B.E. (TU, PU, PoU, KU)

Insights on INFORMATION SYSTEM

ER. ANKU JAISWAL

Scanned with CamScanner

CONTENTS

CHAPTER – 1

INFORMATION SYSTEM 8

1.1 Information System 2

1.2 The Five Primary Components 3

1.3 Classification of Information System 4

1.4 Information Systems (IS) vs Information Technology (IT) 7

1.5 Expanding Roles of IS (History of IS) 9

1.6 MIS: Management Information System 10

1.7 A Framework for Information Systems Architecture 11

1.8 Qualities of Information System 12

1.9 IS Resources 13

1.10 Balanced Scorecard 14

PRACTICE QUESTIONS ***16***

CHAPTER – 2

CONTROL, AUDIT AND SECURITY OF INFORMATION SYSTEM 14

2.1 Control of Information System 17

2.2 Digital Firm 20

2.3 Audit of Information System 21

2.4 Security of Information System 23

2.5 Layered Security 28

2.6 E-Commerce Security 29

2.7 Threats 30

2.8 SSL Certificates 32

2.9 EV (Extended Validation) SSL Certificates 34

Scanned with CamScanner

2.10 Remote Based Authentication (RBA) 35
2.11 Content Control and Policy Based Encryption 37
PRACTICE QUESTIONS 38

CHAPTER – 3

ENTERPRISE MANAGEMENT SYSTEM

8

3.1 Enterprise Management System 39
3.2 Enterprise Software 40
3.3 ERP (Enterprise Resource Planning) 42
3.4 Supply Chain 46
3.5 Customer Relationship Management (CRM) 48
3.6 Enterprise Information Management (EIM) 49
3.7 Enterprise IT Management (EITM) 51
3.8 Enterprise Information Systems (EIS) 51
3.9 Enterprise Software 52
3.10 Enterprise Engineering 53
3.11 Enterprise Integration 53
3.12 Alignment Process 55
3.13 Electronic Organisms 56
PRACTICE QUESTIONS 57

CHAPTER – 4

12

DECISION SUPPORT AND INTELLIGENT SYSTEM

4.1 Decisions 58
4.2 Decision Support System 58
4.3 Group Decision Support System 62
4.4 Enterprise Decision Support System 63
4.5 Executive Decision Support System 63
4.6 Knowledge Management 64
4.7 Knowledge Management System (KMS) 65
4.8 Expert System 66
4.9 Knowledge Engineer 71
4.10 Intelligent System 71
4.11 Artificial Intelligence 73
4.12 Neural Networks (NN) 74
4.13 Virtual Reality (VR) 76
4.14 Data Mining 77
4.15 Data Warehouse 79
4.16 OLTP vs. OLAP 81
PRACTICE QUESTIONS 82

1Q

1Q

1 Q

CHAPTER – 5

PLANNING FOR IS

5

5.1 Information System Planning 83
5.2 Information System Planning 84
5.3 Information Systems Planning 86
5.4 Three Levels of Planning 90
PRACTICE QUESTIONS 90

CHAPTER – 6

5

IMPLEMENTATION OF INFORMATION SYSTEM

6.1 Change Management 91
6.2 Stages of Change 98
6.3 Managerial Issues 98
6.4 Critical Success Factors (CSFS) 99
PRACTICE QUESTIONS 99

Scanned with CamScanner

CHAPTER – 7

WEB BASED INFORMATION SYSTEM AND NAVIGATION

14

7.1 Structure of the Web 100
7.2 Link Analysis 102
7.3 Searching the Web 103
7.4 Web Navigation 105
7.5 Web Mining 105
7.6 Collaborative Filtering 110
7.7 Recommender System 112
7.8 Collective Intelligence 115
PRACTICE QUESTIONS 116

CHAPTER – 8

14

SCALABLE AND EMERGING INFORMATION SYSTEM TECHNIQUES

8.1 Techniques for Voluminous Data 117
8.2 Scalability IS in Terms of IS 119
8.3 Techniques for Scalability in IS 121
8.4 Cloud Computing Technologies and their Types 123
8.5 MapReduce 129
8.6 Hadoop System 131
8.7 Data Management in the Cloud 134
8.8 Information Retrieval 135
8.9 Link Analysis in Cloud Setup 136
8.10 Page Rank 136
PRACTICE QUESTIONS 140

- CASE STUDY FOR INFORMATION SYSTEM 141
- CASE STUDY 1 141
- CASE STUDY 2 141
- CASE STUDY 3 142
- CASE STUDY 4 142
- CASE STUDY 5 & 6 143
- CASE STUDY 7 144
- CASE STUDY 8 144
- CASE STUDY 9 144
- CASE STUDY 10 144
- SOLVED QUESTIONS 145
- SOLVED CASE STUDY 191
- REFERENCES 210

Scanned with CamScanner

INFORMATION SYSTEM

Data: Data are the facts and figures that can be recorded and have implicit meaning in the given context. A company may record its employee's name, address, telephone number, salary etc. are the examples of data. They are simply recorded raw facts either stored in database, spreadsheets or any other storage media.

Information: When data are processed, interpreted, organized, structured or presented so as to make them meaningful or useful, they are called information. A collection of facts organized in such a way that they have additional value beyond the value of the facts themselves. It is processed data, which includes data that possess context, relevance, and purpose. An organization may collect data related to sales, purchases and other economic activities through various department, these data when processed and organized become source of various information such as profit, loss, revenue, forecasting of future sales and revenue and to understand economic status of organization. This information are utilized at various level inside the organization to make their respective business decision. Thus, information now days play major role in supporting decision process and to roll out future plans and objectives.

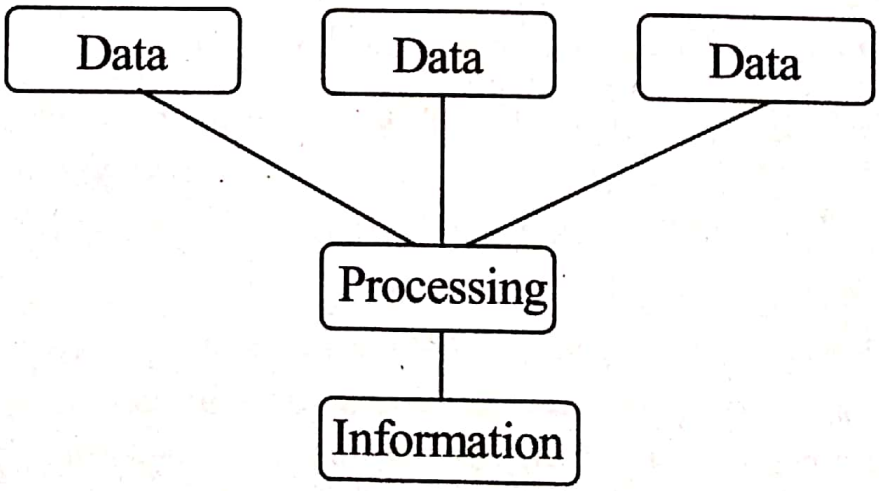


Fig 1: Data and Information

Scanned with CamScanner

The history of temperature readings all over the world for the past 100 years is data. If this data is organized and analyzed to find that global temperature is rising, then that is information.

Data Versus Information

DATA	INFORMATION
Data are simply facts or figures.	Processed data is called information.
Measured in bits and bytes.	Measured in meaningful units like time, quantity, etc.
It never depends on Information	It depended on Data.
It can't be used for decision making	It is widely used for decision-making.
The data collected by the researcher, may or may not be useful.	Information is useful and valuable as it is readily available to the researcher for use.
Data is the property of an organization and is not available for sale to the public.	Information is available for sale to the public.

1.1 Information System

"Information systems (IS) are the study of complementary networks of hardware and software that people and organizations use to collect, filter, and process, create, and distribute data."[1]

"Information systems are combinations of hardware, software, and telecommunications networks that people build and use to collect, create, and distribute useful data, typically in organizational settings."[2]

"Information systems are interrelated components working together to collect, process, store, and disseminate information to support decision making, coordination, control, analysis, and visualization in an organization."[3]

Business firms and other organizations rely on information systems to carry out and manage their operations, interact with their customers and suppliers, and compete in the marketplace. Information systems

are used to run inter organizational supply chains and electronic markets. For example, corporations use information systems to process financial accounts, to manage their human resources, and to reach their potential customers with online promotions.

The IS's designer is concerned with how to use computer systems effectively in producing data for the right person at the right time. 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.

1.2 The Five Primary Components

- **Hardware**

  Information systems hardware is the part of an information system you can touch - the physical components of the technology. Computers, keyboards, disk drives, IPads, and flash drives are all examples of information systems hardware.

- **Software**

  Software is a set of instructions that tells the hardware what to do. Software is non-tangible - it cannot be touched. When programmers create software programs, what they are really doing is simply typing out lists of instructions that tell the hardware what to do. There are several categories of software, with the two main categories being system software and application software. The system software consists of low level programs that runs and interact directly with the hardware, where as the application software run on the top and managed by the system software. Examples of the application software include: MS Word, spreadsheet, various database management system etc. the end users interact directly with them.

- **Data**

  Data is defined as a collection of facts. For example, your street address, the city you live in, and your phone number are all pieces of data. Like software, data is also intangible. By themselves, pieces of data are not really very useful. But

Scanned with CamScanner

aggregated, indexed, and organized together into a database, data can become a powerful tool for businesses.

- **Network**

  These are computing devices that communicate with each other to share information and resources. Different elements need to be connected to each other, especially if many different people in an organization use the same information system.

- **People**

  People are the most important component in most computer based information systems. These are users who combine the hardware, software, data, and networks with procedures to produce quality information. From the front-line help-desk workers, to systems analysts, to programmers, all the way up to the chief information officer (CIO), the people involved with information systems are an essential element that must not be overlooked.

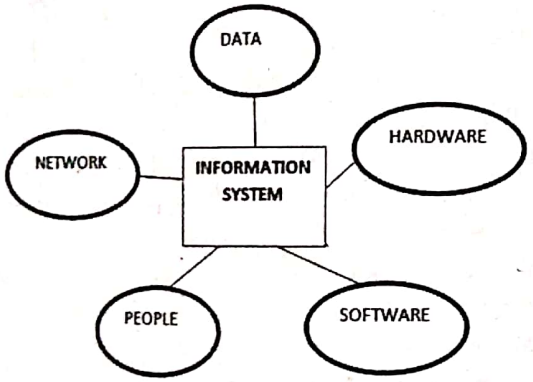


Fig 2: Component of Information System

1.3 Classification of Information System

Information system based on their usage in an organisation can be classified in to two major categories: operations support system and mangement support system.

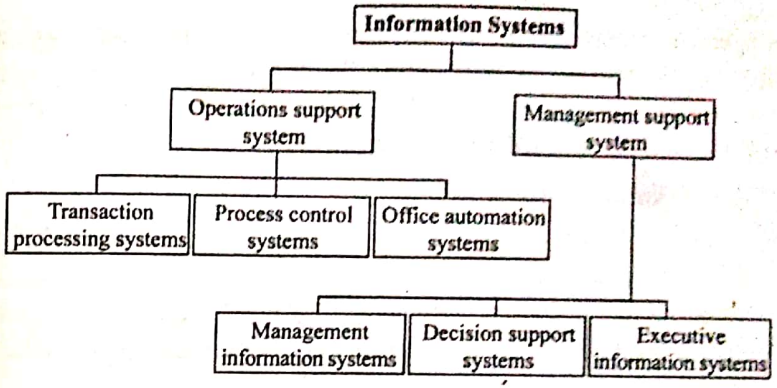


Fig 3: Classification of IS

Operations Support System

The role of operations support system is to efficiently process business transaction, control industrial process support enterprise communication and update corporate database. The Information system generally deals with the support of business operations known as operations support system, produces a variety of information product for external and internal use, however such products produced by OSS needs further processing for efficient use by manager. Major categories are:

- Transaction processing system
- Process control system
- Office automation system

Transaction Processing System

A transaction processing system provides a way to collect, process, store, display modify or cancel transactions. Most of these systems allow multiple transactions to take place simultaneously. The data that this system collects is usually stored in databases, which can be used to produce reports such as billing, wages, inventory summaries; manufacturing schedules, or checks registers.

Process Control System

It is a category of OSS in which decision about a physical production processes are automatically made by computer through routine

Scanned with CamScanner

decisions that control operational process. Ex: A petroleum-refining center uses electronic sensors, which are linked to the computers to continuously monitor chemical processes and make instant adjustments that control the refined process.

Office Automation Systems

These systems are computer based information systems that collect, process, store and transmit electronic messages, documents and other forms of communications among individuals, work groups and organizations. It also helps to automate office procedures and enhance office communications and productivity.

Management Support System

The management support system produces information that is consumed at managerial level to assists in management level planning and decision making. The system deploys various data analysis tools and technologies to derive the information that are critical in managerial decision making process. Major categories are:

- Management Information System
- Decision Support Systems
- Executive Information System

Management Information System

A management information system is an information system that uses the data collected by the transaction processing system and uses this data to create reports in a way that managers can use it to make routine business decisions in response to problems. Some of the reports that this information system creates are summary, exception and ad hoc reports. All this is done to increase the efficiency of managerial activity.

Decision Support System

A decision support system helps make decisions by working and analyzing data that can generate statistical projections and data models. This system gives support rather than replacing a manager's judgment while improving the quality of a manager's decision. A DSS helps to solve problems while using external data.

Executive Information System

Executive information systems are strategic-level information systems that are found at the top of the pyramid. They help executives and senior managers analyze the environment in which the organization operates, to identify long-term trends, and to plan appropriate courses of action. The information in such systems is often weakly structured and comes from both internal and external sources. Executive Information System are designed to be operated directly by executives without the need for intermediaries and easily tailored to the preferences of the individual using them.

1.4 Information Systems (IS) vs Information Technology (IT)

The term "information system" is used to describe people, process and system designed to create, store, manipulate, distribute and disseminate information. The information system acts as intermediary and relates business and computer science. The information technology is defined as the study, design, implementation, management and support for computer based information system. It describes all the hardware, software, networking and various technology related to the working of the information system.

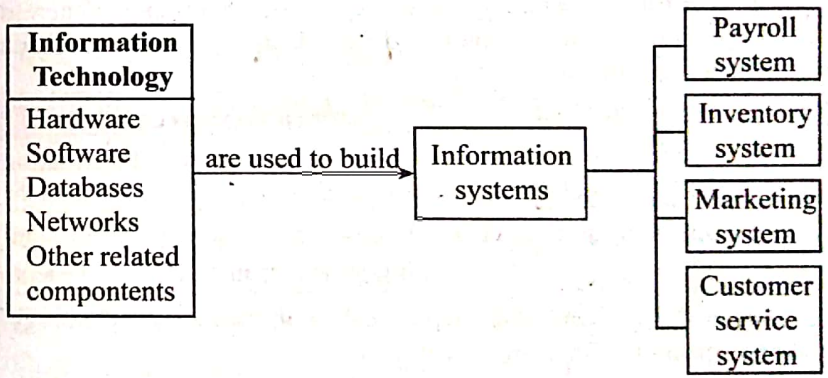


Fig 4: IS vs. IT

- Information technology can be considered as a subset of information systems. It deals with the technology part of any information system, and as such deals with hardware, servers, operating systems and software etc.

Scanned with CamScanner

- A system is always a combination of people, machines, processes and technology. And IT is just a part of the system. Designing of a system takes much more than technology as people and processes are also involved.
- 'Information systems' is in essence bridging the gap between business and the ever growing field of computers. On the other hand, information technology is all about managing technology and making use of it for the betterment of business.

Some of the Points for Differentiating IS and IT are:

- **Origin:** Information systems have been in existence since pre-mechanical era in form of books, drawings, etc. However, the origin of information technology is mostly associated with invention of computers.
- **Development:** Information systems have undergone great deal of evolution, i.e. from manual record keeping to the current cloud storage system. Similarly, information technology is seeing constant changes with evermore-faster processor and constantly shrinking size of storage devices.
- **Business Application:** Businesses have been using information systems for example in form of manual books of accounts to modern TALLY system. The mode of communication has also gone under big change, for example, from a letter to email. Information technology has helped, drive efficiency across organization with improved productivity and precision manufacturing.

Information System	Information Technology
It is combination of people, data, hardware, software and network.	It is a subset of Information System.
Focus on Business Improvement.	Focus on Technology –Design, Implementation and Development
It deals, with bridging the gap between business and field of computer.	It deals with managing technology
Designing a system is much more complex and takes more time.	It is less time taking.
Example: Hardware, Software, Database	Example: Payroll System, Marketing System

1.5 Expanding Roles of IS (History of IS)

- Data Processing: 1950s-1960s
- Management Reporting: 1960s-1970s
- Decision support: 1970s-1980s
- Strategic and End User Support: 1980s-1990s
- 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) are used to access, organize, summarize and display information for supporting routine decision making in the functional areas in the form of report. 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-making process.

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.

Global Internetworking: 1990s-2000s

Global interworking involves 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.

Scanned with CamScanner

1.6 MIS: Management Information System

A management information system (MIS) is a computer system consisting of hardware and software that serves as the backbone of an organization's operations. An MIS gathers data from multiple online systems, analyzes the information, and reports data to aid in management decision-making. Academically, the term is commonly used to refer the study of how individuals, groups, and organizations evaluate, design, implement, manage, and utilize systems to generate information to improve efficiency and effectiveness of decision making, including systems termed decision support systems, expert systems, and executive information systems. Most business schools (or colleges of business administration within universities) have an MIS department, alongside departments of accounting, finance, management, marketing, and sometimes others, and grant degrees (at undergrad, masters, and PhD levels) in MIS. Routine reports are preprogrammed and run at intervals or on demand while others are obtained using built-in query languages; display functions built into the system are used by managers to check on status at desk-side computers connected to the MIS by networks.

Benefits of having an MIS System

- Decision makers need information to make effective decisions. Management Information Systems (MIS) make this possible.
- MIS systems facilitate communication within and outside the organization – employees within the organization are able to easily access the required information for the day to day operations. Facilities such as Short Message Service (SMS) & Email make it possible to communicate with customers and suppliers from within the MIS system that an organization is using.
- Record keeping – management information systems record all business transactions of an organization and provide a reference point for the transactions.

Components of MIS

The major components of a typical management information system are :

- People – People who use the information system
- Data – The data that the information system records

- Business Procedures – Procedures put in place on how to record, store and analyze data
- Hardware – These include servers, workstations, networking equipment, printers, etc.
- Software – These are programs used to handle the data. These include programs such as spreadsheet programs, database software, etc.

1.7 A Framework for Information Systems Architecture

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. Information systems architecture provides a foundation for organizing the various components of any information system you care to develop. Different people have different views of the system. Managers, users, and technical specialists each view the system in different ways, and in different levels of detail. We call these people stakeholders in the system. They can be broadly classified into four groups:

- 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
- *System owners* pay for the system to be built and maintained. They own the system, set priorities for the system, and determine policies for its use. In some cases, system owners may also be system users.
- *System users* are the people who actually use the system to perform or support the work to be completed. In today's team-oriented business world, system users frequently work side-by-side with system designers.
- *System designers* are the technical specialists who design the system to meet the users requirements. In many cases, system designers may also be system builders.

Scanned with CamScanner

- *Systems builders* are the technical specialists who constr test, and deliver the system into operation.

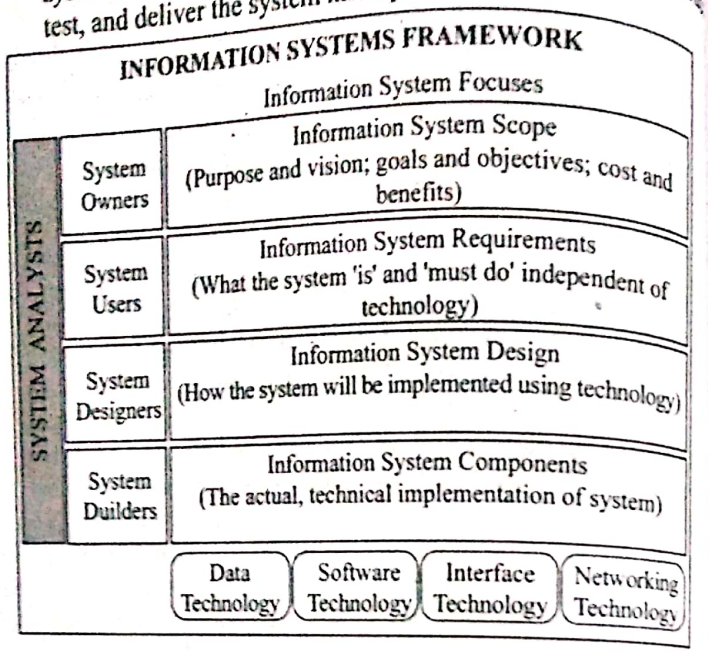


Fig 6: IS Architecture

1.8 Qualities of Information System

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

- **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.
- **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.

- **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.
- **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.
- **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.

1.9 IS Resources

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

Information Systems Resources are :

- **Hardware**

  Information systems hardware is the part of an information system you can touch – the physical components of the technology. Computers, keyboards, disk drives, and flash drives are all examples of information systems hardware.

- **Software**

  Software is a set of instructions that tells the hardware what to do. Software is non-tangible – it cannot be touched. When programmers create software programs, what they are really doing is simply typing out lists of instructions that tell the hardware what to do. There are several categories of software, with the two main categories being system software and application software. The system software consists of low level programs that runs and interact directly with the hardware, where as the application software run on the top and managed

Scanned with CamScanner

by the system software. Examples of the application softwa include: MS Word, spreadsheet, various database manageme system etc. the end users interact directly with them.

- **Data**

  Data is defined as a collection of facts. For example, your stre address, the city you live in, and your phone number are a pieces of data. Like software, data is also intangible. B themselves, pieces of data are not really very useful. B aggregated, indexed, and organized together into a databa data can become a powerful tool for businesses.

- **Network**

  These are computing devices that communicate with each othe to share information and resources. Different elements need t be connected to each other, especially if many different peopl in an organization use the same information system.

- **People**

  People are the most important component in most compute based information systems. These are users who combine the hardware, software, data, and networks with procedures t produce quality information. From the front-line help-desk workers, to systems analysts, to programmers, all the way up t the chief information officer (CIO), the people involved wit information systems are an essential element that must not b overlooked.

1.10 Balanced Scorecard

The balanced scorecard is a management tool that provide stakeholders with a comprehensive measure of how the organization i progressing towards the achievement of its strategic goals.

Balanced Scorecard Concept

The balanced scorecard was first introduced by accounting academic Dr. Robert Kaplan and business executive and theorist Dr. David Norton. It was first published in 1992 in the Harvard Business Review article "The Balanced Scorecard—Measures That Drive Performance." Both Kaplan and Norton took previous metric performance measures and adapted them to include nonfinancial information.

Balanced Scorecard

- Balances financial and non-financial measures.
- Balances short and long-term measures.
- Should contain just enough data to give a complete picture of organizational performance.
- Leads to strategic focus and organizational alignment.

Why ?

- To achieve strategic objectives.
- To provide quality with fewer recourses.
- To eliminate non-value added efforts.
- To track progress.
- To align customer priorities.
- To evaluate process changes.
- To continually improves.

Perspectives of Balanced Scorecard

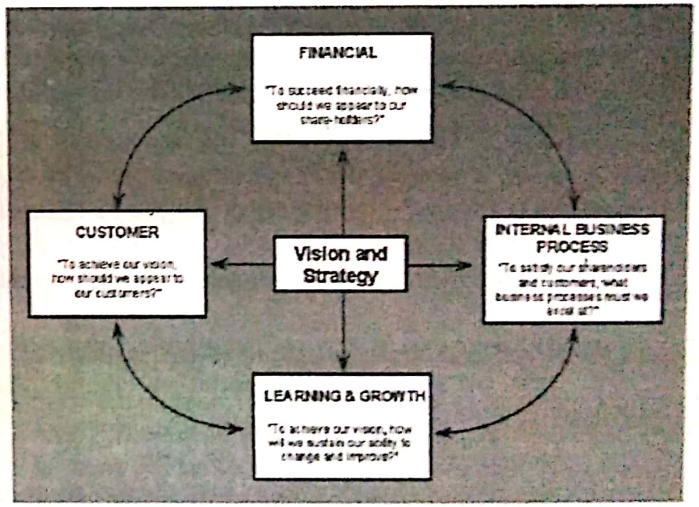


Fig 7: Perspective of Balanced Scorecard

Measures

- To determine how effectively and efficiently the process or service satisfies the customer.

Scanned with CamScanner

- To identify improvement opportunities.
- To make decisions based on FACT and DATA.

Measurements should

- Evaluate the quality of processes.
- Track our improvements.
- Focus our effort on our customers.
- Support our strategies.

Scorecard Potential Pitfalls

- Lack of well Defined Strategy
- Too much focus on the lagging measures
- Use of Generic Metrics
- Self-serving manager

Balanced Scorecard Importance

- Key performance helps to measure strategy at all levels of an organization.
- The methodology facilitates communication and understanding of business goals and strategies at all level of an organization.
- It enables executives to truly execute their strategy by identifying what should be done and measured.

PRACTICE QUESTIONS

1. What is an Information System? Explain the different types of Information System that supports operations and management decision-making.
2. Why does Organization need Information System? Differentiate IS and typical DBMS?
3. Differentiate Information System and Information Technology.
4. Explain relationship between balanced Scorecard and Information System.
5. Explain IS architecture with proper diagram.
6. What are the various IS component? Also discuss about IS resources.
7. What are various quality of IS.
8. Explain different types of IS with proper classification.
9. How does IS help in managing Decision making?

CHAPTER – 2

CONTROL, AUDIT AND SECURITY OF INFORMATION SYSTEM

2.1 Control of Information System

Information system control is defined as methods, policies, and procedures that ensure protection of organization's assets, accuracy and reliability of records, and operational adherence to management standards. Protection of information resources requires a well-designed set of controls. Computer systems are controlled by a combination of general controls and application controls.

General controls govern the design, security, and use of computer programs and the security of data files in general, throughout the organization's information technology infrastructure. On the whole, general controls apply to all computerized applications and consist of a combination of hardware, software, and manual procedures that create an overall control environment.

Application controls are specific controls unique to each computerized application, such as payroll or order processing. They consist of controls applied from the business functional area of a particular system and from programmed procedures.

General Controls

General controls include software controls, physical hardware controls, computer operations controls, data security controls, controls over the systems implementation process.

- **Software controls:** Monitor the use of system software and prevent unauthorized access of software programs, system software, and computer programs. System software is an important control area because it performs overall control functions for the programs that directly process data and data files.
- **Hardware controls:** Ensure that computer hardware is physically secure and check for equipment malfunction.

Scanned with CamScanner

Computer equipment should be specially protected against fires and extremes of temperature and humidity. Organizations that are dependent on their computers also must make provisions for backup or continued operation to maintain constant service.

- **Computer operations controls:** Oversee the work of the computer department to ensure that programmed procedures are consistently and correctly applied to the storage and processing of data. They include controls over the setup of computer processing jobs, computer operations, backup and recovery procedures for processing that ends abnormally.
- **Data security controls:** Ensure that valuable business data files on either disk or tape are not subject to unauthorized access, change, or destruction while they are in use or in storage.
- **Implementation controls:** Audit the systems development process at various points to ensure that the process is properly controlled and managed. The systems development audit looks for the presence of formal reviews by users and management at various stages of development; the level of user involvement at each stage of implementation; and the use of a formal cost-benefit methodology in establishing system feasibility. The audit should look for the use of controls and quality assurance techniques for program development, conversion, and testing and for complete and thorough system, user, and operations documentation.
- **Administrative controls:** Formalize standards, rules, procedures, and control disciplines to ensure that the organization's general and application controls are properly executed and enforced.

Application Controls

Application controls include both automated and manual procedures that ensure that only authorized data are completely and accurately processed by that application. Application controls can be classified as (1) input controls, (2) processing controls, and (3) output controls.

Input controls check data for accuracy and completeness when they enter the system. There are specific input controls for input

authorization, data conversion, data editing, and error handling. Processing controls establish that data are complete and accurate during updating. Run control totals, computer matching, and programmed edit checks are used as processing controls. Output controls ensure that the results of computer processing are accurate, complete, and properly distributed. Following table provides more detailed examples of each type of application control. Not all of the application controls discussed here are used in every information system. Some systems require more of these controls than others, depending on the importance of the data and the nature of the application. Input controls: check data for accuracy and completeness when they enter the system. There are specific input controls for input authorization, data conversion, data editing and error handling.

Name of Control	Type of Application Control	Description
Control totals	Input, processing	Totals established beforehand for input and processing transactions. These totals can range from a simple document count to totals for quantity fields, such as total sales amount (for a batch of transactions). Computer programs count the totals from transactions input or processed.
Edit checks	Input	Programmed routines that can be performed to edit input data for errors before they are processed. Transactions that do not meet edit criteria are rejected. For example, data might be checked to make sure they are in the right format (for instance, a nine-digit social security number should not contain any alphabetic characters).

Scanned with CamScanner

Name of Control	Type of Application Control	Description
Computer matching	Input, processing	Matches input data with information held on master or suspense files and notes unmatched items for investigation. For example, a matching program might match employee time cards with a payroll master file and report missing or duplicate time cards.
Run control totals	Processing, output	Balance the total of transactions processed with total number of transactions input or output.
Report distribution logs	Output	Documentation specifying that authorized recipients have received their reports, checks, or other critical documents.

2.2 Digital Firm

"A digital firm is one in which nearly all of the organization's significant business relationships with customers, suppliers, and employees are digitally enabled, and key corporate assets are managed through digital means."

When a firm goes digital, it's not about just adding a computer system to the mix but throwing a computer system at outdated business processes is exactly the wrong thing to do. A truly digital firm has several characteristics that distinguish it from most of the firms claiming to be digitized:

- Significant business relationships with customers, suppliers, and employees are digitally enabled and mediated.
- Core business processes are accomplished through digital networks and span the entire organization or link multiple organizations.

- Key corporate assets –intellectual property, core competencies, and financial and human assets – are managed through digital means.
- Internal and external environments are quickly recognized and dealt with.

And the number one reason digital firms experience greater opportunities for success and profits is because they view information technology as the —core of the business and the primary management tool.

Protecting the Digital Firm

- **High-availability computing:** Tools and technologies enabling system to recover from a crash
- **Disaster recovery plan:** Runs business in event of computer outage
- **Load balancing:** Distributes large number of requests for access among multiple servers
- **Mirroring:** Duplicating all processes and transactions of server on backup server to prevent any interruption
- **Clustering:** Linking two computers together so that a second computer can act as a backup to the primary computer or speed up processing

2.3 Audit of Information System

A financial audit is an objective examination and evaluation of the financial statements of an organization to make sure that the records are a fair and accurate representation of the transactions they claim to represent. It can be done internally by employees of the organization, or externally by an outside firm. Audit is an appraisal activity carried out by people who are not actively involved in performing the activity under appraisal. It aims at prevention and detection of abuse of the corporate resources.

The information system audit is conducted to evaluate the information systems and suggest measures to improve their value to the business. The information system audit can be used as an effective tool for evaluation of the information system and controlling the computer

Scanned with CamScanner

abuse. Information system audit is carried out by professionals who are not only well versed with the complex information system issues but also know how to relate them to the business.

The process of information system audit involves four steps:

- **Measuring vulnerability of information system:**

  The first step in the process of information system audit is the identification of the vulnerability of each application. Where the probability of computer abuse is high, there is a greater need for an information system audit of that application. The probability of computer abuse would depend upon the nature of the application and the quality of controls.

- **Identification of sources of threat:**

  Most of the threats of computer abuse are from the people. The information system auditor should identify the people who might pose a threat to the information sys-tems. These people include system analysts, programmers, data entry operators, data providers, users, vendors of hardware, soft-ware and services, computer security specialists, PC users, etc.

