Chapter 1 Introduction to Information System

1.1 Classification and Evolution of Information System

What is Information System?

- Information system is the system that arranges data, processes, information technology and manpower of an organization so as to collect, process, store and provide the information that helps to sustain and support organizational progress in the competitive world.
- It converts the raw data within the organization into the valuable information needed for organizational goal.

What is data and information?

- Data is the raw fact that has no any significance and use if not processed.
- Data is processed to generate information.
- Information is those processed data that provides some meaning and can be used for decision making and organizational improvements.

Types of information

1. Operational Information:

- Those information that are needed to perform the daily operations of an organization are called operational information.
- Eg: daily account of sales and stocks, low stock item list, profit and loss account and so on.
- These information are used by line managers to facilitate proper survival of the organization by taking appropriate decisions that would facilitates the user's expectations and needs.

2. Tactical Information:

- Those information that are used to generate tactics to facilitate improvement in profit and performance of an organization.

- It is generally collected on weekly or monthly basis.
- Eg: Whether to provide discount on some stocks.
- It is used by middle managers.

3. Strategic Information:

- Those information needed for long term improvement and planning are called strategic information.
- It includes information on organizational expansion and new opportunities.
- Such information is generally unstructured.
- It is taken on quarterly or yearly basis.
- It is used by the top level managers.
- It helps in decisions like whether to open a branch, whether to initiate customer cards, etc.

4. . Statutory Information:

- Those information that are needed to provide to the government about the organizational structure and status are called statutory information.
- Eg: VAT filling, audits and so on.

Types of Information System

1. Transactional Processing System (TPS):

- TPS is the information system that is responsible for collection and processing of data related to daily transactions within an organization.
- It provides automation of the transactions maintained manually in an organization.
- It is the system that directly deals with the customers of an organization because transaction takes place between the organization and the customers.
- This system provides huge volume of information than other systems.
- It is generally used by the low level staffs who interacts with the customers directly.

2. Management Information System (MIS):

- MIS is an information system that aids the management of an organization to successfully coordinate with other members to develop broad long term vision of an organization.
- It provides various reports of an organization that helps to plan for the proper management system within an organization.
- It is designed with the motto that a proper management within an organization always leads towards success.
- It provides information that improves the management strategies of an organization.
- Eg: Observation of the sales and demands using statistical analysis and control production on that basis to meet the user demand.

3. Decision Support System (DSS):

- DSS is the information system that is responsible to generate information required to take decisions for long term strategic plans within an organization.
- It uses data from various internal sources (sales, production, finance) as well as external sources (population trend, raw material, pricing).
- Since decision making affects the future of an organization, there must be precise collection of data for information generation which is one of the most tedious task.
- It uses various tools like statistical analysis, simulation models, data mining, artificial intelligence and so on.

4. Expert System (ES):

- ES is the extended decision support system that collects the knowledge from the expert of the respective domain and then simulates that collected knowledge base to the one who do not even have the domain knowledge.
- It is the application of artificial intelligence that helps in improvement of the decision making process.

5. Office Information System (OIS):

- OIS is the information system that is responsible to provide the appropriate information to appropriate person at appropriate time.

- Information should be obtained by all the members of an organization as well as the customers but the information that each one obtain may differ based on their necessity to emphasize their work flow.
- This managed information flow is aided by OIS.
- It uses various technologies like electronic form, office automation, electronic messaging and so on.

6. Personal and Work Group Information System:

- - Personal Information System is the information system that is designed to meet the needs of a single user so as to boost the productivity of an individual.
- - Work group information system is the information system that are designed to meet the needs of a work group so as to boost the productivity of a group.

Classification of Information System

Based on Breath of Support:

- This classification of information system is based on how much support one gets from an organization structure.
- It includes functional, enterprise wide and inter-organizational information system.
- Functional system is organized within the traditional departments.
- Enterprise information system is organized over several departments of an organization or within overall enterprise.
- Inter-organizational information system connects two or more organizations.

Based on Organizational Level

- This classification includes clerical level, operational level and knowledge-work level information system.
- Clerical level includes the information that are collected by the workers through direct communication with the customers and supports the managers.
- Operational level includes the line managers that carry out daily operations of an organization in an efficient manner.

- Knowledge-work level includes the advisors and top level managers that are domain experts, who are responsible to integrate the information and knowledge for organizational goals.

1.2 IS in Functional Areas

The information system is used in following functional areas:

1. Production
2. Marketing
3. Material purchase and store
4. Accounting
5. Education
6. Human Resource Development
7. Research and Development

1.3 Information System Architecture

The architecture of information system consist of five major components:

1. Hardware
2. Software
3. Database
4. Network
5. People

- Hardware consists of input and output devices, processor, operating system and media devices that facilitate the functioning of an information system. The different hardware helps to collect the raw data and store the data as well as information.
- Software consists of various programs and procedures that eases and manages the information system. It basically helps to manage the raw data and useful information. It is responsible for conversion of raw data into useful information.

- Network consists of various hubs, communication media and network devices. It is responsible for communication and transparency of the information flow within an organization.
- Database consists of data organized in the required structure. It is responsible for structuring of data and information of an information system of an organization.
- People consists of device operators, network administrators and system specialists as well as managers who are responsible to make proper use of information system for organizational progress.

1.4 Qualities of Information System

1. The information generated by the system must be accurate. For information accuracy, the input raw data should be correctly chosen and should be processed properly as per the necessity of the organization.
2. The information system should be able to integrate all the necessary data and generate the information needed by the whole areas of the organization.
3. The information should be generated on time and should be provided to the right people at right time.
4. All the information should be revealed to the right people even if it is bad because such information can lead to future improvements of an organization.

5. The information should be customer oriented so as to meet the customer demands and gain trust from the customer for the sake of organizational growth.
6. The information system must have all the data from the past up to the present time so as to generate the fruitful results.
7. The information system should use attractive and noticeable formats like graphs to represent the results that help in quick and better understanding about the organization.

1.5 Managing Information System Resources

- Information system consists of many resources and one of them is supply chain.
- Supply chain is the description of how the materials, information, finance and services flow from the raw materials suppliers to the organization to the end customers using services generated by the organization.
- Generally, information system uses software solutions to manage the information system resources.
- The software solutions for managing supply chain activities includes:

1. Enterprise Resource Planning (ERM)
2. Supply Chain Management (SCM)
3. Customer Relationship Management (CRM)

1.6 Balanced Scorecard - Case Studies

- Balanced scorecard is the document that contains all the information needed for an organization required for its success from organizational goals to strategic planning.
- It helps to keep the organization balanced from all perspective like customer perspective, financial perspective, strategic perspective and future growth perspective.
- It should be cascaded to all the units of an organization to ensure successful strategic implementation.
- It analyzes the mission and vision of an organization so as to develop necessary strategic themes and strategy map to ensure those themes.
- The strategic initiatives are followed to achieve the vision of an organization taking care of the performance measures and meeting the customer demand for growth of an organization.

Benefits of Using Balanced Scorecard

1. 1. Better strategic planning:

- Balanced scorecard provides a powerful framework to build and communicate appropriate strategy based on the organizational needs and goals.
- Through strategy map, consensus is reached over a set of interrelated strategic objectives.
- It ensures performance of current state and identification of future performance for the strategy.

2. Improved Strategy Communication and Execution:

- Balanced scorecard contains the strategy and all its interrelated objectives in a single paper that allows communicating the organizational strategy within the organization as well as outside of the organization.
- The operational units and managers can get the greater picture of the organizational strategy and can easily understand what needs to be implemented.
- This makes the staffs to implement the strategy in an efficient way as it is already understood by all.
- The managers hence engage the staff for the strategic execution and can manage the work flow in an efficient way.

3. Better Management Information:

- Balanced scorecard ensures design of key performance measures for various strategic objectives ensuring organization to measure what really matters to an organization.
- It ensures that the organization gets quality management information which when used can guide management process and decision making process that eventually affects the organizational goal.

4. Improved Performance Reporting

- Balanced scorecard provides the strategic planning and implementation details.
- This helps to design the appropriate performance measures required to meet the strategic objectives.
- The proper choice of performance measures improves the performance reports in a transparent manner.

- Such performance reports are communicated both internally and externally so as to flow information to all and gain feedback for performance improvement.

5. Better Strategic Alignment:

- Balanced scorecard provides the strategic objectives to an organization.
- Cascading the balanced scorecard to all the organizational units helps to create link between strategy and operation.
- This ensures that all the organizational units are working towards the common strategy.

6. Better Organizational Alignment:

- Balanced scorecard provides foundation for the strategic planning.
- It helps to align the organizational processes such as budgeting, risk management and so on with strategic priorities.
- This helps to make the organization strategy focused.

From Perspective to Strategic Objectives to Performance Measures

Perspective	Strategic Objectives	Performance Measures
Financial	Increase shareholder value Increase profit generated by each salesperson	Net income Return on assets Return on sales Return on equity Product cost per unit Customer cost per unit Profit per salesperson
Customer	Acquire new customer Retain customer Develop profitable customers	Number of new customers Percentage of customer retained Customer profitability
Internal Business Process	Improve Manufacturing quality Introduce new products Minimize invoice error rate On time delivery by suppliers Increase proprietary products	Percentage of error free invoice Percentage of defective product units Percentage of on time delivery Number of patents
Learning and Growth	Increase information system capabilities Enhance employee skills	Percentage of processes with real time feedback Employee turnover rate Average job related training hours per employee

Chapter 2 Control, Audit and Security of Information System

2.1 Audit of Information System

- Information audit is the element of information management that is responsible to ensure that the information within an organization is managed and well organized.
- Information audit is the process to discover, monitor, analyze and evaluate the information flow within an organization so as to implement, maintain and improve the organizational information management.

