

Course Contents
Unit-09: Fundamentals of Database (4 Hrs.)
<ul style="list-style-type: none">• Introduction• Database• Database System• Database Management System (DBMS)• Database System Architectures• Database Applications• Introduction to<ul style="list-style-type: none">• Data Warehousing,• Data mining, and• BigData

Course Contents
Unit-09: Fundamentals of Database (4 Hrs.)
<ol style="list-style-type: none">1. Define: (1) Database and Database management System (2) Data redundancy, (3) Data consistency2. Define: (1) Data model, (2) Schema, (3) Instance, (4) Entity, (5) Attribute3. Define a database system. List the components of database system.4. Explain in detail the components of database system.5. Explain the architecture of database system in detail.6. Name some database applications.7. What is Data warehousing? What is Data Mining?8. What is Bigdata? Explain.

Course Contents
What is Database?
<ul style="list-style-type: none">• Manual Database consists of filing cabinet, papers and people. There are formal filing method used to in/out the information from the filing cabinet. We might use calculator or computer spreadsheet to analyze the data further or to report it.• A Computer Database is nothing more than an automated version of the filing-and-retrieval functions of a manual paper filing system.

Course Contents
What is Database?
DataBase is a collection of information (data & associated objects) which help to organise the related information in a logical fashion for easy access and retrieval.
<i>Database, basically is nothing more than a computer based record-keeping system. The overall purpose of database is to record and maintain information required by organisation in decision making.</i>

Fundamentals of Database
DATABASE:
<ul style="list-style-type: none">> Database is a repository or collection of logically related, and similar data.> Database stores similar kind of data that is organized in a manner that the information can be derived from it, modified, data added, or deleted to it, and used when needed.
A database is defined as:
<ol style="list-style-type: none">(1) a collection, or repository of data,(2) having an organized structure, and(3) for a specific purpose.

Course Contents
What is Database?
<ul style="list-style-type: none">• Data that is stored more or less permanently in a computer, we term a database. The software that allows one or many persons to use and/or modify this data is a <i>database management system</i> (DBMS).• A Database Management System (DBMS) is a collection of programs that enables an organization to store, modify and extract information from the databases.• A Relational Data Base Management System (RDBMS) stores data in many related Tables. The user can ask complex questions from one or more of these related tables and answers come back to the user as Forms and Reports.

Fundamentals of Database

Database Management System:

- DBMS is a software system for creating, organizing and managing the database. DBMS handles all access to the database and manages the database. Managing the database implies that it provides a convenient environment to the user to perform operations on the database for creation, insertion, deletion, updating, and retrieval of data.
- DBMS defines the scope of the use of database. This keeps data secure from unauthorized access.
- The functionality of DBMS includes : (1) the database that contains interrelated data, and (2) a set of programs to access the data.
- The DBMS implements the three schema architecture - internal schema at internal level, conceptual schema at conceptual level, and external schema at external level.

Some of the common commercial DBMSs are—Oracle Database, IBMs DB2, Microsoft's SQL Server and Microsoft Access. MySQL is a popular open source DBMS.

Fundamentals of Database

Database System:

- A database system integrates the collection, storage, and dissemination of data required for the different operations of an organization, under a single administration.
- A database system is a computerized record keeping system. The purpose of the database system is to maintain the data and to make the information available on demand.

All organizations or enterprises have some basic common functions. They need to collect and store data, process data, and disseminate data for their various functions depending on the kind of organization. Some of the common functions include payroll, sales report etc.

Fundamentals of Database

Database System:

Components of Database System : A database system has four main components

- Users
 - (1) Application Programmers,
 - (2) End Users, and
 - (3) Data Administrators.
- Hardware
- Software
 - (1) DataBase Management System (DBMS),
 - (2) Application software, and
 - (3) User Interface.
- Data

Course Contents

Advantages (Characteristics) of Database?

- Redundancy can be reduced
- Inconsistency can be avoided

Course Contents

Advantages (Characteristics) of Database?

Figure 3.4
One possible database state for the COMPANY relational database schema.

