Department of CSE

COURSE NAME: DBMS
COURSE CODE:23AD2102R

Topic: **INTRODUCTION TO DATABASES**

Session - 1

AIM OF THE SESSION

To familiarize students with the basic concept Database Management Systems

INSTRUCTIONAL OBJECTIVES

This Session is designed to: Characteristics of DBMS

LEARNING OUTCOMES

At the end of this session, you should be able to: basic knowledge of DBMS

KEY CONCEPTS

- Basic Terminology (Data, Information)
- History of Database systems
- Approaches to store data
- File based Approach
- Problems with traditional Approach
- Database Approach
- Operations on database
- Components of Database
- Database Management Systems (DBMS)
- Database Characteristics
- Differences between File and Database Approaches
- Advantages & Disadvantages of DBMS

BASIC TERMINOLOGY (DATA, INFORMATION)

The terms "data" and "information" are often used interchangeably, but they actually aren't the same.

WHAT IS DATA ?

- Data is defined as **a collection of individual facts or statistics**. (While "datum" is technically the singular form of "data," it's not commonly used in everyday language.)
- Data can come in the form of
  - **text,**
  - **observations,**
  - **figures,**
  - **images,**
  - **numbers,**
  - **graphs, or**
  - **symbols**.
- **For example**, data might include **individual prices, weights, addresses, ages, names, temperatures, dates, or distances, etc**.

WHAT IS DATA?

- Data is a **raw form of knowledge** and, on its own, doesn't carry any significance or purpose. In other words, you have to **interpret data for it to have meaning**. Data can be simple—and may even **seem useless until it is analyzed, organized, and interpreted**.
- There are two main types of data:
- **Quantitative data** is provided in numerical form, like the weight, volume, or cost of an item.
- **Qualitative data** is descriptive, but non-numerical, like the name, sex, or eye color of a person.
- **Data can be classified further into different types:**
- **Structured Data**: Data that fits into predefined, organized formats with a consistent structure, such as data stored in relational databases using tables, rows, and columns.
- **Unstructured Data**: Data that does not have a predefined structure or format, such as text documents, images, videos, social media posts, or sensor data.
- **Semi-Structured Data**: Data that has some structure but may not fit neatly into a rigid schema. It includes data formats like XML, JSON, or CSV files.
- **Metadata: Data** that provides information about other data. It describes the characteristics, properties, or attributes of the main data. For example, the metadata of a file can include its size, creation date, author, or file type.

WHAT IS INFORMATION?

- **Information** is defined as knowledge gained through study, communication, research, or instruction.
- Essentially, **information** is the result of analyzing and interpreting pieces of data. Whereas **data** is the individual figures, numbers, or graphs, information is the perception of those pieces of knowledge.
- For example, a set of data could include temperature readings in a location over several years. Without any additional context, those temperatures have no meaning. However, when you analyze and organize that information, you could determine seasonal temperature patterns or even broader climate trends. Only when the data is organized and compiled in a useful way can it provide information that is beneficial to others.

THE KEY DIFFERENCES BETWEEN DATA VS INFORMATION

- Data is a collection of facts, while information puts those facts into context.
- While data is raw and unorganized, information is organized.
- Data points are individual and sometimes unrelated. Information maps out that data to provide a big-picture view of how it all fits together.
- Data, on its own, is meaningless. When it's analyzed and interpreted, it becomes meaningful information.
- Data does not depend on information; however, information depends on data.
- Data typically comes in the form of graphs, numbers, figures, or statistics. Information is typically presented through words, language, thoughts, and ideas.
- Data isn't sufficient for decision-making, but you can make decisions based on information.

HISTORY OF DATABASE SYSTEMS

- **Databases** are a foundational element of the modern world. We interact with them even without knowing it — any time we buy something online, or log in to a service, or access our bank accounts, and so on.
- The concept of a database existed long before computers.
  - In these times, data was stored in journals, in libraries, and in hundreds of filing cabinets.
  - Everything was recorded via paper — and that meant **it took up space, was hard to find, and difficult to back up.**
- And then computers became available, and with them, the opportunity for better data management.

