Chapter 1

Information System

Definition

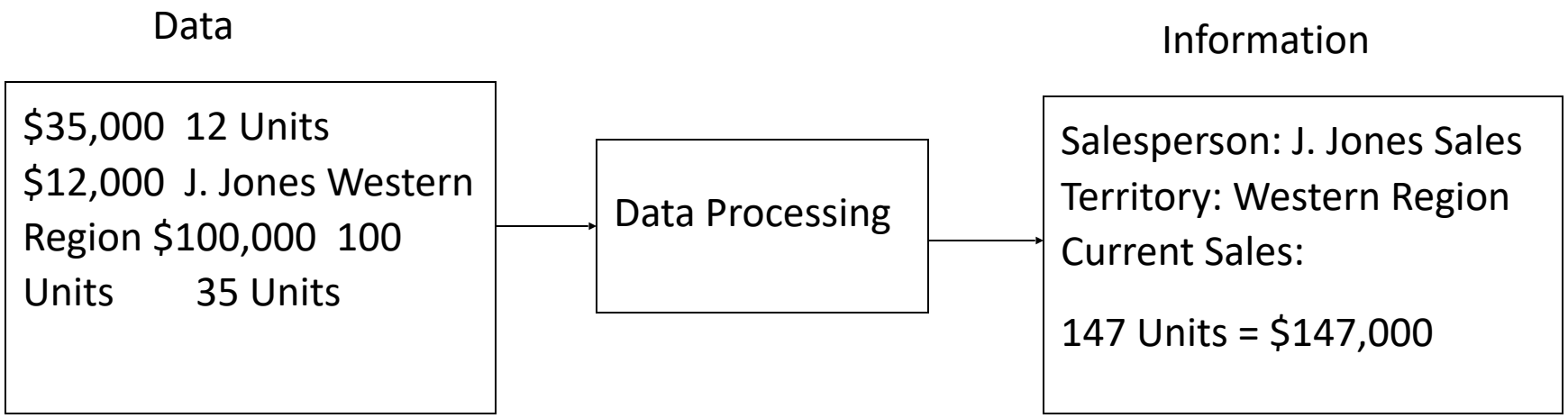
Data

- Raw facts such as an employee's name and number of hours worked in a week, inventory part numbers or sales orders.

Information

- A collection of facts organized in such a way that they have additional value beyond the value of the facts themselves.

Data Vs Information



Information System

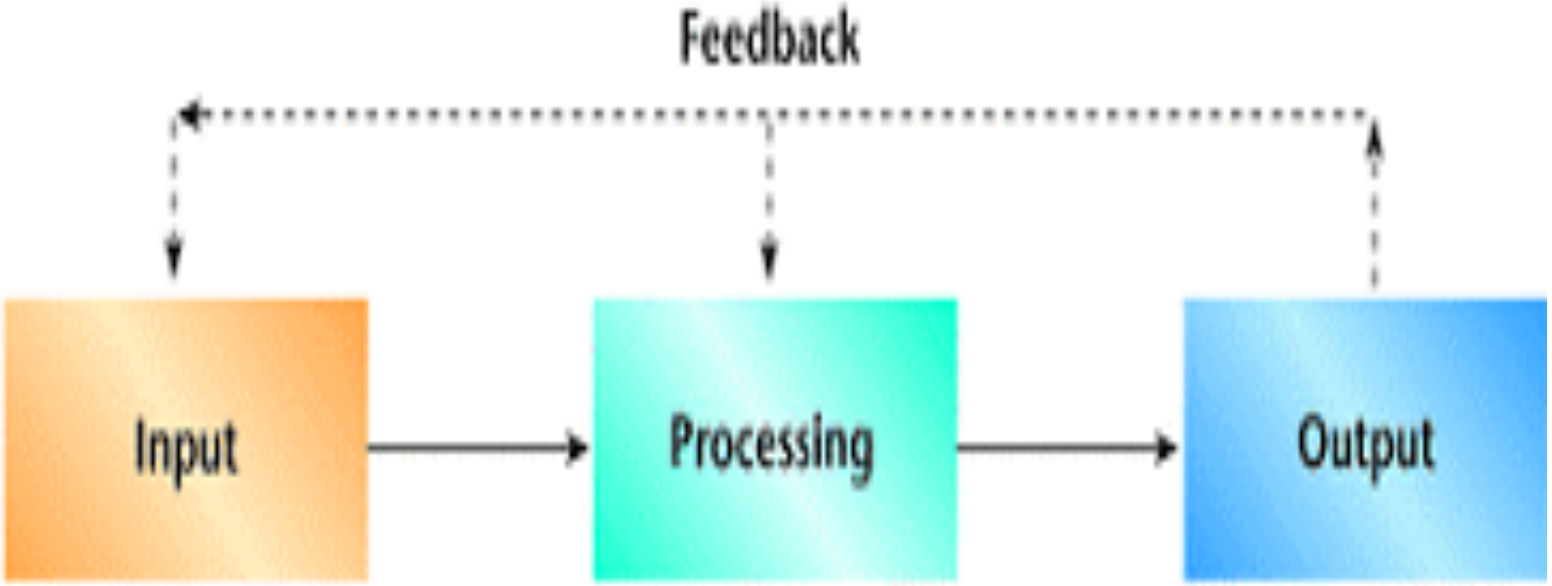
- Information System is the study of complementary networks of hardware and software that people and organizations use to collect, filter, process, create, and distribute data.
- The IS's designer is concerned with how to use computer systems effectively in producing data for the right person at the right time.

Definition

Information Systems

- An information system is typically considered to be a set of interrelated elements or components that collect(input), manipulate(processes), and disseminate (output) data and information and provide a feedback mechanism to meet an objective.
  - Open System
  - Close System

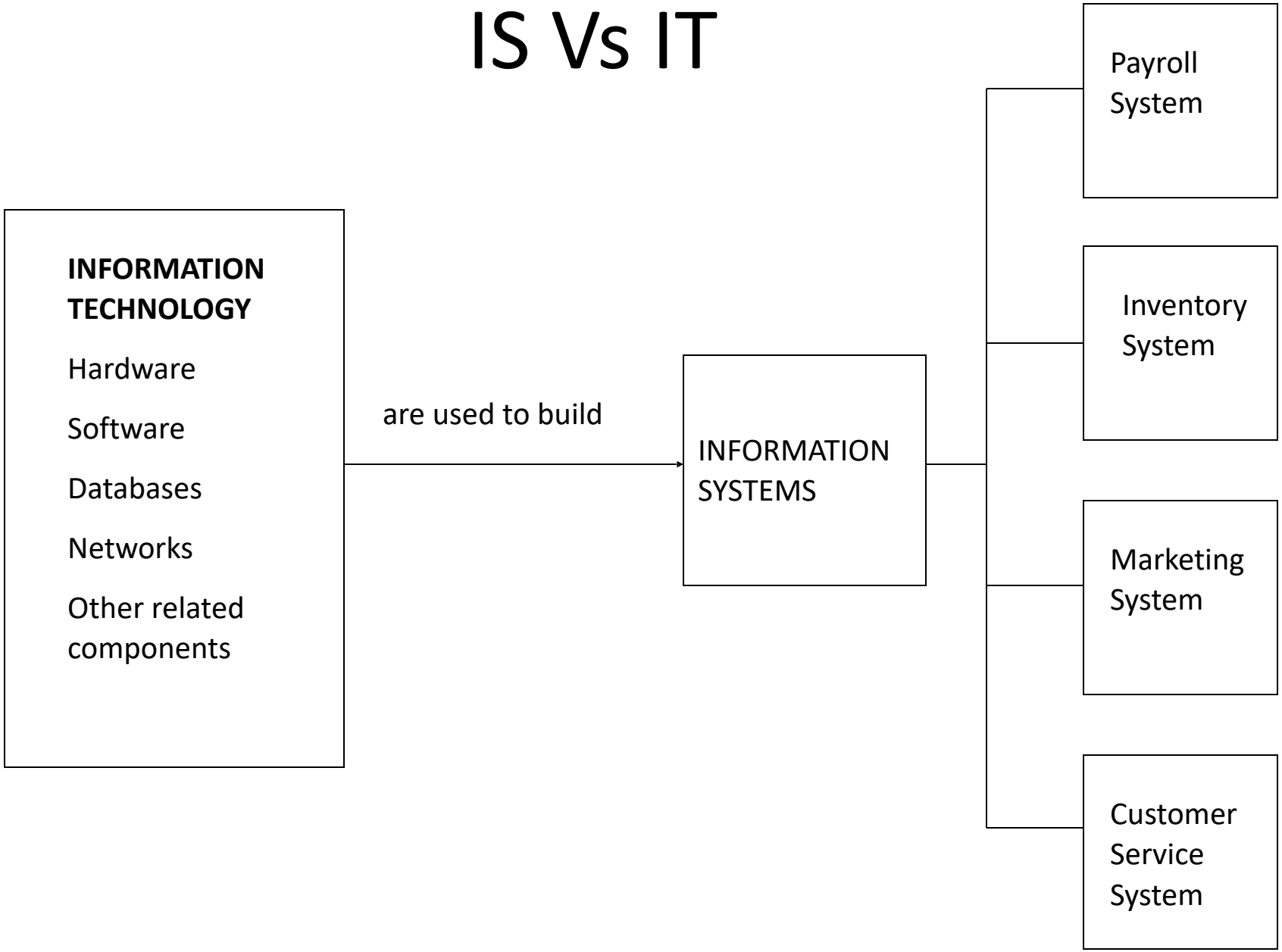
Information System



Computer-based Information System

An Information System is an organized combination of people, hardware, software, communication networks and the data resources that collects, transforms and disseminates information in a organization.

IS Vs IT


INFORMATION TECHNOLOGY
Hardware
Software
Databases
Networks
Other related components
are used to build
INFORMATION SYSTEMS
Payroll System
Inventory System
Marketing System
Customer Service System

Expanding Roles of IS

1. Data Processing: 1950s-1960s
2. Management Reporting: 1960s-1970s
3. Decision support: 1970s-1980s
4. Strategic and End User Support: 1980s-1990s
5. Global Internetworking: 1990s-2000s

Data Processing: 1950s-1960s

The first business application of computers (in the mid- 1950s) performed repetitive, high-volume, transaction-computing tasks. The computers "crunched numbers" summarizing and organizing transactions and data in the accounting, finance, and human resources areas. Such systems are generally called transaction processing systems (TPSs)

Management Reporting: 1960s-1970s

Management Information Systems (MISs): these systems access, organize, summarize and display information for supporting routine decision making in the functional areas. Office Automation Systems(OASs): such as word processing systems were developed to support office and clerical workers.

Decision support: 1970s-1980s

Decision Support Systems: were developed to provide computer based support for complex, non-routine decision.

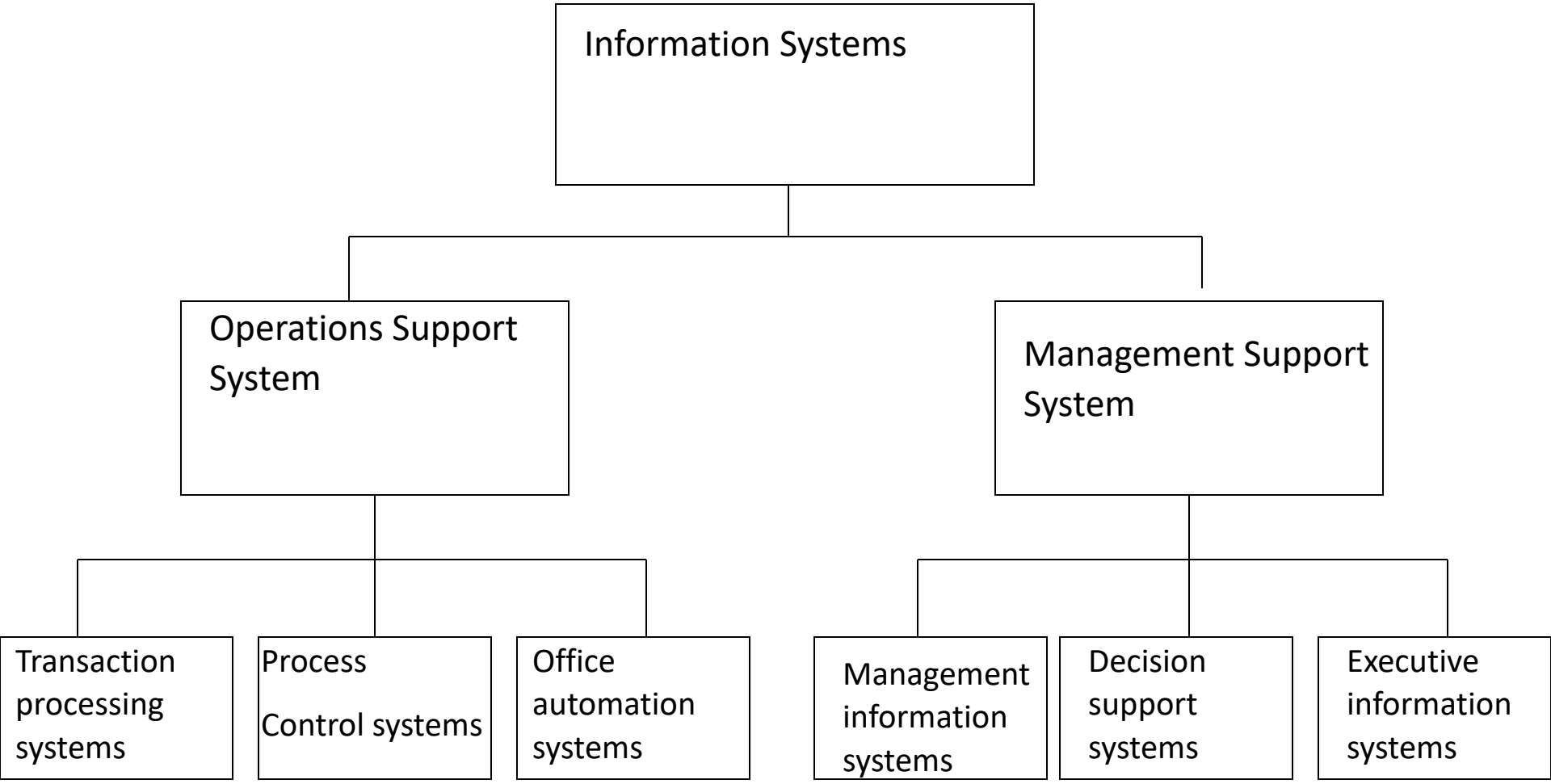
Strategic and End User Support: 1980s-1990s

- The use or development of information systems by the principal users of the systems' outputs, such as analysts, managers, and other professionals.
- Intelligent Support System (ISSs): Include expert systems which provide the stored knowledge of experts to non-experts, and a new type of intelligent system with machine- learning capabilities that can learn from historical cases.
- Knowledge Management Systems: Support the creating, gathering, organizing, integrating and disseminating of organizational knowledge.

Global Internetworking: 1990s-2000s

- Mobile Computing: Information systems that support employees who are working with customers or business partners outside the physical boundaries of their company; can be done over wire or wireless networks.

Classification of IS



1. **Operations support systems** process data generated by business operations

Major categories are:

i) Transaction processing systems

ii) Process control systems

iii) Office automation systems

2. Management Support Systems provide information and support needed for effective decision making by managers

Major categories are:

i) Management Information System

ii) Decision Support Systems

iii) Executive Information System

1. Operations Support System

i) Transaction processing systems

- Process business exchanges
- Maintain records about the exchanges
- Handle routine, yet critical, tasks
- Perform simple calculations

ii) **Process control systems** monitor and control industrial processes.

iii) **Office automation systems** automate office procedures and enhance office communications and productivity.

2. **Management support systems** provide information and support needed for effective decision making by managers

Major categories are:

i) Management information systems

- Routine information for routine decisions
- Operational efficiency
- Use transaction data as main input
- Databases integrate MIS in different functional areas

ii) Decision Support System

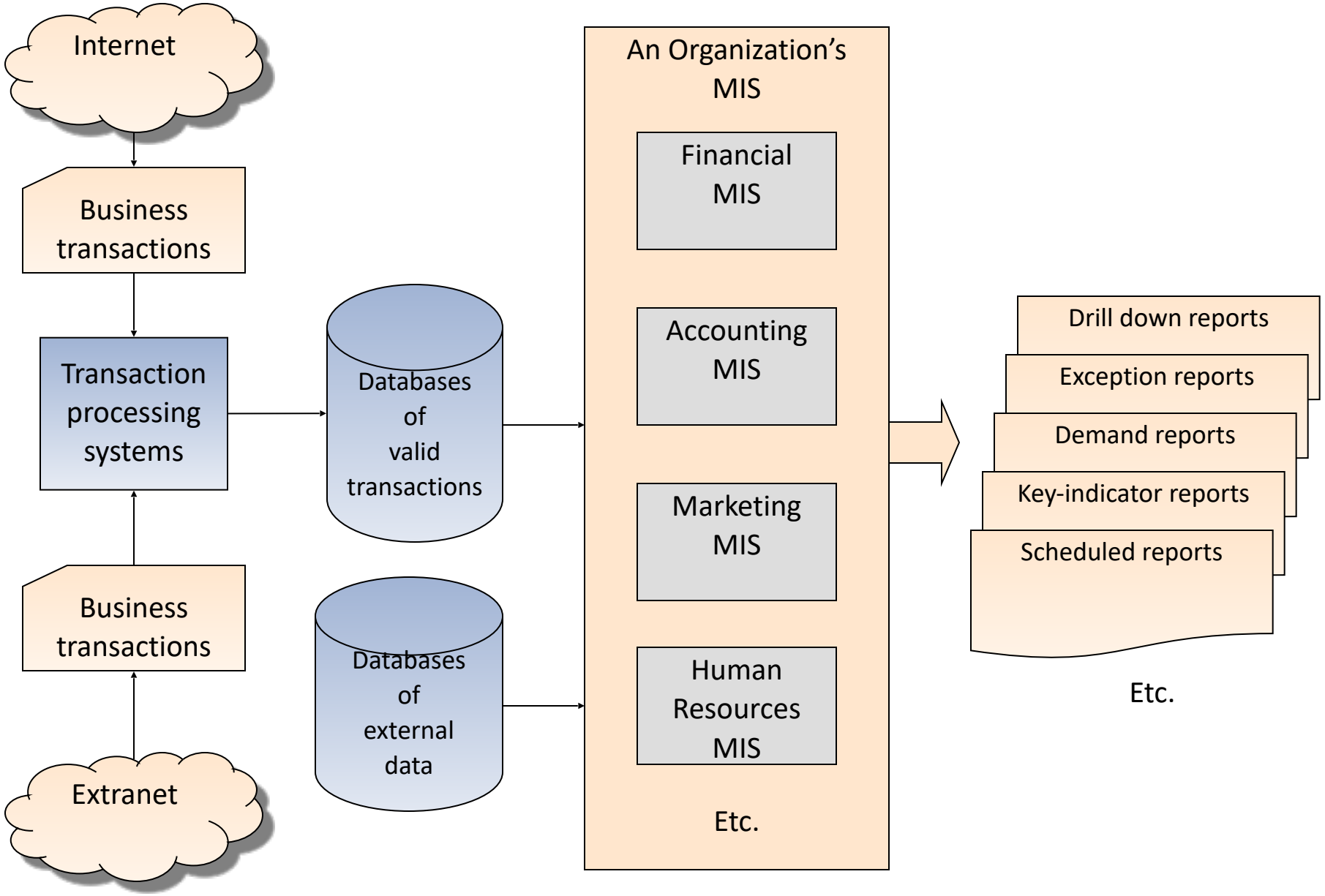
- Interactive support for non-routine decisions or problem
- End-users are more involved in creating a DSS than an MIS

iii) Executive information systems

provide critical information tailored to the information needs of executives

Functional Aspects

MIS is an integrated collection of functional information systems, each supporting particular functional areas.


Internet
Business transactions
Transaction processing systems
Business transactions
Extranet
Databases of valid transactions
Databases of external data
An Organization's MIS
Financial MIS
Accounting MIS
Marketing MIS
Human Resources MIS
Etc.
Drill down reports
Exception reports
Demand reports
Key-indicator reports
Scheduled reports
Etc.

MIS

Financial MIS

- Provides financial information to all financial managers within an organization.

Marketing MIS

- Supports managerial activities in product development, distribution, pricing decisions, and promotional effectiveness

MIS

Human Resource MIS

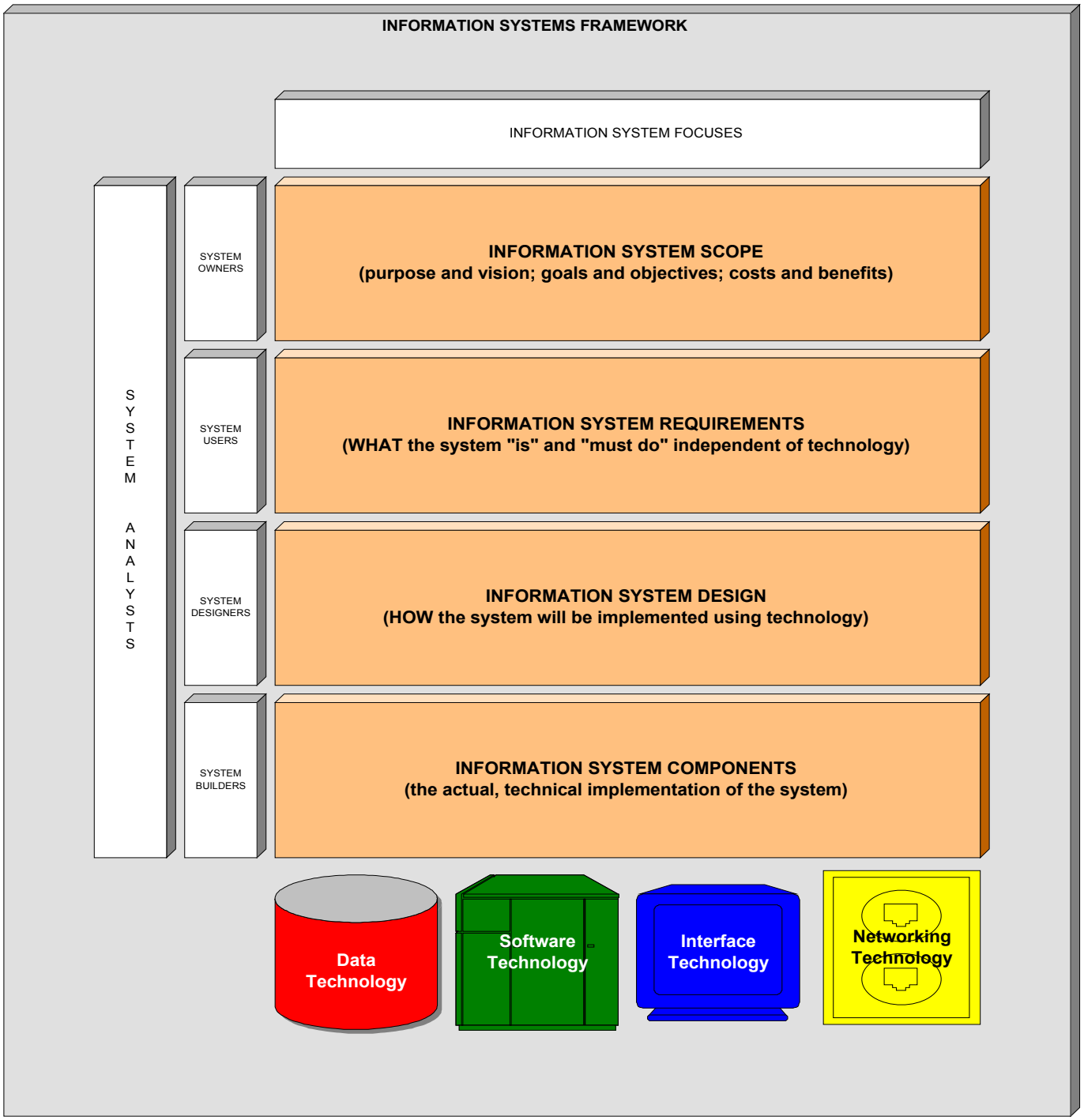
- Concerned with all of the activities related to employees and potential employees of the organization.

Accounting MIS

- Provides aggregated information on accounts payable, accounts receivable, payroll, and other applications.

A Framework For Information Systems Architecture

- **What is an Information Systems Architecture?**
  - An **information systems architecture** provides a unifying framework into which various people with different perspectives can organize and view the fundamental building blocks of information systems.
  - Stakeholders have different views of the system and each has something "at stake" in determining the success of the system.
  - Stakeholders can be broadly classified into four groups:
    - System Owners
    - System Users
    - System Designers
    - System Builders



Information System Perspectives

Perspectives - The People Side of Information Systems

- **What are Information Workers?**
  - The term **information worker** (also called knowledge worker) was coined to describe those people whose jobs involve the creation, collection, processing, distribution, and use of information.
- **System Owners**
  - **System owners** are an information system's sponsors and chief advocates. They are usually responsible for budgeting the money and time to develop, operate, and maintain the information system. They are also ultimately responsible for the system's justification and acceptance.

Perspectives - The People Side of Information Systems

- **System Users**
  - **System users** are the people who use (and directly benefit from) the information system on a regular basis – capturing, validating, entering, responding to, storing, and exchanging data and information.
  - There are many classes of system users including:
    - Internal Users
      - Clerical and service workers
      - Technical and professional staff
        - **Knowledge workers** are a subset of information workers whose responsibilities are based on a specialized body of knowledge.
      - Supervisors, middle managers, and executive managers

Perspectives - The People Side of Information Systems

- **System Users**
  - There are many classes of system users including: (continued)
    - Remote and Mobile Users
    - External Users
- **System Designers**
  - **System designers** translate users' business requirements and constraints into technical solutions. They design the computer files, databases, inputs, outputs, screens, networks, and programs that will meet the system users' requirements. They also integrate the technical solution back into the day-to-day business environment.

Perspectives - The People Side of Information Systems

- **System Builders**
  - **System builders** construct the information system components based upon the design specifications from the system designers. In many cases, the system designer and builder for a component are one and the same.
  - The applications programmer is the classic example of a system builder.
- **The Role of the System Analyst**
  - For the system owners and users, the analyst typically constructs and validates their views.
  - For the system designers and builders, the analyst (at the very least) ensures that the technical views are consistent and compatible with the business views.

Qualities of Information System

Management information systems (MIS) is an organized approach to gathering information from company operations and making a strategic management decision. Developing quality characteristics for gathering information is essential to making solid management decisions.

The main qualities of good management information system are:

1.Relevance

2. Accuracy

3. Timely

4. Exhaustive

5. Cost-Effective

•

Main Qualities of Good Management Information System

1. Relevance: Information should be relevant to the strategic decision that company management is currently reviewing. Because companies may review several business opportunities at one time, avoiding information not relating to the decision is essential.

2. Accuracy: MIS information should be accurate and avoid any probable costs. Making decisions based on estimates can lead to cost overruns or lower profits from future operations.

3. Timely: Many management decisions are based on information from a certain time period, such as quarterly or annual periods. Information outside of the requested time frame may skew information and lead to an improperly informed decision.

4. Exhaustive: MIS information gathering should resemble an upside-down triangle. The early stages of information gathering should be exhaustive, including all types of company information. As management narrows its decision-making process, the information is refined to include only the most relevant pieces.

