
UNIT 1
DATA WAREHOUSING
AND ITS APPLICATIONS

Syllabus

- **Introduction to data warehouse**
- **Need for Data warehousing**
- **Data warehousing Benefits**
- **Application of data warehouse**
- **Data Warehouse characteristics**, **Architecture** and **Components**
- **Data Mart**
- **Classification of data mart**
- Implementation
- **Gathering the business requirement**
- **Planning and project management**
- **Project principles**
- **Data warehouse readiness assessment,**
- **Project team**
- **Selecting the operating system**
- **Selecting the database software**
- Selecting the tools

Introduction to data warehouse

- A database often contains information or data collection that is generally stored electronically in a computer system.
- It is easy to access, manage, modify, update, monitor, and organize the data. Data is stored in the tables of the Database.
- The process of consolidating data and analyzing it to obtain some insights has been around for centuries, but we just recently began referring to this as data Warehousing.
- Data Warehouse is used to collect and manage data from various sources, in order to provide meaningful business insights.
- A data warehouse is usually used for linking and analyzing heterogeneous sources of business data.
- The data warehouse is the center of the data collection and reporting framework developed for the BI system.
- Data warehouse systems are real-time repositories of information, which are likely to be tied to specific applications.
- Data warehouses gather data from multiple sources (including databases), with an emphasis on storing, filtering, retrieving and in particular, analyzing huge quantities of organized data.
- The data warehouse operates in information-rich environment that provides an overview of the company, makes the current and historical data of the company available for decisions, enables decision support transactions without obstructing operating systems, makes information consistent for the organization, and presents a flexible and interactive information source.

Need for Data warehousing

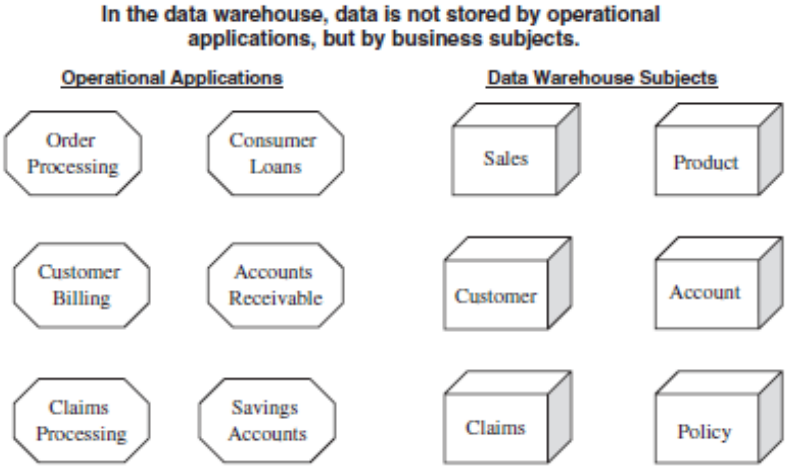
- Data warehouse is used in the largest and most complex businesses around the world.
- In demanding situations, good decision making becomes critical.
- Significant and relevant data is required to make decisions.
- This is possible only with the help of a well-designed data warehouse.
- **Enhancing the turnaround time for analysis and reporting:** Data warehouse allows business users to access critical data from a single source enabling them to take quick decisions. They need not waste time retrieving data from multiple sources. The business executives can query the data themselves with minimal or no support from IT which in turn saves money and time.
- **Improved Business Intelligence:** Data warehouse helps in achieving the vision for the managers and business executives. Outcomes that affect the strategy and procedures of an organization will be based on reliable facts and supported with evidence and organizational data.
- **Benefit of historical data:** Transactional data stores data on a day to day basis or for a very short period of duration without the inclusion of historical data. In comparison, a data warehouse stores large amounts of historical data which enables the business to include time-period analysis, trend analysis, and trend forecasts.
- **Standardization of data:** The data from heterogeneous sources are available in a single format in a data warehouse. This simplifies the readability and accessibility of data. For example, gender is denoted as Male/ Female in Source 1 and m/f in Source 2 but in a data warehouse the gender is stored in a format which is common across all the businesses i.e. M/F.
- **Immense ROI (Return On Investment):** Return On Investment refers to the additional revenues or reduces expenses a business will be able to realize from any project.