HISTORY OF DATABASE SYSTEMS

- **The 1960s – beginnings**
  - ✓ Charles Bachman designed the first computerised database in the early 1960s. This first database was known as the **Integrated Data Store, or IDS.**
  - ✓ This was shortly followed by the **Information Management System**, a database created by IBM.

Both databases were forerunners of the 'navigational database'.

Navigational databases required users to navigate through the entire database to find the information they wanted.

There are two main models of this:

- **hierarchical model (**data is organised like a family tree**)**
- **network model (**a record to have more than one parent and child record**)**

SESSION INTRODUCTION

- **The 1970s – relational databases**

A relational database is one that shows the relationship between different data records.

- **searchable**
- **more space-efficient (reduced data storage costs)**

- Ted Codd defined the relational data model
  - ✓ Won the ACM Turing Award for this work
- IBM then released their relational database, named as, System R
  - ✓ The first in the history of databases to use structured query language (SQL)

SESSION INTRODUCTION

- **The 1980s – growth and standardisation**
  - emergence of Object-oriented database management systems (OODBMS)
  - view data as 'objects'
  - work with programming languages that supported the 'object-oriented' approach
- **The 1990s – the internet**
  - object oriented database systems grow more popular in the 90s.
  - the creation of the World Wide Web
  - demand for client-server database systems
  - the creation of MySQL in 1995

- **The 2000s – NoSQL**
  - In 1998, the term NoSQL (not only structured query language) was coined.
  - It refers to databases that use query language other than SQL to store and retrieve data.
  - NoSQL databases are useful for **unstructured data**, and they saw a growth in the 2000s.
  - This is a notable development in the history of databases because NoSQL allowed for faster processing of larger, more varied datasets
- **The 2010s – distributed databases and cybersecurity**
  - The 2010s were a decade of increased data awareness, with the rise of big data and an increased emphasis on data protection.
  - Big data was a major buzzword of the 2010 — and big data meant big databases to house it.

APPROACHES TO STORE DATA

There are two approaches for storing data in computers:

- File-based approach.
- Database-approach.

FILE BASED APPROACH

- A **filesystem** is a method for storing and organizing computer files and the data they contain to make it easy to find and access them.

- File systems may use a **storage device** such as a hard disk or CD ROM and involve maintaining the physical location of the files.

- Programmers used **programming languages** such as COBOL, C++ to write applications that directly accessed files to perform data management services and provide information for users.

UNIVERSITY DATABASE IN FILE BASED SYSTEM

General Office	Library	Hostel	Account Office
Rollno	Rollno	Rollno	Rollno
Name	Name	Name	Name
Class	Class	Class	Class
Father_Name	Address	Father_Name	Address
Date_of_birth	Date_of_birth	Date_of_birth	Phone_No
Address	Phone_No	Address	Fee
Phone_No	No_of_books_issued	Phone_No	Installments
Previous_Record	Fine	Mess_Bill	Discount
Attendance	etc.	RoomNo	Balance
Marks		etc.	Total
Etc.			etc.

PROBLEMS IN TRADITIONAL APPROACH

- Data Security
- Data Redundancy
- Data Isolation
- Program / Data Dependence
- Lack of Flexibility
- Concurrent Access Anomalies

DATABASE APPROACH

Database: "**A collection of related data**"

"**A database system is referred as a collection of related data, which contains information about one particular organization**"

DATABASE DEFINITIONS

- The database is a single, large repository of data, which can be used simultaneously by many departments and users.

- The related information when placed is an organized form, then it is called a database.

- It can also defined as
  - Specifying data types, structures and constraints of the data to be stored.

DATABASE

- A database system is **a collection of related information stored**, so that it is available to many users for different purposes.
- A database system is essentially nothing more than **a computerized record-keeping system**. i.e. it is a computerized system whose overall purpose is to **maintain information and make that information available on demand**.
- A database system is **an electronic filing system**, as it is a repository for a collection of computerized data files.

UNIVERSITY DATABASE IN DATABASE BASED SYSTEM

General Office	Library	Hostel	Account Office
Rollno	Rollno	Rollno	Rollno
Name	No_of_books_issued	RoomNo	Fee
Class	Fine	Mess_Bill	Installments
Father_Name	etc.	etc.	Discount
Address			Balance
Phone_No			Total
Date_of_birth			etc.
Previous_Record			
Attendance			
Marks			
etc.			

