

ADM 1370- Information Systems

Purpose: To provide accurate, timely and useful information

- Each element must be present and all of the elements must work together.

An information system consists of five parts including

- People
- Procedures
- Software
- Hardware, and
- Data

Information system (IS) = IT plus procedures, and people that produce and utilize information.

IT = Hardware +Software + Data

- Products
- Methods
- Inventions
- Standards

Avoid a common mistake

Do not try to buy an IS; you cannot do it

You can buy IT

Buy or lease hardware license programs and databases, even obtain pre designed procedures.

Ultimately, people execute those procedures to employ that new IT.

Any new system requires training tasks, overcoming employee resistance change and managing employees as they utilize new systems.

Information systems aid the Decision - Making process:

- Improve productivity
- Monitor organizational performance
- Planning and Decision - Making
- Enhance Competitive Advantage

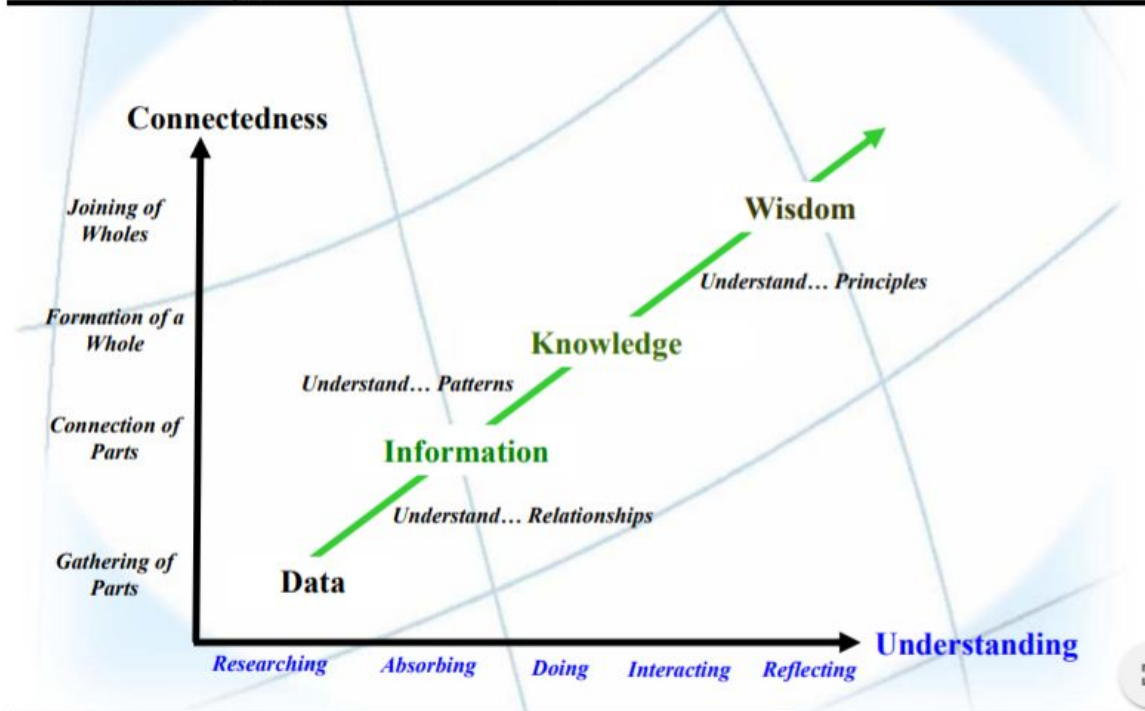
Information System:

- Set of interrelated components
- Collect, process,store and distribute information
- Support decision-making, coordination and control

Data versus information

Data are streams of raw facts

Information is data shaped into meaningful form



Data -----> Information-----> Knowledge

Three core activities of information systems:

Inputs: Captures raw data from organization or external environment

Processing : Converts raw data into meaningful form

Output: Transfers processes information to people or activities that use it

Advanced information systems also incorporate one additional functionality:

- Feedback

Output returned to appropriate members of organization to help evaluate or correct input stage

It is important to remember that these functionalities are not limited to technologies:
Can you think of these activities in a management context ?

Management Information Systems and Analytics

MIS: Study of people, Technology and Organizations

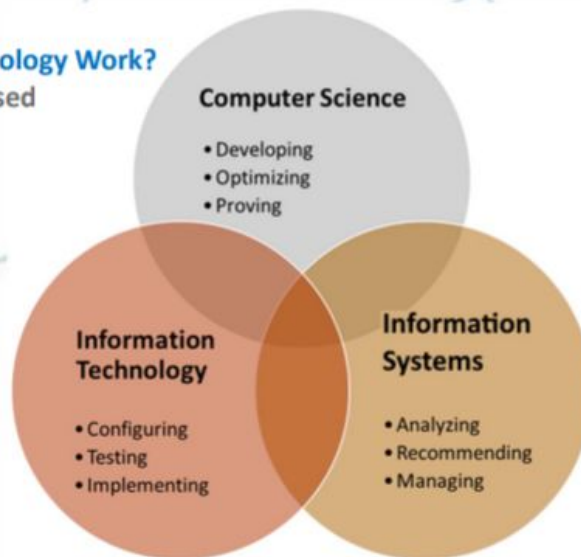
Analytics: Tools and Techniques for Data- Driven Decision Making

As an IS manager for your business function, you will play a vital role in the implementation and administration of technology within your divisions and for the benefit of your organization

- You will plan, coordinate and direct research on the computer related activities of firms
- You will consult with other managers, help determine the goal of an organization and then implement technology to meet those goals
- You will coordinate with pertinent people about the technical aspects such as software development, network security and internet operations.