Benefits to data warehouse

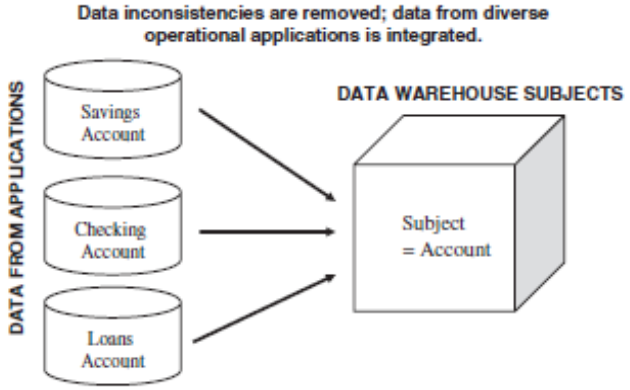
- **Scalability** - Businesses today cannot survive for long if they cannot easily expand and scale to match the increase in the volume of daily transactions DW is easy to scale, making it easier for the business to stride ahead with minimum hassle.
- **Access to Historical Insights** - Though real-time data is important, historical insights cannot be ignored when tracing patterns. Data warehousing allows businesses to access past data with just a few clicks. Data that are months and years old can be stored in the warehouse.
- **Works On-Premises and on Cloud** - Data warehouses can be built on-premises or on cloud platforms. Enterprises can choose either option, depending on their existing business system and the long-term plan. Some businesses rely on both.
- **Better Efficiency -** Data warehousing increases the efficiency of the business by collecting data from multiple sources and processing it to provide reliable and actionable insights. The top management uses these insights to make better and faster decisions, resulting in more productivity and improved performance.
- **Improved Data Security -** Data security is crucial in every enterprise. By collecting data in a centralized warehouse, it becomes easier to set up a multilevel security system to prevent the data from being misused. Provide restricted access to data based on the roles and responsibilities of the employees.
- **Increase Revenue and Returns -** When the management and employees have access to valuable data analytics, their decisions and actions will strengthen the business. This increases the revenue in the long run.
- **Faster and Accurate Data Analytics -** When data is available in the central data warehouse, it takes less time to perform data analysis and generate reports. Since the data is already cleaned and formatted, the results will be more accurate.

Characteristics to data warehouse

- **Subject-oriented**

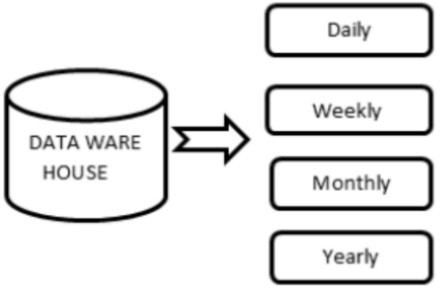


- **Integrated**

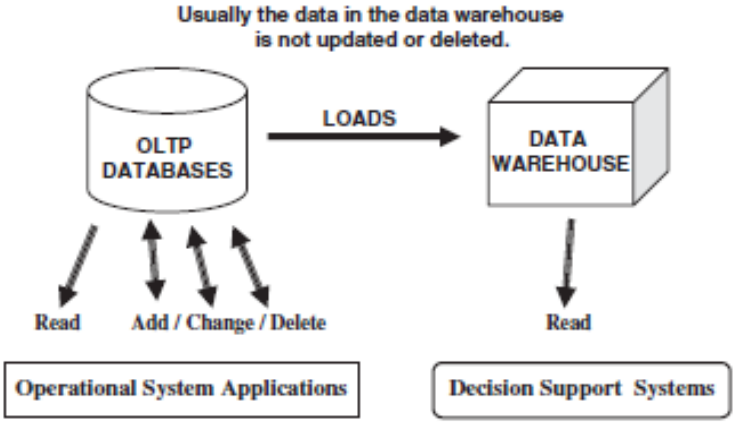


Characteristics to data warehouse

- **Time-Variant**



- **Non-Volatile**



Characteristics to data warehouse

Characteristic	Description	Example
Subject-oriented	Focuses on specific themes (subjects) for decision-making, handling the data warehousing process with well-defined themes. Eliminates unnecessary information for accurate demonstrations.	Sales, Products, Customers, Account
Integrated	Involves consolidating data from various sources into a common system, ensuring consistency, readability, and coding standards. Data is shared across multiple database repositories.	Integration of data from various sources into a relational database.
Time-Variant	Stores information at different intervals (weekly, monthly, yearly), providing a historical perspective. Data is non-modifiable once stored.	Data stored with timestamps at weekly, monthly, or yearly intervals.
Non-Volatile	Data is permanent and not erased or removed when new data is added. Maintains a substantial amount of data and analyzes it within warehouse technologies.	Data warehouse remains separate and non-volatile, not reflecting regular changes in operational databases.

