

PRELUDE

- ✓ Data mining is the process of extracting hidden predictive information from a large database. As more data are gathered, with the amount of data doubling every year, data mining is becoming an increasingly important tool to transform this data into information.
- ✓ It is commonly used in a wide range of profiling practices, such as marketing, fraud detection and scientific discovery. Data mining can be applied to data sets of any size.
- ✓ Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge driven decisions.
- ✓ Data mining sometimes called data or knowledge discovery is the process of analyzing data from different perspectives and summarizing it into useful information.
- ✓ Data mining software is an analytical tool for analyzing data. It allows users to analyze data from many different dimensions, 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.

Data mining consists of five major elements:

- ✓ Extract, transform, and load transaction data onto the data warehouse system.
- ✓ Store and manage the data in a multidimensional database system.
- ✓ Provide data access to business analysts and information technology professionals.
- ✓ Analyze the data by application software.
- ✓ Present the data in a useful format, such as a graph or table.

ADVANTAGES OF DATA MINING

- ✓ **Automated prediction of trends and behaviors :**
 - Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data, quickly.
- ✓ **Automated Discovery of previously unknown patterns:**
 - Data mining tools sweep through databases and identify previously hidden patterns in one step.
 - E.g. – analysis of retail sales data to identify apparently unrelated products that are often purchased together.
- ✓ **Database can be larger in both depth and breadth:**
 - The databases can have more columns and rows. High performance data mining allows users to explore full depth of a database, without pre-selecting a subset of variables. The data mining database contain larger samples (more rows) as they yield lower estimation errors and variance, and allow users to make conclusion about small but important segments of a population.

Data mining techniques can yield the benefits of automation on existing software and hardware platforms and can be implemented on new systems as existing platforms are upgraded and new products are developed. When data mining tools are implemented on high performance parallel processing systems, they can analyze massive databases in minutes. Faster processing means the users can experiment with more models to understand complex data. High speed makes it practical for users to analyze huge quantities of data. Larger databases, in turn, yield improved predictions.

TECHNOLOGIES USED IN DATA MINING

- ✓ **Artificial Neural Networks:** Non-linear predictive models that learn through training and resemble biological neural networks in structure.
- ✓ **Genetic Algorithms:** Optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution.
- ✓ **Decision Trees:** Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID) . CART and CHAID are decision tree techniques used for classification of a dataset. They provide a set of rules that you can apply to a new (unclassified) dataset to predict which records will have a given outcome. CART segments a dataset by creating 2-way splits while CHAID segments using chi square tests to create multi-way splits. CART typically requires less data preparation than CHAID.
- ✓ **Nearest Neighbor Method:** A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where k is 1). Sometimes called the k-nearest neighbor technique. It is also called case based reasoning.
- ✓ **Rule Induction:** The extraction of useful if-then rules from data based on statistical significance.
- ✓ **Data Visualization:** The visual interpretation of complex relationships in multidimensional data. Graphics tools are used to illustrate data relationships.
- ✓ **Evolutionary Programming:** This is the most promising branch of data mining at present. The underlying idea of the method is that the system automatically formulates hypothesis about the dependence of the target variable on other variables in the form of programs expressed in an internal programming language.

PRELUDE

- ✓ GIS is a computer based tool for mapping and analyzing things that exist and events that happens earth.
- ✓ GIS allows us to view, understand, question, interpret, and visualize data in many ways that reveal relationships, patterns, and trends in the form of maps, globes, reports, and charts.
- ✓ A GIS helps us answer questions and solve problems by looking at your data in a way that is quickly understood and easily shared.
- ✓ A GIS captures, stores, analyzes, manages, and presents data that refers to or is linked to location.
- ✓ It is an information system that integrates, stores, edits, analyzes, shares, and displays geographic information.
- ✓ GIS applications are tools that allow users to create interactive queries (user created searches), analyze spatial information, edit data, maps, and present the results of all these operations.
- ✓ GIS is a computer based information system used to digitally represent and analyze the geographic features present on the Earth's surface and the events taking place on it.
- ✓ "Every object present on the Earth can be geo-referenced", is the fundamental key of associating any database to GIS. Here, term 'database' is a collection of information about things and their relationship to each other, and 'geo-referencing' refers to the location of a layer or coverage in space defined by the co-ordinate referencing system.
- ✓
- ✓ Geographic information science is the science underlying the geographic concepts, applications and systems, taught in degree and GIS Certificate programs at many universities.