- **Identification of high risk points:**

  The next step in the process of information system audit is to identify the occasions, points or events when the information system may be penetrated. These points may be when a transaction is added, altered or deleted. The point of high-risk point may also be the occasion when a data or program file is changed or the operation is faulty.

- **Check for computer abuse:**

  The last step in the process is to conduct the audit of high potential points keeping the view the activities of the people who could abuse the information system for the applications that are highly vulnerable.

Scope of Audit

The information system audit may encompass almost all the resources of IT infrastructure. Thus, it will involve evaluation of hardware, application of software, the data resources and the people. However,

one of the most important resources that attract the attention of an information system auditor is the application software.

Information System Auditors

An information system auditor is the link between software development team and the management. His role is different from the system analyst who interacts to help in development of application software. The information system auditor evaluates the review of every project on behalf of the management. The information system auditor is associated right from the feasibility study of information system development project to the implementation stage. In fact, the information system auditor gives the clearance for implementation after due review and evaluation of software package.

2.4 Security of Information System

Security means protection of data from accidental or intentional modification, destruction or disclosure to unauthorized persons. Information systems security, more commonly referred to as INFOSEC, refers to the processes and methodologies involved with keeping information confidential, available, and assuring its integrity.

It also refers to:

- Access controls, which prevent unauthorized personnel from entering or accessing a system.
- Protecting information, no matter where that information is, i.e. in transit (such as in an email) or in a storage area.
- The detection and remediation of security breaches, as well as documenting those events.

The Information Security Triad: Confidentiality, Integrity, Availability

Confidentiality

When protecting information, we want to be able to restrict access to those who are allowed to see it; everyone else should be disallowed from learning anything about its contents. This is the essence of confidentiality. For example, federal law requires that universities restrict access to private student information. The university must be sure that only those who are authorized have access to view the grade records.

Scanned with CamScanner

Integrity

Integrity is the assurance that the information being accessed has not been altered and truly represents what is intended. Just as a person with integrity means what he or she says and can be trusted to consistently represent the truth, information integrity means information truly represents its intended meaning. Information can lose its integrity through malicious intent, such as when someone who is not authorized makes a change to intentionally misrepresent something. An example of this would be when a hacker is hired to go into the university's system and change a grade. Integrity can also be lost unintentionally, such as when a computer power surge corrupts a file or someone authorized to make a change accidentally deletes a file or enters incorrect information.

Availability

Information availability is the third part of the CIA triad. Availability means that information can be accessed and modified by anyone authorized to do so in an appropriate timeframe. Depending on the type of information, appropriate timeframe can mean different things. For example, a stock trader needs information to be available immediately, while a sales person may be happy to get sales numbers for the day in a report the next morning. Companies such as Amazon.com will require their servers to be available twenty-four hours a day, seven days a week. Other companies may not suffer if their web servers are down for a few minutes once in a while.

Tools for Information Security

In order to ensure the confidentiality, integrity, and availability of information, organizations can choose from a variety of tools. Each of these tools can be utilized as part of an overall information-security policy, which will be discussed in the next section.

- **Authentication**

  Authentication is used to ensure that the person accessing the information is, indeed, who they present themselves to be. Authentication can be accomplished by identifying someone through one or more of three factors: something they know, something they have, or something they are. For example, the most common form of authentication today is the user ID and password. In this case , the authentication is done by confirming something that the user knows (their ID and password). But this form of authentication is easy to compromise (see sidebar) and stronger forms of authentication are sometimes needed. Identifying someone only by something they have, such as a key or a card, can also be problematic. When that identifying token is lost or stolen, the identity can be easily stolen. The final factor, something you are, is much harder to compromise. This factor identifies a user through the use of a physical characteristic, such as an eye-scan or fingerprint. Identifying someone through his or her physical characteristics is called biometrics. A more secure way to authenticate a user is to do multi-factor authentication. By combining two or more of the factors listed above, it becomes much more difficult for someone to misrepresent themselves. An example of this would be the use of an RSA Secure ID token. The RSA device is something you have, and will generate a new access code every sixty seconds. To log in to an information resource using the RSA device, you combine something you know, a four-digit PIN, with the code generated by the device. The only way to properly authenticate is by both knowing the code and having the RSA device.

- **Access Control**

  Once a user has been authenticated, the next step is to ensure that they can only access the information resources that are appropriate. This is done through the use of access control. Access control determines which users are authorized to read, modify, add, and/or delete information. Several different access control models exist. Here we will discuss two: the access control list (ACL) and role-based access control (RBAC). For each information resource that an organization wishes to manage, a list of users who have the ability to take specific actions can be created. This is an access control list, or ACL.

Scanned with CamScanner

For each user, specific capabilities are assigned, such as read, write, delete, or add. Only users with those capabilities are allowed to perform those functions. If a user is not on the list, they have no ability to even know that the information resource exists.

ACLs are simple to understand and maintain. However, they have several drawbacks. The primary drawback is that each information resource is managed separately, so if a security administrator wanted to add or remove a user to a large set of information resources, it would be quite difficult. And as the number of users and resources increase, ACLs become harder to maintain. This has led to an improved method of access control called role-based access control, or RBAC. With RBAC instead of giving specific users access rights to an information resource, users are assigned to roles and then those roles are assigned the access. This allows the administrators to manage users and roles separately, simplifying administration and, by extension, improving security.

- **Encryption**

Many times, an organization needs to transmit information over the Internet or transfer it on external media such as a CD or flash drive. In these cases, even with proper authentication and access control, it is possible for an unauthorized person to get access to the data. Encryption is a process of encoding data upon its transmission or storage so that only authorized individuals can read it. This encoding is accomplished by a computer program, which encodes the plain text that needs to be transmitted; then the recipient receives the cipher text and decodes it (decryption). In order for this to work, the sender and receiver need to agree on the method of encoding so that both parties can communicate properly. Both parties share the encryption key, enabling them to encode and decode each other's messages. This is called symmetric key encryption. This type of encryption is problematic because the key is available in two different places. An alternative to symmetric key

encryption is public key encryption. In public key encryption, two keys are used: a public key and a private key. To send an encrypted message, you obtain the public key, encode the message, and send it. The recipient then uses the private key to decode it. The public key can be given to anyone who wishes to send the recipient a message. Each user simply needs one private key and one public key in order to secure messages. The private key is necessary in order to decrypt something sent with the public key.

Potential Threats to Security

- Natural disasters such as fire, floods, earthquakes.
- Accidents such as disk crashes, file erasure by inexperienced operators.
- Theft/erasure of data by disgruntled employees.
- Frauds by changing programs, data by employees.
- Industrial espionage.
- Viruses/Worms.
- Hackers who break into systems connected to the internet.
- Denial of service attacks by flooding with mail.

How to Protect Data/programs?

- Regular back up of data bases every day/or week depending on the time criticality and size.
- Incremental back up at shorter intervals.
- Backup copies kept in safe remote location-particularly necessary for disaster recovery.
- Duplicate systems run and all transactions mirrored if it is a very critical system and cannot tolerate any disruption before storing in disk.
- Physical locks.
- Password system.
- Biometric authentication (e.g. Finger print).

Scanned with CamScanner

- Encrypting sensitive data/programs.
- Identification of all persons who read or modify data an logging it in a file.
- Training employees on data care/handling and security.
- Antivirus software.
- Firewall protection when connected to internet.

2.5. Layered Security

Layered security refers to security systems that use multipl components to protect operations on multiple levels, or layers. Th term can also be related to the term defense in depth, which is base on a slightly different idea where multiple strategies and resources a used to slow, block, delay or hinder a threat until it can be complete neutralized. The central idea behind layered security or defense is th in order to protect systems from a broad range of attacks, usin multiple strategies will be more effective. Layered security ca involve security protocols at the system or network levels, at th application level, or at the transmission level, where security exper may focus on data in use over data at rest. Layered security effor attempt to address problems with different kinds of hacking o phishing, denial of service attacks and other cyber attacks, as well a worms, viruses, malware and other kinds of more passive or indire system invasions. Layered security may also be known as layere defense.

Consumer Layered Security Strategy

- Extended validation (EV) SSL certificates
- Multifactor authentication (also sometimes known as versati or two-factor authentication)
- Single sign-on (SSO)
- Fraud detection and risk-based authentication
- Transaction signing and encryption
- Secure Web and e-mail
- Open fraud intelligence network

Enterprise Layered Security Strategy

- Workstation application whitelisting
- Workstation system restore solution
- Workstation and network authentication
- File, disk and removable media encryption
- Remote access authentication
- Network folder encryption
- Secure boundary and end-to-end messaging
- Content control and policy-based encryption

2.6 E-Commerce Security

Electronic commerce or ecommerce is a term for any type of business, or commercial transaction that involves the transfer of information across the Internet. E-commerce security is the protection of e-commerce assets from unauthorized access, use, alteration, or destruction.

Dimensions of E-Commerce Security

- **Integrity:**

  This means the information served or transferred has not been changed or altered by an unauthorized entity. Example of this would be spoofed email. Intercepting wire transfer for some bank account to other bank account.

- **Nonrepudiation:**

  The ability to make sure those e-commerce participants can't deny their online transactions. User using pin for applying credit card to start the transaction.

- **Authenticity:**

  The ability to establish the identity of the entity with which one is dealing on the internet. Spoofing email, using someone else identity to send email would be good example of authenticity breach.

Scanned with CamScanner

- **Confidentiality:**

  The ability to make sure that the data is only available intended users. Sniffing breach in which a program is used steal information in a network like company files and reports.

- **Privacy:**

  The ability to control the information that the users provides e- commerce merchant about themselves. Hacking attempt the database of the ecommerce website and getting the cre card numbers of the users.

- **Availability:**

  This is the ability to make sure the e-commerce webs continues to do the function for which it was made. D (denial of service attack) in which the hacker floods the serv with fake requests and the server crashes.

2.7 Threats

Anyone with the capability, technology, opportunity, and intent to harm. Potential threats can be foreign or domestic, internal or extern state-sponsored or a single rogue element. Terrorists, inside disgruntled employees, and hackers are included in this.

Intellectual property threats

- use existing materials found on the Internet without t owner's permission, e.g., music downloading, domain na (cybersquatting), software pirating

Client computer threats

- Trojan horse
- Active contents
- Viruses

Communication channel threats

- Sniffer program
- Backdoor
- Spoofing
- Denial-of-service

Server threats

- Privilege setting
- Server Side Include (SSI), Common Gateway Interface (CGI)
- File transfer
- Spamming

Countermeasure

A procedure that recognizes, reduces, or eliminates a threat.

Intellectual property protection

- Legislature
- Authentication

Client computer protection

- Digital certificate
- Browser protection
- Antivirus software
- Computer forensics expert

Communication channel protection

- Encryption
- Protocol
- Digital signature
- Server protection
- Access control and authentication

How to Minimize Security Threats?

Perform a risk assessment-a list of information assets and their value to the firm

Develop a security policy-a written statement on:

- what assets to protect from whom?
- why these assets are being protected?
- who is responsible for what protection?
- which behaviors are acceptable and unacceptable?

Scanned with CamScanner

Develop an implementation plan-a set of action steps to achieve security goals

Create a security organization-a unit to administer the security policy

Perform a security audit-a routine review of access logs and evaluation of security procedures

2.8 SSL Certificates

SSL stands for Secure Socket Layer. It is the standard security technology for establishing an encrypted link between a web server and a browser. This link ensures that all data passed between the web server and browsers remain private and integral. To be able to create an SSL connection a web server requires an SSL Certificate. SSL Certificates are small data files that digitally bind a cryptographic key to an organization's details. When installed on a web server, it activates the padlock and the https protocol (over port 443) and allows secure connections from a web server to a browser. Typically, SSL is used to secure credit card transactions, data transfer and logins, and more recently is becoming the norm when securing browsing of social media sites.

SSL Certificates Binds together:

- A domain name, server name or hostname.
- An organizational identity (i.e. company name) and location.
- An organization needs to install the SSL Certificate onto its web server to initiate secure sessions with browsers.
- Depending on the type of SSL Certificate applied for, the organization will need to go through differing levels of vetting.
- Once installed, it is possible to connect to the website over https://www.domain.com, as this tells the server to establish a secure connection with the browser.
- Once a secure connection is established, all web traffic between the web server and the web browser will be secure. Browsers tell visitors a website is SSL secure via several visible trust indicators:

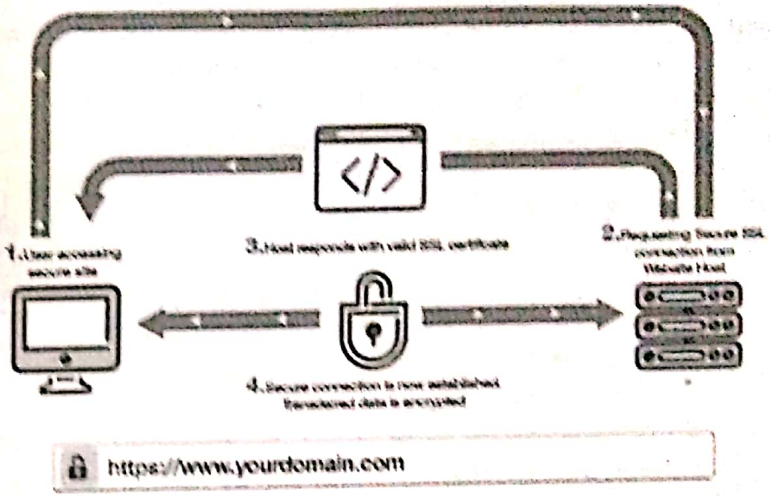


Fig 1: SSL Certificate

Types of SSL Certificates

Why there are different types of Certificates?

- Some organizations need SSL simply for confidentiality, e.g. encryption.
- Some organizations wish to use SSL to enhance trust in their security and identity, e.g. they want to show customers they have been vetted and are a legitimate organization.

There are basically three types of SSL certificates:

- **OV SSL Certificates:** This process assures the validity of a Web site by verifying that the applicant is a legitimate business. Before issuing the SSL certificate, the CA performs a rigorous

Scanned with CamScanner

validation procedure, including checking the applica business credentials (such as the Articles of Incorporation) a verifying the accuracy of its physical and Web addresses.

- **DV SSL Certificates:** The validation procedure is less rigoro for a Domain Validated SSL Certificate. When issuing Domain Validated SSL Certificate, the CA checks only that t applicant's name and contact information matches t registration information in the WHOIS database for the doma name associated with the applied for SSL Certificate.
- **EV SSL Certificates:** The Certificate application process itse is more thorough and the validation criteria more rigorous f EV certification, whose applicants, at least initially, are limite to certain types of business entities and government agencies.

2.9 EV (Extended Validation) SSL Certificates

Extended Validation, or EV SSL, raises the bar on standard SS validation processes, incorporating some of the highest standards i identity assurance to establish the legitimacy of online entitie Certificate Authorities put applicant websites through rigorou evaluation procedures and meticulous documentation checks t confirm their authenticity and ownership. This systemati authentication process, also known as the Extended Validatio Standard, is based on a set of guidelines prescribed for CAs to adher to when they receive a request for a digital certificate from a organization or business entity.

The objective of the EV issuance process is to enable users t distinguish legitimate websites from phishing sites, building their trus in online commercial transactions and increasing participation.

These guidelines include:

- Establishing the legal, physical and operational existence of the entity
- Verifying that the entity's identity matches official records like incorporation and business licensing information
- Confirming that the entity owns or has exclusive rights to use the domain mentioned in the application for certification

- Confirming that the request for an EV certificate has been authorized by the entity

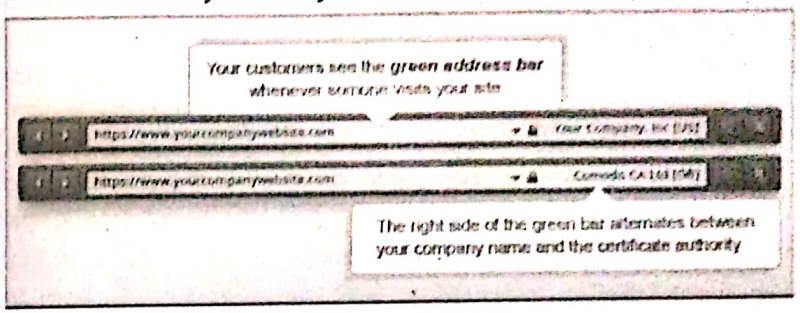


Fig 2: EV Certificate

2.10 Remote Based Authentication (RBA)

Remote access is the ability to get access to a computer or a network from a remote distance through wired or wireless connection. Authentication is the method of proving the subject's identity. E.g.: Password, Passphrase, PIN

Why is RBA used?

- To prevent accessing private data and information transferring between server and users i.e. Channel attack.
- Direct attacks from hackers into network.
- Brute force, software attacks

Authentication Methods:

- **Biometrics**

  Biometrics are biological measurements — or physical characteristics — that can be used to identify individuals. Fingerprint mapping, facial recognition, and retina scans are all forms of biometric technology, but these are just the most recognized options.

- **Passwords**

  Password protection is a security process that protects information accessible via computers that needs to be protected from certain users. Password protection allows only those with an authorized password to gain access to certain information.

Scanned with CamScanner

- **Cognitive Passwords**

  A cognitive password is a form of knowledge-based authentication that requires a user to answer a question presumably something they intrinsically know, to verify their identity. Cognitive password systems have been researched for many years and are currently commonly used as a form of secondary access. They were developed to overcome the common memorability vs. strength problem that exists with the traditional password.

- **Card Based**

  A smart card, chip card, or integrated circuit card (ICC) is a physical electronic authorization device, used to control access to a resource. It is typically a plastic credit card sized card with an embedded integrated circuit. Many smart cards include a pattern of metal contacts to electrically connect to the internal chip. Others are contactless, and some are both. Smart cards can provide personal identification, authentication, data storage, and application processing.

- **One-Time or Dynamic Passwords (token based)**

  OTP security tokens are microprocessor-based smart cards or pocket-size key fobs that produce a numeric or alphanumeric code to authenticate access to the system or transaction. This secret code changes every 30 or 60 seconds, depending on how the token is configured. Mobile device apps, such as Google Authenticator, rely on the token device and PIN to generate the one-time password for two-step verification. OTP security tokens can be implemented using hardware, software or on demand. Unlike traditional passwords that remain static or expire every 30 to 60 days, the one-time password is used for one transaction or login session.

Importance Today?

- Today everything is electronics and internet based like e-banking, e-commerce, e-learning, e-governance, m-banking etc.
- Companies have many branches worldwide so data and information are distributed among branches offices.

- User do transaction remotely using internet using different handheld devices.
- All information of enterprises is centralized at server which is shared/distributed remotely among concerned people worldwide.

2.11 Content Control and Policy Based Encryption

Services for the security of email content in an organization are called content control. Email content like: credit card no., account information, organization vital information and customer vital information needs to be protected.

Email Content Control: This service contains the rules that defines:

- whether an email is to be encrypted or not
- which block or header in email to been encrypted

Policy Based Encryption (PBE)

- It is a service that encrypts specific emails based on policy.

  Set of rules designed to analyze all email
- PBE uses the Email Content Control rules to identify which email needs to be encrypted.
- The PBE Service is managed through the same control panel that you use to manage your Anti-Virus.
- PBE service is closely integrated with the Email Content Control service.

Policy Based Encryption Benefits

- Automatically applies email encryption based on the organization's email security policies.
- Data loss prevention and email messages security policies are consistently and accurately applied.
- Eliminates email encryption key management, backup and administration burdens.

Scanned with CamScanner

PRACTICE QUESTIONS

1. What are different security threats while deploying Information System over extranet of a business firm? What are the technologies for mitigating the security risks in Information System?
2. What are different threats to Information System in terms of security? Discuss different controls and security measures (including physical and logical) that are to be applied to protect information system?
3. Explain the security component in IS. Explain.
4. Explain Enterprise Layered security strategy with its types in detail. Why audit and control in information are important.
5. What is multi layer security strategy? Discuss how multi layer security strategy can be applied to protect e- commerce system.
6. Define physical and logical security that can be applied in Information system. What is the key threat and vulnerabilities against with any information system should be protected? What are different controls that can be applied to protect information system against such threats?
7. How do you define cyber crime and theft of intellectual property right? Discuss with suitable example.
8. List and describe the most common threats against contemporary information system. What are different physical and logical security measures that organization can apply to protect information system for different threats.
9. Define security and control of information. Explain the technologies and tools for protecting information resources.
10. Explain SSL certificates for client server connection over the web. What do you mean by Extended Validation?
11. What is an Information audit? What benefits will you obtain after going through Information audit?

ENTERPRISE MANAGEMENT SYSTEM

"Enterprise applications are about the display, manipulation, and storage of large amounts of often complex data and the support or automation of business processes with that data." -M. Fowler

Enterprise resource planning (ERP) is business management software that allows an organization to use a system of integrated applications to manage the business.

Supply chain management (SCM) is the management of an interconnected or interlinked between network, channel and node businesses involved in the provision of product and service packages required by the end customers in a supply chain.

Customer relationship management (CRM) is a model for managing a company's interactions with current and future customers.

3.1 Enterprise Management System

Enterprise Management Systems (EMS), sometimes just known, as Enterprise Systems (ES) are enterprise-scale application software packages, which address the different software, needs of large organizations - similar to ERP. They allow IT teams to support and manage large, complex and sometimes geographically-dispersed IT infrastructure and-applications.

Most EMS solutions will include core business functions like financial processing, HR management, Customer Relationship Management (CRM), budgeting, sales activity and logistics and supply chain management. Most EMS tools are presented as a single package, but can also be bespoke systems created around an organization's specific needs.

Enterprise Management Systems are designed primarily for large organizations, as the name suggests, and are usually not suitable for small or medium organizations due to implementation costs and the less complex IT infrastructure often found in smaller businesses.

Scanned with CamScanner

- **Enterprise** – an entire company, everything, all-inclusive
- **Management** – The monitoring and controlling of entities
- **Systems** – Information Technology Infrastructure, hardw and software, data, information, and processes

3.2 Enterprise Software

Enterprise software is an over-arching term for any software used large organizations (whether business or government). It is conside to be an essential part of a computer-based information system, and provides business-oriented tools such as online payment process and automated billing systems. Enterprise software is also referred as enterprise application software. By developing specific-purpo oriented software, the enterprises stand to gain through increase business logic functionality as well as productivity. Be it la industrial houses, schools, multi-specialty hospitals, governmen agencies, retailers or other business groups, the functioning in gene will be based on specific modules or systems which should essentia be addressed.

The enterprise software includes many functions and abilities:

- Salespeople can log client information, including personal a product information and client location in the sales pipeline
- Customer service representatives can track client issues a communication alongside internal responses and updates.
- Internal-facing employees can take client specifications to bui the product.
- Managers and C-level executives can track client and employe performance.

Common uses of enterprise software:

- Order processing
- Procurement
- Scheduling
- Customer information management
- Energy management
- Accounting

Enterprise software offers computer-based business tools such as:

- **CRM (customer relationship management)**

  Customer relationship management software is a contact list with a brain. This type of enterprise software works by giving you an invaluable overview of where you are with each client, pulling together customer contact information and details of when and how you've been in touch. The best CRM services include features like workflow automation, dedicated document libraries, role-based user access, multi-currency support, email integration and tracking, and lead capture from websites and custom reporting.

- **Project Management Tool**

  These are designed to help your collaborate more effectively, even if your team is scattered all over the globe. They make sure everyone involved has clearly defined targets and deadlines, and can drop in at any time to check on progress and make sure they have what they need from their colleagues to get their own work done.

  Usually, tools like these give an overview of all active projects, but allow you to dive into each one to check every milestone, every detail, and every upcoming task you need to complete to keep your progress on schedule.

- **BI (business intelligence)**

  BI software works by pulling in data from all your data sources, harmonizing and combining it, and allowing you to run queries that extract actionable insights from these disparate sources. These insights are presented in dashboards that tell you in a glance how the business is doing and how much progress you are making .

- **ERP (Enterprise Resource Planning)**

  An ERP system basically takes all the different modules of the business, digitizes them, and creates complex networks between them that mean everyone is working from the same core records and up-to-date information. The system also integrates with other important enterprise software tools, including CRM, BI,

Scanned with CamScanner

financial information, inventories, asset and supply cha management technology, materials resource management, poi of sale data

- **Marketing Automation**

  This is why enterprise software for marketing automation is a handy. These apps and tools allow you to automa communication with your audience via email, SMS, socia media and/or digital ads; quickly and easily segment your list create targeted marketing campaigns; and set things up so you contact targets after a specified action or an optimal period time.

Other tools include:

- Human Resource Management,
- Online Payment Processing,
- Automated Billing Systems,
- Enterprise Content Management,
- IT Service Management,
- Email Marketing System, Call Center Support System, etc.

3.3 ERP (Enterprise Resource Planning)

Enterprise resource planning (ERP) is business process managemen software that allows an organization to use a system of integrated applications to manage the business and automate many back office functions related to technology, services and human resources.

ERP software typically integrates all facets of an operation : including product planning, development, manufacturing, sales and marketing in a single database, application and user interface.

Function of ERP

- Bookkeeping & Accounting
- Human Resource Management
- Planning Production
- Supply Chain management

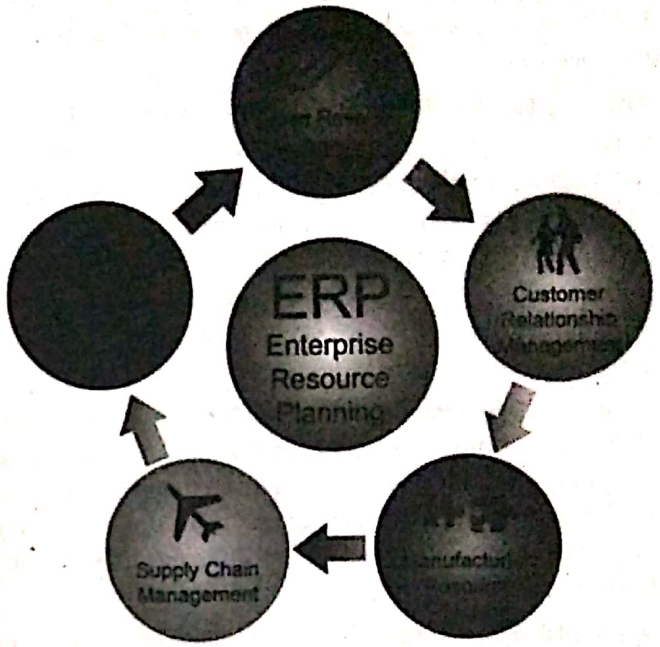


Fig 1: ERP Component

ERP Software

Enterprise resource planning or ERP software is a suite of applications that manages core business processes, such as sales, purchasing, accounting, human resource, customer support, CRM and inventory. It's an integrated system as opposed to individual software designed specifically for business process.

- 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.

Scanned with CamScanner

ERP can be fully integrated or customized to specific processes. A typical ERP system covers key business processes and consists of the following modules:

- **HRM (Human Resource Management) :** This module features standard HRMS tools like time tracker, timesheet, and database for employee records, job profiles, and skills matrix. HRM module may also include performance reviews and payroll system. The latter is closely integrated with the finance management module to manage wages, travel expenses and reimbursements. Some ERP solutions also feature a training or LMS function under HRM.
- **Financial management:** This module manages your capital inflow and outflow. It covers standard Accounting and Finance transactions like expenditures, general ledger, balance sheet, bank reconciliation, tax management, and payments. This also generates financial reports for different departments and business units.
- **Sales and Marketing:** The module handles sales workflows like sales inquiries, quotations, sales orders and sales invoices. The more advanced ERP also features taxation rules and shipping tracker. The Sales and CRM modules work together to speed up the sales cycle and earn the company more profits.
- **CRM (Customer Resource Management):** The CRM module helps you to boost customer service and, eventually, profit per capita. It manages leads, opportunities and customer issues. Likewise, it provides a 360-degree profile of your customers by consolidating data like their social media activities, purchase history and past interactions with support reps. In an ERP setup, CRM is closely integrated with Sales module to fast track conversions.
- **Manufacturing:** Sometimes referred to as engineering or production, this module helps businesses make manufacturing more efficient in areas such as product planning, material sourcing, daily production monitoring, and product forecasting. Some of the key functionalities in this module are bill of material, production scheduling, shop floor control, and distribution planning. The module is tightly integrated with SCM and inventory modules especially in areas like product planning and inventory control.
- **SCM (Supply Chain Management):** This module covers key aspects in your supply chain including purchase order management. It manages product flow from production to consumer and, occasionally, vice-versa for returns or recalls. A key feature of the SCM module is process automation, which streamlines your entire supply chain and makes it adaptive to sudden market shifts.
- **Inventory:** Also called material management module, it helps you measure stock targets, standardize replenishments, and other inventory goals. It uses product serial numbers to track and locate items in your organization. This module is closely integrated with the Purchase module.
- **Purchasing:** This module manages the processes involved in materials procurement. These include: supplier listings; quotation requests and analysis; purchase orders; Good Receipt Notes; and stock updates. As such, it functions closely with SCM or Inventory modules.

Enterprise ERP Trends

The ERP field can be slow to change, but the last couple of years have unleashed new technology trends which are fundamentally shifting the entire area. The following new and continuing computing trends have an impact on the growth of enterprise ERP software:

- **Mobile ERP**

  Executives and employees want real-time access to information, regardless of where they are. It is expected that businesses will embrace mobile ERP for the reports, dashboards and to conduct key business processes.
- **Cloud ERP**

  The cloud has been advancing steadily into the enterprise for some time, but many ERP users have been reluctant to place data in the cloud. Those reservations have gradually been evaporating, however, as the advantages of the cloud become apparent.

Scanned with CamScanner

- **Social ERP**

  There has been much hype around social media and important —or not — it is to add to ERP systems. Certain vendors have been quick to seize the initiative, adding social media packages to their ERP systems with much fanfare. some wonder if there is really much gain to be had integrating social media with ERP.

- **Two-tier ERP**

  Enterprises once attempted to build an all-encompassing ERP system to take care of every aspect of organizational system. But some expensive failures have gradually brought about change in strategy – adopting two tiers of ERP.

3.4 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. A supply chain is a network between a company and its suppliers to produce and distribute a specific product to the final buyer. This network includes different activities, people, entities, information, and resources. The supply chain also represents the steps it takes to get the product or service from its original state to the customer. Companies develop supply chains so they can reduce the costs and remain competitive in the business landscape.

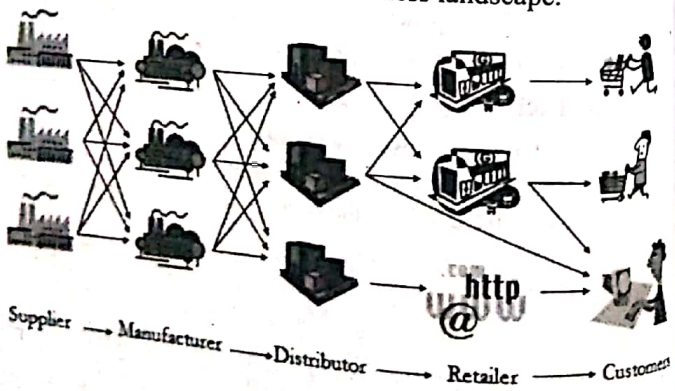


Fig 2: Supply Chain

A supply chain involves a series of steps involved to get a product or service to the customer. The steps include moving and transforming raw materials into finished products, transporting those products, and distributing them to the end user. The entities involved in the supply chain include producers, vendors, warehouses, transportation companies, distribution centers, and retailers. The elements of a supply chain include all the functions that start with receiving an order to meeting the customer's request. These functions include product development, marketing, operations, distribution, finance, and customer service.

