

## Chapter 5

### Database Management System

#### *Introduction:*

A database management system (DBMS) is a software package designed to define, manipulate, retrieve and manage data in a database.

Some other DBMS examples include:

- MySQL
- SQL Server
- Oracle
- dBASE
- FoxPro

#### *Data Hierarchy:*

The components of the data hierarchy are listed below.

A data field holds a single factor attribute of an entity. Consider a date field, e.g. "19 September 2004". This can be treated as a single date field (e.g. birthdate), or three fields, namely, day of month, month and year.

A record is a collection of related fields. An Employee record may contain a name field(s), address fields, birthdate field and so on.

A file is a collection of related records. If there are 100 employees, then each employee would have a record (e.g. called Employee Personal Details record) and the collection of 100 such records would constitute a file (in this case, called Employee Personal Details file).

Files are integrated into a database. This is done using a Database Management System. If there are other facets of employee data that we wish to capture, then other files such as Employee Training History file and Employee Work History file could be created as well.

#### *Illustration of the data hierarchy*

An illustration of the above description is shown in this diagram below.

Hierarchy	Example
Database	<b>Employee Database</b> <div>Employee Details File      Training Records File</div> <div>Salary File</div>
File	<b>Employee Details File</b> <div> <div>EMP_NAME      JOB TITLE      DATE EMPLOYED</div> <div> <div>Alice Carter      Lecturer      31 Mar 2002</div> <div>Faridah bte Hassan      Sales Manager      9 Aug 2013</div> <div>Jeffrey Tan      Lecturer      19 Sep 2004</div> <div>Steve Willis      HR Manager      23 Dec 2005</div> </div> </div>
Record	<b>Employee Record</b> <div> <div>EMP_NAME      JOB TITLE      DATE EMPLOYED</div> <div>Jeffrey Tan      Lecturer      19 Sep 2004</div> </div>
Field	<b>Employee Name Field</b> <div>EMP_NAME</div> <div>Jeffrey Tan</div>
Byte	01001010 (Letter J in ASCII)
Bit	0

Note: EMP = employee

Source: Jeffrey T. Tan Wikipedia original contributor for Data Hierarchy. 9 Aug 2013  
Permission is given to freely use this diagram in its entirety & unedited.

#### Data Hierarchy Diagram – with Employee Database example

The following terms are for better clarity.

With reference to the example in the above diagram.

Data field label = Employee Name or EMP\_NAME

Data field value = Jeffrey Tan

The above description is a view of data as understood by a user e.g. a person working in Human Resource Department.

The above structure can be seen in the hierarchical model, which is one way to organize data in a database. In terms of data storage, data fields are made of bytes and these in turn are made up of bits.

#### Traditional File Approach/traditional file system:

File processing systems was an early attempt to computerize the manual filing system that we are all familiar with. A file system 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 as storage devices such as a hard disk or CD-ROM and involve maintaining the physical location of the files.

### *Limitations of the File Processing System | File-Based Approach*

*There are following problems associated with the File Based Approach:*

*1. Separated and Isolated Data: To make a decision, a user might need data from two separate files. First, the files were reevaluated by analysts and programmers to determine the specific data required from each file and the relationships between the data and then the applications could be written in a programming language to process and extract the needed data. Imagine the work involved if data from several files was needed.*

*2. Duplication of data: Often the same information is stored in more than one file. Uncontrolled duplication of data is not required for several reasons, such as:*

- Duplication is wasteful. It costs time and money to enter the data more than once*
- It takes up additional storage space, again with associated costs.*
- Duplication can lead to loss of data integrity; in other words the data is no longer consistent. For example, consider the duplication of data between the Payroll and Personnel departments. If a member of staff moves to a new house and the change of address is communicated only to Personnel and not to Payroll, the person's payslip will be sent to the wrong address. A more serious problem occurs if an employee is promoted with an associated increase in salary. Again, the change is notified to Personnel but the change does not filter through to Payroll. Now, the employee is receiving the wrong salary. When this error is detected, it will take time and effort to resolve. Both these examples, illustrate inconsistencies that may result from the duplication of data. As there is no automatic way for Personnel to update the data in the Payroll files, it is difficult to foresee such inconsistencies arising. Even if Payroll is notified of the changes, it is possible that the data will be entered incorrectly.*