APPLICATIONS

Geographic information system technology can be used for

- ❖ Scientific investigations
- ❖ Resource management
- ❖ Asset management
- ❖ Archaeology
- ❖ Environmental impact assessment
- ❖ Urban planning
- ❖ Cartography (Study and practice of making geographical maps)
- ❖ Criminology
- ❖ Geographic history
- ❖ Marketing
- ❖ Logistics

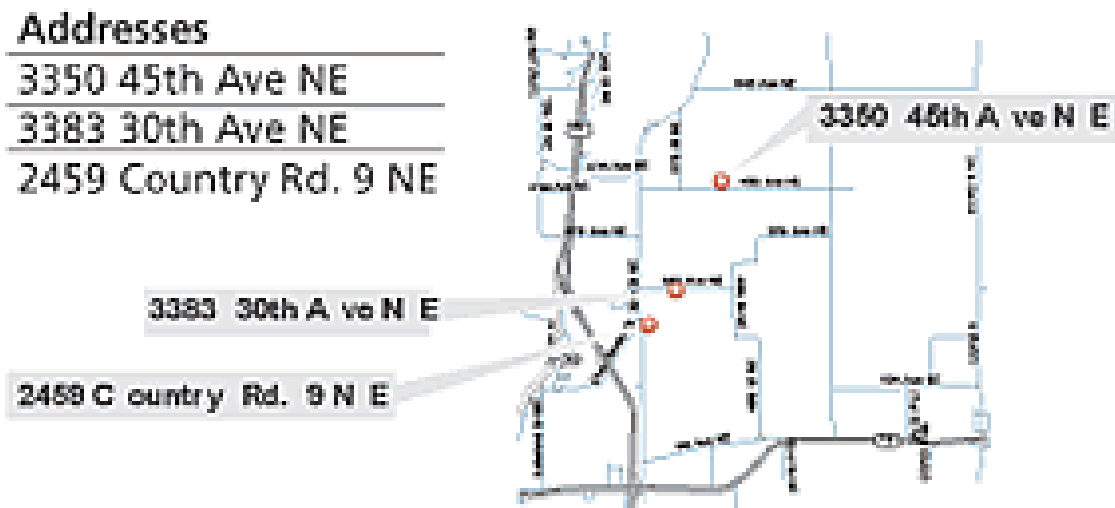
For example, GIS might allow emergency planners to easily calculate emergency response times in the event of a natural disaster, GIS might be used to find wetlands that need protection from pollution, or GIS can be used by a company to site a new business location to take advantage of a previously under-served market.

THREE VIEWS OF A GIS

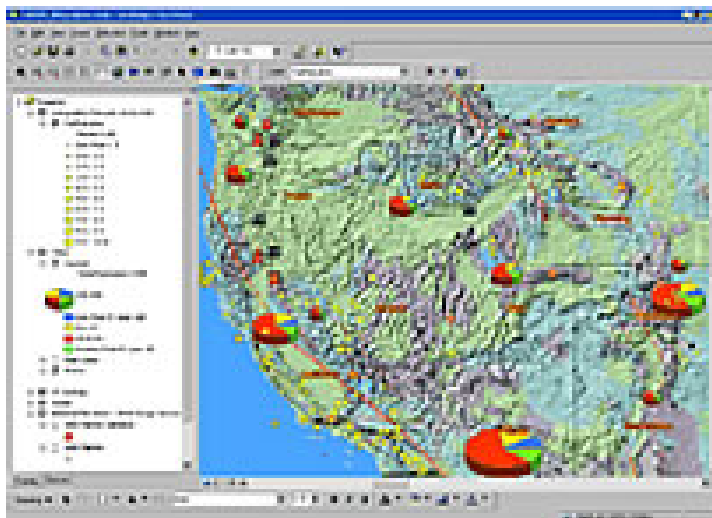
A GIS is most often associated with a map. A map, however, is only one way you can work with geographic data in a GIS, and only one type of product generated by a GIS. A GIS can provide a great deal more problem-solving capabilities than using a simple mapping program or adding data to an online mapping tool (creating a "mash-up").