Benefits of Information Audit:

- It examines the information against the criteria under the identified purpose of the audit to meet the standard compliance.
- It determines the user information needs.
- It lists the information resources available within an organization.
- It identifies the costs and benefits of the information resources available.
- It provides information about the working structure of the information system of an organization.
- It produces report that recommends for the information handling problems.
- It helps organization to make use of information for strategic planning and implementations.
- It aids in decision making and support.
- It enables organization to be dynamic i.e. adapt to necessary changes.
- Information audit helps to identify problems like data redundancy, duplication, inconsistency and cost to store and utilize data and information.
- Information audit helps to identify hidden assets of an organization, skills and expertise of staffs, market for further expansion and so on that would expand organizational opportunities.

Methodological Approach to Information Audit

1. Cost Benefit Method:

- It lists information system options and compares them on the basis of their cost and perceived benefits.

2. Geographical Approach:

- It identifies the components of information system and maps them in relation to one another to identify and meet system needs.

3. Hybrid Approach:

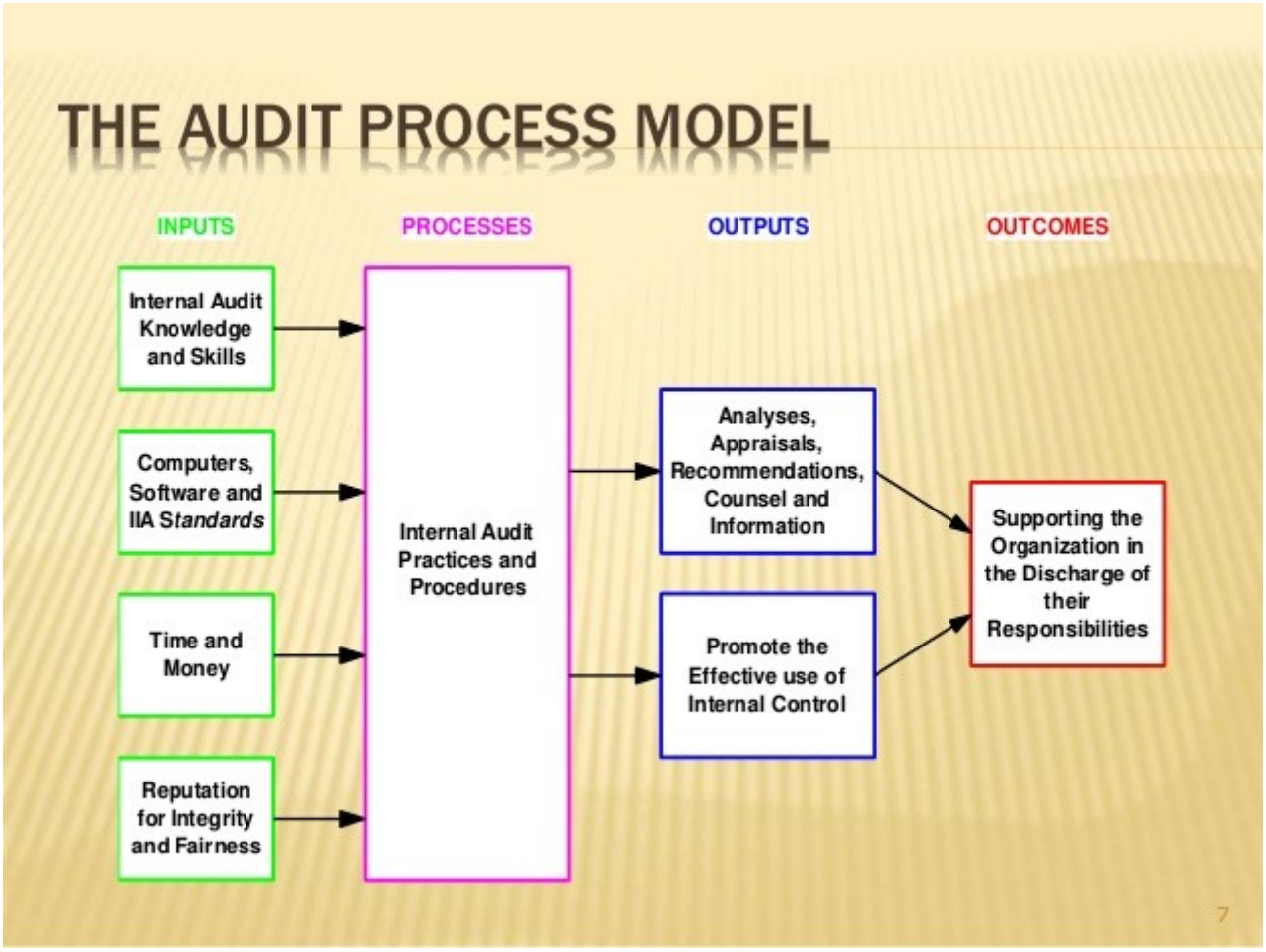
- It takes both geographical approach and cost benefit method into consideration.
- It emphasizes on the control and management procedures for organizational strategy.

4. Management Information Audit:

- It focuses on reports related to the management information.

5. Operational Information Audit:

- It focuses on the efficiency and effectiveness with which information resources are used and accounted for.
- It measures reliability of information system and compliance with obligations, regulations and standards.



Problems in Information Audit

- Support of senior management is very crucial for information audit and in most cases such support is not provided.
- It is difficult to decide whether to use internal auditors or external consultants.
- It is very tedious task to collect and gather necessary information for auditing.
- The information audit time span depends up on the size of an organization.
- It is difficult to establish costs and value of information

2.2 Security of Information System

Threats to Information System:

- Information system is subjected to threats because it is easy and safe to get information which can benefit for the hackers.

- It is very easy to get information from the system if the system is not properly secured.
- Such attacks are less prone to tracing as there is no necessity of physical presence.
- The various threats to information system are as follows:

1. Phising
2. DDOS Attack
3. Spyware
4. Key Logging
5. Man-in-the middle attack
6. SQL Injection
7. Session Hijacking
8. Payload
9. Identity Theft

2.3 Layered Security Strategy

Layered Security

- Layered security is the practice of combining multiple mitigating security controls to protect the resources and data.
- It assures that the information possessed by an organization is not compromised with the attackers.
- Multiple levels of security is provided to make the information system secure from various cyber threats.

Consumer Layered Security Strategy:

1. Extended Validation SSL Certificate
2. Two factor Authentication
3. Single sign-on
4. Fraud detection
5. Secure web

Enterprise Layered Security Strategy:

1. Workstation application whitelisting
2. Workstation system restore solution
3. Workstation and network authentication
4. File, disk and removable media encryption
5. Remote access authentication
6. Network folder encryption
7. Secured end-to-end messaging
8. Content control and policy based encryption

2.4 Extended Validation and SSL Certificate

SSL

- SSL stands for Secured Socket Layer.
- It provides a secure transport connection between web server and client.
- It enhances security of data transfer by developing a dedicated communication path between server and client through which encrypted data are communicated between client and server.
- Generally, SSL is used to transfer sensitive data which when compromised can lead to disastrous results.
- Such data includes log in details, credit card details, payment details and so on.
- The communication path itself is encrypted such that no one can listen to that path expect for the client and server certified by the SSL certificate.
- The website secured using SSL certificate gets https instead of http.
- The SSL certificate consists of server name, its public key, IP number, and an expiration date.
- The certificate is signed with a 1024 bit key by the certificate authority.
- All traffic within the SSL communication is encrypted.

Working of SSL:

- The browser attempts to connect to a web server secured with SSL by requesting for the web server's identity.
- The web server then sends the browser a copy of its SSL certificates.
- The browser checks whether or not it can trusts the SSL certificate. If so, it sends a message to the web server.
- The web server then sends back a digitally signed acknowledgement to start an SSL encrypted session.
- The encrypted data is then shared between the browser and the web server via a dedicated and encrypted communication channel.

Extended Validation:

- Extended validation is a certificate used for HTTPS websites and software that proves the legal entity controlling the websites or software package.
- To obtain EV certificate, verification of the requesting entity's identity is required by a certificate authority.
- It increases the security due to the identity validation process, which is indicated within the certificate by the policy identifier.

2.5 Remote Access Authentication

- Remote access authentication is the process by which a certified computer user can securely have network access and privileges even if the network is geographically separated.
- It makes use of the digital certificate that contains information to identify the user to the server and provides the credentials.
- The remote connection should be established to the network by the user as the network and the user computer are not physically connected to each other. Such remote connected is initiated by dial up connection or connection through Internet or connection through wireless medium.
- Once the credentials within the digital certificate is verified by the server, the server provides access to the remote computer to access its resources and services.

Steps to obtain remote access authentication:

1. Temporary network connection:

- To initiate remote access to the server, the user computer must establish a temporary network connection to the server.
- It may be through dial up, Internet or wireless connection.
- The network connection is established securely using encryption for remote access protocols.
- Proper encryption of communication channel is established to prevent from hijacking of authorized sessions and authorized user's credentials.

2. Establishing proper privileges:

- Once the network connection is successful, proper privileges to the requesting user should be established.
- For this, three steps are performed namely authentication, authorization and accounting.
- The server provides user identification to its user, computer or network device to enable remote access.
- Authentication is the process by which the server checks whether the user which requests for the remote access is eligible for the purpose or not by matching the user supplied identification credentials to the stored credentials.
- Authentication is generally accomplished via username and password system. Other methods include tokens, static biometrics like fingerprints and dynamic biometrics like voice matching.
- Authorization is the process of determining the permissions that are granted to the authenticated users. It determines whether an authenticated user has permission to use a particular resource of the server or not.

Telnet:

- Telnet is the TCP/IP protocol standard that allows users to log on remotely and access resources as if the user had a local terminal connection to the server.
- The major threat of using Telnet is that it uses TCP/IP connection for information flow that has less security.
- The Telnet uses TCP port 23 for connection.

Secure Shell (SSH)

- SSH is a remote access system that provides a secure transport between machines using an SSH daemon at each end for secure login and secure file transfer.
- SSH provides higher level of security as it supports different encryption protocols, cryptographic host authentication and integrity protection.
- The authentication services are host based.