5. Cost-Effective: The MIS needs to be a cost-effective and efficient system for gathering information. Most of these systems are developed internally, creating costs that cannot be passed to clients.

IS Resources

- The individual **data** being processed through the use of **hardware** and **software** and shared through **network** connection has allowed us to utilize more **information** in less time.

Information Systems Resources

- Networks …connected in some manner that allows to sharing of resources
- Hardware and Peripheral Devices …tangible and can be touched
- Software …intangible and can't be touched physically
- Data …one piece of a record
- People …work together to create usable information

Control Audit And Security Of Information System

Units

1 Controls in Information systems
2 Need and methods of auditing Information systems
3 Testing Information systems
4 Security of Information systems

Learning Goals

- Why controls are necessary in Information systems
- Methods of controlling Information systems
- How controls are introduced in Information systems
- Why Information systems need auditing
- How are systems audited
- The methods used to test Information systems
- How the security of an Information system is ensured

Motivation For Controls

- It is very important to ensure the reliability of reports produced by an information system
- If unreliability is seen by users the entire credibility of the system is lost
- Ensuring reliability is not difficult for small systems but when a system has to handle massive data it is a challenge
- Systematic controls are thus essential when a system is designed

Motivation For Audits

- Many organizations are now entirely dependent on computer based information system
- These information systems contain financial data and other critical procedures
- It is essential to protect the systems against frauds and ensure that sound accounting practices are followed
- It is necessary to trace the origin and fix responsibilities when frauds occur
- Audit methods primary purpose is to ensure this.

Motivation For Testing

- Systems contain many individual subsystems
- Usually sub-systems and programs are individually tested
- However when a whole system is integrated unforeseen errors may be seen
- Thus before releasing a system the entire operational system should be tested for correctness and completeness

Motivation For Security

- Systems contain sensitive data about the organization and also about persons working in the organization
- Sensitive data should be protected from spies, thieves or disgruntled employees.
- Thus access should be carefully controlled and provided only on a need to know basis
- When computers are networked corruption may take place due to viruses
- Services may be disrupted due to denial of service attacks
- Thus systems should be designed with appropriate security measures.

Motivation For Disaster Recovery

- Organizations depend on Information systems for their entire operations
- It is thus essential to ensure continuity of service when unforeseen situations such as disk crashes,fires,floods and such disasters take place.
- Thus it is essential to ensure quick recovery from disasters and ensure continuity of service.

Control Audit and Security of Information System

• **CONTROL**- Method to ensure that a system processes data as per design and that all data is included and are correct

• **AUDIT AND TESTING** - Ensure that the system is built as per specifications and that processed results are correct. Protect systems from frauds.

• **SECURITY**- Protection of data resources, programs, and equipment from illegal use, theft, vandalism, accidents, disasters etc.

Need Of Controls

- Information systems handle massive amounts of data – accidents such as not including some data can cause serious damage
- Incorrect data entry can lead to high monetary losses
- Credibility in the information system may be lost if errors are found in operational systems

Objectives Of Controls

- To make sure data entering the computer are correct
- Check clerical handling of data before it is input to a computer
- Provide means of detecting and tracing errors which occur due to bad data or bad program
- Ensure legal requirements are met
- To guard against frauds

Control Techniques

• **ORGANIZATIONAL MEASURES**

Well defined responsibility for input preparation, delivery output use, operation and maintenance

- Changes in program and data (if any) should be documented
- Performance of task and recording must be by different persons to prevent frauds

Control Techniques

- **INPUT PREPARATION CONTROL**
  - -Sequence numbering
  - -Batch controls
  - -Data entry and verification
  - -Record totals
  - -Self checking digits

Auditing Technology for Information Systems

A. Review of Systems Documentation

B. Test Data

C. Integrated-Test-Facility (ITF) Approach

D. Parallel Simulation

E. Audit Software

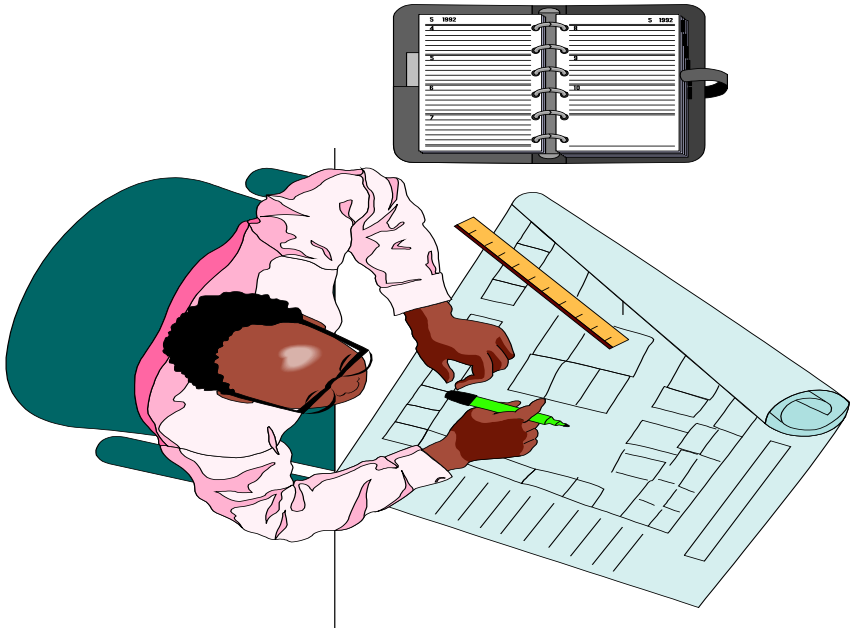
F. Embedded Audit Routines

G. Mapping

H. Extended Records and Snapshots

A. Review of Systems Documentation

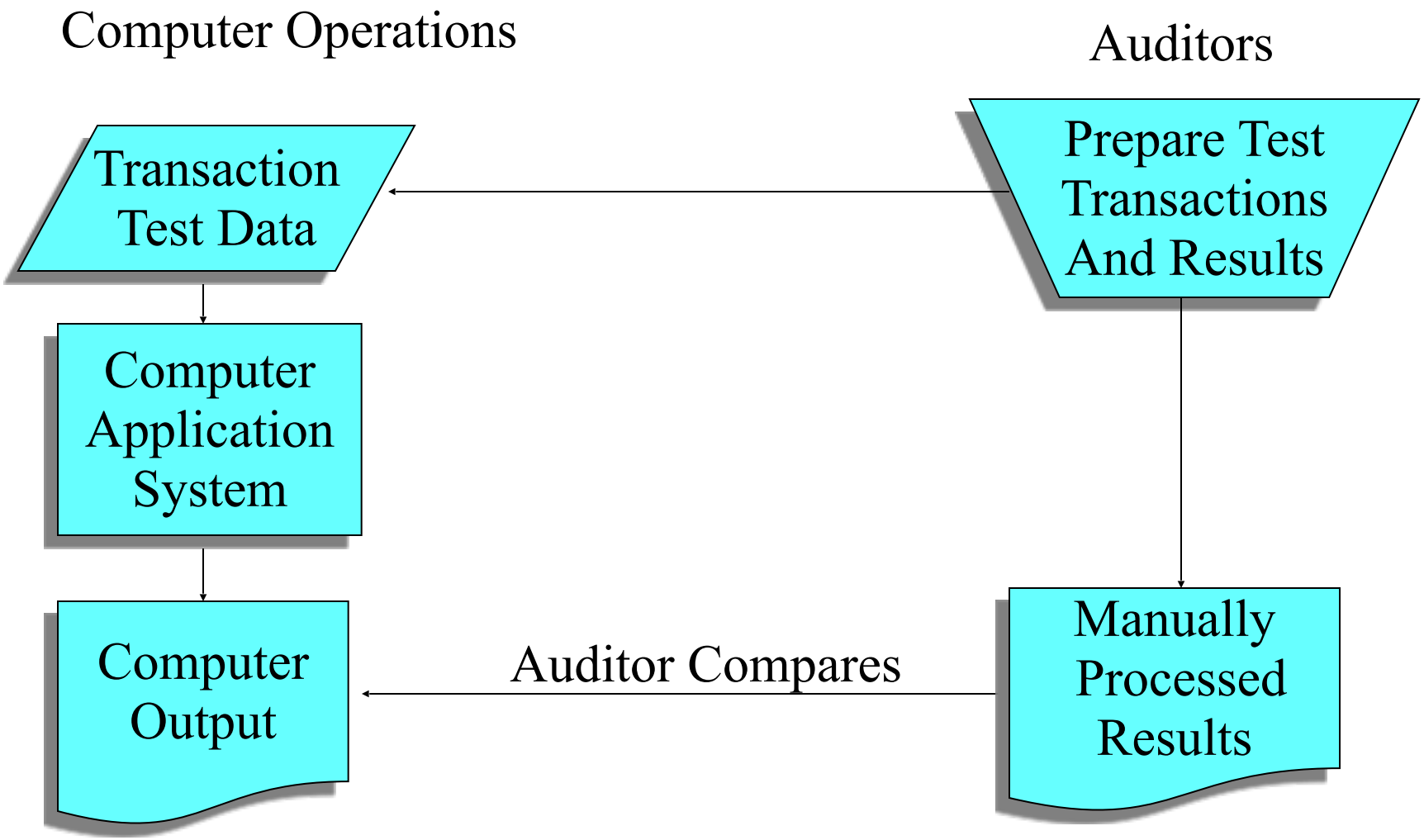
The auditor reviews documentation such as narrative descriptions, flowcharts, and program listings. In desk checking the auditor processes test or real data through the program logic.



B. Test Data

The auditor prepares input containing both valid and invalid data. Prior to processing the test data, the input is manually processed to determine what the output should look like. The auditor then compares the computer-processed output with the manually processed results.

Illustration of Test Data Approach

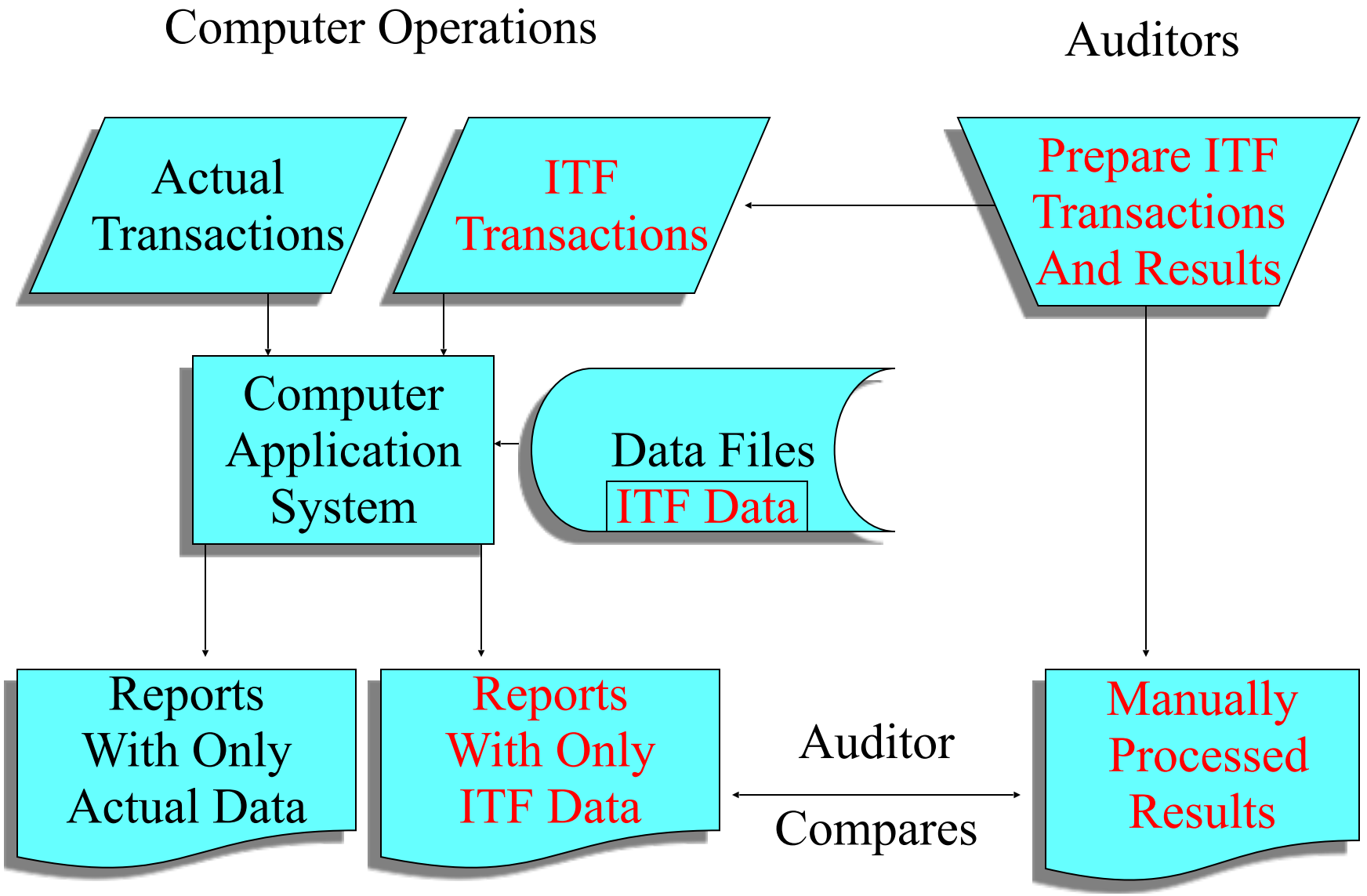


C. Integrated Test Facility (ITF) Approach

A common form of an ITF is as follows:

1. A dummy ITF center is created for the auditors.
2. Auditors create transactions for controls they want to test.
3. Working papers are created to show expected results from manually processed information.
4. Auditor transactions are run with actual transactions.
5. Auditors compare ITF results to working papers.

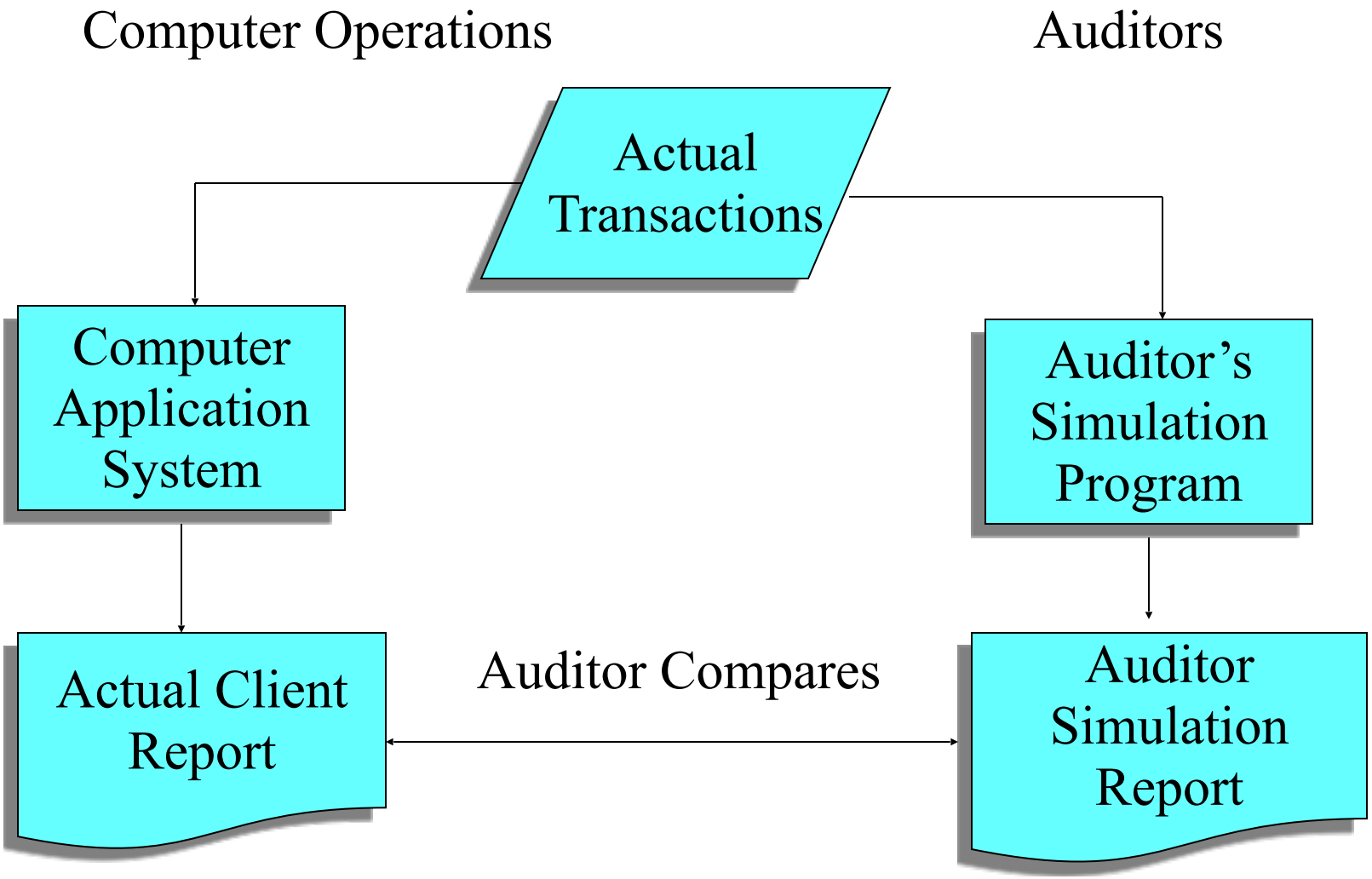
Illustration of ITF Approach



D. Parallel Simulation

The test data and ITF methods both process test data through real programs. With parallel simulation, the auditor processes real client data on an audit program similar to some aspect of the client's program. The auditor compares the results of this processing with the results of the processing done by the client's program.

Illustration of Parallel Simulation



E. Audit Software

Computer programs that permit computers to be used as auditing tools include:

1. Generalized audit software

   Perform tasks such as selecting sample data from file, checking computations, and searching files for unusual items.

2. P.C. Software

   Allows auditors to analyze data from notebook computers in the field.

F. Embedded Audit Routines

1. In-line Code – Application program performs audit data collection while it processes data for normal production purposes.
2. System Control Audit Review File (SCARF)– Edit tests for audit transaction analysis are included in program. Exceptions are written to a file for audit review.

What is Security?

- “The quality or state of being secure—to be free from danger”
- A successful organization should have multiple layers of security in place:
  - Physical security
  - Personal security
  - Operations security
  - Communications security
  - Network security
  - Information security

Types of Security

- Physical Security: To Protect physical items, object or areas
- Personal Security: To protect the individual or group of individuals who are authorized
- Operation Security: To protect the details of a particular operation or activities.

Types of Security

- Communication Security: To protect communication media, technology and content
- Network Security: To protect networking components, connections and contents
- Information Security: To protect information assets

What is Information Security?

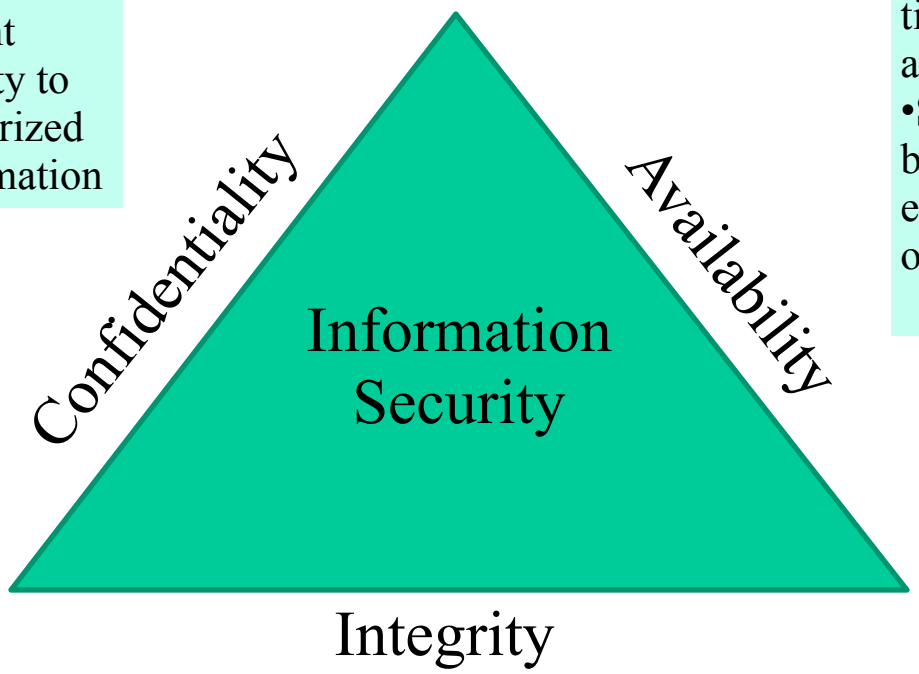
- The protection of information and its critical elements, including systems and hardware that use, store, and transmit that information
- Necessary tools: policy, awareness, training, education, technology
- C.I.A. triangle was standard based on confidentiality, integrity, and availability
- C.I.A. triangle now expanded into list of critical characteristics of information

CIA triangle “Security triad”

- Confidentiality: Making sure that those who should not see information
- Integrity: Making sure that the information hasn’t been changed from it’s original
- Availability: Making sure that the information is available for use when you need it.

Information Security C.I.A triangle

•Data and information is classified into different levels of confidentiality to ensure that only authorized users access the information

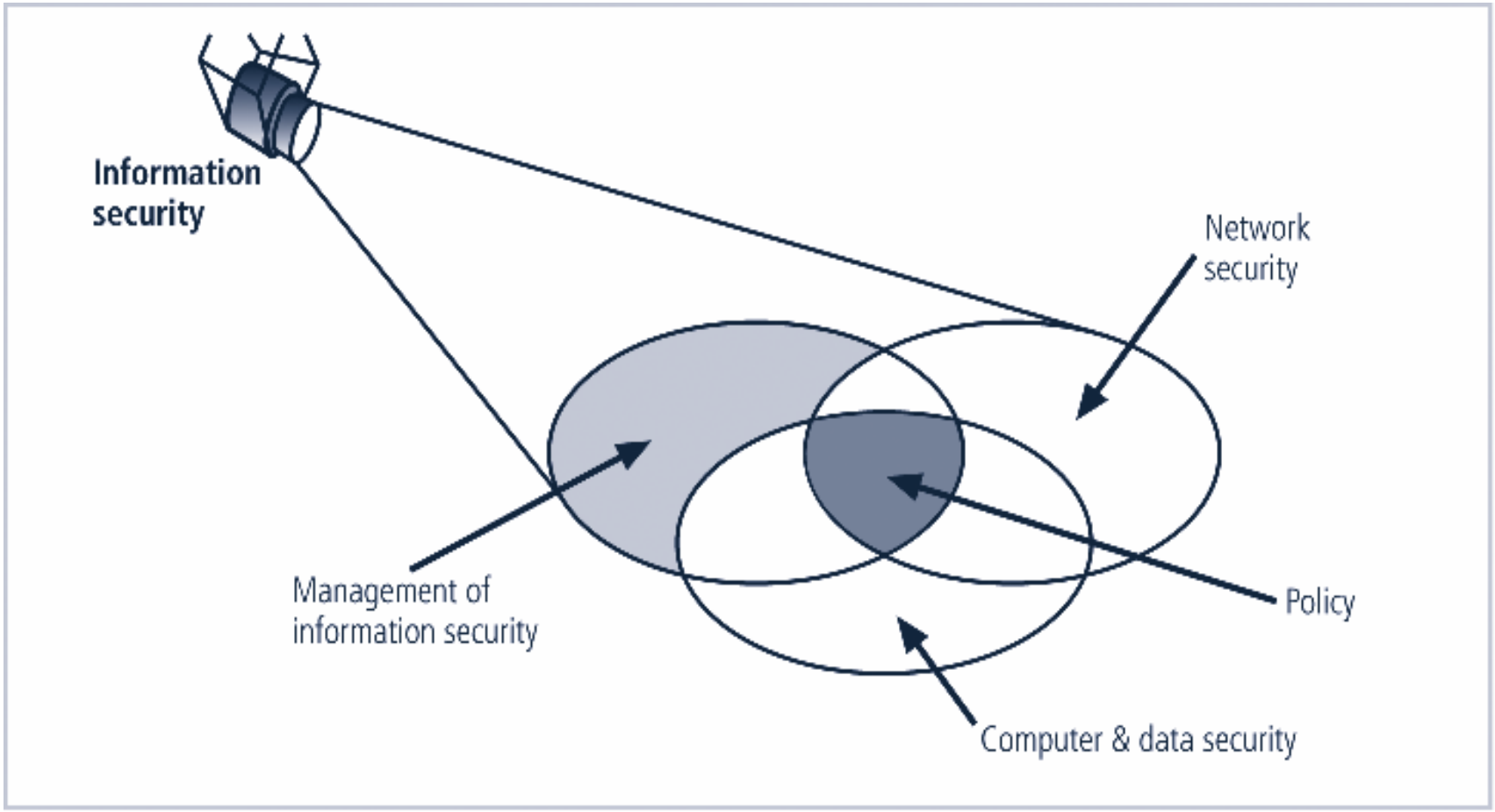


•System is available at all times only for authorizes and authenticated persons.
•System is protected from being shut down due to external or internal threats or attacks

•Data and information is accurate and protected from tampering by unauthorized persons.
•Data and information is consistent and validated.

