RAMAIAH
Institute of Technology

M.S. Ramaiah Institute of Technology
(Autonomous Institute, Affiliated to VTU)
Department of Computer Science and Engineering

Course Name: Database Systems
Course Code: CS52
Credits: 3:1:0
UNIT 1

Term: October 2021– February 2022

Reference:
Elmasri, R., Shamkant B. Navathe, R.
Fundamentals of Database Systems

RAMAIAH
Institute of Technology

Unit I

Introduction

Characteristics of Database approach

Actors on the Scene

Workers behind the scene

Advantages of using DBMS approach

Data models

schemas and instances

Three-schema architecture and data independence

Database languages and interfaces

The database system environment

Centralized and client-server architectures

Classification of Database Management systems,

Entity-Relationship Model:

Using high level Conceptual data models for Database Design

A Sample Database Application

Entity types, Entity sets Attributes and Keys

Relationship types, Relationship Sets,

Roles and Structural Constraints

Weak Entity Types.

RAMAIAH Institute of Technology

INTRODUCTION (1)

- Database - collection of related data (known facts that can be recorded and that have implicit meaning)
- A database has the following implicit properties:
  - represents some aspect of the real world, called the miniworld or the universe of discourse (DoD). Changes to the miniworld are reflected in the database.
  - logically coherent collection of data with some inherent meaning.
  - designed, built, and populated with data for a specific purpose. It has an intended group of users and applications
- A database can be of any size and of varying complexity – address book, library, tax payers information

RAMAIAH Institute of Technology

INTRODUCTION (2)

- Traditional databases applications – information that is stored and accessed is either textual or numeric ex:- Bank, hotel or airline reservation, library, book, toy, online shopping, supermarket.
- Multimedia databases - store pictures, video clips, and sound messages.
- Geographic information systems (CIS) can store and analyze maps, weather data, and satellite images.
- Data warehouses and online analytical processing (OLAP) systems - decision making.
- Real-time and active database technology - controlling industrial and manufacturing processes.
- World Wide Web - search for information.
- Used in business, electronic commerce, engineering, medicine, law, education, and library science.

RAMAIAH
Institute of Technology

INTRODUCTION (3)

- A database management system (DBMS) is a collection of programs that enables users to create and maintain a database.
- Facilitates the processes of defining, constructing, manipulating, and sharing databases among various users and applications.
- Defining - specifying the data types, structures, and constraints
- Constructing - storing the data itself on some storage medium
- Manipulating - querying, updating, and generating reports from the data.
- Sharing - multiple users and programs to access the database concurrently.
- Protection - hardware or software malfunction and security protection against unauthorized or malicious access.
- Maintain - evolve as requirements change over time.
- DBMS software - General-purpose or special-purpose.

RAMAIAH
Institute of Technology

INTRODUCTION (4)

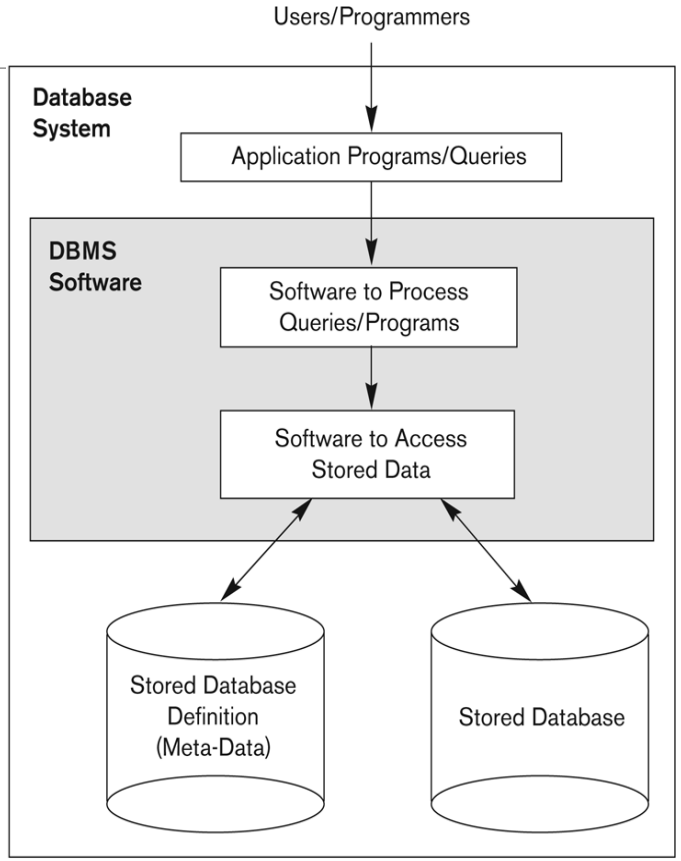

Figure 1.1
A simplified database system environment.