2.6 Content Control / Content Filtering

- Content filtering is the process of controlling what content is permitted to the user.
- It is generally used to restrict material delivered over the Internet via Web, email or other means.
- It determines what content to make available or what content to block.

Implementation of content filtering

The content filtering can be implemented in various ways as follows:

1. Browser based filter:

- It is implemented by using the third party browser extension.
- On implementing such filters, the browser blocks the restricted content to be displayed to the users.

2. E-mail filter:

- It is used to filter the information contained in the mail body or in the mail header.
- The e-mail header or body or attachments are classified into accepted or rejected using the predefined rules or through artificial intelligence.
- Only the accepted mails are shown to the users.
- In most cases, the rejected messages are sent to the spam for manual review from the users.

3. Client side filter:

- It is generally installed on each client computer to allow content filtering.

- Such filter can be managed, disabled or uninstalled by the user having administrative privileges on the system.

4. Content-limited ISP:

- It includes the ISP that offer access to only a portion of Internet content.
- Any users who pursue services from such ISP are subjected to restrictions.

5. Network based filtering:

- It is implemented at transport layer (transport proxy) or application layer (web proxy).
- It filters outbound as well as inbound information within a certain network.
- All the clients of such network should accept the protocols and are subjected to restrictions.

6. Search engine filters:

- Search engine itself provides the facility for safety filter.
- When safety filter is activated, the inappropriate links that appears in the search lists are filtered out.

2.7 Policy Based Encryption

- Policy based encryption is the service that allows customers to set up filters based on the content of the messages.
- The customers are able to set criteria for acceptance of the messages.
- The messages get encrypted only if they meet the defined criteria.
- All the messages to external recipients are first routed to the special gateway.
- The gateway checks the compliance of all the messages to policy settings.
- Based on the defined conditions or policies, the messages are encrypted, send to the receiver, discarded or returned to the sender.

2.8 Security in e-commerce transaction

1. Network security
2. Proper identification
3. Encryption

4. Authorization
5. Host and application security
6. Transaction security
7. Human error and malice
8. Communication security

Chapter 3 Enterprise Management System

3.1 Enterprise Management System (EMS)

Enterprise Information System

- Enterprise is the term that encompasses the large business community including all the members of that business.
- Enterprise incorporates various functions across the large business including human resources, engineering design, production, controlling, maintenance, sales, finance, quality and so on.
- Enterprise system is a large scale application software packages that support business processes, information flows, reporting and data analytics in complex organization.
- Enterprise system is designed to manage large volumes of critical data and to provide high levels of transaction performance and data security.
- Enterprise information system is the information system that allows companies to integrate information across operation on a company.
- Enterprise management system is an enterprise information system designed to coordinate all the resources, information and activities needed to complete business process.

Components of EMS

The components of enterprise management system are as follows:

1. Enterprise Resource Planning
2. Supply Chain Management
3. Customer Relations Management

Necessity of Enterprise System:

1. It is able to integrate all the functionality of an enterprise to achieve faster performance.
2. It provides automation on information system that helps to track the organizational status in real time.
3. It provides information to the right people at right time.
4. It helps in transparent flow of information across all the units of an enterprise.

3.2 Enterprise Resource Planning (ERP)

- An ERP system is an attempt to integrate all the functionality and operations across a company to a single computer system that can serve all those operations as per the necessity.
- ERP system integrates customers, suppliers to the enterprise operation.
- It intends to carry out all the business processes of an organization from a single control interface or computer following a set of best practices.

Features of ERP:

1. Financial management: (Provides financial functionality and analysis reports for different departments)
2. Human resource management (It provides functionality for personnel management, organization management, payroll management, time management, personal development)
3. Manufacturing (It provides functionality like bill of material, sales and distribution plan)
4. Supplier and purchase order management (It is integrated to supply chain process to control production planning)

Benefits of ERP:

1. It integrates all the data in a single source that makes the data and information to flow easily and transparently across the intended departments.

2. It is a real time system that considers customers as well as suppliers along with the internal processes as a part of the enterprise that helps to improve internal as well as external communication and understand about the demands on timely manner.
3. It helps in increasing productivity of an organization based on the customer current demands.
4. As all the information is integrated to a single computer system, it minimizes the operating costs of the information system.
5. It helps to trace the past and present business activities and lay foundation to improve those activities in future for achieving enterprise progress.
6. It helps the enterprise to survive in this competitive world by gaining trust from the customers as the customers are directly involved as a part of an enterprise.
7. It helps to make better use of available resources.

Challenges to ERP Implementation:

1. The ERP technical capabilities are limited.
2. The ERP system is inconsistent with existing business processes.
3. It increases implementation and maintenance costs of an enterprise.
4. The overall employee responsibilities should be changed based on ERP system which may be time consuming and difficult for the staffs as well as managers to adapt to.
5. Implementing ERP requires technical knowledge as well as resources. If the enterprise do not possess such personnel within the enterprise, it would cost more.
6. The enterprise strategy should be implemented and executed as per the ERP system.

3.3 Supply Chain Management (SCM)

- SCM deals with the control of materials, information and financial flow in a network of suppliers, manufacturers, distributors and customers.
- It helps to track goods across departments within an organization in real time.
- It is responsible for managing how and from where the raw materials will enter into the organization so as to produce a product; and how and to whom the produced services will be delivered.
- It is responsible for all the activities from production to end customers.
- Supply chain management helps to link the organization with the suppliers and the customers so as to meet the demand of the users through balanced supply of services as per the need.
- The supply chain includes a chain as follows:

  Supplier ======> Storage ==> Manufacturer ==> Storage ==>
  Distributor ==> Retailer ==> Customer

Features of SCM:

1. Process Customer Requirements (It enhances speed of customer requirement processing. It is responsible to check for availability of raw materials, product manufacturing and passing the product to the logistic team. The software tracks the entire process.)
2. Inventory management (It helps to effectively manage the quantity of stocked goods)
3. Purchase order processing (It reduces time and effort needed to generate and manage purchase orders)
4. Supplier relationship management (It is responsible for strategic planning and managing of all supplier interactions. The software is used to assess the supplier assets and capabilities so as to choose the best business strategy)
5. Warehouse Management (It helps to effectively support a warehouse management system for movement and storage of products.)

Functionality of SCM:

1. Sourcing of raw materials and component for a product or skills.
2. Manufacturing or creating a finished product or service.
3. Ensuring the product or service is transported, warehoused and is always available for the targeted users.
4. Timely delivery to the end consumer.

Challenges of SCM:

1. SCM is not able to perform all the functions that an organization hopes for.
2. Customer demands for the product immediately which is difficult to handle by the SCM.
3. The delivery of the product to the end customer is difficult for which an organization should collaborate with others.
4. If SCM system is not reliable, an organization may face problems to track about their stocks and further productions.

3.4 Customer Relationship Management (CRM)

- Customer Relationship Management refers to the practices, strategies and technologies that organization use to manage and analyze customer interactions and data throughout the customer lifecycle, with the goal of improving business relationships with customers, assisting in customer retention and driving sales growth.
- It is an approach to manage organization's interaction with current and potential customers.
- It uses data analysis about customer's history with a company.
- It collects information from different channels or points of contact between customer and the company such as website, telephone, live chat, mail, social media, marketing materials and so on.

- It also provides staffs who directly face customers with the detail information about customer, their purchase history, buying preferences and concerns so as to make the interaction more efficient and effective.
- It is used to build and manage customer relationships through marketing, observing relationships as they mature through distinct phases, managing these relationships at each stage and recognizing that the distribution of a value of a relationship to the firm is not homogeneous.

Features of CRM

1. 1. Marketing Automation:
- It is responsible to enhance marketing efforts to customers.
- For eg: the CRM system can automatically send the customers with the marketing materials via social media or email.
2. Sales Force Automation:
- It is responsible to track all the contacts and follow ups between customer and sales person so as to prevent duplicate efforts.
3. Contact Center Automation:
- It is responsible to automate the customer contact with the company.
- It can be implemented via a recorded audio or through chat bots.
4. Location Based Services:
- The CRM system include technology that can create geographic marketing campaigns based on customer's physical locations.

Benefits of CRM

1. It enhances ability to target profitable customers.
2. It integrates assistance across different channels.
3. It enhances sales force efficiency and effectiveness.
4. It helps in customizing products and services based on user demands.
5. It improves service to the customers.
6. It helps to connect customer and all the units of an organization together in a single platform.

7. It helps to improve the relationship of an organization with the customer that increases organizational reputation and trust among the customers.

3.5 Role of IS and IT in Enterprise Management

- Information systems help an organization to make adequate use of its data, reduce workload and assist with compliance with various mandatory regulations.

Information Storage and Analysis

- Through the adoption of information systems, enterprises can make use of sophisticated and comprehensive databases that can contain all imaginable pieces of data about the organization.
- Information systems store, update and even analyze the information, which the company can then use to pinpoint solutions to current or future problems.
- These systems can integrate data from various sources, inside and outside the company, keeping the company up to date with internal performance and external opportunities and threats.

Assist Decision Making

- An organization’s management team uses information systems to formulate strategic plans and make decisions for the organization's survival and progress.
- The analysis of data and comparison to market trends helps organizations to analyze the adequacy and quality of their strategic decisions.

Assist Business Processes

- Information systems helps businesses in developing a larger number of value added-systems in the company.
- For example, a company can integrate information systems with the manufacturing cycle to ensure that the output it produces complies with the requirements of the various quality management standards.

- The use of information systems simplifies business processes and removes unnecessary activities.
- Information systems add controls to employee processes, ensuring that only users with the applicable rights can perform certain tasks.
- Information systems eliminate repetitive tasks and increase accuracy, allowing employees to concentrate on more high-level functions.
- Information systems provides means for better project planning and implementation through effective monitoring and comparison against established criteria.