OPERATIONS ON DATABASES

- To add new information
- To view or retrieve the stored information
- To modify or edit the existing
- To remove or delete the unwanted information
- Arranging the information in a desired order etc.

COMPONENTS OF DATABASE

Database system involves four major components:

1. **Data**
2. **Hardware**
   - ***Secondary storage*** (magnetic disks) together with ***the I/O devices, device controllers, I/O channels,*** to hold the stored data.
   - ***Processor(s)*** and associated ***Main memory,*** to support the execution of the database system software.
3. **Software**
   - The ***database manager (DB manager)*** or, more usually ***database management system (DBMS).***
4. **Users**
   - ***Casual end-users, Naïve users, DBA, System analyst, Sophisticated users, Database designers, Application Programmers.***

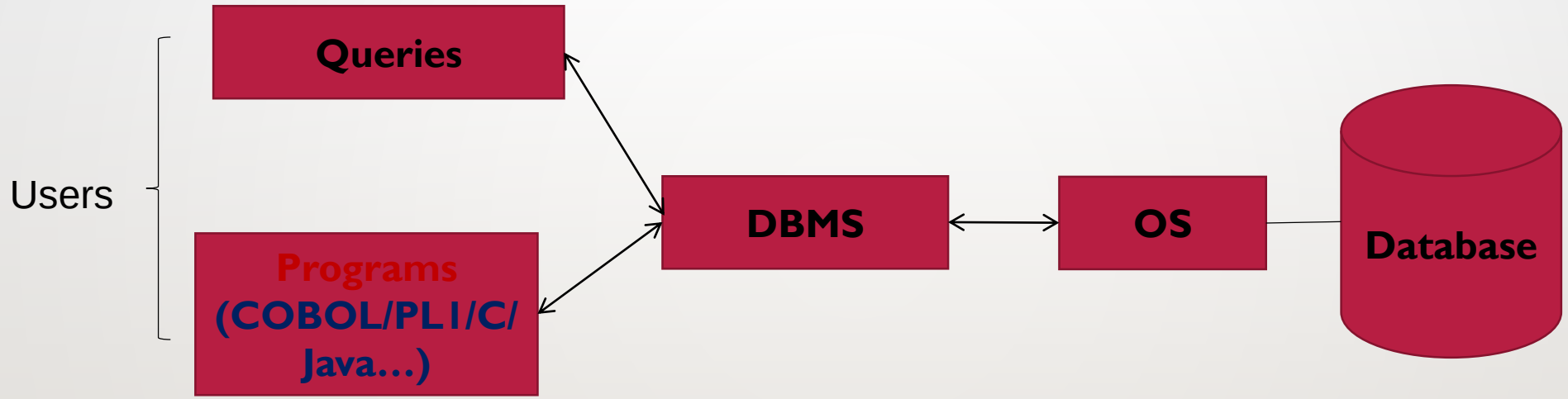
DATABASE MANAGEMENT SYSTEM (DBMS)

- DBMS is the major **software component** of database system to **manage database**, thus, it is called as **database manager** (DB manager).
- All **requests/queries** from users for accessing to the database are **handled by the DBMS**.
- DBMS **interprets and processes user's requests to retrieve information** from a Database.

- DBMS is a **collection of interrelated files** and a set of programs that **allows several users to access and modify these files.**
- DBMS is a **software package/ system** to facilitate the **creation and maintenance of a computerized database.**

DBMS

❖The DBMS acts as **an intermediary/interface** between the **Database and the Users**.