Critical Characteristics of Information

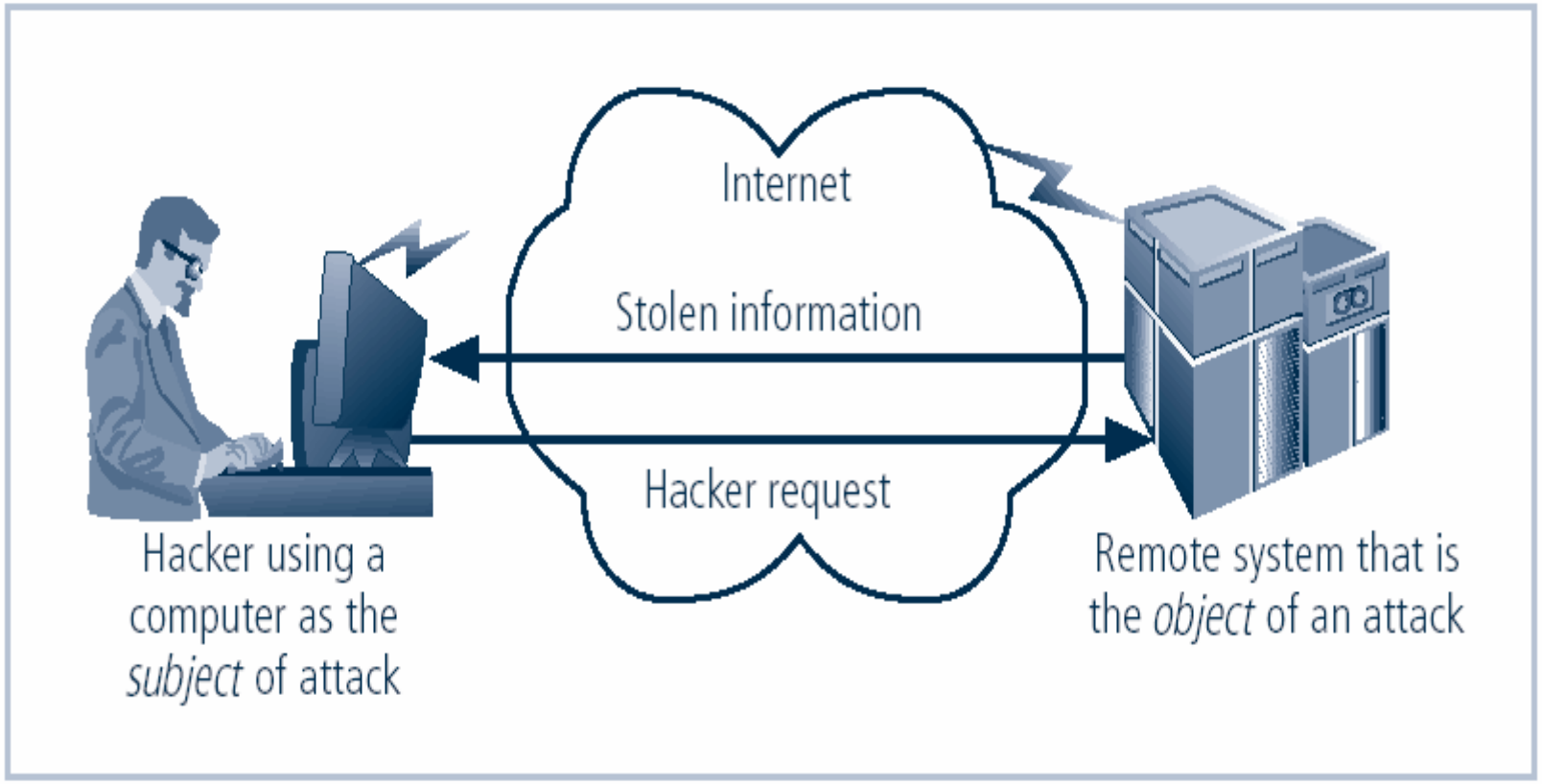
- The value of information comes from the characteristics it possesses:
  - Availability
  - Accuracy
  - Authenticity
  - Confidentiality
  - Integrity
  - Utility
  - Possession



Components of Information Security

Securing Components

- Computer can be subject of an attack and/or the object of an attack
  - When the subject of an attack, computer is used as an active tool to conduct attack
  - When the object of an attack, computer is the entity being attacked



Computer as the Subject and Object of an Attack

A Final Word

- Treat your password like you treat your toothbrush. Never give it to anyone else to use, and change it every few months.

LAYERED SECURITY

- Layered security, in its simplest form, consists of stacking security solutions, one on top of the other, to protect a computer from current, and zero day malware attacks.
- Malware: It refers to software programs designed to damage or do other unwanted actions on a computer system.

  Examples of malware include viruses, worms, trojan horses and spyware.

Why do we need it?

- To providing adequate computer system protection.
- Gaps exist in protection capabilities in even the most sophisticated security applications.

A CONSUMER LAYERED SECURITY APPROACH

- **Backup:** Consider where you would be if your layered security strategy failed. If you've ever lost critical data to a malware infection, no doubt you already consider it of primary importance.
- Free backup utilities are readily available
  - ❖ Hard Drive Cloning is Easy with Free Ease us Disk Copy
  - ❖ Free Drive Image XML- the best way to backup data

Consumer Layered

- **Firewall** – is an application, or a hardware appliance, designed to block unauthorized access to your computer from the Internet, at the same time permitting authorized communications.
- **Antimalware** – A front line antimalware application is absolutely critical to avoid system infection.
- For free antimalware click "Tech Thoughts Top & Free antimalware application"

Consumer Layered

- **Antivirus** – An antivirus application is another critical component in a layered defense strategy to ensure that if a malicious program is detected, it will be stopped dead in its tracks!
- **Web Browser Security** – Install a free Internet Browser add-on such as WOT(Web of Trust). WOT tests web sites you are visiting for spyware, spam, viruses, browser exploits, unreliable online shops, phishing, and online scams, helping you avoid unsafe web sites.

Enterprise Layered Security

- A modern enterprise security strategy uses a layered identity approach as the underpinning of its security.
- All enterprise systems, applications, information systems, facilities, buildings and rooms are assigned as enterprise risk.
- As the user digitally or physically approaches higher risk applications or a physical location the stronger authentication is used.
- As consider the enterprise firewall and the use of Id and passwords for login.

Implementing a Layered Identity Strategy: Enterprise Layered

- This could take the form of digital certificates, security tokens, smart cards and biometrics. It could also take the form of transactional security.
- While the user may successfully use their Id and password, the transaction security software would examine the IP address that the user is coming in from, their geographic position, the time of day, the type of physical computer the user is using and their behavioral pattern.
- If any of these differ from the past, then system alarm bells may start ringing resulting in the user being asked more personal questions, the action being stopped.

Extended Validation (EV) Certificate

- An extended validation (EV) certificate is a data security or anti-fraud measure recommended in 2006 by the Certificate Authority Browser Forum.
- Certificate Authority/Browser Forum (CAB Forum): An open voluntary association of certification authorities and software developers.
- The first version of the *Extended Validation SSL Certificate Guidelines* was ratifed in June 2007.
- The EV identity verification process requires the applicant to prove exclusive rights to use a domain, confirm its legal, operational and physical existence, and prove the entity has authorized the assurance of the Certificate.

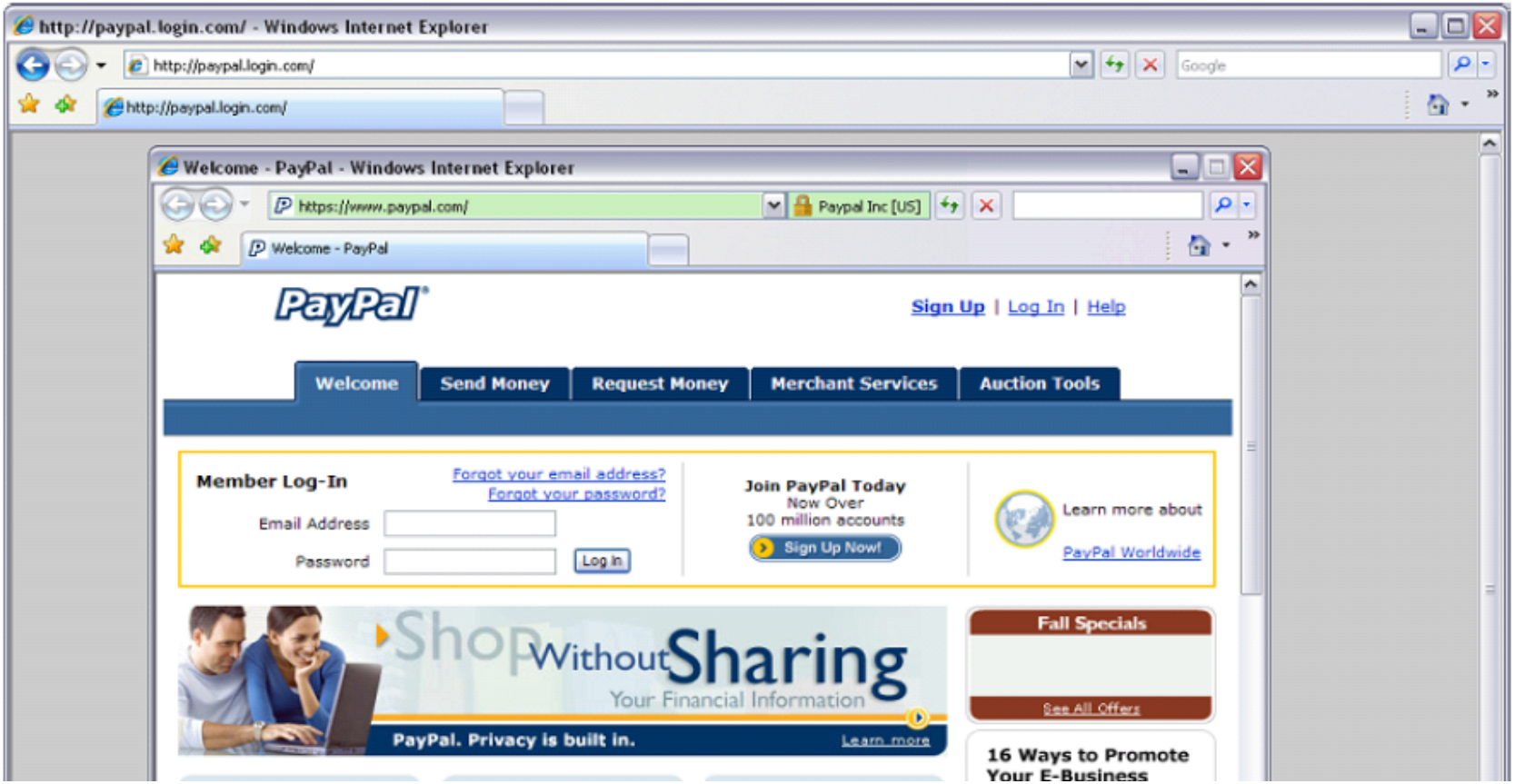
Extended Validation (EV) Certificate

- EV Certificates provide a higher level of validation and are available to all business and government entities, but are not available to individuals.
- The EV process is more rigorous and detailed than for any other Certificate and will require additional steps, which may include obtaining signatures from several people within the applying company, legal verification of the business's existence, etc.

The primary purposes of an EV Certificate

- To identify the legal entity that controls a website which provide a reasonable assurance to the user of an Internet browser that the website which the user is accessing is controlled by a specific legal entity identified in the EV Certificate by name, address of Place of Business, and Registration Number.
- To enable encrypted communications with a website which facilitate the exchange of encryption keys in order to enable the encrypted communication of information over the Internet between the user of an Internet browser and a website.

Designed for Banks and Large E-commerce sites



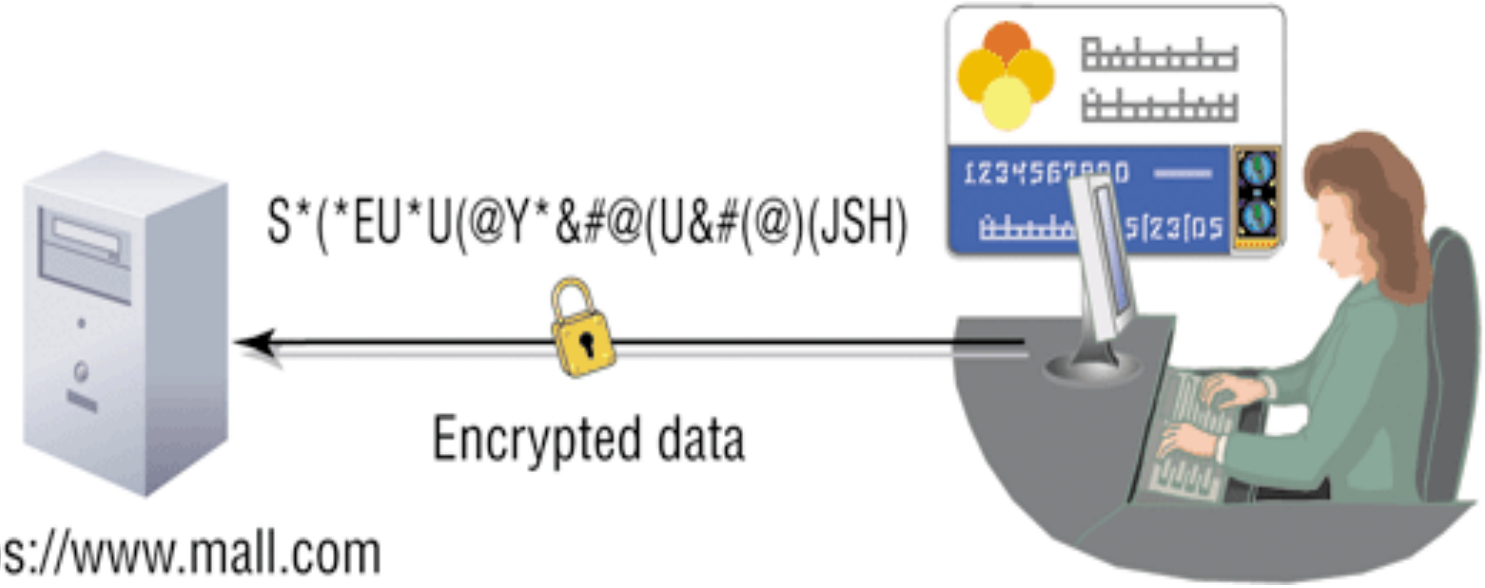
SSL Certificate

Secure Sockets Layer (SSL)

- Digital certificates combined allows for encrypted communications to occur between Web browser and Web server

FIGURE 8.34 • Secure Sockets Layer (SSL)

SSL encrypts data sent over the Web and verifies the identity of the Web server.



What is SSL Certificate?

- SSL (Secure Sockets Layer) is the transaction security protocol used by websites to protect online communications.
- The most common use of SSL is to provide protection for confidential data, such as personal details or credit card information, entered into a website.
- Ecommerce security cannot be an after-thought in your business plans. Today's online shoppers look for the visual cues provided by SSL Certificates, such as the closed padlock and the “https”.

Extended Validation SSL Certificate

- **Show your customers that your site is secure.** Our Extended Validation SSL Certificate features our instant verification green address bar, so your customers can easily see that they're protected. Provide your customers with the highest level of online assurance.
- **When customers see their address bar change to green, they know they can trust your business.** That's because the inspection process for an Extended Validation SSL is more extensive than for any other type of security certificate, verifying your organization's identity, the validity of your request and the overall legitimacy of your business.

- **How can I recognize websites using EV SSL Certificates?**

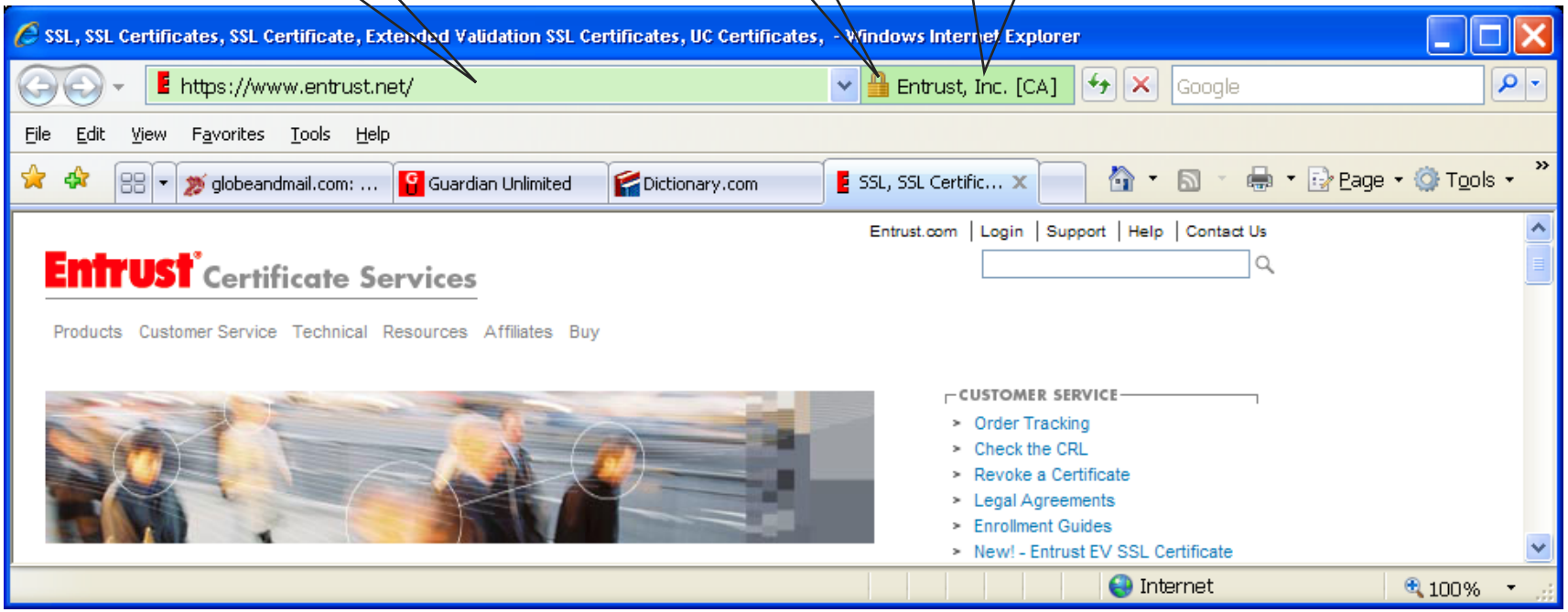
  A website using EV SSL Certificate will activate highly visible indicators directly on the browser address bar:
- **The green address bar, https:// and the padlock icon**
- **The name of the Organization that owns the website and the name of the Certification Authority that issued the EV SSL Certificate.**

Websites using EV SSL Certificates

Assumed name, registered name and country alternating with the issuer's name

Golden padlock

Green address bar



Remote Access Authentication

- Remote access authentication is the process whereby computer users can securely communicate with a network.
- A shared theme to all of these methods is the use of a digital certificate that contains information that identifies the user to a server and provides their credentials.
- Remote access authentication protocols make it safer to conduct business online as well as use ATMs.

RADIUS

- Most modern wireless networks do user authentication using Remote Authentication Dial-In User Service (RADIUS) protocol.
- RADIUS handles the overall authentication process of the user's session on the wireless device as well as also handling the authorization and auditing.
- The RADIUS system takes the (EAP) Extensible Authentication Protocol Authentication Method, challenges the user with the appropriate authentication method, receives the authentication response and then verifies it.

RADIUS

- If the authentication is successful, the RADIUS server will then authorize IP addresses, the tunneling protocol used to create virtual private networks.
- Further, the RADIUS server keeps tracks of when a user session begins and ends.
- For senior executives, who do require quite open access to the applications and information systems via their wireless device.
- Issue them with something like a secureID from (Rivest-Shamir-Adleman)RSA one time password generator and have the executives be required to enter this in order to authenticate their wireless device to the network. RSA algorithm

Policy-based encryption

- The Policy Based Encryption gateway automatically encrypts specific emails based on company-defined policies – that is, a set of rules designed to analyze all email, and encrypt any email that matches the pre-defined conditions.
- The concept of policy-based encryption is a promising paradigm for trust establishment and authorization in large-scale open environments like the Internet and Mobile Networks.
- On policy-based encryption which allow to encrypt a message according to a policy so that only entities fulfilling the policy are able to decrypt the message.

Policy-based encryption

- More generally, policy-based encryption belongs to an emerging family of encryption schemes sharing the ability to integrate encryption with access control structures.
- A policy-based encryption scheme has to fulfill two primary requirements: on one hand, provable security under well defined attack models.
- On the other hand, efficiency, especially when dealing with the conjunctions and disjunctions of credential-based conditions.

E-Commerce

- Electronic commerce
  - Systems that support electronically executed business transactions
  - The fundamental purpose of e-commerce is to execute online transactions
- E-commerce is not new; however, recent rapid development of the Internet is surely responsible for the popularity of e-commerce.
- The new way of commerce through the Internet creates vast opportunities, but at the same time, it poses challenges.

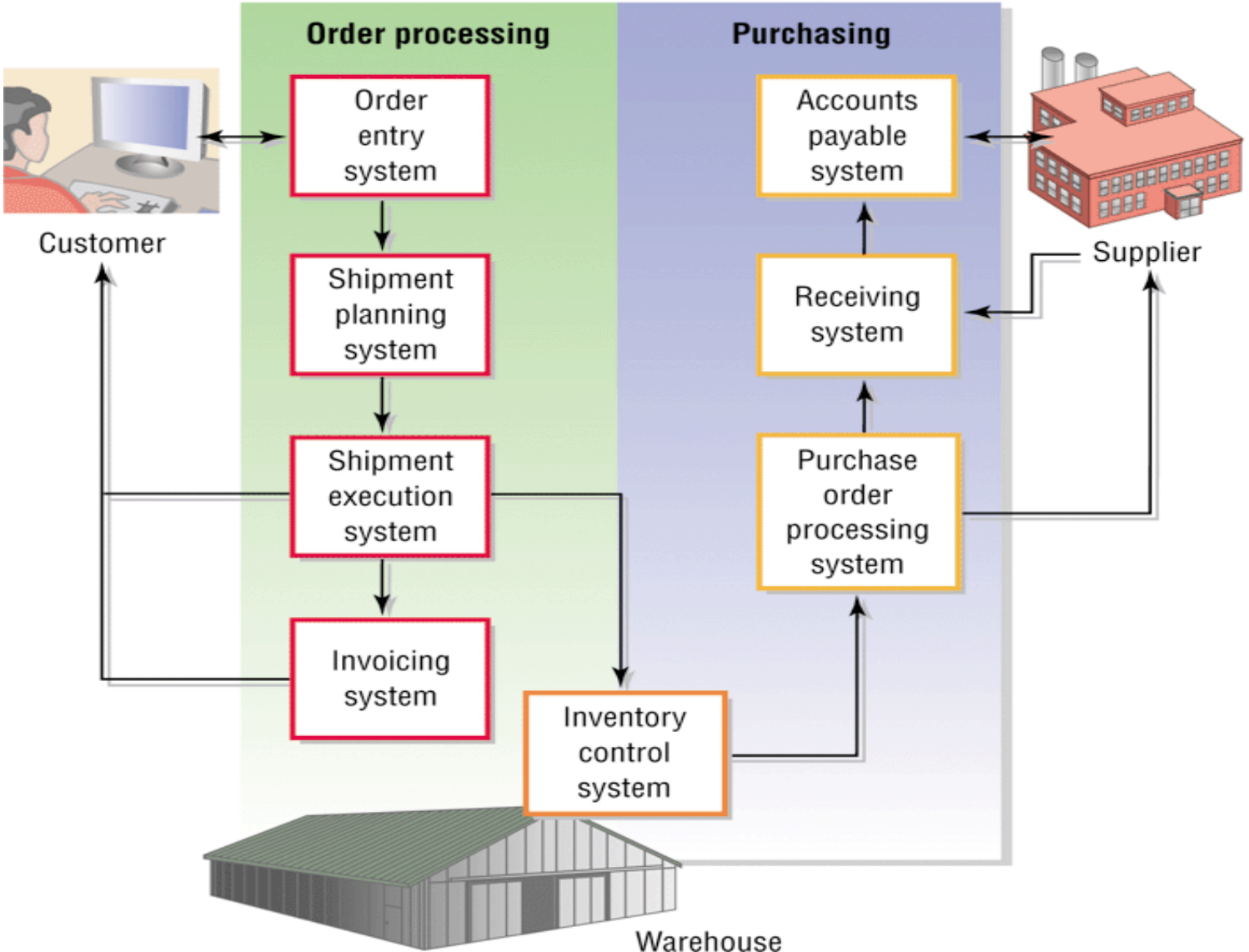
Types of E-Commerce

- Business-to-consumer e-commerce (B2C)
  - Connects individual consumers with sellers , cutting out the middleman
  - E.g. Amazon.com
- Business-to-business e-commerce (B2B)
  - Supports business transactions on across private networks, the Internet, and the Web
- Consumer-to-consumer e-commerce (C2C)
  - Connects individual sellers with people shopping for used items
  - E.g. ebay.com

Different Transaction Processing for Different Needs

FIGURE 8.6 • Transaction processing system interaction

Transaction processing typically makes use of many interconnected systems and subsystems.



Thank You

Chapter 3

Enterprise Management Systems

Enterprise Management System

- EMS is concerned with control, monitoring and the management of IT infrastructure and applications in order to optimize IT service delivery in Company.
- EMS is wide information system designed to coordinate all the resources, information and activities needed to complete business processes.
- Enterprise – an entire company, everything, all-inclusive
- Management – The monitoring and controlling of entities
- Systems – Information Technology Infrastructure, hardware and software, data, information, and processes

Enterprise Software

- Enterprise software is any software used in large organizations (whether business or government).
- It is considered to be an essential part of a computer-based information system, and it provides business-oriented tools such as online payment processing and automated billing systems.
- Enterprise software is also referred to as enterprise application software.

Enterprise Resource Planning

- ERP is business process management software that allows an organization to use a system of integrated applications to manage the business and automate back office functions.
- ERP software integrates all facets of an operation, including product planning, development, manufacturing processes, sales and marketing.
- Some of ERP's functions include:
  - Bookkeeping & Accounting
  - Human Resource Management
  - Planning Production
  - Supply Chain management

ERP Components

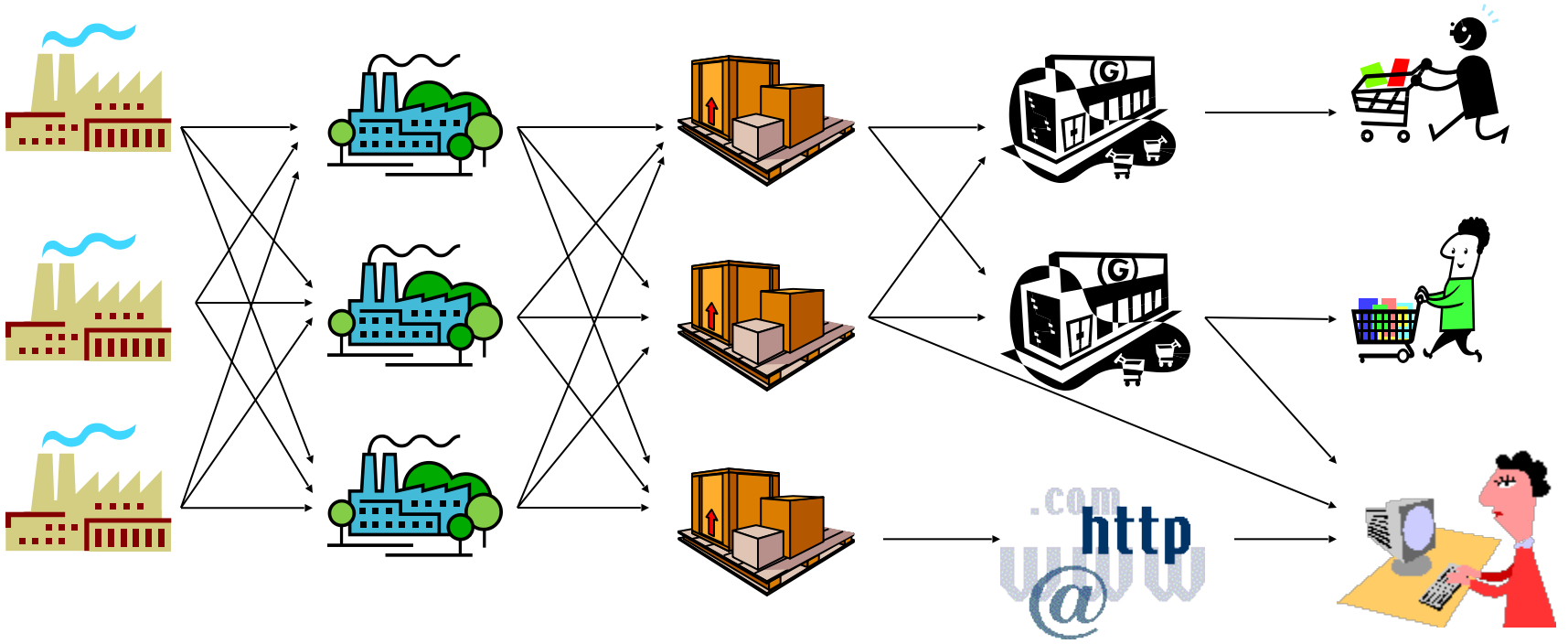
Enterprise Resource Planning Software

- It helps an organization to integrate information flows, operations and processes all resources accessible, for example, materials, work force, machine and money.
- The focus of ERP is on resource management within constraints to maximize the return on investment.
- These data are then stored in a unified database, which are the key for the success of this software solution.
- The ERP package design is built on the principle of Best Practices.
- ERP Software: SAP, Supply Chain Management, CRM

Supply Chain

- A **supply chain** is the system of organizations, people, activities, information and resources involved in moving a product or service from supplier to customer.
- Supply chain activities transform raw materials and components into a finished product that is delivered to the end customer.