3.6 Enterprise Engineering

- Enterprise engineering is the discipline, principles and practices to design whole or part of an enterprise.
- It is an enterprise life cycle oriented discipline for the identification, design and implementation of enterprises and their continuous evolution.
- It examines each aspect of the enterprise such as business process, information flow, material flow and organizational structure.

Methodologies

The formal methodologies and methods used to offer organization reusable business process solutions are as follows:

1. Computer Integrated Manufacturing Open Systems Architecture (It provides templates to encode business, people and information technology of an enterprise requirements.)
2. Integrated Definition (It shows business process flows through a variety of decomposed business functions with corresponding information inputs, outputs and actors.)
3. Petri Nets (It is used to model manufacturing systems and provide formalisms for the modeling of concurrent systems with the ability to create simple state representation, concurrent system transitions and allow duration of transitions.)

4. Unified Enterprise Modeling Language (It is the object oriented enterprise modeling tool in which emphasis is placed on the usage of enterprise objects from which complex enterprise systems are made.)
5. Enterprise Function Diagrams (It is a modeling technique for the representation of enterprise functions and their corresponding interactions. It provides easy to use and detailed representation about a business process and its corresponding functions, inputs, outputs and triggers.)

3.7 Electronic Organization

- Electronic organizations are the organizations established and operated, based on the Internet and other related technologies in an environment referred to as Internet Culture.
- These organizations will be placing the Internet at the center of their business and encouraging ubiquitous use of network technologies.
- All the activities of electronic organization is based on the Internet.
- It includes developing strategies for running Internet based companies improving communication between employees, customers, and suppliers and collaborating with partners to electronically coordinate design and production.
- For example: The bus system that provides passengers access to information and booking over the Internet is an example of electronic organization.
- Electronic organizations affect four components of the management process: planning, organizing, leading and controlling.

Features of electronic organization:

1. It does not have central geographic location.
2. It interacts through computer networks and technologies.
3. It removes necessity of location specific services and face to face communication.
4. Management of production process is done through electronic links.
5. Support to customer is extended through electronic information system.

3.8 Loose Integration vs Full Integration

Loose Integration

- Loosely integrated system is one in which each of its components has, or makes use of, little or no knowledge of the definitions of other separate components.
- Components in a loosely integrated system can be replaced with alternative implementations that provide the same services.
- It provides a system that have less constraints to platform and environment.

Tight Integration

- Tight integration is the system that works on the inter connectivity and inter-processing of two or more systems simultaneously to deliver a cohesive/integrated solution.
- The entire logic is distributed across several hardware and software components, which all need to be operational and connected to deliver the business logic/process.
- For example, a bank ATM machine depends on the ATM machine hardware, built-in firmware/applications and the primary banking application to allow a customer to withdraw cash or access any ATM-specific services.

3.9 Process Alignment

- Process alignment is the activity of resolving differences between business processes.
- It means harmonizing components of one process between each other to reach their better collaboration and coordination.
- It makes sure that every component of a process fits to other components so they have no disagreement with each other, but they work together consistently

and are in synergy (coordination to provide combined outcome) towards process outcomes.

- It helps to provide the desired outcomes of a process to the organization on time and with quality.
- It comprises identifying ineffective areas in the processes and redesigning them to be more productive and less error prone.
- Opportunities for process alignment can be detected by the process of auditing.

Ways for process alignment:

1. Retooling of ineffective process areas.
2. Documenting the process requirements and workflows.
3. Education and training of the process staff.
4. Implementing the best practices and innovation.
5. Attaining a better operational visibility and measurability.

Chapter 4 Decision Support and Intelligent System

4.1 Decision Support System

- Decision support system (DSS) is a computer based information system that supports organizational decision making activities.
- A decision support system (DSS) is a computer program application that analyzes business data and presents it so that users can make business decisions more easily.
- It helps to make decisions about unstructured and semi-structured problems.
- A properly designed DSS is an interactive software-based system intended to help decision makers compile useful information from a combination of raw data, documents, and personal knowledge, or business models to identify and solve problems and make decisions.

Based on relationship with user

1. Passive DSS

- It is a system that aids the process of decision making but that can not bring the explicit decision solutions.

2. Active DSS

- It is a system that is able to bring explicit decision solutions for a problem.

3. Cooperative DSS

- It is a system that allows iterative process between human and system to achieve best solution.

Based on Mode of Assistance

1. Communication driven DSS

- It enables cooperation, supporting more than one person working on a shared task.

2. Data driven DSS

- It emphasizes access to and manipulation of a time series of internal company data and, sometimes, external data.

3. Document driven DSS

- It manages, retrieves, and manipulates unstructured information in a variety of electronic formats.

4. Knowledge driven DSS

- It provides specialized problem-solving expertise stored as facts, rules, procedures, or in similar structures.

5. Model driven DSS

- It emphasizes access to and manipulation of a statistical, financial, optimization, or simulation model.

Components of DSS

1. Database:

- It is a well organized collection of current and historical data from a number of applications and groups.
- It provides easy access to data and information from variety of applications.
- It is able to maintain data integrity.
- The data are extracted from various relevant databases and stored especially for decision support system only.

2. Model:

- A model represents an abstract representation of different components and relationships of a phenomenon.
- The model can be classified as:

  (a) Behavioral Model

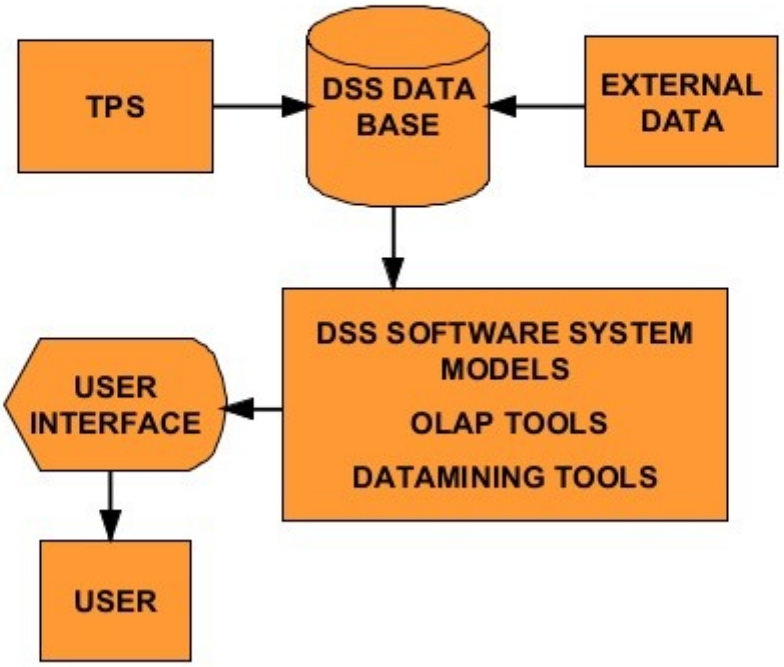
  (b) Management Science Model

  (c) Operation Research Model

3. User Interface:

- It allows interaction between the user of the system and the DSS database and model.
- The user interface is responsible to deal with the end user of the system providing them with friendly interface.

COMPONENTS OF DSS



Behavioral Model:

- It focuses on studying and understanding the trends among the variables.
- It tends to find the correlation and regression among the variables.

Management Science Model:

- It is based on the principle of management and accounting.
- It includes budget system, cost accounting, inventory management and so on.

Operations Research Model:

- It is based on the different mathematical formula.
- It represents the different real life problems depending on the various variables and parameters in the form of algebraic expressions.
- Eg: linear programming, material requirement planning, mathematical programming techniques and so on.

Analytical Model of DSS:

- It focuses on identifying the problem and implementing the solution.
- The steps in analytical model of DSS are as follows:

1. Problem identification
2. Define objectives
3. Pre-decision.
4. Generate alternatives
5. Evaluate alternatives
6. Make a choice
7. Implement choice
8. Evaluate choice

Difference between DSS and MIS:

DSS	MIS
• DSS provides support for unstructured or semi structured decision making. • DSS makes use of qualitative data for decision making. • DSS focuses on leadership and decision making. • DSS helps an organization to choose right path for its progress. • Flow of information is only upward in case of DSS.	• MIS provides support for structured decision making. • MIS makes use of quantitative data for decision making. • MIS focuses on information gathered and report planning. • MIS helps to achieve operational efficiency. • Flow of information is on both sides (up and down) in case of MIS.

4.2 Group Decision Support System (GDSS)

- Group Decision Support System is an interactive computer based system that facilitates a number of decision-makers (working together in a group) in finding solutions to problems that are unstructured in nature.
- GDSS takes inputs from multiple users interacting simultaneously with the systems to arrive at a decision as a group.
- It improves the quality and effectiveness of the group meetings.
- It reduces time of high level managers to reach at efficient decisions.

Advantages over DSS:

1. Better comprehensive consideration of the problems and the various relating issues.
2. Better group understanding of the problem.
3. Less likelihood of argue with the decision made.
4. Better group commitment to the decision.
5. Better communication to/with the implementers.

