DBS211
Introduction to Database Systems

WEEK- 1

Agenda

- **Data vs. Information**
- File Systems
- Database
- Database Management System
- Database Design

Data vs. Information

- Data:
  - Unprocessed information
  - i.e. customer, course, employee
  - Building blocks of information
- Information:
  - Information reveals the meaning of data
  - Produced by processing data which is related
  - i.e. credit card statement, phone bill, video rental
- Accurate and timely information is need for decision making
- Good decision making is key to an organization's survival

File Systems

- Keep information stored on a computer is by saving various files on your computer

- Programs were written to access data.
  - i.e. to get information about a customer, a program would be written to do just that (and nothing else).
  - Program would go against a specific file (or files, depending on it's purpose)

- If other information was needed, another program was written

- Programmer had to specify exactly how the processing of the data needed to be done

File Systems

MASTER FILE DIRECTORY

USER FILE DIRECTORY
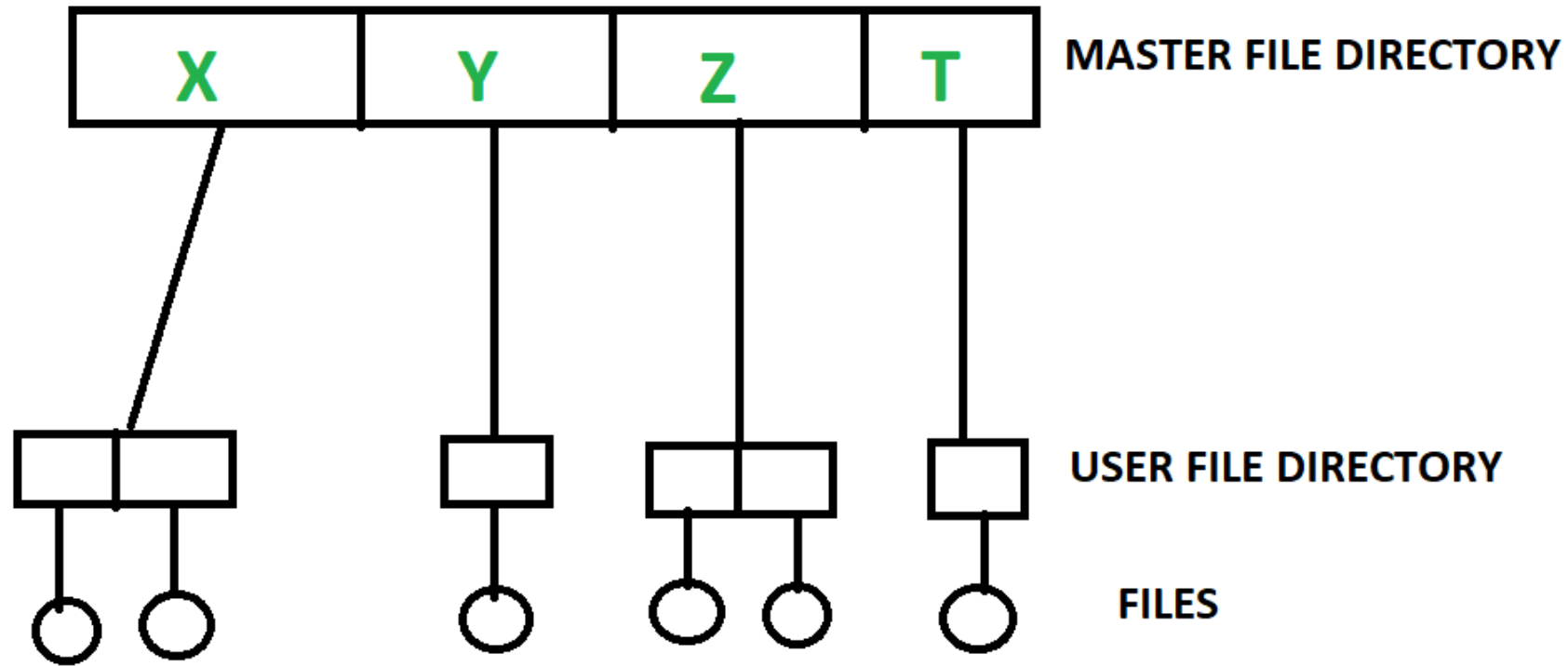
FILES

File Systems (Example)

