**THE CATHOLIC UNIVERSITY OF EASTERN AFRICA**

**COMP 214: DATABASE SYSTEMS**

**INTRODUCTION**

**Definitions**

* **Data**

Generally, data is a [set](http://en.wikipedia.org/wiki/Set_%28mathematics%29) of [values](http://en.wikipedia.org/wiki/Value_%28computer_science%29) of [qualitative](http://en.wikipedia.org/wiki/Qualitative_data) or [quantitative](http://en.wikipedia.org/wiki/Quantitative_data) [variables](http://en.wikipedia.org/wiki/Variable_and_attribute_%28research%29). Data is [measured](http://en.wikipedia.org/wiki/Measurement), [collected and reported](http://en.wikipedia.org/wiki/Data_reporting), and [analyzed](http://en.wikipedia.org/wiki/Data_analysis), whereupon it can be [visualized](http://en.wikipedia.org/wiki/Data_visualization) using graphs or images. Data as a general [concept](http://en.wikipedia.org/wiki/Concept) refers to the fact that some existing [information](http://en.wikipedia.org/wiki/Information) or [knowledge](http://en.wikipedia.org/wiki/Knowledge) is [represented](http://en.wikipedia.org/wiki/Knowledge_representation_and_reasoning) or [coded](http://en.wikipedia.org/wiki/Code) in some form suitable for better usage or [processing](http://en.wikipedia.org/wiki/Data_processing).

Data is also defined as Symbols or [signals](http://www.businessdictionary.com/definition/signal.html) that are [input](http://www.businessdictionary.com/definition/input.html), stored, and processed by a computer, for [output](http://www.businessdictionary.com/definition/output.html) as usable information.

* **Information**

It’s that which informs, i.e. an answer to a question

At its most fundamental, information is any propagation of cause and effect within a system. Information is [conveyed](http://en.wikipedia.org/wiki/Conveyed_concept) either as the content of a [message](http://en.wikipedia.org/wiki/Message) or through direct or indirect [observation](http://en.wikipedia.org/wiki/Observation) of something. That which is [perceived](http://en.wikipedia.org/wiki/Perception) can be construed as a message in its own right, and in that sense, information is always conveyed as the content of a [message](http://en.wikipedia.org/wiki/Message).

Information can be [encoded](http://en.wikipedia.org/wiki/Code) into various forms for [transmission](http://en.wikipedia.org/wiki/Transmission_%28telecommunications%29) and [interpretation](http://en.wikipedia.org/wiki/Language_interpretation) (for example, information may be encoded into [signs](http://en.wikipedia.org/wiki/Sign_%28semiotics%29), and transmitted via [signals](http://en.wikipedia.org/wiki/Signal_%28electronics%29)). It can also be [encrypted](http://en.wikipedia.org/wiki/Encrypted) for safe storage and communication.

Therefore, Information can be defined as ; [Data](http://www.businessdictionary.com/definition/data.html) that is (1) [accurate](http://www.businessdictionary.com/definition/accurate.html) and timely, (2) specific and [organized](http://www.businessdictionary.com/definition/organized.html) for a purpose, (3) presented within a [context](http://www.businessdictionary.com/definition/context.html) that gives it meaning and relevance, and (4) can [lead](http://www.businessdictionary.com/definition/sales-lead.html) to an increase in understanding and decrease in [uncertainty](http://www.businessdictionary.com/definition/uncertainty.html).

* **Database**

A shared collection of logically related data (and a description of this data), designed to meet the information needs of an organization. Logically related data comprises entities, attributes, and relationships of an organization’s information.

**Characteristics of Database**

The data in a database should have the following features:

* **Organized/Related.** It should be well organized and related.
* **Shared.** Data in a database are shared among different users and applications.
* **Permanent or Persistence.** Data in a database exist permanently in the sense the data can live beyond the scope of the process that created it.
* **Validity/integrity/Correctness.** Data should be correct with respect to the real-world entity that they represent.
* **Security.** Data should be protected from unauthorized access.
* **Consistency.** Whenever more than one data element in a database represents related real-world values, the values should be consistent with respect to the relationship.
* **Non-redundancy:** No two data items in a database should represent the same real-world entity.
* **Independence.** Data at different levels should be independent of each other so that the changes in one level should not affect the other levels.
* **Easily Accessible.** It should be available when and where it is needed i.e. it should be easily accessible.
* **Recoverable.** It should be recoverable in case of damage.
* **Flexible to change.** It should be flexible to change.
* **Advantages of Database**

1. Reducing Data Redundancy - The file-based data management systems contained multiple files that were stored in many different locations in a system or even across multiple systems. Because of this, there were sometimes multiple copies of the same file which lead to data redundancy.  This is prevented in a database as there is a single database and any change in it is reflected immediately. Because of this, there is no chance of encountering duplicate data.
2. Sharing of Data- In a database, the users of the database can share the data among themselves. There are various levels of authorization to access the data, and consequently the data can only be shared based on the correct authorization protocols being followed.
3. Data Integrity- Data integrity means that the data is accurate and consistent in the database.
4. Data Security- Data Security is vital concept in a database. Only authorized users should be allowed to access the database and their identity should be authenticated using a username and password.
5. Privacy-The privacy rule in a database means only the authorized users can access a database according to its privacy constraints.
6. Backup and Recovery - Database Management System automatically takes care of backup and recovery.
7. Data Consistency- All data appears consistently across the database and the data is same for all the users viewing the database. Any changes made to the database are immediately reflected to all the users and there is no data inconsistency.

