



# National Institute of Technology Goa

राष्ट्रीय प्रौद्योगिकी संस्थान गोवा

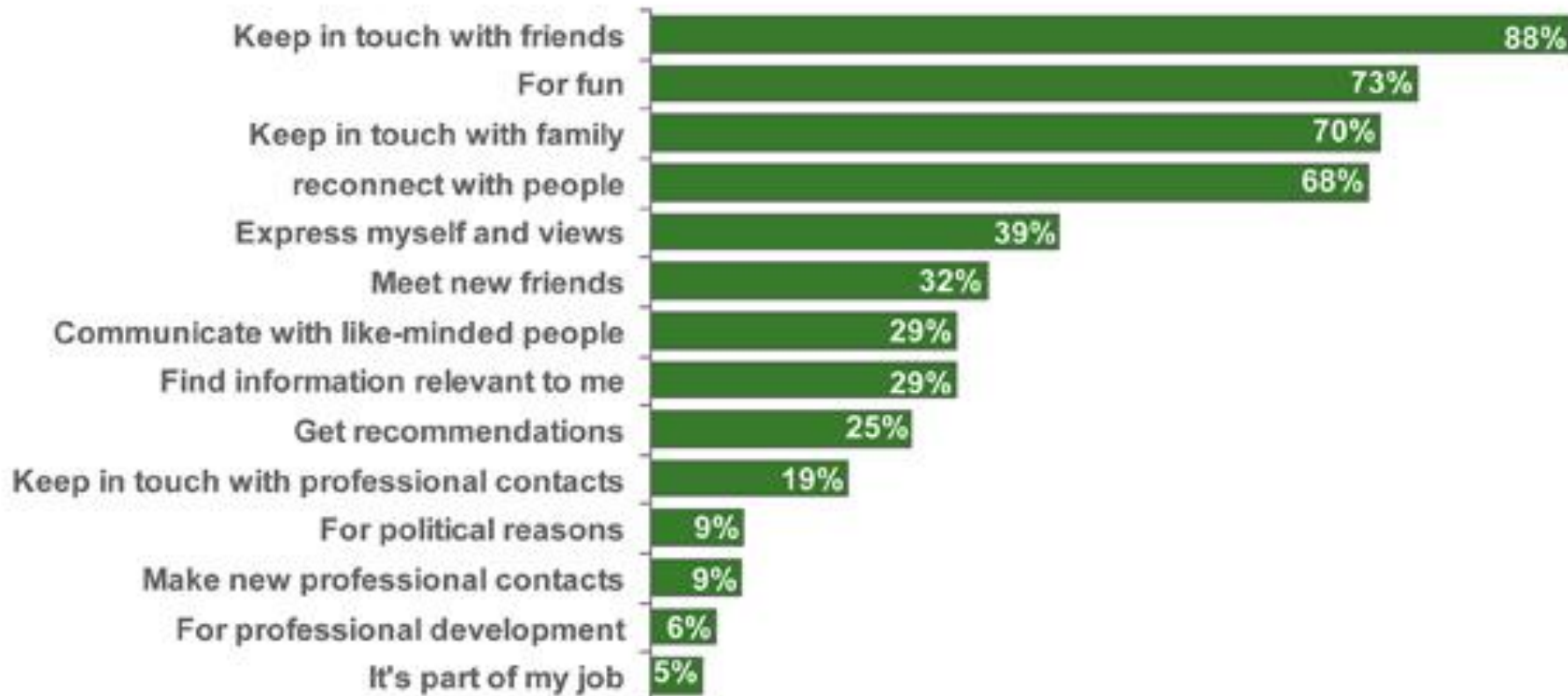
<b>Subject Code</b> CS 301	<b>Database Systems (DS)</b>	<b>Credits: 4 (3-1-0)</b> <b>Total hours: 56</b>
<b>Course Objectives</b>	This course covers the relational database systems RDBS - the predominant system for business, scientific and engineering applications at present. The topics are reinforced using tools such as Oracle server in labs. The course includes entity-relation model, normalization, relational model, relational algebra, and data access queries as well as an introduction to SQL.	