Why does the Technology Work?

- Technology-Focused
- Theory Oriented



Computer Science

- Developing
- Optimizing
- Proving

Information Technology

- Configuring
- Testing
- Implementing

Information Systems

- Analyzing
- Recommending
- Managing

What Technology Would Work Best?

- Business-Focused
- Application Oriented
- Cross-Functional

How Does the Technology Work?

- Deployment-Focused
- Use Oriented

Social Media

Social media is a group of internet based applications that build on the ideological and technological foundations of web 2.0, and that allow the creation and exchange of User generated content.

Web 2.0

- Software as a continually updated service that gets better the more people use it consuming and remixing data from multiple sources, including individual users while providing their own data and services in a form that allows remixing by others creating network effects through an architecture of participation.

User - Generated Content

- Various forms of media content that are publicly available and created by end- users.

3 Basic Requirements

- **Needs to be published either on a publicly accessible website or on a social networking site accessible to a selected group of people**
- **Needs to show a certain amount of creative effort**
- **Needs to have been created outside of professional routines and practices.**

Characteristics of Web 2.0:

- The ability to tap into the collective intelligence of users
- Data is made available in new or never- intended ways
- Relies on user- generated and user- controlled content and data
- Lightweight programming techniques and tools let nearly anyone act as a Web site developer (End - User Development)
- The virtual elimination of software- upgrade cycles makes everything a perpetual beta or work in progress. Applications can be designed quickly to meet changing needs.

Which of these expressions best describes Web 2.0

Ans. Read- Write Web

Read Write Execute Web is Web 3.0

Eg. Cloud Computing

Types of Social Media

Web 2.0 is the whole set, while social media is a subset

- Relationship Networks
- Media Sharing
- Online Reviews
- Discussion Forums (One of the oldest forms of SM)
- Social Publishing
- Social Bookmarking (Often Niche)
- Interest- Based Networks
- Social Commerce (Platforms for Buying and Selling)

Social media has led to the convergence of information and communication:

- More channels for communication as well as media consumption
- Lines between personal communication and media consumption becoming blurry
- Social media is being used for many other purposes than were originally intended or envisioned.

Groundswell- Spontaneous movement of people using online tools to connect, take charge of their own experience and get what they need- information, support, ideas, products, and bargaining power - from each other.”

Eg. Crowdfunding, The #METOO movement

For something to become groundswell, it starts off as a viral item that gains momentum.

Social media has led to the Convergence of information and communication.

Adopting a Social business aptitude:

- Maximize the potential of social media by using it across business functions
- Rather than using social media as an inside out promotional medium, leverage the outside- in conversations as well.
- Harness big social data to glean insights for product research, competitive analysis and prediction of customer needs.
- Update corporate governance policies and guidelines.

Structured and Unstructured Data:

Structured data refers to information with a high degree of organization, such that inclusion in a relational database is seamless and readily searchable by simple, straightforward search engine algorithms or their search operations

Unstructured data is essentially the opposite

The lack of structure makes compilation and analysis a time and energy-consuming task.

- Traditional Approach

Structured and Repeatable Analysis

- Business Users: Determine what questions to ask
- IT: Structures the data to answer that question

Monthly sales reports

Profitability analysis

Customer Surveys

- Big Data Approach

Iterative and Exploratory Analysis

IT delivers a platform to enable creative discovery

Business Users Explore what questions could be asked

Brand Sentiment

Product Strategy

Maximum asset utilization

The 4 Challenges (V's) to data analysis

Volume: An overwhelming amount of data

Variety: Different Variety of Information in the form of different media such a video, audio and images

Velocity: Information coming a very high speeds

Veracity: The trustworthiness of user generated content

Social Network Analysis Tools (SNA's)

- The mapping and measuring of relationships and flows between people, groups, organizations, computers or other information or knowledge processing entities.
- The nodes in the network are the people and groups, whereas the links show relationships or flows between the nodes.

Nodes are the people in a social network, edges or vertices are the relationships between them.

- SNA'S provide both visual and mathematical analysis of relationships.