*3. Data Dependence: In file processing systems, files and records were described by specific physical formats that were coded into the application program by programmers. If the format of a certain record was changed, the code in each file containing that format must be updated. Furthermore, instructions for data storage and access were written into the application's code. Therefore, changes in storage structure or access methods could greatly affect the processing or results of an application.*

*In other words, in file based approach application programs are data dependent. It means that, with the change in the physical representation (how the data is physically represented in disk) or access technique (how it is physically accessed) of data, application programs are also affected and need modification. In other words application programs are dependent on the how the data is physically stored and accessed.*

*If for example, if the physical format of the master/transaction file is changed, by making the modification in the delimiter of the field or record, it necessitates that the application programs which depend on it must be modified.*

*Let us consider a student file, where information of students is stored in text file and each field is separated by blank space as shown below:*

*IRahat35Thapar*

*Now, if the delimiter of the field changes from blank space to semicolon as shown below*

: 1;Rahat;35;Thapar

Then, the application programs using this file must be modified, because now it will token the file id on semicolon, but earlier it was blank space.

4. Difficulty in representing data from the user's view: To create useful applications for the user, often data from various files must be combined. In file processing it was difficult to determine relationships between isolated data in order to meet user requirements.

5. Data Inflexibility: Program-data interdependency and data isolation, limited the flexibility of file processing systems in providing users with ad-hoc information requests

6. Incompatible file formats: As the structure of files is embedded in the application programs, the structures are dependent on the application programming language. For example, the structure of a file generated by a COBOL program may be different from the structure of a file generated by a 'C' program. The direct incompatibility of such files makes them difficult to process jointly.

7. Data Security. The security of data is low in file-based systems because, the data is maintained in the flat file(s) is easily accessible. For Example: Consider the Banking System. The Customer Transaction file has details about the total available balance of all customers. A Customer wants information about his account balance. In a file system it is difficult to give the Customer access to only his data in the file. Thus enforcing security constraints for the entire file or for certain data items are difficult.

8. Transactional Problems. The File-based system approach does not satisfy transaction properties like Atomicity, Consistency, Isolation and Durability properties commonly known as ACID properties.

For example: Suppose, in a banking system, a transaction that transfers Rs. 1000 from account A to account B with initial values of A and B being Rs. 5000 and Rs. 10000 respectively. If a system crash occurred after the withdrawal of Rs. 1000 from account A, but before depositing of amount in account B, it will result in an inconsistent state of the system. It means that the transaction should not execute partially but wholly. This concept is known as Atomicity of a transaction (either 0% or 100% of transaction). It is difficult to achieve this property in a file-based system.

9. Concurrency problems. When multiple users access the same piece of data at the same interval of time then it is called as concurrency of the system. When two or more users read the data simultaneously there is no problem, but when they like to update a file simultaneously, it may result in a problem.

For example: =

Let us consider a scenario where in transaction T1 a user transfers an amount of 1000 from Account A to B (initial value of A is 5000 and B is 8000). In meanwhile, another transaction T2, tries to display the sum of account A and B is also executed. If both the transactions run in parallel it may result in inconsistency as shown below:

T1	T2	Status
withdraw 1000 from account A	Display sum of account A and B.	A is updated to 4000.
Deposir 1000 in account B		It results 4000-8000=12,000/ It show a loss of Rs 1000.  Be is update to 9000.

The above scheduler results in inconsistency of database and it shows Rs. 12,000 as sum of accounts A and B instead of Rs. 13,000. The problem occurs because second concurrently running transaction T2, reads A and B at intermediate point and computes its sum, which results in inconsistent value.

