sales data for a retailer), create the database using SQL (on Oracle or PostgreSQL), populate it with data, and use JDBC to implement several simple user interfaces to the database (As done software tools – I, web based interface for the project is needed).
- Tutorial/Quizzes

Books

1). **A Silberschatz, H. F. Korth, and S Sudarshan, "Database System Concepts", 6/E,TMH, 2006**

2). McFadden, F. Hoffer, Prescott : M. B “Modern Database Management”, 8/E, Benjamin/Cummings Inc,2006

3). C. J Date, “An Introduction to Database Systems”, Publisher: Addison, Wesley, 8/E, 2003

4). Raghu Ramakrishnan and Gehrke: “Database Management System”, 3/E, WCB/McGraw-Hill,2003

5). Margaret H Dunham , “Data Mining. Introductory and advanced topics”, Pub: Pearson Education, 2003

6). Ivan Bayross, SQl, PL/SQL The Programming Language of Oracle, BPB Publications

End of Presentation

Terminology

- **Database**
  - Collection of related data (logically coherent)
  - Known facts that can be recorded and that have implicit meaning
  - Represents some aspect(s) of the real world
  - Built for a specific purpose
- Examples of large databases
  - Amazon, Organization

Terminology (cont'd.)

- **Database Management System (DBMS)**
  - Collection of programs
  - Enables users to create and maintain a database
  - Allows multiple users and programs to access and manipulate the database concurrently
  - Provides protection against unauthorized access and manipulation
  - Provides means to evolve database and program behaviour as requirements change over time
- Examples of database management systems
  - Microsoft's Access
  - IBM's DB2
  - Microsoft's SQL Server
  - Oracle
  - MySQL

Utilization of DBMS

- **Define** a database
  - Specify the data types, structures, and constraints of the data to be stored
  - Uses a *Data Definition Language*
- **Meta-data**
  - Database definition or descriptive information
  - Stored by the DBMS in the form of a *database catalog* or *data dictionary*
- Phases for designing a database:
  - **Requirements specification and analysis**
  - **Conceptual design**
    - e.g., using the *Entity-Relationship model*
  - **Logical design**
    - e.g., using the *relational model*
  - **Physical design**

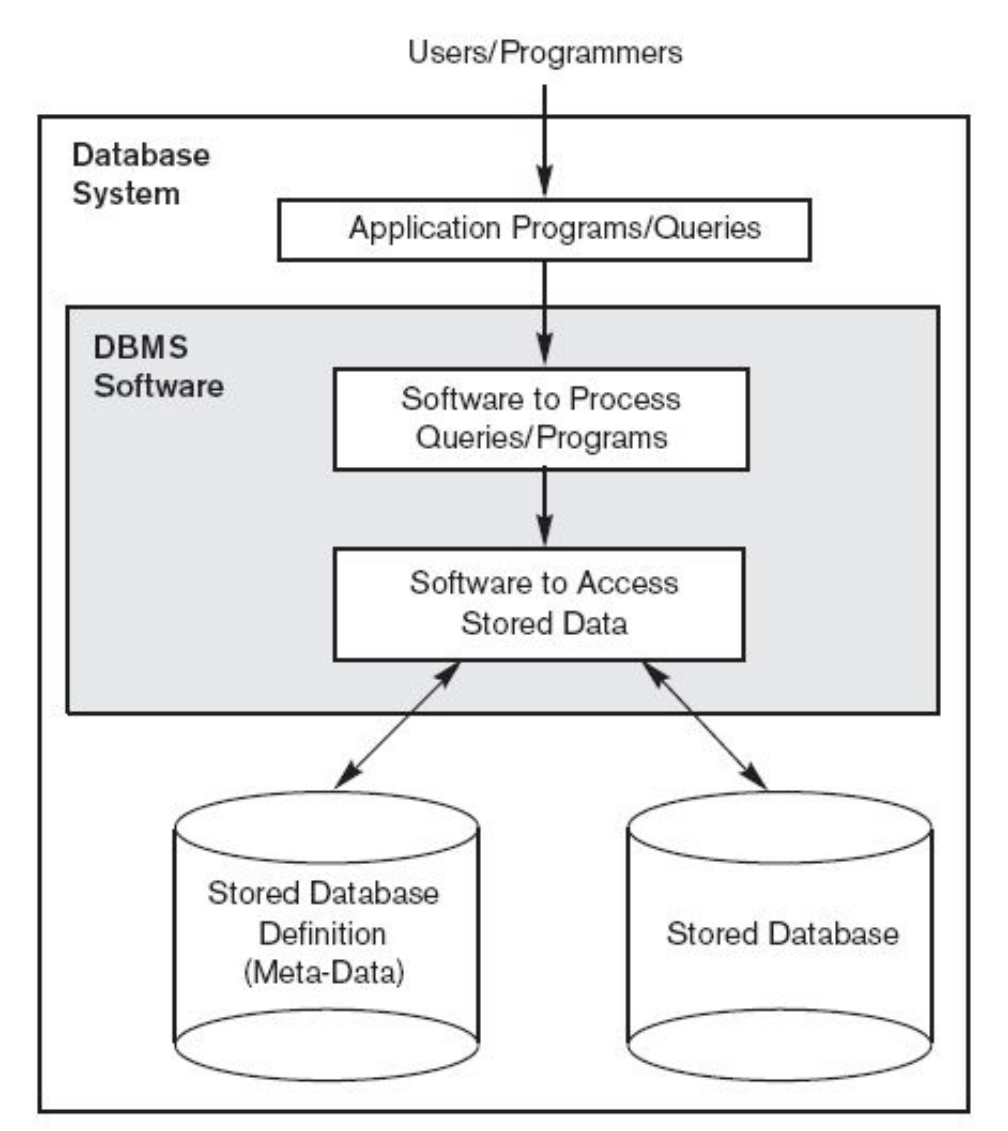
Utilization of DBMS (cont'd.)