<b>Module 1</b>	<b>12 Hours</b>
Introduction: An overview of database management system, database system vs file system, database system concept and architecture, data model schema and instances, data independence and database language and interfaces,(DDL,DML,DCL), overall database structure, database users. Data modeling using the Entity Relationship model: ER model concepts, notation for ER diagram, mapping constraints, keys, specialization, generalization, aggregation, reduction of an ER diagrams to tables, extended ER model, relationship of higher degree.	
<b>Module 2</b>	<b>14 Hours</b>
Relational data Model and Language: Relational data model concepts, integrity constraints, entity integrity, referential integrity, key constraints, domain constraints, relational algebra, relational calculus, tuple and domain calculus.Introduction on SQL: Characteristics of SQL, advantage of SQL, SQL data type and literals, types of SQL commands, SQL operators and their procedure, tables, views and indexes, queries and sub queries, aggregate functions, insert, update and delete operations, joins, unions, intersection, minus, cursors, triggers, procedures in SQL/PL SQL.	
<b>Module 3</b>	<b>18 Hours</b>
Data Base Design & Normalization: Functional dependencies, primary key, foreign key, candidate key, super key, normal forms, first, second, third normal forms, BCNF, 4th Normal form,5th normal form, loss less join decompositions, canonical cover, redundant cover, synthesis the set of relation , MVD, and JDs,inclusion dependence, transaction processing concept, transaction system, testing ofserializability, serializability of schedules, conflict & view serializable schedule, recoverability, Recovery from transaction failures, log based recovery, deadlock handling.	
<b>Module 4</b>	<b>12 Hours</b>
Concurrency Control Techniques: Concurrency control, locking techniques for concurrency control, 2PL, time stamping protocols for concurrency control, validation based protocol, multiple granularity, multi version schemes and recovery with concurrent transaction. Storage: Introduction, secondary storage devices, tertiary storage, buffering of blocks, structure of files, file organization, indexing and hashing, types of single level ordered indexes, multilevel indexes, dynamics multilevel indexes using B-trees and B+- Trees, database security.	

<b>Reference books</b>	<ol style="list-style-type: none"><li>(1) Korth, Silberschatz, “Database System Concepts”, 4<sup>th</sup> ed., TMH, 2003.</li><li>(2) Elmsari and Navathe, “Fundamentals of Database Systems”, 4<sup>th</sup> ed., A. Wesley, 2004</li><li>(3) Raghu Ramakrishnan , Johannes Gehrke, “ Database Management Systems”, 3<sup>rd</sup> Edition, McGraw- Hill, 2003.</li><li>(4) J D Ullman, “Principles of database systems”, Computer Science Press, 2001.</li></ol>
----------------------------	--

# Social Networking drivers

## Why do you use social networking sites?



Base: Visit social network last 30 days  
Simmons New Media Study, Fall 2009





# DATA NEVER SLEEPS

How Much Data Is Generated Every Minute?

By 2015, it is estimated that there will be 2.1 billion people on the planet. Of that, 1.5 billion will be using mobile devices. That's a lot of data. And it's growing. Every minute, 217 terabytes of data are generated. That's 217 million gigabytes. That's 217 billion megabytes. That's 217 trillion bytes. That's a lot of data. And it's growing.



WITH NO SIGNS OF SLOWING, THE DATA KEEPS GROWING.

There are now more than 2.1 billion people on the planet. And it's growing. Every minute, 217 terabytes of data are generated. That's 217 million gigabytes. That's 217 billion megabytes. That's 217 trillion bytes. That's a lot of data. And it's growing.