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. Supply chain management is the management of the flow of goods and services and includes all processes that transform raw materials into final products. It involves the active streamlining of a business's supply-side activities to maximize customer value and gain a competitive advantage in the marketplace. SCM represents an effort by suppliers to develop and implement supply chains that are as efficient and economical as possible. Supply chains cover everything from production to product development to the information systems needed to direct these undertakings.

SCM attempts to centrally control or link the production, shipment, and distribution of a product. By managing the supply chain, companies are able to cut excess costs and deliver products to the consumer faster. Keeping tighter control of internal inventories, internal production, distribution, sales, and the inventories of company vendors does this. SCM is based on the idea that nearly every product that comes to market results from the efforts of various organizations that make up a supply chain. Although supply chains have existed for ages, most companies have only recently paid attention to them as a value-add to their operations.

Benefits of using SCM

- Higher Efficiency Rate
- Reduce Cost Effects

Scanned with CamScanner

- Raise Output
- Raised Your Business Profit Level
- Boost Cooperation
- Lowers Delay in Processes
- Enhanced Supply Chain Network

3.5 Customer Relationship Management (CRM)

Customer relationship management is a strategy for managing all your company's interactions with current and prospective customers. CRM form a bond between a company and its customers. CRM enables your company to increase productivity, close more business, and improve customer satisfaction and retention.

CRM Strategies

Fig 3: 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

Goals of CRM

- **Consolidate customer information**: CRM gives you the ability to consolidate customer information and documents into a single database which can be easily accessed and used by all members of your organization.
- **Obtain better lead management**: Having one convenient place for your sales teams to input, analyze and track leads allow for more effective management of this data.
- **Automate tasks**: By automating tasks associated with marketing, sales and call centers, you can streamline these efforts and better prioritize your agent's time.
- **Manage your workforce**: Through enhanced tracking of employee information such as job performance stats and reviews, you can better manage and optimize the efforts of your sales and marketing teams.

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

3.6 Enterprise Information Management (EIM)

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

Scanned with CamScanner

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.

Process of EIM

- Organize
- Evaluate
- Develop
- Implement
- Manage
- Audit

Benefits of implementing EIM in an Organization:

- **Business Insight:** Manage growing volumes of information across all formats and channels for optimized strategic analysis and decision-making
- **Process Velocity:** Integrate existing information, processes and applications for increased process velocity, efficiency, and agility
- **Business Impact:** globalization, social decision making, and the "consumerization" of IT to create exceptional experiences for customers, users, and partners
- **Information Governance:** Enforce information governance policies and ensure compliance
- **Information Security:** Protect valuable enterprise information from outside intruders and inside leaks

Challenges of EIM

- **Accurate** – Lack of control over how data is created, updated, and stored causes serious information errors .
- **Complete** – An inability to reach all sources and combine and consolidate the data contained within them hinders the thoroughness of the information that stakeholders use to carry out their day-to-day activities .
- **Consistent** – Fragmented business applications often result in multiple versions of the truth and can negatively impact cross-departmental coordination, resulting in misinformed or incorrect business decisions .
- **Relevant** – To be valuable, information must first be useful. If employees cannot access the relevant, contextual data they need to perform their jobs, they will be less productive .
- **Timely** – Managed information becomes worthless if the information is not delivered in a timely manner when needed.

3.7 Enterprise IT Management (EITM)

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.

3.8 Enterprise Information Systems (EIS)

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)

Scanned with CamScanner

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

3.9 Enterprise Software

Enterprise software, also known as enterprise application software (EAS), is purposed-designed computer software used in the furtherance of the needs and objectives of the organizations. Purpose to use the EAS can vary widely as in a business, schools, interest-based user groups and clubs, retailers, or government. Enterprise software describes a collection of computer programs with common business applications, tools for modeling how the entire organization works, and development tools for building applications unique to the organization. Enterprise application software performs business functions such as order processing, procurement, production scheduling, customer information management, energy management and accounting. It is typically hosted on servers and provides simultaneous services to a large number of users, typically over a computer network.

Different types of industry standard Enterprise Software product categories have emerged, and these are shown below:

- Accounting software
- Business intelligence
- Business process management
- Content management system (CMS)
- Customer relationship management (CRM)
- Master data management (MDM)
- Enterprise resource planning (ERP)
- Enterprise asset management (EAM)

3.10 Enterprise Engineering

Enterprise engineering is an 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. It encompasses the application of knowledge, principles, and disciplines related to the analysis, design, implementation and operation of all elements associated with an enterprise. In essence it is an interdisciplinary field, which combines systems engineering and strategic management as it seeks to engineer the entire enterprise in terms of the products, processes, and business operations. The view is one of continuous improvement and continued adaptation as firms, processes and markets develop along their life cycles. This total systems approach encompasses the traditional areas of research and development, product design, operations and manufacturing as well as information systems and strategic management.

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.

3.11 Enterprise Integration

- Integration of markets
- Integration between several development and manufacturing sites

Scanned with CamScanner

- 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

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.

- 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.

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)

3.12 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.
- 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.

Scanned with CamScanner

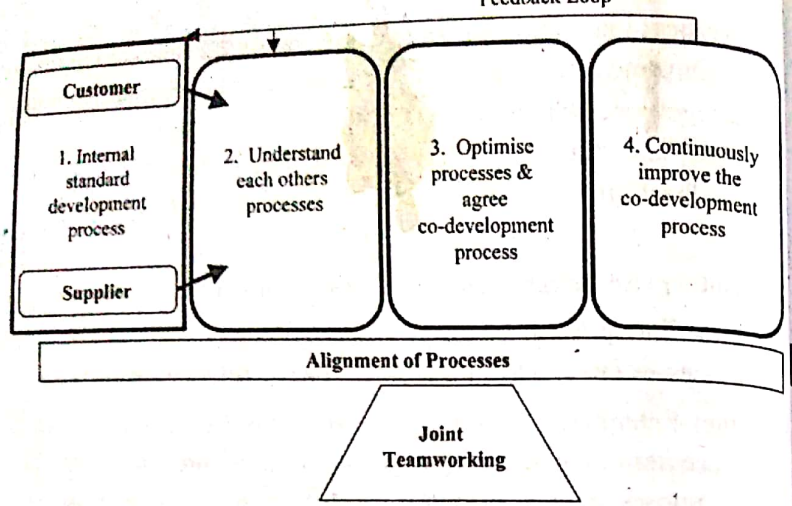


Fig 4: Alignment Process

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.

3.13 Electronic Organisms

A digital or electronic organism is a self-replicating computer program that mutates and evolves. All electronic organism:

- have to be complex, because they have to contain all the creative infrastructure necessary for their creation, reproduction, maintenance and action,
- but they can easily afford to be complex, because there is no need for detailed communication with a programmer.
- have the ability to react immediately to unforeseen challenges, without the need for a programmer to recognize the situation and deal with it by modifying a program.

- Electronic Organisms do so by recurrence to fundamental goals and organizing principles, just as programmers do so now.
- Electronic organisms will live, grow and evolve in the rapidly growing world of installed computers and networks, just as microbes, plants and animals live in natural ecosystems.
- Strong forces are pushing technology towards electronic organisms.

PRACTICE QUESTIONS

1. What is difference between traditional Information system and Enterprise System Explain with example?
2. What do you understand by electronic organization (digital firm)? Discuss the concept of Enterprise management system.
3. What is the use of ERP system in enterprise? Discuss key feature and functionality of an ERP software system.
4. Explain in detail Enterprise Resource planning. Supply chain management is a top strategic objective for much organization. Explain.
5. Explain the role of IS and IT in enterprise Management System.
6. What do you understand by an Enterprise System Discuss the architecture of Enterprise systems? What reasons do you think that organization is motivated to establish enterprise system?
7. What is the use of CRM systems in enterprise? Discuss key features and functionality of a CRM system.
8. How does the Enterprise system differ conventional information system? Describe various IT tools that are used to enhance Customer Relationship Management (CRM) system in a web-based environment.
9. What do you understand by enterprise system? Discuss the architecture of enterprise System. For what reasons do you think the organization is motivated to establish enterprise system?
10. Write short note on: ERP, CRM, SCM

Scanned with CamScanner

CHAPTER

DECISION SUPPORT AND INTELLIGENT SYSTEM

4.1 Decisions

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.

Four different stages in decision-making are

- Intelligence
- Design
- Choice
- Implement

4.2 Decision Support System

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. Decision Support System 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 this phase again and again till the satisfactory solution is found.

It is an integrated set of computer tool that allows a decision maker to interact directly with computers to create information, which is 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. It helps decision makers and to identify and solve problems and 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
- The nature of the decision is such that the decision makers need a variety of information methods of decision making a change from time to time.

Following are some typical information that decision support system might gather and present:

- Comparative sales figures between one week and the next
- Projected revenue figures based on new product sales assumptions
- The consequences of different decision alternatives, given past experience in a context that is described

Decision Making Stages

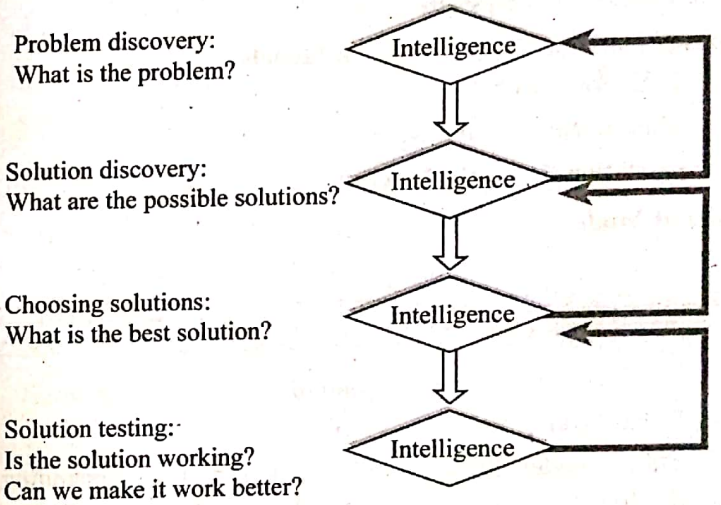


Fig 1: Decision making stages

DSS: Types of Decisions

structured semistructured.
unstructured

Decision may vary from structured to unstructured. Structured decisions are made in the situations, which are fully understood while, in unstructured decisions the situations are uncertain and unclear. Structured decisions are generally made for routine tasks, for instance the hiring of new IT specialists in a firm while unstructured decisions

Scanned with CamScanner

are made for a sudden one-shot kind of situations, for in
dealing with a labor strike in a factory.

Structured decisions are made for specified processes like spec
manufacturing processes while unstructured decisions are ma
general processes. Structured decisions have a well-de
methodology for finding a solution and have the data to
decisions. They are usually straight forward and made on a re
basis, an example of a structured decision in my company is wh
or not to withdraw funds from an international account depend
the current exchange rate while Unstructured decision rel
knowledge and/or expertise and often require data and mode
solve, an example of an unstructured decision in my company is
types of new content should be created and what market shou
targeted.

Decision Support System Models

Types of Decision Support System Models:

1. Behavioral Models
2. Management Science Models
3. Operation Research Models

Types of Models

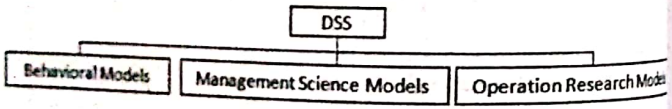


Fig 2: Types of DSS

1. **Behavioral Models:**

These model Focuses on studying and understanding
different behavior/trends amongst the variables. A behav
model is built by observing the previous behavior of an enti
a system; the resulting model can then be used to predict f
behavior and performance. Examples of such a model ca
trend analysis, co-relation, regression etc.

- The decision maker can make the decisions for
behavioral relationships.

For e.g.: 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 is designed and computerized to evaluate customer's buying behavioral).

2. **Management Science Models**

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

  For e.g.: 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 e.g.: Sales Budget, Production Budget

3. **Operation Research Models**

These are the mathematical models in providing guidelines to managers for making effective decisions within the state of the current information and are based on mathematical formula.

Some applications

- Telecommunications/Road/Rail Network Design
- Organization Supply Chain Strategy
- Revenue, Pricing and Promotions
- Demand Forecasting
- Project Planning

For e.g.: 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.

Scanned with CamScanner

Benefits of Decision Support System

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

4.3 Group Decision Support System

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

There are four configurations of group members are possible:

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

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 e.g. 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

4.4 Enterprise Decision Support System

Enterprise decision support system is an enterprise approach that applies analytical and rule-based systems to manage and deploy all operational decisions, such as relationships with employees, suppliers and customers.

The computerized EDSS movement altered the enterprise decision-making process by incorporating information-based decisions based on historical behavioral data, as well as prior decisions and their outcomes.

4.5 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:

Scanned with CamScanner

- Rapid access to timely information;
- Direct access to management reports;
- Very user friendly and supported by graphics.

Executive Decision Support Systems support the informational role of executives.

- **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.

4.6 Knowledge Management

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

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.

Sources of Knowledge of an Organization

- Intranet
- Data warehouses and knowledge repositories
- Decision support tools
- Groupware for supporting collaboration
- Networks of knowledge workers
- Internal expertise

4.7 Knowledge Management System (KMS)

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 Management System

- **Create :** Knowledge must be created either within or outside the organization.
- **Store :** Knowledge can be stored somewhere, 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.
- **Acquire :** Once the knowledge source is found, the user will then go through the act of actually acquiring it
- **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.

Purpose of KMS

- Improved performance
- Competitive advantage
- Innovation
- Sharing of knowledge
- Integration
- Continuous improvement by –
  - Driving strategy
  - Starting new lines of business
  - Solving problems faster
  - Developing professional skills
  - Recruit and retain talent

Scanned with CamScanner

4.8 Expert System

Expert System is an extension of the decision support system. Exp system is an information system application that captures knowledge and expertise of a problem solver or decision maker, then simulates the 'thinking' of that expert for those who have expertise. A Knowledge Based System is a computer program uses artificial intelligence to solve problems within a specializ domain that ordinarily requires human expertise. It involv classification, diagnosis, monitoring, design, scheduling, and plann for specialized tasks.Knowledge-based system is a more general the expert system. Expert systems are implemented with artific intelligence technology, often called expert system shells.

Expert System application areas in action

1. Medical Diagnosis
2. Telephone Network Maintenance

Capabilities of Expert Systems

The expert systems are capable of :

- Advising
- Instructing and assisting human in decision making
- Demonstrating
- Deriving a solution
- Diagnosing
- Explaining
- Interpreting input
- Predicting results
- Justifying the conclusion
- Suggesting alternative options to a problem

They are incapable of :

- Substituting human decision makers
- Possessing human capabilities
- Producing accurate output for inadequate knowledge base
- Refining their own knowledge

Expert System Architecture

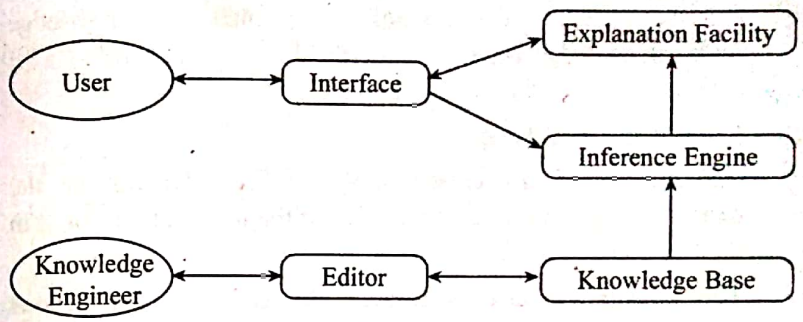


Fig 3: Expert System Architecture

Components of Expert Systems (ES)

The components of ES include:

- Knowledge Base
- Inference Engine
- User Interface

a. Knowledge Base

It contains domain-specific and high-quality knowledge.

Knowledge is required to exhibit intelligence. The success of any ES majorly depends upon the collection of highly accurate and precise knowledge.

What is Knowledge?

The data is collection of facts. The information is organized as data and facts about the task domain. Data, information, and past experience combined together are termed as knowledge.

Components of Knowledge Base

The knowledge base of an ES is a store of both, factual and heuristic knowledge.

- Factual Knowledge – It is the information widely accepted by the Knowledge Engineers and scholars in the task domain.
- Heuristic Knowledge – It is about practice, accurate judgement, one's ability of evaluation, and guessing.

Scanned with CamScanner

Knowledge representation

It is the method used to organize and formalize the knowledge in the knowledge base. It is in the form of IF-THEN-ELSE rules.

Knowledge Acquisition

The success of any expert system majorly depends on the quality, completeness, and accuracy of the information stored in the knowledge base.

The knowledge base is formed by readings from various experts, scholars, and the **Knowledge Engineers**. The knowledge engineer is a person with the qualities of empathy, quick learning, and case analyzing skills.

He acquires information from subject expert by recording, interviewing, and observing him at work, etc. He then categorizes and organizes the information in a meaningful way, in the form of IF-THEN-ELSE rules, to be used by interference machine. The knowledge engineer also monitors the development of the ES.

b. Inference Engine

Use of efficient procedures and rules by the Inference Engine is essential in deducting a correct, flawless solution. In case of knowledge-based ES, the Inference Engine acquires and manipulates the knowledge from the knowledge base to arrive at a particular solution.

In case of rule based ES, it:

- Applies rules repeatedly to the facts, which are obtained from earlier rule application.
- Adds new knowledge into the knowledge base if required.
- Resolves rules conflict when multiple rules are applicable to a particular case.

To recommend a solution, the Inference Engine uses the following strategies:

- Forward Chaining
- Backward Chaining

Forward Chaining

It is a strategy of an expert system to answer the question, **"What can happen next?"**

Here, the Inference Engine follows the chain of conditions and derivations and finally deduces the outcome. It considers all the facts and rules, and sorts them before concluding to a solution.

This strategy is followed for working on conclusion, result, or effect. For example, prediction of share market status as an effect of changes in interest rates.

Backward Chaining

With this strategy, an expert system finds out the answer to the question, **"Why this happened?"**

On the basis of what has already happened, the Inference Engine tries to find out which conditions could have happened in the past for this result. This strategy is followed for finding out cause or reason. For example, diagnosis of blood cancer in humans.

c. User Interface

User interface provides interaction between user of the ES and the ES itself. It is generally Natural Language Processing so as to be used by the user who is well-versed in the task domain. The user of the ES need not be necessarily an expert in Artificial Intelligence.

It explains how the ES has arrived at a particular recommendation. The explanation may appear in the following forms –

- Natural language displayed on screen.
- Verbal narrations in natural language.
- Listing of rule numbers displayed on the screen.

The user interface makes it easy to trace the credibility of the deductions.

Requirements of Efficient ES User Interface

- It should help users to accomplish their goals in shortest possible way.

Scanned with CamScanner

- It should be designed to work for user's existing or desired work practices.
- Its technology should be adaptable to user's requirements, not the other way round.
- It should make efficient use of user input.

Expert Systems Limitations

No technology can offer easy and complete solution. Large systems are costly, require significant development time, and computer resources. ES have their limitations, which include –

- Limitations of the technology
- Difficult knowledge acquisition
- ES are difficult to maintain
- High development costs

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

- 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

4.9 Knowledge Engineer

A knowledge engineer is a computer scientist who knows how to design and implement programs that incorporate artificial intelligence techniques. 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. 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:

An expert system is judged to be entirely successful when it operates on the level of a human expert.

4.10 Intelligent System

An intelligent system is a machine with an embedded, Internet-connected computer that has the capacity to gather and analyze data and communicate with other systems. Other criteria for intelligent systems include the capacity to learn from experience, security, connectivity, the ability to adapt according to current data and the capacity for remote monitoring and management. They help improve energy efficiency, quality, and flexibility of these systems. Intelligent systems are complex and use a wide range of technologies: artificial intelligence, cyber security, natural language processing, deep learning, embedded CPUs, distributed storage, wireless networking and graphical signaling.

Scanned with CamScanner

Applications of intelligent systems

Intelligent systems are poised to fill a growing number of role today's society, including:

- Factory automation
- Field and service robotics
- Assistive robotics
- Military applications
- Medical care
- Education
- Entertainment
- Visual inspection
- Character recognition
- Human identification using various biometric modalities (e face, fingerprint, iris, hand)
- Visual surveillance
- Intelligent transportation

Challenges in intelligent systems

Research in intelligent systems faces numerous challenges, many which relate to representing a dynamic physical wo computationally.

1. **Uncertainty**: Physical sensors/effectors provide limited, no and inaccurate information/action. Therefore, any actions t system takes may be incorrect both due to noise in the senso and due to the limitations in executing those actions.
2. **Dynamic world**: The physical world changes continuous requiring that decisions be made at fast time scales accommodate for the changes in the environment.
3. **Time-consuming computation**: Searching for the optimal pa to a goal requires extensive search through a very large sta space, which is computationally expensive. The drawback spending too much time on computation is that the world ma change in the meantime, thus rendering the computed pla obsolete.
4. **Mapping**: A lot of information is lost in the transformati from the 3D world to the 2D world. Computer vision must de

with challenges including changes in perspective, lighting and scale; background clutter or motion; and grouping items with intra/inter-class variation.

4.11 Artificial Intelligence

Artificial intelligence (AI) is an area of computer science that emphasizes the creation of intelligent machines that work and react like humans. It has become an essential part of the technology industry.

Research associated with artificial intelligence is highly technical and specialized. The core problems of artificial intelligence include programming computers for certain traits such as:

- Knowledge
- Reasoning
- Problem solving
- Perception
- Learning
- Planning
- Ability to manipulate and move objects

Areas of Artificial Intelligence

Three major areas

- Cognitive science
- Robotics
- Natural interfaces

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

Scanned with CamScanner

4.12 Neural Networks (NN)

Neural networks are a system of programs and data structures t approximates the operation of the human brain and is good recognizing subtle (sensitive), hidden and newly emerging patte within complex data interpreting incomplete inputs. Neural networ with their remarkable ability to derive meaning from complicated imprecise data, can be used to extract patterns and detect trends t are too complex to be noticed by either humans or other compu techniques. A trained neural network can be thought of as an "expe in the category of information it has been given to analyze. T expert can then be used to provide projections given new situations interest and answer "what if" questions.

Other advantages include:

- **Adaptive learning:** An ability to learn how to do tasks bas on the data given for training or initial experience.
- **Self-Organization:** A NN can create its own organization representation of the information it receives during learni time.
- **Real Time Operation:** NN computations may be carried out parallel, and special hardware devices are being designed a manufactured which take advantage of this capability.
- **Fault Tolerance via Redundant Information Coding:** Parti destruction of a network leads to the corresponding degradatio of performance. However, some network capabilities may b retained even with major network damage.

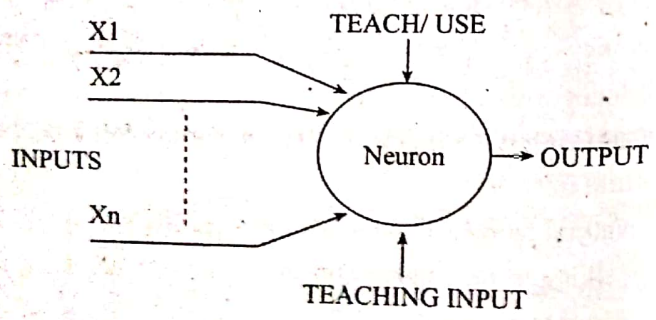


Fig 4: Neural network

Key Advantages of Neural Networks:

NNs have some key advantages that make them most suitable for certain problems and situations:

- NNs have the ability to learn and model non-linear and complex relationships, which is really important because in real-life, many of the relationships between inputs and outputs are non-linear as well as complex.
- NNs can generalize-After learning from the initial inputs and their relationships, it can infer unseen relationships on unseen data as well, thus making the model generalize and predict on unseen data.

Unlike many other prediction techniques, NN does not impose any restrictions on the input variables (like how they should be distributed). Additionally, many studies have shown that ANNs can better model data with high volatility and non-constant variance, given its ability to learn hidden relationships in the data without imposing any fixed relationships in the data. This is something very useful in financial time series forecasting (e.g. stock prices) where data volatility is very high.

A few applications:

NNs, due to some of its wonderful properties have many applications:

1. **Image Processing and Character recognition:** Given NNs ability to take in a lot of inputs, process them to infer hidden as well as complex, non-linear relationships, NNs are playing a big role in image and character recognition. Character recognition like handwriting has lot of applications in fraud detection (e.g. bank fraud) and even national security assessments. Image recognition is an ever-growing field with widespread applications from facial recognition in social media, cancer detention in medicine to satellite imagery processing for agricultural and defense usage. The research on NN now has paved the way for deep neural networks that forms the basis of "deep learning" and which has now opened up all the exciting

Scanned with CamScanner

and transformational innovations in computer vision, speech recognition, natural language processing—famous examples being self-driving cars.

2. **Forecasting:** Forecasting is required extensively in everyday business decisions (e.g. sales, financial allocation between products, capacity utilization), in economic and monetary policy, in finance and stock market. More often, forecasting problems are complex, for example, predicting stock prices is a complex problem with a lot of underlying factors (some known, some unseen). Traditional forecasting models throw up limitations in terms of taking into account these complex, non-linear relationships. NNs, applied in the right way, can provide robust alternative, given its ability to model and extract unseen features and relationships. Also, unlike these traditional models, NN doesn't impose any restriction on input and residual distributions.

4.13. Virtual Reality (VR)

Virtual reality is the use of computer technology to create a simulated environment. Unlike traditional user interfaces, VR places the user inside an experience. Instead of viewing a screen in front of them, users are immersed and able to interact with 3D worlds. By simulating as many senses as possible, such as vision, hearing, touch, even smell, the computer is transformed into a gatekeeper to this artificial world. The only limits to near-real VR experiences are the availability of content and cheap computing power.

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 tele operation or tele presence.

- **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.
- **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.

4.14 Data Mining

Data mining is a process used by companies to turn raw data into useful information. Data mining is the search for relationships and global patterns that exist in large databases but are 'hidden' among the vast amount of data. Data Mining is also known as knowledge Discovery in Databases (KDD). 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.

Scanned with CamScanner

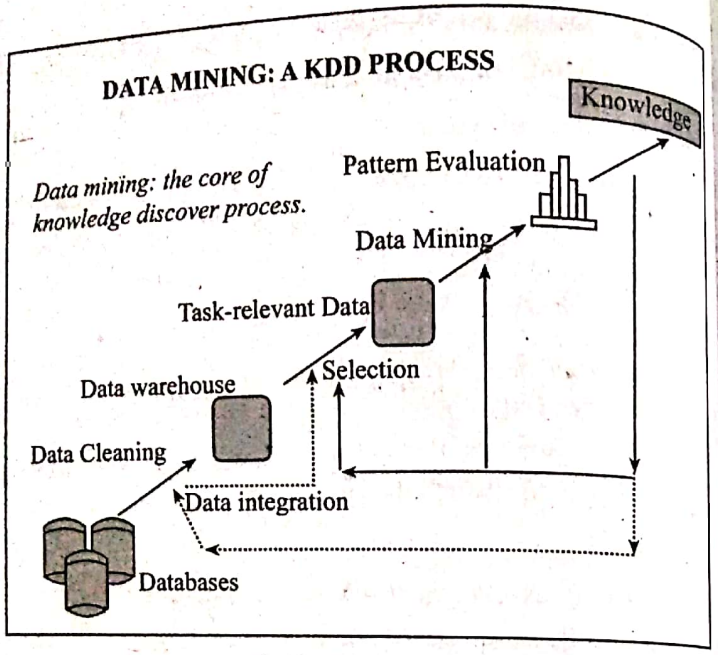


Fig 5: KDD Process

Steps of 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"
- Credit ratings/targeted marketing
- Fraud detection
- Customer relationship management

4.15 Data Warehouse

A data warehousing is defined as a technique for collecting and managing data from varied sources to provide meaningful business insights. It is a blend of technologies and components which aids the strategic use of data.

It is electronic storage of a large amount of information by a business which is designed for query and analysis instead of transaction processing. It is a process of transforming data into information and making it available to users in a timely manner to make a difference.

Characteristics

- Subject oriented, Integrated, Time Variant, Non-volatile
- Web-based, Relational/multidimensional, Client/server, Real-time
- Include metadata

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

Scanned with CamScanner

- Constructed by integrating multiple, heterogeneous da sources, relational databases, flat files, on-line transactio records
- Data cleaning and data integration techniques are applied.
- Ensure consistency in naming conventions, encoding structure attribute measures, etc. among different data sources
- **E.g., Hotel price:** currency, tax, breakfast covered, etc.

Data Warehouse—Integrated

- Constructed by integrating multiple, heterogeneous da sources: relational databases, flat files, on-line transactio 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

4.16 OLTP vs. OLAP

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

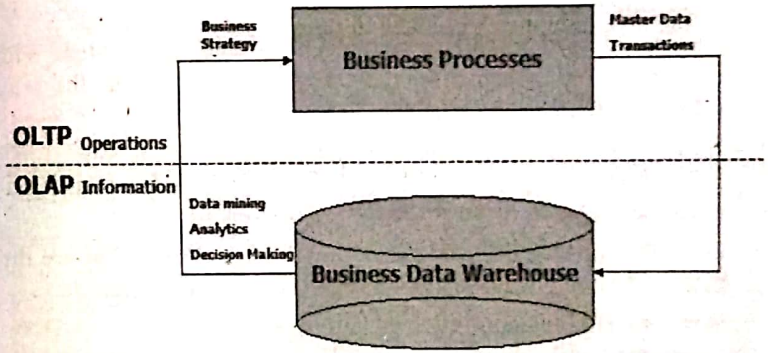


Fig 6: OLAP and OLTP

- OLTP (On-line Transaction Processing) is characterized by a large number of short on-line transactions (INSERT, UPDATE, DELETE). The 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. In OLTP database there is detailed and current data, and schema used to store transactional databases is the entity model (usually 3NF).

Scanned with CamScanner

- OLAP (On-line Analytical Processing) is characterized by relatively low volume of transactions. Queries are often very complex and involve aggregations. For OLAP systems, response time is an effectiveness measure. OLAP applications are widely used by Data Mining techniques. In OLAP databases there is aggregated, historical data, stored in multi-dimensional schemas (usually star schema).

PRACTICE QUESTIONS

1. Define and explain Decision Support System (DSS) with suitable diagram. Discuss key characteristics of GDSS system.
2. What is Data Warehouse? Why do you require data warehouse along with operational database? How is it related with data mining?
3. Discuss four different areas of Information system that use data analytics and related techniques.
4. Describe GDSS for modern organization. How do they help organizational learning?
5. How do you differentiate between MIS and DSS system with respect to the information they present to users and the way information are processed?
6. How does decision support system help an organization? Explain the component of DSS. List some scope of operations research.
7. Describe the role of Analytical Modeling in DSS. What are the differences between MIS and Executive Support System?
8. Write short notes on:
   a. OLTP vs OLAP
   b. Data Mining and Data Warehouse

CHAPTER – 5

PLANNING FOR IS

5.1 Information System Planning

- **Business planning** – the process of identifying the firm's goals, objectives, and priorities + developing action plans for accomplishing them.
- **Information systems planning** – the part of business planning concerned with developing the firm's information systems resources

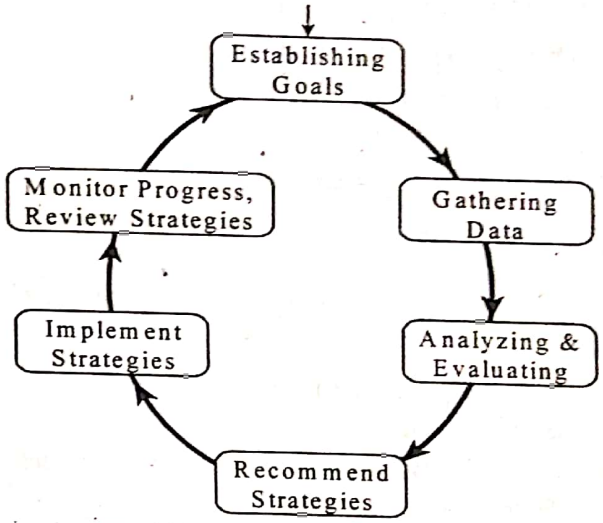


Fig 1: Information System Planning

Challenges in Business Planning

- Foreseeing and assessing opportunities
- Assuring consistency with organizational plans and objectives
- Building systems
- Maintaining information system performance
- Collaborating with IT professionals

Principles for IS Planning

- Support the firm's business strategy with appropriate technical architecture

Scanned with CamScanner

- Evaluate technology as a component of a larger system
- Recognize life cycle costs, not just acquisition costs

5.2 Information System Planning

- Information system planning is a formal process that develo plan for developing and managing information systems that w 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? (Importance of plan)

- To obtain resources
- To align Information System with the business
- To identify needed applications
- To establish goal, schedule, and milestone in order to tra progress
- To provide an opportunity for communication with t management and user management

Approaches to Planning (4 - 6 marks)

- **Top-down Planning**

  Focuses on organizational goals first, then on the needs business units
- **Bottom-up Planning**

  Focuses on needs of business units first, then on organization 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 impleme 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

- 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-Range, Medium-Range, 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.