* **Database Management System (DBMS)**

A software system that enables users to define, create, and maintain the database and that provides controlled access to this database.

**Advantages of DBMSs**

* Control of data redundancy
* Data consistency
* More information from the same amount of data
* Sharing of data
* Improved data integrity
* Improved security
* Enforcement of standards
* Economy of scale
* Balanced conflicting requirements
* Improved data accessibility and responsiveness
* Increased productivity
* Improved maintenance through data independence
* Increased concurrency
* Improved backup and recovery services

**Disadvantages of DBMSs**

* Complexity
* Size
* Cost of DBMS
* Additional hardware costs
* Cost of conversion
* Performance
* Vulnerability to attack
* Higher impact of a failure

**What is a Database Management System (or DBMS)?**

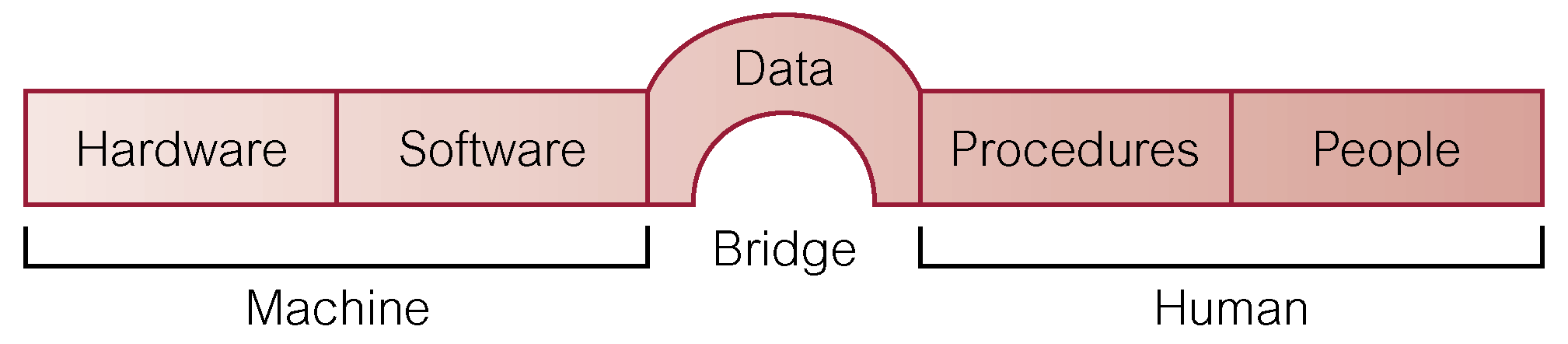
Organizations employ Database Management Systems (or DBMS) to help them effectively manage their data and derive relevant information out of it. A DBMS is a technology tool that directly supports data management. It is a package designed to define, manipulate, and manage data in a database.

Some general functions of a DBMS:

* Designed to allow the definition, creation, querying, update, and administration of databases
* Define rules to validate the data and relieve users of framing programs for data maintenance
* Convert an existing database, or archive a large and growing one
* Run business applications, which perform the tasks of managing business processes, interacting with end-users and other applications, to capture and analyze data

Some well-known DBMSs are Microsoft SQL Server, Microsoft Access, Oracle, SAP, and others.

**Components of DBMS Environment**



1. **Hardware**

-Can range from a PC to a network of computers.

1. **Software**

-DBMS, operating system, network software (if necessary) and also the application programs.

1. **Data**

-Used by the organization and a description of this data called the schema.

1. **Procedures**

-Instructions and rules that should be applied to the design and use of the database and DBMS.

1. **People**
2. **COMPONENTS OF DBMS in depth**

DBMS have several components, each performing very significant tasks in the database management system environment. Below is a list of components within the database and its environment.

Software Icon***Software***  
This is the set of programs used to control and manage the overall database. This includes the DBMS software itself, the Operating System, the network software being used to share the data among users, and the application programs used to access data in the DBMS.

Hardware Icon***Hardware***  
Consistsof a set of physical electronic devices such as computers, I/O devices, storage devices, etc., this provides the interface between computers and the real world systems.

Data Icon***Data***  
DBMS exists to collect, store, process and access data, the most important component. The database contains both the actual or operational data and the metadata.

Procedures Icon***Procedures***  
These are the instructions and rules that assist on how to use the DBMS, and in designing and running the database, using documented procedures, to guide the users that operate and manage it.

Access Language***Database Access Language***

This is used to access the data to and from the database, to enter new data, update existing data, or retrieve required data from databases. The user writes a set of appropriate commands in a database access language, submits these to the DBMS, which then processes the data and generates and displays a set of results into a user readable form.