The global internet population

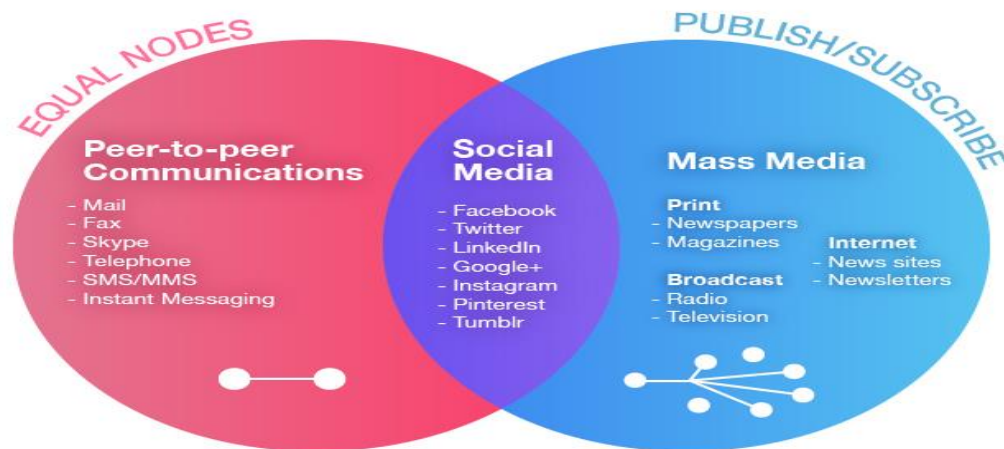
2.1 BILLION PEOPLE

# SOCIAL MEDIA

**Social media is defined as a group of Internet-based applications that allow the creation and exchanges of user-generated content.**

**Social media gives users an easy-to-use way to communicate and network with each other on an unprecedented scale.**

**Facebook, the social networking site, recorded more than 845 million active users as of December 2011. In the third quarter of 2012, the number of active Facebook users was 1 billion. As of the first quarter of 2015, Facebook had 1.44 billion monthly active users..**