- Simple business example:
  - A customer *orders* products from a Retailer
  - Sales Department would *take the order* information from the customer
  - Sales Department would *send a request* to the Shipping Department to deliver a product to a customer
  - The Shipping Department would *enter in the information* about the customer order and create a shipping document
  - The Shipping Department *would print the shipping* document and send the products to the Customer

File Systems (Cont'd)

- Possible required files to store data:

  - Files required in Sales Department:
    - *Customer, Product, Sale*

  - Files required Shipping Department:
    - *Customer, Shipping Document*

Can anybody see a problem with this?

Where do you go to look u p a customer phone number? (both sales and shipping have customer data)

Where do you get the product code for a specific product? (Sales, Shipping and Warehousing all have product data)

Data Redundancy

- Same data appears in more than one location
  - Customer data exists in two different files:

  - **Sales Department:**
    - ***Customer***, *Product*, *Sale*

  - **Shipping Department:**
    - ***Customer***, *Shipping Document*

- Results in data inconsistency
- Data anomalies result when changes to redundant data are not made successfully.

Example:

Courses Tables

Course Code	Course Name	Student ID	Student Name
dbs211	intro to dbase systems	900111111	John Smith
dbs211	intro to dbase systems	900222222	**Allyson Jones**
web222	web programming fundamentals	900111111	John Smith

Students Table

Student ID	Student Name	Date of Birth	Phone
900111111	**John Smith**	May 16, 2002	9055551111
900222222	**Allyson Jones**	Sept 12, 2001	4165552222
900333333	**Raj Patel**	Feb 28, 1999	2865553333

File Systems

- What if you wanted some information from a file quickly?

- Could not be done using a traditional file system.

- Why? Programs specified exact structure of a file. Think of the complexity if information was required from multiples files!

File Systems

- What if you wanted to change the structure of a file?

- All the programs that used that file would have to be changed.

Data Anomalies

- Modification anomalies
- Insertion anomalies
- Deletion anomalies

Modification Anomaly

Employees' Skills

Employee ID	Employee Address	Skill
426	87 Sycamore Grove	Typing
426	87 Sycamore Grove	Shorthand
519	94 Chestnut Street	Public Speaking
519	96 Walnut Avenue	Carpentry

- A modification anomaly occurs when changing one records data leaves another records data in error.
- Change the address for employee 519 – has to be done for all rows with employee ID 519.
- What if this employee address is in another table (file) – has to be done for all tables (files)

Deletion Anomaly

Faculty and Their Courses

Faculty ID	Faculty Name	Faculty Hire Date	Course Code
389	Dr. Giddens	10-Feb-1985	ENG-206
407	Dr. Saperstein	19-Apr-1999	CMP-101
407	Dr. Saperstein	19-Apr-1999	CMP-201

DELETE

▶ Dr. Giddens is not teaching ENG-206 this semester

▶ We could lose the hire date if we remove this row to reflect current faculty and their courses

Insertion Anomaly

Faculty and Their Courses

Faculty ID	Faculty Name	Faculty Hire Date	Course Code
389	Dr. Giddens	10-Feb-1985	ENG-206
407	Dr. Saperstein	19-Apr-1999	CMP-101
407	Dr. Saperstein	19-Apr-1999	CMP-201
424	Dr. Newsome	29-Mar-2007	**?**

- We can record the details of any faculty member who teaches at least one course in 2007
- We cannot record the details of a newly hired faculty member who has not yet been assigned to teach any courses