An integrated Social Media Strategy (ISMs)

Plan

Do

Check

Act

PDCA

Social Media Strategic Metrics:

Customer Dialog Metrics

- Genuine Dialog provides the basis for building awareness and word of mouth

Share of Voice: **Brand Mentions/Total Mentions X (Brand + Competitor A,B,C)= Share of Voice**

Audience Engagement: **Comments + Shares + Trackbacks/Total Views = Audience Engagement**

Conversation Reach: **Total People Participating/ Total Audience Exposure = Conversation Reach**

- Conversation Reach might double count people

-Needs to be between 6-10% to be considered a good campaign)

Web 1.0 vs Web 2.0

Web 1.0 vs. Web 2.0	
Web 1.0	Web 2.0—the Social Web
Static pages	Dynamic pages
Author controlled content	User controlled content
Computers	Computers, cell phones, televisions, PDAs, game systems, car dashboards
Users view content	Users create content
Individual users	User communities
Marketing goal: <i>influence</i>	Marketing goal: <i>relationships</i>
Data: single source	Data: multiple sources, e.g., mashups

Salient Features of Web 2.0

Information Sharing, Collaboration and Interactive Functionality of the web:

- Inter Connectivity and Interactivity of web-delivered content

Architecture based on Participation

Popular Examples of Web 2.0 based applications:

Social Networking Sites, Video Sharing Sites, wikis, blogs and Folksonomies (an index of information based on multiple people coming together eg. hashtags)

Mashups and APIs

Table 1. Comparison of Web 2.0 and related legacy technologies.

Web 2.0 service	Related legacy service	Revolutionary?
Blogs	Bulletin board systems (BBS) and threaded news groups	Not particularly; however, greater ownership and easier use are possible.
Image sharing (such as Flickr)	Image sharing Web sites	No, but many more images are available now and it's arguably easier to search for them.
Wikis	Personal Web sites	Somewhat. Wikis may be useful in work groups and other moderated environments
Really simple syndication (RSS)	None	Yes, due to RSS' ability to deliver granular news on demand.
Social networks	Personal Web sites	Not particularly. Admittedly, contemporary social networks are more user friendly.
Mash-ups	None	Yes, because of their ability to combine content to form new content.
Podcasts and vodcasts	File servers with Web-exposed content	Somewhat, because of their ability to subscribe to chosen granular content.
Folksonomies	Legacy search engines (such as Webcrawler)	Yes, because of the power to find new information based on other users' searches.

Rich Internet Applications are a predominant feature of web 2.0

Attributes:

- Web applications that have many of the characteristics of desktop software:
- Eg. powerful interfaces that provide the responsiveness of traditional desktop applications.

Typically delivered either through specialized browsers, browser plug-ins or client virtual machines

Uses:

- Rendering multiple forms of content (text, audio, video etc.) in an integrated fashion.
- Engaging users via interactive user-friendly interfaces
- Performing complex data visualization, including dynamic charting or graphical presentation of data.

Example of an Enabling Technology:

AJAX (Asynchronous JavaScript & XML)

Enterprise 2.0 refers to Web 2.0 technologies used form some business purpose:

- Promote collaboration and knowledge exchange among employees, consultants and company partners.

A wiki is a website or similar online resources which allows users to add and edit content collectively.

- Collectively share and edit a body of knowledge
- Ongoing process of creation and collaboration
- Knowledge Management- e.g Wikipedia, Enterprise wikis

They are a dynamic collectively authored set of web pages

Characteristics of a Wiki application:

- Web based
- Interactive
- Collaborative
- Iterative

Wikis have been considered to be an enabling technology for knowledge contributions, storage and exchange.