# Classification of Social Media

## Nine different types of social media:

- **Online social networking**

facebook

- **Blogging**



- **Micro-Blogging**



- **Wikis**



- **Social News**



reddit

Wikis

- **Social Bookmarking**



StumbleUpon

- **Media Sharing**

You Tube

- **Opinion, Reviews, and Ratings**



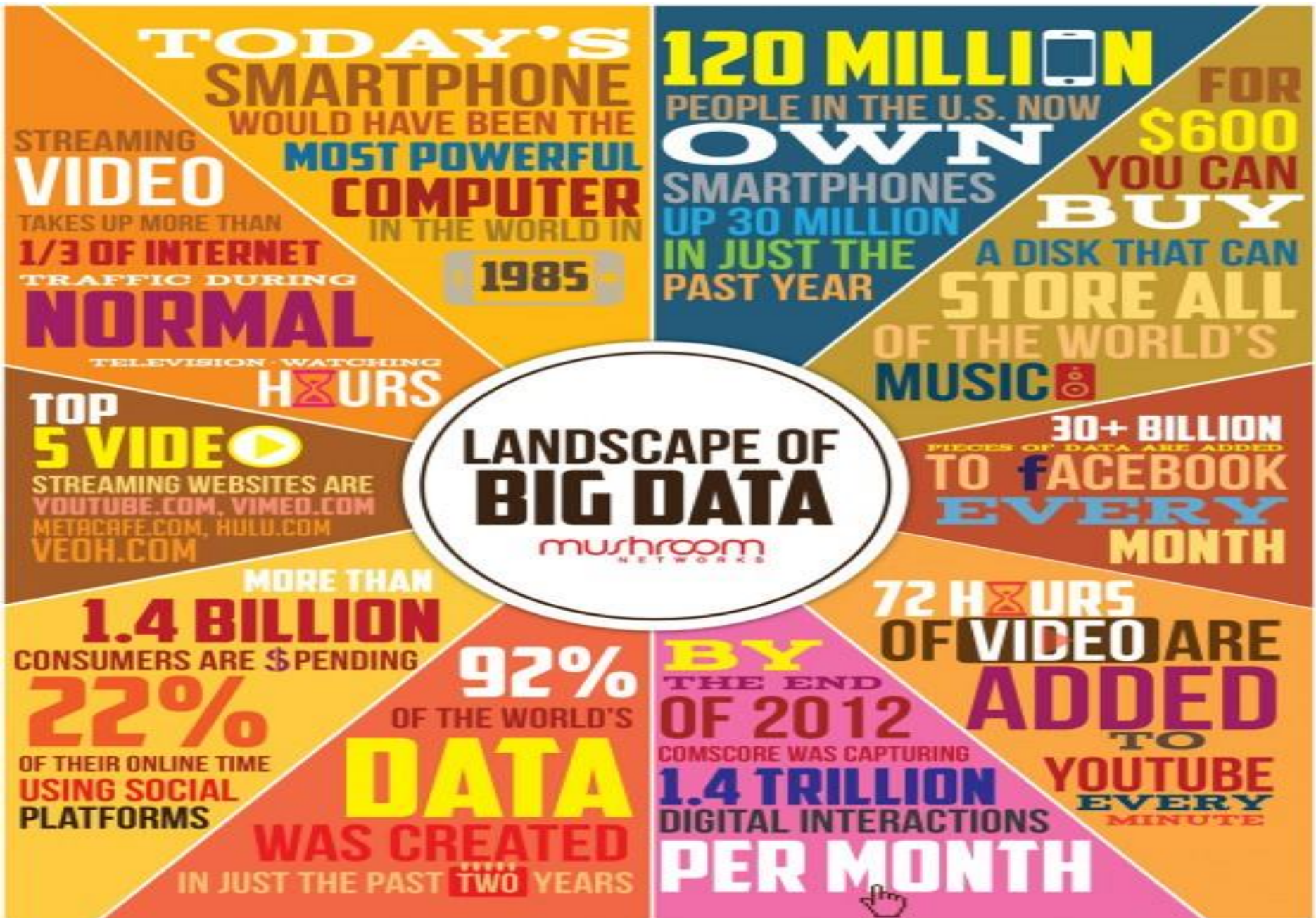
Epinions.com

- **Answers**