RAMAIAH Institute of Technology

INTRODUCTION (5)

UNIVERSITY database for maintaining information concerning students, courses, and grades in a university environment.

STUDENT - stores data on each student

COURSE - stores data on each course

SECTION file - stores data on each section of a course,

GRADE_REPORT - stores the grades

PREREQUISITE - stores the prerequisites of each course.

RAMAIAH
Institute of Technology

INTRODUCTION (6)

STUDENT

Name	Student_number	Class	Major
Smith	17	1	CS
Brown	8	2	CS

COURSE

Course_name	Course_number	Credit_hours	Department
Intro to Computer Science	CS1310	4	CS
Data Structures	CS3320	4	CS
Discrete Mathematics	MATH2410	3	MATH
Database	CS3380	3	CS

SECTION

Section_identifier	Course_number	Semester	Year	Instructor
85	MATH2410	Fall	07	King
92	CS1310	Fall	07	Anderson
102	CS3320	Spring	08	Knuth
112	MATH2410	Fall	08	Chang
119	CS1310	Fall	08	Anderson
135	CS3380	Fall	08	Stone

Figure 1.2
A database that stores student and course information.

GRADE_REPORT

Student_number	Section_identifier	Grade
17	112	B
17	119	C
8	85	A
8	92	A
8	102	B
8	135	A

PREREQUISITE

Course_number	Prerequisite_number
CS3380	CS3320
CS3380	MATH2410
CS3320	CS1310

RAMAIAH
Institute of Technology

CHARACTERISTICS OF DATABASE APPROACH (1)

Student details maintained by faculty, account section and dept. office may have different attributes.

Having multiple copies using traditional approach will lead to redundancy and inconsistency.

Characteristics of Database approach are:

- Self-describing nature of a database system
- Insulation between programs and data, and data abstraction
- Support of multiple views of the data
- Sharing of data and multiuser transaction procesing

RAMAIAH Institute of Technology

CHARACTERISTICS OF DATABASE APPROACH (2)
Self Describing Nature of a Database System

- Complete definition or description of the database structure and constraints – **metadata**
- **DBMS catalog** stores metadata
- Catalog is used by the DBMS software and database users
- In traditional file processing, structure is declared in the application programs.

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 enumerared type with all known majors. XXXXNNNN is used to define a type with four alpha characters followed by four digits

Figure 1.3
An example of a database catalog for the database in Figure 1.2.

RAMAIAH Institute of Technology

CHARACTERISTICS OF DATABASE APPROACH (3)
Insulation between Programs and Data, and Data Abstraction

- **program-data independence** – no need to change program if the structure of the file changes. For example – adding DoB to student data.
- An **operation** (also called a function or method) is specified in two parts.
- The **interface** of an operation includes the operation name and the data types of its arguments.
- The **implementation** of the operation is specified separately.
- **Program-operation independence** – changing the implementation of the operation without affecting its interface. For ex – calculating grade of student.
- **Data abstraction** - The characteristic that allows program-data independence and program-operation independence.
- **Data model** – is used to provide data abstraction to database users, it provides conceptual view without disclosing storage details.

RAMAIAH Institute of Technology

CHARACTERISTICS OF DATABASE APPROACH (4)
Support of Multiple Views of the Data

- Many users - require a different perspective or view of the database.
- A view may be a subset of the database or it may contain virtual data that is derived from the database files but is not explicitly stored.
- Some users may not need to be aware of whether the data they refer to is stored or derived.

RAMAIAH
Institute of Technology

CHARACTERISTICS OF DATABASE APPROACH (5)
Sharing of Data and Multiuser Transaction Processing

- A multiuser DBMS -allows multiple users to access the database at the same time.
- Concurrency control software – Controlled manner updation of database
- Example – Bus ticket reservation
- OLTP – Online Transaction Processing
- Transaction - executing program or process that includes one or more database accesses, such as reading or updating of database records.
- Transaction properties –Atomicity, Consistency, Isolation and Durability

RAMAIAH Institute of Technology

ACTORS ON THE SCENE (1)
Database Administrators

- Organisation's primary resource – database, secondary resource - DBMS and related software.
- Administering these resources is the responsibility of the database administrator (DBA).
- The DBA -
  - authorizing access to the database
  - coordinating and monitoring its use
  - acquiring software and hardware resources as needed.
  - breach of security
  - poor system response time.
- In large organizations, the DBA is assisted by a staff that helps carry out these functions