EMPLOYEE							
Empid	Empname	SSN	Birthdate	Address	Sex	Salary	Supervisor
123456789	Smith	123456789	1968-01-08	731 Franklin, Houston, TX	M	30000	333445555
987654321	Wing	333445555	1955-12-08	638 West, Houston, TX	M	40000	888885555
555554444	Zhang	999887777	1958-01-19	3221 Castle, Spring, TX	F	25000	987654321
888884444	Walker	987654321	1941-09-20	291 Barn, Dallas, TX	F	42000	988885555
666664444	Harner	999884444	1952-09-15	975 Five Oak, Houston, TX	M	38000	333445555
454544321	English	454544321	1972-07-21	5631 Rice, Houston, TX	F	25000	333445555
987654321	Jallop	987654321	1968-02-29	1801 Dallas, Houston, TX	M	25000	987654321
888885555	Berg	888885555	1927-11-10	450 Stone, Houston, TX	M	50000	NULL

DEPARTMENT				DEPT LOCATIONS	
Deptname	Department	Mgr empid	Mgr start date	Location	City
Research	5	333445555	1988-05-22	Houston	
Administration	4	987654321	1988-01-01	Dallas	
Headquarters	1	888885555	1981-06-19	Houston	

WORKS ON				PROJECT			
Empid	Projid	Hours	Deptname	Project	Priority	Location	Deptname
123456789	1	30.0	5	Product1	1	Dallas	5
123456789	2	15.0	5	Product1	2	Houston	5
666664444	3	40.0	5	Product2	3	Houston	5
454544321	1	20.0	5	Component1	10	Dallas	4
454544321	2	20.0	5	Component2	20	Houston	4
333445555	3	15.0	5	Newsletters	30	Dallas	4
333445555	10	10.0	5				
333445555	20	10.0	5				
999887777	30	30.0	5				
999887777	10	10.0	5				
987654321	10	30.0	5				
987654321	30	15.0	5				
987654321	20	20.0	5				
987654321	30	20.0	5				
987654321	20	15.0	5				
888885555	30	10.0	5				

DEPENDENT					
Empid	Dependent empid	Sex	Birthdate	Relationship	Relationship
333445555	Alice	F	1988-04-05	Daughter	
333445555	Phredson	M	1989-10-25	Son	
333445555	Jay	F	1988-05-03	Son	
987654321	Alvin	M	1942-02-28	Spouse	
123456789	Michael	M	1989-01-04	Son	
123456789	Alice	F	1988-12-30	Daughter	
123456789	Elizabeth	F	1987-05-05	Spouse	

COMPANY DATABASE

Fundamentals of Database

Characteristics of Database Approach:

Main characteristics of the database approach:

- Data Redundancy is Minimized: Database system keeps data at one place in the database. The data is integrated into a single, logical structure. Different applications refer to the data from the centrally controlled location. The storage of the data, centrally, minimizes data redundancy.
- Data Inconsistency is Reduced: Minimizing data redundancy using database system reduces data inconsistency too. Updating of data values becomes simple and there is no disagreement in the stored values.
- Data is Shared: Data sharing means sharing the same data among more than one user. Each user has access to the same data, though they may use it for different purposes. Authorized users are permitted to use the data from the database. Users are provided with views of the data to facilitate its use.

Fundamentals of Database

Characteristics of Database Approach:

- **Data Independence:** It is the separation of data description (metadata) from the application programs that use the data. Data descriptions are stored in a central location called the data dictionary. This property allows an organization's data to change and evolve (within limits) without changing the application programs that process the data.
- **Data Integrity maintain:** Stored data is changed frequently for variety of reasons such as adding new data item types, and changing the data formats. The integrity and consistency of the database are protected using constraints on values that data items can have.
- **Data Security improve:** The database is kept secure by limiting access to the database by authorized personnel. Authorized users are generally restricted to the particular data they can access, and whether they can update it or not. Access is often controlled by passwords.

Fundamentals of Database

Characteristics of Database Approach:

- **Backup and Recovery Support :** Backup and recovery are supported by the software that logs changes to the database. This support helps in recovering the current state of the database in case of system failure.
- **Standards are Enforced:** Since the data is stored centrally, it is easy to enforce standards on the database. Standards could include the naming conventions, and standard for updating, accessing and protecting data. Tools are available for developing and enforcing standards.
- **Application Development Time is Reduced:** The database approach greatly reduces the cost and time for developing new business applications. Programmer can focus on specific functions required for the new application, without having to worry about design, or low-level implementation details; as related data have already been designed and implemented. Tools for the generation of forms and reports are also available.

