UNIT-1

Learning Objective

- Escalating Need for strategic information
- Data warehouse Architecture
- Defining the business requirements

Learning Objective cont..

- Failures of Past decision-support systems,
- Operational versus decision-support systems
- Data warehousing – the only viable solution

Escalating need for Strategic Information

- Our discussion of need for strategic information will not be complete unless, we study the opportunities provided by strategic information and risk facing a company with out such information.
- Who needs strategic information in an enterprise?
- What exactly do we mean by strategic information?

Escalating need for Strategic Information Cont..

The executives and managers who are responsible for keeping the enterprise competitive, need information to make proper decisions. They need information to formulate the business strategies, establish goals, set objectives and monitor results.

Some business Objectives are

- Retain the present customer base
- Increase the customer base by 15% over the next 15 years.

For making such objectives managers needs information for the following purpose:-

- To get in depth knowledge of their company's operations.
- Monitor how the business factor change over time.
- Compare their company's performance relative to competition and to industry bench marks.

Some business Objectives are cont..

Strategic information is not for running the day to day operations of the business. It is not intended to a shipment, a post a withdrawal from a bank account. Strategic information is for more important for the continued health and survival of the corporation.

Some business Objectives are cont..

Characteristics of Strategic Information

Integrated:- Must have a single, enterprise wide view

Data Integrity:- Information must be accurate and must conform to business rule.

Accessible:- Easily accessible with intuitive access path and responsive for analysis.

Credible (believable):-Every business factor must have one and only one value.

Timely:- Information must be available with in the stipulated time frame.

Information Crisis.

If you are working in the IT Dept. of big or small organization. Think of the various computer applications in your company. Think all the data bases and the Quantities of data that support the operation of your company. How many year's worth of customer data is saved and available? How many years' worth of financial data is kept in storage?10years?15 years? Where is all this data ? On one platform? In legacy systems? In Client/server applications.?

Information Crisis cont..

We Faces two facts

- Organizations have lots of data.
- IT systems are NOT effective at turning all the data into useful strategic information.

In our organization we have lot of data , then why our executives and managers uses this data for making strategic decisions?

Information Crisis cont..

Most of the company faces Information crises due to the reason the data is available is NOT readily usable for strategic decision making. Reason for this:- the data of an enterprise is spread across many type of in compatible structures and systems, for proper decision making on over all corporate strategies and objectives we need information integrated from all systems.

Data needed for strategic decision making must be in a format suitable for analyzing trends.

Technology Trends

As for example the name of the computer department in an enterprise went from **"data processing"** to **"MIS"** (Management Information System) then to **"Information Systems"** and more recently to **"Information technology".**

We have seen the explosive Changes n the following areas:

Opportunities and Risks

Examples of the opportunities made available to companies through the use of strategic information:

A community- based pharmacy that competes on a national scale with more than 800 franchised pharmacies coast to coast gains in-depth understanding of what customers buy, resulting in reduced inventory levels, improved effectiveness of promotions and marketing campaigns and improved profitability for the company.

Opportunities and Risks cont..

On the other hand, consider the following cases where risks and threats of failures existed before strategic information was made available for analysis and decision making.

For a large utility company that provided electricity to about 25 million consumers in five mid-Atlantic states in the US, deregulation could result in a few winners and lot of losers. Remaining competitive and perhaps even surviving itself depended on centralizing strategic information from various sources, streamlining data access, and facilitating analysis of the information by the business units.

Failures of Past Decision Support System

If marketing department is concern about performance of the west cost region. The marketing Vice President wants to get some reports from the IT department to analyze the performance over the past two years, Product by Product, and compared to monthly targets. CEO wants to deliver as soon as possible to manager and manager immediately go to the sub ordinate, to give marketing report. There is no report available you have to gather the data from multiple application and start from scratch, then it is required to serve several application which is on different platforms. So, Some times these reports lacks the actual agenda, which causes in consistencies among the data obtained from different applications.

Failures of Past Decision Support System Cont..