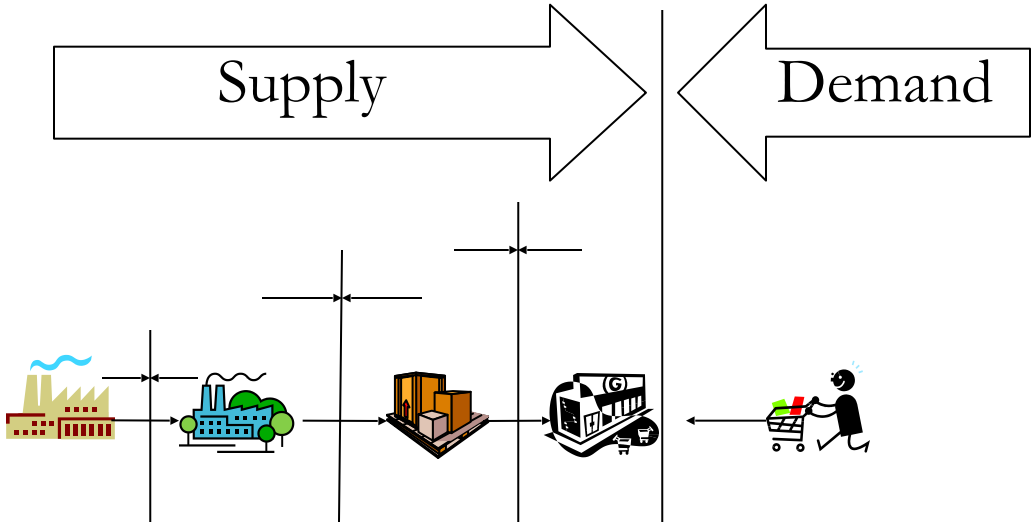
Supply Chain



Supplier → Manufacturer → Distributor → Retailer → Customers

Supply Chain Management

Supply Chain Management is
the design and management of processes
across organizational boundaries
with the goal of matching supply and demand
in the most cost effective way.

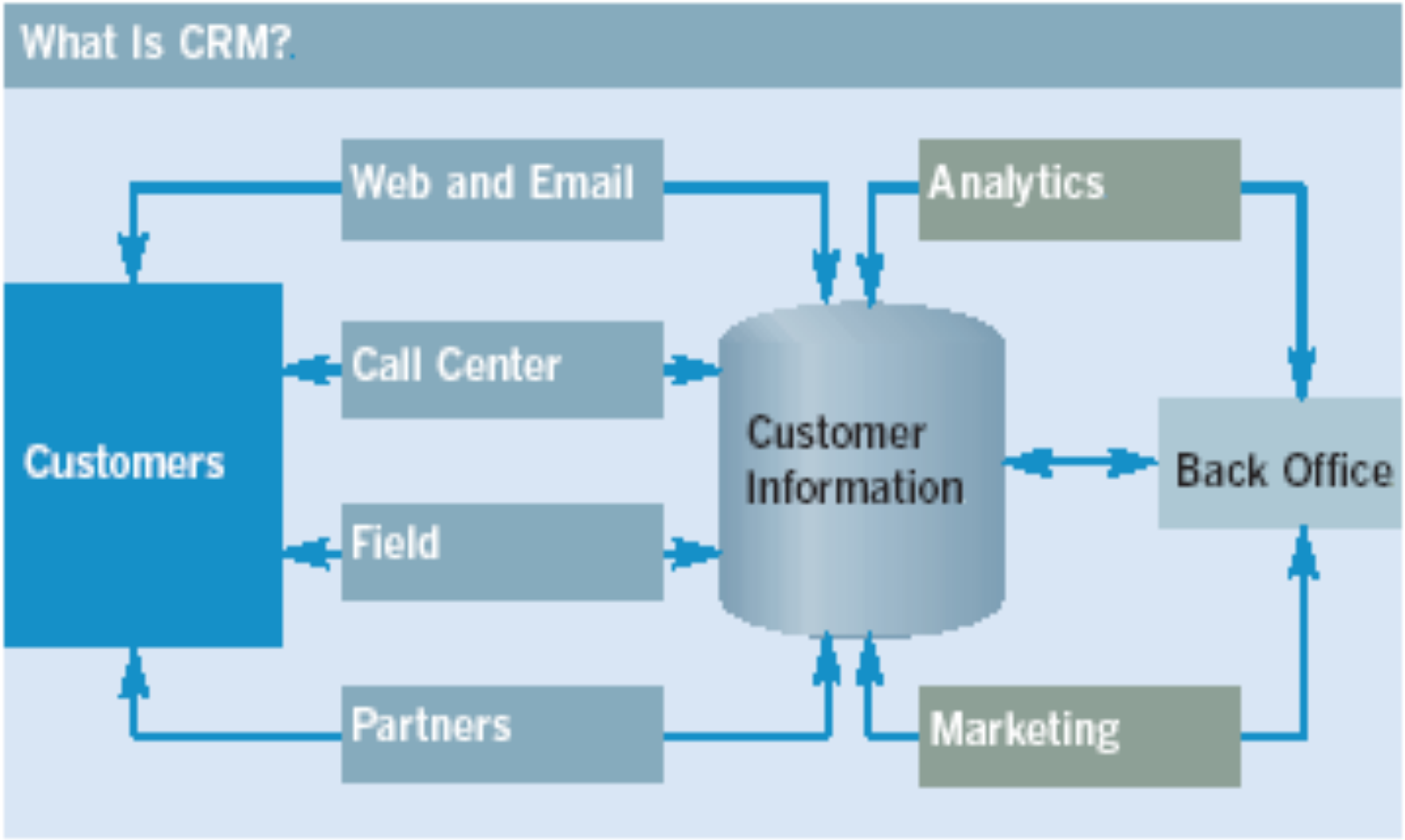


Matching Supply and Demand

Customer Relationship Management

- Customer Relationship Management is a strategy for managing all your company's interactions with current and prospective customers.
- CRM formation of *bonds* between a company and its customers.
- CRM enables your company to increase productivity, close more business, and improve customer satisfaction and retention.

Model of Customer Relationship Management



CRM Strategies

Customer Retention Marketing Techniques

- Customization: Changing the product (not just the marketing message) according to user preferences
- Customer co-production: Allows the customer to interactively create the product
- Customer service tools include:
  - Frequently asked questions
  - Real-time customer service chat systems
  - Automated response systems

Benefits of using CRM

- Centralized customer interaction
- Improved customer support and satisfaction
- High rate of customer retention
- Increase revenue and referrals from existing customers
- Improve your products/services
- Measure and optimize your performance
- Boost new business

Enterprise Information Management

- Enterprise information management (EIM) is a set of business processes, disciplines and practices used to manage the information created from an organization's data.
- EIM initiatives seek to build efficient and agile data management operations with capabilities for information creation, capture, distribution and consumption.
- The goal is to provide and preserve information as a business asset that remains secure, easily accessible, meaningful, accurate and timely.

Enterprise IT Management

- **EITM** is a strategy conceived and developed by Computer Associates International which details how organizations can transform the management of IT in order to maximize business value.
- Strategy for increasing the business relevance of the IT function, EITM considers the need for IT organizations to start operating as a service-based business.
- Ensuring investments are prioritized according to business strategy and that operational efficiencies can be more quickly realized and costs reduced when IT processes are integrated and automated.

Role of IT in Enterprise Management

- Enterprise IT Management was developed in response to a growing need by IT organizations to gain more value from investments made in IT capabilities, infrastructure and resources.
- EITM proposes a set of capabilities that enable IT to better govern, manage and secure the IT services delivered to the business.
- IT/IS as asset, “strategic weapon”, “nervous system” (strategic level) Vs tool, commodity (operational level)

Enterprise Information Systems

- Enterprise information system (EIS) is a system that serves an entire enterprise or at least two functional departments in:
  - Business intelligence (BI)
  - Enterprise resource planning (ERP)
  - Knowledge management (KM)
  - Partner relationship management (PLM)
  - Business process management (BPM)
  - Customer relationship management (CRM)

Role of IS in Enterprise Management

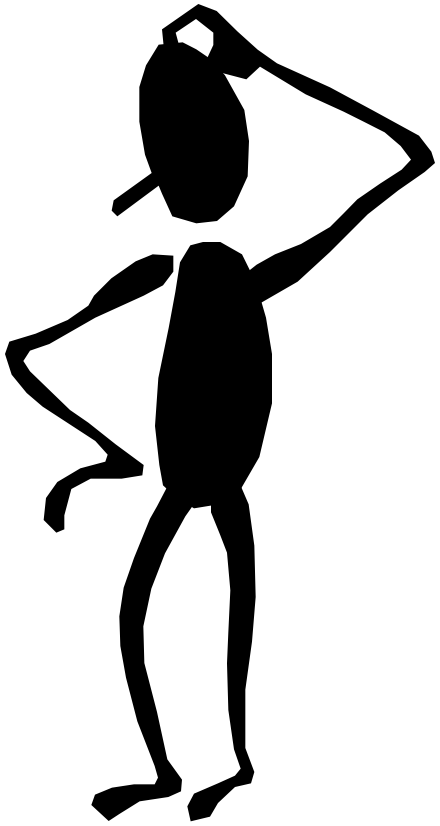
- Help to unify the firm's structure and organization: One organization
- Management: Firm wide knowledge-based management processes
- Technology: Unified platform
- Business: More efficient operations & customer-driven business processes
- Supporting the major business functions: sales and marketing, manufacturing and production, finance and accounting, and human resources

Role of IS and IT in Enterprise Management

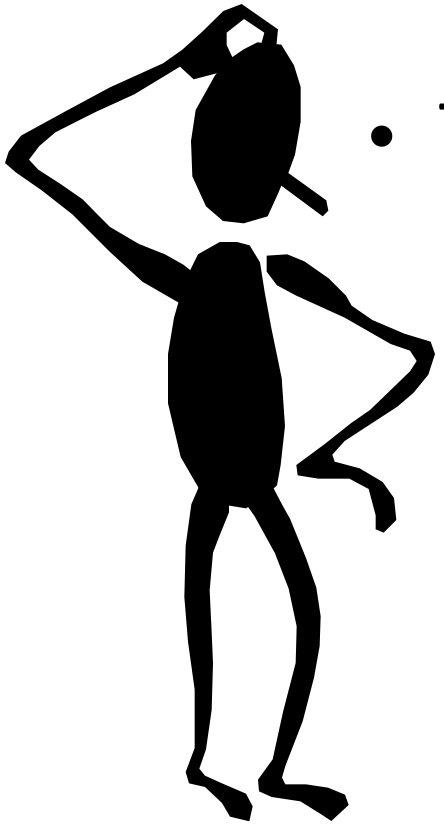
- Reduce Costs/ Improve Productivity
- Improve Customer Satisfaction/ Loyalty
- Create Competitive Advantage
- Generate growth
- Streamline Supply Chain
- Global Expansion

What do Enterprise Engineers do?

- Identify and Integrate best and most successful ways to change an enterprise



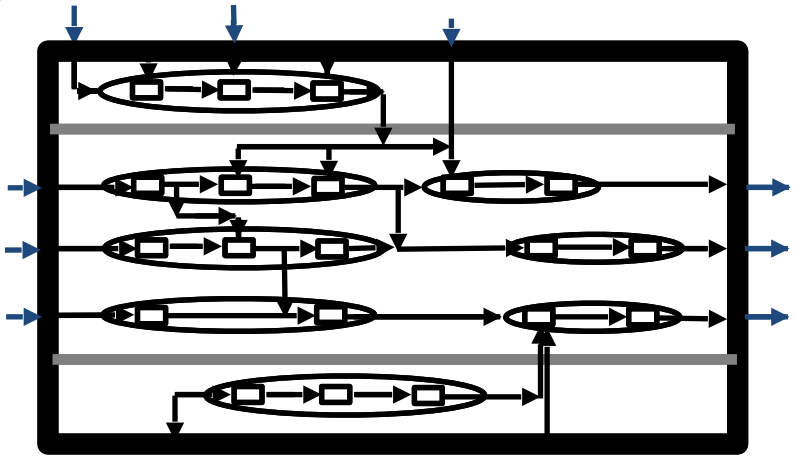
What do Enterprise Engineers do?



- Two aspects
  - Understand new mechanisms
    - New ways of organizing work
    - New Corporate Architectures must be understood
  - Understand methods that can change an enterprise

Two questions Enterprise Engineers always ask

- What should the enterprise be?
- How do we get there from here?



Enterprise Engineering

- Enterprise Engineering is integrated set of disciplines for building or changing an enterprise, its processes, and systems.
- It integrates the most powerful change methods and makes them succeed.
- The goal is a human-technological partnership of maximum efficiency in which learning takes place at every level.

Goal of the Enterprise Engineer

- Identify and integrate the most valuable and successful ways to change an enterprise, and to take them into a professional discipline with a teachable methodology and measures of effectiveness.

Need for Enterprise Integration

- Integration of markets
- Integration between several development and manufacturing sites
- Integration between suppliers and manufacturers
- Integration of design and manufacturing
- Integration of multi-vendor hardware and software components

Basic principles for integration

- Provide the vision, right information, resources, and responsibility
- Empowered people
- A comprehensive and effective communication networks
- Democratization and dissemination of information
- Freely shared information

Two major issues of Enterprise Integration

- How to motivate employee
- How to provide employee with the right information to do their jobs

Types of Integration

- Loose Integration versus Full Integration
- Horizontal Integration versus Vertical Integration
- Intra-Enterprise Integration versus Inter-enterprise Integration
- System Integration, Application Integration, and Business Integration

Loose Integration versus Full Integration

- Loose Integration - If two systems can merely exchange information with one another with no guarantee that they will interpret this information the same way

- Full integration - Two systems are fully integrated if and only if
  - the specificities of any one of these systems are only known to the system itself and not by the other one,
  - the two systems both contribute to a common task, and
  - the two systems share the same definition of each concept they exchange

Horizontal Integration versus Vertical Integration

- Horizontal Integration - concerning physical and logical integration of business processes from product demand to product shipment, regardless of the organizational boundaries.

  Concerning the technological flow
- Vertical Integration - concerns integration between the various management levels of the enterprise, i.e. decision-making integration, where a management level defines the set of constraints for its lower management levels, which in turn send feedback information to their upper management level, and so on.

  Concerning the decision flow

Intra-Enterprise Integration versus Inter-Enterprise Integration

Intra-Enterprise Integration - the integration of the business processes internal to a given enterprise. (Full integration)

Inter-Enterprise Integration - the integration of business processes of a given enterprise with business processes of other enterprises, or even sharing some parts of business processes by different cooperative enterprises. (Loose integration)

System Integration, Application Integration, and Business Integration

- Physical System Integration - concerning System communication
- Application Integration - concerning Interoperability of applications
- Business integration - concerning Business process coordination

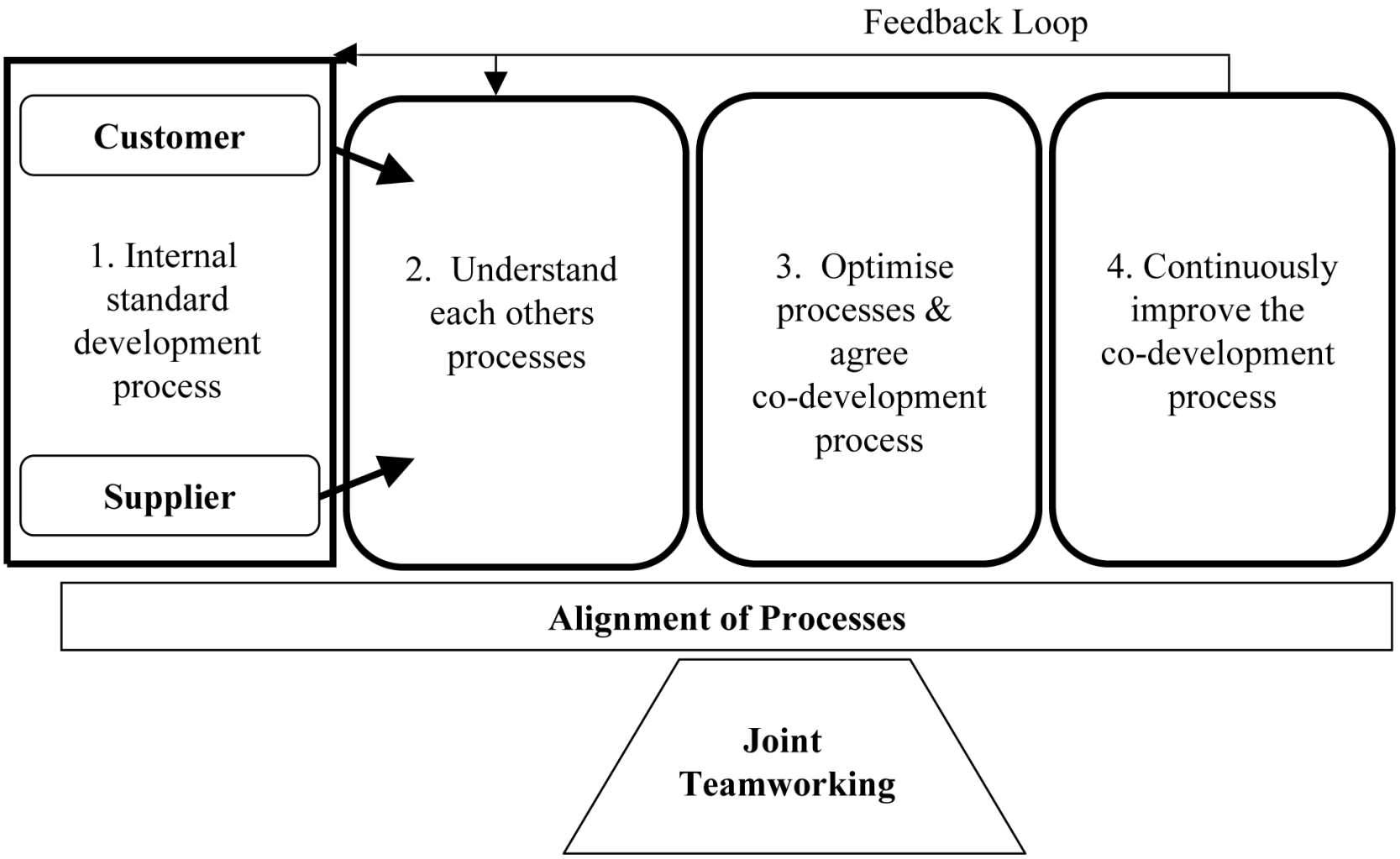
Alignment Process

- Developing a common understanding among the key stakeholders of the purpose and goals of the project and the means and methods of accomplishing those goals is called the **Alignment Process**.
- It is important to accomplish this alignment during the initiation phase.
- Project managers usually conduct a start-up meeting that is sometimes called a kickoff meeting.

Alignment Process (Continued)

- The agenda and duration of the start-up meeting depends on the complexity level of the project.
- Projects with a limited scope and short duration may engage in a session start-up meeting over lunch.
- A medium-complexity project will require more-hour meeting while a high-complexity project cannot achieve alignment in a single meeting. Alignment can require several days of activities.

Alignment Process



Objective of Alignment Process

- The purpose of the alignment process is to develop a common understanding of the purpose, agree on the means and methods, and establish trust.
- The components of the alignment process are discussions of the purpose, goals, participant roles, methods of tracking progress and costs, methods of managing change, and building trust.
- The effects of a lack of trust are delays caused by fact checking or missing information that was not shared because the person's discretion was not trusted to handle sensitive information.

Final Words

"Doing your best is not enough."

W. Edwards Deming

You must know what to do, how to do it and be willing to pay the price to do it.

Decision Support and Intelligent Systems

Decision Making

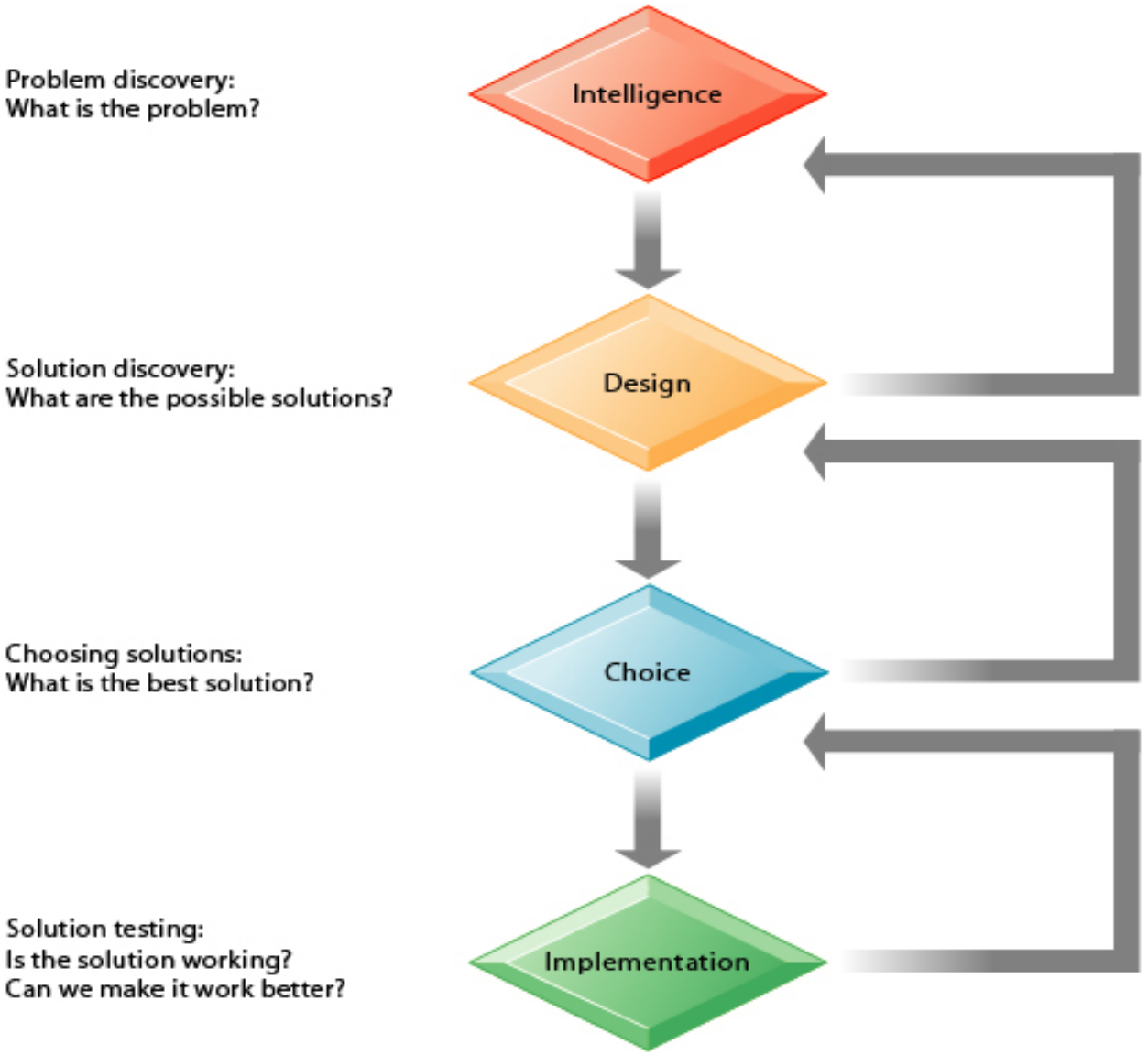
- Information is used to make decisions. Decision making is not a single activity that takes place all at one.
- The process consists of several different activities that take place at different times.
- The decision maker has to identify and understand problems.
- Once perceived, solutions must be designed; once solutions are designed, choices have to be made about a particular solution; finally, the solution has to be carried out and implemented.

  Four different stages in decision making are
  - Intelligence
  - Design
  - Choice
  - Implement

Decision Support System

- The design support system basically helps the information system in the intelligence phase to identify the problem and then go to the design phase for solution.
- The choice of selection criteria varies from problem to problem.
- It is required to go through these phase again and again till the satisfactory solution is found.

Decision Making Stages



Decision Support System

- A **decision support system** is an integrated set of computer tool that allows a decision maker to interact directly with computers to create information and it useful in making semi-structured and unstructured decisions.
- The software components for decision-support systems are a language system which enables the user to interact with the decision-support system, a problem-processing system.
- A properly designed DSS is an interactive software-based system planned to help decision makers and to identify and solve problems and make decisions.

Continued

- DSS serve the management, operations, and planning levels of an organization and help to make decisions, which may be rapidly changing and not easily specified in advance.
- DSS are used to collect data, analyze and shape the data that is collected, and make sound decisions or construct strategies from analysis whether computers, databases, or people are involved generally it does not matter.
- The nature of the decision is such that the decision makers need a variety of information. The reason for changing the demands is also because the methods of decision making a change from time to time.

DSS Types of Decisions

- Structured / Programmed Decisions :
  - Schedule decisions
  - Organization develops specific process for handling
  - Rules of decision making system are predetermined

DSS Types of Decisions

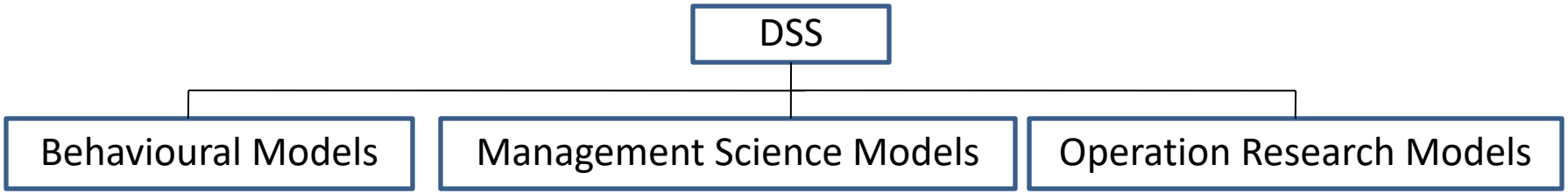
- Unstructured / Non-programmed Decisions :
  - Repetitive decisions
  - Handled by general problem solving process
  - Decision taken by Decision Support Systems
  - Rules of decision making system are not fixed or predetermined
  - It requires every time the user has to go through the decision-making cycle.
  - Decision support systems can be built in case of programmable decision situation.

DSS Types of Decisions

- Semistructured Decisions:
  - Some decision procedures can be specified in advance, but not enough to lead to a definite recommended decision

Decision Support System Models

- Types of Decision Support System Models :
  - Behavioural Models
  - Management Science Models
  - Operation Research Models



Types of Models

Behavioural Models

- Behavioural Models :
  - The decision maker can make the decisions for such behavioral relationships.