Query Processor Icon***Query Processor***

This transforms the user queries into a series of low level instructions. This reads the online user’s query and translates it into an efficient series of operations in a form capable of being sent to the run time data manager for execution.

Run Time***Run Time Database Manager***

Sometimes referred to as the database control system, this is the central software component of the DBMS that interfaces with user-submitted application programs and queries, and handles database access at run time. Its function is to convert operations in user’s queries. It provides control to maintain the consistency, integrity and security of the data.

Data Manager Icon***Data Manager***

Also called the cache manager, this is responsible for handling of data in the database, providing a recovery to the system that allows it to recover the data after a failure.

Database Engine Icon***Database Engine***

The core service for storing, processing, and securing data, this provides controlled access and rapid transaction processing to address the requirements of the most demanding data consuming applications. It is often used to create relational databases for online transaction processing or online analytical processing data.

Data Dictionary Icon***Data Dictionary***

This is a reserved space within a database used to store information about the database itself. A data dictionary is a set of read-only table and views, containing the different information about the data used in the enterprise to ensure that database representation of the data follow one standard as defined in the dictionary.

Report Writer Icon***Report Writer***

Also referred to as the report generator, it is a program that extracts information from one or more files and presents the information in a specified format. Most report writers allow the user to select records that meet certain conditions and to display selected fields in rows and columns, or also format the data into different charts.

1. **CHARACTERISTICS**

Traditionally, data was organized in file formats. DBMS was a new concept then, and all the research was done to make it overcome the deficiencies in traditional style of data management. A modern DBMS has the following characteristics −

* **Real-world entity**

A modern DBMS is more realistic and uses real-world entities to design its architecture. It uses the behavior and attributes too. For example, a school database may use students as an entity and their age as an attribute.

* **Relation-based tables**

DBMS allows entities and relations among them to form tables. A user can understand the architecture of a database just by looking at the table names.

* **Isolation of data and application**

A database system is entirely different than its data. A database is an active entity, whereas data is said to be passive, on which the database works and organizes. DBMS also stores metadata, which is data about data, to ease its own process.

* **Less redundancy**

DBMS follows the rules of normalization, which splits a relation when any of its attributes is having redundancy in values. Normalization is a mathematically rich and scientific process that reduces data redundancy.

* **Consistency**

Consistency is a state where every relation in a database remains consistent. There exist methods and techniques, which can detect attempt of leaving database in inconsistent state. A DBMS can provide greater consistency as compared to earlier forms of data storing applications like file-processing systems.

* **Query Language**

DBMS is equipped with query language, which makes it more efficient to retrieve and manipulate data. A user can apply as many and as different filtering options as required to retrieve a set of data. Traditionally it was not possible where file-processing system was used.

* **ACID Properties**

DBMS follows the concepts of **A**tomicity, **C**onsistency, **I**solation, and **D**urability (normally shortened as ACID). These concepts are applied on transactions, which manipulate data in a database. ACID properties help the database stay healthy in multi-transactional environments and in case of failure.

* **Multiuser and Concurrent Access**

DBMS supports multi-user environment and allows them to access and manipulate data in parallel. Though there are restrictions on transactions when users attempt to handle the same data item, but users are always unaware of them.

* **Multiple views**

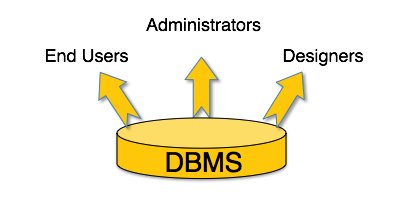
DBMS offers multiple views for different users. A user who is in the Sales department will have a different view of database than a person working in the Production department. This feature enables the users to have a concentrate view of the database according to their requirements.

* **Security**

Features like multiple views offer security to some extent where users are unable to access data of other users and departments. DBMS offers methods to impose constraints while entering data into the database and retrieving the same at a later stage. DBMS offers many different levels of security features, which enables multiple users to have different views with different features. For example, a user in the Sales department cannot see the data that belongs to the Purchase department. Additionally, it can also be managed how much data of the Sales department should be displayed to the user. Since a DBMS is not saved on the disk as traditional file systems, it is very hard for miscreants to break the code.

1. **USERS**

A typical DBMS has users with different rights and permissions who use it for different purposes. Some users retrieve data and some back it up. The users of a DBMS can be broadly categorized as follows −



1. **Administrators**

Administrators maintain the DBMS and are responsible for administrating the database. They are responsible to look after its usage and by whom it should be used. They create access profiles for users and apply limitations to maintain isolation and force security. Administrators also look after DBMS resources like system license, required tools, and other software and hardware related maintenance.

1. **Designers**

Designers are the group of people who actually work on the designing part of the database. They keep a close watch on what data should be kept and in what format. They identify and design the whole set of entities, relations, constraints, and views.

1. **End Users**

End users are those who actually reap the benefits of having a DBMS. End users can range from simple viewers who pay attention to the logs or market rates to sophisticated users such as business analysts.