10. Poor data modeling of real world. The file based system is not able to represent the complex data and inter file relationships, which results in poor data modeling properties.

#### Data Duplication

Data duplication occurs when an exact copy of a piece of data is created

For example, copy and pasting an item called "MyPicture.jpg"

- o The new pasted item contains the exact same data as the original picture

On different Operating Systems, the naming convention for copies will change (e.g. "MyPicture2.jpg" or "MyPicturecopy.jpg")

Data duplication provides benefits such as providing us with the ability to backup copies of files and create multiple versions of a file (which may be required for progress reporting or other information)

The duplication of data is often intentional and used primarily for creating backups

Data duplication on a database may result in data redundancy, and thus an inefficient and inconsistent database

#### Data Inconsistency

Data inconsistency is a condition that occurs between files when similar data is kept in different formats in two different files, or when matching of data must be done between files. As a result of the data inconsistency, these files duplicate some data such as addresses and names, compromising data integrity.

The duplicated data, also known as the redundant data, creates unreliable information because the chances of having a value changed in one file are high, but on the other file the value remains the same. This condition of inconsistency is often experienced when using the traditional file processing, and it is very expensive and difficult to rectify such inconsistencies.

## *LackOfDataIntegration*

*Data integrity is the overall completeness, accuracy and consistency of data. This can be indicated by the absence of alteration between two instances or between two updates of a data record, meaning data is intact and unchanged. Data integrity is usually imposed during the database design phase through the use of standard procedures and rules. Data integrity can be maintained through the use of various error-checking methods and validation procedures.*

*The concept of data integrity ensures that all data in a database can be traced and connected to other data. This ensures that everything is recoverable and searchable. Having a single, well defined and well-controlled data integrity system increases stability, performance, reusability and maintainability. If one of these features cannot be implemented in the database, it must be implemented through the software.*

## *DataDependence*

*Data dependence means that one or more attributes uniquely identify other attributes of a relation. In more simple terms we can say that some data values are dependent on other data values in order to get recognized.*

*example*

*rollno name*

*In this example rollno. will be unique for each student but two students can have the same name. Suppose we want to know about the student whose name is Amit but there are two students with name Amit, so in this case name doesn't uniquely identify the student (we have confusion about which student's information we want). But if we take rollno. (say 101) then we will get information of one student whose name is Amit and rollno. is 101, hence no confusion. Therefore name is dependent on rollno.*

## *ProgramDependence*

*An application program that deals with data stored externally to it (such as in a file or a database) includes in its source code some structural definition of that data. The extent to which that program is exposed to changes made to that external source is called data dependence. A program is exposed, in this sense, if some change to the external source invalidates the program and thus necessitates changes to its source code ("unproductive maintenance").*

*By*

*"change to the external source" we normally mean structural changes of any kind—we assume application programs are immune to mere changes in the data content of the external source, such as addition and deletion of records and updates made to existing records.*

## Objective Of Database, Advantage Of database System

### The database management system

has promising potential advantages, which are explained below:

1. **Controlling Redundancy:** In file system, each application has its own private files, which cannot be shared between multiple applications. This can often lead to considerable redundancy in the stored data, which results in wastage of storage space. By having a centralized database most of this can be avoided. It is not possible that all redundancy should be eliminated. Sometimes there are sound business and technical reasons for maintaining multiple copies of the same data. In a database system, however, this redundancy can be controlled.
2. **Integrity can be enforced:** Integrity of data means that data in a database is always accurate, such that incorrect information cannot be stored in a database. In order to maintain the integrity of data, some integrity constraints are enforced on the database. ADBMS should provide capabilities for defining and enforcing the constraints.
3. **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. At such times the data is said to be inconsistent. So, if the redundancy is removed, chances of having inconsistent data is also removed.
5. **Standards can be enforced:** Since DBMS is a central system, so a standard can be enforced easily at company level, department level, national level or international level. The standardized data is very helpful during migration or interchanging of data. The file system is an independent system so a standard cannot be easily enforced on multiple independent applications.
6. **Restricting unauthorized access:** When multiple users share a database, it is likely that some users will not be authorized to access all information in the database. For example, account office data is often considered confidential, and hence only authorized persons are allowed to access such data. In addition, some users may be permitted only to retrieve data, whereas others are allowed both to retrieve and to update. Hence, the type of access operation (retrieval or update) must also be controlled.
7. **Solving Enterprise Requirement than Individual Requirement:** Since many types of users with varying levels of technical knowledge use a database, a DBMS should provide a variety of user interfaces. The overall requirements of the enterprise are more important than the individual user requirements. So, the DBA can structure the database system to provide an overall service that is "best for the enterprise".
8. **Providing Backup and Recovery:** ADBMS 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 program, the recovery subsystem is responsible for making sure that the database is restored to the state it was in before the program started executing.
9. **Cost of developing and maintaining system** is lower: It is much easier to respond to

unanticipated requests when data is centralized in a database than when it is stored in a conventional file system. Although the initial cost of setting up of a database can be large, but the cost of developing and maintaining application programs to be far lower than for similar service using conventional systems. The productivity of programmers can be higher in using non-procedural languages that have been developed with DBMS than using procedural languages.

10. **Data Model can be developed:** The centralized system is able to represent the complex data and interfile relationships, which results better data modeling properties. The data modeling properties of relational model is based on Entity and their Relationship, which is discussed in detail in chapter 4 of the book.

11. **Concurrency Control:** DBMS systems provide mechanisms to provide concurrent access of data to multiple users.

### Disadvantages of DBMS

The disadvantages of the database approach are summarized as follows:

1. **Complexity:** The provision of the functionality that is expected of a good DBMS makes the DBMS an extremely complex piece of software. Database designers, developers, database administrators and end-users must understand this functionality to take full advantage of it. Failure to understand the system can lead to bad design decisions, which can have serious consequences for an organization.

2. **Size:** The complexity and breadth of functionality makes the DBMS an extremely large piece of software, occupying many megabytes of disk space and requiring substantial amount of memory to run efficiently.

3. **Performance:** Typically, a File Based system is written for a specific application, such as invoicing. As a result, performance is generally very good. However, the DBMS is written to be more general, to cater for many applications rather than just one. The effect is that some applications may not run as fast as they used to.

4. **Higher impact of a failure:** The centralization of resources increases the vulnerability of the system. Since all users and applications rely on the availability of the DBMS, the failure of any component can bring operations to a halt.

5. **Cost of DBMS:** The cost of DBMS varies significantly, depending on the environment and functionality provided. There is also the recurrent annual maintenance cost.

6. **Additional Hardware costs:** The disk storage requirements for the DBMS and the database may necessitate the purchase of additional storage space. Furthermore, to achieve the required performance it may be necessary to purchase a larger machine, perhaps even a machine dedicated to running the DBMS. The procurement of additional hardware results in further expenditure.

7. **Cost of Conversion:** In some situations, the cost of DBMS and extra hardware may be insignificant compared with the cost of converting existing applications to run on the new DBMS and hardware. This cost also includes the cost of training staff to use the new systems and possibly the employment of specialist staff to help with conversion and running of the system. T



his  
cost is one of the main reasons why some organizations feel tied to their current systems and cannot switch to modern database technology.

## Types of Database Structure

There are several types of database management systems. Here is a list of seven common database management systems:

1. Hierarchical databases
2. Network databases
3. Relational databases
4. Object-oriented databases
5. Graph databases
6. ER model databases
7. Document databases
8. NoSQL databases

### Hierarchical Databases

In a hierarchical database management system (hierarchical DBMS) model, data is stored in a parent-child relationship nodes. In a hierarchical database, besides actual data, records also contain information about their groups of parent/child relationships.