It is also possible the person from IT dept. create a report from single application for his/her convenience, so such information may not be helpful in strategic decisions making.

So, from the above scenario we come to know that when information is scattered in different places with forms, it is difficult to use the available information in strategic Decisions.

Operational Vs Decision Support Systems

The fundamental reason for the in ability to provide strategic information is that we have been trying all along to provide strategic information from the operational systems. These operational systems such as order processing, inventory control, claims processing, out patient billing , and so on are not designed or intended to provide strategic information.

Making the wheels of Business Turn

Operational Systems are online transaction Processing (OLTP) systems. These are the systems terms that are used to run the day to day core business of the company. They are so called bread and butter systems. Operational systems make the wheels of business turn by

- **Take an order**
- **Process a claim**
- **Make a shipment**
- **Generate an invoice**
- **Receive cash**
- **Reserve an air line seat**

They support the basic business processes of the company

Watching the Wheels of Business Turn

DSS –Decision Support Systems are not meant to run the core business processes. They are used to watch how the business runs and then make strategic decision to improve the business.

- **Shows the top-selling products.**
- **Shows the problem region.**
- **Shows the highest margins**
- **Alert whenever a district sells below target.**

Data Ware housing- The only viable Solutions

Till now we all came to know that we do need different types of DSS to provide Strategic information. The type of information needed for strategic decision making is different from available from operational systems. We need a new type of system environment for the purpose of providing strategic information for analysis, discerning trends and monitoring performance.

Data Ware housing- The only viable Solutions cont..

Let we see the desirable features and processing requirements of this new type of system environment.

- Data Base designed for analytical tasks.
- Data from multiple applications.
- Easy to use and Conducive to long interactive sessions by users.
- Content updated periodically and stable
- Content to include current and historical data
- Ability for users to run queries and get results online.
- Ability for users to initiative reports.

Processing Requirements in the New Environment

- **Running of Simple queries and report against current and historical data.**
- **Ability to perform "What if " Analysis in many different ways.**
- **Ability to Query, step back, analyze, and then continue to process to any desired length.**
- **Spot historical trends and apply them for future results.**

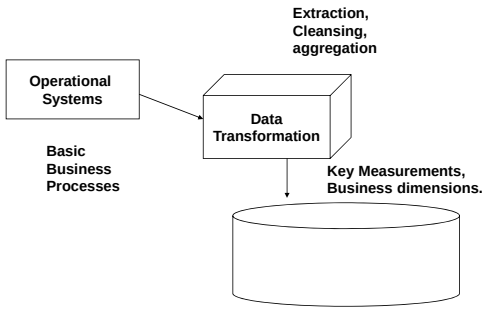
Data Warehouse

Business Intelligence at the data Ware House

This new systems environment that users desperately need to obtains strategic information happens to be new paradigm of data warehousing. Enterprises that are building data ware houses are actually building this new system environment.

This new environment is kept separate from the system environment supporting the day to day operations. The data warehouse holds the business intelligence for the enterprise to enable strategic decision making.

Data Warehouse Cont..



Conclusion

Escalating Need for strategic information

Now, we reached to the conclusion that data ware housing is the only viable solution for providing Strategic information. We arrived at this conclusion based on the functions of the new system environment called the Data Warehouse.

Data Warehouse Architecture

- Defining Features
- Data warehouses and data marts
- Overview of the components
- Metadata in the data warehouse

Defining Features

Key Defining Features of the Data ware house based on these Definitions. What about the nature of the Data in the Data Warehouse? How is this Data Different from the Data in any operational System? Why does it have to be different? How is the Data content in the Data Ware house used?

Defining Features Cont..

Subject-Oriented Data: In operational Systems we store data by individual applications. In the Data sets for an order processing application, we keep the data for that particular application. These data sets provide the Data for all the functions for entering orders, Checking stock, Verifying customer's credit, and assigning the order for shipment.

But in Data Ware house, Data is stored by subjects. Business Subjects differ from organization to organization.

Defining Features Cont..

For a Manufacturing Company Sales, shipments and inventory are critical Business Subjects.