Key Elements of IS Planning

- Corporate mission statement
- Vision for IT within organization

Scanned with CamScanner

- IS strategic and tactical plans
- Operations plan to achieve mission and vision
- Budget to ensure resources are available

Steps of Information Systems Planning

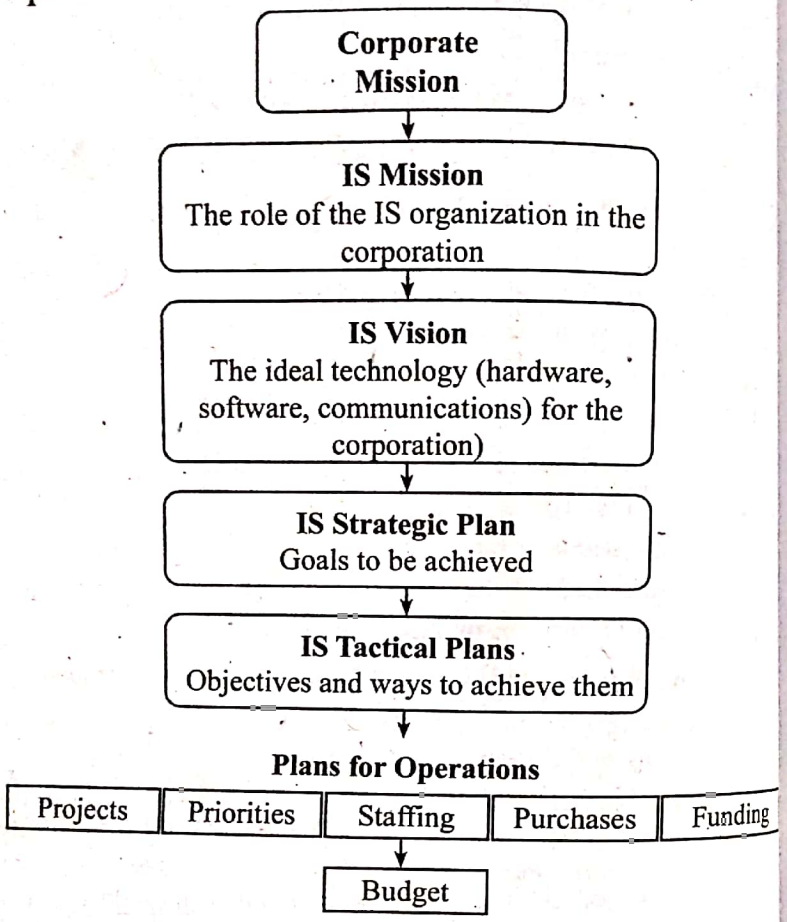


Fig 2: Steps of Information System Planning

5.3 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.

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

Types of IS Planning

a. 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.
- 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.

Fig 3: Strategic Planning

Scanned with CamScanner

Strategic Information System

- Strategic Information System is a system that companies alter their business strategy. It is used accelerate the reaction time to environmental changes aid the company in achieving a competitive advantage 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 advantages over the competitive forces it faces in the global marketplace.

Typical Strategic-Level Job Titles

- Top-level Managers
- General Manager
- Steering Committee
- CEOs or Presidents
- Corporate Boards
- Board of Directors

b. Tactical IS Planning

- IS tactical plan describes how goals will be met and by whom
- 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 development projects, and develops allocation plans financial and technology resources.

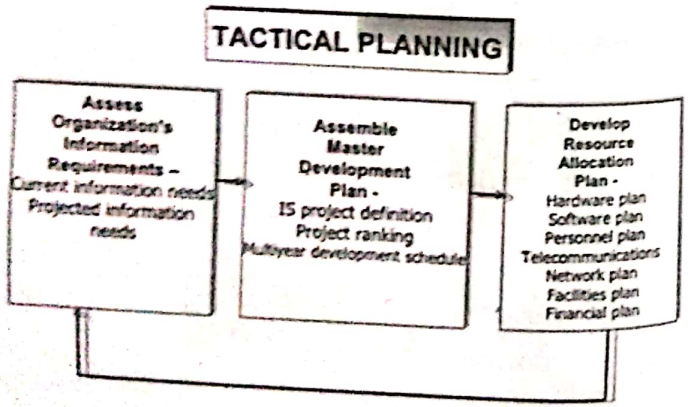


Fig 4: Tactical Planning

Important Factors in IS Tactical Planning

- Flexibility
- Connectivity
- Standardization
- Compatibility
- Scalability
- Total Cost of Ownership

Tactical-Level Job Titles

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

c. 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.

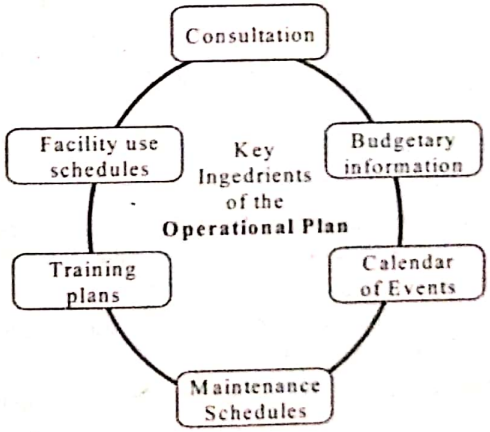


Fig 5: Operational IS Planning

Scanned with CamScanner

5.4 Three Levels of Planning (6 marks)

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 bro and genera
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 a ideas, b not ve 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, wh where a when)

PRACTICE QUESTIONS

1. What do you mean by strategic information system? What a different benefits of using strategic information system?
2. Explain change management IS planning.
3. What do you understand by balanced scorecard approa (BSC)? Discuss the importance of BSC for managing implementation.
4. Why IS planning is needed for an organization. Expla Strategic Information System.
5. Differentiate Strategic Plan, Tactical Plan and Operational Pla

CHAPTER – 6

IMPLEMENTATION OF INFORMATION SYSTEM

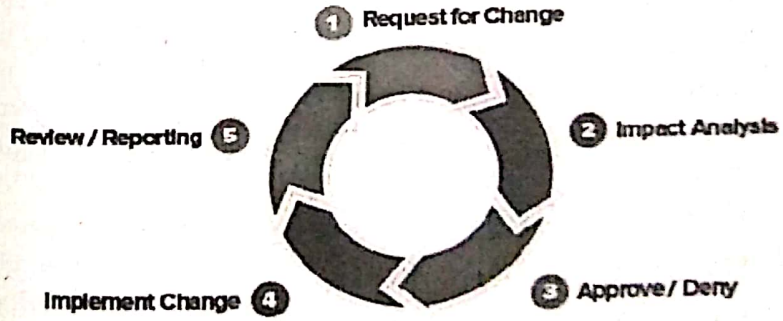
Organization Transformation

Major organizational changes such as transformation to e-business are referred to as organization transformation. This process usually requires change management. Change refers to an organization with a "new face," whose business processes, structure, strategy, and procedures are completely different from the old one. Such a radical transformation can be a lengthy, expensive, and complex process, which may involve organizational learning, changes in management and personnel, creation of a new structure, and employee retraining.

6.1 Change Management

Refers to the implementation, control and guidelines to introduce change into organizations. Changing business processes, organizational structure and operating procedures are interrelated and depending upon the magnitude of the change can be met with employee resistance. Since change is a learning process its impact can be minimized if properly managed.

Change Management Process



Scanned with CamScanner

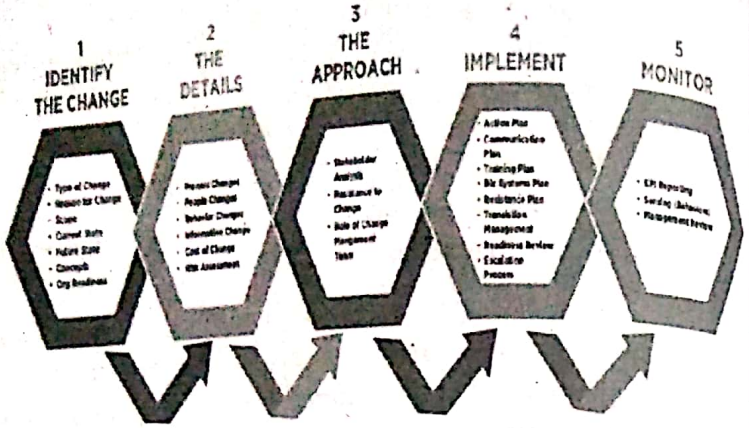


Fig 1: Change Management Process

Possible Responses of Change

- Productivity declines as people become more consumed with the change being introduced.
- Passive resistance festers.
- Active resistance emerges and sabotages the change.
- Valued employees leave the organization, a very costly proposition in terms of the lost contribution and the cost to replace them.
- Employees become disinterested in the current state and the future state.
- Employees begin arguing about change and the direction of the company.
- People are left to wonder why the change is happening.
- More people begin taking sick days or not showing up for work.
- People find work-around to avoid implementing the new way of doing things.
- Employees revert back to the old way of doing things.
- Changes are not fully implemented.
- Changes are scrapped and cancelled due to the lack of support throughout the organization

- Divides are created in the organization between 'us' and 'them'.
- The organization builds a history of failed and painful changes.
- Many types of risk are created - risk to the project, to the organization, to the employees involved and to the individuals supporting or chartering the change.

Benefits of Effectively Managed Change

- Employees have a solid understanding of why change is happening.
- Employees engage in both the solution and the change.
- Training is used to build knowledge after employees have made the personal decision to support the change.
- Resistance is identified and dealt with early in the process.
- Senior leaders demonstrate their own and the organization's commitment to the change.
- Communications are segmented and customized for different audiences, answering the questions that they care about.
- Momentum is built throughout different areas and levels within the organization.
- Changes are less painful to the organization and to the employees.
- A coalition of support among senior leaders and managers creates momentum throughout the organization.
- Probability of meeting project objectives is increased.
- The organization begins to build a history of successful change, creating a better 'backdrop' for the next change initiative.

Change Management Tools

Flowcharting

Flowcharting is a tool that lays out any process as an easy to read and easy to use diagram. The value of flowcharts is in taking complex processes and simplifying them. in flowcharting the current process, team members are able to map out the process clearly, discover where

Scanned with CamScanner

waste and rework are happening that they could not see before, a frequently identify quick and effective changes that can be made improve the process.

ADKAR Analysis

The ADKAR model in change management is used as a coaching to to make sure that the people or employees involved in the proce support and believe in the change. ADKAR stands for,

- Awareness
- Desire
- Knowledge
- Ability
- Reinforcement

- **Spread awareness**

  Clearly communicate to the employees what change occurring and why.

- **Inspire desire**

  Build desire in your employees to support the change so th would do their part naturally rather than forcing them to do so.

- **Impart knowledge**

  Make sure to help employees learn how to support the chang through training programs, job aids, tutorials, coachi programs and checklists provided by the company.

- **Improve ability**

  Help employees further improve their ability through feedba and evaluations.

- **Reinforce the changes**

  Make sure that employees don't go back to their old ways. order to reinforce the changes, give incentives and rewards employees.

Culture Mapping

A culture map is a tool that is used to visualize the company cultu which is influenced by values, norms, employee behavior etc. It hel

you discover information that is crucial to your change initiative, such as who are the positive enablers and how you can minimize risks during a project.

Step 1: Identify the subcultures

Identify the groups that belong to the different subcultures in your organization. These could be IT, finance, marketing, sales, etc teams in your organization. Identify 5-6 people that can best represent these subcultures.

Step 2: Carry out interviews

Interview each group to know the cultural blockers and enablers reinforced intentionally or unintentionally by the management.

Step 3; Organize the information

Organize the information such as the behaviors, enablers, blockers, current and desired outcomes etc. in a cultural map like the one below for quick analysis.

Force Field Analysis

Step 1: Identify the subcultures

Identify the groups that belong to the different subcultures in your organization. These could be IT, finance, marketing, sales, etc teams in your organization. Identify 5-6 people that can best represent these subcultures.

Step 2: Carry out interviews

Interview each group to know the cultural blockers and enablers reinforced intentionally or unintentionally by the management.

Step 3; Organize the information

Organize the information such as the behaviors, enablers, blockers, current and desired outcomes etc. in a cultural map like the one below for quick analysis.

Step 1: Identify the subcultures

Identify the groups that belong to the different subcultures in your organization. These could be IT, finance, marketing, sales, etc teams in your organization. Identify 5-6 people that can best represent these subcultures.

Scanned with CamScanner

Step 2: Carry out interviews

Interview each group to know the cultural blockers and enablers reinforced intentionally or unintentionally by the management.

Step 3; Organize the information

Organize the information such as the behaviors, enablers, blockers, current and desired outcomes etc. in a cultural map like the one below for quick analysis.

Stakeholder Analysis

In the stakeholder analysis, you identify stakeholders and categorize them based on different factors such as geography, income groups, occupation, legal requirements etc.

How to use it

Step 1: Identify the stakeholders

Think of all the people who might be affected by your project. Remember to consider those who come from outside of your company as well.

Step 2: Prioritize the stakeholders

Your list of stakeholders may include people who may be affected by and care about your project and, who may be unaffected by and be not that interested in your project.

Step 3: Stakeholder map

Now that you have prioritized your stakeholders, you can use a stakeholder map to easily categorize them.

Gantt Charts

A Gantt chart is constructed with a horizontal axis representing the total time span of the project, broken down into increments (for example, days, weeks, or months) and a vertical axis representing the tasks that make up the project (for example, if the project is outfitting your computer with new software, the major tasks involved might be conduct research, choose software, install software). Horizontal bars of varying lengths represent the sequences, timing, and time span for each task.

Goals of Change Management

- Convergence of user and system readiness
- Optimize organization's social (cultural) and technical infrastructures
- Create a learning organization
  - Continuous improvement
  - Customer focus
  - Adaptable to changes in the environment (external and internal)

Key Principles of Change Management

- Steering Committee Must Have Product and Process Champions or Stakeholders
- Communication needs to multi-channeled (up is as valued as down)
- Conflict resolution and knowledge acquisition are collaborative(information sharing not hoarding)
- Implementation Team (change management team) has a structure and roles of individuals are defined
- Resistance to Change is always there but may be hard to detect until after Go Live (Be Proactive)

Challenges and Pitfalls

- Change process must be planned and managed
- Magnitude of change is ALWAYS greater than imagined
- Resistance to change appears in different forms -- must be addressed
- In global organizations cultural factors need special consideration
- Change in organizations occurs in stages -- need strategies for EACH stage

Scanned with CamScanner

6.2 Stages of Change

a. **Boston Consulting Group 5 Stage Model:**

- Stagnation (building desire to change)
- Anticipation (acceptance of change is highest here)
- Implementation (critical time for commitment -most pa occurs here)
- Determination (expectations get real - people begin incorporate change into daily routine)
- Fruition (acceptance begins to climb back up)

b. **Lewins' 3 Stage Model:** Unfreeze, Change and Re-freeze

Unfreezing:

- Develop awareness of change
- Users see or have a need to change (internal felt need)
- Driving forces for change outweigh the Restraining Forces

Change:

- Involve users in project team
- Train the users
- Develop change agents that can help resisters

Refreeze

- the changes are used all the time, and that they ar incorporated into everyday business.

6.3 Managerial Issues

- Getting IT ready for the future
- Organizing for planning
- Fitting the IT architecture to the organization
- IT architecture planning
- IT policy
- Ethical and legal issues
- Other ethical issues may involve sharing of computin resources

- IT strategy
- Be a leader in technology
- Be a follower
- Be an experimenter, on a small scale
- Integration
- Failures

6.4 Critical Success Factors (CSFS)

CSFs are those few things that must go right in order to ensure the organization's survival and success. Critical success factors vary by industry categories manufacturing, service, or government and by specific industries within these categories. Sample questions asked in the CSF approach are:

- What objectives are central to your organization?
- What are the critical factors that are essential to meeting these objectives?
- What decisions or actions are key to these critical factors?
- What variables underlie these decisions, and how are they measured?
- What information systems can supply these measures?

PRACTICE QUESTIONS

1. Why do you think change management is important during IS implementation? What are different change management tactics that is to be applied during execution of change management?
2. What do you mean by strategic information system? What are different benefits of using strategic information system?
3. Explain Next Generation balanced Scorecard
4. Explain change management IS planning.
5. Why change management is considered as important activity during Information System implementation? Discuss different change management practices which should be practiced during implementation.

Scanned with CamScanner

CHAPTER-

WEB BASED INFORMATION SYSTEM AND NAVIGATION

7.1 Structure of the Web

The Web is the common name for the World Wide Web, a subset of the Internet consisting of the pages that can be accessed by a Web browser. Many people assume that the Web is the same as the Internet, and use these terms interchangeably. It has no engineered architecture or structure, but rather a virtual network of content and hyperlinks and consist a large, strongly connected core in which every page can reach every other by a path of hyperlinks. The structure of Web is explained below:

Different nodes are upstream, downstream and tendril nodes

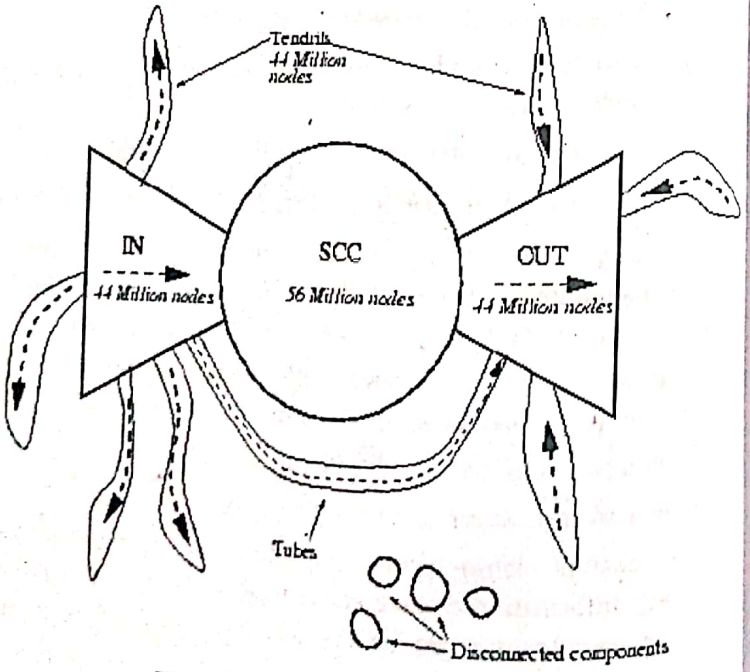


Fig 1: The Bow-Tie Structure of Web

Different nodes:

- In the SCC
- In the "inbound" part
- In the "outbound" part
- Tendrils
- Disconnected nodes

In 1999, after the Web had been growing for the better part of a decade, Andrei Broder and his colleagues set out to build a global map of the Web, using strongly connected components (SCC).

TheBow-Tie Structure-The step include to position all the remaining SCCs in relation to the giant one. This involves classifying nodes by their ability to reach and be reached from the giant SCC. The first two sets in this classification are the following.

1. **IN:**

   Nodes that can reach the giant SCC but cannot be reached from it – i.e., nodes that are "upstream" of it.

2. **OUT:**

   Nodes that can be reached from the giant SCC but cannot reach it –i.e., nodes are "downstream" of it.

3. The strongly connected components or disconnected components of an arbitrary directed graph form a partition into sub-graphs that are themselves strongly connected.

4. **Tendrils:**

   The "tendrils" of the bow-tie consist of (a) the nodes reachable from IN that cannot reach the giant SCC, and (b) the nodes that can reach OUT but cannot be reached from the giant SCC.

5. **Disconnected:**

   Finally, there are nodes that would not have a path to the giant SCC even if we completely ignored the directions of the edges. These belong to none of the preceding categories.

Scanned with CamScanner

A central core contains pages between which users can surf easily. Another large cluster, labelled 'in', contains pages that link to the core but cannot be reached from it. These are often new pages that have not yet been linked to. A separate 'out' cluster consists of pages that can be reached from the core but do not link to it, such as corporate websites containing only internal links. Other groups of pages, called 'tendrils' and 'tubes', connect to either the in or out clusters, or both, but not to the core, whereas some pages are completely unconnected. To illustrate this structure, the researchers picture the web as a plot shaped like a bow tie with finger-like projections.

7.2 Link Analysis

Link analysis is a data analysis technique used in network theory that is used to evaluate the relationships or connections between network nodes. These relationships can be between various types of objects (nodes), including people, organizations and even transactions.

Link analysis is essentially a kind of knowledge discovery that can be used to visualize data to allow for better analysis, especially in the context of links, whether Web links or relationship links between people or between different entities. Link analysis is often used in search engine optimization as well as in intelligence, in security analysis and in market and medical research.

- A means of finding authoritative, relevant sources on the Web has proven useful in the design of improved search engines
- A data analysis technique used to evaluate relationships (connections) between nodes
- leading to improved methods for accessing and understanding the available information
- Web pages can be defined as hubs and authorities.
- Characteristic patterns for to identify communities of pages on the same topic

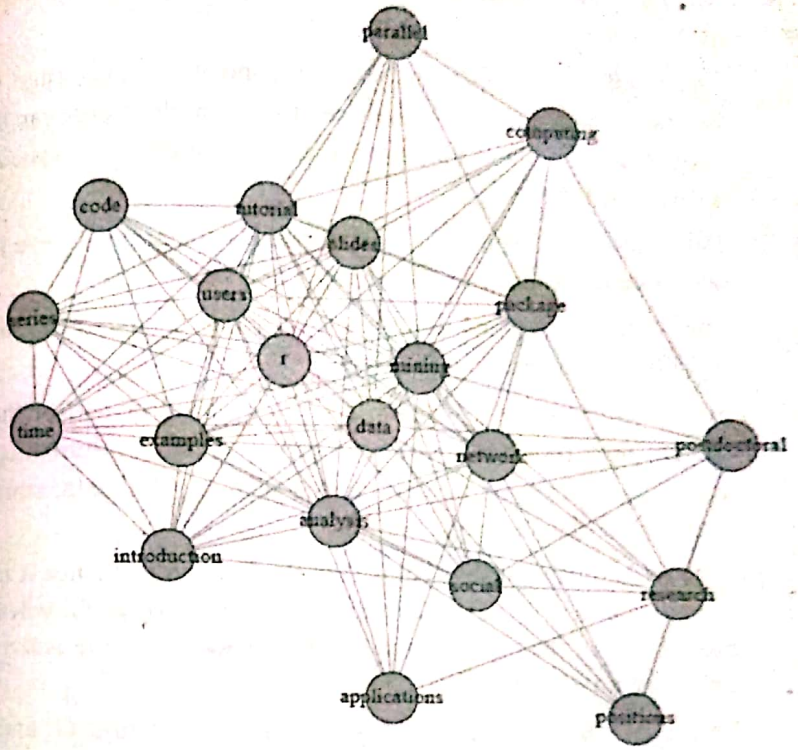


Fig 2: Link Analysis

7.3 Searching the Web

A web search engine is a software system designed to search for information on the World Wide Web. Searching in the web is done by search engine

A search engine operates in the following order:

- Web crawling
- Indexing
- Searching

Web search engines work by storing information about many web pages, which they retrieve from the HTML markup of the pages. These pages are retrieved by a Web crawler (sometimes also known as a spider) — an automated Web crawler which follows every link on

Scanned with CamScanner

the site. The site owner can exclude specific pages by using robots.txt.

- The search engine then analyzes the contents of each page to determine how it should be indexed (for example, words can be extracted from the titles, page content, headings, or special fields called meta tags).
- Data about web pages are stored in an index database for use in later queries. A query from a user can be a single word.
- The index helps find information relating to the query as quickly as possible.
- Some search engines, such as Google, store all or part of the source page (referred to as a cache) as well as information about the web pages, whereas others, such as AltaVista, store every word of every page they find.
- This cached page always holds the actual search text since it is the one that was actually indexed, so it can be very useful when the content of the current page has been updated and the search terms are no longer in it.
- This problem might be considered a mild form of linkrot, and Google's handling of it increases usability by satisfying user expectations that the search terms will be on the returned webpage.
- Increased search relevance makes these cached pages very useful as they may contain data that may no longer be available elsewhere.
- Most Web search engines are commercial ventures supported by advertising revenue and thus some of them allow advertisers to have their listings ranked higher in search results for a fee. Search engines that do not accept money for their search results make money by running search related ads alongside the regular search engine results. The search engines make money every time someone clicks on one of these ads.

7.4 Web Navigation

Web navigation refers to the process of navigating a network of information resources in the World Wide Web, which is organized as a hypertext or hypermedia.

- The user interface that is used to do so is called a web browser.
- A central theme in web design is the development of a web navigation interface that maximizes usability.
- A website's overall navigational scheme includes several navigational pieces such as global, local, supplemental, and contextual navigation; all of these are vital aspects of the broad topic of web navigation.
- Hierarchal navigation systems are vital as well since it is the primary navigation system.
- It allows for the user to navigate within the site using levels alone, which is often seen as restricting and requires additional navigation systems to better structure the website.
- The global navigation of a website, as another segment of web navigation, serves as the outline and template in order to achieve an easy maneuver for the users accessing the site, while local navigation is often used to help the users within a specific section of the site.
- All these navigational pieces fall under the categories of various types of web navigation, allowing for further development and for more efficient experiences upon visiting a webpage.

7.5 Web Mining

Web mining is the application of data mining techniques to discover patterns from the Web. With the rapid growth of World Wide Web, Web mining becomes a very hot and popular topic in Web research. E-commerce and E-services are claimed to be the killer applications for Web mining, and Web mining now also plays an important role for E-commerce website and E-services to understand how their websites and services are used and to provide better services for their customers and users.

Scanned with CamScanner

Applications of Web Mining:

- E-commerce Customer Behavior Analysis
- E-commerce Transaction Analysis
- E-commerce Website Design
- E-banking
- M-commerce
- Web Advertisement
- Search Engine
- Online Auction.

Web mining can be divided into three different types, which are

- Web usage mining,
- Web content mining
- Web structure mining.

a. Web Usage Mining

Web usage mining is the process of extracting useful information from server logs e.g. use Web usage mining is the process of finding out what users are looking for on the Internet. Web Usage Mining focuses on techniques that could predict the behavior of users while they are interacting with the WWW. Web usage mining, discover user navigation patterns from web data, tries to discovery the useful information from the secondary data derived from the interactions of the users while surfing on the Web. Web usage mining collects the data from Web log records to discover user access patterns of web pages. There are several available research projects and commercial tools that analyze those patterns for different purposes. The insight knowledge could be utilized in personalization, system improvement, site modification, business intelligence and usage characterization. Some users might be looking at only textual data, whereas some others might be interested in multimedia data. Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs of Web-based applications. Usage data captures

the identity or origin of Web users along with their browsing behavior at a Web site.

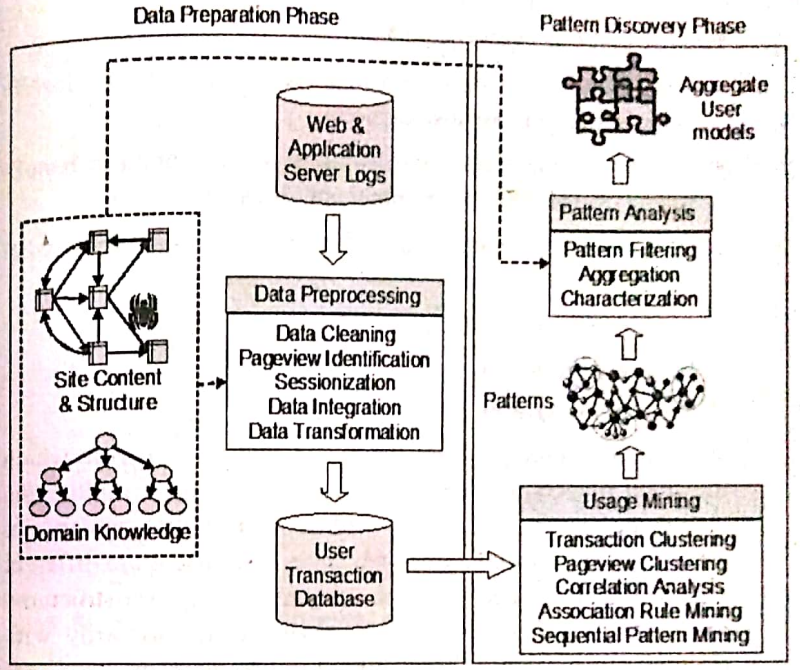


Fig 3: Web Usage Mining

Classification of Web Usage mining

- **Web Server Data:**
  The user logs are collected by the Web server. Typical data includes IP address, page reference and access time.
- **Application Server Data:**
  Commercial application servers have significant features to enable e-commerce applications to be built on top of them with little effort. A key feature is the ability to track various kinds of business events and log them in application server logs.
- **Application Level Data:**
  New kinds of events can be defined in an application, and logging can be turned on for them thus generating histories

Scanned with CamScanner

of these specially defined events. It must be noted, however, that many end applications require a combination of one or more of the techniques applied in the categories above.

Pros & Cons of Web Usage Mining

- Government agencies are using this technology to classify threats and fight against terrorism.
- The predicting capability of mining applications can benefit society by identifying criminal activities.
- Better customer relationship by giving them exactly what they need.
- attract and retain customers

Cons: Invasion of Privacy

b. Web Content Mining

Web content mining targets the knowledge discovery, in which the main objects are the traditional collections of multimedia documents such as images, video, and audio, which are embedded in or linked to the web pages. It is also quite different from Data mining because Web data are mainly semi-structured and/or unstructured, while Data mining deals primarily with structured data. Web content mining is also different from Text mining because of the semi-structure nature of the Web, while Text mining focuses on unstructured texts. Web content mining thus requires creative applications of Data mining and / or Text mining techniques and also its own unique approaches. In the past few years, there was a rapid expansion of activities in the Web content mining area. This is not surprising because of the phenomenal growth of the Web contents and significant economic benefit of such mining. However, due to the heterogeneity and the lack of structure of Web data, automated discovery of targeted or unexpected knowledge information still present many challenging research problems. Web content mining could be differentiated from two points of view: Agent-based approach or Database approach. The first approach aims on improving the information finding and filtering. The second approach aims on modeling the data on the Web into more

structured form in order to apply standard database querying mechanism and data mining applications to analyze it.