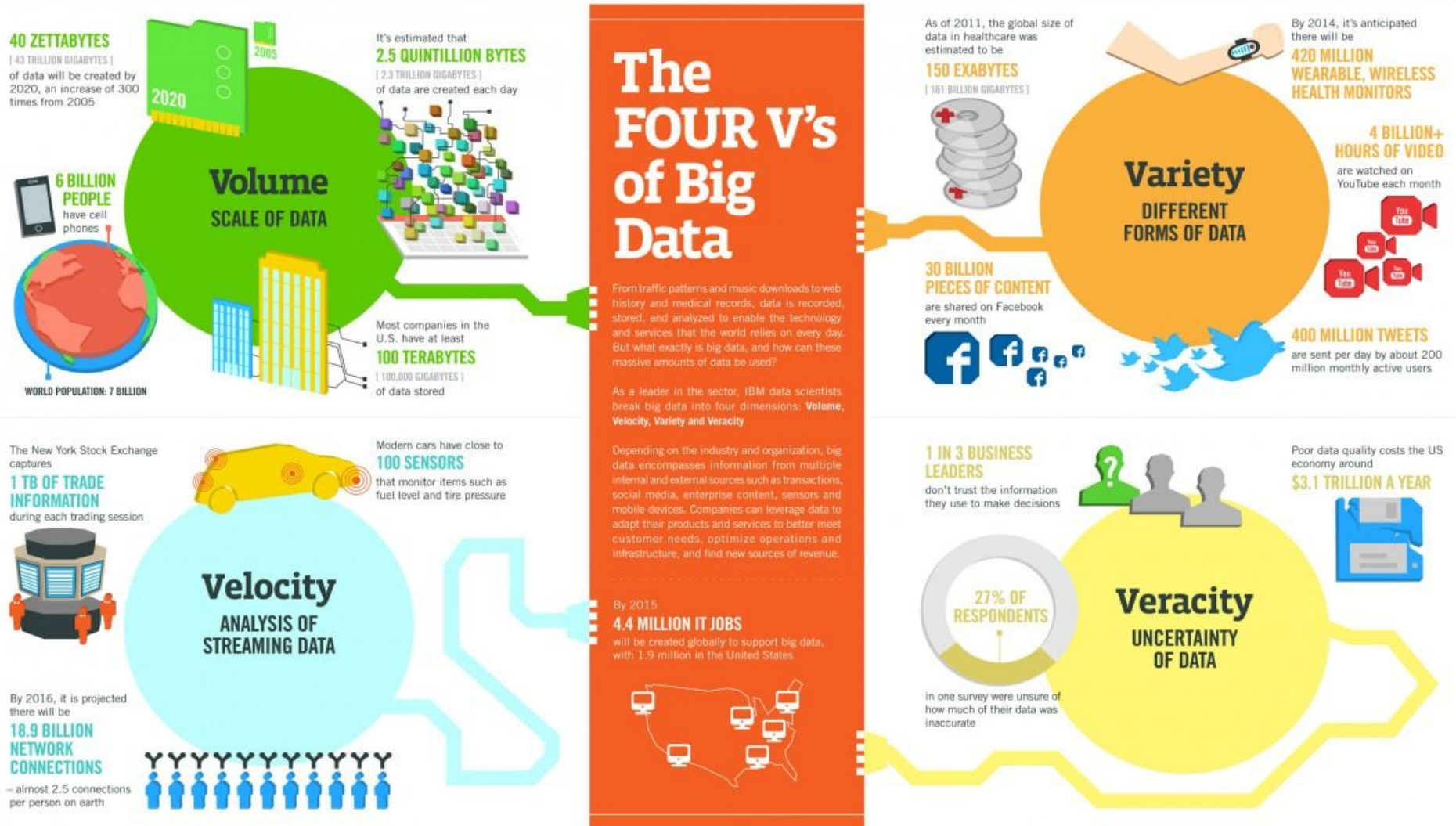
# Introduction



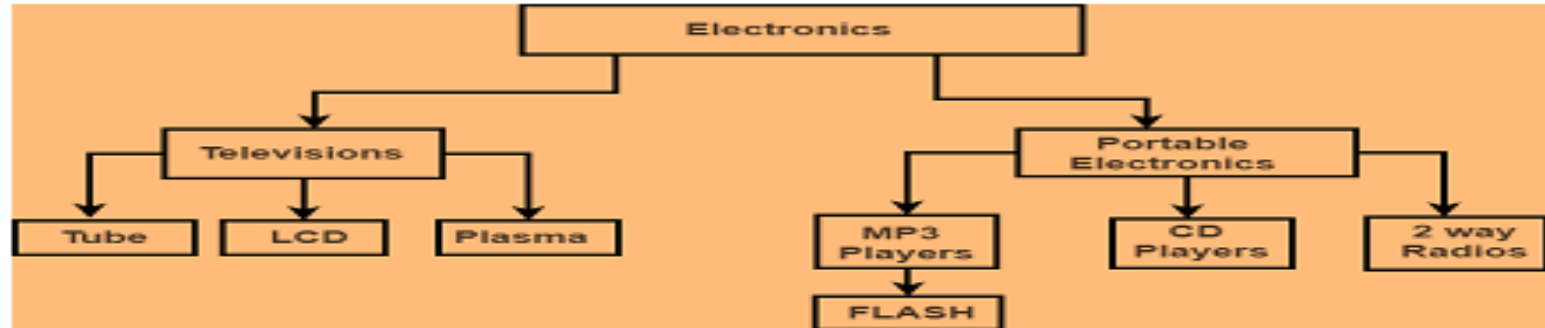


## Introduction...contd.

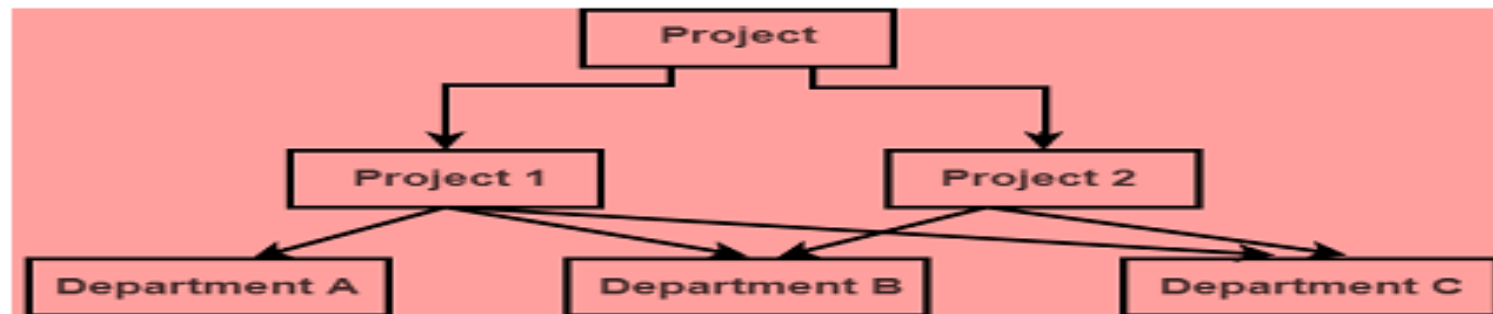
- Mining big data streams faces four principal challenges: volume, velocity, variety and veracity.



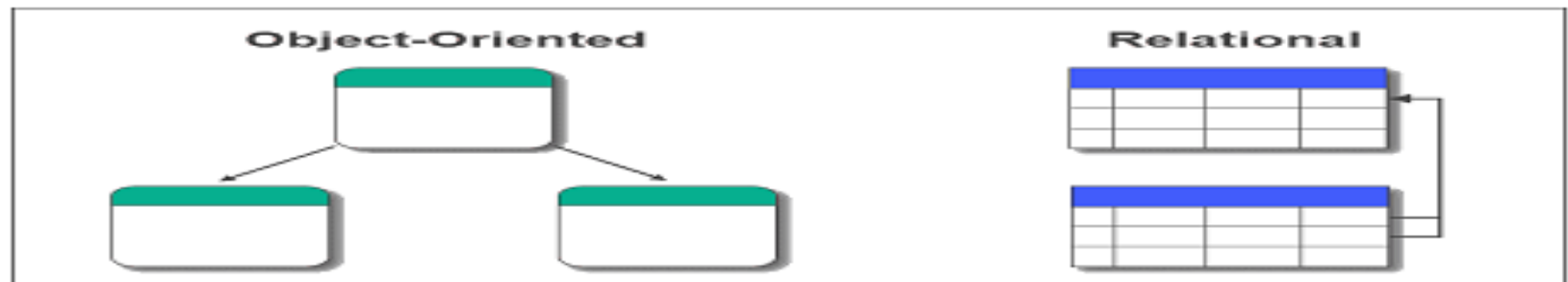
## Higherarchial Database



## Network Database



## Object Oriented and Relational Database



# Database Systems

- **Data:**

Data are recorded representation of physical objects, abstract things, events and facts which affects in decision making.

- **Data Structure:**

Data structure is a particular way of organizing data in a computer so that it can be used efficiently.

Different kinds of data structures are suited to different kinds of applications, and some are highly specialized to specific tasks.