Agenda

▶ Data vs. Information

▶ File Systems

▶ **Database**

▶ Database Management System

▶ Database Design

Database

- **Database**
  - A structure that contains logically related data in a *single repository*
  - centralized and unique
- A Database contains:
  - End user component – data
 - Metadata – data about data

Types of Databases

- Centralized
  - Supports one or more users at a time
  - Database is in one physical location
  - Database could be on a pc, a mid-range, or a main-frame
- Distributed:
  - Data is distributed at several physical locations
  - Database at each physical locations can vary (i.e. one location might have the database on a mid-range, one might have it on a pc)

Database vs. File System
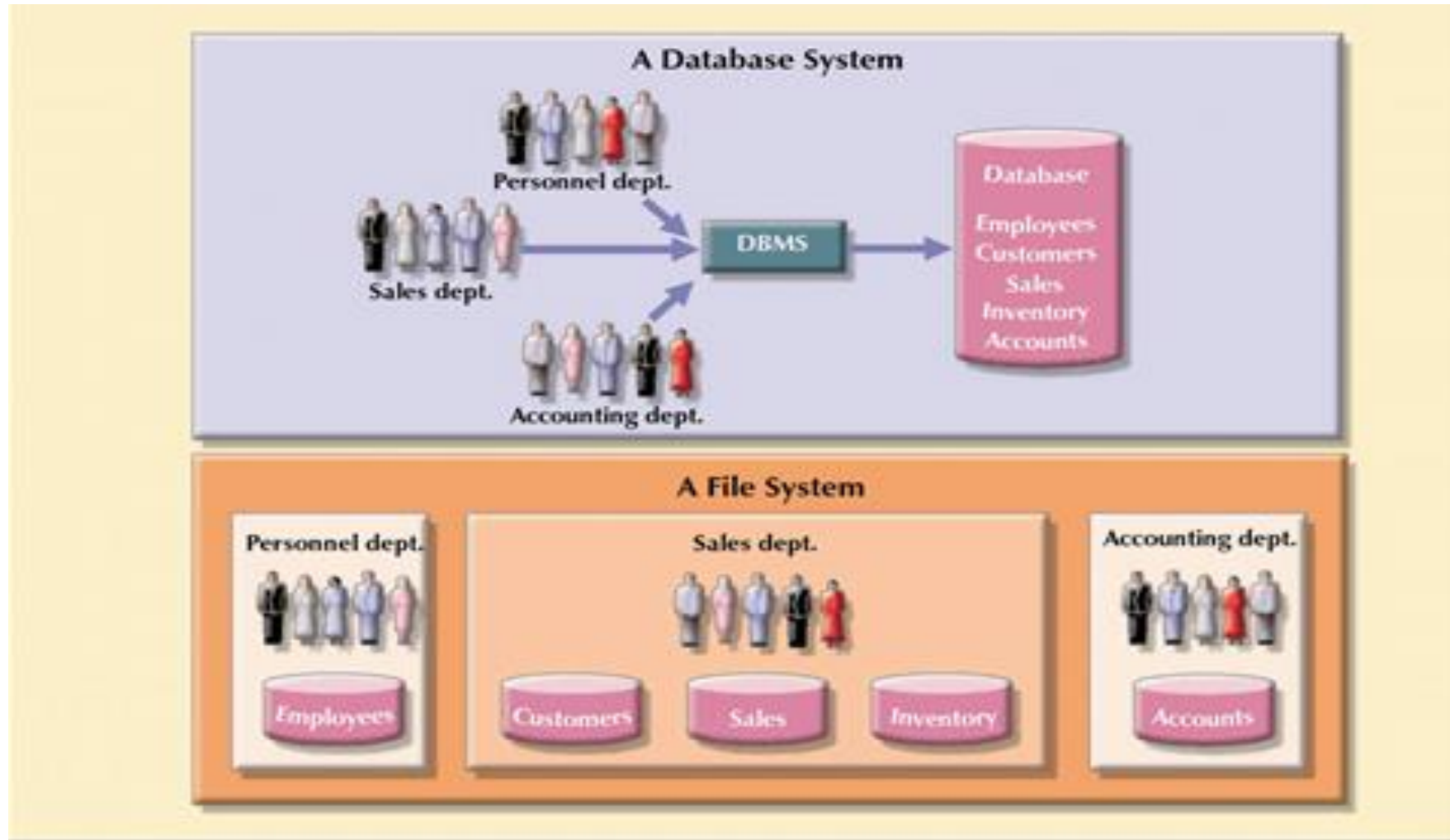
Database Uses

- ***Production or Transaction***
  - Supports day-to-day operations

- ***Decision Support***
  - Information for tactical or strategic decision making

- ***Data Warehouse***