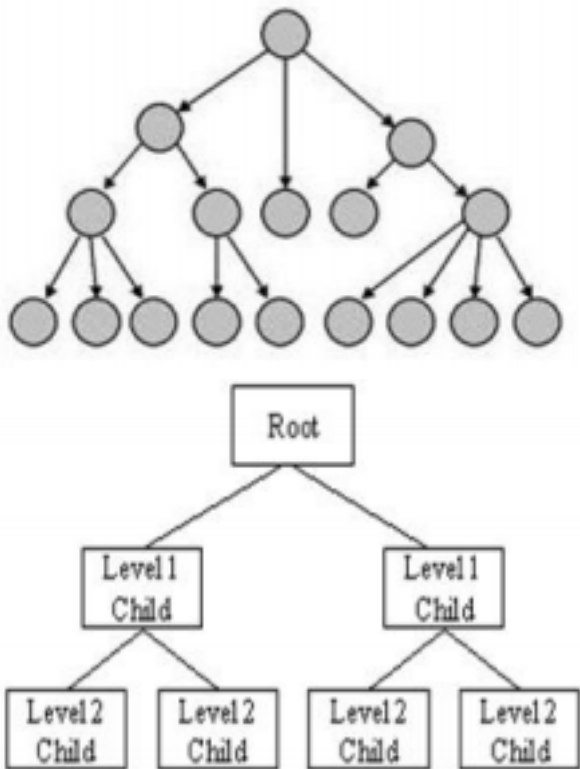
In a hierarchical database model, data is organized into a tree-like structure. The data is stored in the form of a collection of fields where each field contains only one value. The records are linked to each other via links into a parent-child relationship. In a hierarchical database model, each child record has only one parent. A parent can have multiple children.

To retrieve a field's data, we need to traverse through each tree until the record is found.

The hierarchical database system structure was developed by IBM in the early 1960s. While hierarchical structure is simple, it is inflexible due to the parent-child one-to-many relationship.

Hierarchical databases are widely used to build high performance and availability applications usually in banking and telecommunications industries.

The IBM Information Management System (IMS) and Windows Registry are two popular examples of hierarchical databases.



### Advantage

Hierarchical database can be accessed and updated rapidly because in this model structure is like a tree and the relationships between records are defined in advance. This feature is a two edged.

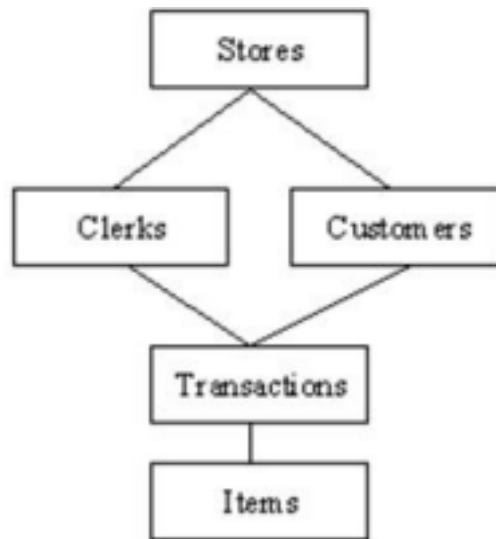
### Disadvantage

This type of database structure is that each child in the tree may have only one parent, and relationships or linkages between children are not permitted, even if they make sense from a logical standpoint. Hierarchical databases are so in their design. It can adding a new field or record requires that the entire database be redefined.

### Network Databases

Network database management systems (Network DBMSs) use a network structure to create a relationship between entities. Network databases are mainly used on large digital computers. Network databases are hierarchical databases but unlike hierarchical databases where one node can have one parent only, a network node can have a relationship with multiple entities. A network database looks more like a cobweb or an interconnected network of records.