A GIS can be viewed in three ways:

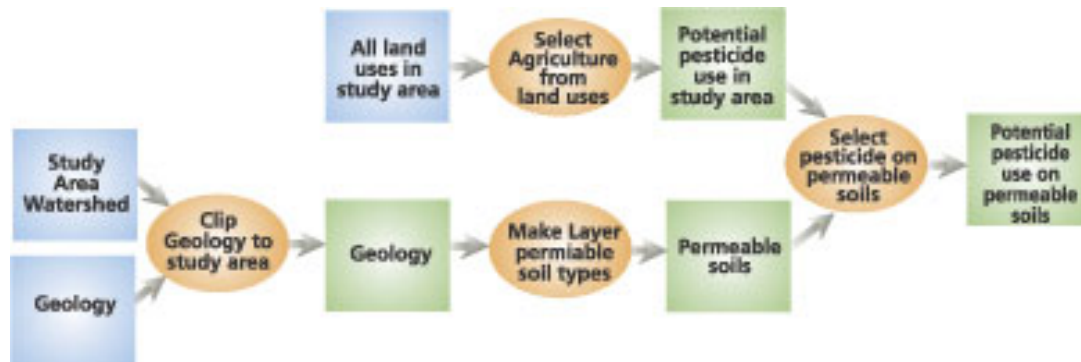
1. The Database View: A GIS is a unique kind of database of the world—a geographic database (geodatabase). It is an "Information System for Geography." Fundamentally, a GIS is based on a structured database that describes the world in geographic terms.



2. The Map View: A GIS is a set of intelligent maps and other views that show features and feature relationships on the earth's surface. Maps of the underlying geographic information can be constructed and used as "windows into the database" to support queries, analysis, and editing of the information.



3. The Model View: A GIS is a set of information transformation tools that derive new geographic datasets from existing datasets. These geoprocessing functions take information from existing datasets, apply analytic functions, and write results into new derived datasets.



COMPONENTS OF GIS

A working GIS integrates five key components: hardware, software, data, people and methods.

Hardware:

Hardware consists of the technical equipment needed to run a GIS including a computer system with enough power to run the software, enough memory to store large amounts of data, and input and output devices such as scanners, digitizers, GPS data loggers, media disks, and printers.

It is the computer on which a GIS operates. Today, GIS software runs on a wide range of hardware types, from centralized computer servers to desktop computers used in stand alone or networked configurations.



Software:

GIS software provides the functions and tools needed to store, analyze and display geographic information. Key software components are:

- ❖ Tools for the input and manipulation of geographic information
- ❖ A database management system
- ❖ Tools that support geographic query, analysis and visualization
- ❖ A graphical user interface for easy access to tools

There are many different GIS software packages available today. All packages must be capable of data input, storage, management, transformation, analysis, and output, but the appearance, methods, resources, and ease of use of the various systems may be very different. Today's software packages are capable of allowing both graphical and descriptive data to be stored in a single database, known as the object-relational model. Before this innovation, the geo-relational model was used. In this model, graphical and descriptive data sets were handled separately. The modern packages usually come with a set of tools that can be customized to the users' needs (Lo, 2002).