Web Content Mining Problems/Challenges :

- **Data/Information Extraction:** Extraction of structured data from Web pages, such as products and search results is a difficult task. Extracting such data allows one to provide services. Two main types of techniques, machine learning and automatic extraction are used to solve this problem.
- **Web Information Integration and Schema Matching:** Although the Web contains a huge amount of data, each web site (or even page) represents similar information differently. Identifying or matching semantically similar data is a very important problem with many practical applications.
- **Opinion extraction from online sources:** There are many online opinion sources, e.g., customer reviews of products, forums, blogs and chat rooms. Mining opinions (especially consumer opinions) is of great importance for marketing intelligence and product benchmarking.
- **Knowledge synthesis:** Concept hierarchies or ontology are useful in many applications. However, generating them manually is very time consuming. A few existing methods that explores the information redundancy of the Web will be presented. The main application is to synthesize and organize the pieces of information on the Web to give the user a coherent picture of the topic domain.
- **Segmenting Web pages and detecting noise:** In many Web applications, one only wants the main content of the Web page without advertisements, navigation links, copyright notices. Automatically segmenting Web page to extract the main content of the pages is interesting problem.

c. Web Structure Mining

Web structure mining focuses on analysis of the link structure of the web and one of its purposes is to identify more preferable documents. The different objects are linked in some way. The intuition is that a hyperlink from document A to document B

Scanned with CamScanner

implies that the author of document. A thinks document that contains worthwhile information. Web structure mining helps in discovering similarities between web sites or discovering important sites for a particular topic or discipline or in discovering web communities. Simply applying the traditional processes and assuming that the events are independent can lead to wrong conclusions. However, the appropriate handling of the links could lead to potential correlations, and then improve the predictive accuracy of the learned models.

The goal of Web structure mining is to generate structural summary about the Web site and Web page. Technically, Web content mining mainly focuses on the structure of inner-document, while Web structure mining tries to discover the link structure of the hyperlinks at the inter-document level. Based on the topology of the hyperlinks, Web structure mining will categorize the Web pages and generate the information, such as the similarity and relationship between different Web sites.

Web structure mining can also have another direction - discovering the structure of Web document itself. This type of structure mining can be used to reveal the structure (schema) of Web pages; this would be good for navigation purpose and make it possible to compare/integrate Web page schemes. This type of structure mining will facilitate introducing database techniques for accessing information in Web pages by providing a reference schema.

7.6 Collaborative Filtering

Collaborative filtering is the process of filtering for information or patterns using techniques involving collaboration among multiple agents, viewpoints, data sources, etc.

- Collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users
- Applications of collaborative filtering typically involve very large data sets.

It is applied to many different kinds of data including:

- sensing and monitoring data, such as in mineral exploration, environmental sensing over large areas or multiple sensors
- financial data, such as financial service institutions that integrate many financial sources
- in electronic commerce and web applications where the focus is on user data, etc

- Collaborative filtering algorithms often require
  - users' active participation
  - an easy way to represent users' interests to the system
  - algorithms that is able to match people with similar interests.

Methodology

- Collaborative filtering systems can be reduced to two steps:
  - Look for users who share the same rating patterns with the active user (the user whom the prediction is for).
  - Use the ratings from those like-minded users found in step 1 to calculate a prediction for the active user
  - A specific application of this is the user-based Nearest Neighbor algorithm.
- Alternatively, item-based collaborative filtering invented by Amazon.com (users who bought x also bought y), proceeds in an item-centric manner:
  - Build an item-item matrix determining relationships between pairs of items
  - Infer the tastes of the current user by examining the matrix and matching that user's data

Challenges in Collaborative Filtering

- **Data Sparsity:** based on large datasets, matrix used for collaborative filtering could be extremely large and sparse, which brings about the challenges in the performances of the recommendation.
- **Scalability:** Eg: millions of customers then complex

Scanned with CamScanner

- **Synonyms:** tendency of a number of the same or very similar items to have different names or entries. Most recommender systems are unable to discover this latent association and thus treat these products differently.
- **Grey sheep:** refers to the users whose opinions do not consistently agree or disagree with any group of people and thus do not benefit from collaborative filtering.
- **Shilling attacks:** people may give lots of positive ratings for their own items and negative ratings for their competitors.

Cold Start Problem

- Cold start is a potential problem in computer-based information systems which involve a degree of automated data modeling.
- It concerns the issue that the system cannot draw any inferences for users or items about which it has not yet gathered sufficient information.
- In the collaborative filtering approach, the recommender system would identify users who share the same preferences (e.g. rating patterns) with the active user, and propose items which the like-minded users favored (and the active user has not yet seen). Due to the cold start problem, this approach would fail to consider items which no-one in the community has rated previously.

Solution

- In scenarios involving interface agents, the cold start problem may be overcome by introducing an element of collaboration amongst agents assisting various users (Community based recommendation).
- This way, novel situations may be handled by requesting other agents to share what they have already learnt from their respective users

7.7 Recommender System

This system is a subclass of information filtering system that seek to predict the 'rating' or 'preference' that user would give to an item. It is

mostly used on movies, music, news, books, research articles, search queries, social tags, and products. A recommendation system is an extensive class of web applications that involves predicting the user responses to the options.

Often termed as Recommender Systems, they are simple algorithms which aim to provide the most relevant and accurate items to the user by filtering useful stuff from of a huge pool of information base. Recommendation engines discovers data patterns in the data set by learning consumers choices and produces the outcomes that co-relates to their needs and interests.

In Real time examples are like Amazon, they have been using a recommendation engine for suggesting the goods or products that customers might also like.

Produces recommendation by one of two ways:

- Through collaborative filtering
- Through content based filtering

a. Collaborative Filtering

Collaborative filtering systems are based on the assumption that if a user likes item A and another user likes the same item A as well as another item, item B, the first user could also be interested in the second item. Hence, they aim to predict new interactions based on historical ones. There are two types of methods to achieve this goal: *memory-based* and *model-based*.

- **Memory-based**

  There are two approaches: the first one identifies clusters of users and utilizes the interactions of one specific user to predict the interactions of other similar users. The second approach identifies clusters of items that have been rated by user A and utilizes them to predict the interaction of user A with a different but similar item B. These methods usually encounter major problems with large sparse matrices, since the number of user-item interactions can be too low for generating high quality clusters.

Scanned with CamScanner

- **Model-based**

  These methods are based on machine learning and data mining techniques. The goal is to train models to be able to make predictions. For example, we could use existing user-item interactions to train a model to predict the top-5 items that a user might like the most. One advantage of these methods is that they are able to recommend a larger number of items to a larger number of users, compared to other methods like memory-based. We say they have large *coverage*, even when working with large sparse matrices.

Issues with collaborative filtering systems

There are two main challenges that come up with these systems:

1. **Cold start**: We should have enough information (user-item interactions) for the system to work. If we setup a new e-commerce site, we cannot give recommendations until users have interacted with a significant number of items.
2. **Adding new users/items to the system**: Whether it is a new user or item, we have no prior information about them since they don't have existing interactions.

These problems can be alleviated by asking users for other type of data at the time of sign-up (gender, age, interests, etc), and using meta information from the items in order to be able to relate them to other existing items in the database.

b. **Content Based Filtering**

These systems make recommendations using a user's item and profile features. They hypothesize that if a user was interested in an item in the past, they will once again be interested in it in the future. Similar items are usually grouped based on their features. User profiles are constructed using historical interactions or by explicitly asking users about their interests. There are other systems, not considered purely content-based, which utilize user personal and social data.

One issue that arises is making obvious recommendations because of excessive specialization (user A is only interested in categories B, C, and D, and the system is not able to

recommend items outside those categories, even though they could be interesting to them).

Another common problem is that new users lack a defined profile unless they are explicitly asked for information. Nevertheless, it is relatively simple to add new items to the system. We just need to ensure that we assign them a group according to their features.

7.8 Collective Intelligence

It is a shared or group intelligence that emerges from the collaboration, collective efforts, and competition of many individuals. The concept is used in sociology, business, computer science and mass communications.

According to Pierre Lévy "*It is a form of universally distributed intelligence, constantly enhanced, coordinated in real time, and resulting in the effective mobilization of skills.*"

- Collective intelligence is a mass collaboration. In order for this concept to happen, four principles need to exist:
  - **Openness:** Sharing ideas and intellectual property: By allowing others to share ideas, it gains significant improvement.
  - **Peering:** Peering where users are free to modify and develop programs, provided that they make it available for others.
  - **Sharing:** Companies have started to share some ideas while maintaining some degree of control over others.
  - **Acting Globally:** The internet is widespread, therefore a company has no geographical boundaries and may access new markets, ideas and technology.

Collective intelligence can be harnessed from social media through a variety of means and can be beneficial to your organization. Surveys and polls are available on sites like Facebook and LinkedIn, allowing you to identify trends and patterns in people's opinions. By monitoring the Likes, shares, and comments on social networking sites, you will eventually see certain patterns arise in people's viewpoints that show the popularity of one opinion over another. To

Scanned with CamScanner

give an example, let's say that you owned a shoe company and wanted to identify which product line would sell the most. By uploading photos to a social bookmarking site , people can click a link indicating they like the product . These photos can be shared with others on other social networking sites like Facebook, increasing exposure to the product. By monitoring the reactions of people, you'll see trends where a majority of people liked one shoe over another, and thereby predict that it will sell better than others. While the results might not have a guarantee, they would tend to be more accurate than the opinion of a single or small group of decision makers.

PRACTICE QUESTIONS

1. What is collaborative filtering and recommender system in web based IS? Explain.
2. What is web? How is it organized? Discuss the working method of search engine with respect to the structure of web.
3. Discuss the concept of collaborative filtering with suitable practical examples.
4. What do you understand by web mining? Discuss web mining with respect to content, usage and structure mining.
5. Discuss the structure of web. What are different processes we might follow to find information in the web?
6. Describe link analysis for web based environment. What is page rank? What do you mean by cold start problem in collaborative filtering?
7. What do you mean by collective intelligence?
8. What is recommender system? Discuss different examples of recommender system.
9. Define collective intelligence. How does it help refinement of knowledge management system of an organization? Explain with suitable example.

CHAPTER – 8

SCALABLE AND EMERGING INFORMATION SYSTEM TECHNIQUES

8.1 Techniques for Voluminous Data

1. Big Data

Big data is an evolving term that describes a large volume of structured, semi-structured and unstructured data that has the potential to be mined for information and used in machine learning projects and other advanced analytics applications.

Big data is often characterized by the 3Vs: the extreme *volume* of data, the wide *variety* of data types and the *velocity* at which the data must be processed. Those characteristics were first identified by Gartner analyst Doug Laney in a report published in 2001. more recently, several other Vs have been added to descriptions of big data, including *veracity*, *value* and *variability*. Although big data doesn't equate to any specific volume of data, the term is often used to describe Terabytes, Petabytes and even Exabyte of data captured over time.

Understanding three Vs of Big Data

Such voluminous data can come from myriad different sources, such as: business transaction systems, customer databases, medical records, internet clickstream logs, mobile applications, social networks, the collected results of scientific experiments, machine-generated data and real-time data sensors used in internet of things environments. Data may be left in its raw form or preprocessed using data mining tools or data preparation software before it's analyzed.

The three Vs of Big data as defined by industry analyst Doug Laney are:

Scanned with CamScanner

- **Volume.** Organizations collect data from a variety of sources, including business transactions, social media and information from sensor or machine-to-machine data. In the past, storing it would've been a problem – but new technologies (such as Hadoop) have eased the burden.
- **Velocity.** Data streams in at an unprecedented speed and must be dealt with in a timely manner. RFID tags, sensors and smart metering are driving the need to deal with torrents of data in near-real time.
- **Variety.** Data comes in all types of formats – from structured, numeric data in traditional databases to unstructured text documents, email, video, audio, stock ticker data and financial transactions.

Following diagram show the 6 Vs of Big data:

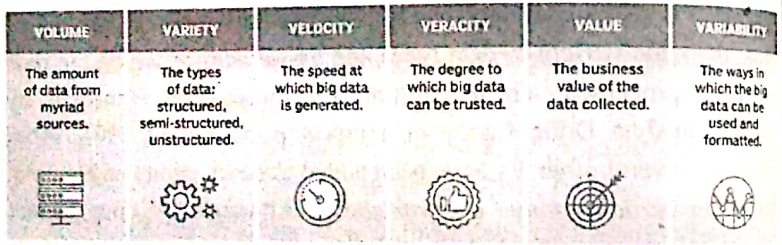


Fig 1: Six V's of Big Data

Big data also encompasses a wide variety of data types, including:

- Structured data in SQL databases and data warehouses
- Unstructured data, such as text and document files held in Hadoop clusters, or NoSQL systems,
- Semi-structured data, such as web server logs or streaming data from sensors.

How big data is stored and processed

The need to handle big data velocity imposes unique demands on the underlying compute infrastructure. The computing power required to quickly process huge volumes and varieties of data can overwhelm a single server or server cluster. Organizations must apply adequate

processing capacity to big data tasks to achieve the required velocity. This can potentially demand hundreds or thousands of servers that can distribute the processing work and operate collaboratively in a clustered architecture.

Public cloud providers offer big data capabilities through managed services that include highly distributed Apache Hadoop compute instances, the Apache Spark processing engine and related big data technologies. Amazon Elastic MapReduce (EMR) from Amazon Web Services (AWS) is one example of a big data service that runs in a public cloud; others include Microsoft's Azure and Google Cloud Dataproc. In cloud environments, big data can be stored in the Hadoop Distributed File System (HDFS) or in lower-cost cloud object storage, such as Amazon Simple Storage Service (S3); NoSQL databases are another option in the cloud for applications that are a good fit for them.

2. Cloud Computing

8.2 Scalability IS in Terms of IS

The ability of a system, network, or process to handle a growing amount of work in a capable manner or the ability to accommodate growth

There are 2 types of Scaling:

- **Vertical Scaling:** also referred to as "scale-up" approach is an attempt to increase the capacity of a single machine :
  - By adding more processing power
  - By adding more storage More memory etc.

  Vertical Scaling is most commonly used in applications and products of middle-range as well as small and middle-sized companies. One of the most common examples of Virtual Scaling is to buy an expensive hardware and use it as a Virtual Machine hypervisor (VMware ESX).

  Vertical Scaling usually means upgrade of server hardware. Some of the reasons to scale vertically includes increasing IOPS

Scanned with CamScanner

(Input / Output Operations), amplifying CPU/RAM capacity, as well as disk capacity.

However, even after using virtualization, whenever an improved performance is targeted, the risk for downtimes with it is much higher than using horizontal scaling.

- **Horizontal Scaling:** Horizontal Scaling also referred to as "scale-out" is basically the addition of more machines or setting up a cluster or a distributed environment for your software system. This usually requires a load-balancer program, which is a middle-ware component in the standard 3-tier client-server architectural model. Scaling horizontally involves adding more processing units or physical machines to your server or database. It involves growing the number of nodes in the cluster, reducing the responsibilities of each member node by spreading the key space wider and providing additional end-points for client connections. Horizontal Scaling has been historically much more used for high level of computing and for application and services.

Although this does not alter the capacity of each individual node, the load is decreased due to the distribution between separate server nodes.

Some of the reasons why organizations should choose to scale horizontally include increasing I/O concurrency, reducing the load on existing nodes, and increasing disk capacity.

The Internet and particular web services have boosted the use of Horizontal Scaling. Most giant companies that provide well-known web services like Google (Gmail, YouTube), Yahoo, Facebook, EBay, Amazon etc. are using heavily horizontal scaling.

Horizontal scaling means that you scale by adding more machines into your pool of resources whereas Vertical scaling means that you scale by adding more power (CPU, RAM) to an existing machine. Horizontal-scaling or scale dynamically is quite easy as you can add more machines into the existing pool. Vertical-scaling on the contrary is often limited to the capacity

of a single machine. Scaling beyond that capacity results in downtime and comes with an upper limit.

8.3 Techniques for Scalability in IS

- **Distributed Information System**

Distributed information systems represent an increasingly important trend to computer users. Distributed processing is a technique for implementing a single logical set of processing functions across a number of physical devices, so that each performs some part of the total processing required. Distributed processing is often accompanied by the formation of a distributed database. A distributed database exists when the data elements stored at multiple locations are interrelated, or if a process (program execution) at one location requires access to data stored at another location.

- **Grid Information System**

Grid information system is an architecture that combines computer resources from various domains to reach a main objective. In grid computing, the computers on the network can work on a task together, thus functioning as a supercomputer. Typically, a grid works on various tasks within a network, but it is also capable of working on specialized applications. It is designed to solve problems that are too big for a supercomputer while maintaining the flexibility to process numerous smaller problems. Computing grids deliver a multiuser infrastructure that accommodates the discontinuous demands of large information processing.

- **Parallel Information Processing System**

Parallel information processing system is a type of computing architecture in which several processors execute or process an application or computation simultaneously. Parallel computing helps in performing large computations by dividing the workload between more than one processor, all of which work through the computation at the same time. Most supercomputers employ parallel computing principles to operate.

Scanned with CamScanner

- **P2P Systems**

Peer-to-peer system (P2P system) is a commonly used computer networking architecture in which each workstation, or node, has the same capabilities and responsibilities. It is often compared and contrasted to the classic client/server architecture, in which some computers are dedicated to serving others. P2P may also be used to refer to a single software program designed so that each instance of the program may act as both client and server, with the same responsibilities and status.

- **Web Information Searching and Retrieval**

Information retrieval (IR) is the activity of obtaining information system resources that are relevant to an information need from a collection of those resources. Searches can be based on full-text or other content-based indexing. Information retrieval is the science of searching for information in a document, searching for documents themselves, and also searching for the metadata that describes data, and for databases of texts, images or sounds. An IR system is a software system that provides access to books, journals and other documents; stores and manages those documents. Web search engines are the most visible IR applications.

- **Content Delivery Networks (CDN)**

A content delivery network (CDN) refers to a geographically distributed group of servers, which work together to provide fast delivery of Internet content. A CDN allows for the quick transfer of assets needed for loading Internet content including HTML pages, JavaScript files, style sheets, images, and videos. The popularity of CDN services continues to grow, and today the majority of web traffic is served through CDNs, including traffic from major sites like Facebook, Netflix, and Amazon. A properly configured CDN may also help protect websites against some common malicious attacks, such as Distributed Denial of Service (DDOS) attacks.

- **Mobile and Wireless Database System**

Mobile Database is a database that is transportable, portable and physically separate or detached from the corporate database server but has the capability to communicate with those servers from remote sites allowing the sharing of various kinds of data.

With mobile databases, users have access to corporate data on their laptop, PDA, or other Internet access device that is required for applications at remote sites.

8.4 Cloud Computing Technologies and their Types

Cloud computing is the on-demand delivery of compute power, database, storage, applications, and other IT resources via the Internet with pay-as-you-go pricing.

Whether you are using it to run applications that share photos to millions of mobile users or to support business critical operations, a cloud services platform provides rapid access to flexible and low cost IT resources. With cloud computing, you don't need to make large upfront investments in hardware and spend a lot of time on the heavy lifting of managing that hardware. Instead, you can provision exactly the right type and size of computing resources you need to power your newest idea or operate your IT department. You can access as many resources as you need, almost instantly, and only pay for what you use.

How Does Cloud Computing Work?

Cloud computing gives you access to servers, storage, databases, and a broad set of application services over the Internet. A cloud services provider such as Amazon Web Services, owns and maintains the network-connected hardware required for these application services, while you provision and use what you need via a web application.

Five Characteristics of Cloud Computing

a. **On-demand Self-service**

Cloud computing resources can be provisioned without human interaction from the service provider. In other words, a manufacturing organization can provision additional computing

Scanned with CamScanner

resources as needed without going through the cloud service provider. This can be a storage space, virtual machine instances, database instances, and so on.

b. **Broad Network Access**

Cloud computing resources are available over the network and can be accessed by diverse customer platforms. It other words, cloud services are available over a network—ideally high broadband communication link—such as the internet, or in the case of a private clouds it could be a local area network (LAN).

c. **Multi-tenancy and Resource Pooling**

Cloud computing resources are designed to support a multi-tenant model. Multi-tenancy allows multiple customers to share the same applications or the same physical infrastructure while retaining privacy and security over their information. It's similar to people living in an apartment building, sharing the same building infrastructure but they still have their own apartments and privacy within that infrastructure. That is how cloud multi-tenancy works. Resource pooling means that multiple customers are serviced from the same physical resources. Providers' resource pool should be very large and flexible enough to service multiple client requirements and to provide for economy of scale. When it comes to resource pooling, resource allocation must not impact performances of critical manufacturing applications.

d. **Rapid Elasticity and Scalability**

One of the great things about cloud computing is the ability to quickly provision resources in the cloud as manufacturing organizations need them. And then to remove them when they don't need them. Cloud computing resources can scale up or down rapidly and, in some cases, automatically, in response to business demands. It is a key feature of cloud computing. The usage, capacity, and therefore cost, can be scaled up or down with no additional contract or penalties.

e. **Measured service**

Cloud computing resources usage is metered and manufacturing organizations pay accordingly for what they have used.

Resource utilization can be optimized by leveraging charge-per-use capabilities. This means that cloud resource usage—whether virtual server instances that are running or storage in the cloud—gets monitored, measured and reported by the cloud service provider. The cost model is based on "pay for what you use"—the payment is variable based on the actual consumption by the manufacturing organization.

Benefits of Cloud

- **Cost**

Cloud computing eliminates the capital expense of buying hardware and software and setting up and running on-site datacenters—the racks of servers, the round-the-clock electricity for power and cooling, and the IT experts for managing the infrastructure. It adds up fast.

- **Speed**

Most cloud computing services are provided self service and on demand, so even vast amounts of computing resources can be provisioned in minutes, typically with just a few mouse clicks, giving businesses a lot of flexibility and taking the pressure off capacity planning.

- **Global scale**

The benefits of cloud computing services include the ability to scale elastically. In cloud speak, that means delivering the right amount of IT resources—for example, more or less computing power, storage, bandwidth—right when they're needed, and from the right geographic location.

- **Productivity**

On-site datacenters typically require a lot of "racking and stacking"—hardware setup, software patching, and other time-consuming IT management chores. Cloud computing removes the need for many of these tasks, so IT teams can spend time on achieving more important business goals.

- **Performance**

The biggest cloud computing services run on a worldwide network of secure datacenters, which are regularly upgraded to

Scanned with CamScanner

the latest generation of fast and efficient computing hardware. This offers several benefits over a single corporate datacenter, including reduced network latency for applications and greater economies of scale.

- **Reliability**

  Cloud computing makes data backup, disaster recovery, and business continuity easier and less expensive because data can be mirrored at multiple redundant sites on the cloud provider's network.

- **Security**

  Many cloud providers offer a broad set of policies, technologies, and controls that strengthen your security posture overall, helping protect your data, apps, and infrastructure from potential threats.

Cloud Computing Deployment Models

a. Public Clouds

The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.

A public cloud is the obvious choice when:

- Your standardized workload for applications is used by lots of people, such as e-mail.
- You need to test and develop application code.
- You need incremental capacity (the ability to add compute resources for peak times).
- You're doing collaboration projects

b. Private Clouds

The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.

When is a Private Cloud for you?

- You need data sovereignty but want cloud efficiencies
- You want consistency across services
- You have more server capacity than your organization can use
- Your data center must become more efficient
- You want to provide private cloud services

c. Hybrid Clouds

The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

Here are a couple of situations where a hybrid environment is best:

- Your company wants to use a SaaS application but is concerned about security.
- Your company offers services that are tailored for different vertical markets. You can use a public cloud to interact with the clients but keep their data secured within a private cloud.
- You can provide public cloud to your customers while using a private cloud for internal IT.

d. Community Clouds

The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises.

Here are some situations where a community cloud is best:

- Government organizations within a state that need to share resources

Scanned with CamScanner

- A private HIPAA compliant cloud for a group of hospitals or clinics
- Telco community cloud for Telco DR to meet specific FCC regulations

Cloud Computing Service Model

a. 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

b. 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:

- Users use resources but have no control of underlying cloud infrastructure
- Users pay for what they use
- Flexible scalable infrastructure without extensive pre-planning

c. Platform as a Service (PAAS)

Platform as a Service (PAAS) is a cloud computing model that delivers applications over the internet. In a PAAS model, a cloud provider delivers hardware and software tolls, usually those needed for application development, to its users as a

service. A PAAS provider hosts the hardware and software on its own infrastructure.

Some of the main characteristics of PAAS are :

- Scalability and auto-provisioning of the underlying infrastructure.
- Security and redundancy.
- Build and deployment tools for rapid application management and deployment.
- Integration with other infrastructure components such as web services, databases, and LDAP.
- Multi-tenancy, platform service that can be used by many concurrent users.
- Logging, reporting, and code instrumentation.
- Management interfaces and/or API.

8.5 MapReduce

A MapReduce is a data processing tool which is used to process the data parallelly in a distributed form. It was developed in 2004, on the basis of paper titled as "MapReduce: Simplified Data Processing on Large Clusters," published by Google.

The MapReduce is a paradigm, which has two phases, the mapper phase, and the reducer phase. In the Mapper, the input is given in the form of a key-value pair. The output of the Mapper is fed to the reducer as input. The reducer runs only after the Mapper is over. The reducer too takes input in key-value format, and the output of reducer is the final output.

Steps in Map Reduce

- The map takes data in the form of pairs and returns a list of <key, value> pairs. The keys will not be unique in this case.
- Using the output of Map, sort and shuffle are applied by the Hadoop architecture. This sort and shuffle acts on these list of <key, value> pairs and sends out unique keys and a list of values associated with this unique key <key, list(values)>.

Scanned with CamScanner

- An output of sort and shuffle sent to the reducer phase. The reducer performs a defined function on a list of values for unique keys, and Final output <key, value> will be stored/displayed.

MapReduce and Hadoop Systems

Apache Hadoop MapReduce is a framework for processing large data sets in parallel across a Hadoop cluster. Data analysis uses a two step map and reduce process. The job configuration supplies map and reduce analysis functions and the Hadoop framework provides the scheduling, distribution, and parallelization services.

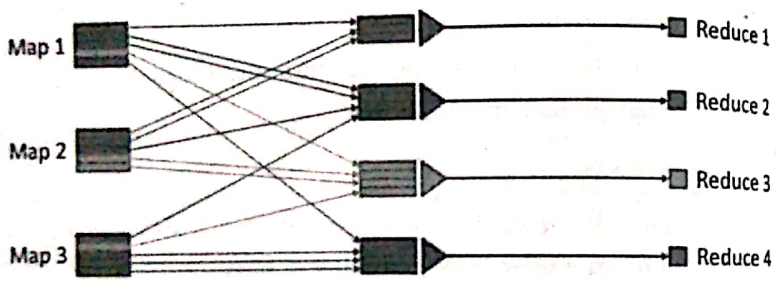


Fig 2: Map Reduce

The top level unit of work in MapReduce is a job. A job usually has a map and a reduce phase, though the reduce phase can be omitted. For example, consider a MapReduce job that counts the number of times each word is used across a set of documents. The map phase counts the words in each document, then the reduce phase aggregates the per-document data into word counts spanning the entire collection.

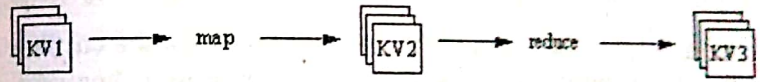
During the map phase, the input data is divided into input splits for analysis by map tasks running in parallel across the Hadoop cluster. By default, the MapReduce framework gets input data from the Hadoop Distributed File System (HDFS). Using the MarkLogic Connector for Hadoop enables the framework to get input data from a MarkLogic Server instance.

The reduce phase uses results from map tasks as input to a set of parallel reduce tasks. The reduce tasks consolidate the data into final

results. By default, the MapReduce framework stores results in HDFS. Using the MarkLogic Connector for Hadoop enables the framework to store results in a MarkLogic Server instance.

Although the reduce phase depends on output from the map phase, map and reduce processing is not necessarily sequential. That is, reduce tasks can begin as soon as any map task completes. It is not necessary for all map tasks to complete before any reduce task can begin.

MapReduce operates on key-value pairs. Conceptually, a MapReduce job takes a set of input key-value pairs and produces a set of output key-value pairs by passing the data through map and reduce functions. The map tasks produce an intermediate set of key-value pairs that the reduce tasks uses as input. The diagram below illustrates the progression from input key-value pairs to output key-value pairs at a high level:



Though each set of key-value pairs is homogeneous, the key-value pairs in each step need not have the same type. For example, the key-value pairs in the input set (KV1) can be (string, string) pairs, with the map phase producing (string, integer) pairs as intermediate results (KV2), and the reduce phase producing (integer, string) pairs for the final results (KV3).

8.6 Hadoop System

Apache Hadoop is a collection of open-source software utilities that facilitate using a network of many computers to solve problems involving massive amounts of data and computation. It provides a software framework for distributed storage and processing of big data using the MapReduce programming model. Originally designed for computer clusters built from commodity hardware still the common use—it has also found use on clusters of higher-end hardware. All the modules in Hadoop are designed with a fundamental assumption that

Scanned with CamScanner

hardware failures are common occurrences and should be automatically handled by the framework.

The core of Apache Hadoop consists of a storage part, known as Hadoop Distributed File System (HDFS), and a processing part which is a MapReduce programming model. Hadoop splits files into large blocks and distributes them across nodes in a cluster. It then transfers packaged code into nodes to process the data in parallel. This approach takes advantage of data locality, where nodes manipulate the data they have access to. This allows the dataset to be processed faster and more efficiently than it would be in a more conventional supercomputer architecture that relies on a parallel file system where computation and data are distributed via high-speed networking.

The base Apache Hadoop framework is composed of the following modules:

- *Hadoop Common* – contains libraries and utilities needed by other Hadoop modules;
- *Hadoop Distributed File System (HDFS)* – a distributed file-system that stores data on commodity machines, providing very high aggregate bandwidth across the cluster;
- *Hadoop YARN* – introduced in 2012 is a platform responsible for managing computing resources in clusters and using them for scheduling users' applications
- *Hadoop MapReduce* – an implementation of the MapReduce programming model for large-scale data processing.

The term Hadoop is often used for both base modules and submodules and also the ecosystem, or collection of additional software packages that can be installed on top of or alongside Hadoop, such as Apache Pig, Apache Hive, Apache HBase, Apache Phoenix, Apache Spark, Apache ZooKeeper, Cloudera Impala, Apache Flume, Apache Sqoop, Apache Oozie, and Apache Storm.

Apache Hadoop's MapReduce and HDFS components were inspired by Google papers on MapReduce and Google File System.

The Hadoop framework itself is mostly written in the Java programming language, with some native code in C and command

line utilities written as shell scripts. Though MapReduce Java code is common, any programming language can be used with Hadoop Streaming to implement the map and reduce parts of the user's program. Other projects in the Hadoop ecosystem expose richer user interfaces.

Some Prominent Users of Hadoop Systems:

- Yahoo!
- Facebook
- Amazon
- Google
- Twitter

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

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.

Scanned with CamScanner

8.7 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
- 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.

There are the 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
- Data management is carried out by Cloud Data Management Interface (CDMI)
- CDMI is standard that specifies protocols for self-provisioning, administering and accessing cloud storage

8.8 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.
- 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 ?

Stopwords / Stoplist

- Function words do not bear useful information for IRof, in, about, with, I, although, …
- **Stoplist:** contain Stopwords, not to be used as index
  - Prepositions

Scanned with CamScanner

- 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
  computer
  compute
  computes
  computing
  computed
  computation

8.9 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.

8.10 Page Rank

PageRank (PR) is an algorithm used by Google Search to rank websites in their search engine results. PageRank was named after

Larry Page, one of the founders of Google. PageRank is a way of measuring the importance of website pages.

According to Google:

"PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites."