- **Populate** a database
  - Insert data
- **Query**
  - Interaction causing some data to be retrieved
  - Uses a *Query Language*
- **Manipulate** a database
  - Query and update the database to understand
  - Generate reports
  - Uses a *Data Manipulation Language*
- **Application program**
  - Accesses database by sending queries and updates to DBMS
- **Transaction**
  - An atomic unit of queries and updates that must be executed as a whole
    - e.g., buying a product, transferring funds

Utilization of DBMS (cont'd.)

- **Reorganize** a database
  - Changes the metadata rather than the data
  - More drastic than data updates
    - May require massive changes to the data
    - May require changes to some application programs
  - Uses the *Data Definition Language again*

Database System Overview



An Example

- Movie database
  - Information concerning movies, actors, awards
- **Data records**
  - Film
  - Person
  - Role
  - Honors
- Define structure of each type of record by specifying **data elements** to include and **data type** for each element
  - String (sequence of alphabetic characters)
  - Numeric (integer or real)
  - Date (year or year-month-day)
  - Monetary amount
  - etc.

An Example (cont'd.)

- Populate MOVIES database
  - Store data to represent each film, actor, director, award, role

Film

title	genre	year	director	runtime	budget	gross
The Company Men	drama	2010	John Wells	104	15,000,000	4,439,063
Lincoln	biography	2012	Steven Spielberg	150	65,000,000	181,408,467
War Horse	drama	2011	Steven Spielberg	146	66,000,000	79,883,359
Argo	drama	2012	Ben Affleck	120	44,500,000	135,178,251
Fire Sale	comedy	1977	Alan Arkin	88	1,500,000	0

Person

name	birth	city
Ben Affleck	1972	Berkeley
Alan Arkin	1934	New York
Tommy Lee Jones	1946	San Saba
John Wells	1957	Alexandria
Steven Spielberg	1946	Cincinnati
Daniel Day-Lewis	1957	Greenwich

Honours

movie	award	category	winner
Lincoln	Critic's Choice	actor	Daniel Day-Lewis
Argo	Critic's Choice	director	Ben Affleck
Lincoln	Screen Actors Guild	supporting actor	Tommy Lee Jones
Lincoln	Screen Actors Guild	actor	Daniel Day-Lewis
Lincoln	Critic's Choice	screenplay	Tony Kushner
Argo	Screen Actors Guild	cast	Argo
War Horse	BMI Flim	music	John Williams

Role

actor	movie	persona
Ben Affleck	Argo	Tony Mendez
Alan Arkin	Argo	Lester Siegel
Ben Affleck	The Company Men	Bobby Walker
Tommy Lee Jones	The Company Men	Gene McClary
Tommy Lee Jones	Lincoln	Thaddeus Stevens
Alan Arkin	Fire Sale	Ezra Fikus

An Example (cont'd.)