Common GIS Software: GRASS, Quantum GIS, ILWIS, ARC View, ARC GIS, MapWindow GIS

Data:

Perhaps the most time consuming and costly aspect of initiating a GIS is creating a database. There are several things to consider before acquiring geographic data. It is crucial to check the quality of the data before obtaining it. Errors in the data set can add many unpleasant and costly hours to implementing a GIS and the results and conclusions of the GIS analysis most likely will be wrong. Several guidelines to look at include:

- ❖ **Lineage** – This is a description of the source material from which the data were derived, and the methods of derivation, including all transformations involved in producing the final digital files. This should include all dates of the source material and updates and changes made to it.
- ❖ **Positional Accuracy** – This is the closeness of an entity in an appropriate coordinate system to that entity's true position in the system. The positional accuracy includes measures of the horizontal and vertical accuracy of the features in the data set.
- ❖ **Attribute Accuracy** – An attribute is a fact about some location, set of locations, or features on the surface of the earth. This information often includes measurements of some sort, such as temperature or elevation or a label of a place name. The source of error usually lies within the collection of these facts. It is vital to the analysis aspects of a GIS that this information be accurate.
- ❖ **Logical Consistency** - Deals with the logical rules of structure and attribute rules for spatial data and describes the compatibility of a datum with other data in a data set. There are several different mathematical theories and models used to test logical consistency such as metric and incidence tests, topological and order related tests. These consistency checks should be run at different stages in the handling of spatial data.
- ❖ **Completeness** – This is a check to see if relevant data is missing with regards to the features and the attributes. This could deal with either omission errors or spatial rules such as minimum width or area that may limit the information.

People

The people are the component who actually makes the GIS work. They include a plethora (surplus) of positions including GIS managers, database administrators, application specialists, systems analysts, and programmers. They are responsible for maintenance of the geographic database and provide technical support. People also need to be educated to make decisions on what type of system to use. People associated with a GIS can be categorized into: viewers, general users, and GIS specialists.

- ❖ **Viewers** are the public at large whose only need is to browse a geographic database for referential material. These constitute the largest class of users.
- ❖ **General Users** are people who use GIS to conducting business, performing professional services, and making decisions. They include facility managers, resource managers, planners, scientists, engineers, lawyers and business entrepreneurs.
- ❖ **GIS specialists** are the people who make the GIS work. They include GIS managers, database administrators, application specialists, systems analysts, and programmers. They are responsible for the maintenance of the geographic database and the provision of technical support to the other two classes of users.

Procedures

A successful GIS operates according to a well-designed plan and business rules, which are the models and operating practices unique to each organization.

- ❖ Procedures include how the data will be retrieved, input into the system, stored, managed, transformed, analyzed, and finally presented in a final output.
- ❖ The procedures are the steps taken to answer the question that is needed to be resolved. The ability of a GIS to perform spatial analysis and answer these questions is what differentiates this type of system from any other information systems.

The transformation processes includes such tasks as adjusting the coordinate system, setting a projection, correcting any digitized errors in a data set, and converting data from vector to raster or raster to vector.

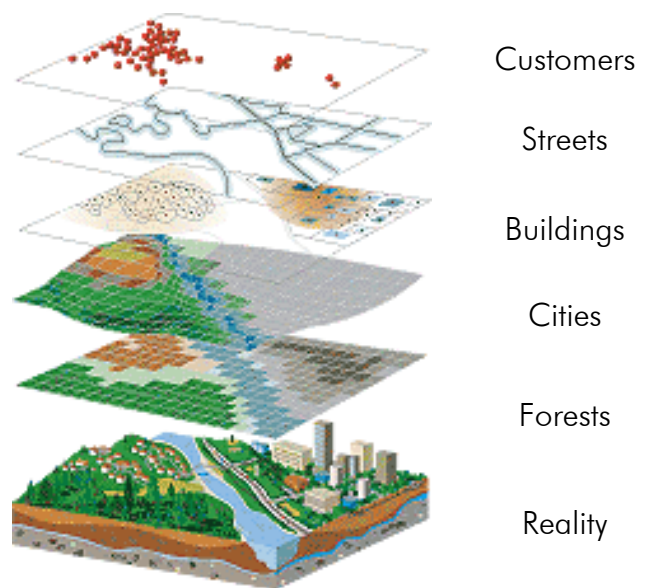
How GIS WORKS?

GIS store information about the world as a collection of thematic layers that can be linked together by geography. This simple but extremely powerful and versatile concept has proven invaluable for solving many real-world problems from tracking delivery vehicles, to recording details of planning applications, to modeling global atmospheric circulation.