*In network databases, children are called members and parents are called occupier. The difference between each child or member can have more than one parent.*



*The approval of the network data model is similar to a hierarchical data model. Data in a network database is organized in many-to-many relationships.*

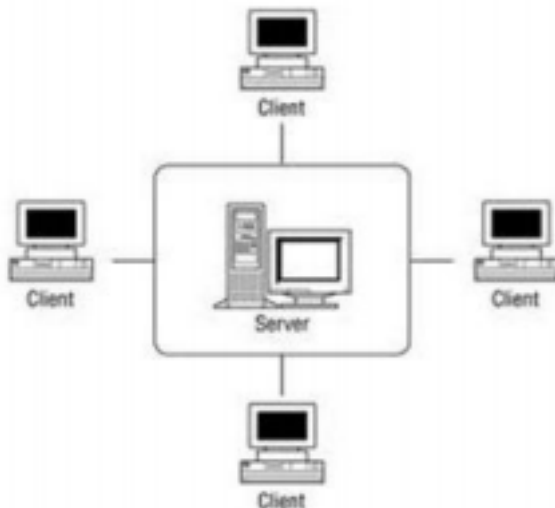
*The network database structure was invented by Charles Bachman. Some of the popular network databases are Integrated Data Store (IDS), IDMS (Integrated Database Management System), Raima Database Manager, TurboIMAGE, and Univac DMS-1100.*

### *Relational Databases*

*In relational database management systems (RDBMS), the relationship between data is relational and data is stored in a tabular form of columns and rows. Each column in a table represents an attribute and each row in a table represents a record. Each field in a table represents a data value.*

*Structured Query Language (SQL) is the language used to query a RDBMS including inserting, updating, deleting, and searching records.*

*Relational databases work on each table has a key field that uniquely indicates each row, and that these key fields can be used to connect one table of data to another.*



*Relational databases are the most popular and widely used databases. Some of the popular DDBMS are Oracle, SQL Server, MySQL, SQLite, and IBM DB2.*

*The relational database has two major reasons*

- 1. Relational databases can be used with little or no training.*
- 2. Database entries can be modified without specifying the entire body.*

*Properties of Relational Tables*

*In the relational database we have to follow some properties which are given below.*

*It's Values are Atomic  
In Each Row is alone.  
Column Values are of the Same thing.  
Columns are undistinguished.  
Sequence of Rows is Insignificant.  
Each Column has a common Name.*

*RDBMS are the most popular databases in the world.*

*Object-Oriented Model*

*It takes more than storage of programming language objects. Object DBMS's increase the semantics of the C++ and Java. It provides full-featured database programming capability, while containing native language compatibility. It adds the database functionality to object programming languages. This approach is the analogical of the application and database development into a constant data model and language environment. Applications require less code, use more natural data modeling, and code bases are easier to maintain. Object developers can write complete database applications with a decent amount of additional effort.*

*The object-oriented database derivation is the integrity of object-oriented programming language systems and consistent systems. The power of the object-oriented databases comes*

es

from the cyclical treatment of both consistent data, as found in databases, and transient data, as found in executing programs.



Object-oriented databases use small, recyclable separated software called objects.

The

objects themselves are stored in the object-oriented database. Each object contains two elements:

1. Piece of data (e.g., sound, video, text, or graphics).
2. Instructions, or software programs called methods, for what to do with the data.

Object-oriented database management systems (OODBMs) were created in early 1980s. Some

OODBMs were designed to work with OOP languages such as Delphi, Ruby, C++, Java, and Python.

Some popular OODBMs are TORNADO, Gemstone, ObjectStore, GBase, VBase, InterSystems Cache, Versant Object Database, ODABA, ZODB, Poet, JADE, and Informix.

Disadvantage of Object-oriented databases

1. Object-oriented databases have these disadvantages.
2. Object-oriented databases are more expensive to develop.
3. In the most organizations are unwilling to abandon and convert from those databases

. Benefits of Object-oriented databases

The benefits to object-oriented databases are compelling. The ability to mix and match reusable objects provides incredible multimedia capability.