    For eg: The trend (development) analysis, forecasting and statistical analysis models.
  - The trend analysis indicates how different variables behave in trend setting in the past and hence in future.

    - In Market Research method, they can forecast or judge the behavior of the customers buying decisions. (i.e. The questionnaire are designed and computerized to evaluate customer's buying behavioral).

Management Science Models

- Management Science Models :
  - These models are developed on the principles of the business management, accounting and economics.

    For eg: the budgetary systems, cost accounting system, inventory management system.
  - In the budgetary system, budgets are used for planning and control.
  - In all the organization, budgets are prepared with the use of graphical representation in the form of line charts or bar charts.
  - For eg : Sales Budget, Production Budget

Operation Research Models

- **Operation Research Models :**
  - The Operation Research models are the mathematical models in providing guidelines to managers for making effective decisions within the state of the current information, or in seeking further information if current knowledge is insufficient to reach a proper decision.

  Some applications

  - Telecommunications/Road/Rail Network Design
  - Organization Supply Chain Strategy
  - Just-in-Time Manufacturing Planning
  - Retail Shop floor Layout
  - Revenue, Pricing and Promotions
  - Demand Forecasting
  - Project Planning

  For eg : Linear Programming is mathematical modeling technique useful for guiding quantitative decisions in business planning, industrial engineering, and—to a lesser extent in the social and physical sciences.

Benefits of Decision Support System

- **Benefits of Decision Support System :**
  - Ability to view data and sensing the problem through the different view.
  - Ability to understand and evaluate the business performance.
  - Ability to understand the problem and its result, and ability to judge the impact on business.
  - Ability to evaluate the impact of any change in the business performance and enabling to focus on the areas where impact is negative.
  - Ability to view the complex scenario or problem and to analyze it and develop alternatives to solve the problem.
  - Ability to make a better decisions due to quick analysis.
  - Ability to control the risk exposure in decisions.

Group Decision Support System

- Group Decision Support System
  - Information technology supports decision-making where there is a group participation. Such decision support system is called as Group Decision Support Systems (GDSS).
  - GDSS has come to mean computer software and hardware used to support group functions and processes.
  - GDSS supports Manager and Staff working in groups.

Group Decision Support System

- There are four configurations of group members are possible.
    - Group members in one room operating on network with common display screen to share the display for all members.
    - Group members sit at their respective locations and use their desktop to interact with other members.
    - Group members are in different cities and they come together through teleconferencing or video conferencing with prior planning GDSS operations.
    - Group members are at remote locations may be in different countries and they come together through long distance telecommunication network.

GDSS Time/Place Environment

Same-Time **Same-Place** (Most widely used GDSS-computers with projectors, voting tools)	**Same-Time** **Different-Place** (team room, tools, audio conferencing, screen sharing, chat)
Different-Time **Same-Place** (audio/video conferencing, document sharing)	**Different-Time** **Different-Place** (voice mail, email, bulletin boards)

WHY USE GDSS?

- High level managers can spend 80% of their time making decisions in groups. Applied correctly, GDSS can reduce this time, arriving at a better decision faster.
- GDSS provides the hardware, software, databases and procedures for effective decision making.

Advantages of GDSS

- Parallel Communication – eliminate monopolizing, providing increased participation, better decisions
- Automated record keeping – no need to take notes, they're automatically recorded
- Ability for virtual meetings – only need hardware, software and people connected
- Portability - Can be set up to be portable…laptop
- Global Potential - People can be connected across the world.

Disadvantages of GDSS

- Cost – infrastructure costs to provide the hardware and software/room/network connectivity can be very expensive
- Security – especially true when companies rent the facilities for GDSS; also, the facilitator may be a lower level employee who may leak information to peers
- Technical Failure – power loss, loss of connectivity, relies heavily on bandwidth and LAN/WAN infrastructure properly setup system should minimize this risk

Enterprise Decision Support System

- The Enterprise Decision Support System is a multi-disciplinary organization that provides technical and program/project leadership for Departmental Information Technology, applications and systems.
- Enterprise use decision support systems to quickly perform complex analyses over large amounts of data whose results are used to inform critical business decisions.

Enterprise and Decision Making

Executive Decision Support System

- **Executive Decision Support System** is a computer-based technology designed specifically for the information needs of top executives level and provides for:
  - Rapid access to timely information;
  - Direct access to management reports;
  - Very user friendly and supported by graphics.
- **Executive Decision Support Systems** support the informational roles of executives.

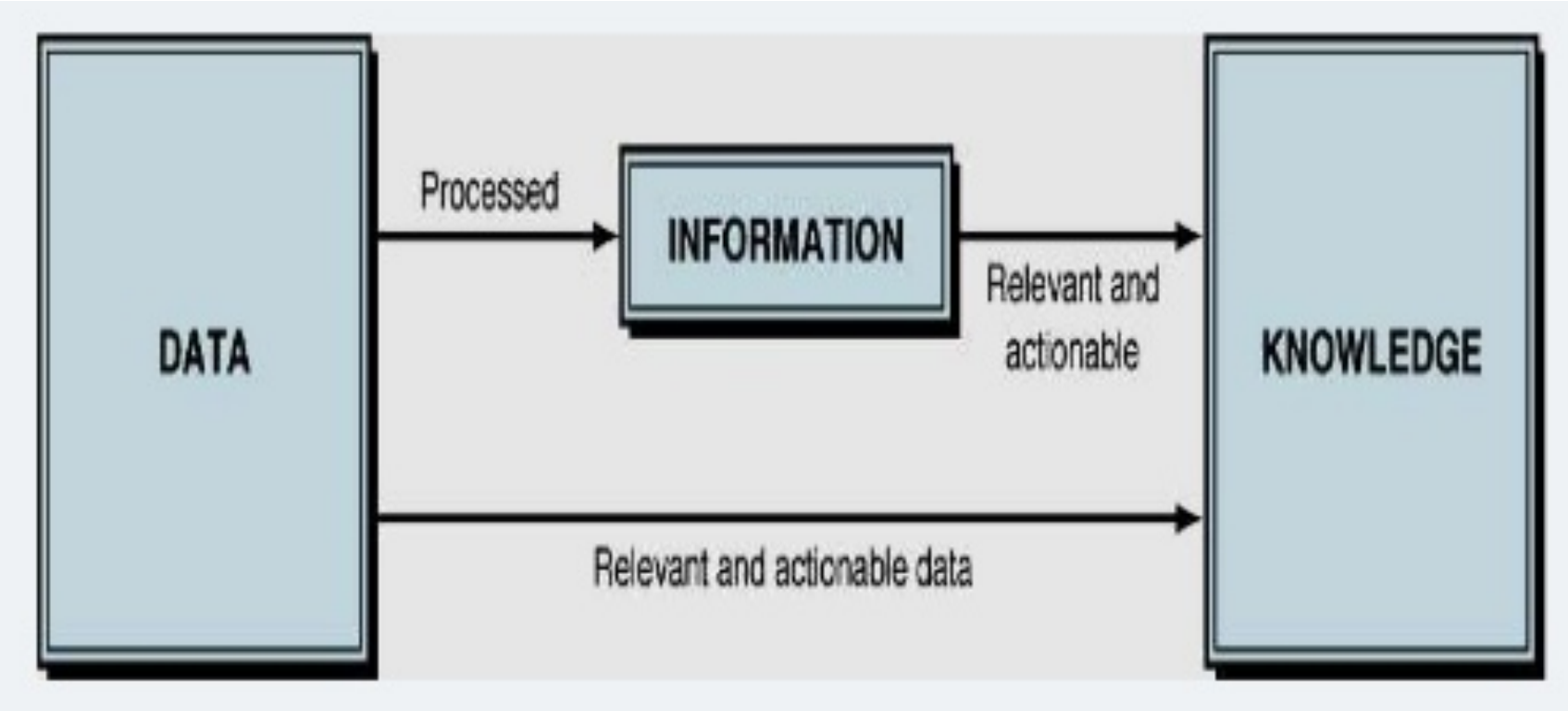
ESS(Continued)

- **Executive Support Systems** – specialized decision support systems designed to meet the needs of senior management.
- **Exception reporting** – reporting of only the results that deviate from a set of standards.
- **Drill down reporting** – investigating information in increasing detail.
- Include analysis support, communications, office automation and intelligence support.

Knowledge

- **Knowledge** is something that comes from information processed by using data.
- Knowledge is applied by knowledge workers who are involved in a particular job or task.
- People use their knowledge in making decisions as well as many other actions
- Knowledge is information that is contextual, relevant, and actionable.

Data, Information and Knowledge



Data, Information and Knowledge

- Data

  refers to isolated facts such as individual measurement No meaning on their own
- Information

  fact about situation, person, events
- Knowledge

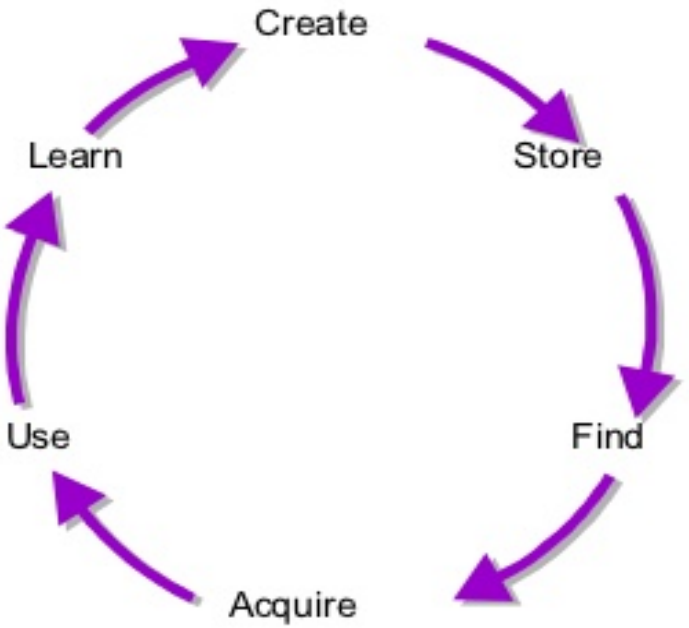
  experience

Knowledge Management

- **Knowledge Management** (KM) comprises a range of strategies and practices used in an organization to identify, create, represent, distribute, and enable adoption of insights and experiences
- KM is a process that helps organizations identify, select, organize, disseminate, and transfer important information and expertise that are part of the organization's memory.
- KM is the process of systematically and actively managing and leveraging stores of knowledge in an organization

Knowledge Management System

- Knowledge Management System refers to a (generally IT based) system for managing knowledge in organizations for supporting creation, capture, storage and distribution of information.



Knowledge Life Cycle

Knowledge Management System

- Knowledge Management System
  - **Create :** Knowledge must be created either within or outside the organization. This is typically comprised of iterative tacit and explicit loops until the knowledge is ready for distribution for those outside the creating group.
  - **Store :** Knowledge can be stored somewhere, either tacitly or explicitly so that it is accessible for others to find and use.
  - **Find :** Those who need the specific knowledge must then find out where it is, when they need it, by searching in the right places and asking the right people.

Knowledge Management System

- **Acquire :** Once the knowledge source is found, the user will then go through the act of actually acquiring it. This will involve gaining personal knowledge from other humans or documented sources.
- **Use :** Once acquired, the knowledge can be put to use towards some productive purpose.
- **Learn :** Having been used, perhaps repeatedly, the user will learn what worked well and not so well as a result of applying the knowledge gained. This can then be taken as significant input into further iterations of the knowledge creation and distribution process.

Knowledge Management System

- IT that helps gather, organize, and share business knowledge within an organization
- Hypermedia databases that store and disseminate business knowledge. It may also be called **knowledge bases.**
- Best practices, policies, business solutions entered through the enterprise knowledge portal.

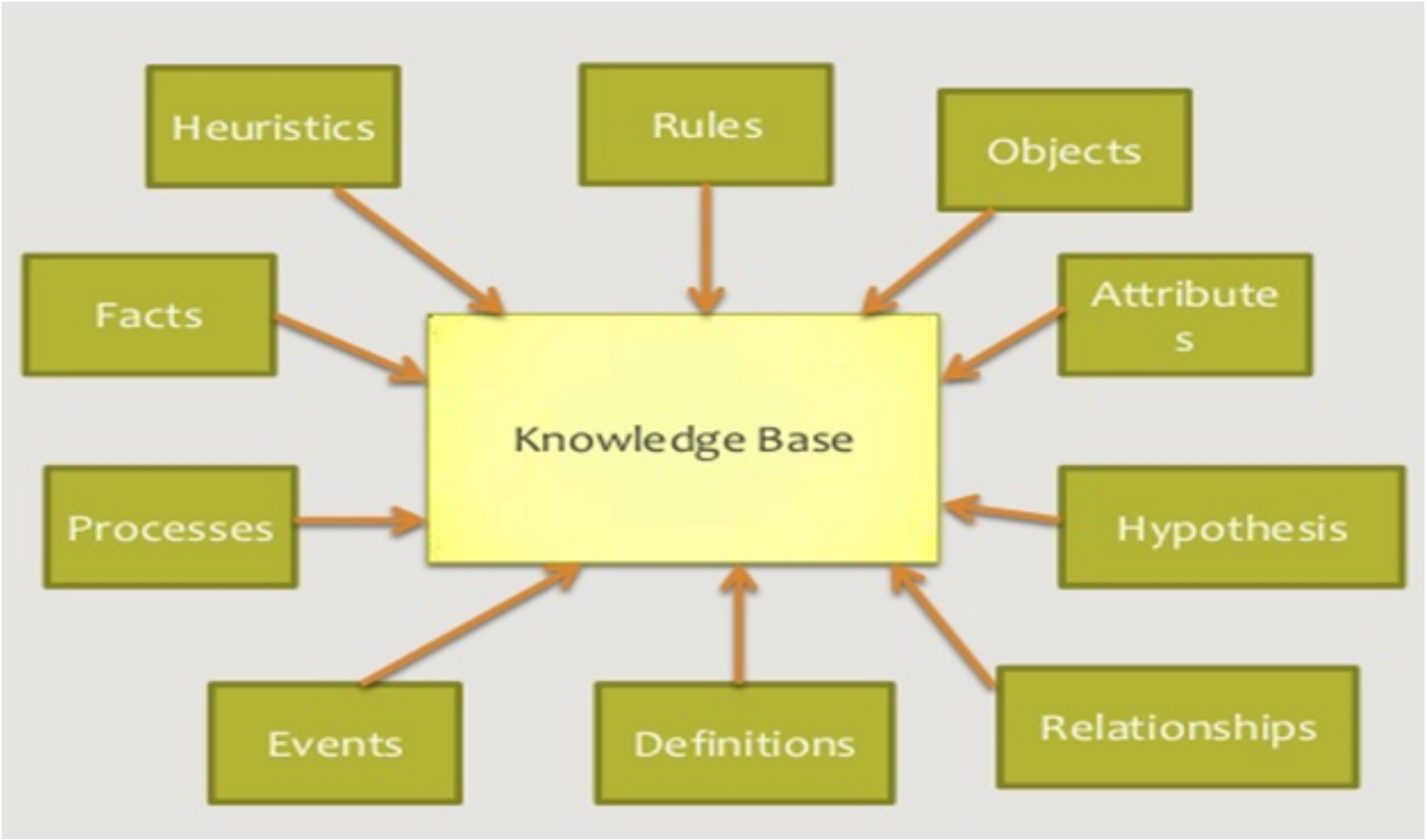
Knowledge Based Expert System

- A Knowledge Based System is a computer program that uses artificial intelligence to solve problems within a specialized domain that ordinarily requires human expertise.
- Typical tasks for expert systems involve classification, diagnosis, monitoring, design, scheduling, and planning for specialized tasks.
- Knowledge-based system is a more general than the expert system.

KBS as Real world Problem Solvers

- Problem-solving power does not lie with smart reasoning techniques nor clever search algorithms but domain dependent real-world knowledge.
- Real-world problems do not have well-defined solutions
- KBS allow this knowledge to be represented and creates an explained solution.
- KBS draws upon the knowledge of human experts captured in a knowledge-base to solve problems that normally require human expertise.
- Uses Heuristic (cause-and-effect) rather than algorithms KBS as real-world problem solvers

Knowledge Base

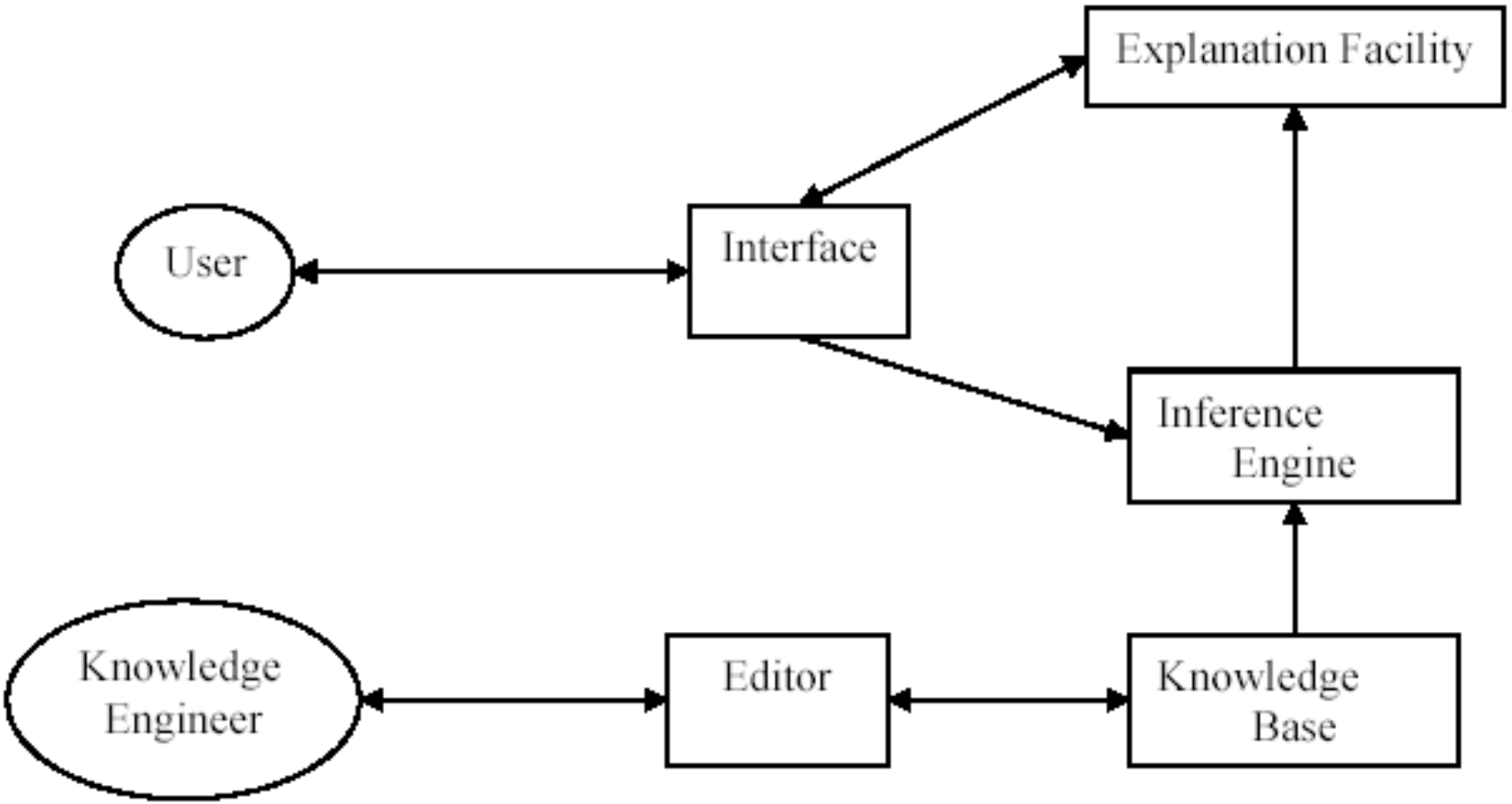


Expert System

- Expert System is an extension of the decision support system.
- **Expert system** is an information system application that captures the knowledge and expertise of a problem solver or decision maker, and then simulates the 'thinking' of that expert for those who have less expertise.
- Expert systems are implemented with **artificial intelligence** technology, often called expert system shells.
- Expert System application areas in action

    1. Medical Diagnosis
    2. Telephone Network Maintenance
    3. Detection of Common Metals

Expert System Architecture



Architecture of a Typical Expert System

Inference Engine

- An inference engine tries to derive answers from a knowledge base.
- It is the brain of the expert systems that provides a methodology for reasoning about the information in the knowledge base, and for formulating conclusions.
- User Interface
  - It enables the user to communicate with an expert system.

Knowledge Engineer

- Knowledge Engineer
  - A knowledge engineer is a computer scientist who knows how to design and implement programs that incorporate artificial intelligence techniques.
- Knowledge Engineering
  - The art of designing and building the expert systems is known as Knowledge Engineering, knowledge engineers are its practitioners.
  - Knowledge engineering relies heavily on the study of human experts in order to develop intelligent & skilled programs.

Knowledge Engineer

- The engineer then translates the knowledge into a computer- usable language, and designs an inference engine, a reasoning structure, that uses the knowledge appropriately.
- He/she also determines how to integrate the use of uncertain knowledge in the reasoning process, and what kinds of explanation would be useful to the end user.
- When the expert system is implemented, it may be:
  - The inference engine is not just right
  - Form of representation of knowledge is awkward
- An expert system is judged to be entirely successful when it operates on the level of a human expert.

Expert Systems

Characteristics of Expert System

- Like a human expert, an expert system is expected to
- Be specialist : know facts and procedural rules
- Use heuristics : interpolate from known facts
- Justify its conclusions : to establish credibility and confidence.
- The user can ask: be able to learn : be able to absorb new knowledge and apply it estimate the reliability of its answer.

Benefits of Expert System

- Benefits of Expert System
  - Increased output and productivity
  - Decreased decision making time
  - Increased process and product quality
  - Reduced downtime (machine failure detect and repair time)
  - Capture of scarce expertise
  - Easier equipment operation
  - Elimination of the need for expensive equipment
  - Operations in hazardous environments (no human required.)
  - Ability to work with incomplete or uncertain information
  - Knowledge transfer to remote locations
  - Enhancement of other information systems

Intelligent System

- **Intelligent systems** is a term that describes the various commercial applications of AI.
- **Artificial Intelligence (AI)** is a subfield of computer science concerned with:
  - studying the thought processes of humans;
  - recreating those processes via machines, such as computers and robots.
- **Behavior by a machine that, if performed by a human being, would be considered intelligent.**
- **Turing test** is a test for artificial intelligence, in which a human interviewer, conversing with both an unseen human being and an unseen computer, cannot determine which is which: named for British Mathematician **AI** pioneer (Alan Turing).

Artificial Intelligence

- A field of science and technology based on disciplines such as computer science, biology, psychology, linguistics, mathematics, & engineering
- At the machine level, think like human and act like human means Artificial Intelligence.
- Goal is to develop computers that can think, see, hear, walk, talk, and feel.
- Major thrust to development of computer functions normally associated with human intelligence reasoning, learning, problem solving

Domains of Artificial Intelligence

- Domains of AI
  - Three major areas
    - Cognitive science
    - Robotics
    - Natural interfaces
- Cognitive science
  - Focuses on researching how the human brain works & how humans think and learn
  - Applications
    - Expert systems
    - Adaptive learning systems
    - Fuzzy logic systems
    - Neural networks
    - Intelligent agents

Domains of Artificial Intelligence

- Robotics
  - Produces robot machines with computer intelligence and computer controlled, human like physical capabilities
- Natural interfaces
  - Natural language and speech recognition
  - Talking to a computer and having it understand
  - Virtual reality

Neural Networks

- **Neural networks** is a system of programs and data structures that approximates the operation of the human brain.
- **Neural networks** are particularly good at recognizing subtle, hidden and newly emerging patterns within complex data as well as interpreting incomplete inputs.

Virtual Reality

- Virtual reality is plainly speaking, seeing an imaginary world, rather than the real one. Seeing, hearing, smelling, testing, feeling. The imaginary world is a simulation running in a computer. The sense data is fed by some system to our brain.
- A medium composed of interactive computer simulations giving users the feeling of being present in the simulations.

Why Virtual Reality is needed?

- Operations in dangerous environments
  - There are still many examples of people working in dangerous or hardship environments that could benefit from the use of VR-mediated teleportation.
  - Workers in radioactive, space, or toxic environments could be relocated to the safety of a VR environment where they could handle any hazardous materials without any real danger using teleoperation or telepresence.

Why Virtual Reality is needed?

- Why Virtual Reality is needed?
  - Scientific Visualization
    - Scientific Visualization provides the researcher with immediate graphical feedback during the course of the computations and gives him/her the ability to 'steer' the solution process.
    - Application at NASA Ames Research Center is the Virtual Planetary Exploration. It helps planetary geologists to remotely analyze the surface of a planet. They use VR techniques to roam planetary terrains.
- NASA VR Mars navigation simulation Geologists remotely analyzing the surface of a planet at NASA

Why Virtual Reality is needed?

- Why Virtual Reality is needed?
  - Medicine
    - Until now experimental research and education in medicine was mainly based on dissection and study of plastic models. Computerized 3D human models provide a new approach to research and education in medicine. Experimenting medical research with virtual patients will be a reality.
    - We will be able to create not only realistic looking virtual patients, but also histological and bone structures. With the simulation of the entire physiology of the human body.

Why Virtual Reality is needed?

- Why Virtual Reality is needed?
  - Education and training
    - The most common example is the flight simulator. This type of simulator has shown the benefits of simulation environments for training. They have lower operating costs and are safer to use than real aircraft.
    - They also allow the simulation of dangerous scenarios not allowable with real aircraft.