Components of GDSS

1. Hardware:

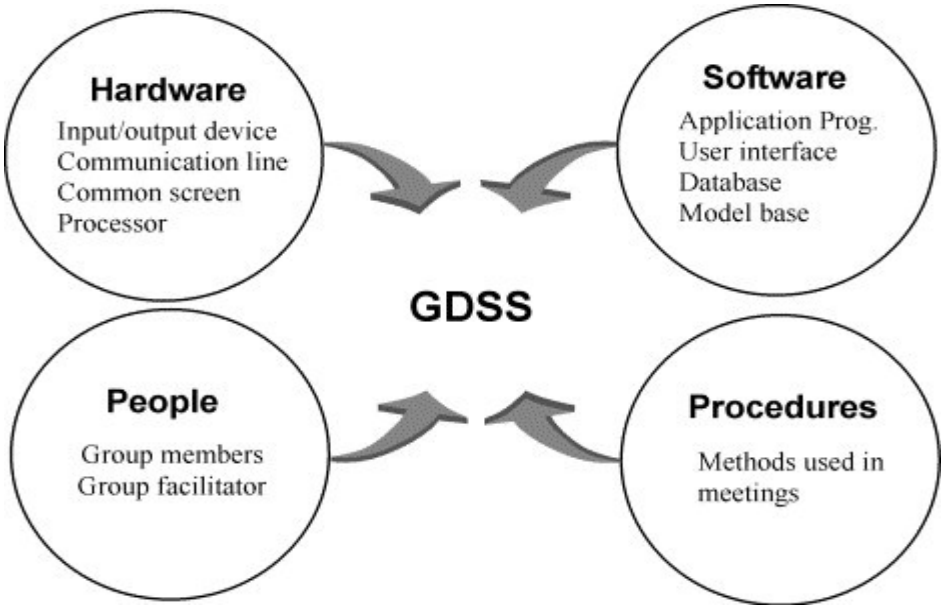
- It includes electronic hardware like computer, equipment used for networking, electronic display boards and audio visual equipment.
- It also consists of conference infrastructures.
- All these hardware facilitates the support for group decision support system.

2. Software Tools:

- It includes various tools and techniques that helps the decision makers to plan, organize ideas, gather information, establish priorities, take decisions and document the meeting proceedings.
- The tools and techniques consist of electronic questionnaire, idea organizer, priority setting tools, policy formation tools, modeling tools, dialogue manager, database management system and so on.

3. People:

- It includes members to participate in the meeting, a trained facilitator who helps with the proceedings of the meeting, and an expert staff to support the hardware and software.



Features of GDSS

1. It is easy and simple to use because of its interactive and user friendly user interface.
2. It facilitates users at different locations to make decisions as a group that results in better decisions.
3. The facilitator provides general support to the group and helps them to use the system.
4. It emphasizes on unstructured and semi structured decisions.
5. It supports all phases of decision making (intelligence, design, choice and implementation).
6. The participants are able to share their ideas more openly.

4.3 Enterprise and Executive Decision Support System (EDSS)

- Executive support system is a specialized decision support system that serves the information that is needed by the various top executives.
- It assists the top level executives in taking and performing various types of decisions.

- It also includes communication, office automation, analysis support and so on.

Characteristics of ESS:

1. Informational characteristics
   i. Flexibility and ease of use.
   ii. Provides the timely information with the short response time and also with the quick retrieval.
   iii. Produces the correct information.
   iv. Produces the relevant information.
   v. Produces the validated information.
2. User interface/orientation characteristics
   i. Consists of the sophisticated self help.
   ii. Contains the user friendly interfaces consisting of the graphic user.
   iii. Can be used from many places.
   iv. Offers secure reliable, confidential access along with the access procedure.
   v. Is very much customized.
   vi. Suites the management style of the individual executives.
3. Managerial / executive characteristics
   i. Supports the over all vision, mission and the strategy.
   ii. Provides the support for the strategic management.
   iii. Sometimes helps to deal with the situations that have a high degree of risk.
   iv. Is linked to the value added business processes.
   v. Supports the need/ access for/ to the external data/ databases.
   vi. Is very much result oriented in the nature.

Benefits of ESS:

1. Achievement of the various organizational objectives.
2. Facilitates access to the information by integrating many sources of the data.
3. Facilitates broad, aggregated perspective and the context.
4. Offers broad highly aggregated information.

5. User's productivity is also improved to a large extent.
6. Communication capability and the quality are increased.
7. Provides with the better strategic planning and the control.
8. Facilitates pro active rather than a reactive response.
9. Provides the competitive advantage.
10. Encourages the development of a more open and active information culture.
11. The cause of a particular problem can be founded.

4.4 Knowledge Management and Knowledge Based Expert System

- An expert system is an intelligent program that solves problems in a narrow problem area by using expert specific knowledge rather than an algorithm.
- It simulates the decision making process of a human expert in a specific domain.

Features

a) Reasoning capacity
b) Cope with uncertainty
c) Use of knowledge not data
d) Symbolic knowledge representation
e) Use meta knowledge
f) Use user interface

Role of Expert System in Information System

- Organization can realize benefits in consistency, accuracy, and reliability in problem-solving activities.
- It replicates the human expert knowledge which may be facts or heuristic information.
- It provides faster and accurate results than the human expert if designed properly.

- An organization should face significant loss in losing an expert. Such system overcome this fact.
- Improvements in reliability and quality frequently appear when expert systems distribute expert advice, opinion, and explanation on demand.
- It is able to handle enormously complex task without loosing its quality of output.
- Expert systems can reduce production downtime and, as a result, increase output and quality.

4.5 AI Neural Networks, Virtual Reality and Intelligent Agents

Virtual Reality

- Virtual reality (VR) is a computer technology that generate realistic images, sounds and other sensations that simulate a user's physical presence in a virtual or imaginary environment.
- With the help of virtual reality in information system, the users can quickly get the broader view of the problem, the alternative solutions and the effects of those solutions in the future trends.

Areas of Information System that uses Artificial Intelligence

1. Decision Support System
2. Strategy Formation
3. Business Automation
4. Human Resource Development

4.6 Data Mining and Data Warehousing

What is Data Mining?

- Data mining is the process of finding patterns in the given set of data.
- It is also defined as the process of extracting information from huge sets of data.

- Such patterns generally provides some meaningful information to the intended users.
- It uses statistical analysis.
- The applications of data mining are as follows:
    1. Market analysis
    2. Fraud detection
    3. Customer retention
    4. Production control

What is Data Warehousing?

- Data warehousing is the process of aggregating data from multiple sources into a common repository.
- Data warehouse is a database which is kept separate from the operational database, which basically consists of consolidated historical data.
- A data warehouse helps executives to organize, understand, and use their data to take strategic decisions.
- The features of a data warehouse are as follows:
    1. Subject oriented
    2. Integrated
    3. Time variant
    4. Non-volatile

Relation of Data Mining and Data Warehousing:

- Data warehousing is the process in which the data from different sources are aggregated to a single database.
- Data mining is the process of analyzing the data stored in the data warehouse to generate some meaningful patterns.
- Data mining is the process that is incomplete without data warehousing.
- The general flow of data analysis is:

    Data from different source ====> Data warehouse =====> Data mining =====> Useful patterns

Necessity of Data Warehousing

A data warehouses is kept separate from operational databases due to the following reasons:

1. An operational database is constructed for well-known tasks and workloads such as searching particular records, indexing, etc. In contract, data warehouse queries are often complex and they present a general form of data.
2. Operational databases support concurrent processing of multiple transactions. Concurrency control and recovery mechanisms are required for operational databases to ensure robustness and consistency of the database.
3. An operational database query allows to read and modify operations, while a data warehouse query needs only read only access of stored data.
4. An operational database maintains current data. On the other hand, a data warehouse maintains historical data.

4.7 OLAP and OLTP

On-Line Transaction Processing (OLTP):

- Online Transaction Processing is a information system type that prioritizes transaction processing, dealing with operational data.
- These systems gather input information and store them on a database, in a large scale.
- he main emphasis for OLTP systems is put on very fast query processing, maintaining data integrity in multi-access environments and an effectiveness measured by number of transactions per second.
- Eg: A banking transaction system

On-Line Analytical Processing (OLAP):

- Online analytical processing is a computer technology term referring to systems focused on analysing data in a specific database.
- Example: In a hospital there is 20 years of very complete patient information stored. Someone on the administration wants a detailed report of the most common deseases, sucess rate of treatment, intership days and a lot of relevant data. For this, we apply OLAP operations to our data warehouse with historical

information, and throught complex queries we get these results. Then they can be reported to the administration for further analysis.

Differences:

	OLTP System Online Transaction Processing (Operational System)	OLAP System Online Analytical Processing (Data Warehouse)
Source of data	Operational data; OLTPs are the original source of the data.	Consolidation data; OLAP data comes from the various OLTP Databases
Purpose of data	To control and run fundamental business tasks	To help with planning, problem solving, and decision support
What the data	Reveals a snapshot of ongoing business processes	Multi-dimensional views of various kinds of business activities
Inserts and Updates	Short and fast inserts and updates initiated by end users	Periodic long-running batch jobs refresh the data
Queries	Relatively standardized and simple queries Returning relatively few records	Often complex queries involving aggregations
Processing Speed	Typically very fast	Depends on the amount of data involved; batch data refreshes and complex queries may take many hours; query speed can be improved by creating indexes
Space Requirements	Can be relatively small if historical data is archived	Larger due to the existence of aggregation structures and history data; requires more indexes than

		OLTP
Database Design	Highly normalized with many tables	Typically de-normalized with fewer tables; use of star and/or snowflake schemas
Backup and Recovery	Backup religiously; operational data is critical to run the business, data loss is likely to entail significant monetary loss and legal liability	Instead of regular backups, some environments may consider simply reloading the OLTP data as a recovery meth

Chapter 5 Planning for IS

5.1 Information System Planning

Necessity of IS Planning

- Information system environment is very complex. So, planning for information system is very much crucial to get most out of it.
- The plan describes the structure and content of the information system and how it is to be developed.
- The information system plan should be based on the organizational strategic plan. It helps in design of suitable information system that would assist in achieving organizational goals.
- It also helps to understand the current status of the organizational information system and what can be done to improve the information system to achieve reliability and quality.
- Information system planning is done by Chief Information Officer.

5.2 Strategic Information System

- Strategic information system is the information system that is responsible to implement the business strategy for organizational progress.
- Such system applies the information system resources to strategic business opportunities in such a way that the computer system affect the organizational products and activities.
- It links business and computer strategies.
- It shapes an organizational competitive strategy.
- It helps an organization to gain a competitive advantage and meet other strategic enterprise objectives.