- **Database:**

Database is an interrelated data which related to a particular enterprise.

or

Database is an collection of data organized in a particular way.

or

In simple terms we can say database is an interrelated data.

# Database Systems...contd.

- **DBMS is a collection of interrelated data and set of programs to access those data.**
- **The primary goal of DBMS is to provide a way to store and retrieve database formation that is both convenient and efficient.**
- **The database system must ensure the safety of information stored, despite system crashes or attempts unauthorized access.**

## **An environment that is both convenient and efficient to use 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
- **Databases touch all aspects of our lives**



# Database Systems...contd.

- **Database:**

**Database is an interrelated data which related to a particular enterprise.**

**or**

**Database is an collection of data organized in a particular way.**

**or**

**In simple terms we can say database is an interrelated data.**

- **DBMS is a collection of interrelated data and set of programs to access those data.**
- **The primary goal of DBMS is to provide a way to store and retrieve database formation that is both convenient and efficient.**
- **The database system must ensure the safety of information stored, despite system crashes or attempts unauthorized access.**

# Database Systems...contd.

- **Database:**

**Database is an interrelated data which related to a particular enterprise.**

**or**

**Database is an collection of data organized in a particular way.**

**or**

**In simple terms we can say database is an interrelated data.**

- **DBMS is a collection of interrelated data and set of programs to access those data.**
- **The primary goal of DBMS is to provide a way to store and retrieve database formation that is both convenient and efficient.**
- **The database system must ensure the safety of information stored, despite system crashes or attempts unauthorized access.**

# Database Systems...contd.

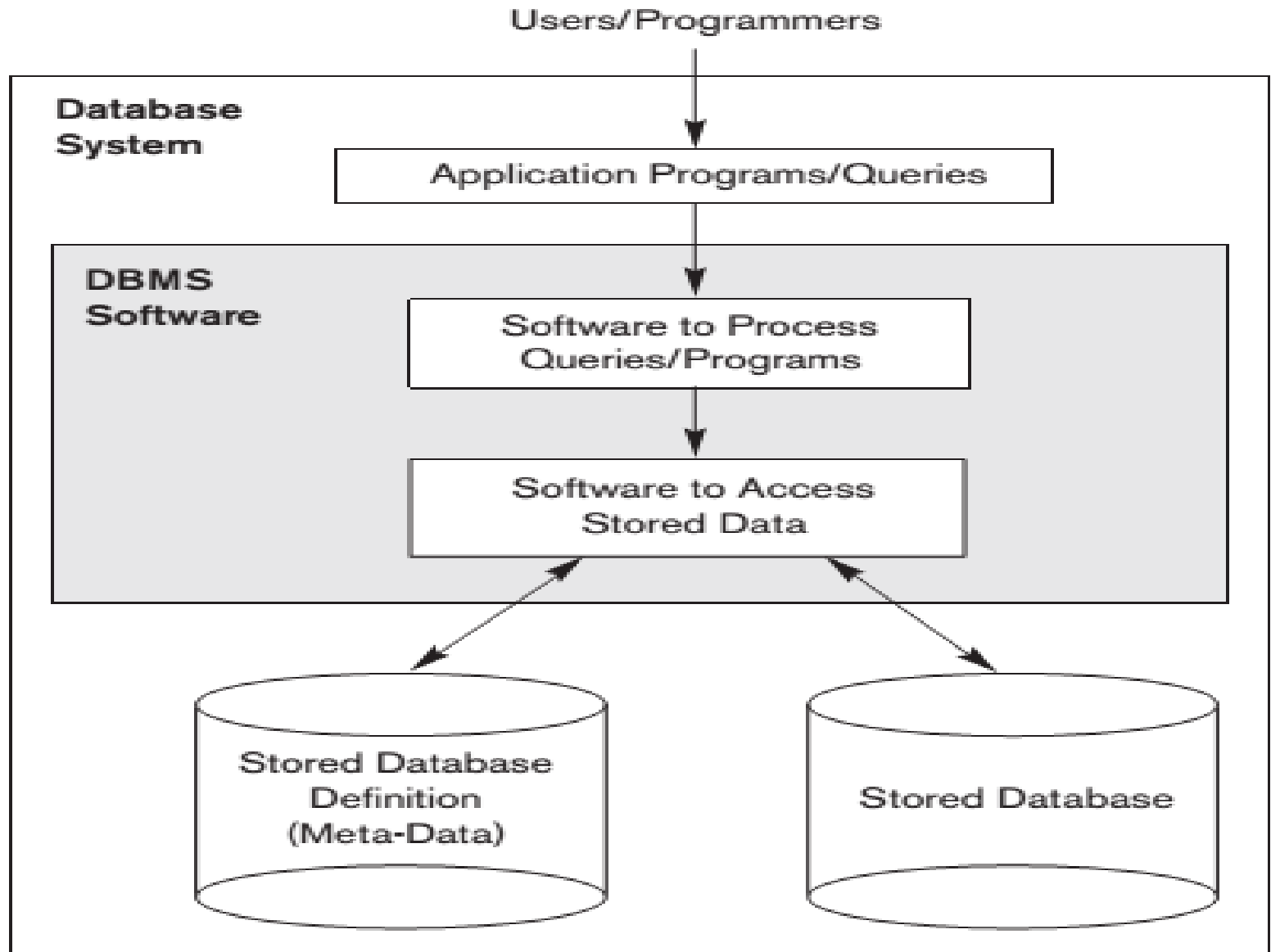
- The DBMS is a general-purpose software system that facilitates the processes of **defining, constructing, manipulating, and sharing databases** among various users and applications.
- **Defining:**
  - Defining a database involves specifying the data types, structures, and constraints of the data to be stored in the database.
  - The database definition or descriptive information is also stored by the DBMS in the form of a database catalog or dictionary; it is called meta-data.
- **Constructing**