2) Integrated Data: For Proper Decision Making, we need to retrieve all the relevant data from various applications. The data in the Data Ware Houses Comes from Several Operational Systems. Source of storage of Data is different data base, files.

Here are some of the items that would need standardization.

- **Naming Conventions**
- **Codes**
- **Data Attributes**
- **Measurements**

Defining Features Cont..

3)Time-Variant Data: The Data in the Data Warehouse is meant for analysis and Decision making. If a user is looking at the buying pattern of a specific customer, the user needs data not only about the current purchase, but on the past purchases as well when a user wants to find out the reasons for the drop in sales in North Division, the user needs all the sales data for that Division over a period extending back in time.

Defining Features Cont..

The time-variant nature of the Data in a Data Warehouse.

Allows for analysis of the past.

Relates information to the present.

Enables forecasts for the future.

4) Non Volatile Data: We add, change, or delete Data from an operational system as each transaction happens But do not usually update the data in the data ware house, once the data is captured in the data ware house, we do not run individual transactions to change the data there. Data updates are common place and operational Database; not so in data ware house. The Data in a Data ware house is not as volatile as the data in the Operational Databases

Defining Features Cont..

5) Data Granularity: Data Granularity in a Data ware house refers to the level of details.

We find it efficient to keep data summarized at different levels. Depending on the Query, you can then go to the particular level of details and satisfy the query. The lower the level of detail, the finer the data granularity, if we want to keep data in the lowest level of details, we have to store a lot of Data in the Data Warehouse.

Data Ware Houses and Data Marts

If we follow the literature on the data ware houses for the past few years, You would, no doubt, have come across the terms **"Data Ware house"** and **"Data marts"**.

Data Ware Houses and Data Marts Cont..

Many who are new to this paradigm are confused about these terms, some authors and venders use the two terms synonymously. Some make distinctions that are not clear enough. At this point, it would be worth while for us to examine these two terms and take our position.

Should we build a large Data Ware house and then let that repository feed data into local, Departmental Data marts? On the Other hands, should you build individual local data marts, and combine them to form our over all data Warehouse?

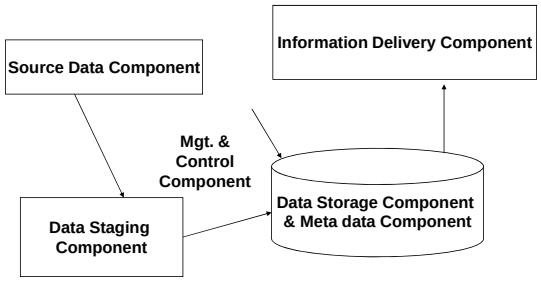
Data Ware Houses and Data Marts Cont..

Data Ware House	Data Mart
Enterprise-wide	Departmental
Union of all Data marts	A Single Business Process.
Data Received from Staging Area	Facts and Dimensions
Structure for corporate view of Data	Technology optimal for data access and analysis.
Organized on E-R model	Structure to Suit the departmental View of data

Overview of The Components

Data Ware house Architecture

Data Ware house Architecture



Data Ware house Architecture cont..

1. Source Data Component: Source data coming into the data warehouse may be grouped into four broad categories, Which are as follows:

Production Data: This category of data comes from various operational systems of the enterprise.

Internal Data: In every organization, user keep their "private" spread sheets, documents, customer profiles and some times even departmental Databases.

Data Ware house Architecture cont..

This data is known as Internal Data which is also useful in a Data ware house.

Archived Data: In operational systems, we periodically take the old data and store it in archived files. The Data in these archived files is referred to as Archived Data.

External Data: In this Category, the data included the data from the external sources. For Example: Market share data of competitors.

Data Ware house Architecture cont..

2) Data Staging Component: When we extracted data from various operational systems and from external source, we have to prepare the data for storing in the data ware house. The Extracted data coming from several disparate sources needs to be changed, converted and made ready in a format that is suitable to be stored for querying and analysis.

Data Ware house Architecture cont..