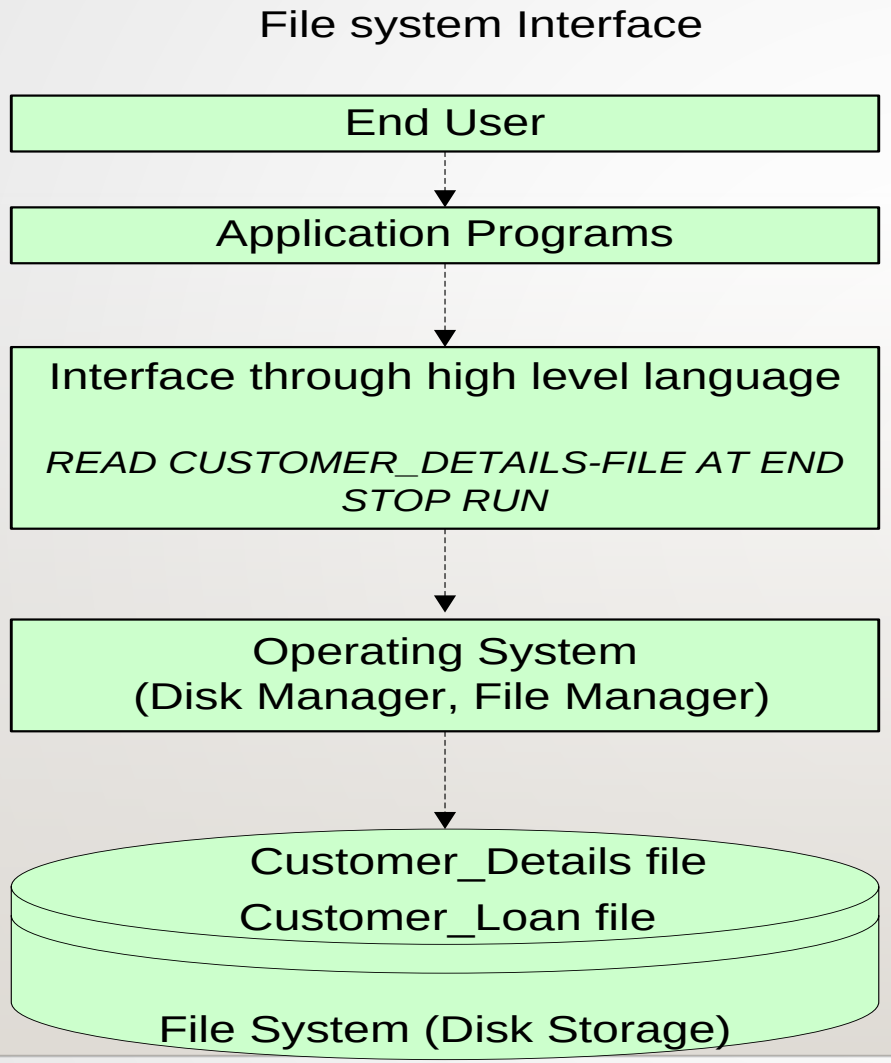
DBMS as an interface between physical Database and User's requests