Constructing the database is the process of storing the data on some storage medium that is controlled by the DBMS.
- **Manipulating:**

Manipulating a database includes functions such as querying the database to retrieve specific data, updating the database to reflect changes in the mini-world.

# Database Systems...contd.

## Database System Environment





# Database Systems...contd.

## 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

## 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

# Student Database

# Database Systems...contd.

- In the early days, database applications were built directly on top of file systems
- **Drawbacks of using file systems to store data**
  - **Data redundancy and inconsistency**
    - Multiple file formats, duplication of information in different files
  - **Difficulty in accessing data**
    - Need to write a new program to carry out each new task
  - **Data isolation** - Multiple files and formats
  - **Integrity problems**
    - Integrity constraints (e.g. account balance  $> 0$ ) become “buried” in program code rather than being stated explicitly
    - Hard to add new constraints or change existing ones

# Database Systems...contd.

- **Atomicity of updates**

- Failures may leave database in an inconsistent state with partial updates carried out

- **Example:**

Transfer of funds from one account to another should either complete or not happen at all

- **Concurrent access by multiple users**

- Concurrent access needed for performance
- Uncontrolled concurrent accesses can lead to inconsistencies

- **Example:**

Two people reading a balance and updating it at the same time

- **Security problems**

- Hard to provide user access to some, but not all, data
- Database systems offer solutions to all the above problems

### **Characteristics of the Database Approach**

- **Self-describing nature of a database system**

Database system contains not only the database itself but also a complete definition or description of the database structure and constraints (Meta Data).

- **Insulation between programs and data, and data abstraction**

The structure of data files is stored in the DBMS catalog separately from the access programs.

- **Support of multiple views of the data**

- A database typically has many users, each of whom may 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.

- **Sharing of data and multiuser transaction processing**

- Data for multiple applications (online transaction processing (OLTP)) is to be integrated and maintained in a single database.
- The DBMS must include concurrency control software to ensure that several users trying to update the same data do so in a controlled manner so that the result of the updates is correct.