  - Historical data

Advantages of Database Processing

- Ability to get more information from the same amount of data
  - i.e. all customer related data is in one customer place
- Sharing data
  - Data is available to be use by authorized users
- Controlling Redundancy
  - Only a single copy of the data exists

More Advantages

- Balancing
  - Databases are structured to benefit all users in the organization, not just a single group
- Expanding security
  - Access to data can be password protected or can restrict access to data – read only, update

Some examples of popular
Relational DBMSs are:

Advantages Continued

- Increasing Productivity
  - Ability to write ad hoc queries
  - Users don't have to know exact structure of the data
- Provides data independence
  - Can change the structure of the data without having to change the programs that access it

Disadvantages of Database Processing

- Large file size
  - Files now include metadata information

- Increased complexity
  - Need to take into consideration data design, security, backup and recovery, integrity

More Disadvantages

- Greater impact of failure
  - Data is now shared, if a failure occurs, many users are impacted
- More difficult recovery
  - If data can be updated by many users at the same time, how can it be recovered to the correct state for all users correctly

Agenda

- Data vs. Information
- File Systems
- Database
- **Database Management System**
- Database Design

Database Management System (DBMS)

▶ Collection of programs that manages database structure and controls access to the database (and ultimately the data)

▶ Manages sharing of data among multiple applications and users

▶ Data is more consistent

▶ Ability to do ad hoc querying

Database Management System (DBMS)

FIGURE 1.2 THE DBMS MANAGES THE INTERACTION BETWEEN THE END USER AND THE DATABASE
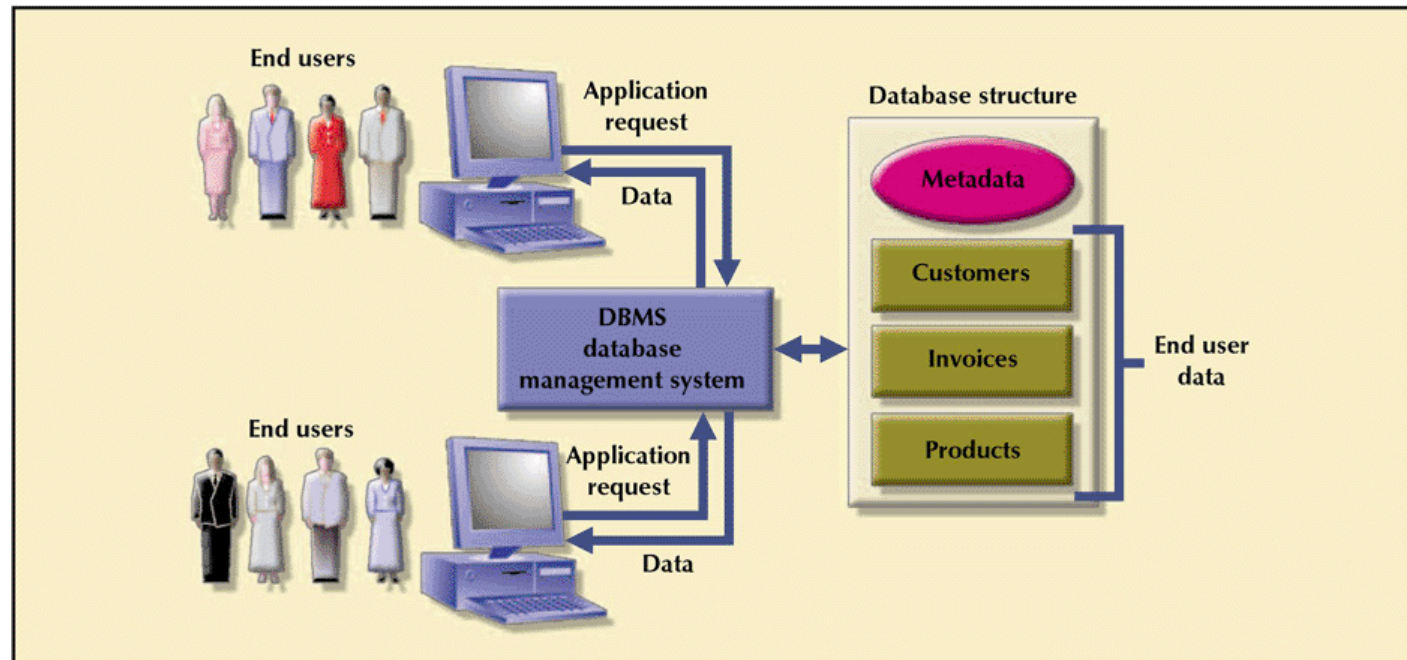
Importance of a DBMS