How it started

- Sir Tim Berners-Lee stored new websites in notebook
- Websites containing links to new websites
- As they grew search were added
- Hence Search Engines were created

Algorithm

The PageRank algorithm outputs a probability distribution used to represent the likelihood that a person randomly clicking on links will arrive at any particular page. PageRank can be calculated for collections of documents of any size. It is assumed in several research papers that the distribution is evenly divided among all documents in the collection at the beginning of the computational process. The PageRank computations require several passes, called "iterations", through the collection to adjust approximate PageRank values to more closely reflect the theoretical true value.

Problem:

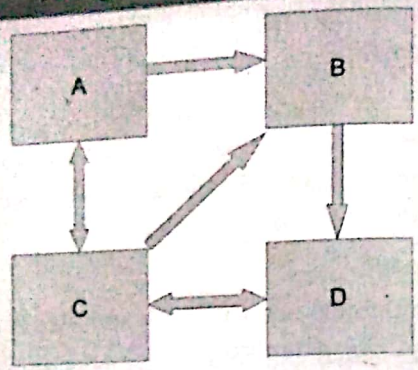
How to sort the search result?

Page Rank : Formula

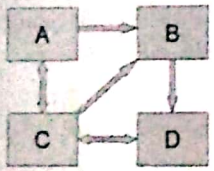
- $PR(A) = (1-d)+d(PR(T1)/C(T1)+...+PR(Ti)/C(Ti))$
  - PR(A) is the PageRank of page A
  - PR(Ti) is the PageRank of pages Ti which link to page A
  - C(Ti) is the number of outbound links on page Ti
  - d is a damping factor which can be set between 0 and 1.

Scanned with CamScanner

Page Rank : Example



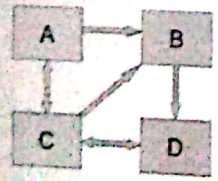
Page Rank : Example



	Iteration 0	Iteration 1	Iteration 2	Page Rank
A	1/4			
B	1/4			
C	1/4			
D	1/4			

Initial Page Rank of all Pages =
1/(Total number of page)

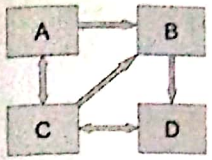
Page Rank : Example



	Iteration 0	Iteration 1	Iteration 2	Page Rank
A	1/4	1/12		
B	1/4			
C	1/4			
D	1/4			

Let d = 1
PR(A) = (1-d) +d(PR(C)/C(C))
PR(A) = (1-1) + 1((1/4)/3)

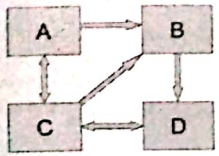
Page Rank : Example



	Iteration 0	Iteration 1	Iteration 2	Page Rank
A	1/4	1/12		
B	1/4	2.5/12		
C	1/4			
D	1/4			

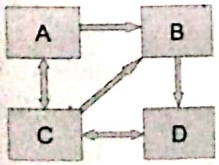
PR(B) = (1-d) +d (PR(A)/C(A) + PR(C)/C(C))
PR(B) = 0 + (1/4)/2 + (1/4)/3
PR(B) = 2.5/12

Page Rank : Example



	Iteration 0	Iteration 1	Iteration 2	Page Rank
A	1/4	1/12		
B	1/4	2.5/12		
C	1/4	4.5/12		
D	1/4	4/12		

Page Rank : Example



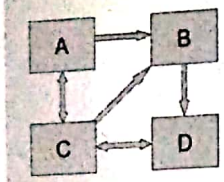
	Iteration 0	Iteration 1	Iteration 2	Page Rank
A	1/4	1/12	1.5/12 + 0.125	4
B	1/4	2.5/12	3/12 + 0.1667	3
C	1/4	4.5/12	4.5/12 + 0.375	1
D	1/4	4/12	4/12 + 0.3333	2

According to Google about 100 iteration is required for PageRank Algorithm.

Scanned with CamScanner

Page Rank : Optimization

- Initial Page Rank as 1
- Using power method(Matrix)



$$\underset{H}{\begin{vmatrix} 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/3 & 0 \\ 1/2 & 0 & 0 & 1 \\ 0 & 1 & 1/3 & 0 \end{vmatrix}} \times \underset{V}{\begin{vmatrix} 1 \\ 1 \\ 1 \\ 1 \end{vmatrix}} = \underset{V_1}{\begin{vmatrix} 1/3 \\ 5/6 \\ 1/2 \\ 1/3 \end{vmatrix}}$$

$v_1 = Hv$

$v_2 = Hv_1 = H(Hv) = H^2v$

$v_n = H^nv$

PRACTICE QUESTIONS

1. What is Cloud Computing? Explain in detail.
2. What are the different issues that should be taken into consideration while choosing cloud as IS platform?
3. What are different techniques and methods developed to handle voluminous data? Discuss two such techniques.
4. What is key challenge in handling the voluminous data in cloud architecture?
5. What are characteristics of voluminous data? Describe how Hadoop and Map Reduce work on big data over distributed architecture. 'Cloud is basically a client server model in essence, do you agree with this statement?
6. Discuss cloud computing technologies and their types.
7. Differentiate between Map Reduce and Hadoop.
8. Illustrate what type of cloud technologies would you recommend for a small business firm with high level of accuracy and security required for business transaction?

CASE STUDY FOR INFORMATION SYSTEM

- Each group must have 4-5 members.
- For every case study topic each group should prepare a case study report and a presentation slide and should be presented by each group members.

CASE STUDY 1

PREPARE A BALANCED SCORE CARD FOR ANY ORGANIZATION.

CASE STUDY 2

DISCUSS ABOUT ANY ONE SSL SERVICE PROVIDER:

1. Comodo SSL
2. DigiCert
3. Entrust Datacard
4. GeoTrust
5. GlobalSign
6. GoDaddy
7. Network Solutions
8. RapidSSL
9. SSL.com
10. Thawte

Discuss about the real world scenario for using above SSL certificate for below area (you can also choose your own area):

i. School Gives Parents Secure Web Based Access to Student Data with SSL Certificates

ii. Travel Company Uses SSL Certificates to Build Trust with Customers and Grow Business Online

iii. Healthcare Portal Secures Medical Data with SSL Certificates to Build Trust with Patients and Providers

iv. New E-commerce Site Quickly Builds Customer Confidence with SSL Certificates

Scanned with CamScanner

Each topic should include:

i. Industry Name

ii. Key Challenges

iii. Solution using any one SSL service provider.

iv. Result

CASE STUDY 3

A CASE STUDY ON INTELLECTUAL PROPERTY RIGHT

What is Intellectual Property Right?

Discuss about a real case of Intellectual Property Right. It should contain the following topics:

Title of Case

Country

History of Case

Claim by the other party

How was the case solved?

CASE STUDY 4

CLOUD COMPUTING

What is cloud computing?

List of various cloud service provider.

- PhoenixNAP
- Amazon Web Services
- Microsoft Azure
- Google Cloud Platform
- Adobe
- VMware
- IBM Cloud
- Rackspace
- Red Hat
- Salesforce

- Oracle Cloud
- SAP
- Verizon Cloud
- Navisite
- Dropbox
- Egnyte

Discuss about any one service provider in detail. Is there any Cloud Service provider in Nepal? If any list it and discuss.

CASE STUDY 5 & 6

CHANGE MANAGEMENT

Change Management is one of the risky tasks. It can be beneficial of can degrade the organizational success.

Study about few case studies in which organization have implemented changes and also discuss whether the particular change was beneficial or not for organization.

Some of the companies with successful change: [include 5 students]

Nokia

Coca-Cola

Toyota

GE

Amazon

Some of the companies with successful change: [include 5 students]

McDonald's

Kodak

Pan-Am

Scanned with CamScanner

CASE STUDY 7

CLOUD COMPUTING

Cloud Computing although being one of the secure and reliable technology to provide various services to user as infrastructure, platform and software, have failed in many way.

Discuss about some cases or some cloud service provider where there were security breaches or who failed to provide secure service to their users. Also highlight the particular case.

CASE STUDY 8

WEB MINING

There are three types of web mining:

- Web usage mining
- Web content mining
- Web structure mining

Discuss about each and also discuss how these mining used in real world application with proper example.

CASE STUDY 9

"THE CYBER SECURITY RISKS ON SOCIAL MEDIA"

We are addicted to use various social sites like Twitter, Facebook, LinkedIn and many more. Discuss about some security issues related to these sites. Each student should discuss about one social site. You can include as many students.

CASE STUDY 10

CLOUD COMPUTING

How is the following task done in cloud?

A. Data management in cloud

B. Information retrieval in cloud

Explain each topic in detail with proper techniques and example.

SOLVED QUESTIONS

1. Why do organizations need Information System? Differentiate between Information System and typical database management system.

Ans. Information System is the study of complementary networks of hardware and software that people and organizations use to collect, filter, and process, create and distribute data. Information System can help all kinds of organizations to improve the efficiency and effectiveness of their business processes, managerial decision making, and workgroup collaboration, thus strengthening their competitive positions in a rapidly changing marketplace. Information System helps to collect all the related data for the organization and store them in an easy and efficient manner. It also makes all the information available to the organization all the time which helps to take decision easily and properly. Information System also helps to connect the organization to their clients and help the organization improve their products as per the client's requirement. So, Information System can help all kinds of organization to improve efficiency and effectiveness of their business processes, managerial decision making and workgroup collaboration. Hence, organiza-tions need Information System.

Information System	Database Management System
1. Information System is the study of complementary networks of hardware and software that people and organizations use to collect, filter, and process, create and distribute data.	1. A Database Management System is a suite of computer software providing the interface between users and a database or databases.
2. Information System can store both structured and unstructured data.	2. Database Management Systems can only store structured data.

Scanned with CamScanner

Information System	Database Management System
3. Information System may contain database management system.	3. Database Management System can be subset of Information System.
4. Information system is used to take decision based on meaningful information.	4. Database system is used to store, arrange & collect data in useful format.

2. What do you understand by Information System security? Describe both technology and human based safeguards for Information Systems.

Ans. Information System consists of huge collection of critical and important data. Information System security means protection of such critical data from accidental or intentional modification including related hard-ware and software, destruction or disclosure to unauthorized persons.

To assure that the data of an Information System is secured, certain safeguard must be implemented. These safeguard may be technological or human based.

Technological safeguards are the safeguards applied with the help of software and hardware devices. Some technology based safeguards are explained as follows:

- **Firewalls**: It is a software or a hardware device which is used to allow and deny access to the Information System through network. Some firewalls are also designed to detect intrusion and take proper action against it.
- **Encryption**: Use of encryption methods to transfer data ensures that only authorized person who has the decryption key can view the data without being modified or spoofed by any third party.
- **Virus Monitoring and Protection**: Viruses and malwares are major risk for Information Systems. They may create, modify or even destroy data stored in an Information System. They may also slow down the systems. Use of anti-

virus and anti-malware software may prevent and detect such viruses and malwares.

- **Audit-Control Software**: Audit-Control Software are used to keep track of computer activity. It can detect any suspicious activity and can help to take actions accordingly.

Human based safeguards are those safeguards which are applied with the help of human. Some human based safeguards are explained as follows:

a. Use of federal and state laws as well as ethics
b. Use of formal investigative techniques to evaluate digital information
c. Evaluation of storage devices for traces of illegal activity
d. Restoration of deleted files

3. Explain the process of developing an I.S. security plan.

Ans. Information System consists of huge collection of critical and important data. Information System security means protection of such critical data from accidental or intentional modification including related hard-ware and software, destruction or disclosure to unauthorized persons. For the proper security of Information System, a security plan must be developed. An information security strategic plan can position an organization to mitigate, transfer, accept or avoid information risk related to people, processes and technologies. An established strategy also helps the organization adequately protect the confidentiality, integrity and availability of information. To develop such Information Security plan the following process needs to be done:

- Develop the information security policies
- Communicate the information security policies
- Identify critical information assets and risks
- Test and reevaluate risks
- Obtain stakeholder support

Scanned with CamScanner

4. How can you justify the huge expenditures on implementation of Enterprise Management System? Explain with practical example. Explain the push-pull view of Supply Chain Management System (SCM). [5+3]

Ans. Enterprise Management System is business process management software that allows an organization to use a system of integrated applications to manage the business and coordinate all the resources, information and activities needed to complete the business processes. But implementation of such Enterprise Management System can be expensive which can be a major disadvantage. But Enterprise Management System has many advantages which makes implementing such system important.

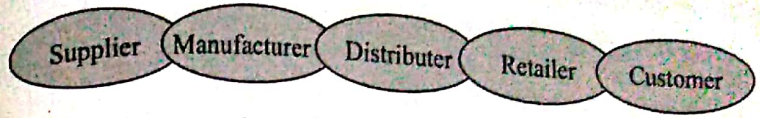
Some of such ad-vantages are as follows:

- Enterprise Management System gives the ability to share information between all components or departments of the organization.
- Enterprise Management System also helps to optimize the business process of the organization.
- It also gives the access to accurate and reliable information on time.
- It helps to synchronize all the departments with each other.
- It improves the chance of success of an organization and its products.

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. Supply chain processes fall into one of two categories depending on the timing of their execution relative to customer demand.

Push View: In push view execution is initiated in anticipation of customer orders i.e. the product for the customer is produced by the manufacturer before the customer request hoping that the customer uses the product.

Push to Customer



Supplier → Manufacturer → Distributer → Retailer → Customer

Fig. Push Process

Pull View: In the pull view execution is initiated in response to a customer order i.e. the product or service for the customer is only provided after the customer request for that product.

Pull from Customer

Supplier ←Manufacturer ←Distributer ←Retailer ←Customer

Fig.: Pull Process

5. Explain about Enterprise and Executive decision support system (DSS) with examples.

Ans. Enterprise Decision Support System is a computer-based technology used by mostly multi-disciplinary organization which provides technical and project/program leadership for Departmental Information Technology, applications and systems. It uses decision support systems to quickly perform complex analyses over large amounts of data. The results of Executive Decision Support System are used to inform critical business decisions.

Executive Decision Support System is a computer-based technology designed specifically for the information needs of top executive level and provides for rapid access to timely information, direct access to management reports, very user friendly and supported by graphics. Executive Decision Support System also supports the informational roles of the executives.

Scanned with CamScanner

6. Differentiate between knowledge management (KM) and knowledge based expert systems.

Ans. Some of the differences between Knowledge Management System and Knowledge Based expert Systems are as follows:

Knowledge Management System	Knowledge Based expert Systems
i. Knowledge Management System refers to a (generally IT based) system for managing knowledge in organizations for supporting creation, capture, storage and distribution of information.	i. A Knowledge Based System is a computer program that uses artificial intelligence to solve problems within a specialized domain that ordinarily requires human expertise.
ii. It is used in an organization to identify, create, represent, distribute, and enable adoption of insights and experiences	ii. It is used for classification, diagnosis, monitoring, design, scheduling, and planning for specialized tasks.
iii. Knowledge Management System mostly designed to be generic.	iii. Knowledge Based expert System is mostly designed to be specific.
iv. It is used to capture the knowledge from different sources.	iv. It is used to implement such captured knowledge into useful applications.

6. Explain the role of planning in the development of business / IT strategies architectures and application. What do you mean by Operational Information Systems? [5+3]

Ans. Business planning is the process of identifying the firm's goals, objectives, and priorities. It also helps developing action plans for accomplishing them.

Information systems planning on the other hand is the part of business planning concerned with developing the firm's information systems resources. Planning helps in many things and helps in success of the product and organization. Some of

the roles of planning in the development of business and IT strategies architectures and applications are as follows:

- To detect and reduce risks in the project.
- To identify the required resources beforehand and to obtain it on time.
- To map Information System with business.
- To establish goal, schedule, and milestone in order to track progress.
- To provide an opportunity for communication with top management and user management

Operational IS Planning involves resource allocation, that is, allocating both IS application development resources and operational resources. It concerns the performance targets and specific tasks, schedules, and budgets to achieve short-range objectives, usually of one year or not more than two years in duration. The final product of the resource allocation phase should be an IS budget and operating plan. Generic format of Operational IS Plan is:

- Service objectives and overview
- Application development and maintenance plan
- Operations Plan
- Technical Support Plan
- Standard Practices Program
- Staffing and Organization Plan, etc.

7. Explain the critical success factor (CSFs) while implementing information system applications and explain the sources of CSFs. [6]

Ans. CSFs are those few things that must go right in order to ensure the organization's survival and success. Critical success factors vary by industry categories—manufacturing, service, or government—and by specific industries within these categories. Some of such critical factors are as follows:

Scanned with CamScanner

Top Management Commitment

It is vital for the successful IS implementation. In the absence of such commitment even minor issues becomes unmanageable. So, it is important that top management commitment for such IS implementation is not only present but is also communicated well within the organization, so that there is no ambiguity about the commitment.

IS Team Composition

The implementation of IS is done by a team. The team consists of technical and management specialists from the vendor organization as well as people from IS department and other stakeholder departments of the client organization. The constitution of the team is a very important and a critical factor in the successful implementation of IS. The team must consist of representatives from the user com-munity, people who were part of the planning team, designing team in the client organization, other stakeholders in the client organization and people from the human resources department.

IT Teamwork

The team that is created must work in a cohesive manner with a sense of purpose and respect in all.

IS Planning Quality

An IS project is as successful as its plan. If the planning was faulty, the IS implementation will no doubt be faulty.

Diligent Project Management

The manner in which the implementation project is handled is also another critical success factor.

Modern project management techniques must be used to estimate accurate cost and time schedules which must then be strictly monitored and adhered to.

8. What are the pros and cons of collective intelligence? Give some examples.

Ans. Collective intelligence (CI) is shared or group intelligence that emerges from the collaboration, collective efforts, and competition of many individuals. Collective Intelligence has

both pros and cons. Some pros and cons of collective intelligence are listed below:

Pros:

i. Using CI, problem solving becomes much easier and faster. For example – Two people trying to solve a problem is faster

ii. Different strategies can be found out by each individuals and the best one can be used. For example – Two people can solve a same problem in two different ways

iii. Each individuals can view and obtain different problems that may happen in the future. For exam-ple – A group of individuals may see different problems that may occur in the future

Cons:

i. There may be crisis between different individuals. For example – There might be confusion who is the leader and who to follow

ii. If more than one strategy is found to solve a problem, finding which one is better can be another problem.

9. Describe the characteristics for the design of the typical recommender system. Explain the different types of recommender systems?

Ans. A recommender system is a system which provides recommendations to a user. They are the subclass of information filtering system that seek to predict the 'rating' of 'preference' that user would give to an item (such as music, book, movies, etc.) or social element (like people or groups) they had not yet considered, using a model built from the characteristics of an item (content-based approaches) or the user's social environment (collaborative filtering approaches). Some of the characteristics for the design of the typical recommender system are as follows:

i. It should predict the rating of item for the user.

ii. It should build a model to recommend items using user's behavior and item's characteristics.

Scanned with CamScanner

iii. It should help user to discover items they might not have found otherwise.

iv. It should enhance user experience.

There are mainly three types of recommender systems. They are explained below:

i. **Collaborative Recommender System:**

It builds a model from a user's past behavior and use that model to predict items (or ratings for items) that the user may have an interest in. The key is to find users or user group whose behavior match with the current user. More the users and rating, the result becomes better.

ii. **Content based Recommender System:**

It utilize a series of discrete characteristics of an item in order to recommend additional items with similar properties. In this type of recommender system, user profiling is the key. Matching "user preference" with the "item characteristics" works for textual information.

10. Explain the working mechanisms of Map Reduce for big data with example.

Ans. Map Reduce is a programming paradigm used in Big Data technologies. It allows massive scalability across large number of servers. It was originally developed by Google for indexing Web pages and replacing their original indexing algorithms and heuristics in 2004. It is most widely used in Hadoop and Google File System. Map Reduce actually refers to two separate and distinct tasks that Hadoop programs perform which are Map and Reduce. Map job takes a set of data and converts it into another set data. While Reduce job takes the output from map as input and combines those tuples into smaller set of tuples. Map Reduce is a software framework that allows developers to write programs that process massive amounts of un-structured data in parallel across a distributed cluster of processors or stand-alone computers.

Let us take an example word count in which user has to count number of each word in a document. Let the document contain

text "Map Reduce is used in Big Data. Big Data is Map Reduce."

Then in map reduce first the Map Job is conducted. In this example, Map Job sets a counter for each word as 1. So the count becomes:

Map	1
Reduce	1
Is	1
Used	1
In	1
Big	1
Data	1
Big	1
Data	1
Is	1
Map'	1
Reduce	1

This becomes the output of the Map Job. Then Map Job is completed and Reduce Job starts. This output is used as input for the Reduce Job. In this example, Reduce Job takes common word and creates a kind of dictionary with Word as key and its count as value. The count is then added to get the total value of each word. So the final output becomes as follows:

Map	2
Reduce	2
Is	2
In	1
Used	1
Big	2
Data	2

11. Mention the types of cloud computing. Explain about Data Management in the cloud.

Ans. There are mainly three types of cloud computing as listed below:

Scanned with CamScanner

i. **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.

ii. **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.

iii. **PaaS (Platform as a Service):** This provides a platform allowing customers to develop, run, and man-age applications without the complexity of building and maintaining the infrastructure typically associated with developing and launching an app. For example AWS Elastic Beanstalk, Windows Azure, Google App Engine, Apache Stratos etc.

Data management applications are potential candidates for deployment in the cloud. In Industry, enterprise database system have significant up-front cost that includes both hardware and software costs. In Academia, managing, processing and sharing mass-produced data in the cloud can be done. Many "Cloud Killer Apps" are in fact data-intensive. Batch processing as in Map Reduce, OLTP in automated business applications, OLAP in data mining or machine learning are few process that needs to be executed 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. There are three characteristics of a cloud computing environment.

i. Compute power is elastic, but only if workload is parallelizable.

ii. Data is stored at untrusted host.

iii. Data is replicated, often across large geographic distances

Data management is carried out by Cloud Data Management Interface (CDMI). CDMI is standard that specifies protocols for self-provisioning, administering and accessing cloud storage.

12. What are critical success factors? Explain with example. What are the factors one needs to take for successful change management in an organization?

Ans: Critical Success Factors (CSFs) is a management term for an element that is necessary for an organization or project to achieve its mission. It is a critical factor or activity required for ensuring the success of a company or an organization. The term was initially used in the world of data analysis and business analysis. Critical success factors are those few things that must go well to ensure success for a manager or an organization and, therefore, they represent those managerial or enterprise areas that must be given special and continual attention to bring about high performance. CSFs include issues vital to an organization's current operating activities and to its future success. For example, CSFs for a successful Smart Phone Company can be attributes, product features or service details for decent design (else unattractive to buy), had to recognize brand (if no one heard of company, on sales), had to have new innovative features, etc.

Change management in business can be a complicated issue. Major adjustments to the way in which a business operates or provides its products and services must be carefully managed. There are several different ways to manage change in the workplace – but here are just a few key tips that the experts would recommend if you're looking to lead your business through a change process.

Scanned with CamScanner

Plan: Before undertaking any change, a business needs to have a clear and valid reason and mission for the change process. Make sure any undertaking, no matter its size or impact, is realistic, measurable and achievable – there is no use in creating a difficult situation if there is no significant gain to be made for the organization or its stakeholders.

Communicate: The most important factor in changing anything within a business, small or large, is to communicate the change, and its many factors, to the key stakeholders within the organization – staff, first and foremost, but also outside stakeholders like shareholders and consumers.

Integrate: If it's possible for the change program of your business to be adjusted, then open up your plans for consultation with your stakeholder group. The staff of an organization are the people who work with a particular product, or encounter customers, or provide services – whatever the case may be for your organization. Because they have direct experience, they are likely to have valuable information and suggestions to offer to a proposed change.

Motivate: One of the biggest challenges with change management processes comes in the form of motivation. How are you meant to motivate staff when there are major changes occurring within an organization? Unfortunately, some changes are not necessarily beneficial – in tough economic changes, occurrences like mergers and downsizing can create an unmotivated environment in some businesses.

Be Confident: Staff are often guilty of reverting to old ways throughout a change process – both effective and ineffective. A change process needs to be organization wide, so have faith in the changes being implemented and show your trust by limiting opportunities for staff to revert to their old habits.

Manage the Workload: Change can bring an increased workload in many organizations, especially if there are changes to technology or company procedure. As a manager in a change situation, be mindful that if you are changing the way in which

your staff operate or produce work, they may be less efficient than they previously were – simply because they are most likely learning new habits again.

Evaluate: As a manager or executive throughout a change management situation, it's crucial that the organization, its passage through the change program, and the wider organization's reactions to that change program are evaluated at every feasible opportunity.

14. What is the difference between traditional information system and enterprise system? Explain with example.

Ans:

Traditional Information System	Enterprise System
• The traditional approach usually consists of custom built data processes and computer information systems tailored for a specific business function.	• Enterprise system is the overall combination of computer hardware and software that a business uses to organize and run its operations.
• It is not beneficial for large funding. Systems were designed to be cheap in order to save on cost	• The initial implementation of ES is difficult and not cost effective.
• simple: Matched existing business processes and functions. Company's were not as interested in funding complicated information systems	• The system is complex and changing or transferring all the data from the present system to ES will slightly exhaust your money, resource and time.
• Lack of data sharing and unnecessary doubling of efforts for processing and managing data	• data can be shared with various departments of the company.

Scanned with CamScanner

• it take very high costs in the long run. Doubling of work as well as excessive maintenance costs	• The risk of wasting time and money on the completion of parallel jobs being addressed in various departments within a company is eliminated by this system.
• Data inconsistency: Redundant data stored in various locations was usually never stored the same way. Formatting was not centrally managed	• Improved data entry: Due to the integration of all systems in an organization, there is little or no entry of redundant data and processes throughout the system.

15. Explain Next Generation balanced Scorecard.

Ans: Traditional balanced scorecard is no longer enough to manage discrete functions separately. They lack the most important element of non-financial performance namely "sustainability concepts" for measuring corporate impacts across the Triple Bottom Line (Environmental, Social, Economic). To implement the concepts that were missing in traditional scorecard, next generation balanced scorecard came into existence.

Next Generation balanced Scorecard is a version of the traditional balanced scorecard, a structured report, supported by proven design methods and automated tools, that can be used by managers to keep track of the execution of activities by the staff within their control, and to monitor the consequences arising from these actions

The third-generation version was developed in the late 1990s to address design problems inherent to earlier generations. It is distinguished by the components making up the balanced scorecard and the design process used to develop these components.

Components

A next-generation balanced scorecard has four main components:

- **A destination or Vision statement**

  This is a one or two page] description of the organization at a defined point in the future, assuming the current strategy has been successfully implemented. The descriptions of the successful future are segmented into perspectives for example financial & stakeholder expectations, customer & external relationships, processes & activities, organization & culture

- **A strategic linkage model**

  This is a version of the traditional strategy map that typically contains 12-24 strategic objectives segmented into two perspectives, activities and outcomes, analogous to the logical framework. Linkages indicate hypothesized causal relations between strategic objectives.

- A set of definitions for each of the strategic objectives.
- A set of definitions for each of the measures selected to monitor each of the strategic objectives, including targets.

Design Process

The design process for third-generation balanced scorecard requires the active involvement of the management team who will eventually use the balanced scorecard. The managers themselves make all decisions about the balanced scorecard content. The process begins with the development of a 'destination or Vision statement' to build management consensus on longer term strategic goals; this document is then used to build a 'strategic linkage model', describing the shorter term management priorities, both the strategic activities to complete and the strategic outcomes to achieve. Once the 'strategic objectives' are decided in outline these are assigned 'owners' from within the management team, who further define each objective, plus the measures and targets associated with achieving each objective.

Advantages

The main improvements with this third generation of balanced scorecard relate to the improved "ownership" of the objectives

Scanned with CamScanner

by managers, because objectives, goals and outcomes are refined by the responsible managers themselves. All measures and targets must be reviewed to ensure that they are interconnected (dependencies identifies) and congruent (appropriately prioritized and support one another).

16. Explain the importance of web mining and its types.

Ans. Web mining is the process of using data mining techniques and algorithms to extract information directly from the Web by extracting it from Web documents and services, Web content, hyperlinks and server logs. The goal of Web mining is to look for patterns in Web data by collecting and analyzing information in order to gain insight into trends, the industry and users in general.

Types of Web mining:

- **Web content mining** — This is the process of mining useful information from the contents of Web pages and Web documents, which are mostly text, images and audio/video files. Techniques used in this discipline have been heavily drawn from natural language processing (NLP) and information retrieval.
- **Web structure mining** — This is the process of analyzing the nodes and connection structure of a website through the use of graph theory. There are two things that can be obtained from this: the structure of a website in terms of how it is connected to other sites and the document structure of the website itself, as to how each page is connected.
- **Web usage mining** — This 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.

17. What do you mean by cold start problem in collaborative filtering? Explain with example.

Ans: The term "Cold start" derives from cars/automotive in cold environment. When it's really cold, the engine has problems

with starting up, but once it reaches its optimal operating temperature, it will run smoothly. With recommendation engines, the "cold start" simply means that the circumstances are not yet optimal for the engine to provide the best possible results. In Ecommerce, there are two distinct categories of cold start: product cold start and user cold starts. Here we're going to examine both.

The right target audience for an advertisement is best calculated by looking at the former visitors for the ad. According to the basic assumption of the collaborative filtering concept, if an ad was already popular with a certain group of people, then others that fit the group's profile are likely to respond well to the ad. However, each time a new ad is placed on your site, it goes through the cold start phase due to the lack of valuable user interactions. User actions are incredibly important since these determine the future of both product-to-product and personalized user-history-based recommendations. If there aren't enough user actions for a certain ad to set the foundations for accurate recommendations, the engine will not know when to display this particular ad. So, we could say, that the more interaction an ad has collected, the easier it is for the recommendation system to qualify and target. News sites, auction sites, Ecommerce stores and classified sites all experience the product cold start.

So, Cold start (in collaborative filtering) is a potential problem in computer-based information systems which involve a degree of automated data modeling. Specifically, it concerns the issue that the system cannot draw any inferences for users or items about which it has not yet gathered sufficient information.

18. Explain in detail about Enterprise management system (EMS) with practical example. What are the benefit of using EMS?

Ans. Enterprise Management Systems (EMS) Are Large- scale Application Software Packages That Support Business Processes, Information Flows, Reporting, And Data Analytics

Scanned with CamScanner

in complex organizations. Enterprise systems integrate a number of different applications, protocols and formats. In doing so, an enterprise system allows companies to integrate business processes, such as sales, deliveries and accounts receivable, by sharing information across business functions and employee hierarchies.

Some of examples of EMS are:

Enterprise Resource Planning: The enterprise resource planning system integrates software applications just as a company integrates business processes, such as purchasing, finance, human resources and inventory management. Within an ERP system, the integrated software modules, such as sales, quality management and accounts receivable, communicate and share data.

Customer relationship management systems were developed to address the need to raise a sales department's productivity and make the management of a company's customers an effective way to increase sales. With CRM functions, such as sales opportunity management, a company learns more about its customers' needs and buying behavior and combines this information with market information to enhance the quality of the company's marketing plans and sales forecasts.

Supply Chain Management: A supply chain refers to the collection of people, tasks, equipment, data and other resources required to produce and move products from a vendor to a customer. Supply chain management refers to the management of supply chain activities by the supply chain firms in an effective and efficient way that will provide a company with strategic advantage.

Benefits of using EMS:

- Improved efficiency and reduced operational costs
- Improved effectiveness
- Improved visibility into operations and performance
- Increased control and governance
- Better service and customer experience

- Improved access and communication channels, plus more effective communication
- Improved accountability, even across business-function boundaries
- Better understanding of what services are needed and provided
- Improved collaboration within and across business functions

19. Explain Next Generation balanced Scorecard.

Ans. Traditional balanced scorecard is no longer enough to manage discrete functions separately. They lack the most important element of non-financial performance namely "sustainability concepts" for measuring corporate impacts across the Triple Bottom Line (Environmental, Social, Economic). To implement the concepts that were missing in traditional scorecard, next generation balanced scorecard came into existence.