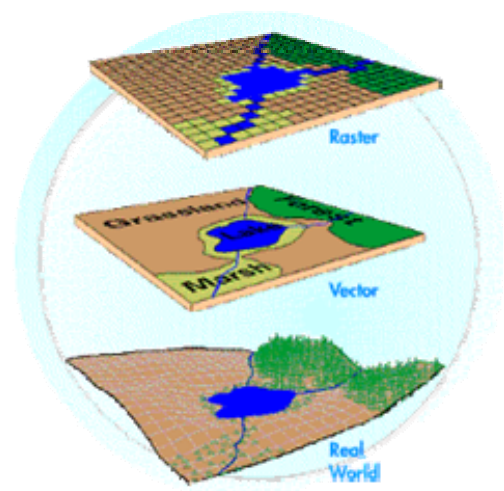
Geographic information contains either an explicit geographic reference, such as a latitude and longitude or national grid coordinate, or an implicit reference such as an address, postal code, census tract name, forest stand identifier, or road name. An automated process called geocoding is used to create explicit geographic references (multiple locations) from implicit references (descriptions such as addresses). These geographic references allow you to locate features, such as a business or forest stand, and events, such as an earthquake, on the earth's surface for analysis.

Vector and Raster Models

Geographic information systems work with two fundamentally different types of geographic models--the "vector" model and the "raster" model. In the vector model, information about points, lines, and polygons is encoded and stored as a collection of x,y coordinates. The location of a point feature, such as a bore hole, can



How GIS works (y layers)?



be described by a single x,y coordinate. Linear features, such as roads and rivers, can be stored as a collection of point coordinates. Polygonal features, such as sales territories and river catchments, can be stored as a closed loop of coordinates.

The vector model is extremely useful for describing discrete features, but less useful for describing continuously varying features such as soil type or accessibility costs for hospitals. The raster model has evolved to model such continuous features. A raster image comprises a collection of grid cells rather like a scanned map or picture. Both the vector and raster models for storing geographic data have unique advantages and disadvantages. Modern GISs are able to handle both models.

DATA FOR GIS

Base Maps: Includes streets and highways; boundaries for census, postal and political areas, rivers, lakes, parks and landmarks and place names.

Business Maps and Data: Includes data related to census/demography, consumer products, financial services, health care, telecommunication, emergency preparedness, crime, advertising, transportation and business establishment.

Environmental Maps and Data: Includes data related to the weather, environmental risk, satellite imagery, topography and natural resources.

General Reference Maps: Includes world and country maps and data

GIS AND RELATED TECHNOLOGIES

GISs are closely related to several other types of information systems, but it is the ability to manipulate and analyze geographic data that sets GIS technology apart. Although there are no hard and fast rules about how to classify information systems, the following discussion should help differentiate GIS from desktop mapping, computer-aided design (CAD), remote sensing, DBMS, and global positioning systems (GPS) technologies.

Desktop Mapping

A desktop mapping system uses the map metaphor (symbol) to organize data and user interaction. The focus of such systems is the creation of maps: the map is the database. Most desktop mapping systems have more limited data management, spatial analysis, and customization capabilities. Desktop mapping systems operate on desktop computers such as PCs, Macintoshes, and smaller UNIX workstations.

CAD

CAD systems evolved to create designs and plans of buildings and infrastructure. This activity required that components of fixed characteristics be assembled to create the whole structure. These systems require few rules to specify how components can be assembled and very limited analytical capabilities. CAD systems have been extended to support maps but typically have limited utility for managing and analyzing large geographic databases.

Remote Sensing and GPS

Remote sensing is the art and science of making measurements of the earth using sensors such as cameras carried on airplanes, GPS receivers, or other devices. These sensors collect data in the form of images and provide specialized capabilities for manipulating, analyzing, and visualizing those images. Lacking strong geographic data management and analytical operations, they cannot be called true GISs.

DBMS

Database management systems specialize in the storage and management of all types of data including geographic data. DBMSs are optimized to store and retrieve data and many GISs rely on them for this purpose. They do not have the analytic and visualization tools common to GIS.