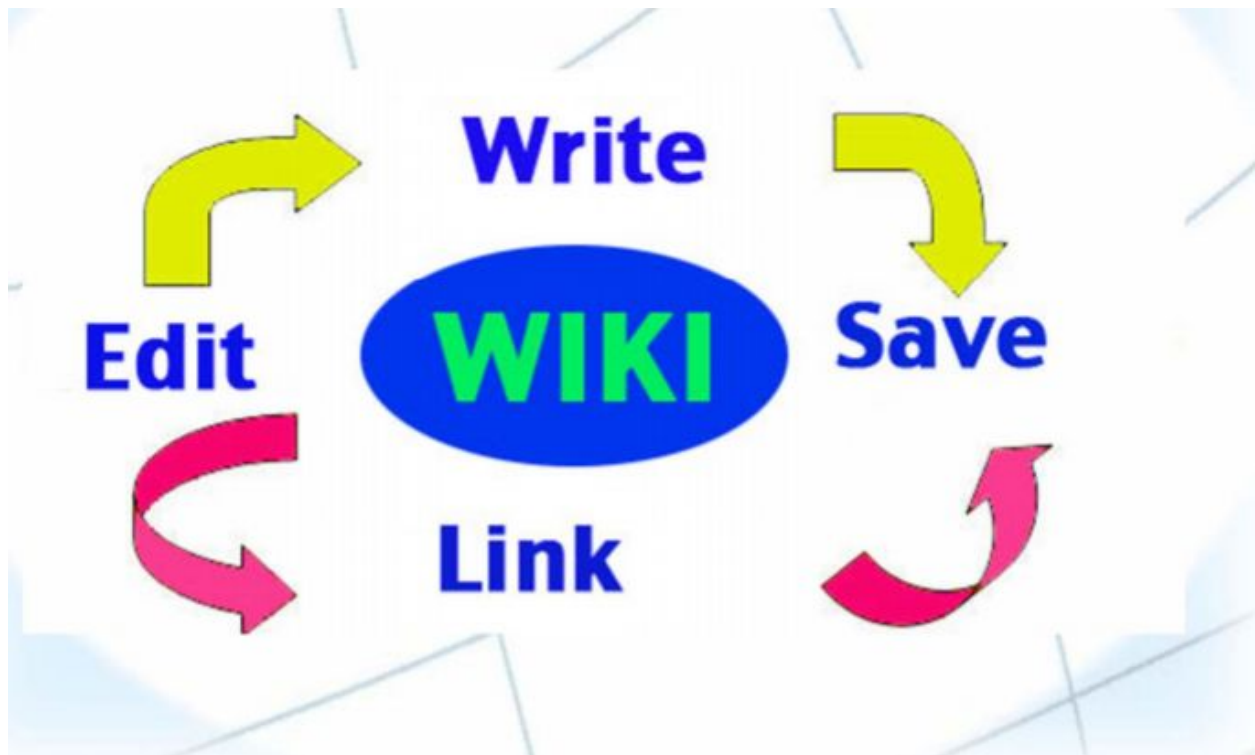
Wikis are collaboration tools

- Blogs and chats are more turned towards conversation

Wikis are intended to maintain a series of unique documents and their content evolves

Wikis have built in version control

- No changes can be made without creating a record of who made those changes
- Reversion to an earlier version is always possible



RSS (Rich Site Summary) (Really Simple Syndication)

- A standardized data format to publish frequently updated works such as blog entries and news headlines.
- Website material is made available to end users or other sites through web feeds.

A web feed is data format used for providing users with frequently updated content.

Content distributors syndicate a web feed, thereby allowing users to subscribe to it.

RSS FEEDS are usually accessed through an aggregator tool:

- Making a collection of web feeds accessible in one spot is known as aggregation, which is performed by an aggregator.
- Aggregators can be scheduled to check for new content periodically. Web feeds are an example of pull technology, although they may appear to push content to the user.

Widgets

- Mini web applications or Web add-ons
- When embedded in other website, they allow those websites to access content or functionality from other widget provider website
- Snippets of codes from third-party sites that are included on your own website or web application

Examples:

- Weather Information Widget
- Calculator widget
- Currency converter widget

Collaborative Consumption: Sharing Economy, you could be a buyer or seller at different points of time.

Collective Intelligence: Allowing people to come together and make decisions off of what these people are telling you eg. Twitter Poll, Reddit

Collaborative Intelligence: Wikipedia or Wikis in general, people collaborate together and create something new

Mashups

The word "Mashup" describes when data from two or more sources are combined into an integrated application. Applications or interfaces generated by combining content, presentation or application functionality from disparate sources.

Content used in mashups is typically sourced from a third party via a public interface or API (application programming interface)

API (Application Programming Interface)

- An abstraction that defines and describes an interface for interaction with a set of functions used by components used of a software system.
- Abstraction is the process of taking away or hiding or removing characteristics from an object in order to reduce it to a set of essential characteristics.
- For the consumer or user, abstraction helps focus on the essential elements with unwanted detail omitted.

Mashups are applications generated by combining content, presentation or application functionality from disparate sources.

Web mashups are usually composite web applications partially constructed from the service and content from other web sites.

Mashups have an emphasis on GUI and programming less specification.

- The concept of mashups originated from the understanding that the number of applications available on the web and the needs to combine them to meet user requirements are growing very rapidly.

YOU DON'T NEED TO BE A DEVELOPER TO CREATE A MASHUP, relies heavily on End-User Development

Examples of Mashups: Padmapper.com Portwiture.com

Some mashups may not be characterized as Composite Applications:

- Mashups may also be used to access a single resource to mine data or migrate content (repurpose existing data and information according to your needs)
- Creating mashups is all about finding data, functionality, and services and using them to both solve problems and create opportunities.

Mashups have been regarded as a Disruptive Technology:

- Mashups will be a major force in the next few years...

The disruptiveness come in the end-user environment, where business users with little technical ability will be able to create their own mashups and assemble them in dashboards. This will introduce security and privacy challenges for the IT industry.”

- Gartner Research

Mashups are about simplicity, usability, and ease of access. This simplicity has the upper hand over feature completeness or full extensibility.

The tools for constructing mashups have begun to reach a level of usability where even non-technical users can build their own solutions.

Mashups are sometimes erroneously referred to as a Web API

What is an API ?

An abstraction (freedom from representational qualities) that defines and describes an interface for interaction with a set of functions used by components of a software system. An API may include details about the calling conventions , protocols and format of data transfers, but hide details about internal algorithms and program computations.

Example:

Google Maps

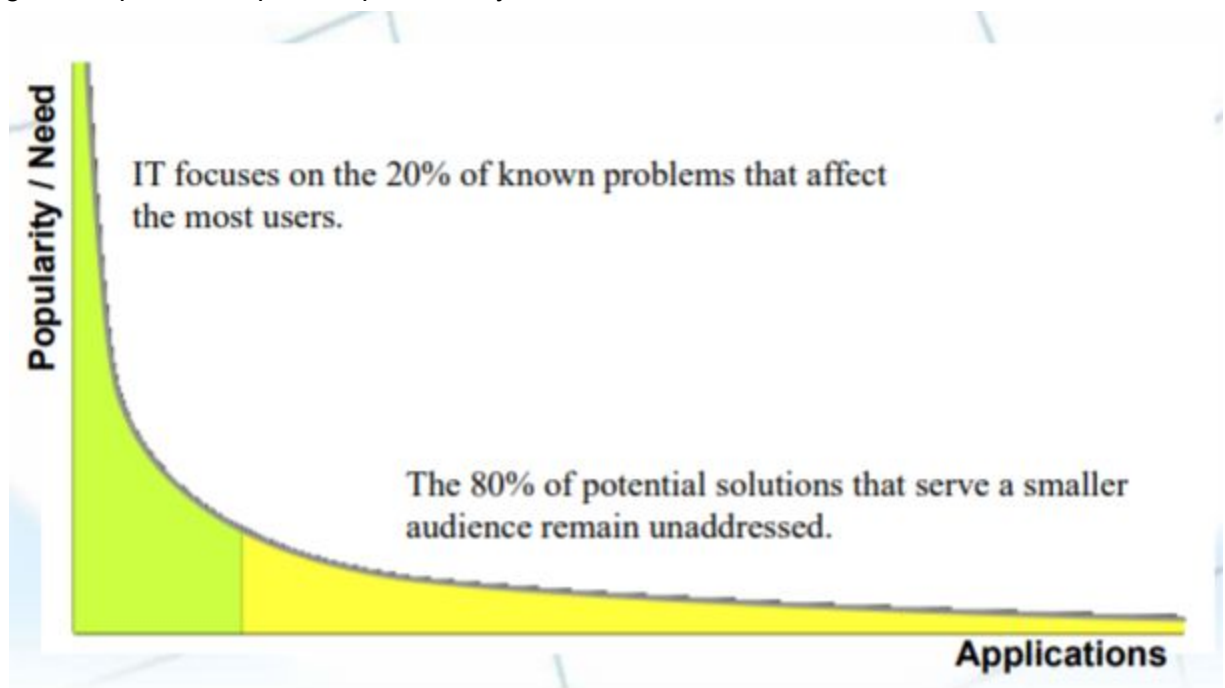
How to Call an API:

Typically by using a URL with all the necessary information supplied within the URL

- The bits of information being supplied in the url as inputs to the API are referred to as "Parameters"
- The parameters typically follow a "?" and are noted in the form of name = value

Traditionally, application development dollars are directed towards those projects and enhancements demanded by the largest group of users.

- This practice of catering to the masses doesn't necessarily lead to an outcome with the greatest positive impact on productivity.



Armed with powerful new tools that leverage existing data application resources, developers and power users can quickly assemble products to target the long tail.

Mashups are often used as a “Proof of Concept” .

Module 2

Microsoft Excel & Spreadsheet Applications

- A computer program used to analyze, and present (quantitative) data
- A spreadsheet program.

A spreadsheet is a collection of text and numbers laid out in a rectangular grid.

- Often used in business for accounting, budgeting, financial analysis, inventory management and other functions.

What can Excel be used for?

- Business Use Context
- Personal Use Context

What other alternatives exist ?

- Other Vendors

An electronic spreadsheet program such as Excel aids a multitude of problem-solving and decision-making processes through providing:

- Data management features
- Automatic calculation functions
- Presentation tools
- Decision analysis functions

Excel:

- A computerized spreadsheet application used to build and manipulate worksheets and workbooks.

Worksheet:

- A spreadsheet that may contain data including text, numbers, formulas, charts etc.
- Sometimes a charts based worksheet is referred to as a “Chartsheet”

Workbook:

A collection of related worksheets within one file

Planning for Good Workbook and Worksheet Design

- Decide on the purpose of the spreadsheet and how it will be constructed.
- Make it obvious where data is to be entered:

Use titles, headings, instructions, color schemes, and forms to designate areas for data-entry.

Wherever possible, setup formulas and use cell references for calculations.

Allow excel to do what it was designed for, i.e automatic calculations

- Try to minimize hard-coded numbers in your calculations.

Test multiple times to make sure the results are what you expect:

- Check your results against your mental model of what the results should be.

Format the worksheet so it looks

Use referenced values from cells instead of literal values.

Data Delivery Functions in Excel

- These function can be used

Logical Test: Anything that uses a comparison operator to compare two or more sets of values.

The difference between a logical function and a conditional function is that a logical function only produces true or false values. Conditional functions can return other values after it tests for certain criteria.

Summarizing Data through Subtotals & Grouping

Subtotals:

- Excel provides a number of features that enable you to organize large groups of data into more manageable groups.
- Data in a list can be summarized by inserting a subtotal.
- Before you can subtotal, however, you must first sort the list by the field on which you want the list subtotaled.

Grouping:

- If you have a list of data that you want to group and summarize, you create an outline.
- Grouping refers to organizing data so that it can be viewed as a collapsible and expandable outline.
- To group data, each column must have a label in the first row and the column must contain similar facts. The data must be sorted by the column of columns for that group.

Modelling Approach to Problem-Solving

- Models are usually simplified versions of the things they represent:

A valid model accurately represents the relevant characteristics of the object or decision being studied

Types of Models:

- Mental

- Visual
- Physical
- Mathematical

A computer model:

- A set of mathematical relationships and logical assumptions implemented in a computer as an abstract representation of a real-world object or phenomenon.
- Spreadsheets provide the most convenient way for business people to build computer models.