Next Generation balanced Scorecard is a version of the traditional balanced scorecard, a structured report, supported by proven design methods and automated tools, that can be used by managers to keep track of the execution of activities by the staff within their control, and to monitor the consequences arising from these actions

The next-generation version was developed in the late 1990s to address design problems inherent to earlier generations. It is distinguished by the components making up the balanced scorecard and the design process used to develop these components.

Components

A next-generation balanced scorecard has four main components:

- **A destination or Vision statement**

  This is a [one or two page] description of the organization at a defined point in the future, assuming the current strategy has been successfully implemented. The descriptions of the successful future are segmented into perspectives for

Scanned with CamScanner

example financial & stakeholder expectations, customer & external relationships, processes & activities, organization & culture

- **A strategic linkage model**

  This is a version of the traditional strategy map that typically contains 12-24 strategic objectives segmented into two perspectives, activities and outcomes, analogous to the logical framework. Linkages indicate hypothesised causal relations between strategic objectives.

- A set of definitions for each of the strategic objectives.
- A set of definitions for each of the measures selected to monitor each of the strategic objectives, including targets.

Design Process

The design process for next-generation balanced scorecard requires the active involvement of the management team who will eventually use the balanced scorecard. The managers themselves make all decisions about the balanced scorecard content. The process begins with the development of a 'destination or Vision statement' to build management consensus on longer term strategic goals; this document is then used to build a 'strategic linkage model', describing the shorter term management priorities, both the strategic activities to complete and the strategic outcomes to achieve. Once the 'strategic objectives' are decided in outline these are assigned 'owners' from within the management team, who further define each objective, plus the measures and targets associated with achieving each objective.

Advantages

The main improvements with this next generation of balanced scorecard relate to the improved "ownership" of the objectives by managers, because objectives, goals and outcomes are refined by the responsible managers themselves. All measures and targets must be reviewed to ensure that they are interconnected (dependencies identifies) and congruent (appropriately prioritized and support one another).

20. What is the use of CRM system in enterprise? Discuss key features and functionality of a CRM system.

Ans. **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. Customer relationship management (CRM) is a term that refers to practices, strategies and technologies that companies use to manage and analyse customer interactions and data throughout the customer lifecycle, with the goal of improving customer service relationships and assisting in customer retention and driving sales growth.

The use of CRM system in enterprise are:

a. Improved Informational Organization

The more you know about your customers, the better you'll be able to provide them with the kind of positive experience that really pays off. Everything that they do, and every interaction that they have with your organization needs to be identified, documented, and recorded. To do this, you need to move beyond the sticky-notes and disorganized filing cabinets, and start utilizing advanced organizational technology that can not only accurately quantify and categorize data for easy future reference, but also make that data available across departments. Thanks to CRM this all becomes a possibility, it allows you to store a vast list of customers and any important information regarding them. Access to their file is even more convenient than before due to the cloud, so no matter who it is that is helping the customer in question, they'll have the same actionable data instantly available. This will result in less wasted time for clients and employees.

Scanned with CamScanner

b. CRM for Enhanced Communication

As mentioned above, CRM makes it possible for any employee to provide the same high level of service, by having access to the same customer data. After all, even if your customers have a single, main point of contact, there's a good chance that at some point that contact may not be available, and the client will be forced to have to work with someone new. When that happens, many customers face the unhappy prospect of having to 'start fresh' with someone who doesn't understand their own unique preferences and issues. CRM does away with this concern, by making detailed customer information communicable to whomever might need it. As such, it won't matter who it is that is currently assisting the client, because they'll be working from the same information. And given that CRM is cloud-based and accessible from any device with an internet connection, the communication benefits of mobile CRM are not limited to the office.

c. CRM Improves Your Customer Service

Your time is valuable, but so is your customers' time. And, should your customers experience a problem that needs resolution, they're going to be unhappy unless that problem can be taken care of quickly. With CRM, as soon as a customer contacts your company, your representatives will be able to retrieve all available activity concerning past purchases, preferences, and anything else that might assist them in finding a solution. In many cases, your more experienced representatives, armed with past information and history, will be able to locate a solution within the first few minutes, thanks to an accessible database of potential issues. And, should a solution not be readily apparent, then bringing in other representatives, or even crowdsourcing for answers through customer portals, is a simple matter. With CRM, customer support becomes a walk in the park.

d. Automation of Everyday Tasks

Completing a sale is never as easy as just getting a customer to agree to commit. Along with the surface details of any sale, there are hundreds of smaller tasks that must be completed in order for everything to function properly. Forms need to be filled out, reports need to be sent, legal issues need to be addressed—these ancillary chores are a time consuming, yet vital aspect of the sales process. The best CRM systems are designed to take the burden of many of these tasks from off the shoulders of your employees, thanks to the magic of automation. This means that your representatives will be able to focus more of their efforts towards closing leads and resolving customer pain points, while the automated CRM system takes care of the details.

e. Greater efficiency for multiple teams

Automatically stored communication allows you to view emails, calendar and phone call details in one easily accessible place. Add that to the ability for multiple teams to access the same information, it simply sky rockets the amount of achievable progress. Sales, marketing, and customer service teams can share valuable information about clients to continue to funnel them down the pipeline to get the desired result of closing a sale, knowledge of new products, or excellent customer service. Every department can now tag team to get the right information to the right individual. With this new found ease, teams can seamlessly work together to improve the bottom line.

f. Improved Analytical Data and Reporting

Miscalculated data should not be the reason you cannot succeed, with CRM this is no longer a possibility. CRM systems store information in one place which leads to improved analysing of the data as a whole. Easily integrated with different tools or plugins, you have the ability to generate automatic reports to maximize your time. Personalize your dashboard views to quickly locate information needed such as customer information, sales

Scanned with CamScanner

goals, and performance reports to reach untapped opportunities. With better reporting data you can make resourceful and effective decisions to reap the rewards in customer loyalty and long run profitability.

The key features and functionality of CRM are:

a. Contact Management

Organizes customer data in a centralized location. Contact details, demographics, transactions, customer service interactions, and other customer data are easily saved and retrieved.

b. Reporting and dashboards

Consolidates streams of data in meaningful ways that help you to spot patterns, trends, behaviors, and other market insights. Many CRM solutions can also generate real-time data that makes reporting more timely and accurate.

c. Lead management

Manages leads from prospecting to conversion. It lets you track sales activities, deals, and tasks and move them across sales stages. It also connects to contact management to let you view customer interactions and profiles in one window. Most CRM solutions allow you to customize the pipeline stages.

d. Deals and Tasks

CRM lets you add, assign, and track actionable items and feed them into the sales pipeline. Deals and tasks can be automatically populated with customer data or updated with real-time information.

e. Campaign management

Manages campaigns from start to end, including setting up target customers, automating workflows, and measuring success based on clearly defined metrics.

f. Email tracking

Tracks opened or clicked emails and usually works with popular email clients and mailers like Gmail, Outlook, and

MailChimp. Email tracking can be a part of campaign management or sold separately.

g. Social media management

Tracks social media activity of your audience, such as, likes and dislikes, comments, mentions, and shares. It can also be a part of campaign management or sold separately.

h. Mobile version

Many CRM solutions have Android and iOS apps, while others have mobile responsive sites.

20. Discuss the concept of collaborative filtering with suitable practical examples.

Ans. Collaborative filtering is the process of filtering or evaluating an item through the opinions of other people. CF is a technique used by recommender systems. It brings together the opinions of large interconnected communities on the web, supporting filtering of substantial quantities of data. Collaborative filtering systems produce automatic predictions (filtering) about the interests of an individual by collecting preferences or information from several other individuals. The underlying assumption of the collaborative filtering approach is that if a person A has the same opinion as a person B on a issue, A is more likely to have B's opinion on a different issue than a randomly chosen person. There are 3 types of ratings in a collaborative filtering system.

- Scalar ratings: consist of either numerical rating from 1-5 stars in accordance to strongly agree, agree, neutral,, disagree, strongly disagree.
- Binary ratings: either agree/disagree or good/bad.
- Unary ratings: indicate whether a user has observed or purchased an item, or has rated the item positively.

CF works in the following way:

Many people must participate so that a person with similar interests is easily found. The users rate the items and their interests are recorded. The ratings may be explicit (buying or

Scanned with CamScanner

rating an item) or implicit (browsing time, number of clicks). Nearest neighbor matching is used to find the people with similar interests. Items that neighbors rate highly but you haven't rated are recommended to you. The user can then rate the recommended items.

Types:

i. User-based collaborative filtering: A specific application of this is the user-based Nearest Neighbor algorithm.

- Look for users who share the same rating patterns with the active user (the user whom the prediction is for).
- Use the ratings from those like-minded users found in step 1 to calculate a prediction for the active user.

ii. Item-based collaborative filtering invented by Amazon.com (users who bought x also bought), proceeds in an item-centric manner:

- Build an item-item matrix determining relationships between pairs of items
- Using the matrix, and the data on the current user, infer his taste

iii. User behavior based- works on implicit observations of normal user behavior (as opposed to the artificial behavior imposed by a rating task). In these systems you observe what a user has done together with what all users have done (what music they have listened to, what items they have bought) and use that data to predict the user's behavior in the future or to predict how a user might like to behave if only they were given a chance. These predictions then have to be filtered through business logic to determine how these predictions might affect what a business system ought to do. It is, for instance, not useful to offer to sell somebody some music if they already have demonstrated that they own that music or, considering another example, it is not useful to suggest more travel guides for Paris to someone who already bought a travel guide for this city.

Problems:

- **Cold start problem** - A collaborative filtering system doesn't automatically match content toone's preferences. Unless achieving good diversity and independence of opinions, one point of view will always dominate another in a particular community. As in the personalized recommendation scenario, new user or new item can always cause a problem, as there will be insufficient data for the collaborative filtering to work. In order to make accurate recommendations, the system must first learn the user's preferences from user's past voting or rating activities. For a new item added, the collaborative filtering system would not be able to recommend it until it's rated by a substantial number of users.
- **Scalability**- As database size increases, the recommendation computation becomes computationally more intensive.
- **Synonyms**- refer to the tendency of a number of the same or very similar items to have different names or entries. Most recommender systems are unable to discover this latent association and thus treat these products differently.
- **Grey sheep**- refers to the users whose opinions do not consistently agree or disagree with any group of people and thus do not benefit from collaborative filtering. Black sheep are the opposite group whose idiosyncratic tastes make recommendations nearly impossible.
- **Shilling attacks**- people may give lots of positive ratings for their own items and negative ratings for their competitors. It is often necessary for the collaborative filtering systems to introduce precautions to discourage such kind of manipulations.
- **Diversity**- Collaborative filters are expected to increase diversity because they help us discover new products among myriad choices. Some algorithms, however, may unintentionally do the opposite. Because collaborative filters recommend products based on sales or ratings, they cannot

Scanned with CamScanner

usually recommend products with limited historical data. This creates a rich-get-richer effect for popular products, leading to less diversity.

21. Differentiate between Map Reduce and Hadoop.

MapReduce	Hadoop
1. It is a programming paradigm used in Hadoop.	1. It is an open source software framework that support data-intensive distributed applications.
2. It allows data to distributed across a large cluster.	2. It support running applications on large clusters with both reliability and data motion to application
3. It is a software framework that allow developers to write a program that process massive amounts of unstructured data in parallel across a distributed cluster of processor or stand-alone computers.	3. It utilize Hadoop distributed File system written in Java.
4. It refers two separate: Map and Reduce.	4. It can process any kind of data such as Structured, Semi-Structured and Un-Structure.
5. Map Job: takes a set of data and converts it into another set of data. Reduce Job: takes the output from map as input and combine those tuples into smaller set of tuples.	5. Processing Coupled with Data storage.
6. Its framework is Fault-tolerant.	6. Fault Tolerance is nothing but Replication of the Data Blocks.

22. What is Cloud Computing? Explain in detail.

Ans. Cloud Computing is the practice of using a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computer. It is a model for enabling ubiquitous network access to a shared pool of configurable computing resources. Cloud computing and storage solutions provide users and enterprises with various capabilities to store and process their data in third-party data centers. These platforms hide the complexity and details of the underlying infrastructure from users and applications by providing very simple graphical interface or API (Applications Programming Interface). It is synonymous to distributed computing over a network and is able to run programs on many computers simultaneously despite heterogeneity.

A number of characteristics define cloud data, applications services and infrastructure:

- Remotely hosted: Services or data are hosted on remote infrastructure.
- Ubiquitous: Services or data are available from anywhere.
- Commodified: The result is a utility computing model similar to traditional that of traditional utilities, like gas and electricity - you pay for what you would want!

23. What are the different issues that should be taken into consideration while choosing cloud as IS platform?

Ans. The different issues that should be taken into consideration while choosing cloud as IS platform are as follows:

- Data Security and Privacy Requirement
- Budget Requirements
- Type of cloud - public, private or hybrid
- Data backup requirements
- Training requirements
- Dashboard and reporting requirements
- Client access requirements
- Data export requirements

Scanned with CamScanner

24. What are different techniques and methods developed to handle voluminous data? Discuss two such techniques.

Ans. The Techniques and methods to handle voluminous data are:

1. **Processing Big Data:**

   Map reduce, Hadoop is an integrated framework for processing and storing Big data.

2. **Analysis and querying of data:**

   WibiData, PLATFORA, PIG

3. **Business Intelligence:**

   Hive

4. **Storage:**

   Cloud storage, Column-oriented database, schema-less database

5. **Machine Learning :**

   Apache Mahout, SkyTree

1. **MapReduce:**

   MapReduce is the key algorithm that the Hadoop MapReduce engine uses to distribute work around a cluster.

   Mapper function: A map transform function is provided to transform an input data row of key and value to an output key/value:

   · map(key1,value) -> list <key2, value2>

   That is, for an input it returns a list containing zero or more (key, value) pairs: The output can be a different key from the input. The output can have multiple entries with the same key

   Reduce function: A reduce transform is provided to take all values for a specific key, and generate a new list of the reduced output.

   · reduce(key2, list<value2>) -> list <value3>

2. **Hadoop:**

   Hadoop is an open source framework for distributed storage and processing of large sets of data on commodity hardware.

Hadoop enables businesses to quickly gain insight from massive amounts of structured and unstructured data. It is used in maintaining, scaling and analyzing large scale of data. This data can be structured or unstructured.

3. **PIG:**

   PIG is a platform for analyzing large data sets. PIG's language, PIG Latin, lets one specify a sequence of transformation functions like merge, filter, grouping etc. Apart from built-in functions it also provides facility for user defined functions to do special-purpose processing. PIG's language allows for query execution over data stored on a Hadoop cluster, instead of a "SQL-like" language.

4. **Hive**

   Hive enables traditional BI applications to run queries against a Hadoop cluster. It was developed originally by Facebook, but has been made open source for some time now, and it's a higher-level abstraction of the Hadoop framework that allows anyone to make queries against data stored in a Hadoop cluster just as if they were manipulating a conventional data store. It makes Hadoop more useful for BI users.

5. **Column-Oriented Database:**

   Conventional, row-oriented databases are best fit for online transaction processing with high update speeds, but they fall short on query performance as the data volumes grow and as data becomes more unstructured. Column-oriented databases store data with a focus on columns, instead of rows, allowing for huge data compression and very fast querying.

6. **Schema-Less Database or NoSQL Database:**

   There are several database types that fit into this category, such as key-value stores and document stores, which focus on the storage and retrieval of large volumes of unstructured, semi-structured, or even structured data. They achieve performance gains by doing away with some (or all) of the restrictions traditionally associated with conventional

Scanned with CamScanner

databases, such as read-write consistency, in exchange for scalability and distributed processing.

7. **Clouds for Big Data:**

Most of the above technologies demand cloud, ie , many of the products and platforms mentioned are either entirely cloud-based or have cloud versions themselves. Most cloud vendors are already offering hosted Hadoop clusters that can be scaled on demand according to their user's needs. Big Data and cloud computing go hand-in-hand. Cloud computing enables companies of all sizes to get more value from their data by enabling faster analytics at minimal costs. This, in turn improves the company's productivity and thus returns the costs invested.

25. What is key challenge in handling the voluminous data in cloud architecture?

Ans:
- **Security**: With cloud computing, we can heavily dependent on the service provider for security. Cloud service providers can claim to provide complete security for access, compliance, data segregation, backup, recovery, etc.
- **Data Ownership**: What happens to our data when it leaves our organization to reside in the cloud? Companies who move to the cloud probably will not completely lose track of their data but they are likely to lose some level of ownership and control. It is important to understand who can access the data and for what purpose.
- **Lock-in and Interoperability**: Now days each service offering provides its own unique way for the cloud to interact with applications, data and clients. It can be very difficult to use multiple vendors and to have consistent cloud services.
- **Standard Architecture**: There is no standard open architecture being used for cloud services. Each of the major cloud providers (Amazon Web Services, Salesforce Force, Google App Engine and Microsoft Azure) imposes different architectures that are dissimilar to the common architectures

currently used for enterprise apps. Although a customer's functional and technical requirements may be satisfied, the lack of standards will compromise the customer's ability to consistently migrate from one service provider to another and may need a complete rewrite of their software to do so.

- **Enterprise Support and Service Maturity**: Cloud computing services may not provide the levels of reliability, manageability, and support required by large enterprises.
- **Loss of data**: Data stored in the cloud can be replicated across multiple machines and backed up. However, not all cloud services have the same redundancy for disaster recovery.
- **Return on Investment**: The expectation is that external cloud computing can reduce costs. However, the cost advantage for large enterprises may not be as clear as for SMEs. Currently, many large enterprises can reap the benefits of significant economies of scale in their own internal IT operations.
- **Requirement for online connectivity**: Cloud computing is impossible if you cannot connect to the internet. A dead internet connection means it is not possible to work, and in areas where internet connections are few or inherently unreliable, this could be a problem.

26. What are characteristics of voluminous data? Describe how Hadoop and Map Reduce work on big data over distributed architecture. 'Cloud is basically a client server model in essence, do you agree with this statement?

Ans: **Characteristic of voluminous data are as follows:-**

- **Volume:** Volume Refers to the vast amounts of data generated every second. We are not talking Terabytes but Zetta bytes or Bronto bytes. These volumes of data sets are too large to store and analyze using traditional database technology. New big data tools use distributed systems so that we can store and analyze data across databases that are dotted around anywhere in the world.

Scanned with CamScanner

- **Variety:** Different Types: Variety describes different formats of data that do not lend themselves to storage in structured relational database systems. These include a long list of data such as documents, emails, social media text messages, video, still images, audio, graphs, and the output from all types of machine-generated data from sensors, devices, RFID tags, machine logs, cell phone GPS signals, DNA analysis devices, and more. This type of data is characterized as unstructured or semi-structured and has existed all along. Variety is also used to mean data from many different sources, both inside and outside of the company.
- **Velocity:** Data-In-Motion: Data scientists like to talk about data-at-rest and data-in-motion. One meaning of Velocity is to describe data-in-motion, for example, the stream of readings taken from a sensor or the web log history of page visits and clicks by each visitor to a web site. This can be thought of as a fire hose of incoming data that needs to be captured, stored, and analyzed.
- **Lifetime of Data Utility**: A second dimension of Velocity is how long the data will be valuable. Is it permanently valuable or does it rapidly age and lose its meaning and importance. Understanding this dimension of Velocity in the data you choose to store will be important in discarding data that is no longer meaningful and in fact may mislead.
- **Value:** Although Value is frequently shown as the fourth leg of the Big Data stool, Value does not differentiate Big Data from not so big data. It is equally true of both big and little data that if we are making the effort to store and analyze it then it must be perceived to have value.

  There are at least four additional characteristics that pop up in the literature from time to time. All of these share the same definitional problems of Value. That is they may be a descriptor of data but not uniquely of Big Data
- **Veracity:** What is the provenance of the data? Does it come from a reliable source? It is accurate and by extension, complete.

- **Variability:** There are several potential meanings for Variability. Is the data consistent in terms of availability or interval of reporting? Does it accurately portray the event reported? When data contains many extreme values it presents a statistical problem to determine what to do with these 'outlier' values and whether they contain a new and important signal or are just noisy data.
- **Viscosity:** This term is sometimes used to describe the latency or lag time in the data relative to the event being described. We found that this is just as easily understood as an element of Velocity.
- **Virality:** Defined by some users as the rate at which the data spreads; how often it is picked up and repeated by other users or events.

  A Hadoop job consists of the input file(s) on HDFS, m- map tasks and n- reduce tasks, and the output is n- files. The stages of one map-reduce iteration are:
  - mapper (written by programmer)
  - combiner (written by programmer)
  - sort and shuffle (done by Hadoop framework)
  - reduce (written by programmer)

  At each such iteration, there is input read in from HDFS and given to the mapper, and output written out to HDFS by the reducer. Optimizing the MapReduce pipeline often consists of minimizing the I/O transfers.

27. Discuss cloud computing technologies and their types.

Ans: Cloud Computing refers to manipulating, configuring, and accessing the hardware and software resources remotely. It offers online data storage, infrastructure, and application. Cloud computing offers platform independency, as the software is not required to be installed locally on the PC. Hence, the Cloud Computing is making our business applications mobile and collaborative.

Scanned with CamScanner

The types of cloud computing are described below:

1. **Software as a Service (SaaS)**

   SaaS is a software delivery method that provides access to software and its functions remotely as a Web-based service. Instead of paying an upfront fee to purchase and/or license software, SaaS customers pay a recurring (often monthly or annual) fee to subscribe to the service. In general, they can access the SaaS from any Internet-connected device, any time day or night. Well-known examples of SaaS include Salesforce.com, Microsoft Office 365, Google G Suite, Dropbox, Adobe Creative Cloud and others.

2. **Platform as a Service (PaaS)**

   PaaS is a computing platform being delivered as a service. Here the platform is outsourced in place of a company or data center purchasing and managing its own hardware and software layers. Most PaaSes are designed for developers and aim to simplify the process of creating and deploying software. For example, a Web developer might use a PaaS that includes operating system software, Web server software, a database and related Web development tools. The leading PaaS vendors include Amazon Web Services, Microsoft Azure, IBM and Google Cloud Platform.

3. **Infrastructure as a Service (IaaS)**

   Computer infrastructure, such as servers, storage and networking delivered as a service. IaaS is popular with enterprises that appreciate the convenience of having the cloud vendor manage their IT infrastructure. They also sometimes see cost savings as a result of paying only for the computing resources they use. The leading IaaS vendors include Amazon Web Services, Microsoft Azure, IBM and Google Cloud Platform.

   While SaaS, PaaS and IaaS are the three most common types of cloud services, cloud computing vendors sometimes also use other "as a service" labels to describe their offerings. For example, some offer database as a service (DBaaS), mobile back-end as a service (MBaaS), functions as a service (FaaS) or others.

28. Differentiate between Map Reduce and Hadoop.

MapReduce	Hadoop
1. It is a programming paradigm used in Hadoop.	1. It is an open source software framework that support data-intensive distributed applications.
2. It allows data to distributed across a large cluster.	2. It support running applications on large clusters with both reliability and data motion to application
3. It is a software framework that allow developers to write a program that process massive amounts of unstructured data in parallel across a distributed cluster of processor or stand-alone computers.	3. It utilize Hadoop distributed File system written in Java.
4. It refers two separate: Map and Reduce.	4. It can process any kind of data such as Structured, Semi-Structured and Un-Structure.
5. Map Job: takes a set of data and converts it into another set of data. Reduce Job: takes the output from map as input and combine those tuples into smaller set of tuples.	5. Processing Coupled with Data storage.
6. Its framework is Fault-tolerant.	6. Fault Tolerance is nothing but Replication of the Data Blocks.

Scanned with CamScanner

29. Illustrate what type of cloud technologies would you recommend for a small business firm with high level of accuracy and security required for business transaction?

Ans: We recommend a private cloud technology for a small business firm with high level of accuracy and security required for business transaction.

Private cloud is cloud infrastructure operated solely for a single organization, whether managed internally or by a third-party, and hosted either internally or externally. Undertaking a private cloud project requires significant engagement to virtualize the business environment, and requires the organization to reevaluate decisions about existing resources. It can improve business, but every step in the project raises security issues that must be addressed to prevent serious vulnerabilities. Self-run data centers are generally capital intensive. They have a significant physical footprint, requiring allocations of space, hardware, and environmental controls. These assets have to be refreshed periodically, resulting in additional capital expenditures. They have attracted criticism because users "still have to buy, build, and manage them" and thus do not benefit from less hands-on management, essentially "the economic model that makes cloud computing such an intriguing concept"

DIFFERENTIATE OLAP AND OLTP

Data mining is the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis. Data mining tools allow enterprises to predict future trends.

Comparison between OLTP and OLAP:

	OLAP	OLTP
Abbreviation	• It stands for 'Online Analytical Processing'.	• It stands for 'Online Transaction Processing'.
Use	• It is used for Query Processing.	• It is used for Transaction Processing.

	OLAP	OLTP
Data	• It holds historical data. • It stores only relevant data. • It is de-normalized in the analytical process. • It has a large database. • It contains non-volatile data.	• It holds current data. • It stores all data. • It is normalized for efficient transaction process. • It has a small database. • It contains volatile data.
Type	• It is analysis driven.	• It is application driven.
Source	• The data comes from various OLTP sources.	• It is the original source of data.
Purpose	• To help with planning, problem solving, and decision support.	• To control and run fundamental business tasks.
Business	• It reveals the multi-dimensional view of all types of business activities.	• It reveals the ongoing business process.
Updates	• There are periodic long-running batch jobs which refresh the data.	• Short and fast inserts and updates initiated by end users.
Queries	• They are often complex queries involving aggregations.	• They are standardized and simple queries.
Speed	• It is slow depending on the data.	• It is very fast.
Market	• It is customer orientated.	• It is market orientated.
Database design	• It is de-normalized with fewer tables and makes use of star or snowflake schemas.	• It is highly normalized with many tables.

Scanned with CamScanner

	OLAP	OLTP
View	• It represents managerial view.	• It represents clerical or operator view.
Users	• It has few concurrent users.	• It has many concurrent users.

30. What is GDSS? Describe the components, features and tools of GDSS.

Ans: A group decision support system (GDSS) 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. They are designed in such a way that they take input from multiple users interacting simultaneously with the systems to arrive at a decision as a group.

The tools and techniques provided by group decision support system improve the quality and effectiveness of the group meetings. Groupware and web-based tools for electronic meetings and videoconferencing also support some of the group decision making process, but their main function is to make communication possible between the decision makers.

In a group decision support system (GDSS) electronic meeting, each participant is provided with a computer. The computers are connected to each other, to the facilitator's computer and to the file server. A projection screen is available at the front of the room. The facilitator and the participants can both project digital text and images onto this screen.

A group decision support system (GDSS) meeting comprises different phases, such as idea generation, discussion, voting, vote counting and so on. The facilitator manages and controls the execution of these phases. The use of various software tools in the meeting is also controlled by the facilitator.

A Group decision support system (GDSS) is composed of 3 main components, namely hardware, software tools, and people.

- **Hardware**: It includes electronic hardware like computer, equipment used for networking, electronic display boards

and audio-visual equipment. It also includes the conference facility, including the physical setup – the room, the tables and the chairs – laid out in such a manner that they can support group discussion and teamwork.

- **Software Tools**: It includes various tools and techniques, such as electronic questionnaires, electronic brainstorming tools, idea organizers, tools for setting priority, policy formation tool, etc. The use of these software tools in a group meeting helps the group decision makers to plan, organize ideas, gather information, establish priorities, take decisions and to document the meeting proceedings. As a result, meetings become more productive.
- **People**: It compromises the members participating in the meeting, a trained facilitator who helps with the proceedings of the meeting, and an expert staff to support the hardware and software. The GDSS components together provide a favorable environment for carrying out group meetings.

Features of Group Decision Support System (GDSS)

- Ease of Use: It consists of an interactive interface that makes working with GDSS simple and easy.
- Better Decision Making: It provides the conference room setting and various software tools that facilitate users at different locations to make decisions as a group resulting in better decisions.
- Emphasis on Semi-structured and Unstructured Decisions: It provides important information that assists middle and higher-level management in making semi-structured and unstructured decisions.
- Specific and General Support: The facilitator controls the different phases of the group decision support system meeting (idea generation, discussion, voting and vote counting etc.) what is displayed on the central screen and the type of ranking and voting that takes place, etc. In addition, the facilitator also provides general support to the group and helps them to use the system.

Scanned with CamScanner

- Supports all Phases of the Decision Making: It can support all the four phases of decision making, viz intelligence, design, choice and implementation.
- Supports Positive Group Behavior: In a group meeting, as participants can share their ideas more openly without the fear of being criticized, they display more positive group behavior towards the subject matter of the meeting.

Group Decision Support System (GDSS) Software Tools

Group decision support system software tools helps the decision makers in organizing their ideas, gathering required information and setting and ranking priorities. Some of these tools are as follows:

- **Electronic Questionnaire**: The information generated using the questionnaires helps the organizers of the meeting to identify the issues that need immediate attention, thereby enabling the organizers to create a meeting plan in advance.
- **Electronic Brainstorming** Tools: It allows the participants to simultaneously contribute their ideas on the subject matter of the meeting. As identity of each participant remains secret, individuals participate in the meeting without the fear of criticism.
- **Idea Organizer**: It helps in bringing together, evaluating and categorizing the ideas that are produced during the brainstorming activity.
- **Tools for Setting Priority**: It includes a collection of techniques, such as simple voting, ranking in order and some weighted techniques that are used for voting and setting priorities in a group meeting.
- **Policy Formation Tool**: It provides necessary support for converting the wordings of policy statements into an agreement.

31. What is web? How is it organized? Discuss the working method of search engine with respect to the structure of web.

Ans. The World Wide Web (abbreviated WWW or the Web) is an information space where documents and other web resources

are identified by Uniform Resource Locators (URLs), interlinked by hypertext links, and accessible via the Internet.

Generally A web search engine can be defined as a software system that is designed to search for information on the World Wide Web. A search engine maintains the following processes in near real time:

- Web search engines get their information by web crawling from site to site. The "spider" checks for the standard filename robots.txt, addressed to it, before sending certain information back to be indexed depending on many factors, such as the titles, page content, JavaScript, Cascading Style Sheets (CSS), headings, as evidenced by the standard HTML markup of the informational content, or its metadata in HTML meta tags. No web crawler crawls the entire reachable web. Due to infinite websites, spider traps, spam, and other exigencies of the real web, crawlers instead apply a crawl policy to determine when the crawling of a site should be deemed sufficient. Some sites are crawled exhaustively, while others are crawled only partially.
- Indexing means associating words and other definable tokens found on web pages to their domain names and HTML-based fields. The associations are made in a public database, made available for web search queries. A query from a user can be a single word. The index helps find information relating to the query as quickly as possible.
- Typically when a user enters a query into a search engine it is a few keywords. The index already has the names of the sites containing the keywords, and these are instantly obtained from the index. Every page in the entire list must be weighted according to information in the indexes. Then the top search result item requires the lookup, reconstruction, and markup of the snippets showing the context of the keywords matched. Beyond simple keyword lookups, search engines offer their own GUI- or command-driven operators and search parameters to refine the search

Scanned with CamScanner

results. It is also possible to weight by date because each page has a modification time. Most search engines support the use of the boolean operators AND, OR and NOT to help end users refine the search query. The engine looks for the words or phrases exactly as entered. Some search engines provide an advanced feature called proximity search, which allows users to define the distance between keywords. There is also concept-based searching where the research involves using statistical analysis on pages containing the words or phrases you search for. Also, natural language queries allow the user to type a question in the same form one would ask it to a human. A site like this would be ask.com. The usefulness of a search engine depends on the relevance of the result set it gives back. Most search engines employ methods to rank the results to provide the "best" results first. There are two main types of search engine that have evolved: one is a system of predefined and hierarchically ordered keywords that humans have programmed extensively. The other is a system that generates an "inverted index" by analyzing texts it locates. This first form relies much more heavily on the computer itself to do the bulk of the work. Most Web search engines are commercial ventures supported by advertising revenue and thus some of them allow advertisers to have their listings ranked higher in search results for a fee. Search engines that do not accept money for their search results make money by running search related ads alongside the regular search engine results. The search engines make money every time someone clicks on one of these ads.