USAGES OF GIS

Perform Geographic Queries and Analysis

The ability of GIS to search databases and perform geographic queries has saved many companies literally millions of dollars. GIS has helped reduce costs by

- ❖ Streamlining (reorganizing) customer service.
- ❖ Reducing land acquisition costs through better analysis.
- ❖ Reducing fleet maintenance costs through better logistics.
- ❖ Analyzing data quickly.

Improve Organizational Integration

Many organizations that have implemented a GIS have found that one of its main benefits is improved management of their own organization and resources. Because GISs have the ability to link data sets together by geography, they facilitate interdepartmental information sharing and communication. By creating a shared database, one department can benefit from the work of another - data can be collected once and used many times.

As communication increases among individuals and departments, redundancy is reduced, productivity is enhanced, and overall organizational efficiency is improved. Thus, in a utility company the customer and infrastructure databases can be integrated so that when there is planned maintenance, affected customers can be sent a computer-generated letter.



Make Better Decisions

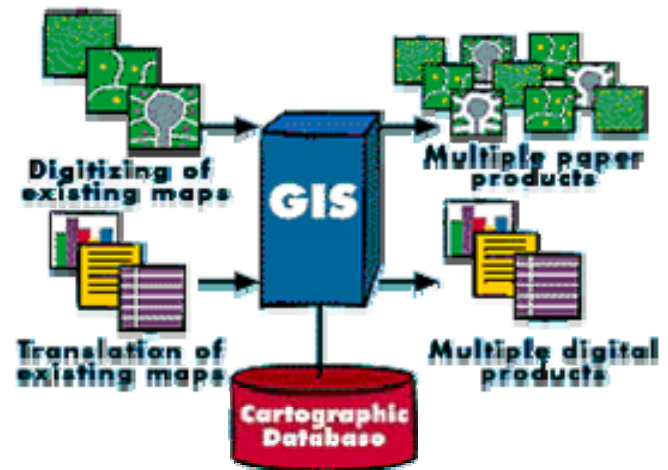
The old adage "better information leads to better decisions" is as true for GIS as it is for other information systems. A GIS, however, is not an automated decision making system but a tool to query, analyze, and map data in support of the decision making process. GIS technology has been used to assist in tasks such as presenting information at planning inquiries, helping resolve territorial disputes and minimize visual intrusion.

GIS can be used to help reach a decision about the location of a new housing development that has minimal environmental impact, is located in a low-risk area, and is close to a population center. The

information can be presented succinctly and clearly in the form of a map and accompanying report, allowing decision makers to focus on the real issues rather than trying to understand the data. Because GIS products can be produced quickly, multiple scenarios can be evaluated efficiently and effectively.

Making Maps

Maps have a special place in GIS. The process of making maps with GIS is much more flexible than are traditional manual or automated cartography approaches. It begins with database creation. Existing paper maps can be digitized and computer-compatible information can be translated into the GIS. The GIS-based cartographic database can be both continuous and scale free. Map products can then be created centered on any location, at any scale, and showing selected information symbolized effectively to highlight specific characteristics.



The characteristics of atlases and map series can be encoded in computer programs and compared with the database at final production time. Digital products for use in other GISs can also be derived by simply copying data from the database. In a large organization, topographic databases can be used as reference frameworks by other departments.

GIS IN EVERYDAY LIFE

Making GIS Data Work for You

- ✓ In today's global community, the more information you have at your fingertips, the easier it is to make an informed decision. In today's high-tech world, information comes in many different ways, from company reports and statistics from down the hall to digital photos and multimedia from across the world.
- ✓ Information can be overwhelming and the need for timely decisions calls not only for innovative ways to access accurate, up-to-the minute information, but also tools to help present the information in useful ways.
- ✓ A geographic information system or GIS allows you to bring all types of data together based on the geographic and locational component of the data.
- ✓ But unlike a static paper map, GIS can display many layers of information that is useful to you.
- ✓ You will be able to integrate, visualize, manage, solve, and present the information in a new way.
- ✓ Relationships between the data will become more apparent and your data will become more valuable.
- ✓ GIS will give you the power to create maps, integrate information, visualize scenarios, solve complicated problems, present powerful ideas, and develop effective solutions like never before.
- ✓ GIS is a tool used by individuals and organizations, schools, governments, and businesses seeking innovative ways to solve their problems.