The 3 major functions need to be performed for getting the data ready.

Data Extraction / Extract the Data: For data extraction we have to employ the appropriate technique to get the suitable data from lot of data received from the operational system for data ware house.

Data Transformation: Data transformation involves many forms of combining pieces of data from the different sources.

This functions ends when we have a collection of integrated data that is cleaned, standardized and summarized. Now, we ready to load data in data warehouse.

Data Ware house Architecture cont..

Data Loading: In this phase movement of large volumes of data using up substantial amount of time.

3)Data Storage Component: The data Storage for the data ware house is a separate repository. The operational systems of our enterprise support the day-to-day operations. The Data repositories of the operational systems typically contain only the current data, while the data repository for a data ware house, we need to keep large volumes of historical data for analysis. We have to keep the data in the data ware house in structures suitable for analysis, and not for quick retrieval of individual pieces of Information.

Data Ware house Architecture cont..

Therefore, the data storage for the data ware house is kept separate from the data storage for operational systems.

4)Informational Delivery Component: To Provide information to the wide community of Data Warehouse users. The information delivery component includes a variety of information delivery. Such as, we may include several information delivery mechanisms, we provide for online queries and reports.

Data Ware house Architecture cont..

The users will enter their requests online and will receive the results online. We may setup delivery of scheduled reports through e-mail or we may make adequate use of our organization intranet for information delivery.

5) Meta Data Component: Meta Data in a Data ware house is similar to the Data dictionary or the Data Catalog in a Data Base Management System, In this Component we keep the information about the logical data Structures, the information about the files and addresses, the information about the indexes.

Data Ware house Architecture cont..

6) Management and Control Component: This component of the data ware house architecture sits on top of all the other components, the mgt. and control component co-ordinates the services and activities with in the data warehouse, on the other hand, it moderates the information delivery to the users. It works with the database mgt. systems and enables data to be properly stored in the repositories. It also monitors the movement of the data into the staging area and from there into the data warehouse storage itself.

Meta Data in the Data Warehouse

The Meta Data component serve as a directory of the contents of our data warehouse.

The meta data in a data warehouse fall in the major categories.

1)Operational Meta Data: Operation meta data gets its data from operational data sources. These sources contains different data structures for storing data from various operational system.

Meta Data in the Data Warehouse cont..

2) Extraction and Transformation Meta Data: Extraction and transformation metadata contains data about the extraction of data from the source system like extraction frequency, extraction methods for data extraction. This also contains the information about all the data transformation that take place in the data staging area.

3) End-User Meta Data: The end-user meta data is the navigational map of the data ware house. It enables the end-users to find information from the data warehouse.

Conclusion

The Data ware house is an informational environment that

- Provides an integrated and total view of the enterprise.
- Makes the enterprise's current and historical information easily available for Decision Making.
- Makes Decision-Support transactions possible with out hindering Operational Systems.
- Renders the Organization's information Consistent.
- Presents a Flexible and interactive Source of Strategic information.

Defining The Business Requirements

- Dimensional analysis
- Information packages
- Requirements gathering methods
- Requirements definition

Dimensional Analysis

Building a data ware house is very different from building an operational system. This become notable especially in the requirements gathering phase. Because of this difference, the traditional methods of collecting requirements that work well for operational systems cannot be applied to data warehouses.

Even though the users cannot fully describe what they want in a data warehouse, they can provide us with very important in sights into how they

Dimensional Analysis cont..

Think about the business. They can tell you what measurement units are important for them. Each user department can let you know how they measure success in that particular department. The user can give you insights into how they combine the various pieces of information for strategic decision making.

Managers think of the business in terms of business dimensions. Various kinds of questions managers are likely to ask for decision making.

Dimensional Analysis cont..

Let us briefly examine these questions.

The marketing Vice President is interested in the revenue generated by her new product, but he/she is not interested in a single number. He/she is interested in the revenue numbers by month, in a certain division, relative to the previous product version, and compared to plan. These are the his/her business dimensions along which he/she wants to analyze his/her numbers.