Benefits of Using Strategic Information System

1. Creating barriers to the competitor's entry:

- It helps to provide a product that is difficult to duplicate.
- This makes the competitors stop competing with the organizational products and services.

2. Database to improve marketing:

- It provides all the previous data about sales through a database.
- It makes use of historical sales and marketing data to find improvement in future market.

3. Customer and Supplier Retention:

- It focuses on the customer and supplier providing them with highly reliable services such that they do not switch over to a competitor.
- It helps in maintaining proper relationships with the customer.
- It provides customers with the dedicated and quality services on time that helps to generate trust towards the organization among the customers.

4. Identify Marketing and Sales Strategy:

- It helps to analyze the historical data on sales and marketing so as to determine new strategy for market improvement.
- It may includes lowering of cost of the materials or services as per the demand so as to increase the customers.

5.3 Tactical Information System

- Tactical information system is the information system that is responsible for planning short range tactics emphasizing the current operation of an organization.
- It helps to know how the organization, in the short time period, can improve its organizational operations.
- Tactical plans are usually developed in the areas of production, marketing, personnel, finance and plant facilities.

5.4 Operational Information System

- Operational information system is the information system that is responsible to aid the daily operations of an organization.
- It focuses on how the organization can improve its daily operations.
- It processes operational information to plan for operational progress.

Chapter 6 Implementation of Information System

6.1 Change Management in Information System

Change Management

- On implementation of information system, many changes may be resulted in an organization.
- Some of the areas that the information system affects are as follows:
  1. Organizational structure
  2. Centralization of Authority
  3. Job content
  4. Relationships
- Change management is the organizational process that analyze and define all the changes faced by an organization and develop programs to reduce the cost and risk and to maximize the benefits of the changes.
- It is a structured approach to facilitates the adoption of new changes within the system.

Importance of Change Management in IS Implementation

- Information system is the dynamic aspect of an organization. Changes are necessary over time.
- The information system's reliability is based on the new technologies for which changes are important to be made.
- Change management helps to manage the possible changes without affecting the current operation of an organization.
- Change management also helps to adopt changes in an efficient manner for organizational progress.

- Change management detects all the possible problems that may be faced by an organization on adopting to new changes and develop strategies to overcome effects of those changes.

Change Management Tactics to be applied:

- Change management process is the sequence of steps that is to be followed to take most out of the changes faced by an organization.
- The tactics to be applied for a successful change management process are as follows:

1. Preparing for the change:

- It involves the process of understanding the problem and the need to change.
- It helps in understanding the environmental factors affecting the business.

2. Managing change:

- The vision is created for the necessary changes to be applied.
- The possible resistance to changes are identified.
- A change management plan is formulated.
- A path for change management is created.
- A project deployment plan is made.

3. Reinforcing change:

- A deployment plan for the necessary changes are executed.
- For executing change, an appropriate change model is used.
- The fundamental changes are made to an organizational system.
- Reviews and feedback are collected about the changes made.
- Defects caused by the changes are identified and the corrective actions are performed to make changes more sustainable for organizational operations.

6.2 Next Generation Balanced Scorecard

Importance of BSC in IS Implementation:

- Balanced scorecard measures and manages the direction of a strategically focused company through financial as well as non-financial performance indicators and key figures.

Next Generation BSC:

- Next generation balanced scorecard is the concept or implementation of the things that are lacking in the previous balanced scorecard which must be implemented for successful strategic achievement by an organization.
- It includes critical alignment, shared responsibility and cause and effect relationships which are not included in the traditional balanced scorecard.
- It implements the concept that the balanced scorecard should be broken down to the underlying levels by taking few of the operational goals and key figures from each individual strategic goals from top to down.
- It then assigns each of the measured variables that can be influenced, to the responsible employees in the sense of the performance agreement.
- It also manages a comprehensive target and key figures library to keep track of the different sectors for strategic controlling.

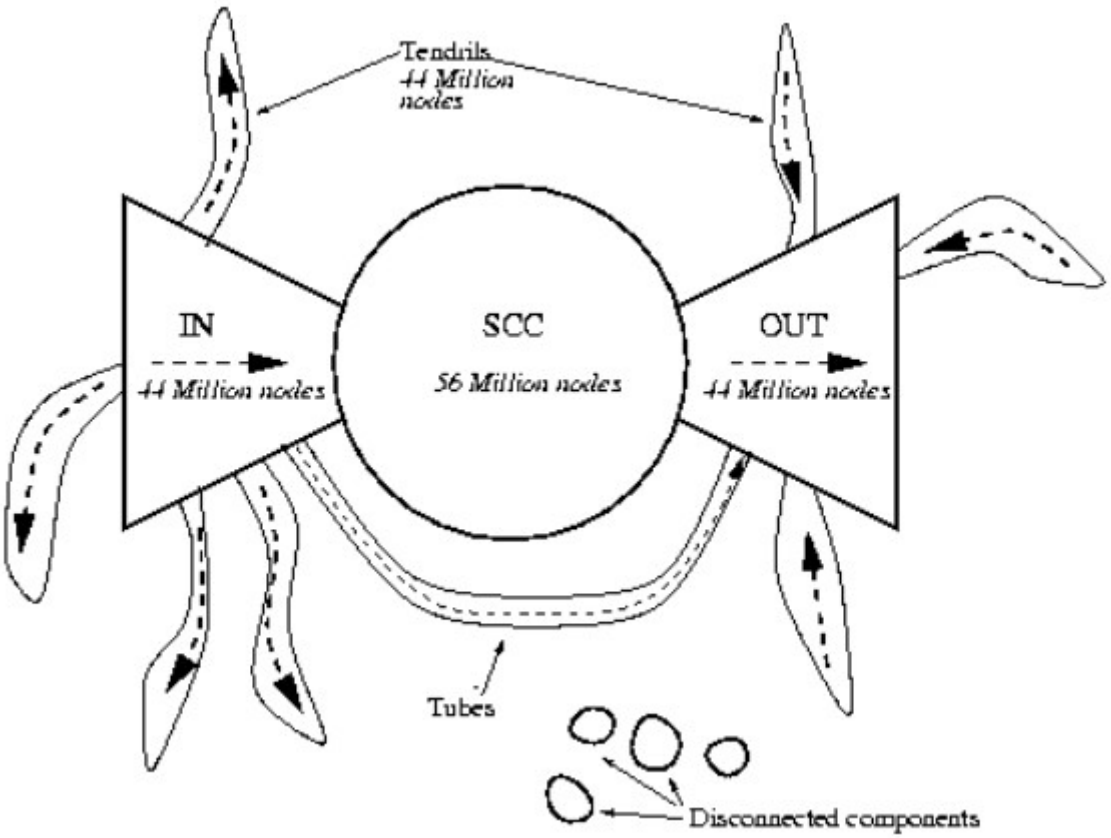
Chapter 7 Web Based Information System and Navigation

7.1 Structure of Web

Bow-Tie Structure of the Web

- A structure of the web can be defined by the directed graph.
- Each node represents the page of the web.
- The directed arrow represents the reachability of one node from another node.
- The Bow-Tie structure of the web consists of following components:

1. Strongly Connected Components (SCC)
2. IN Component
3. OUT Component
4. Tendrils
5. Disconnected Components



Strongly Connected Components (SCC):

- A strongly connected component in a directed graph is a subset of the nodes such that:

a) Every node in the subset has a path to every other node.

b) The subset is not the part of some larger set where every node can reach every other nodes.

- A web consists of a giant strongly connected component.
- The number of web sites and search engines have links leading to directory type pages from which one can reach home pages of almost all of the web sites and also these websites link back to search engines.
- This shows the fact of mutual reachability forming the giant SCC.

IN Component:

- It consists of all the nodes that can reach the giant SCC but can not be reached from it.
- It is the nodes that are upstream of giant SCC.

OUT Component:

- It consists of all the nodes that can be reached from the giant SCC but can not reach giant SCC.
- It is the nodes that are downstream of giant SCC.

Tendrils:

- It consists of:

1. Nodes reachable from IN component that can not reach the giant SCC.
2. Nodes that can reach OUT but can not be reached from giant SCC.

Disconnected Component:

- It consists of the nodes that would not have any path to the giant SCC even if we completely ignored the direction of the edges.

7.2 Link Analysis

Web Link Analysis

- Link is the portion of a web page which refers to other pages.
- Link analysis means the process of analyzing the links present in the web pages.
- Link analysis is very important for the search engines to display their results.
- It helps to analyze whether the links present are active or dead.
- Link analysis helps the analyst to determine whether the search engine is able to find and index the website.
- Link analysis is used by the search engines to compute a composite score for a web page on any given query.

Page Rank

- Page rank is the composite score given by the search engines to the web pages to find and index them when user searches for the query.
- It is the algorithm that is used by the search engines to rank the websites in their results.
- It works by counting the number and quality of links to a page to determine a rough estimate of how important the website is.
- Page rank is a link analysis algorithm and it assigns a numerical weighting to each element of a hyperlinked set of documents, such as the World Wide Web, with the purpose of "measuring" its relative importance within the set.
- Page rank is computed as follows:

1. In a network with n nodes, we assign all nodes the same initial Page Rank, set to be
   1/n.
2. We choose a number of steps k.
3. We then perform a sequence of k updates to the Page Rank values, using the following rule for each update:

a) Basic Page Rank Update Rule: Each page divides its current Page Rank equally
   across its out-going links, and passes these equal shares to the pages it points to. (If a page has no out-going links, it passes all its current Page Rank to itself.) Each page updates its new Page Rank to be the sum of the shares it receives.

7.3 Searching the Web

Search Engine:

- Search engine is the huge database of internet resources that helps to locate information on the World Wide Web.

- Users can search for any information in a search engine by passing query in the form of keywords or phrase.
- The query is then searches in its database and the results are displayed back to the users.