Computer in Business and Industry

Office Automation System

- Is defined as using computer and communications technology to help people better use and manage information.
- Includes all types of computers, telephones, electronic mail and office machines that use microprocessors of other high- technology components.
- People who use office automation are often called knowledge workers – senior executives, managers, supervisors, analysts, engineers and other white- collar office workers
- These systems keep track of the information originating in various operations throughout the company, such as order processing, accounting, inventory and manufacturing.
- Provides knowledge workers with the information – producing systems to collect, analyze, plan, and control information about the many facets of the business, using text, voice, graphics and video display technology.
- People
- Ergonomics
 - The study of how to create safety, comfort and ease of use for the people who use machines.
 - Office tasks involve a great deal of thinking and decision making so it is different from factory automation.
 - Office systems must be flexible and versatile; moreover they must be designed so any knowledge worker, regardless of background can easily use them.
 - With the advent of computers, ergonomics engineers became particularly interested in office automation systems, furniture and environments for the knowledge workers,
 - Best design of keyboards, set eye fatigue level for monitor and specified desk, seating design.
 - Ergonomics has played a significant role in helping people use technology more effectively.

Office Automation Technologies

- Five primary technologies used in managing information in office automation.
 - Text or written words
 - Data, as in numbers or non- text formats
 - Graphics, including drawings, charts and photographs
 - Audio, as in telephone, voice mail or voice recognition systems
 - Video, such as captured images, videotapes or teleconferencing
- In past these forms of information was created using different technologies.
 - Text – conventional typewriters or word processing
 - Data - such as sales report was provided by central computers
 - Graphs and charts – hand written or creating using slide photography.
- It was not possible to combine all those things.
- Computer made it possible to combine them.
- Computer produces electronic documents, which is self- contained works, conveying information that has been created by a knowledge worker and stored in a computer system.

- Electronic documents can be simple as text file or more complex which contained graphics or even audio and video
- Data, sound and images can all be entered into a computer, stored and translated into the kind of output we need.
- Groupware
- The center of this integration are networking and communication systems

Office Automation Systems

- Uses computer – based systems to provide information to help knowledge workers make decisions that benefit the business.
- Comprise of many subsystems.

Text Management Systems

- Computer system designed to work with the written or typewritten word.
- Includes all kind of typewriters, word processing systems, PCs with word processing, desktop publishing and text editing system.
- Are used for tasks like writing documents simple and complex documents comprise of reports, memos, notices etc.

Business Analysis Systems

- Provides data that, when used with the proper software, helps its users better understand the business environment and make more effective decisions.
- Different software tools for performing analysis are
 - Decision Support System (DSS)
 - Helps the knowledge worker to extract information from the various MIS database and reporting systems, analyze it and then formulate a decision or a strategy for business planning
 - Expert System
 - Is a computer system that can store and retrieve data with special problem solving expertise.
 - Executive Support System
 - Is an information system that consolidates and summaries ongoing transactions with the organization.
 - It provides the management with all the information it requires at all times from internal as well as external sources.

Document Management Systems

- Aid in filing, tracking and managing documents, whether they are paper, computer based, micrographics or purely electronic.

Network and Communication Management Systems

- Includes telephone, electronic mails, voice messaging systems, teleconferencing and fax machines. Pbx (private branch exchange)

Management Information System

- A general term for all automated hardware and software used to provide and maintain information
- is the discipline covering the application of people, technologies, and procedures — collectively called information systems — to solving business problems
- can be any organized combinations people, hardware, software, communications networks, and data resources that collects, transforms and disseminates information in an organization
- The term is commonly used to refer to the group of information management methods tied to the automation or support of human decision making, e.g. Decision Support Systems, Expert systems, and Executive information systems.