Course Contents

Database Models?

Data models are a collection of conceptual tools for describing **data, data relationships, data semantics** and **data constraints**. It is a graphical representation of data.

Data models are groups are as:

1. Entity-Relationship Model
2. Hierarchical
3. Network Models.
4. Relational Model:

Fundamentals of Database

Schema and Instances:

- **Schema** is the logical structure of the database. A schema contains information about the descriptions of the database like the names of the record type, the data items within a record type, and constraints. A schema does not show the data in the database. The database schema does not change frequently.
- **Instances** are the actual data contained in the database at a particular point of time. The content of the database may change from time to time.

	SSN	Name	Address	Department	Birthday
STUDENT					

	DName	Dept_Address	Dep_Head	Phone
DEPARTMENT				

Fundamentals of Database

Schema and Instances:

	SSN	Name	Address	Department	Birthday
STUDENT	001	Shankar	Kathmandu	Computer Sc.	04/01/2031
	002	Hari	Lalitpur	Computer Sc.	04/02/2032
	003	Shyam	Lalitpur	Physics	04/03/2033
	004	Ram	Kathmandu	Physics	05/12/2034
	005	Deepak	Kirtipur	Chemistry	06/11/2035

	DName	Dept_Address	Dep_Head	Phone
DEPARTMENT	Computer Sc.	Lalitpur	AAA	4444444
	Physics	Lalitpur	BBB	5555555
	Chemistry	Lalitpur	CCC	6666666

Fundamentals of Database

Database System Architectures

- The architecture of a database system provides a general framework for database systems. Different database systems, small or big, may not support all aspects of the architecture; however, different systems can be matched to the framework.
- ANSI/SPARC study group proposed the architecture for database and is called ANSI/SPARC architecture.
- The purpose of the architecture is to make databases more independent of the application that is using the database.

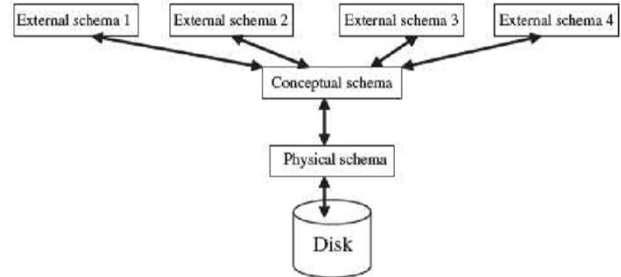
Fundamentals of Database

Database System Architectures: The ANSI/SPARC architecture is divided into three levels (Figure 12.9), as follows:

- Internal (Physical) Level: The internal schema describes the physical storage structure of the database. It is concerned about how data is stored physically. It describes the organization of files, the access path to the database etc. The physical data model is used to describe the physical schema.
- Conceptual Level: The conceptual schema describes the structure of whole database for the users. It describes the entities in the database, their relationships, and constraints. The representation or implementation data model is used to describe the conceptual schema.
- External Level or View Level: provides a user's and application's view of the data. It includes one or more external schema. For a particular user, the external schema describes the structure of the database relevant to it, and hides rest of the information.

Fundamentals of Database

Database System Architectures: The ANSI/SPARC architecture is divided into three levels (Figure 12.9), as follows:



Fundamentals of Database

Database Applications:

Databases range from those designed for a single user with a desktop computer to those on mainframe computers with thousands of users.

The database applications can be for different purposes like

1. Personal databases that support one user with a stand-alone personal computer,
2. Workgroup databases for a small team of people (less than 25) who work in collaboration on a project,
3. Departmental databases designed to support the various functions and activities of a department (a functional unit of an organization), and
4. Enterprise databases to support organization-wide operations and decision making. Data warehouse is an enterprise database.