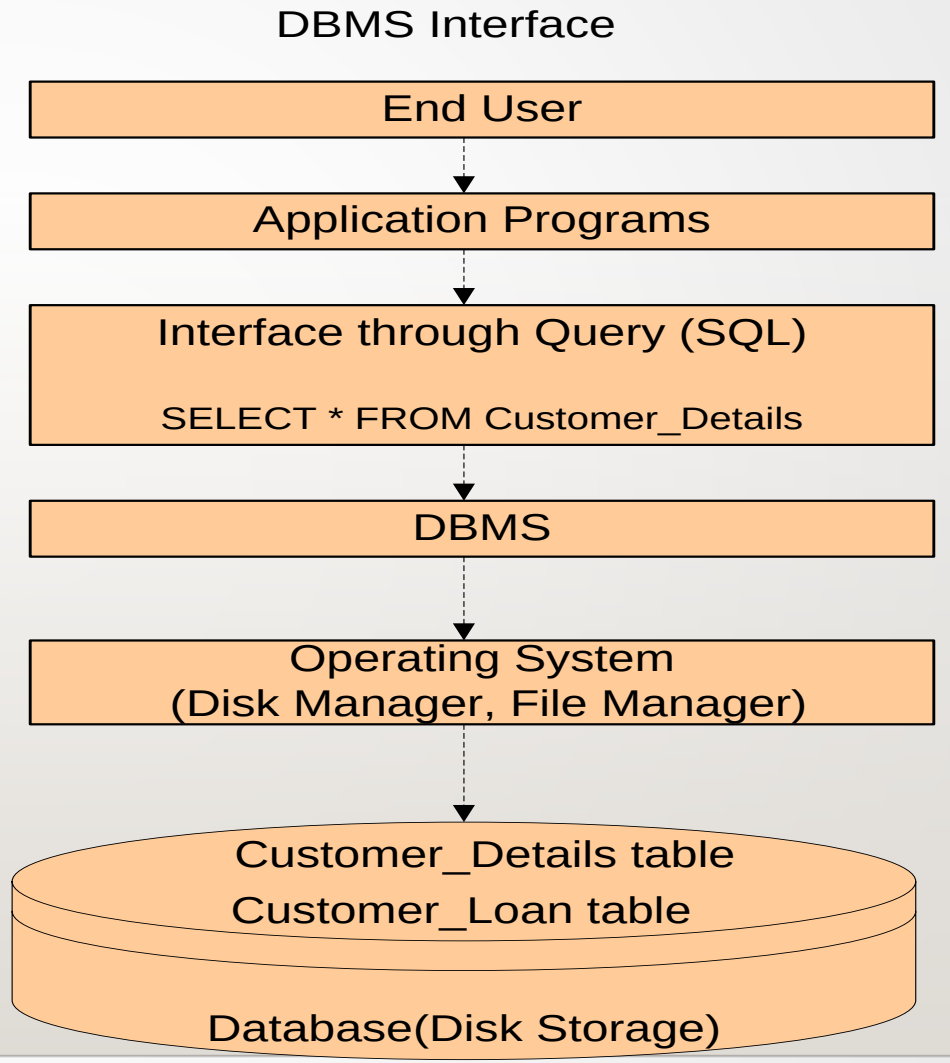
SESSION INTRODUCTION

- DBMS, A database management system is the software system that allows users to define, create and maintain a database and provides controlled access to the data.

- A database management system (DBMS) is basically a collection of programs that enables users to store, modify, and extract information from a database as per the requirements.

Difference Between File and DBMS Operations





APPLICATION AREAS OF DBMS

- Banking
- Airline
- Universities
- Credit card Transactions
- Telecommunications
- Finance
- Sales
- Manufacturing

CLASSIFICATION & APPLICATIONS

Classification of DBMS

1. Based on data models
   - Hierarchical DBMS
   - Network DBMS
   - Relational DBMS
2. Based on number of users
   - Single - user system
   - Multi - user system
3. Based on number of sites
   - Centralized database systems
   - Distributed database systems
4. Based on the purpose
   - General purpose
   - Special purpose

Database Applications:

- Banking: all transactions
- Airlines: reservations, schedules
- Universities: registration, grades
- Sales: customers, products, purchases
- Online retailers: order tracking, customized recommendations
- Manufacturing: production, inventory, orders, supply chain
- Human resources: employee records, salaries, tax deductions

EXAMPLES OF DBMS

- **The top commercial DBMS:**
  - Oracle
  - Microsoft SQL Server
  - IBM DB2
  - Microsoft Access, etc.
- **The top open source DBMS:**
  - MySQL
  - PostgreSQL
  - MongoDB
  - Cassandra
  - Redis
  - SQL Lite, etc.

WHICH DATABASE IS BETTER : COMMERCIAL OR OPEN SOURCE DATABASE ?

- Both Commercial and Open Source database have their own Advantages and Disadvantages.

- In most cases it makes sense to choose Open Source as compared to Commercial Database because :
  - Open Source database is Cost effective.
  - Better quality source code.
  - More secure.
  - More preferred.