Working of Search Engine:

1. The user enters the keyword to search for the required information through a query in the search engine user interface.
2. The search engine looks for the keyword in the index for predefined database instead of going directly to the web to search for the keyword.
3. It then uses software to search for the information in the database. This software component is known as web crawler.
4. Once web crawler finds the pages, the search engine then shows the relevant web pages as a result. These retrieved web pages generally include title of page, size of text portion, first several sentences etc.
5. User can click on any of the search results to open it to get the relevant information.

Architecture of Search Engine:

- Search Engine consists of following components:

1. Content Collection and Refinement
2. Search Core
3. User and Application Interface

7.4 Web Uses Mining

Web Mining:

- Web mining is the data mining technique that is used to discover patterns from the World Wide Web.
- It is the process of gathering information by mining (extracting something useful) the web.

- It is divided into three types:

1. Web Content Mining
2. Web Usage Mining
3. Web Structure Mining

	Web mining			
	Web Content Mining		Web Structure Mining	Web Usage Mining
	IR view	Db View		
View of Data	-Unstructured -Structured	-Semi Structured -Web Site as DB	-Link Structure	-Interactivity
Main Data	-Text documents -Hypertext documents	-Hypertext documents	-Link Structure	-Serves Logs -Browser Logs
Representation	-Bag of words, n-gram Terms, -phrases, Concepts or ontology -Relational	-Edge labeled Graph, -Relational	-Graph	-Relational Table -Graph
Method	-Machine learning -Statistical (including NLP)	-Proprietary algorithms -Association rules	-Proprietary algorithms	-Machine Learning -Statistical -Association rules
Application Categories	-Categorization -Clustering -Finding extract rules -Finding patterns in text	-Finding frequent sub structures -Web site schema discovery	-Categorization -Clustering	-Site Construction -adaptation and management -Marketing -User Modeling

Web Content Mining:

- Web content mining is the process of mining useful information and knowledge from the contents of the web pages and web documents.
- As the web contents are mostly text, images, audio and video files, NLP techniques are mostly used for mining.

Web Usage Mining:

- Web usage mining is the process of extracting patterns and information from server logs to gain insight on user activity including where the users are from, how many clicked what item on the site and the types of activities being done on the site.
- It provides basic insights on how the users are using the web.
- It helps to discover the web usage patterns from the web data to understand and serve the needs of web based applications.

Web Structure Mining:

- Web structure mining is the process of using graph theory to analyze the node and connection structure of a web site.

- It helps to extract the patterns from the hyperlinks in the web.
- It helps to analyze the document structure to describe the structure of the web site.
- Web structure mining can be used for page ranking of the web sites for search engines.

7.5 Recommender System

- A recommender system is an information filtering system that seeks to predict the preference that a user would give to an item.
- It provides the recommendation to the user based on their previous historical data.
- It is assumed to be the alternative to search algorithms as they help the users discover items they might not have found by themselves.
- The recommender system produces a list of recommendation in two ways as follows:

1. Collaborative Filtering
2. Content based Filtering

Collaborative Filtering:

- Collaborative filtering is based on collecting and analyzing a large amount of information on users' behavior, activities or preferences and predicting what users will like based on their similarity to other users.
- It does not depend up on machine analyzable content, so is able to produce accurate recommendation for complex items too.
- It does not require understanding of the content of an item.
- It is based on the assumption that people who agreed in the past will also agree in the future and that they will like similar kinds of items as they liked in the past.
- The data on users' behavior can be collection explicitly (asking user to search, asking a user to rank items and so on) or implicitly (observing the items that a

user views in an online store, analyzing viewing time of an item, keeping record of items that a user purchases online, analyzing social network of user and so on.)

- The collected data is compared to the similar and dissimilar data collected from others and calculates a list of recommended items for the user.

Problems of Collaborative Filtering:

1. 1. Cold Start:

- The system requires a huge amount of existing data on a user so as to make accurate recommendations.
- This problem is termed as cold start.

2. Scalability:

- In the real world system, there are millions of users and products.
- So, to calculate recommendations, a large computational power should be possessed by the system.

3. Sparsity:

- All the users do not rate the items.
- So, even the most popular items may have few ratings.

Real World Examples : Collaborative Filtering

1. Last.fm (It recommends music based on a comparison of the listening habits of similar users.)
2. Facebook, MySpace, LinkedIn (They use collaborative filtering to recommend new friends, groups and other social connections by examining the network of connections between a user and their friends.)

Content Based Filtering:

- Content based filtering is based on a description of the item and a profile of user's preferences.
- Keywords are used to describe an item and a user profile is built to indicate the type of item this user likes.

- It recommends items that are similar to those that a user liked in the past.
- Item presentation algorithm is used to abstract the features of the items in the system.
- User profile are created by focusing on model of user's preference and history of user interaction with the recommender system.
- The system consists of item profile and content based profile of users based on the weighted vector of item features.
- The weights denote the importance of each feature to the user.
- It uses machine learning techniques like Bayesian classifier, decision tree and ANN to estimate the probability that the user is going to like the item.

Problems with Content Based Filtering:

1. It is effective to recommend same type of items as the user is using. For eg: recommending news articles based on browsing of news.

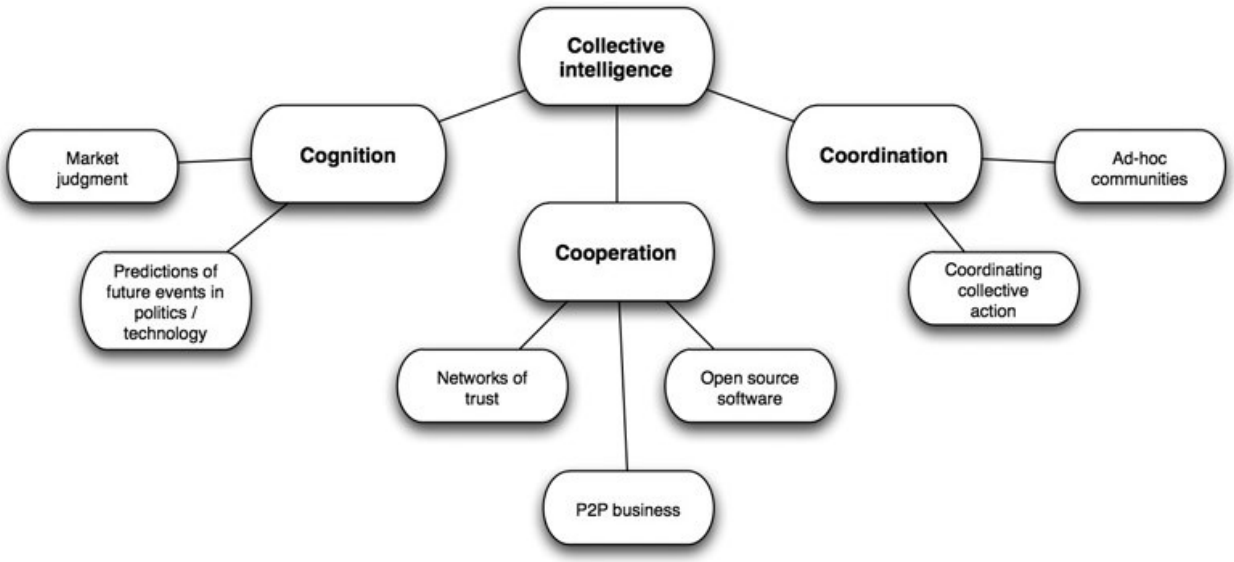
Real World Examples : Content Based Filtering

1. Pandora Radio (It plays music with similar characteristics to that of a song provided by the user as the initial seed)
2. Rotten Tomatoes (Movie recommendation system)

7.6 Collective Intelligence

- Collective intelligence is shared or group intelligence that emerges from the collaboration, collective efforts and competition of many individuals and appears in consensus decision making.
- It is an emergent property between expert and ways of processing information.
- The main goal of collective intelligence is mutual recognition and enrichment of individuals rather than the cult of hypostatized communities.
- In case of computer science, collective intelligence is the capacity of networking information system to enhance the collective pool of social knowledge by simultaneously expanding the extent of human interactions.

- It contributes to the shift of knowledge and power from the individual to the group.
- c factor (general collective intelligence factor) indicates a group's ability to perform a wide range of tasks.



Chapter 8 Scalable and Emerging Information System Technique

8.1 Techniques for Voluminous Data

Voluminous Data

Voluminous data or big data is the application of specialized techniques and technologies used to process very large sets of data. Such data are so large and complex that it becomes difficult to process using traditional data mining techniques and database management tools.

Techniques to analyze Voluminous Data:

Association Rule

- Association rule learning is a method for discovering interesting correlations between variables in large databases.
- It helps to understand closeness of products with each other so as to increase sales.
- It can be used to monitor system logs to detect intruders and malicious activity.
- It helps to extract information about visitors to websites from web server logs.

Classification Tree Analysis

- Statistical classification is the method of identifying categories that a new observation belongs to.
- It requires training set of correctly identified historical data.
- It mainly helps in assigning objects to categories and groups automatically.

Machine Learning

- Machine learning is the method of inducing human like sense to real world inside the machine.
- It provides ability to the computers to learn without being explicitly programmed.

- It helps in predictions based on known properties learned from sets of training data.

Regression Analysis

- Regression analysis is the method that involves manipulating some independent variable to see how it influences a dependent variable.
- It describes how the value of dependent variable changes when the independent variable is varied.
- It is used to understand customer satisfaction against loyalty.

Sentiment Analysis

- Sentiment analysis is the method that determine the sentiments (view) of speakers or writers with respect to a topic.
- It is used in improving services by analyzing guest comments and customers demands.

Characteristics of Voluminous Data:

Volume:

- Volume indicates the quantity of generated and stored data.
- The size of data determines the potential value and insight.
- It also determines whether it is considered to be big data or not.
- The volume of data in the world is increasing exponentially .