Benefits of the modelling approach:

- Economy: It is often less costly to analyze decision problems using models.
- Timeliness: Models often deliver needed information more quickly than their real-world counterparts
- Feasibility: Models can be used to do things that would be impossible.

Models give us insight and understanding that improves decision making.

Most spreadsheet models are very similar to the generic mathematical model:

$$= f(X1, X2, \dots, Xn)$$

Most spreadsheets have input cells (representing $X1$) to which mathematical functions ($f(.)$) are applied to compute an output variable (or Y)

Decisions:

- Refers to possible choices, or courses of action, that we might take.

Outcomes:

- Refers to the consequences of the decisions
- The performance measures we use to evaluate the results of taking action

Structure:

- Refers to the logic and the mathematics that link the elements of the model together

Data:

- Refers to the specific numerical assumptions and values of data used to the model

Best Practices for Creating Spreadsheet Models:

Planning

- Starts at the requirements gathering phase
- Spend more time initially to avoid rework in later stages

Prototyping

- Gradually build and regularly test models
- Start simple and then address complex requirements

Debugging

Identify and fix errors on an ongoing basis during model development

Reviewing

Get someone else to review your models

What-if Analysis:

- The study of how the uncertainty in the output of a model can be attributed to different sources of uncertainty in the model inputs
- Using a spreadsheet, what if analysis involves changing the values in cells to see how those changes will affect the outcome of formulas on the worksheet
- What - if analysis tools in excel include:
 - Scenarios
 - Goal seek
 - Data tables

Solver:

An add in program that searches for the best solution to a problem with several variables
Used to find the best way to allocate resources. Once added, it appears in the Analysis group on the data tab.

- Used to find the best way to allocate resources

Requires three parameters:

- Target cell typically contains a formula that is directly or indirectly based on the adjustable cells and constraints:
 - Dependent Variables and Independent Variables need to be formulated appropriately
- Adjustable cells are the cells whose values are adjusted until the constraints are satisfied
- Constraints specify the restrictions:
 - Typically in the form of inequalities or equations

To perform what-if analysis:

- Change the value of a worksheet cell (the input cell)
- Observe its impact on the calculated cells (the result cells)

To perform goal seeking:

- In the Data tools group on the data tab, click the what-if analysis button, and then click goal seek.

- In the set cell box, select the result cell, and then, in the TO value box, specify its value (goal)
- In the By changing cell box, specify the input cell
- Click the OK button. The value of the input cell changes to set the value of the result cell.

Requirements of Linear Programming:

- LP problems seek to maximize or minimize some quantity expressed as an objective function
- The presence of restrictions, or constraints, limits the degree to which we can pursue our objective.
- There must be alternative courses of action to choose from
- The objective and constraints in linear programming problems must be expressed in terms of linear equations or inequalities.

Decision Analysis Using Solver:

- Define the problem
- Use identified input ranges, output cells and constraints, solver can minimize or maximize the input cell or set the output cell to a particular value.
- Define the parameters using the solver dialog box.

Module 3: Data Management

Data management Practices include all stages of the data life cycle, carry an enterprise-wide scope, span various types of data and information, and provide different levels of granularity.

Data Governance:

- An approach to managing information across an entire organization.
- A formal set of business processes and policies that are designed to ensure that data is handled in a certain, well defined fashion.
- The organization follows unambiguous rules for creating, collecting and handling, and protecting its information.
- Its objective is to make information available, transparent, and useful for the people who are authorized to access it, from the moment it enters an organization until it is outdated and deleted.

Example of Data Governance Practice:

- Master Data Management

Master Data Management:

Master Data: Consistent and uniform set of identifiers and extended attributes that describe the core entities of the enterprise such as customer, product, employee, vendor, geographic location

Transaction Data: Generated and captured by operational systems, describing the business' activities or transactions.

Advantages of Master Data Management:

- Improves search effectiveness
- Increases accuracy
- Streamlines new product entry into the database
- Facilitates processing of transactions

Data Quality Considerations

- Ensuring data quality:
Before new database in place, need to:
 - Identify and correct faulty data
 - Establish better routines for editing data once database in operation

Corporate initiatives to improve data quality:

- Data quality audit:
Structured survey of the accuracy and level of completeness of the data in an information system
 - Survey samples from data files, or
 - Survey end users for perceptions of quality
- Data cleansing:

Software to detect and correct data that are incorrect, incomplete, improperly formatted, or redundant.