CHARACTERISTICS OF THE DATABASE APPROACH

- **Self-describing nature of a database system:**
  - A DBMS **catalog** stores the description of a particular database (e.g. data structures, types, and constraints)
  - The description is called **meta-data.**
  - This allows the DBMS software to work with different database applications.
- **Insulation between programs and data:**
  - Called **program-data independence.**
  - Allows changing data structures and storage organization without having to change the DBMS access programs

CHARACTERISTICS OF THE DATABASE APPROACH

- **Data Abstraction:**
  - A data model is used to hide storage details and present the users with a conceptual view of the database
  - Programs refer to the data model constructs rather than data storage details
- **Support of multiple views of the data:**
  - Each user may see a different view of the database, which describes only the data of interest to that user

CHARACTERISTICS OF THE DATABASE APPROACH

- **Sharing of data and multi-user transaction processing:**
  - Allowing a set of **concurrent users** to retrieve from and to update the database
  - *Concurrency control within the DBMS guarantees that each* transaction is correctly executed or aborted
  - *Recovery subsystem ensures each completed transaction* has its effect permanently recorded in the database
  - OLTP (Online Transaction Processing) is a major part of database applications This allows hundreds of concurrent transactions to execute per second

Advantages of DBMS

- **Integrity can be enforced:** Integrity of data means that data in database is always accurate, such that incorrect information cannot be stored in database
- **Controlling Redundancy:** Duplication of data is almost avoided but not completely.
- **Inconsistency can be avoided:** When the same data is duplicated and changes are made at one site, which is not propagated to the other site, it gives rise to inconsistency and the two entries regarding the same data will not agree

Other Advantages

- Data can be shared.
- Data Independence.
- Concurrent access.
- Recovery from crashes.
- Providing Backup and Recovery.
- Restricting unauthorized access.
- Solving enterprise requirement than individual requirement

Disadvantages of DBMS

- Complexity
- Higher impact of a failure.
- Cost of DBMS.
- Additional Hardware costs.

SESSION INTRODUCTION

DBMS	File System
DBMS is a collection of data. In DBMS, the user is not required to write the procedures.	File system is a collection of data. In this system, the user has to write the procedures for managing the database.
DBMS gives an abstract view of data that hides the details.	File system provides the detail of the data representation and storage of data.
DBMS provides a crash recovery mechanism, i.e., DBMS protects the user from the system failure.	File system doesn't have a crash mechanism, i.e., if the system crashes while entering some data, then the content of the file will lost.
DBMS provides a good protection mechanism.	It is very difficult to protect a file under the file system.
DBMS contains a wide variety of sophisticated techniques to store and retrieve the data.	File system can't efficiently store and retrieve the data.
DBMS takes care of Concurrent access of data using some form of locking.	In the File system, concurrent access has many problems like redirecting the file while other deleting some information or updating some information.

Objectives of DBMS

- Mass Storage: DBMS can store a lot of data and can fetch data when ever it is needed.
- Removes Duplicity: Checks are provided while inserting new records.
- Multiple Users Access: Allows Concurrent Access also.
- Data Protection: No one can alter or modify without privilege of using the data.
- Data Backup and Recovery: Ability to back up and recover all the data in DB.
- Everyone can work on DBMS: No need to be a master. Non technical back ground person can work on it.
- Integrity: Multiple validity checks that makes data completely accurate and consistence.
- Platform Independent: Can run on any platform.

SELF-ASSESSMENT QUESTIONS

1. Data is

(a) Used in decision making
(b) Raw facts or events
(c) Transformed facts
(d) Information

2. What is a database?

(a) Organized collection of information that cannot be accessed, updated, and managed
b) Collection of data or information without organizing
c) Organized collection of data or information that can be accessed, updated, and managed
d) Organized collection of data that cannot be updated

SUMMARY

1. Understand the importance of DBMS
2. History of DBMS
3. Characteristics approach

TERMINAL QUESTIONS

1. Describe the history of DBMS.
2. List out the application areas of DBMS.
3. Analyze DBMS and file systems.
4. Summarize the characteristics of database approach.

REFERENCES FOR FURTHER LEARNING OF THE SESSION

Reference Books:

1. 1. Database System Concepts, Sixth Edition, Abraham Silberschatz, Yale University Henry, F. Korth Lehigh University, S. Sudarshan Indian Institute of Technology, Bombay.
2. Fundamentals of Database Systems, 7th Edition, RamezElmasri, University of Texas at Arlington, Shamkant B. Navathe, University of Texas at Arlington.

Web Link:

1. https://nptel.ac.in/courses/106105175

THANK YOU

(DEEMED TO BE UNIVERSITY)

Team – DBMS