Dimensional Analysis cont..

Similarly, for the marketing manager, his/her business dimensions are product, product category, time (day, week, month), sale district, and distribution channel.

For the financial controller, the business dimensions are budget line, time (month, quarter, year), district and division. The analysis of sales units along the three business dimensions of product, time and geography. These three dimensions are plotted against three axes of coordinates. We can see that the three dimensions from a collection of cubes. In each of the small dimensional cubes, we will find the sales units for that particular slice of time , product and geographical division. The business data of sales units is three dimensional because there are just 3 dimension used in this analysis.

Information Packages-A New Concept

Information Packages, this concept introduce the idea for determining and recording information requirements for a data ware house.

This new concept enable us to

- Define the common subject areas.
- Design key business metrics.
- Decide how data must be presented

Information Packages-A New Concept cont..

- Determine how users will aggregate or roll up.
- Decide the data quantify for user analysis or query.
- Decide how data will be accessed.
- Establish data granularity
- Estimate data ware house size
- Determine the frequency for data refreshing

Requirements Gathering Methods

- Interviews
- Group Sessions or G.D.
- JAD (Joint Application Development)
- Review the existing documents.

Requirements Definition:

Scope And Content:

Formal documentation is often neglected in Computer System Projects. The Project team goes through the requirements definition Phase. They conduct the interviews and GD . They review the existing Documentation. They gather enough material to support the next phases in the system development life cycle. But they skip the detailed documentation of the requirements definition.

Why requirement Gathering

First of all, the requirements definition document is the basis for the next phases. If Project team members have to leave the project for any reason at all, the project will not suffer from people walking away with the knowledge they have gathered. The formal documentation will also validate our findings when reviewed with the users.

Data Sources

The requirement definition document should include the following information:

- Available Data sources
- Data Structures with in the data sources
- Location of the Data Sources
- Data extraction procedures
- Availability of historical data.

Data Sources cont..

- Data Transformation:
- Data Transformation necessarily involve mapping of source data to the data in the data ware house.
- Data Storage:
- Our requirement definition document must include sufficient details about storage requirement.
- Information Delivery:
- Drill-Down Analysis.
- Roll-Up Analysis

Information Package Diagrams

The information packages diagrams crystallize the information requirements for the data warehouse. They contains the critical matrices measuring the performance of the business units, the business dimensions along which the metrics are analyzed, and the details how drill-down & roll-up analyses are done.

Spend as much time as needed to make sure that the information package diagrams are completes & accurate.

Requirements Definition Document Outline

- Introduction (Purpose and Scope of the Project)
- General Requirements description (Like interview Summary)
- Specific Requirements (Storage requirements)
- Information Package
- Other Requirements (Includes Data Loading Methods)
- User Expectations (How the users expect to use the data ware House)
- User Participation (List of tasks in which users expected to participate through out the development life cycle)
- General Implementation Plan: AT this stage, give a high level plan for implementation.

Conclusion

- Business data is dimensional in nature and the users of the data warehouse think in terms of business dimensions.
- A requirements definition for the data warehouse can, therefore, be based on business dimensions such as product, geography, time and Promotion.
- Scope and Content of the requirements definition document include data sources, data transformation, data storage, information delivery, and information package diagrams.

Summary

Companies are desperate for strategic information to counter fiercer competition, extend market share, and improve profitability.

All the past attempts by IT to provide strategic information have been failures. This was mainly because IT has been trying to provide strategic information from operational systems.

We Need a new type of computing environment to provide strategic information. The data warehouse promises to be this new computing environment.

It is a critical to adapt data warehousing to work with ERP packages, Knowledge management, and CRM.

Review Questions

Objective Questions:

1) A data warehouse is which of the following?

a) Can be updated by end users.
b) Contains numerous naming conventions and formats.
c) Organized around important subject areas.
d) Contains only current data.

2)An operational system is which of the following?

a) A system that is used to run the business in real time and is based on historical data.
b) A system that is used to run the business in real time and is based on current data.
c) A system that is used to support decision making and is based on current data.
d) A system that is used to support decision making and is based on historical data.