- For Top Management

  - Provides information necessary for strategic decision making and planning

  - Provides access to external and internal data

  - Provides information on company performance and whether the company is achieving their goals (targets) or not

- For Middle Management

  - Provides information necessary for tactical decision making planning

Importance of a DBMS

- For Operational Management
  - Provides timely information
  - Represents and supports the company operations as closely as possible (operational data)
- For Other Users
  - Provides timely information
  - Produces results within specified performance levels

Agenda

▶ Data vs. Information

▶ File Systems

▶ Database

▶ Database Management System

▶ **Database Design**

Database Design

▶ Defines the database's expected use

▶ Avoids redundant data (unnecessarily duplicated)

▶ Eliminates poorly designed databases

▶ Done within a systems development life cycle (SDLC) framework

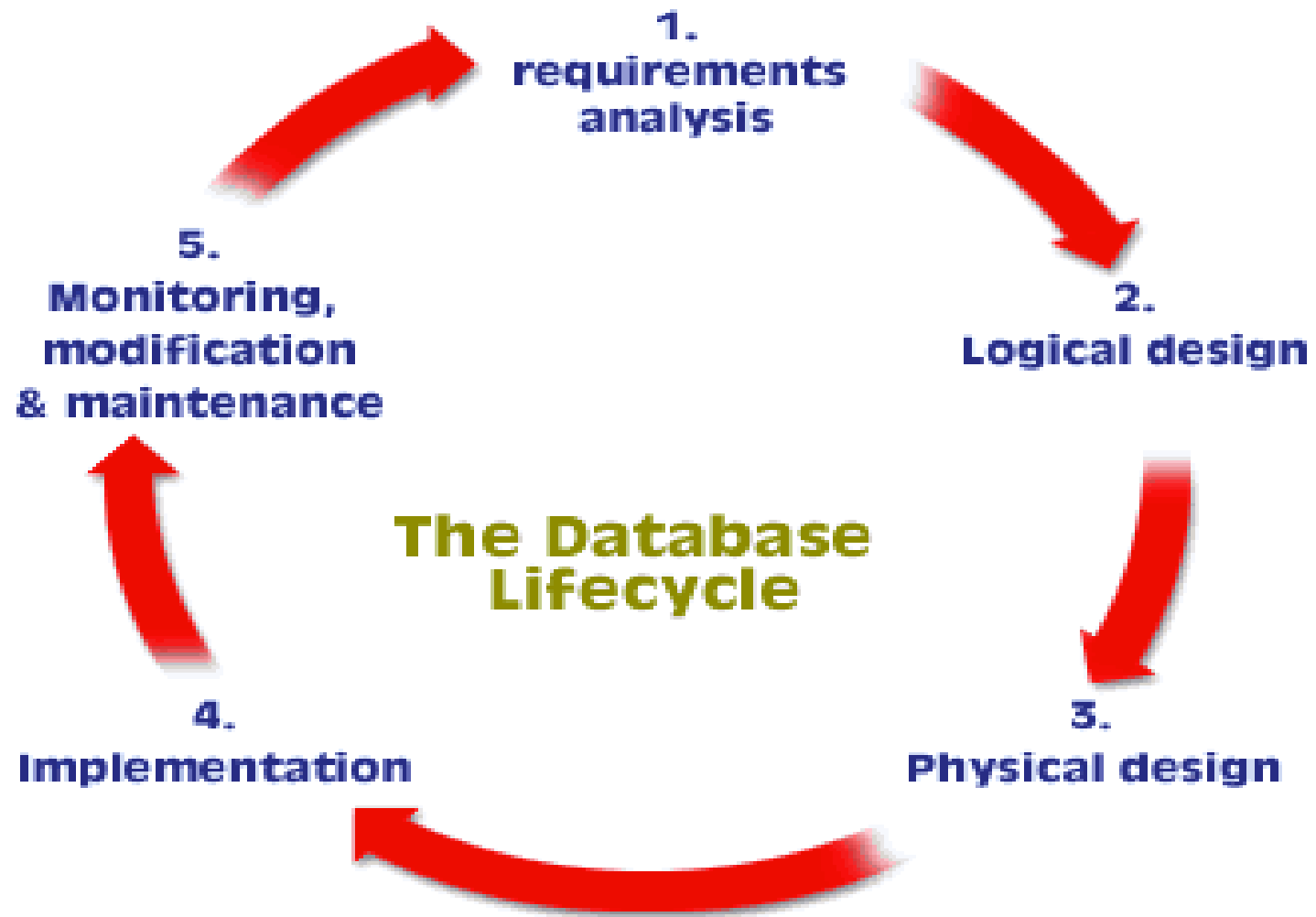▶ Database Design has it's own framework, within the SDLC called the Database Life Cycle (DBLC)

33

Database Design

- Different database design strategies exist: top-down, bottom-up, centralized and decentralized

- The database topic will be covered in database design and modeling lecture.