Data Mining

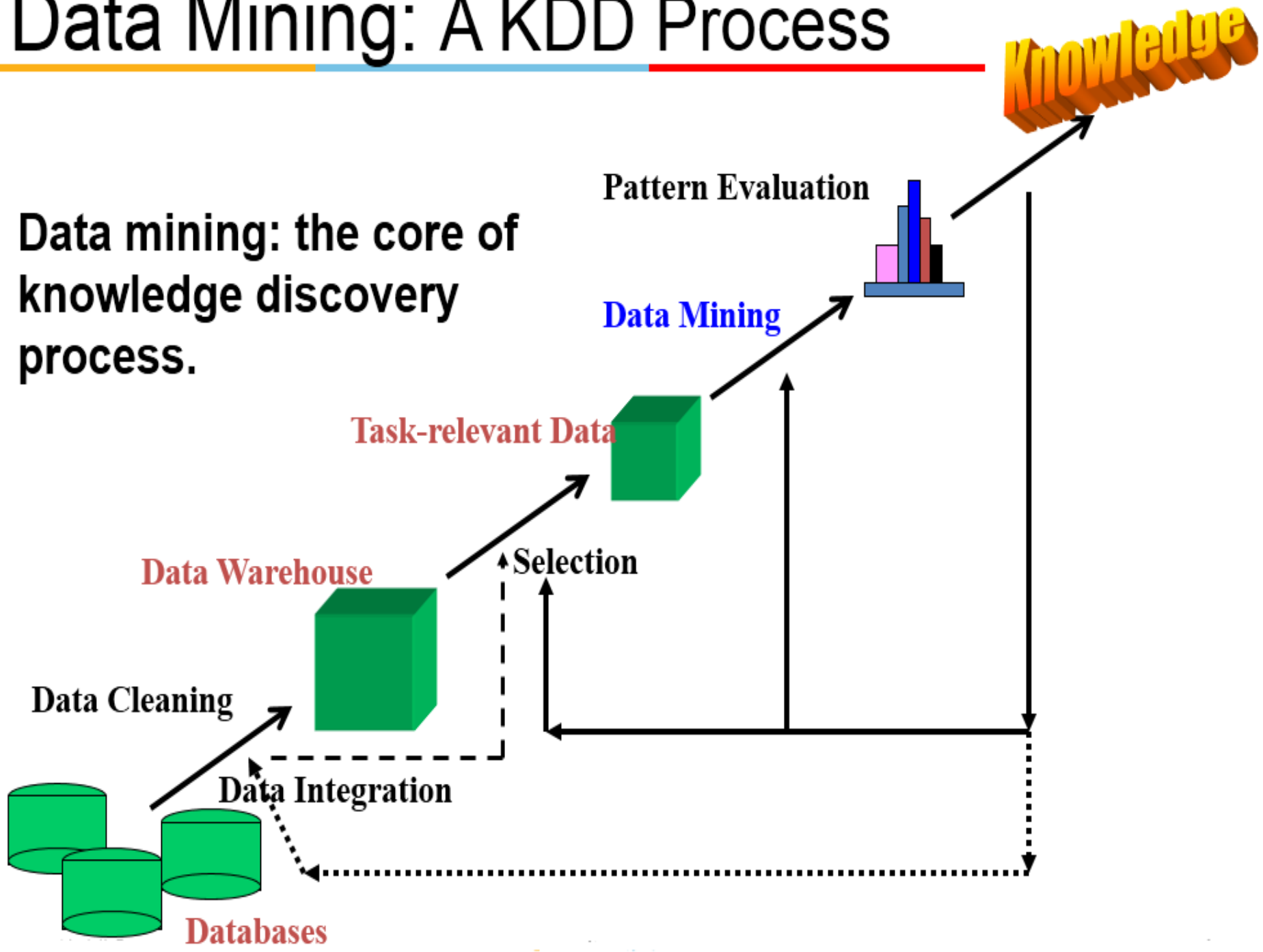
- Data mining is the search for relationships and global patterns that exist in large databases but are `hidden' among the vast amount of data, such as a relationship between patient data and their medical diagnosis.
- Data Mining is also known as knowledge Discovery in Databases(KDD)
- These relationships represent valuable knowledge about the database and the objects in the database and, if the database is a faithful mirror, of the real world registered by the database.

Data Mining

- Data mining refers to using a variety of techniques to identify nuggets of information or decision-making knowledge in bodies of data, and extracting these in such a way that they can be put to use in the areas such as decision support, prediction, forecasting and estimation.
- The data is often voluminous, but as it stands of low value as no direct use can be made of it; it is the hidden information in the data that is useful.

Data Mining: A KDD Process

Data mining: the core of knowledge discovery process.



Steps of a KDD Process

- Learning the application domain:
  - relevant prior knowledge and goals of application
- Creating a target data set: data selection
- Data cleaning and preprocessing: (may take 60% of effort!)
- Data reduction and transformation:
  - Find useful features, dimensionality/variable reduction, invariant representation.
- Choosing functions of data mining
  - summarization, classification, regression, association, clustering.
- Choosing the mining algorithm(s)
- Data mining : search for patterns of interest
- Pattern evaluation and knowledge presentation
  - visualization, transformation, removing redundant patterns, etc.
- Use of discovered knowledge

Why is Data Mining necessary?

- Make use of your data assets
- There is a big gap from stored data to knowledge; and the transition won't occur automatically.
- Many interesting things you want to find cannot be found using database queries

  "find me people likely to buy my products"

  "Who are likely to respond to my promotion"

Why Data Mining

- Credit ratings/targeted marketing:
  - Given a database of 100,000 names, which persons are the least likely to default on their credit cards?
  - Identify likely responders to sales promotions
- Fraud detection:
  - Which types of transactions are likely to be fraudulent, given the demographics and transactional history of a particular customer?
- Customer relationship management:
  - Which of my customers are likely to be the most loyal, and which are most likely to leave for a competitor?

Data Warehouse

- Data Warehouse is a physical repository where relational data are specially organized to provide enterprise-wide, cleansed data in a standardized format.
- Characteristics
  - Subject oriented, Integrated, Time Variant, Non-volatile
  - Web-based, Relational/multidimensional, Client/server, Real-time
  - Include metadata

Data Warehousing is process of constructing and using data warehouses which requires data integration, data cleaning, and data consolidation.

Data Warehouse—Subject-Oriented

- Organized around major subjects, such as customer, product, sales
- Focusing on the modeling and analysis of data for decision makers, not on daily operations or transaction processing
- Provide a simple and concise view around particular subject issues by excluding data that are not useful in the decision support process

Data Warehouse—Integrated

- Constructed by integrating multiple, heterogeneous data sources
  - relational databases, flat files, on-line transaction records
- Data cleaning and data integration techniques are applied.
  - Ensure consistency in naming conventions, encoding structures, attribute measures, etc. among different data sources
    - E.g., Hotel price: currency, tax, breakfast covered, etc.
  - When data is moved to the warehouse, it is converted.

Data Warehouse—Time Variant

- The time horizon for the data warehouse is significantly longer than that of operational systems
  - Operational database: current value data
  - Data warehouse data: provide information from a historical perspective (e.g., past 5-10 years)
- Every key structure in the data warehouse
  - Contains an element of time, explicitly or implicitly
  - But the key of operational data may or may not contain "time element"

Data Warehouse—Nonvolatile

- A physically separate store of data transformed from the operational environment
- Operational update of data does not occur in the data warehouse environment
  - Does not require transaction processing, recovery, and concurrency control mechanisms
  - Requires only two operations in data accessing:
    - initial loading of data and access of data

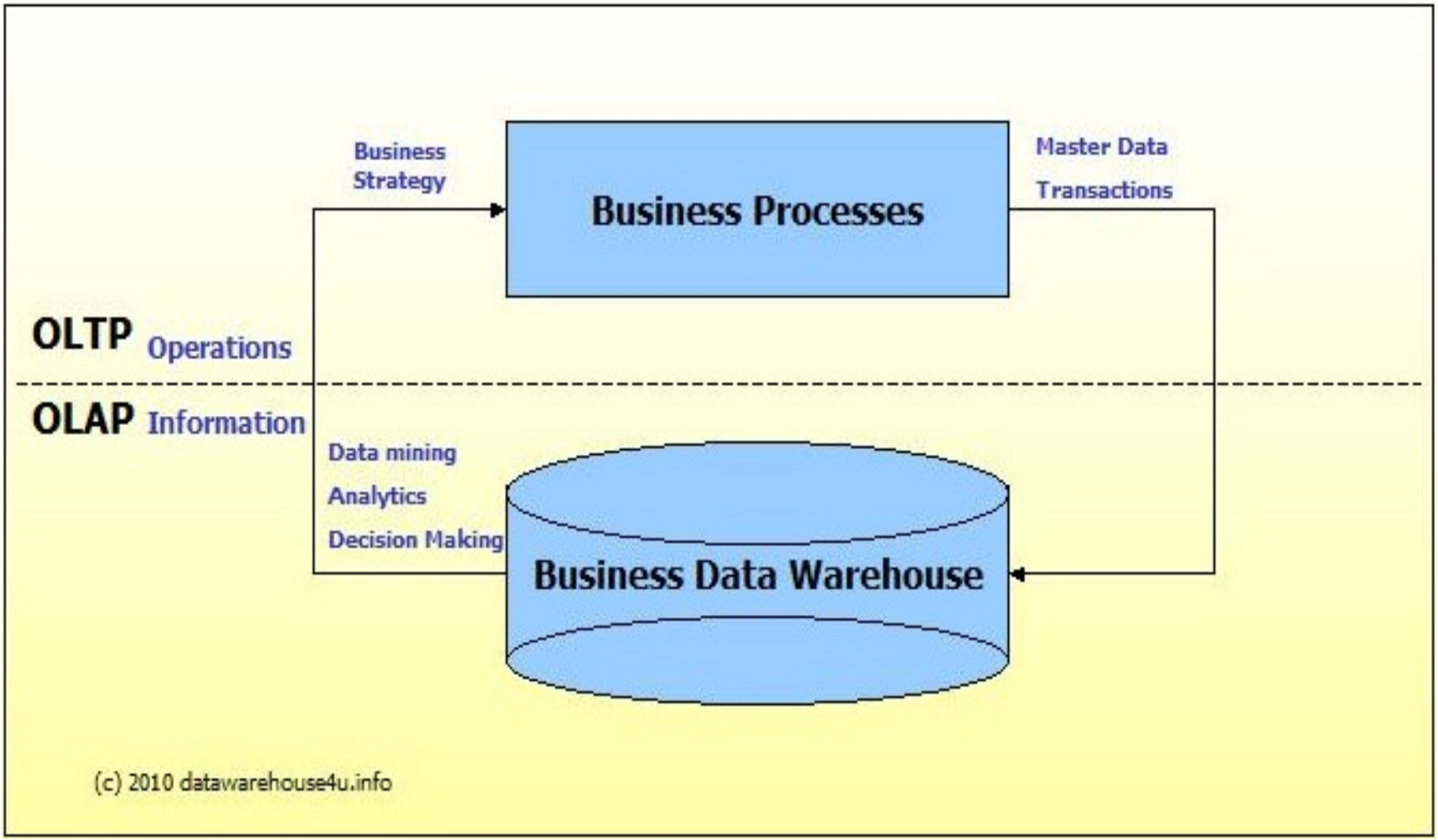
OLTP vs. OLAP

Divide IT systems into transactional (OLTP) and analytical (OLAP).

In general, assume that OLTP systems provide source data to data warehouses, whereas OLAP systems help to analyze it.

- OLTP (on-line transaction processing)
  - Major task of traditional relational DBMS
  - Day-to-day operations: purchasing, inventory, banking, manufacturing, payroll, registration, accounting, etc.
- OLAP (on-line analytical processing)
  - Major task of data warehouse system
  - Data analysis and decision making

OLTP Vs OLAP


Business Strategy
Business Processes
Master Data
Transactions
OLTP Operations
OLAP Information
Data mining
Analytics
Decision Making
Business Data Warehouse
(c) 2010 datawarehouse4u.info

OLTP vs. OLAP

- OLTP: On-Line Transaction Processing
  - Many short transactions (queries + updates)
  - Examples:
    - Update account balance
    - Enroll in course
    - Add book to shopping cart
  - Queries touch small amounts of data (one record or a few records)
  - Updates are frequent
  - Concurrency is biggest performance concern

- OLAP: On-Line Analytical Processing
  - Long transactions, complex queries
  - Examples:
    - Report total sales for each department in each month
    - Identify top-selling books
    - Count classes with fewer than 10 students
  - Queries touch large amounts of data
  - Updates are infrequent
  - Individual queries can require lots of resources

Thank you

Planning for Information System

Chapter - 5

Kantipur Engineering College

Dhapakhel, Lalitpur

What is Information System Planning

- Information system planning is a formal process that develops plan for developing and managing information systems that will support goals of an organization.
  - Information System plan includes:
    - Activities planner believes will help achieve goals.
    - Program for monitoring real-world progress.
    - Means for implementing changes in the plan.

Why Plan?

- To obtain resources
  - Financial
  - Facilities – “Capacity planning”
  - Staff
- To align Information System with the business
- To identify needed applications
- To establish goal, schedule, and milestone in order to track progress
- To provide an opportunity for communication with top management and user management

Approaches to Planning

- Top-down Planning
  - Focuses on organizational goals first, then on the needs of business units
- Bottom-up Planning
  - Focuses on needs of business units first, then on organizational goals

Information System Planning Process

- Establish a mission statement
- Assess the environment
- Set goals and objectives
- Derive strategies and policies
- Develop long-, medium-, and short-range plans implement plans and monitor results

Establish a Mission Statement

- These are services that you are responsible for; it is your place in the organization.
- It is not what you are supposed to achieve, it is who you are and what you do in the company.

Goals and Objectives

- Set goals – what do you want to achieve?
- Set objectives – what are your specific, measurable targets?

Derive strategies and policies

- Strategies for
  - Technology focus
  - Personnel and career development
  - Aligning with the company
  - Funding criteria; how much to spend on IT?
- Policies for
  - Funding criteria; how much to spend on IT?
  - Allocation criteria; priority setting
  - Organizational arrangements
  - Use of outside IT services, outsourcing
  - Selling IT services to outside organizations

Develop long-, medium-, and short-range plans

- Short-Range – the next year, the next budget period; developing and operating current systems

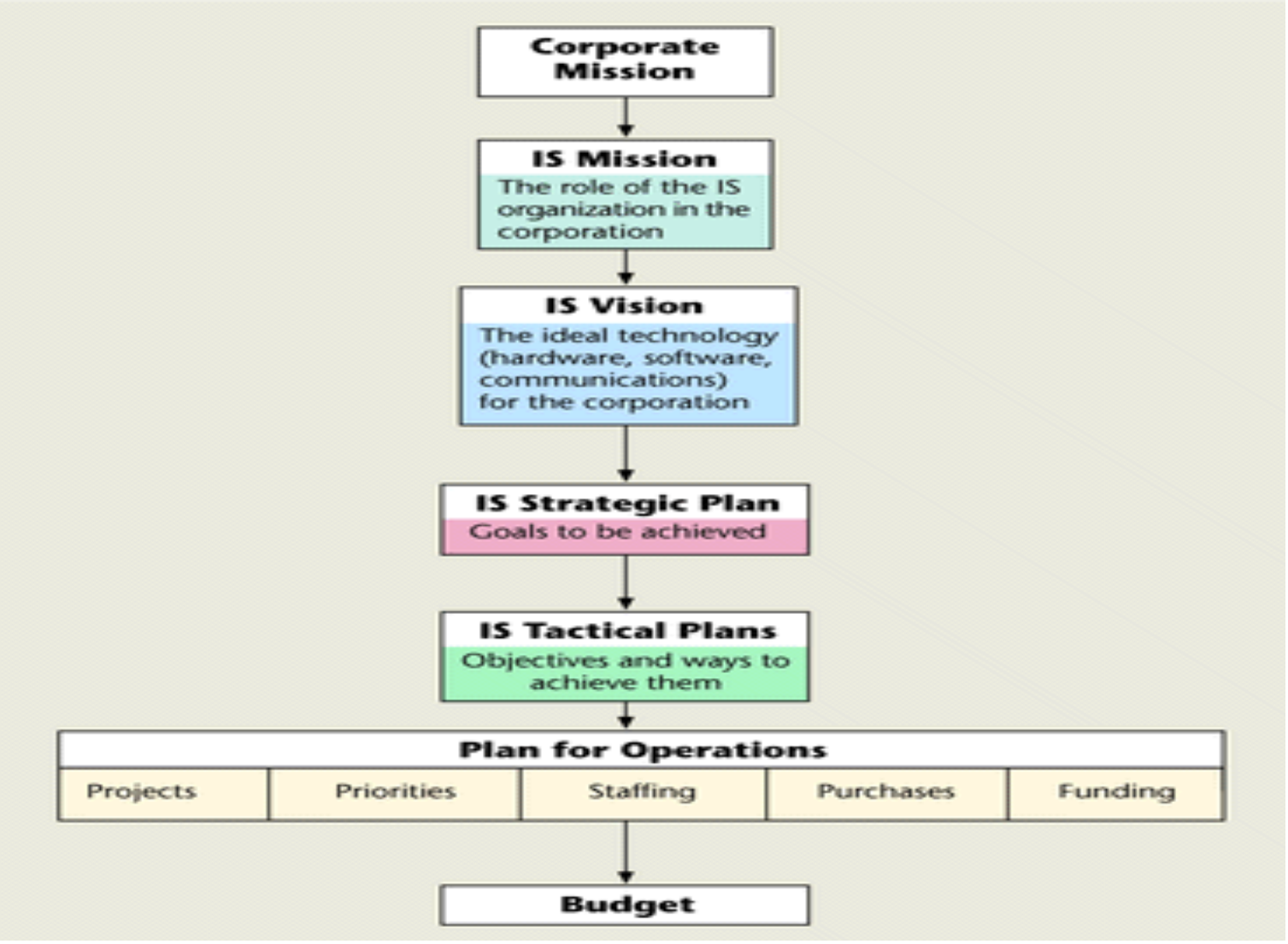
- Medium-Range – committing to development efforts for applications that will take more than one year to complete; meeting management's current information needs, projected into the future for as many years as needed to complete them. This is what most organizations call "Long-Range Planning."

- Long-Range planning – preparing for management's future information needs. These are not application specific; they are investments in infrastructure ; it is creating an information architecture.

What are Key Elements of IS Planning

- Key elements of an IS Plan are
  - Corporate mission statement
  - Vision for IT within organization
  - IS strategic and tactical plans
  - Operations plan to achieve mission and vision
  - Budget to ensure resources are available

Steps of Information Systems Planning



Information Systems Planning

- The Corporate and IS Mission Statements
  - The corporate mission statement details the purpose of the organization and its overall goals.
  - IS mission statement outlines the purpose of ISs in the organization.

Information Systems Planning

- The IT Vision
  - Wish list of what IS managers would like to see in terms of hardware, software, and communications, to contribute to goals of the organization

Information Systems Planning

- Strategic IS Planning
  - IS Strategic plan details what is to be achieved
  - Strategic plans are designed with the entire organization in mind and begin with an organization's mission.
  - Essentially, strategic plans look ahead to where the organization wants to be in three, five, even ten years. Strategic plans, provided by top-level managers, serve as the framework for lower-level planning.

Strategic Information System

- Strategic Information System is a system that helps companies alter their business strategy. It is used to accelerate the reaction time to environmental changes and aid the company in achieving a competitive advantage over its competitors.
- They help in producing low cost quality products.
- The strategic role of IS involves using IT to develop products, services, and capabilities that give company major advantages over the competitive forces it faces in the global marketplace.

Typical Strategic-Level Job Titles

- Top-level Managers
- CEOs or Presidents
- General Manager
- Corporate Boards
- Steering Committee
- Board of Directors
  - will design and execute strategic plans to paint a picture of the desired future and long-term goals of the organization.

Tactical IS Planning

- IS Tactical plan describes how goals will be met and by when
- Tactical IS Plan performed by middle managers responsible for acquisition and allocation of resources for projects according to tactical plans, set out for one or two years.
- Tactical IS Planning is evaluates current and projected information needs of the organization, prioritizes IS development projects, and develops allocation plans for financial and technology resources.

What are important factors in IS tactical planning

- Important Factors in IS Tactical Planning
  - Flexibility
  - Compatibility
  - Connectivity
  - Scalability
  - Standardization
  - Total Cost of Ownership

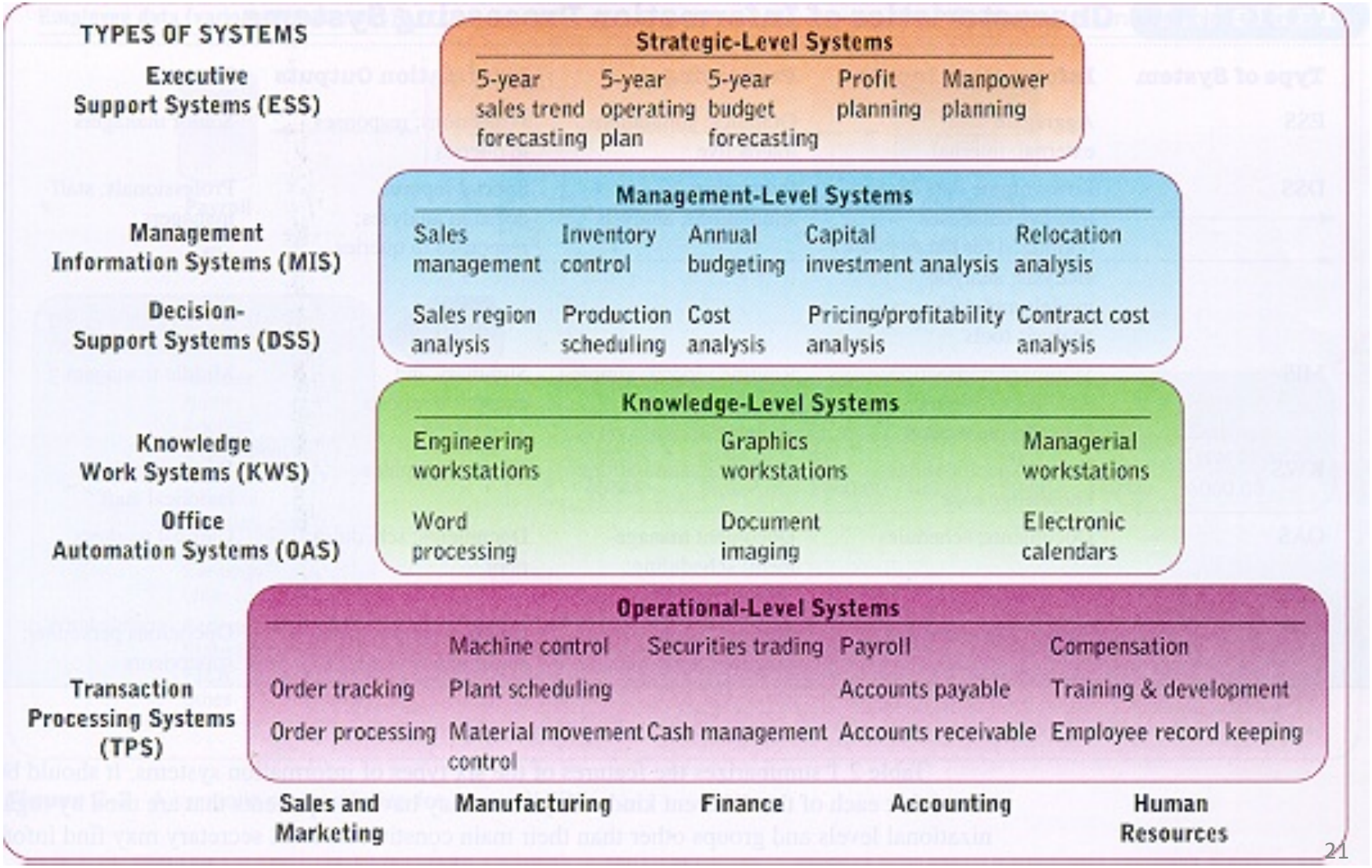
Typical Tactical-Level Job Titles

- Advertising manager
- Personnel manager
- Creative director
- Manager of information systems
- Communications director
- Chief Financial Officer

Operational IS Planning

- Operational Plan prepared by a component of an organization that clearly defines actions it will take to support the strategic objectives and plans of upper management.
- Operational IS Planning develops plans such as annual operating budgets and individual IS project plans.
- Operations IS Planning performed by supervisors of smaller work units concerned with planning and control of short-term (typically, a week or six months) budgets and schedules.

Hierarchy of Planning



Three levels of planning.

Type of Plan	Created By	Scope	Includes	Level of Detail
Strategic Plan	Top Management	Entire organization	Mission of the company, future goals and ambitions	Very broad and general
Tactical Plan	Mid-level Management	Single area of the business as a whole (e.g. a division of the company)	Specific actions to support or work towards the Strategic Plan	Specific actions and ideas, but not very detailed
Operational Plan	Low-level Management	A unit within a single area of the business (e.g. a department within a division)	Specific plans for low level and day-to-day activities and processes that will support and enable the Tactical Plan	Extremely detailed (who, what, where and when)

Thank You

Web Based Information System and Navigation

Chapter-7

KEC, Dhapakhel

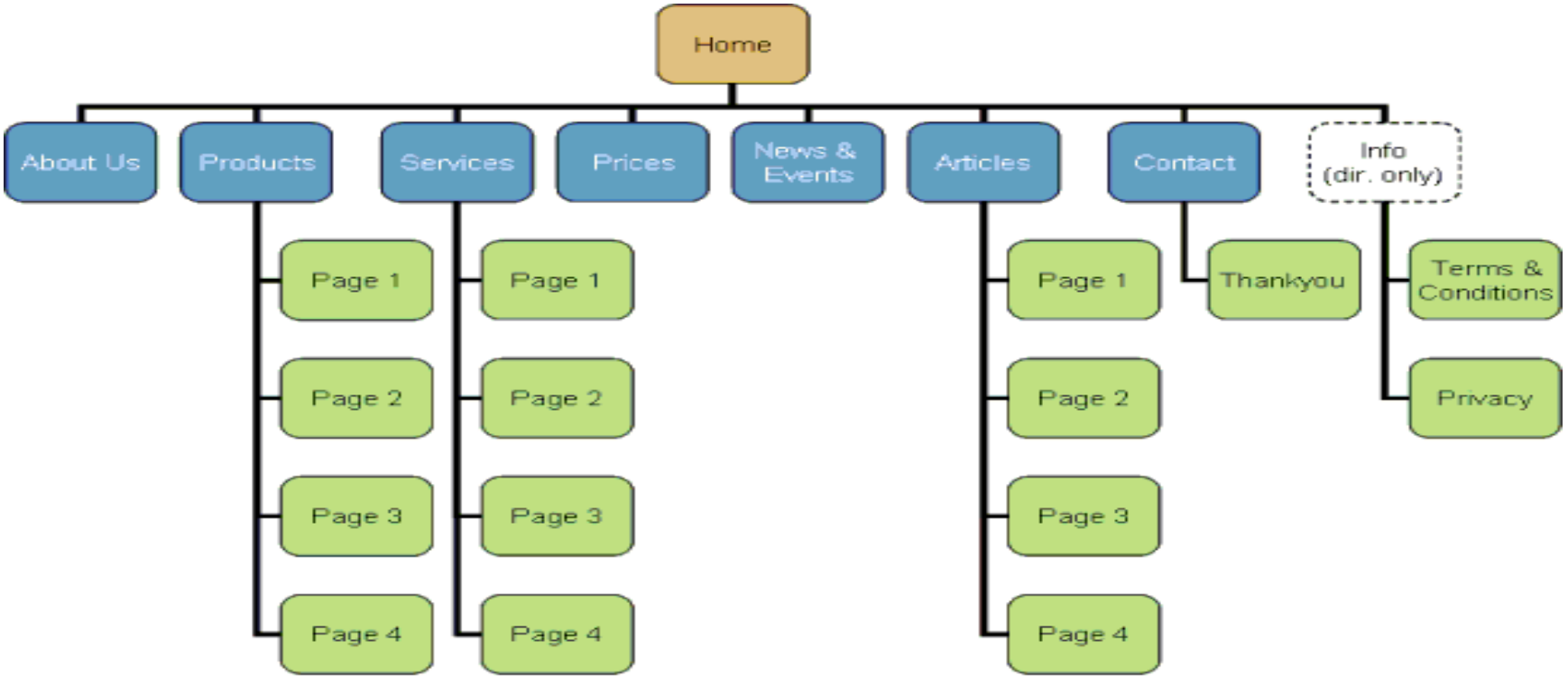
Structure of the Website

- Website structure is the process of defining the look and feel and the navigation of CM *websites*.
- A site structure is normally the role of an information architect, the reality is that everybody from designers to website owners find themselves working on it.
- The website structure consists of three components: Layout Templates, URL patterns, and Linkage Structure.