Review Questions cont..

3)The generic two-level data warehouse architecture includes which of the following?

a) At least one data mart
b) Data that can extracted from numerous internal and external sources
c) Near real-time updates
d) All of the above.

4)The active data warehouse architecture includes which of the following?

a) At least one data mart
b) Data that can extracted from numerous internal and external sources
c) Near real-time updates
d) All of the above.

Review Questions cont..

5)Reconciled data is which of the following?

a) Data stored in the various operational systems throughout the organization.

b) Current data intended to be the single source for all decision support systems.

c) Data stored in one operational system in the organization.

d) Data that has been selected and formatted for end-user support applications.

6)Transient data is which of the following?

a) Data in which changes to existing records cause the previous version of the records to be eliminated

b) Data in which changes to existing records do not cause the previous version of the records to be eliminated

c) Data that are never altered or deleted once they have been added

d) Data that are never deleted once they have been added

Review Questions cont..

7)The extract process is which of the following?

a) Capturing all of the data contained in various operational systems

b) Capturing a subset of the data contained in various operational systems

c) Capturing all of the data contained in various decision support systems

d) Capturing a subset of the data contained in various decision support systems

8)Data scrubbing is which of the following?

a) A process to reject data from the data warehouse and to create the necessary indexes

b) A process to load the data in the data warehouse and to create the necessary indexes

c) A process to upgrade the quality of data after it is moved into a data warehouse

d) A process to upgrade the quality of data before it is moved into a data warehouse

Review Questions cont..

9)The load and index is which of the following?

a) A process to reject data from the data warehouse and to create the necessary indexes

b) A process to load the data in the data warehouse and to create the necessary indexes

c) A process to upgrade the quality of data after it is moved into a data warehouse

d) A process to upgrade the quality of data before it is moved into a data warehouse

10)Data transformation includes which of the following?

a) A process to change data from a detailed level to a summary level

b) A process to change data from a summary level to a detailed level

c) Joining data from one source into various sources of data

d) Separating data from one source into various sources of data

Review Questions cont..

Short answer type Questions

Q1. Explain the need of metadata in a data warehouse?
Q2. What do you mean by Strategic Information?
Q3. Differentiate between Data Warehouse and Data Mart?
Q4. What do you mean by a Web-enabled data warehouse?
Q5. Define OLTP?
Q6. What type of Processing take Place in a data warehouse?
Q7. Define ETL routine?
Q8. What data does an information package contain?
Q9. In which situations can JAD methodology be successful for collecting requirements?
Q10. List various data sources that feed the data warehouse?

Review Questions cont..

Long answer type Questions

Q1. Explain Data warehouse Architecture in detail?
Q2. Explain business Dimensions. Why and how can business dimensions be useful for defining requirements for the data warehouse?
Q3. State any three factors that indicate the continued growth in data warehousing. Can you think of some examples?
Q4. Discuss the top-down and bottom up approach of creating a data warehouse?

Review Questions cont..

Q5. For a commercial bank, name five types of strategic objectives and explain each objective in detail.
Q6. What do you mean by Information Packages and also explain the need for information packages.
Q7. A data warehouse is an environment, not a product. Discuss.
Q8. Explain various type of data ware house meta data in detail.

Review Questions cont..

Q9. For an airlines company, how can strategic information increases the number of frequent flyers? Discuss giving specific details.

Q10. Examine the opportunities that can be provided by strategic information for a medical center. Can you explain five such opportunities

Suggested Reading/References

[1]. Paul Raj Poonia, "Fundamentals of Data Warehousing", John Wiley & Sons, 2003.

[2]. Sam Anahony, "Data Warehousing in the real world: A practical guide for building decision support systems", John Wiley, 2004

[3]. W. H. Inmon, "Building the operational data store", 2nd Ed., John Wiley, 1999.

[4]. Kamber and Han, "Data Mining Concepts and Techniques", Hartcourt India P. Ltd.,2001