Fundamentals of Database

- Introduction to
 - Data Warehousing,
 - Data mining, and
 - BigData

Fundamentals of Database

Data Warehouse and Data Mining:

- Data warehouse is a collection of data designed to support management decision making. It is the main repository of an organization's historical data in its corporate memory.
- The term data warehouse generally refers to the combination of many different databases across an entire enterprise.
- The main purpose of DW is that a data analyst can perform complex queries and analysis, such as data mining on the information without slowing down the operational systems.

Fundamentals of Database

Goal of Data Warehouse:

- Provide access to the data of an organization
- Data consistency
- Capacity to separate and combine data
- Inclusion of tools set to query, analyze and prevent information
- Publish used data
- Drive business re-engineering

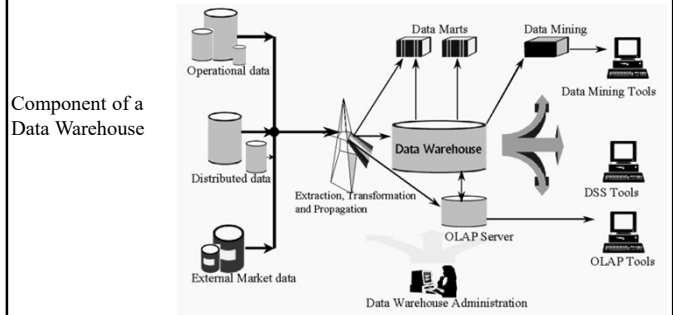
Fundamentals of Database

Characteristics of Data Warehouse:

1. Subject Oriented : Data to be organized towards the major subject areas of organization i.e. an insurance company orient data by customer, premium and claims etc.
2. Integrated : The way data is extracted and transformed is uniform, regardless of the original source.
3. Non-volatile: A data warehouse is not updated in real-time. It is periodically updated via the uploading of data, protecting it from the influence of momentary change.
4. Time-Variant: Data is organized via time-periods (weekly, monthly, annually, etc).

Fundamentals of Database

Data Warehouse and Data Mining:



Fundamentals of Database

USES/ADVANTAGES OF DATA WAREHOUSE

- Easy to retrieve and find information
- More cost-effective decision making
- Better enterprise intelligence
- Fast retrieval of information
- Help management in decision making support system
- Enhance the value of Business
- Enhance customer service

Fundamentals of Database

Data Mining:

Generally Data mining (called knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful to increase revenue, cuts, costs or both.

Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

Fundamentals of Database

Data Mining:

Data mining is the principle of sorting through large amounts of data and picking out relevant information. It is usually used by business intelligence organizations, and financial analysts, but it is increasingly used in the sciences to extract information from the enormous data sets generated by modern experimental and observational methods. It has been described as "the nontrivial extraction of implicit, previously unknown, and potentially useful information from data" and "the science of extracting useful information from large data sets or databases".

Fundamentals of Database

Advantages of Data Mining:

- Data Mining tool is used for searching valuable business information in a large database. Data Mining technology can help business by providing following capabilities:
- Automated prediction of trends and behaviours: Data mining technologies automates the process of finding predicting information in large databases. Statistical analytical tool can use to predict the behaviours / trends of business from analyzing large databases.
- Automated discovery of previously unknown patterns: Data mining tools can also identify hidden patterns like fraudulent credit card transaction; finding fraudulent (untrue) VAT bill as of recent news.
- Databases can be large in both depth and breadth: Usually, analysts must often limit the number of variables they examine when doing hands-on analysis due to time constrains. But with the evolution of data mining technologies, analysis being done in full databases (or samples) and yield lower estimation errors and variance.

Fundamentals of Database

- Introduction to BigData

“Big data is high-volume, high-velocity and/or high-variety information assets that demand cost effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation”

Fundamentals of Database

- Introduction to BigData

Volume:

Large amounts of data, terabytes to zettabyte

Velocity: Large amounts of data from transactions with high refresh rate, time to act on the basis of these data streams will often be very short, batch processing to real time streaming, 30 KB to 30 GB per second

Variety: Data from different data sources (enterprise data, experimental/observational data, social media) , data types (structured, unstructured, and semi structured)

Fundamentals of Database

- Introduction to
 - BigData

Challenges and Characteristics of Big Data (3V + 2V)