RAMAIAH
Institute of Technology

ACTORS ON THE SCENE (2)
Database Designers

- Identify the data to be stored in the database
- Choose appropriate structures to represent and store the data.
- Interact with each potential group of users
- Develop views of the database that meet the data and processing requirements of these groups.
- Each view is then analyzed and integrated with the views of other user groups.
- The final database design must be capable of supporting the requirements of all user groups
- Designing is mostly done before the database is implemented and populated with data.
- In many cases, they are on the staff of the DBA and may be assigned other staff responsibilities later.

RAMAIAH Institute of Technology

ACTORS ON THE SCENE (3)
END USERS (1)

End users are the people whose jobs require access to the database for querying, updating, and generating reports. Categories of end users:

Casual end users

- occasionally access the database
- may need different information each time.
- Use a sophisticated database query language to specify their requests
- middle- or high-level managers or other occasional browsers.
- learn only a few facilities that they may use repeatedly

RAMAIAH Institute of Technology

ACTORS ON THE SCENE (4)
END USERS (2)

Naive or parametric end users -

- Constant querying and updating the database,
- Using standard types of queries and updates-called canned transactions-that have been carefully programmed and tested.
- Bank tellers - check account balances and post withdrawals and deposits.
- Reservation clerks for airlines, hotels, and car rental companies - check availability and make reservations.
- Learn very little about the facilities provided by the DBMS
- Understand only the user interfaces of the standard transactions designed and implemented for their use.

RAMAIAH Institute of Technology

ACTORS ON THE SCENE (5)
END USERS (3)

Sophisticated end users

- Include engineers, scientists, business analysts, and others
- Thoroughly familiarize themselves with the facilities of the DBMS to implement their applications
- Learn most of the DBMS facilities in order to achieve their complex requirements

RAMAIAH Institute of Technology

ACTORS ON THE SCENE (6)
END USERS (4)

Stand-alone users

- Maintain personal databases by using ready-made program packages that provide easy-to-use menu-based or graphics-based interfaces.
- Example - user of a tax package - stores a variety of personal financial data for tax purposes.
- Usually become very proficient in using a specific software package.

RAMAIAH Institute of Technology

ACTORS ON THE SCENE (7)
System Analysts and Application Programmers (Software Engineers)

- Determine the **requirements of end users**, especially naive and parametric end users
- Develop specifications for **canned transactions** that meet these requirements.
- Application programmers implement these specifications as programs; then they **test, debug, document, and maintain** these canned transactions.
- Such analysts and programmers-commonly referred to as software engineers-should be familiar with the **full range of capabilities** provided by the DBMS to accomplish their tasks

RAMAIAH Institute of Technology

WORKERS BEHIND THE SCENE (1)
DBMS system designers and implementers

Design and implement the DBMS modules and interfaces as a software package. Responsible for:

- implementing the catalog
- processing query language, processing the interface
- accessing and buffering data
- controlling concurrency
- handling data recovery and security

The DBMS must interface with other system software, such as the operating system and compilers for various programming languages.

RAMAIAH Institute of Technology

WORKERS BEHIND THE SCENE (2)
Tool Developers

- Design and implement tools-the software packages
- Tools are optional packages that are often purchased separately.
- They include packages for database design, performance monitoring, natural language or graphical interfaces, prototyping, simulation, and test data generation.
- In many cases, independent software vendors develop and market these tools.

RAMAIAH Institute of Technology

WORKERS BEHIND THE SCENE (3)
Operators and maintenance personnel

- System administration personnel who are responsible for the actual running and maintenance of the hardware and software environment for the database system.

Workers behind the scene are instrumental in making the database system available to end users, they typically do not use the database for their own purposes.

RAMAIAH Institute of Technology

ADVANTAGES OF USING THE DBMS APPROACH (1)
Controlling Redundancy (1)

In traditional software development utilizing file processing, every user group maintains its own files for handling its data-processing applications.

This **redundancy** in storing the same data multiple times leads to several problems.

First, entering data on a new student-multiple times: once for each file where student data is recorded. This leads to *duplication of effort.*

Second, *storage space* is *wasted* when the same data is stored repeatedly, and this problem may be serious for large databases.