Database Development Life Cycle
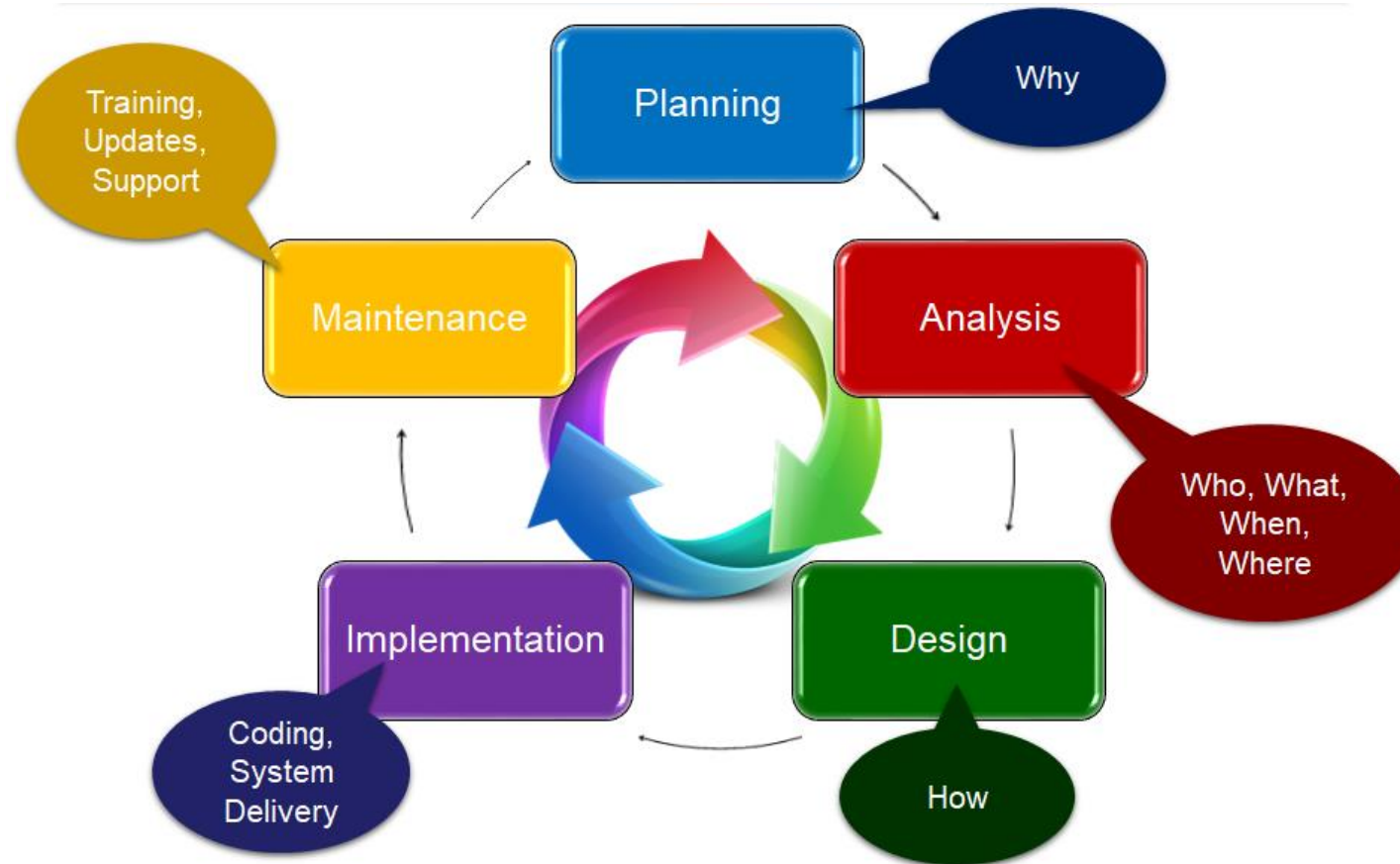
Planning	We have an idea! Why do we need this software, how will it help the organization? Who will champion this project? What are the goals or objectives of the software?
Analysis	Needs analysis: What does the software need to be successful? Should we build it? Do we have the right team to build it? Can we financially afford to built it? Do we have the right technologies available?
Design	Specifications, architecture, choice of languages, frameworks, diagrams, pseudo-code, scoping.
Implementation	Coding the software, creating the graphics, testing the software, actually creating the product, installation and delivery.
Maintenance and Support	Maintenance plan, on-going monitoring and review, what are the next steps, new versions? Training and on-going customer support

The Database Lifecycle

1. requirements analysis
2. Logical design
3. Physical design
4. Implementation
5. Monitoring, modification & maintenance

Requirements Analysis	What does the database need? What information do we need later? How is the database going to be used? What other applications will the database support?
Logical Design	Design Specifications, Data Modeling, Entity Relationship Diagrams, Data Types, Normalization,
Physical Design	Actual process of creating tables and database structure (SQL, DDL)
Implementation	Inserting or importing data, installation on live server, backup systems, and more
Monitoring, modification, and Maintenance	On-going backups, data archival, monitoring, additional queries as required

Next Class

- **Relational Database Design**
  - Relational Databases
  - Relationships
  - Referential Integrity
  - Table Types
  - Various Keys