Layout Template

- Most web pages consist of HTML elements like table, menu, button, image, and input box.
- The layout of a web page describes what HTML elements are included in the page, as well as how these elements are visually distributed in page rendering.
- In a website, pages are generated based on distinguishable templates according to their functions.
- Visually similar pages usually have same function. In this way, user can easily identify a page's function at a glance.

Layout of Site



Home — Home Page. *Eg www.domain.com/content/main.html*

About — Top Level Content – Section Main Page. *Eg www.domain.com/content/about/main.html*

Page — Section Content Page. *Eg www.domain.com/content/about/page1.html*

URL Pattern

- A URL pattern is a generalization of a group of URLs sharing similar syntactic format.
- Some example URL patterns discovered, again, from the ASP.NET Forums.
  - List-of-thread pages
    - ^http://forums\.asp\.net/\d+\.aspx$
    - ^http://forums\.asp\.net/\d+\.aspx\?PageIndex=\d+&forumoptions=\d+:\d+:\d+::$
  - List-of-post pages
    - ^http://forums\.asp\.net/t/\d+\.aspx$
    - ^http://forums\.asp\.net/t/\d+\.aspx\?PageIndex=\d+$
    - ^http://forums\.asp\.net/p/\d+/\d+\.aspx$
    - ^http://forums\.asp\.net/ThreadNavigation\.aspx\?PostID=\d+&NavType=(Previous|Next)$
  - User profile pages
    - ^http://forums\.asp\.net/user/Profile\.aspx\?UserID=\d+$
    - ^http://forums\.asp\.net/members/[^/?]*$
- It is noticed that one layout templates can have more than one related URL pattern. For example, a bookseller website usually designs one template to show a list of books, and provides different query parameters to generate such a list.
- Various query parameters in this scenario will lead to different URL patterns, but the search results are shown with the same template.
- Another common case is duplicate pages, i.e., pages with the same content (and very likely the same layout) but different URLs.

Link Structure

- Based on the layout templates and URL patterns, we can construct a directed graph to represent the website organization structure.
- Each layout template is considered as a node in a graph, and two nodes are linked if there are hyperlinks between the pages belonging to the two nodes.
- The link direction is the same as the related hyperlinks. And each link is characterized with the URL pattern of the corresponding hyperlink URLs.
- It should be noticed that there could be multiple links from one node to another if the corresponding hyperlinks have more than one URL pattern.

Web Structure

- WEB STRUCTURE is a pioneer in "fusion engineering"; fusing design sensitivity with cost consciousness to develop the most cost effective structures in which the traditional separation between architectural design and structures is erased in a seamless harmony of design intent.
- In the completed work, architecture and structure resonate to create a single entity; an essential symbiotic interaction, a fusion that presents itself in works of beauty.

Site Structure

- When confronted with a new and complex information system, users build mental model.
- They use these models to assess relations among topics and to guess where to find things they haven't seen before.
- The success of the organization of web site will be determined largely by how well site's information architecture matches users' expectations.
- A logical, consistently named site organization allows users to make successful predictions about where to find things.
- Consistent methods of organizing and displaying information permit users to extend their knowledge from familiar pages to unfamiliar ones.
- If you mislead users with a structure that is neither logical nor predictable, or constantly uses different or ambiguous terms to describe site features, users will be frustrated by the difficulties of getting around and understanding what you have to offer.
- Don't want user's mental model of web site to look like figure1.

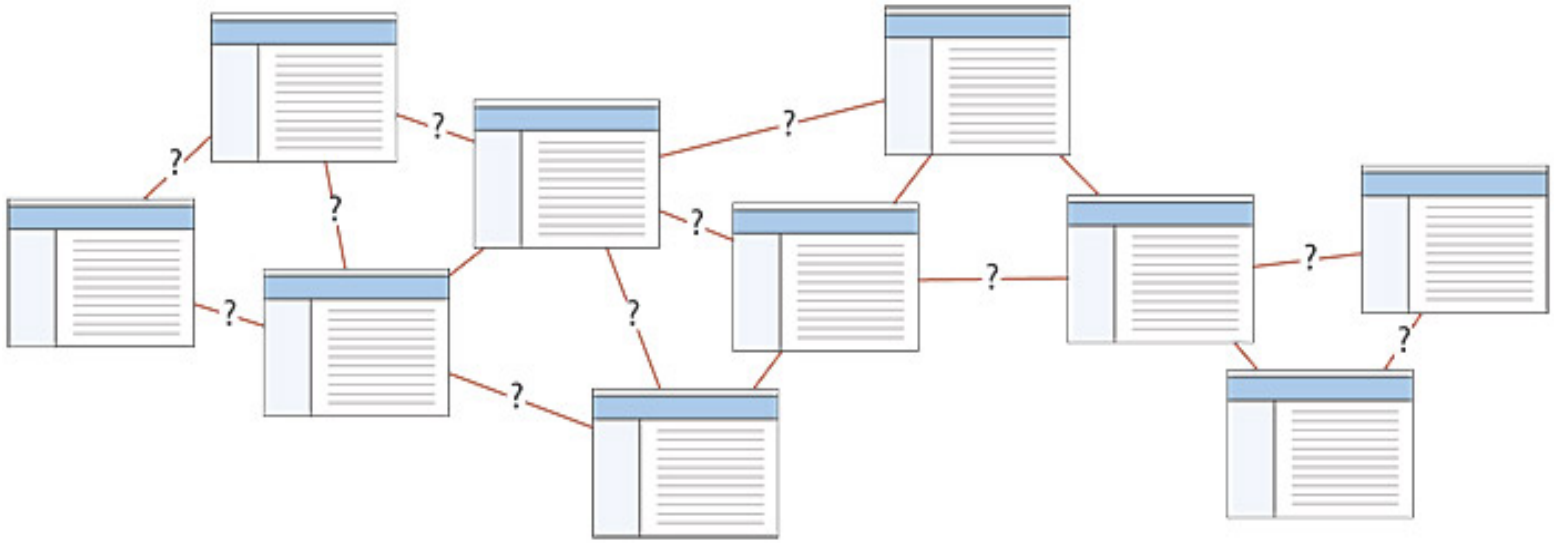


Figure 1 — Don't make a confusing web of links. Designers aren't the only ones who make models of sites. Users try to imagine the site structure as well, and a successful information architecture will help the user build a firm and predictable mental model of site.

Web Page Structure

- A web page constructed using HTML has a basic and essential structure. The page always begins with the **start tag** of the html element and always terminates with the **end tag** of the html element as follows:

  Example 1

  <html>
  ...web page...
  </html>

- All other element tags are 'nested' within the start and end html tags. The web page is then further subdivided into two main sections which are the 'head' and the 'body'.

Web Page Structure

- The head section begins with the <head> start tag and terminates with the</head> end tag.
- Immediately following this comes the <body> start tag and just before the html end tag comes the </body> end tag.
- There is only *one set* of <html>**...**</html> tags,
- *one set* of <head>**...**</head> tags and
- *one set* of <body>**...**</body> tags.

How do Web Pages Change

- Most pages do not change much.
- Larger pages change more often.
- Commercial pages change more often.
- Past change to a web page is a good indicator of future change.
- About 30% of pages are very similar to other pages, and being a near-duplicate is fairly stable.

Link Analysis

- Link analysis is a data-analysis technique used to evaluate connections between nodes. Relationships may be identified among various types of nodes (objects), including organization, people and transactions.
- The analysis of hyperlinks and the graph structure of the Web has been instrumental in the development of web search.
- Link analysis for web search has intellectual antecedents in the field of citation analysis, aspects of which overlap with an area known as Bibliometrics.

Link Analysis

- Link analysis is an important part of site assessment, either your own or competitor's.
- Outbound links are links on your site which refer to other sites, they go beyond the borders of your site.
- Internal links are links which point to another page of your site, i.e. they refer to some place within your site.

Website Navigation

Website navigation is important to the success of website visitor's experience to website.

The website's navigation system is like a road map to all the different areas and information contained within the website.

Types of Website Navigation

- **Hierarchical website navigation**

  The structure of the website navigation is built from general to specific. This provides a clear, simple path to all the web pages from anywhere on the website.

- **Global website navigation**

  Global website navigation shows the top level sections/pages of the website. It is available on each page and lists the main content sections/pages of the website.

- **Local website navigation**

  Local navigation would the links with the text of your web pages, linking to other pages within the website.

Website Navigation Use

- To be consistent throughout the website. The website visitors will learn, through repetition, how to get around the website.
- The main navigation links kept together. This makes it easier for the visitor to get to the main areas of the website.
- Reduced clutter by grouping links into sections. If the list of website navigation links are grouped into sections and each section has only 5-7 links, this will make it easier to read the navigation scheme.
- Minimal clicking to get to where the visitor wants to get to. If the number of clicks to the web page the visitor wishes to visit is minimal, this leads to a better experience.

Website Navigation Use

- Some visitors can become confused or impatient when clicking a bunch of links to get to where they want to be.
- In large websites, this can be difficult to reduce. Using breadcrumbs is one way to help the visitor see where they are within the website and the path back up the navigation path they took.
- Creating the website navigation system at the planning stage of the website will effect the overall design of the web page layout and help develop the overall plan for the website.

Web Mining

- Web mining can be broadly defined as discovery and analysis of useful information from the World Wide Web.
- Based on the different emphasis and different ways to obtain information, web mining can be divided into two major parts: Web Contents Mining and Web Usage Mining.
- **Web Contents Mining** can be described as the automatic search and retrieval of information and resources available from millions of sites and on-line databases though search engines / web spiders.
- **Web Usage Mining** can be described as the discovery and analysis of user access patterns, through the mining of log files and associated data from a particular Web site.

Web Usage Mining

- The automatic discovery of patterns in clickstreams and associated data collected or generated as a result of user interactions with one or more Web sites.

Goals

- To analyze the behavioral patterns and profiles of users interacting with a Web site.
- The discovered patterns are usually represented as collections of pages, objects, or resources that are frequently accessed by groups of users with common interests.

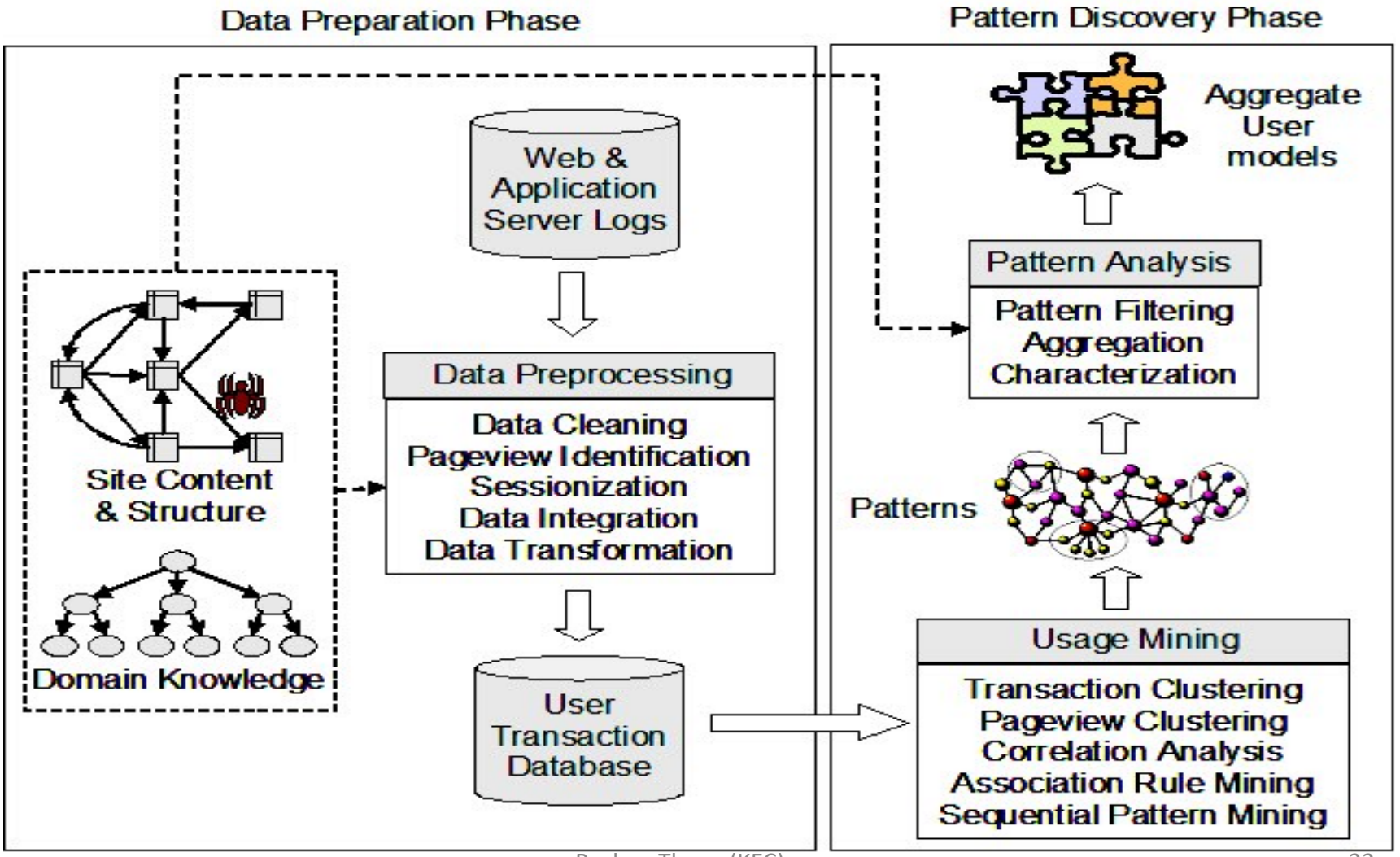
How to perform Web Usage Mining?

- Web usage mining is achieved first by reporting visitors traffic information based on Web server log files and other source of traffic data.
- Web server log files were used initially by the webmasters and system administrators for the purposes of "how much traffic they are getting, how many requests fail, and what kind of errors are being generated", etc. **However, Web server log files can also record and trace the visitors' on-line behaviors.**
- Web log file is one way to collect Web traffic data. The other way is to "sniff" TCP/IP packets as they cross the network, and to "plug in" to each Web server.

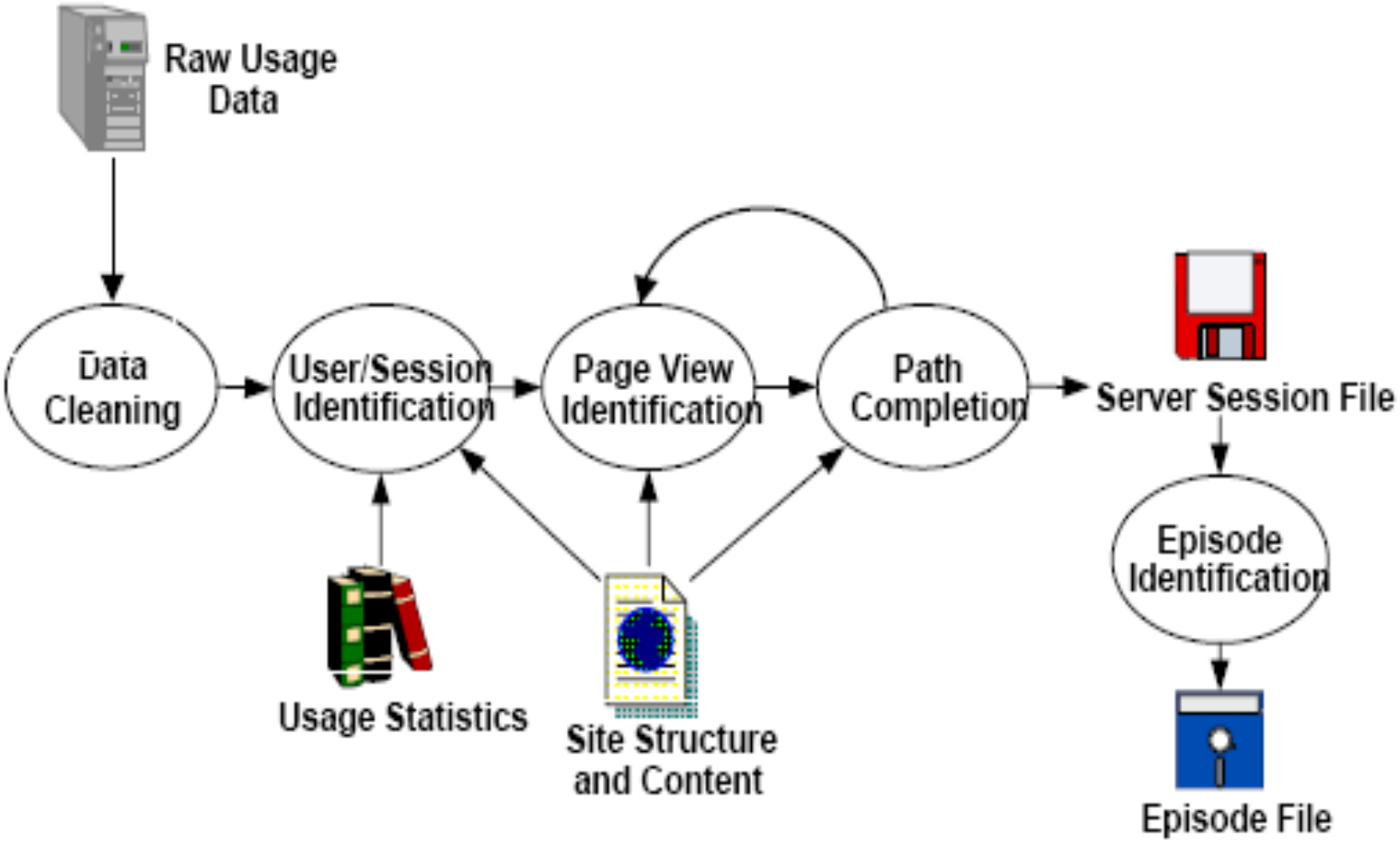
Data in Web Usage Mining

- Data in Web Usage Mining:
  - Web server logs
  - Site contents
  - Data about the visitors, gathered from external channels
  - Further application data
- Not all these data are always available.
- When they are, they must be integrated.
- A large part of Web usage mining is about processing usage/ clickstreams data.
  - After that various data mining algorithm can be applied.

Web usage mining process



Pre-processing of web usage data



Data Cleaning

- Data cleaning
  - remove irrelevant references and fields in server logs
  - remove references due to spider navigation
  - remove erroneous references
  - add missing references due to caching (done after sessionization)

Identify sessions (Sessionization)

- In Web usage analysis, these data are the sessions of the site visitors: the activities performed by a user from the moment she enters the site until the moment she leaves it.
- Difficult to obtain reliable usage data due to proxy servers and dynamic IP addresses, missing references due to caching, and the inability of servers to distinguish among different visits.

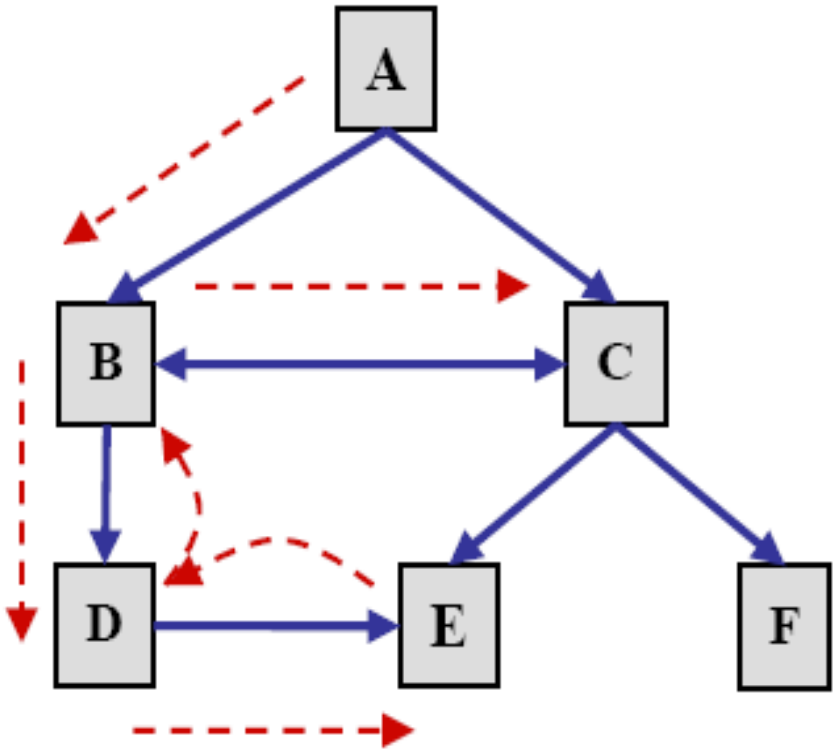
Page View

- A pageview is an aggregate representation of a collection of web objects contributing to the display on a user's browser resulting from a single user action (such as a click-through).
- Conceptually, each pageview can be viewed as a collection of web objects or resources representing a specific "user event" e.g., reading an article, viewing a product page, or adding a product to the shopping cart.

Path Completion

- Client- or proxy-side caching can often result in missing access references to those pages or objects that have been cached.
- For instance,
  - if a user returns to a page A during the same session, the second access to A will likely result in viewing the previously downloaded version of A that was cached on the client-side, and therefore, no request is made to the server.
  - This results in the second reference to A not being recorded on the server logs.

Missing references due to caching



User's actual navigation path:

A → B → D → E → D → B → C

What the server log shows:

URL	Referrer
A	--
B	A
D	B
E	D
C	B

Fig. 12.7. Missing references due to caching.

Path completion

- The problem of inferring missing user references due to caching.
- Effective path completion requires extensive knowledge of the link structure within the site
- Referrer information in server logs can also be used in disambiguating the inferred paths.
- Problem gets much more complicated in frame-based sites.

Collaborating Filtering

- Collaborative filtering(CF) is the process of filtering or evaluating items through the opinions of other people.
- CF technology brings together the opinions of large interconnected communities on the web, supporting filtering of substantial quantities of data.
- Collaborative filtering systems produce predictions or recommendations for a given user and one or more items. Items can consist of anything for which a human can provide a rating, such as art, books, CDs, journal articles, or vacation destinations.

Collaborating Filtering

Ratings in a collaborative filtering system can take on a variety of forms.

- Scalar ratings can consist of either numerical ratings, such as the 1-5 stars provided in ordinal ratings such as strongly agree, agree, neutral, disagree, strongly disagree.
- Binary ratings model choices between agree/disagree or good/bad.
- Unary ratings can indicate that a user has observed or purchased an item, or otherwise rated the item positively.

The absence of a rating indicates that we have no information relating the user to the item (perhaps they purchased the item somewhere else).

Collaborative Filtering

Match people with similar interests as a basis for recommendation.

1) Many people must participate to make it likely that a person with similar interests will be found.
2) There must be a simple way for people to express their interests.
3) There must be an efficient algorithm to match people with similar interests.

How does CF Work?

- Users rate items – user interests recorded.
- Ratings may be:
  - Explicit, e.g. buying or rating an item
  - Implicit, e.g. browsing time, no. of mouse clicks
- Nearest neighbour matching used to find people with similar interests
- Items that neighbours rate highly but that you have not rated are recommended to you
- User can then rate recommended items

Example of CF MxN Matrix with M users and N items (An empty cell is an unrated item)

Items / Users	Data Mining	Search Engines	Data Bases	XML
Alex	1		5	4
George	2	3	4	
Mark	4	5		2
Peter			4	5

Observations

- Can construct a vector for each user (where 0 implies an item is unrated)
  - E.g. for Alex: <1,0,5,4>
  - E.g. for Peter <0,0,4,5>
- On average, user vectors are sparse, since users rate (or buy) only a few items.
- Vector similarity or correlation can be used to find nearest neighbour.
  - E.g. Alex closest to Peter, then to George.

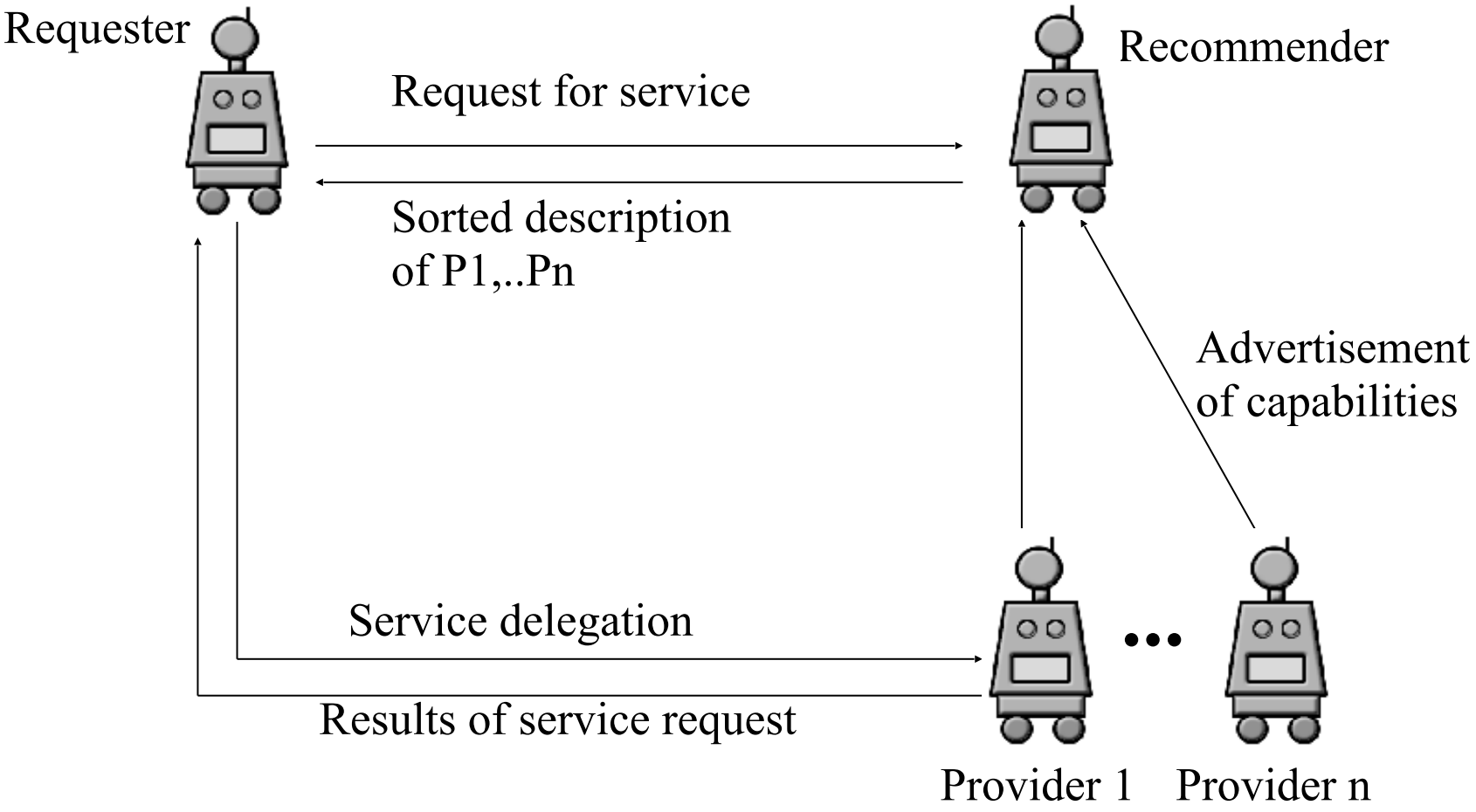
Case Study – Amazon.com

- Item-to-item collaborative filtering
  - Find similar items rather than similar customers.
- Record pairs of items bought by the same customer and their similarity.
  - This computation is done offline for all items.
- Use this information to recommend similar or popular books bought by others.
  - This computation is fast and done online.