Third, files that represent the same data may become *inconsistent.*

It is sometimes necessary to use **controlled redundancy** for improving the performance of queries

RAMAIAH
Institute of Technology

ADVANTAGES OF USING THE DBMS APPROACH (2)
Controlling Redundancy (2)

Figure 1.6
Redundant storage of Student_name and Course_name in GRADE_REPORT. (a) Consistent data. (b) Inconsistent record.

(a)

GRADE_REPORT

Student_number	Student_name	Section_identifier	Course_number	Grade
17	Smith	112	MATH2410	B
17	Smith	119	CS1310	C
8	Brown	85	MATH2410	A
8	Brown	92	CS1310	A
8	Brown	102	CS3320	B
8	Brown	135	CS3380	A

(b)

GRADE_REPORT

Student_number	Student_name	Section_identifier	Course_number	Grade
17	Brown	112	MATH2410	B

RAMAIAH Institute of Technology

ADVANTAGES OF USING THE DBMS APPROACH (3)
Restricting Unauthorized Access

- When multiple users share a large database, it is likely that most users will not be authorized to access all information in the database. Example –Financial data.
- some users may be permitted only to retrieve data, whereas others are allowed both to retrieve and to update
- Typically, users or user groups are given account numbers protected by passwords, which they can use to gain access to the database.
- A DBMS should provide a security and authorization subsystem, which the DBA uses to create accounts and to specify account restrictions.
- The DBMS should then enforce these restrictions automatically.
- only the DBA's staff may be allowed to use certain privileged software, such as the software for creating new accounts.
- Similarly, parametric users may be allowed to access the database only through the canned transactions developed for their use.

RAMAIAH
Institute of Technology

ADVANTAGES OF USING THE DBMS APPROACH (4)
Providing Persistent Storage for Program Objects

- Programming languages typically have complex data structures, such as class definitions in c++ or Java.
- The values of program variables are discarded or explicitly stored into a format suitable for file storage.
- To read this data once more, the programmer must convert from the file format to the program variable structure.
- Object-oriented database systems are compatible with programming languages such as c++ and Java, and the DBMS software automatically performs any necessary conversions.
- A complex object in c++ can be stored permanently in an OODB. Such an object is said to be persistent, since it survives the termination of program execution and can later be directly retrieved by another c++ program.
- The persistent storage of program objects and data structures is an important function of database systems.
- Traditional database systems suffered from impedance mismatch problem, OODB offers data structure compatibility with one or more object oriented programming languages.

RAMAIAH Institute of Technology

ADVANTAGES OF USING THE DBMS APPROACH (5)
Providing Storage Structures for Efficient Query Processing

- DBMS must provide specialized data structures to speed up disk search for the desired records.
- Auxiliary files called indexes are used for this purpose.
- Indexes are typically based on tree data structures or hash data structures, suitably modified for disk search.
- In order to process the database records needed by a particular query, those records must be copied from disk to memory.
- DBMS has a buffering module that maintains parts of the database in main memory buffers or use operating system for the same.
- The query processing and optimization module of the DBMS is responsible for choosing an efficient query execution plan for each query based on the existing storage structures.
- Choice of index creation is part of physical database design and tuning, done by DBA staff.

RAMAIAH
Institute of Technology

ADVANTAGES OF USING THE DBMS APPROACH (6)
Providing Backup and Recovery

- A DBMS must provide facilities for recovering from hardware or software failures.
- The **backup and recovery subsystem** of the DBMS is responsible for recovery.
- For example, if the computer system fails in the middle of a complex update transaction,
- it should be recovered to initial state or resume from the point it was interrupted.

RAMAIAH
Institute of Technology

ADVANTAGES OF USING THE DBMS APPROACH (7)
Providing Multiple User Interfaces

- Because many types of users with varying levels of technical knowledge use a database,
- A DBMS should provide a variety of user interfaces.
- These include query languages for casual users
- programming language interfaces for application programmers
- Graphical user interfaces (GUIs) -forms and command codes for parametric users, and menu-driven interfaces and natural language interfaces for stand-alone users.
- Web GUI interfaces to a database

RAMAIAH Institute of Technology

ADVANTAGES OF USING THE DBMS APPROACH (8)
Representing Complex Relationships among Data

- A database may include numerous varieties of data that are interrelated in many ways.
- A DBMS must have the capability to represent a variety of complex relationships among the data as well as to retrieve and update related data easily and efficiently.