Enforces consistency among different sets of data from separate information systems

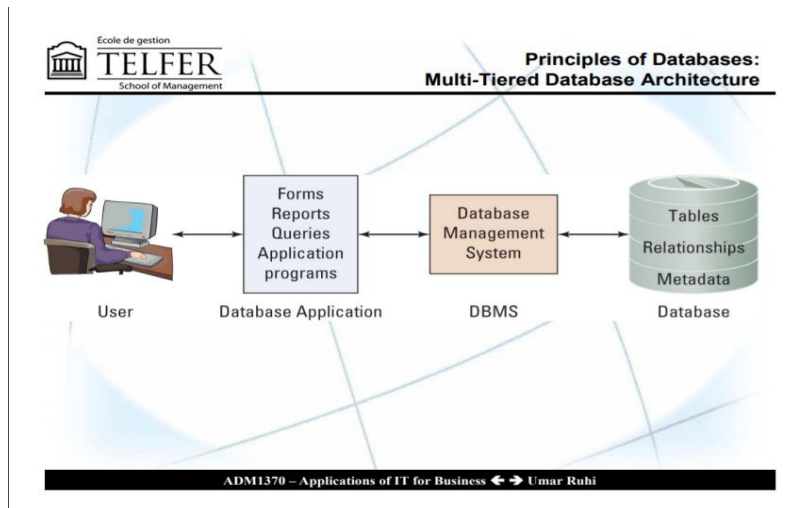
Some sources of low-quality information include:

- Online customers intentionally enter inaccurate information to protect their privacy
- Data or information from different system have different entry standards and formats
- Call center operators enter abbreviated or erroneous information by accident or to save time.
- Third party and external information contains inconsistencies, inaccuracies, and errors.

Potential business effects resulting from low quality information include:

- Inability to accurately track customers

- Difficulty identifying valuable customers



Stored Data Formats:

- Atomic Data:
 - Data in units that cannot be subdivided
 - First Name, Price, etc.
- Binary Large Objects (BLOB):
- Digitized Image and Multimedia (Audio Video)
- Large size
- Stored as a single entity (can't be handled in smaller chunks)

Typical Database Design Phases

1. Requirements Gathering:
 - Customer Interviews and Business Requirements
2. Report Design
 - Output Required (Another Method for Requirements Gathering)
3. Data Design
 - Data items inventory, metadata and primary keys for Data
4. Table Design
 - Entity-Relationship Diagrams (ERD) and Normalization
5. Form Design
 - Input Interface Design

Phases and Techniques for Database design:

Conceptual Design:

- Abstract model of database from a business perspective

Physical Design

- Detailed description of business information needs

Entity-Relationship Diagram (ERD):

Methodology for documenting databases illustrating relationships between database entities

Normalization: Process of creating small stable data structures from complex groups of data

A key tool employed in the design of databases is the entity relationship diagram (ERD)

ERD is a methodology for documenting databases and is used to illustrate the relationships between entities.

Various Notations for ERD exist including the “named diamond” notation and the “crow’s feet” notation.

Entity- A person, place, thing, transaction or event about which information is stored (Customers, Orders, Employees, Locations, etc.)

Relationship:

Association between entities

Degree of a relationship indicates the number of entities associated with a relationship

- Unary relationship exists when an association is maintained within a single entity;
- Binary relationship exists when two entities are associated;
- Ternary relationship exists when three entities are associated

Connectivity:

Entity relationship may be classified as one-to-one, one-to-many, or many to many.

Cardinality:

The maximum number of times an instance of one entity can be associated with an instance in the related entity.

Cardinality can be mandatory single, optional single, mandatory many, or optional many.

Business rules:

Precise descriptions of policies, procedures or principles in any organization which are derived from a description of an organization’s operations.

Can be used to determine the nature of entity relationships.

ERD Notation:

MANDATORY SINGLE



OPTIONAL SINGLE



MANDATORY MANY



OPTIONAL MANY



ERD EXAMPLE:

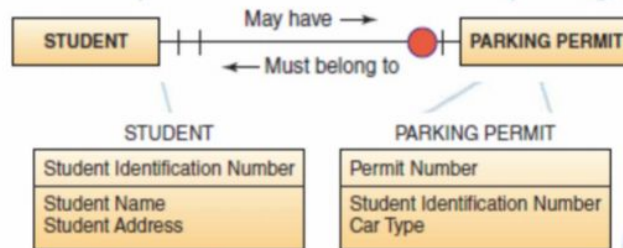
ERD Examples:

Students & Parking Permits:

- One-to-One (1:1) STUDENT-PARKING PERMIT relationship

Business Rule:

- Each Student may only have one parking permit.



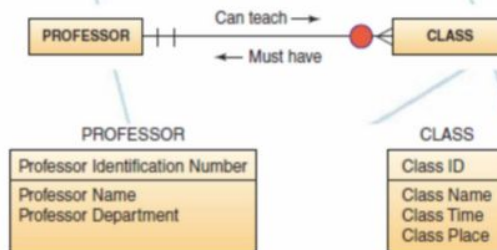
ERD Examples:

Professors & Classes:

- One-to-Many (1:M) CLASS-PROFESSOR relationship

Business Rule:

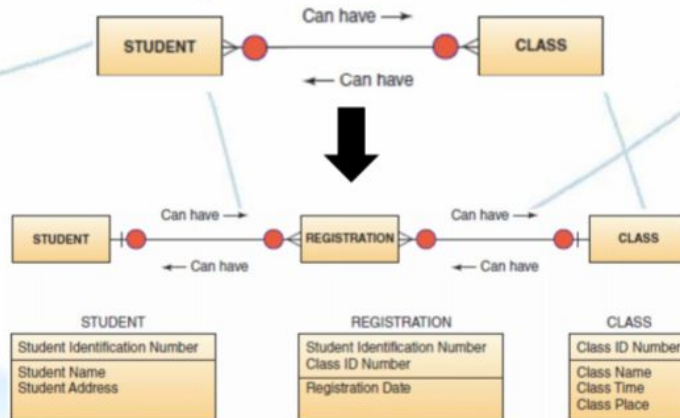
- No Team Teaching Allowed.