Variety:

- Variety indicates the different types and nature of data.
- All the data present in big data analysis may not be of same type.
- Even a single application may be generating variety of data.
- This increases complexity in big data analysis and knowledge extraction.
- The data may be web data, relational data, XML, structured data, streaming data, graph data and so on.
- For efficient extraction of information or patterns, all these variety of data must be linked together and analyzed together.

Velocity:

- Velocity indicates the speed at which data is generated and processed to meet the demands.
- The data obtained is of dynamic nature, so they must be analyzed very fast to provide efficient and effective knowledge.

8.2 Cloud Computing and their types

Cloud

- Cloud is the network basically which is present at remote location.
- It can provide services over public and private networks.

Cloud Computing

- Cloud computing is the method of manipulating, configuring and accessing the hardware and software resources remotely through online data storage, infrastructure and application.
- .It provides platform independent services.

Benefits of Cloud Computing

1. One can access applications over the Internet. It reduces the necessity of installation of software in the system.
2. One can manipulate and configure applications online at any time using any devices.
3. It provides tools for online development and deployment.
4. It provides platform independent resources.
5. It operates at high efficiency with optimal utilization.
6. It provides load balancing services.

Risks of Cloud Computing

1. Cloud computing is provided by third party, this may cause risk to handover the sensitive information to cloud service providers.
2. It is difficult to switch from one cloud service provider to another.

3. The services are accessible by any one from the Internet. So, there may be compromise if necessary security system is not applied.
4. In some cases, data deletion may be insecure or incomplete.

Characteristics of Cloud Computing

1. On demand self service
2. Broad network access
3. Resource pooling
4. Rapid elasticity

Classification based on Access Type

1. **Public Cloud**

- It allows systems and services to be easily accessible to the general public.
- It is insecure due to openness characteristics.

2. **Private Cloud**

- It allows systems and services to be accessible within an organization.
- It is secure because of its private nature.

3. **Community Cloud**

- It allows systems and services to be accessible by a group of organizations.

4. **Hybrid Cloud**

- It is the mixture of public and private cloud.
- The sensitive activities are hosted using private cloud.
- The general activities are hosted using public cloud.

Classification based on Service Type

1. 1. Anything as a Service (XaaS)

- It is the cloud service that provides services related to network, business, identity, database and strategy.

2. Infrastructure as a Service (IaaS)

- It provides infrastructure services to the users.
- The services include virtual machines, servers, storage, networks and so on.

3. Platform as a Service (PaaS)

- It provides the runtime environment for applications, development and deployment tools and so on.
- It includes database, web server, deployment tools and so on.

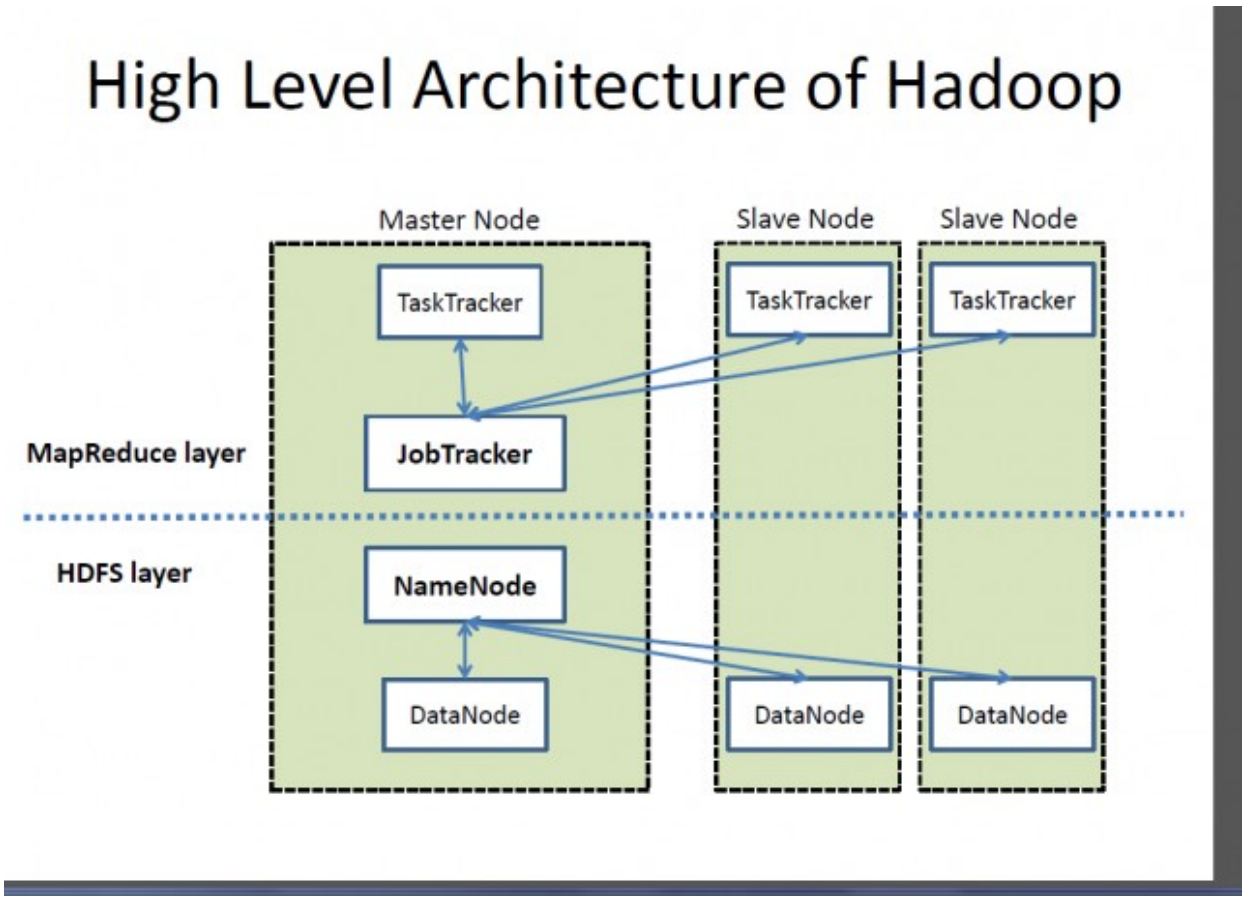
4. Software as a Service (SaaS)

- It provides users to access software applications as a service to the end users.
- It includes CRM, email, games, virtual desktop and so on.

8.3 Map Reduce and Hadoop System

Hadoop System

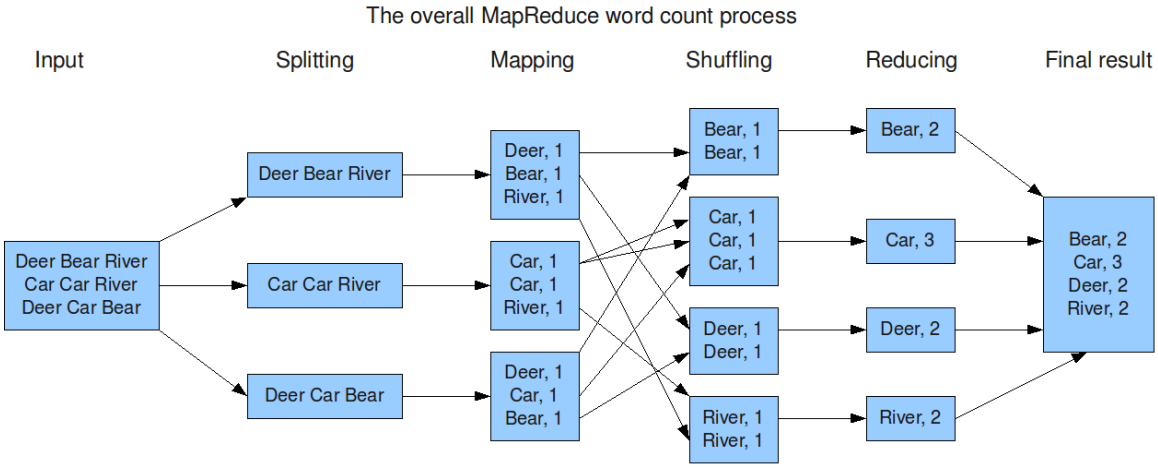
- Hadoop is a framework that allows to process and store huge data sets.
- It is a batch oriented data processing system that works by storing and tracking data across multiple machines and can scale to thousands of servers.
- It is generally used to process huge data sets that are unstructured in nature.
- The data loaded to Hadoop system is split into pieces and spreads across different servers.
- It keeps track of where the data is.
- The complex queries can be performed with faster performance as all the processors are working in parallel.
- For executing such distributed queries, it uses MapReduce.
- It can be divided into two parts: processing and storage.



MapReduce

- MapReduce is the programming model that allows user to process huge data stored in distributed Hadoop system.
- It provides platform to perform distributed and parallel processing on large data sets in a distributed environment.
- It consists of tow distinct tasks: Map and Reduce.
- Map task reads and processes a block of data to produce key-value pairs as intermediate output.
- The output of a Mapper is the input to the Reducer.
- Reduce task receives the key-value pair from multiple map jobs and then aggregates them into a single result set.
- The single result set is the final output of the system.

The working of MapReduce for word count problem is shown in figure below:



8.4 Data Management in the Cloud

Transactional Data Management

- Transactional data are those data that needs ACID property to be guaranteed.
- Such data in the cloud is not a perfect match because of following reasons:

1. Cloud provides shared-nothing architecture but transactional data can not be implemented in such architecture.
2. Since the data are replicated over large geographic distances, it is difficult to maintain the ACID properties.
3. Storing transactional data on an untrusted cloud host arise a lot of risk of data compromization.

Analytical Data Management

- Analytical data means those data that are queried up on for use in business planning, decision support and problem solving.
- The scale of such data is larger as it contains all historical data too.
- Such data are well suited to run in cloud environment due to following reasons:

1. It uses shared-nothing architecture.
2. It does not require ACID properties to be guaranteed.
3. It generally do not contain sensitive data. So, there is no risk of data compromise.