RAMAIAH Institute of Technology

ADVANTAGES OF USING THE DBMS APPROACH (9)
Enforcing Integrity Constraints

- Most database applications have certain integrity constraints that must hold for the data.
- A DBMS should provide capabilities for defining and enforcing these constraints.
- The simplest type of integrity constraint involves specifying a data type for each data item.
- A record in one file must be related to records in other files.
- Another type of constraint specifies uniqueness on data item values.
- Database designer should identify integrity constraints during database design.
- Some constraints can be specified to the DBMS and automatically enforced.
- Other constraints may have to be checked by update programs or at the time of data entry.
- A data item may be entered erroneously and still satisfy the specified integrity constraints. For example grade of a student.
- Such data entry errors can only be discovered manually and corrected later by updating the database.

RAMAIAH Institute of Technology

ADVANTAGES OF USING THE DBMS APPROACH (10)
Permitting Inferencing and Actions Using Rules

- Some database systems provide capabilities for defining *deduction rules* for inferencing new information from the stored database facts.
- Such systems are called deductive database systems.
- For example, there may be complex rules in the miniworld application for determining when a student is on probation.
- These can be specified *declaratively* as rules, which when compiled and maintained by the DBMS can determine all students on probation.
- In a traditional DBMS, an explicit *procedural program code* would have to be written to support such applications. But if the miniworld rules change, it is has to be re-coded.

RAMAIAH
Institute of Technology

ADVANTAGES OF USING THE DBMS APPROACH (11)
Additional Implications of Using the Database Approach (1)

Potential for Enforcing Standards.

- The database approach permits the DBA to define and enforce standards among database users in a large organization.
- This facilitates communication and cooperation among various departments, projects, and users within the organization.
- Standards can be defined for names and formats of data elements, display formats, report structures, terminology, and so on.
- The DBA can enforce standards in a centralized database environment more easily than in an environment where each

RAMAIAH Institute of Technology

ADVANTAGES OF USING THE DBMS APPROACH (12)
Additional Implications of Using the Database Approach (2)

Reduced Application Development Time.

- Developing a new application-such as the retrieval of certain data from the database for printing a new report-takes very little time.
- Designing and implementing a new database from scratch may take more time than writing a single specialized file application.
- However, once a database is up and running, substantially less time is generally required to create new applications using DBMS facilities.
- Development time using a DBMS is estimated to be one-sixth to one-fourth of that for a traditional file system.

RAMAIAH
Institute of Technology

ADVANTAGES OF USING THE DBMS APPROACH (13)
Additional Implications of Using the Database Approach (3)

Flexibility

- It may be necessary to change the structure of a database as requirements change.
- In response, it may be necessary to add a file to the database or to extend the data elements in an existing file.
- Modern DBMSs allow certain types of evolutionary changes to the structure of the database without affecting the stored data and the existing application programs.

RAMAIAH Institute of Technology

ADVANTAGES OF USING THE DBMS APPROACH (14)
Additional Implications of Using the Database Approach (4)

Availability of Up-to-Date Information.

- A DBMS makes the database available to all users.
- As soon as one user's update is applied to the database, all other users can immediately see this update.
- This availability of up-to-date information is essential for many transaction-processing applications, such as reservation systems or banking databases.

RAMAIAH
Institute of Technology

ADVANTAGES OF USING THE DBMS APPROACH (15)
Additional Implications of Using the Database Approach (5)

Economies of Scale.

- The DBMS approach permits consolidation of data and applications, thus reducing the amount of wasteful overlap between activities of data processing personnel in different projects or departments.
- This enables the whole organization to invest in more powerful processors, storage devices, or communication gear, rather than having each department purchase its own (weaker) equipment.
- This reduces overall costs of operation and management.

RAMAIAH
Institute of Technology

Review Questions

1.1. Define the following terms: *data, database, DBMS, database system, database catalog, program-data independence, user view, DBA, end user, canned transaction, deductive database system, persistent object, meta-data, transaction-processing application.*
1.2. What three main types of actions involve databases? Briefly discuss each.
1.3. Discuss the main characteristics of the database approach and how it differs from traditional file systems.
1.4. What are the responsibilities of the DBA and the database designers?
1.5. What are the different types of database end users? Discuss the main activities of each.
1.6. Discuss the capabilities that should be provided by a DBMS.

Exercises

1. Identify some informal queries and update operations that you would expect to apply to the database shown in Figure 1.2.
2. What is the difference between controlled and uncontrolled redundancy? Illustrate with examples.
3. Name all the relationships among the records of the database shown in Figure 1.2.
4. Give some additional views that may be needed by other user groups for the database shown in Figure 1.2.
5. Cite some examples of integrity constraints that you think should hold on the database shown in Figure 1.2.