- Manipulation involves querying and updating
- Examples of queries:
  - List the cast of characters for *Lincoln*.
  - Who directed a *drama* in *2012*?
  - Who directed a film in which he or she also played a role?
  - What awards were won by *War Horse*?
- Examples of updates:
  - Record that *Argo* won a Golden Globe award for best picture.
  - Add another $395,533 to the gross earnings for *Lincoln*.
  - Change the birthplace for *Daniel Day-Lewis* to *London*.
  - Delete *Fire Sale* from the database.

Terminology (cont'd.)

- **Reorganizing** a database
  - Changes the metadata rather than the data
  - More drastic than data updates
    - May require massive changes to the data
    - May require changes to some application programs
  - Uses the *Data Definition Language again*
- Examples:
  - Rename *gross* to be *domestic earnings* and add a new column for *foreign earnings*.
  - Move *director* from FILM to a separate relation DIRECTOR with columns for *person* and *movie*
  - Change *birth* from *yyyy* to *yyyy/mm/dd*
  - Split name in PERSON to separate *surname* from *given names*.
  - Include column *movieID* in FILM (to accommodate remakes and other duplications of film title) and update other relations accordingly.

Database Catalog

RELATIONS

Relation_name	No_of_columns
STUDENT	4
COURSE	4
SECTION	5
GRADE_REPORT	3
PREREQUISITE	2

COLUMNS

Column_name	Data_type	Belongs_to_relation
Name	Character (30)	STUDENT
Student_number	Character (4)	STUDENT
Class	Integer (1)	STUDENT
Major	Major_type	STUDENT
Course_name	Character (10)	COURSE
Course_number	XXXXNNNN	COURSE
….	….	…..
….	….	…..
….	….	…..
Prerequisite_number	XXXXNNNN	PREREQUISITE

Note: Major_type is defined as an enumerated type with all known majors.
XXXXNNNN is used to define a type with four alpha characters followed by four digits.

When Not to Use a DBMS

- More desirable to use regular files for:
  - Simple, well-defined applications with no expected changes at all
  - Small variety of data and/or small amount of data
  - Stringent, real-time requirements that cannot afford DBMS overhead
  - Only single (personal) access to data

- Unlikely that any of these apply to corporate data management
  - In fact, corporations often maintain many databases across many database systems

References

- https://www.postgresql.org/docs/8.0/connect-estab.html

File System vs. Database Management System

File System	Database Management System (DBMS)
1. It is a software system that manages and controls the data files in a computer system.	1. It is a software system used for creating and managing the databases. DBMS provides a systematic way to access, update, and delete data.
2. File system does not support multi-user access.	2. Database Management System supports multi-user access.
3. Data consistency is less in the file system.	3. Data consistency is more due to the use of normalization.
4. File system is not secured.	4. Database Management System is highly secured.
5. File system is used for storing the unstructured data.	5. Database management system is used for storing the structured data.
6. In the file system, data redundancy is high.	6. In DBMS, Data redundancy is low.
7. No data backup and recovery process is present in a file system.	7. There is a backup recovery for data in DBMS.

File System	Database Management System (DBMS)
8. Handling of a file system is easy.	8. Handling a DBMS is complex.
9. Cost of a file system is less than the DBMS.	9. Cost of database management system is more than the file system.
10. If one application fails, it does not affect other application in a system.	10. If the database fails, it affects all application which depends on it.
11. In the file system, data cannot be shared because it is distributed in different files.	11. In DBMS, data can be shared as it is stored at one place in a database.
12. These system does not provide concurrency facility.	12. This system provides concurrency facility.
13. **Example:** NTFS (New technology file system), EXT (Extended file system), etc.	13. **Example:** Oracle, MySQL, MS SQL Server, DB2, Microsoft Access, etc.