ERD Examples:

Students & Classes:

- Many-to-Many (M:M) STUDENT-CLASS relationship.
 - Not Supported in many RDBMS
 - Needs to be broken into two 1:M relationships by using a Junction (Bridge) entity/table.



The relationships tool in MS access allows explicit creation of relationships between tables.

- Referential integrity & Cascading Updates ensures data integrity between parent and child tables.

Database normalization is the process of removing redundant data from your tables into improves storage efficiency, data integrity, and scalability.

It is a method for analyzing and reducing a relational database to its most streamlined form to ensure minimum redundancy, maximum data integrity, and optimal processing performance.

In the relational model, methods exist for quantifying how efficient a database is. These classifications are called normal forms or NF, and there are algorithms for converting a given database between them.

Normalization generally involves splitting existing tables into multiple ones, which must be re-joined or linked each time a query is issued.

1st Normal Form (1NF):

- Identify/assign Primary Keys

However,

- Multiple entities still represented in the same table
- Table contains repeating groups

2nd Normal Form (2NF):

Separate tables for some entities

Functional dependency between primary key and all attributes within the same table.

However,

- Some tables still contain repeating groups
- Still need to separate some entities further.

3rd Normal Form (3NF):

- Separate tables for all entities
- Foreign keys are used to link related tables
- No Data redundancy

Query:

- A set of instructions used for working with data.
- Creating a query is like asking the database a question
- Running a query performs these instructions and provides the answers.
 - The results that a query returns can be sorted, grouped or filtered
- A query can get its data from one or more tables, from existing queries, or from a combination of the two.
- The tables or queries from which a query gets its data are referred to as its record source
- The results from a query are referred to as its recordset.

Methods to Create Queries

- Structured Query Language (SQL):
The most popular query language used for interacting with a database

Allows people to perform complicated searches by using relatively simple statements or key words.

Typical keywords are SELECT (to choose a desired attribute), FROM (to specify the table or tables to be used), and WHERE (to specify conditions to apply in the query).

Query by Example (QBE):

The user fills out a grid or template also known as a form to construct a sample or a description of the data desired. Users can construct a query quickly and easily by using a drag and drop feature in DBMS such as Microsoft Access.

Conducting queries in this manner is simpler than keying in SQL commands

A DBMS has a data manipulation language used to add, change, delete and retrieve data in the database

One such language is SQL

- MS access uses SQL behind a user-friendly interface

DBMS products fall into two broad categories

- Enterprise DBMS and personal/desktop DBMS

Enterprise DBMS

- These products process large operational and analytical database
- These products can support thousands of users at a time, along with multiple database applications.
- IBM's DB2, Microsoft's SQL server and Oracle are examples of enterprise DBMS products.

Personal/Desktop DBMS

- These products are designed for smaller, simpler database applications
- Such products are used for personal or small workgroup applications that involve fewer than 100 users, and normally fewer than 15.
- The great bulk of databases in this category have only a single user.

Data Warehouses and Data Marts

- For business decision-making purpose, data from many sources and over many time periods must be gathered together and organized in a consistent and useful way.
- Data warehouse and data marts support business intelligence applications
- A data warehouse is a copy of transaction data specifically structured for querying analysis reporting and more rigorous data mining.
 - A data warehouse is a repository of historical data that is organized by subject to support decision makers in the organization
 - The data warehouse contains a copy of the transactions which are not updated or changed later directly by the transaction system.
- A data mart is a low-cost, scaled down version of a data warehouse that is designed for the end user needs in a strategic business unit or an individual department.

Extraction, transformation and loading (ETL):

- A process that extracts information from internal and external databases, transforms the information using a common set of enterprise definitions, and loads the information into a data warehouse

Data mart:

- Contains a subset of data warehouse information

- Contains summarized or highly focused portion of data for a specified business unit or group of users
- Contains a snapshot of operational data for specific business contexts.
- Creation of a data mart is predicated on a specific, predefined need for a certain grouping and configuration of select data

Organized by business dimension or subject:

- Data is organized by subject, this arrangement differs from operational databases where data is organized by business processes such as order entry, inventory control
- Use online analytical processing (OLAP):
Involves the analysis of accumulated data by end users, as opposed to operational databases that use online transaction processing (OLTP)
- Integrated:
Data is collected from multiple systems and then integrated around subjects.
- Time variant:
Data warehouses and data marts maintain historical data, unlike transactional systems, which maintain only recent data.

Non-volatile:

End-users or transactional processes cannot change or update the data directly. Warehouses and marts are updated through IT-controlled load processes rather than by users.

Multidimensional:

Typically, the data warehouse or mart uses a multidimensional data structure. A common representation for this multidimensional structure is the data cube.

Multidimensional Data Analysis using OLAP (online analytical processing)

Allows users to analyze large volumes of data and present the same data in different ways.

Each aspect of information represents a dimension