Recommender systems

- Too much information: information overload – consumers have too many options
- A recommender system is a system which provides recommendations to a user
- Applications: Books, music CDs, movies. Even documents, services and other products such as software games

Recommender



Information needed

Information used for recommendations can come from different sources:

- browsing and searching data
- purchase data
- feedback explicitly provided by the users
- textual comments
- expert recommendations
- demographic data

Providing recommendations

Recommendations can take the following forms:

- Attribute-based recommendations: based on syntactic attributes of products (e.g. *science fiction* books)
- Item-to-item correlation (as in shopping basket recommendations)
- User-to-user correlation (finding users with similar tastes)
- Non-personalized recommendations (as in traditional stores, i.e. dish of the day, generic book recommendations etc.)

Recommender systems in e-commerce

- Turning browsers into customers: they can stimulate the users' needs (need identification stage)
- Cross-selling: suggest additional products which may match the user's interests or current shopping basket
- Personalization: personalized services, or the site can be personalized to the user's liking – unique shopping experience
- Keeping customers informed
- Retaining customer loyalty

Collective Intelligence

- A shared or group intelligence that emerges from the collaboration and competition of many individuals.
- Groups of people and computers, connected by the Internet, collectively doing intelligent things. For example, Google technology harvests knowledge generated by millions of people creating and linking web pages and then uses this knowledge to answer queries in ways that often seem amazingly intelligent.
- In Wikipedia, thousands of people around the world have collectively created a very large and high quality intellectual product with almost no centralized control, and almost all as volunteers!

Examples

- One example of collective intelligence would be political parties and the way in which the take the views of the people to form policies, select their candidates and run election campaigns.
- Online multi-player games are another example of collective intelligence. Games such as Halo, Second Life and Call of Duty rely on gamers coming together as a community to form the game's Identity.

Examples

- The online encyclopaedia Wikipedia is one of the best examples of collective intelligence. Anyone can add information to an exiting page or indeed create a new page of information; pages also hyperlink to other areas of the website that people have edited.
- Google is a prominent example of collective intelligence. The search engine is made up of millions of websites, which have been created by people all over the world.

Examples

- The social networking world is perhaps the most popular of collective intelligence. Friend post statuses which then act as newsfeed, which informs other friends of their thoughts. Friends can also recommend other friends, applications and pages to any person on their friend list.

Example

- If a person has a Amazon account they can buy or sell products to other people with accounts this is collective intelligence because the people are making up the website.
- The website also recommends items that may also interest you judging on what you have already looked at which is collective intelligence also.
- Things such as customer reviews can also be heavily influential when choosing a product. You are essentially basing your opinion off of the opinions of other members of the public.

Thank You

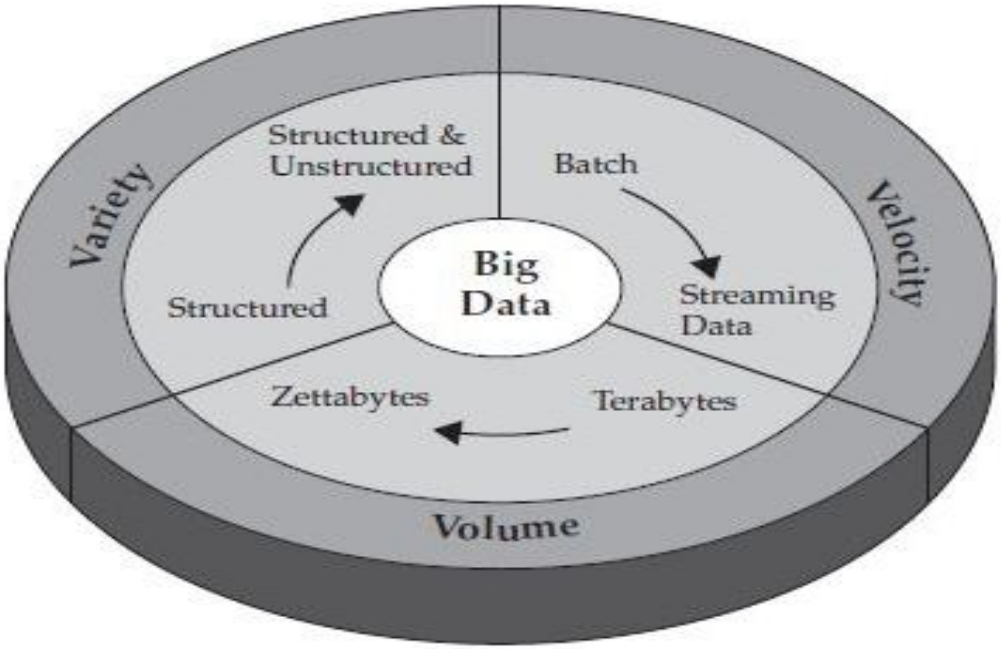
Scalable and Emerging Information System Techniques

Chapter-8

KEC, Dhapakhel

Big Data

- Big Data applies to information that can't be processed or analyzed using traditional processes or tools.



IBM characterizes Big Data by its volume, velocity, and variety—or simply,

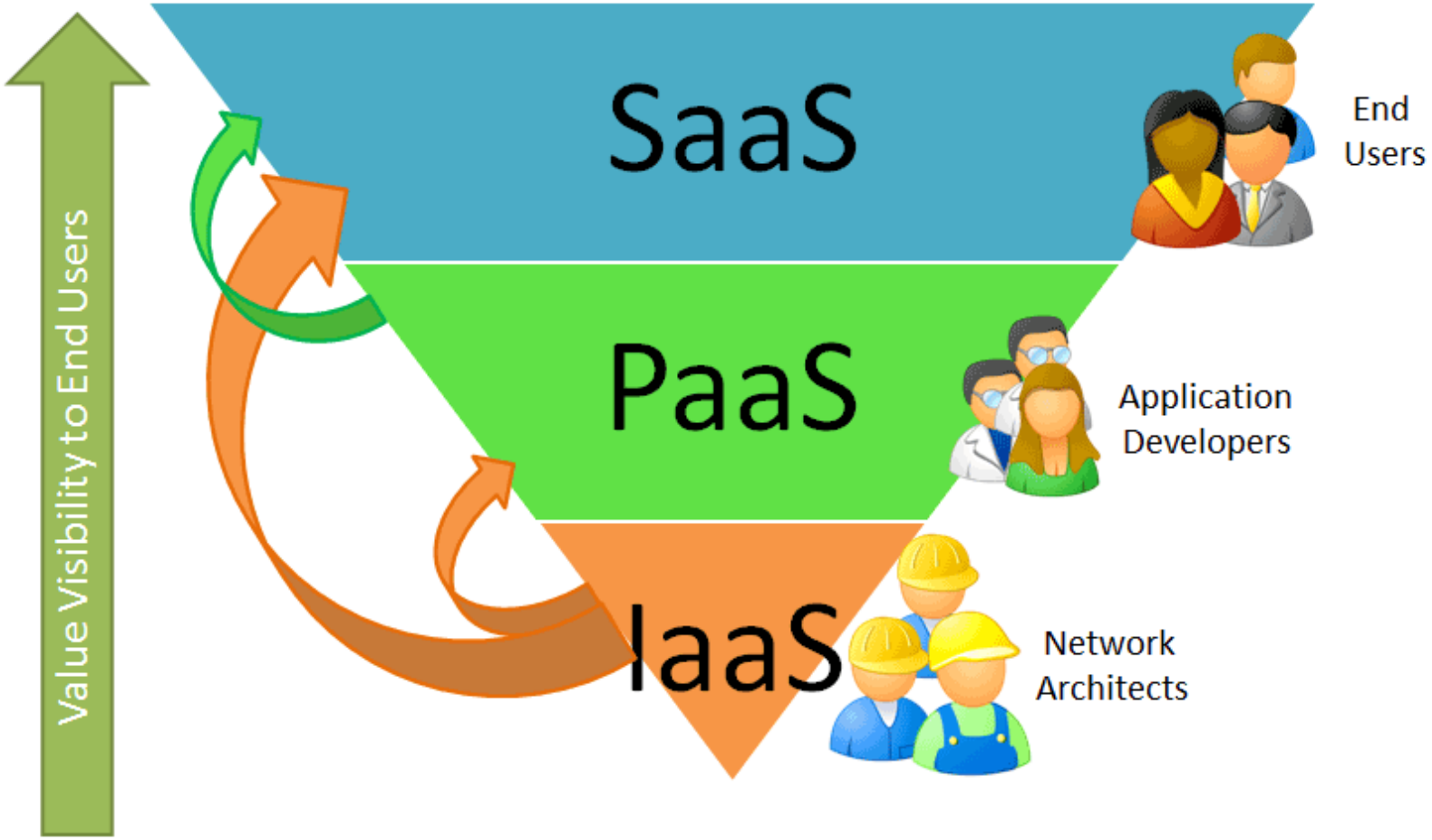
Techniques for Voluminous Data

- Cloud Computing is an efficient method to balance between dealing with voluminous data and keeping costs competitive, is designed to deliver IT services consumable on demand, is scalable as per user need and uses a pay-per-use model.
- Business houses are progressively turn towards retaining core competencies, and shedding the non-core competencies for on-demand technology, business innovation and savings.

Cloud Computing

- Cloud computing consists of hardware and software resources made available on the internet as managed third-party services.
- These services typically provide access to advanced software applications and high-end networks of server computers.
- Cloud computing is comparable to grid computing, a type of computing where unused processing cycles of all computers in a network are connect to solve problems too intensive for any stand-alone machine.

Cloud Computing Layers



Types of Cloud Computing

SaaS (Software as a Service): This is the idea of providing a given application to multiple tenants, typically using the browser which supports business applications of host and delivery type as a service. End Customers for instance Google Doc, Myspace.com

Common features of SaaS:

a) User applications run on cloud infrastructure

b) Accessible by users through web browser

c) Suitable for CRM (Customer Resource Management) applications

d) Supports multi-tenant environment

Types of Cloud Computing

PaaS (Platform as a Service): This is a variant of SaaS. You run your own applications but you do it on the cloud provider's infrastructure. Provides a comprehensive stack for developers to create Cloud-ready business applications.

Developers for instance Force.com, Google App Engine, Azure and Salesforce.com etc

Features of PaaS are:

a) Supports web-service standards

b) Dynamically scalable as per demand

c) Supports multi-tenant environment

Types of Cloud Computing

IaaS (Infrastructure as a Service): These are virtual storage and server options that organizations can access on demand, even allowing the creation of a virtual data center. Delivers computing hardware like Servers, Network, Storage, etc. For instance Rackspace.com, GoGrid.com etc.

Typical features are:

a)Users use resources but have no control of underlying cloud infrastructure

b)Users pay for what they use

c)Flexible scalable infrastructure without extensive pre-planning

Benefit of Cloud Computing

- Reduced Cost : Cloud technology is paid incrementally, saving organizations money.
- Increased Storage: Organizations can store more data than on private computer systems.
- Highly Automated: No longer do IT personnel need to worry about keeping software up to date.
- Flexibility: Cloud computing offers much more flexibility than past computing methods.
- More Mobility: Employees can access information wherever they are, rather than having to remain at their desks.
- Allows IT to Shift Focus: No longer having to worry about constant server updates and other computing issues, government organizations will be free to concentrate on innovation.

MapReduce

- MapReduce is a software framework that allows developers to write programs that process massive amounts of unstructured data in parallel across a distributed cluster of processors or stand-alone computers.
- It was developed at Google for indexing Web pages and replaced their original indexing algorithms and heuristics in 2004.

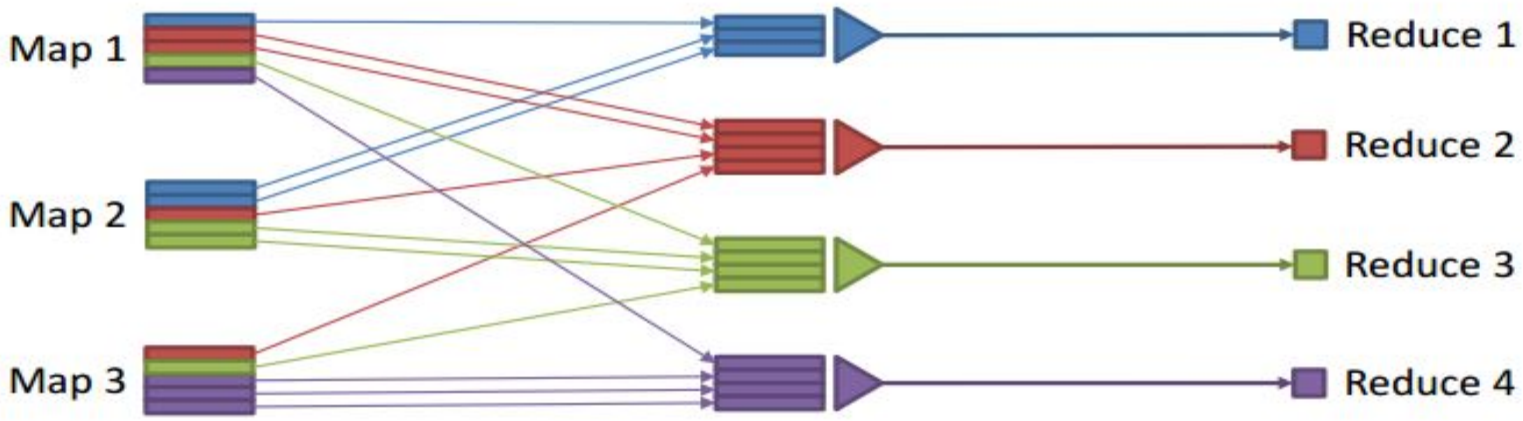
MapReduce and Hadoop Systems

- MapReduce is the heart of Hadoop.
- MapReduce allows data to be distributed across a large cluster, and can distribute out tasks across the data set to work on pieces of it independently, and in parallel.
- This allows big data to be processed in relatively little time.
- Apache has produced an open source MapReduce platform called **Hadoop**.

Framework is divided into two parts

- **Map**, a function that parcels out work to different nodes in the distributed cluster.
- **Reduce**, another function that collates the work and resolves the results into a single value.
- The MapReduce framework is fault-tolerant because each node in the cluster is expected to report back periodically with completed work and status updates. If a node remains silent for longer than the expected interval, a master node makes note and re-assigns the work to other nodes.

MapReduce Interaction



Map functions create a user-defined “index” from source data

- Reduce functions compute grouped aggregates based on index
- Flexible framework
  - users can cast raw original data in any model that they need
  - wide range of tasks can be expressed in this simple framework

Hadoop System

- Developed by Apache as an open source distributed MapReduce platform, based off of Google's MapReduce.
- Runs on a Java architecture framework that supports the processing of large data sets in a distributed computing environment.
- Hadoop allows businesses to process large amounts of data quickly by distributing the work across several nodes.
- Good for Big data sets and on large cluster.

Hadoop - A Key Business Tool

Hadoop System is used by Large Content-Distribution Companies, such as...

Yahoo

- Hadoop is used for many of their tasks, and over 25,000 computers are running Hadoop.

Amazon

- Hadoop is good for Amazon, they have lots of product data, as well as user-generated content to index, and make searchable.

New York Times

- Hadoop is used to perform large-scale image conversions of public domain articles.

Hadoop - A Key Business Tool

Used by Non-Content-Distribution Companies, such as

- Facebook
- eHarmony

Other early adopters include anyone with big data:

- medical records
- tax records
- network traffic
- large quantities of data

Wherever there is a lot of data, a Hadoop cluster can generally process it relatively quickly.

Data Management in the Cloud

- Data management applications are potential candidates for deployment in the cloud
  - Industry: enterprise database system have significant up-front cost that includes both hardware and software costs
  - Academia: manage, process and share mass-produced data in the cloud
- Many “Cloud Killer Apps” are in fact data-intensive
  - Batch Processing as with MapReduce
  - Online Transaction Processing (OLTP) as in automated business applications
  - Offline Analytical Processing (OLAP) as in data mining or machine learning

Data Management in the cloud

- A database system must implement for it to run well in the cloud, in potential database applications to consider for cloud deployment.
- Data management applications are best suited for deployment on top of cloud computing infrastructure.
- Data management is the proper management of a data resource for an organization. Data management consists of a set of theories, concepts, principles, and techniques for properly managing data.
- The primary objective is to support the business information as needs of the organization.

Data Management in Cloud

There are three characteristics of a cloud computing environment.

- Compute power is elastic, but only if workload is parallelizable.
- Data is stored at untrusted host.
- Data is replicated, often across large geographic distances

Information Retrieval

- **Information retrieval** is the activity of obtaining information resources relevant to an information need from a collection of information resources.
- Searches can be based on metadata or on full-text indexing.
- Automated information retrieval systems are used to reduce what has been called “information overload”.

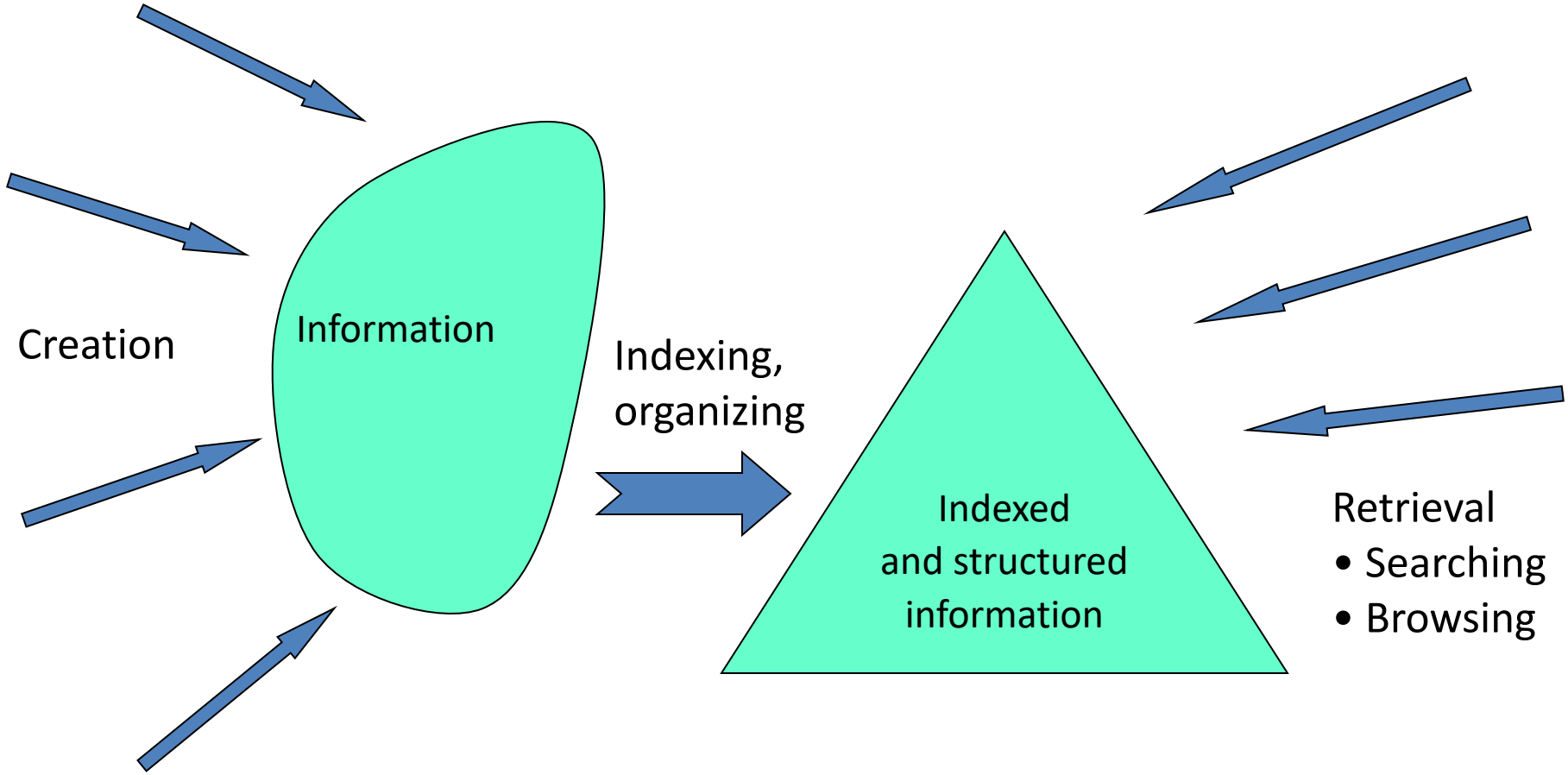
Information Retrieval in the Cloud

- IR user seeks actively information, pulling at it, by means of querying or browsing.
- In tag querying, user enters one or more tags in the search box to obtain an ordered list of resources which were in relation with these tags.
- When a user is scanning this list, the system also provide a list of related tags (i.e. tags with a high degree of co-occurrence with the original tag), allowing hypertext Browsing.

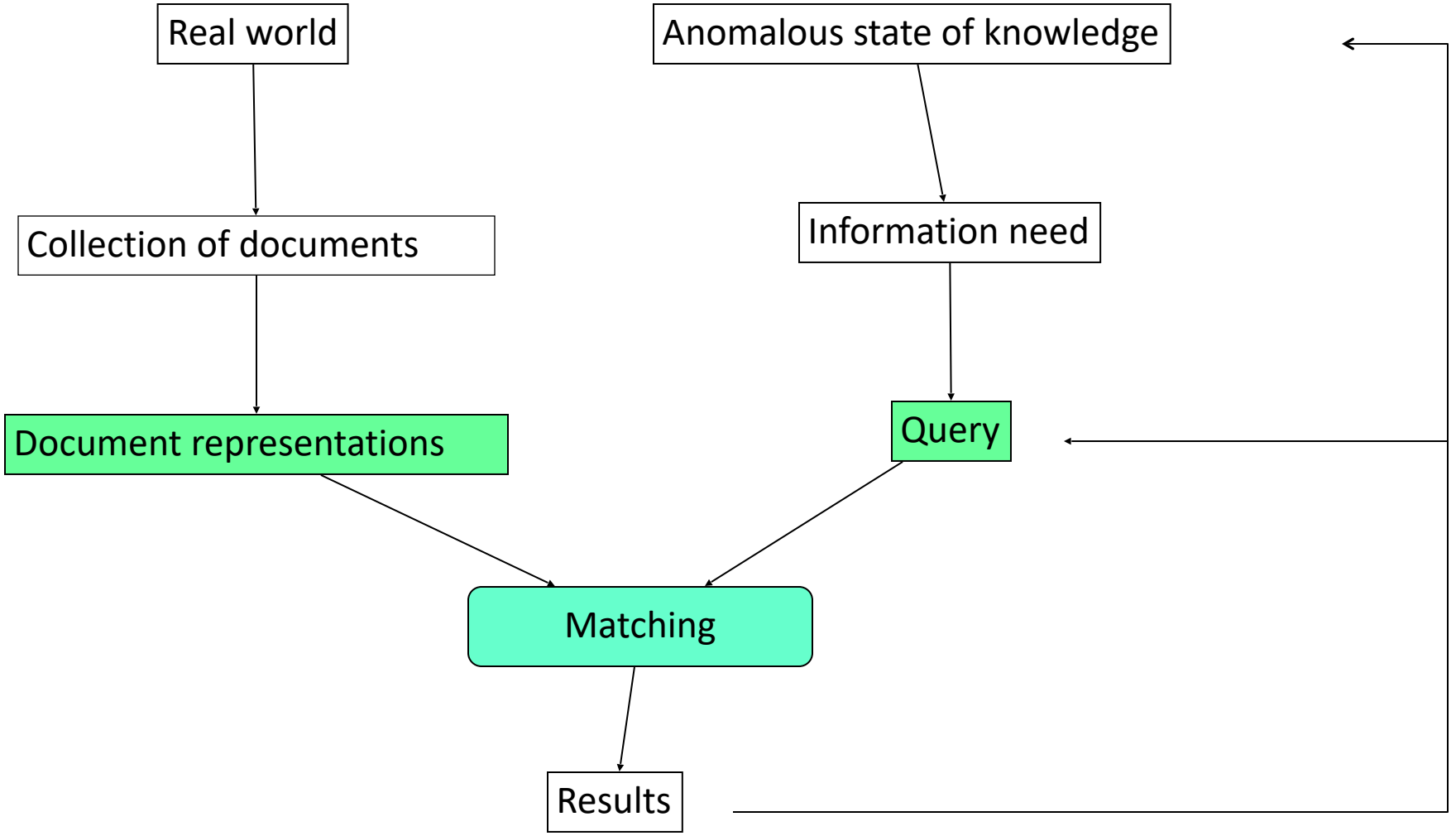
Information Retrieval System

- Typically it refers to the automatic (rather than manual) retrieval of documents
  - Information Retrieval System (IRS)
- Information Retrieval is a research-driven theoretical and experimental discipline
  - The focus is on different aspects of the information–seeking process, depending on the researcher's background or interest:
    - Computer scientist – fast and accurate search engine
    - Librarian – organization and indexing of information
    - Cognitive scientist – the process in the searcher's mind
    - Philosopher – Is this really relevant ?

The stages of IR



The formalized IR process



Stopwords / Stoplist

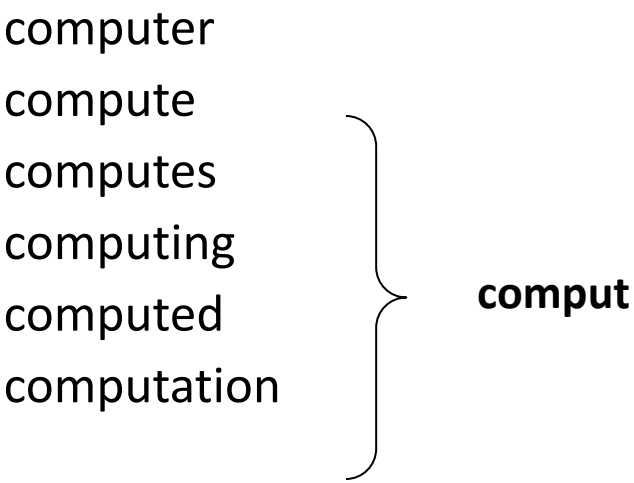
- Function words do not bear useful information for IR

  of, in, about, with, I, although, …
- Stoplist: contain Stopwords, not to be used as index
  - Prepositions
  - Articles
  - Pronouns
  - Some adverbs and adjectives
  - Some frequent words (e.g. document)

- The removal of stopwords usually improves IR effectiveness
- A few “standard” stoplists are commonly used.

Stemming

- Reason:
  - Different word forms may bear similar meaning (e.g. search, searching): create a “standard” representation for them
- Stemming:
  - Removing some endings of word



Link Analysis in Cloud Setup

- The web is not just a collection of documents – its hyperlinks are important!
- A link from page *A* to page *B* may indicate:
  - *A* is related to *B*, or
  - *A* is recommending, citing, voting for or endorsing *B*
- Links are either
  - referential – *click here and get back home*, or
  - Informational – *click here to get more detail*
- Links effect the ranking of web pages and thus have commercial value.

- See More on Chapter 7 for Link Analysis

Thank you