SOLVED CASE STUDY

Case Study 1:

Strategy Map and Balanced Scorecard of NIKE

1. Overview

Nike is an American multinational corporationn that is engaged in the design, development, manufacturing, and worldwide marketing and sales of footwear, apparel, equipment, accessories, and services. The company is headquartered near Beaverton, Oregon, in the Portland metropolitan area. It is the world's largest supplier of athletic shoes and apparel and a major manufacturer of sports equipment, with revenue in excess of US$24.1 billion in its fiscal year 2012. It operates with the aim of delivering innovative products, experiences and services to inspire athletes worldwide.

2. History

Year	Major Events
1950s	• Phil Knight and Bill Bowerman, the founders, meet
1960s	• Blue Ribbon Sports (BRS) was made and founded by Phil Knight • The popular Cortez aka "Dope Mans" is made in Japan
1970s	• The Swoosh logo is created by Carolyn Davidson for $35.00 • The first Nike model shoe to hit the retail market is a soccer/football shoe • A Promo Nike Tee becomes the first apparel item • The famous Waffle Trainer is introduced, which becomes the best selling shoe in the US • Factories for manufacturing are set up in Korea and Taiwan • For the first time, Nike shoes are sold in Asia

Scanned with CamScanner

Year	Major Events
	• Blue Ribbon Sports changes its company name to Nike Inc. • World Headquarters are opened in Beaverton, Oregon
1990s	• The new World Campus sits on 74 acres with 570,000 square feet. • In Portland, Oregon the first Nike Town opens • Nike Town opens in Chicago • Nike Town opens in Atlanta and Orange County • Nike gets distribution rights in Korea and Japan
2000	• The National Football League declines to renew its exclusive apparel licensing arrangement with Nike.
2001	• Nike opens its first Nike Goddess store, a unit targeting women, in Newport Beach, CA.
2003	• Nike purchases Converse Inc. for $ 305 million.
2008	• Nike acquired sports apparel supplier Umbro,

3. **Growth Strategies**

NIKE's generic strategies emphasize product mix diversity whereas intensive growth strategies focus on innovation, improving market position, and increasing market share.

Generic Strategies

- **Cost Leadership Strategy:** Minimize cost and maximize profitability through smart outsourcing solutions.
- **Differentiation Strategy:** Integrate cutting edge design and state of the art technologies for uniqueness on the market.

Intensive Growth Strategies

- **Market Development:** Target new markets and market segment.
- **Product Development:** Bring new products with a focus on leveraging state of the art technologies.
- **Market penetration:** Increase presence by augmenting the number of retailers worldwide.

4. **Mission Statement**

The company aims to "lead-in corporate citizenship through proactive programs that reflect caring for the world family of Nike, our teammates, our consumers, and those who provide services to Nike.

To continue to offer quality products with increasing growth in the industry and expanding globally.

To provide a competitive edge by developing the most technological products whilst keeping in mind fair labour practices and maintaining a competitive advantage with the shareholder's interests, and company profits in mind.

Consider employees are one of the most important assets.

To increase the responsibility towards the environment by evaluating the impact of day to day operation and attempts to change operations that have a negative impact.

- **Vision Statement**

  To bring inspiration and innovation to every athlete in the world.

  To equip every athlete with products that combine performance, quality, and fashion.

  To remain the most authentic, connected and distinctive brand.

- **Products and Portfolio**

  Nike produces a wide range of sports equipment. Their first products were track running shoes. They currently also make shoes, jerseys, shorts, base layers etc. for a wide range of sports including track & field, baseball, ice hockey, tennis, Association football, lacrosse, basketball and cricket. The most recent additions to their line are the Nike 6.0 and Nike SB shoes. The portfolio of brands under Nike are:

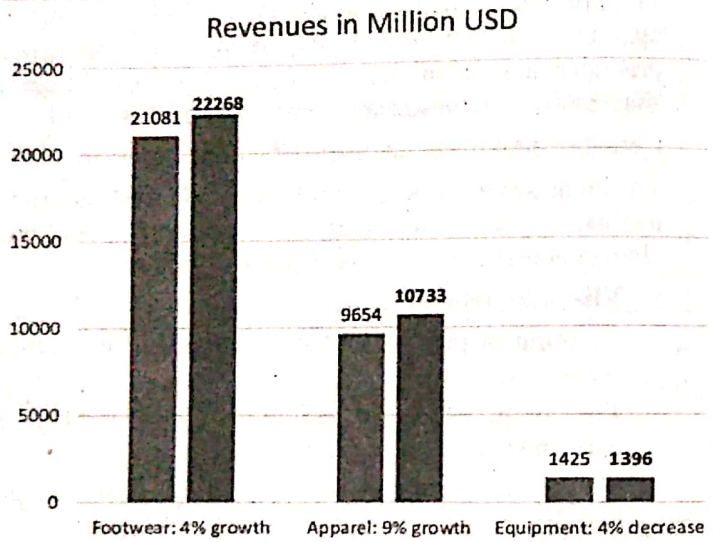
  - Converse Inc
  - Hurley International
  - Jordan Brand
  - Cole Haan
  - Bauer International

Scanned with CamScanner

5. Market Assessment

Revenue Analysis

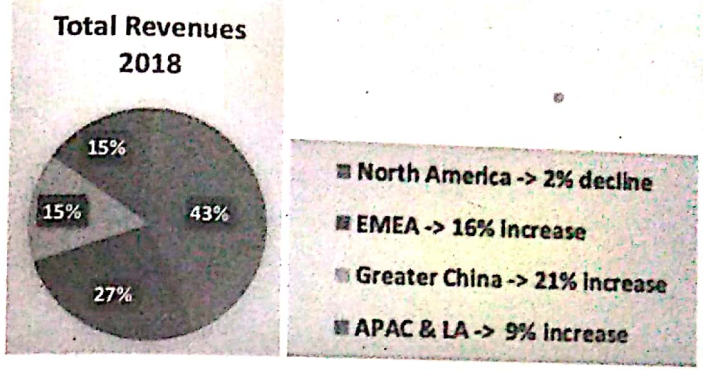
- Nike revenue is growing continuously and achieved a new record for fiscal 2018 by growing 6% to $36.4 billion.
- Revenues of NIKE remained constant (only slightly increase) between 2017 and 2018 in general but equipment revenues dropped slightly by 4%.



- Net income for fiscal 2018 decreased tremendously, in part reflecting the negative impact of weakening foreign currency exchange rates.
- Cash flow from operations skyrocketed over the past 4 years.

In million USD	2018	2017	2016	2015	2014
Revenues	36.397	34.350	32.376	30.601	27.799
Gross profit	15.956	15.312	14.971	14.067	12.446
Gross margin	43.8%	44.6%	46.2%	46.0%	44.8%
Net income	1.933	4.240	3.760	3.273	2.693
Cash flow from operations	4.955	3.846	3.399	4.906	3.158

- North America and EMEA account for almost 2/3rd of Nike's total revenue. Meanwhile, Greater China boasts a double-digit increase in revenue.



6. Customer Environment

Nike has maintained a global customer base through extensive efforts in marketing and sales:

- Nike targets a young and urban international audience with an increasing focus on females
- It reaches out to customers through socially-conscious social media campaigns and partnerships with world-class athletes.
- Nike maintains excellent pre and post-sale customer service
- Value-based pricing has helped to establish a healthy customer environment for the company.

7. Competitive Environment Analysis

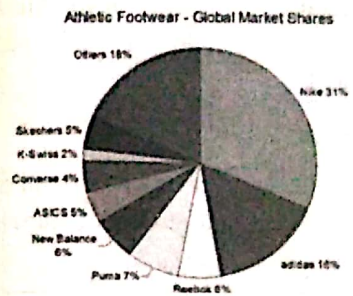
Nike faces major competitions in the market from other sportswear and apparel companies such as Adidas, Reebok, Puma, etc. Adidas competes with Nike being globally present and having high revenues totalling upto around 21.9 billion Euros with net earnings 1.7 billion Euros. Reebok is similarly a huge company which was acquired by Adidas in 2006. Puma has a globally balanced sale with revenue upto 4.7 billion Euros and a net earnings of 187 million Euros. The following table shows a comparison of Nike with two of its major competitors: Adidas and Reebok in terms of various critical success factors:

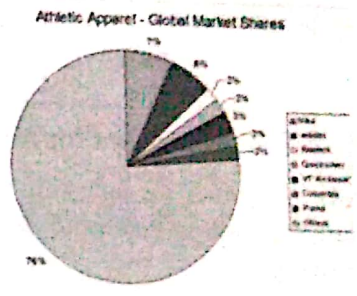
Scanned with CamScanner

		NIKE		ADIDAS		REEBOOK	
CRITICAL SUCCESS FACTORS	WEIGHTS	RATING	W.SCORES	RATING	W.SCORES	RATING	W.SCORES
Domestic factor positioning	0.1	4	0.4	2	0.2	3	0.3
International market positioning	0.1	4	0.4	3	0.3	3	0.3
Consumer loyalty	0.08	3	0.24	3	0.24	3	0.24
Brand recognition	0.1	4	0.4	4	0.4	4	0.4
Price competiveness	0.09	3	0.27	3	0.27	4	0.36
Product Quality	0.07	4	0.28	4	0.28	3	0.21
Relationship with supliers and manufactures	0.07	3	0.21	4	0.28	3	0.21
Product R&D	0.1	4	0.4	3	0.3	3	0.3
Product diversity	0.1	4	0.4	3	0.3	2	0.2
Financial positioning	0.07	3	0.21	3	0.21	2	0.14
Marketing organitational	0.08	4	0.32	4	0.32	3	0.24
Organitational structure	0.04	3	0.12	3	0.12	3	0.12
TOTAL	1		2.85		2.72		2.42

8. Market Share

Nike was the clear market leader, with 31% of the global athletic footwear market in 2007. Looking at the market in the United States, Europe, or Asia reveals a similar picture: Nike's market share in these regions hovers around 36%, followed by Adidas at 20%, with Puma and New Balance as distant third and fourth. The market for athletic apparel is both larger--$49.5 billion in 2005--and more diffuse; the top five firms control only 27% of the market. Nike is, however, also the global leader in apparel, with a 7% market share in 2007.





8 Strategic Planning (SWOT Analysis)

Nike can use the SWOT (Strength, Weakness, Opportunity and Threat) Analysis to perform strategic planning. The various factors under each of the SWOT matrix components have been listed in the table below.

Strengths	Weaknesses
• World's leading brand for sports shoes and apparel	• In Vietnam, the company faced allegations of labour and wage laws with employees
• Launched many other products in collaboration with companies e.g. launched music player, watches etc	• Nike is dependent on its footwear due to other branded products not being as strong.
• Unique and creative marketing campaigns	• Limited presence in developing countries

Scanned with CamScanner

Threats	Opportunities
• Footwear market is so competitive, major competition from other brands	• Expansion into emerging markets
• Fake imitations and replicas cause problems	• Growing segments of sporting industries (women's apparel, parcours, muslims)
• Currency fluctuations of foreign countries	• Enhance brand visibility by opening sports academies across the world for nurturing talent

9. Recommendations for future growth

Nike can pursue the following strategic objectives in order to induce further growth to its already large market share and strong financial position:

- **Extension to new customer segment**
  - Creation of low-priced product variants
  - Trading down with less quality and lower prices
  - Strict differentiation required to maintain premium image
  - Utilization of discounters as distribution channels
- **Commitment to sustainability**
  - Increase the current 71% of recycled materials to 100%: induce customers to send in old products
  - Decentralization & relocation of production facilities to diminish environmental footprint and distribution cost
  - Drive innovations for sustainable production processes
- **Exploit the full potential of promising marketplaces**
  - Intensify activities regarding women's apparel market and Chinese market, e.g. by sponsorships or co-operations
  - Adjust products according to respective preferences
- **Broaden product portfolio**
  - Factor in cultural backgrounds of buyers with special regards to women in the Arabic- Islamic world

  - Design products to suit the different types of veiling (material, style and product range)

10. Balanced Scorecard

	Area of Objectives	Measure	Time Expectation	Accountable
CUSTOMERS				
1.	Customer satisfaction	Customer and online surveys	Quarterly	Managers/ Marketing
2.	Customer Loyalty	Product and purchasing reviews. Memberships and number of returning and new customers.	Quarterly	Marketing
3.	Accessibility	Open more stores in various countries	Biannually-Annually	Marketing
EMPLOYEES				
1.	Improve working conditions	Increase in productivity, employee surveys	Quarterly	CEO
2.	Improve employee training	Increase in productivity and overall operating efficiency	Quarterly	Human Resources
SOCIAL RESPONSIBILITY				
1.	Business Ethics	Endorse positive role model athlete's. Increase	Annually	CEO

Scanned with CamScanner

		promotion of sports and wellness		
2.	Environmentally Friendly	Recycle materials, improve reputation and customer perspective.	Biannually	CEO
6.	Community involvement	Run local sports camps, community/ city events- increase customer awareness	Biannually	Regional Managers
OPERATIONS				
1.	Improve Brand Image	Increase in sales and customer recommendations.	Quarterly	CEO
2.	Product Innovation	Number of new stores, products and marketing		CEO/ Marketing
3.	Market Penetration	Number of stores and sales in new/ other countries	Annually	Marketing
FINANCIAL				
1.	Reduce Cost of production	Decrease in production expenses	Annually	CFO
2.	Increase Revenue	Increase in sales annual	Annually	CFO

Solved By: Anish Bhusal, Avishek Shrestha, Ramesh Pathak and Saramsha Dotel (072-BCT)

Solution of Case Study 2:

thawte

Healthcare Portal Secures Medical Data with Thawte SSL Certificates to Build Trust with Patients and Providers

Scanned with CamScanner

Healthcare Portal Secures Medical Data with Thawte SSL Certificates to Build Trust with Patients and Providers

Health nexxus is a web-based health information and communications portal that enhances interactions between individuals and healthcare providers. Launched in 2004, Health nexxus is owned and operated by Gateway File Systems Inc., a leading Canadian healthcare information systems research and development company. In addition to working with small medical practices and individual providers around the world, Gateway File Systems has partnered with major organizations including Kaiser Permanente, the British Columbia Organ Donor Registry, and Sutter Health located in Northern California.

Challenge

PROTECT MEDICAL RECORDS WHILE BUILDING TRUST WITH PATIENTS AND DOCTORS

With more than 25 years of experience developing systems to manage Electronic Medical Records (EMRs), Gateway File Systems founder Brian Harper understands the power of improving access to healthcare information. "With electronic records, information from doctors and other healthcare service providers can be collected quickly and consolidated into one place," he says. "That way, doctors can get a comprehensive view of their patients' health and provide better care."

In 2004, Harper and his team decided to push the boundaries of traditional EMR systems by offering online access to medical records to both providers and patients. The result was Health nexxus, one of the world's first web-based healthcare information portals. "We knew that giving patients access to their health information online would empower them to be more involved in their own care, a change that would lead to better communication between doctors and patients, and result in better health outcomes overall," Harper says.

Since personal health information is extremely sensitive, Harper also knew that Health nexxus had to be highly secure. "If doctors and patients didn't trust the Health nexxus system, then no one would feel comfortable using it," he says. "That's why we began looking for an SSL solution that not only offered the highest levels of security, but also provided a clear, unmistakable sign that the portal was protected."

SOLUTION SUMMARY

Industry:

- Healthcare information technology

Key Challenges:

- Secure web-based healthcare information portal and protect highly confidential medical records
- Build trust with both health providers and patients, encouraging greater use of the portal
- Provide users with a visible, easy-to-understand signal that patient data is secure

Solution:

- Thawte® SSL Web Server Certificates with EV

Results:

- Provides safe, convenient online access to vital healthcare information
- Green extended validation bar gives users a clear signal that patient data is secure
- Allows portal to meet key healthcare compliance requirements and regulations

Health® nexxus

"We tested EV with some of our doctors and we got excellent feedback. They said, 'yes, the green bar makes a lot of sense. I'm going to look for that on every site from now on.' Opting for Thawte EV SSL certificates has been a great choice for our business."

—Brian Harper, Owner, Health nexxus

Solution

THE GREEN BAR MAKES ONLINE SECURITY EASY TO UNDERSTAND

When Harper and his team developed Health nexxus, they initially used a proprietary security framework to protect patient records. As use of the system grew, Harper decided to switch to SSL security and began a thorough search to find the right provider. "We explored several SSL solutions, but selected Thawte SSL certificates because the company has a long history of providing strong, reliable security," says Harper. "Not only that, but Thawte is widely recognized around the world, making it an ideal choice to establish trust with our international providers and patients."

The fact that Thawte offers cost-effective Extended Validation (EV) SSL protection also played a key role in Harper's decision to choose Thawte SSL certificates. "The green EV bar was a fundamental attraction for us because users can see it right in their browsers and understand that the portal is safe," Harper says.[1] "We tested EV with some of our doctors and we got excellent feedback. They said, 'yes, the green bar makes a lot of sense. I'm going to look for that on every site from now on.' Opting for Thawte EV SSL certificates has been a great choice for our business."

In addition, Harper and his team discovered that Thawte SSL certificates are easy to deploy, giving Health nexxus the flexibility to scale to meet changing business needs. "In my experience, many enterprise technology solutions can be hard to roll out, but Thawte SSL certificates are easy to implement," says Harper. "We're currently focusing on small practices and individual practitioners, but it's reassuring to know that we can easily accommodate an enterprise deployment if we decide to go in that direction."

1. Extended validation SSL certificates trigger a green address bar in high-security browsers.

Results

IMPROVING CARE WITH CONVENIENT, SECURE ACCESS TO HEALTH INFORMATION

By using Thawte SSL Web Server Certificates with EV to secure Health nexxus, Harper and his team have been able to give patients and providers safe and convenient web-based access to health information, increasing patient engagement and improving care. "We want our users to feel like they can depend on the Health nexxus portal, and Thawte SSL certificates are crucial to providing the security and reliability that they have come to expect," says Harper.

With Thawte SSL certificates, Health nexxus has also been able to comply with key government security requirements, underscoring the company's commitment to protecting patient data. "One of the biggest issues with EMR systems is who has access to medical records, especially in the United States where complying with HIPAA – the Health Insurance Portability and Accountability Act – is critical," Harper says. "With Thawte SSL security, we can help ensure that the entry point to health information, the browser, is secure. That is absolutely essential to our business."

Future

In the future, Harper plans to enlarge the geographic scope of the Health nexxus system and offer more web-based services through the portal. "Healthcare is heavily regulated, and we've invested a lot into assessing the compliance requirements that we face as we expand Health nexxus," he says. "With Thawte, we're confident that we have a partner that we can rely on to help us continue to meet all of our online security criteria as we grow."

To learn more, contact our sales advisors:

- Via phone
  - US toll-free: +1 888 484 2983
  - UK: +44 203 450 5486
  - South Africa: +27 21 819 2800
  - Germany: +49 69 3807 89081
  - France: +33 1 57 32 42 68
- Email sales@thawte.com
- Visit our website at https://www.thawte.com/log-in

Protect your business and translate trust to your customers with high-assurance digital certificates from Thawte, the world's first international specialist in online security. Backed by a 17-year track record of stability and reliability, a proven infrastructure, and world-class customer support, Thawte is the international partner of choice for businesses worldwide.

© 2013 Thawte, Inc. All rights reserved. Thawte, the thawte logo, and other trademarks, service marks, and designs are registered or unregistered trademarks of Thawte, Inc. and its subsidiaries and affiliates in the United States and in foreign countries. All other trademarks are property of their respective owners.

Scanned with CamScanner

thawte

School Gives Parents Secure Web-Based Access to Student Data with Thawte SSL Certificates

School Gives Parents Secure Web-Based Access to Student Data with Thawte SSL Certificates

Located in Hong Kong, the Independent Schools Foundation (ISF) is a charitable organization that is dedicated to providing diverse educational opportunities for primary and secondary school students. In 2003, the ISF established a private, non-profit school – the Independent Schools Foundation Academy – to provide an inquiry-based learning environment offering a Chinese and English bilingual immersion education for students in Foundation Year (K) through grade 12.

Challenge

BUILD TRUST WITH PARENTS SO STUDENTS CAN THRIVE

The Independent Schools Foundation Academy prides itself on its innovative approach to education, providing an environment where students are immersed in both English and Chinese to prepare them for success in today's fast-paced global society.

The ISF Academy also takes an inventive approach to communicating with parents, giving them instant access to student records and important school announcements online. "In 2008, we created a web-based portal where parents can log in to check report cards, get homework assignments, find out about upcoming school holidays, and everything else in between," says Adams Fung, IT Manager at the ISF Academy.

Since the portal houses so much sensitive information, Fung knew that it had to be highly secure. "Accessing student data online is incredibly convenient, but parents needed to understand that their children's information would remain confidential before they felt comfortable using the portal," he says.

In order to instill confidence in its web-based student information system, the ISF Academy needed to find an SSL certificate solution that offered highly visible security backed by a well-known brand name that parents would recognize and trust.

1. Extended validation SSL certificates trigger a green address bar in high-security browsers.

SOLUTION SUMMARY

Industry:

- Education

Key Challenges:

- Secure online student information portal and web-based payment processing system cost effectively
- Build trust with parents, allowing them to feel comfortable accessing sensitive data online
- Offer highly visible security that parents recognize immediately

Solution:

- Thawte® SSL Web Server Certificates with EV

Results:

- Allowed The ISF Academy to provide a secure online information portal and payment system
- Green address bar[1] gives parents a clear signal that data is protected, building trust in the school
- Enabled parents to take an active role in children's education, helping to ensure student success

ISF

Scanned with CamScanner

Solution

THE GREEN BAR GIVES PARENTS IMMEDIATE CONFIDENCE

When his team began building the Academy's student information portal, Fung researched several security solutions but selected Thawte based on his prior experience with the company. "We did our due diligence and looked at other providers, but I knew that Thawte offered excellent name recognition at a great price, and I was confident that the certificates would also be very easy to manage," he says.

Because Fung wanted to provide a highly noticeable cue that the portal was secure, he selected Thawte SSL Web Server Certificates with EV (extended validation). "Not every parent knows what to look for to make sure that a website is protected, but the green bar is impossible to miss," he says. "With Thawte EV, parents can immediately see that the portal is secure, giving them even more confidence that their children's sensitive information is safe."

The ISF Academy also uses Thawte SSL certificates to secure online credit card transactions, allowing parents to pay school application fees quickly and easily. "Protecting parents' financial data is just as important as safeguarding student information, and with Thawte SSL certificates we can easily do both," says Fung. "Thawte SSL certificates are easy to install, so we've been able to move more of our services online and offer parents a fast, secure way to manage their accounts on the web."

Results

CONVENIENT, SECURE ACCESS TO STUDENT DATA – ANYTIME

By using Thawte SSL Web Server Certificates with EV to secure its online information portal, the ISF Academy has been able to give parents convenient online access to vital student data, a service that helps parents stay up to date on their children's homework, academic progress, school announcements, and events. "When parents take an active role in their children's education, students are more likely to succeed," Fung says. "With Thawte, we've been able to do more than just secure information online. We've also been able to build trust and open the lines of communication with parents, creating an environment where students can thrive."

> *"Not every parent knows what to look for to make sure that a website is protected, but the green bar is impossible to miss. With Thawte EV, parents can immediately see that the portal is secure and it gives them even more confidence that their children's sensitive information is safe."*
>
> ***—Adams Fung, IT Manager,***
>
> ***The Independent Schools Foundation Academy***

Future

THE FIRST CHOICE FOR SECURING SITES

As the ISF Academy continues to expand the number of web-based services it offers to parents, faculty, and staff, the organization plans to continue using Thawte to provide strong security. "In the future, if we have any system that needs to be secured, Thawte will be our first choice," says Fung. "We've always delivered world-class education, and with Thawte, we can provide world-class security to match."

To learn more, contact our sales advisors:

- Via phone
  - US toll-free: +1 888 484 2983
  - UK: +44 203 450 5486
  - South Africa: +27 21 819 2800
  - Germany: +49 89 3807 89081
  - France: +33 1 57 32 42 68
- Email sales@thawte.com
- Visit our website at https://www.thawte.com/log-in

Protect your business and translate trust to your customers with high-assurance digital certificates from Thawte, the world's first international specialist in online security. Backed by a 17-year track record of stability and reliability, a proven infrastructure, and world-class customer support, Thawte is the international partner of choice for businesses worldwide.

© 2013 Thawte, Inc. All rights reserved. Thawte, the thawte logo, and other trademarks, service marks, and designs are registered or unregistered trademarks of Thawte, Inc. and its subsidiaries and affiliates in the United States and in foreign countries. All other trademarks are property of their respective owners.

thawte

Leading European Travel Company Uses Thawte SSL Certificates to Build Trust with Customers and Grow Business Online

Scanned with CamScanner

Leading European Travel Company Uses Thawte SSL Certificates to Build Trust with Customers and Grow Business Online

In Europe, TUI is synonymous with travel. The corporation operates a broad range of travel companies, from its highly popular and widespread travel agencies, to its airlines, to its Robinson Resort chain, and more than 260 hotels. These businesses serve 30 million customers in 180 countries worldwide, generating €16 billion in revenue and helping TUI become a household name around the globe.

Challenge

SECURE ONLINE TRANSACTIONS SO CUSTOMERS FEEL SAFE BOOKING TRAVEL ONLINE

TUI prides itself on providing superior customer service across all of its nearly 230 brands, so it's no surprise that ensuring trust in its travel booking sites is one of the company's top priorities. "Many people are accustomed to going into a travel agency and working with an agent to book their flights and hotel stays," says Alexander Bialas, Senior Manager at TUI. "People trust their travel agent, so we need to provide the same level of comfort and safety on our travel websites."

Bialas isn't the only one who thinks that online trust and safety are important. According to EyeForTravel, a source for travel industry news, about 60 percent of Europeans agree that trust is the most important consideration for successful online travel resources, rating it higher than other characteristics including simplicity, friendliness, usefulness, and professionalism.

Given these high stakes, TUI needed a flexible, reliable security solution that would protect online transactions while helping to reassure customers that its web-based travel sites were safe and trustworthy.

SOLUTION SUMMARY

Industry:

- Travel

Key Challenges:

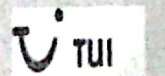
- Protect credit card information and other information from the moment customers submit sensitive data
- Build trust with global customers across more than 120 travel web sites
- Utilize a flexible, reliable online security solution that is respected worldwide

Solution:

- Thawte® SSL Web Server Certificates

Results:

- Established reputation for strong security
- Protects customers on more than 120 travel booking sites
- Experienced substantial growth in online businesses



Solution

A BRAND NAME RECOGNIZED AROUND THE WORLD AND SSL RECOGNIZED BY MORE BROWSERS

In 2003, when the TUI team decided to use SSL certificates to secure its transactional web sites, they researched several providers, but opted for Thawte based on its international reputation for strong, reliable security. "We have customers all over the globe so the fact that Thawte is recognized worldwide was extremely important to us," says Bialas.

Not only that, but TUI also needed a solution that provided native compatibility with a wide range of browsers and applications. "We serve so many customers and our businesses are so diverse that we needed one solution that could be used across a range of software," says Bialas. "With Thawte, we get a solution that offers root ubiquity combined with a wellknown brand name, making it an ideal choice for our web sites."

Results

MORE THAN 120 SECURE TRAVEL WEBSITES AND A GROWING ONLINE BUSINESS

Today, TUI relies on Thawte to secure more than 120 of its sites with SSL certificates. Customers can now book holiday packages, flights, hotels, and car rentals without worry that their credit card and other sensitive information will be compromised. "With Thawte SSL certificates, customers can safely provide payment information and other confidential details," says Bialas. "This makes our travel sites a fast, convenient way to book travel."

With its reputation for security and trust firmly established, TUI now focuses on increasing its web-based business, launching several new sites and increasing volumes at its existing sites.

"With Thawte SSL certificates, customers can safely provide payment information and other confidential details. This makes our travel sites a fast, convenient way to book travel."

—Alexander A. Bialas, Senior Manager, TUI AG

"Overall, our strategy has always been to attract more customers to our websites while maximizing conversions and online revenue," Bialas says. "Not everyone may understand what SSL certificates do, but they have been one of the key building blocks of our online approach."

Future

CONTINUING TO RELY ON THAWTE FOR WORLD-CLASS ONLINE SECURITY

The European online travel market is predicted to be worth €91 billion by 2012. Backed by Thawte, TUI is perfectly positioned to take advantage of that surging demand. Sandra Leonhard, Director Web Strategy & Business Development of TUI Travel plc, predicts the coming years will bring even more profitable growth of TUI Travel's online business. "And as we roll out more travel sites and add capabilities that increase user engagement, trust will continue to be an important factor in our online success," says Leonhard.

Bialas couldn't agree more. "We will continue to rely on Thawte to help ensure the safety of our sites as we continue to expand and enhance our online presence across our travel, hotels and resorts, and cruises sectors," he adds.

To learn more, contact our sales advisors:

- Via phone
  - US toll-free: +1 888 484 2983
  - UK: +44 203 450 5486
  - South Africa: +27 21 819 2800
  - Germany: +49 69 3087 69061
  - France: +33 1 57 32 42 66
- Email sales@thawte.com
- Visit our website at https://www.thawte.com/log-in

Protect your business and translate trust to your customers with high-assurance digital certificates from Thawte, the world's first international specialist in online security. Backed by a 17-year track record of stability and reliability, a proven infrastructure, and world-class customer support, Thawte is the international partner of choice for businesses worldwide.

© 2012 Thawte, Inc. All rights reserved. Thawte, the thawte logo, and other trademarks, service marks, and designs are registered or unregistered trademarks of Thawte, Inc. and its subsidiaries and affiliates in the United States and in foreign countries. All other trademarks are property of their respective owners.

Scanned with CamScanner

REFERENCES

[1] http://indstuds.yolasite.com/resources/02.types_of_is.pdf

[2] Management Information Systems : Managing the Digital Firm http://www.yourarticlelibrary.com/information-technology/process-of-information-system-audit-4-steps/10494

[3] www.ioenotes.edu.np/ioe-syllabus/information-systems-1355

[4] Other references and diagram are taken from various internet source

[5] https://www.guru99.com/difference-information-data.html

[6] Wikipedia entry on "Information Systems," as displayed on August 19, 2012. Wikipedia: The Free Encyclopedia. San Francisco: Wikimedia Foundation. http://en.wikipedia.org/wiki/Information_systems_(discipline).

[7] Excerpted from Information Systems Today - Managing in the Digital World, fourth edition. Prentice-Hall, 2010.

[8] Excerpted from Management Information Systems, twelfth edition, Prentice-Hall, 2012

[9] https://www.geeksforgeeks.org/page-rank-algorithm-implementation/

[10] https://www.esds.co.in/blog/vertical-scaling-horizontal-scaling/#sthash.WnwD7nLw.dpbs

[11] https://sites.google.com/site/assignmentssolved/mca/semester6/mc0088/14

[12] https://www.sisense.com/blog/10-enterprise-software-applications-no-business-without/

[13] https://financesonline.com/erp-software-analysis-features-types-benefits-pricing/#examples

Scanned with CamScanner