Data warehouse applications

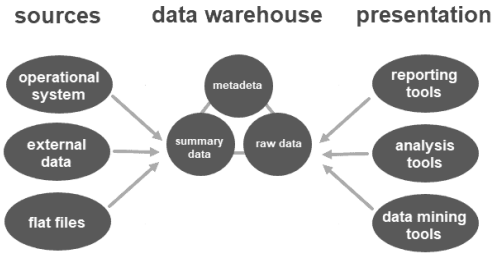
- **Investment and Insurance:** In this sector, data warehousing is used to analyze the customer, market trends and other patterns of data. The two sub-sectors where data warehousing plays an important role are Forex and stock markets.
- **Healthcare**: A data warehousing system is used to forecast outcomes of a treatment generate its reports and share the data with different units. These units can be the research labs, medical units, and insurance providers. Enterprise data warehouses serve as the backbone of healthcare systems as they are updated with recent information which is crucial for saving lives.
- **Retail**: Be it distribution, marketing, examining pricing policies, keeping a track of promotional deals, and finding the pattern in the customer buying trends: data warehousing solves it all. Many retail chains incorporate enterprise data warehousing for business intelligence and forecasting.
- **Social Media Websites**: Social networking sites such as Facebook, Twitter, LinkedIn etc. are based on large data sets analyses. These sites collect data on members, groups; locations etc. and store this information in a single central repository. Data warehouse is necessary to implement the same data, because of its high volume of data.
- **Banking**: Most banks are now using warehouses to see account/cardholder spending patterns. They use this to make special offers, deals, etc. available.
- **Government**: In addition to store and analyze taxes used to detect tax theft, government uses the data warehouse.
- **Airlines**: It is used in the airline system for operational purposes such as crew assignments, road profitability analyses, flight frequency programs promotions, etc.
- **Public sector**: Information is collected in the public sector's data warehouse. It helps government agencies and departments manage their data and records.

Data warehouse Architecture

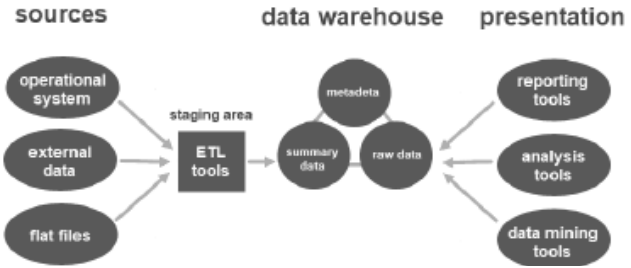
- Data warehouse architecture is a data storage framework's design of an organization.
- It takes information from raw data sets and stores it in a structured and easily digestible format.
- A data warehouse architecture plays a vital role in the data enterprise.
- As databases assist in storing and processing data, and data warehouses help in analysing that data.
- Data warehousing is a process of storing a large amount of data by a business or organization.
- The data warehouse is designed to perform large complex analytical queries on large multi-dimensional datasets in a straightforward manner.
- Data warehouses extract data from different resources, which are in different fonts, convert it into a unique form, and place data in Data Warehouse.
- Data warehouse architecture defines the arrangement of the data in different databases.

Types of Data warehouse Architecture

- Single-tier



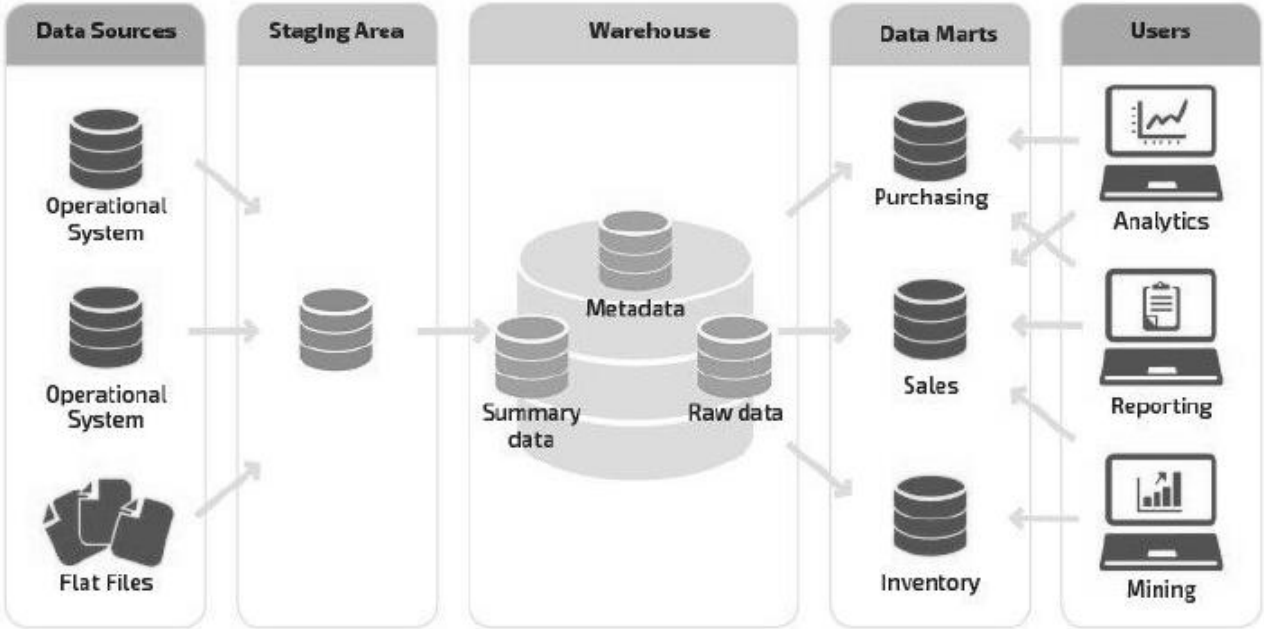
Two-tier



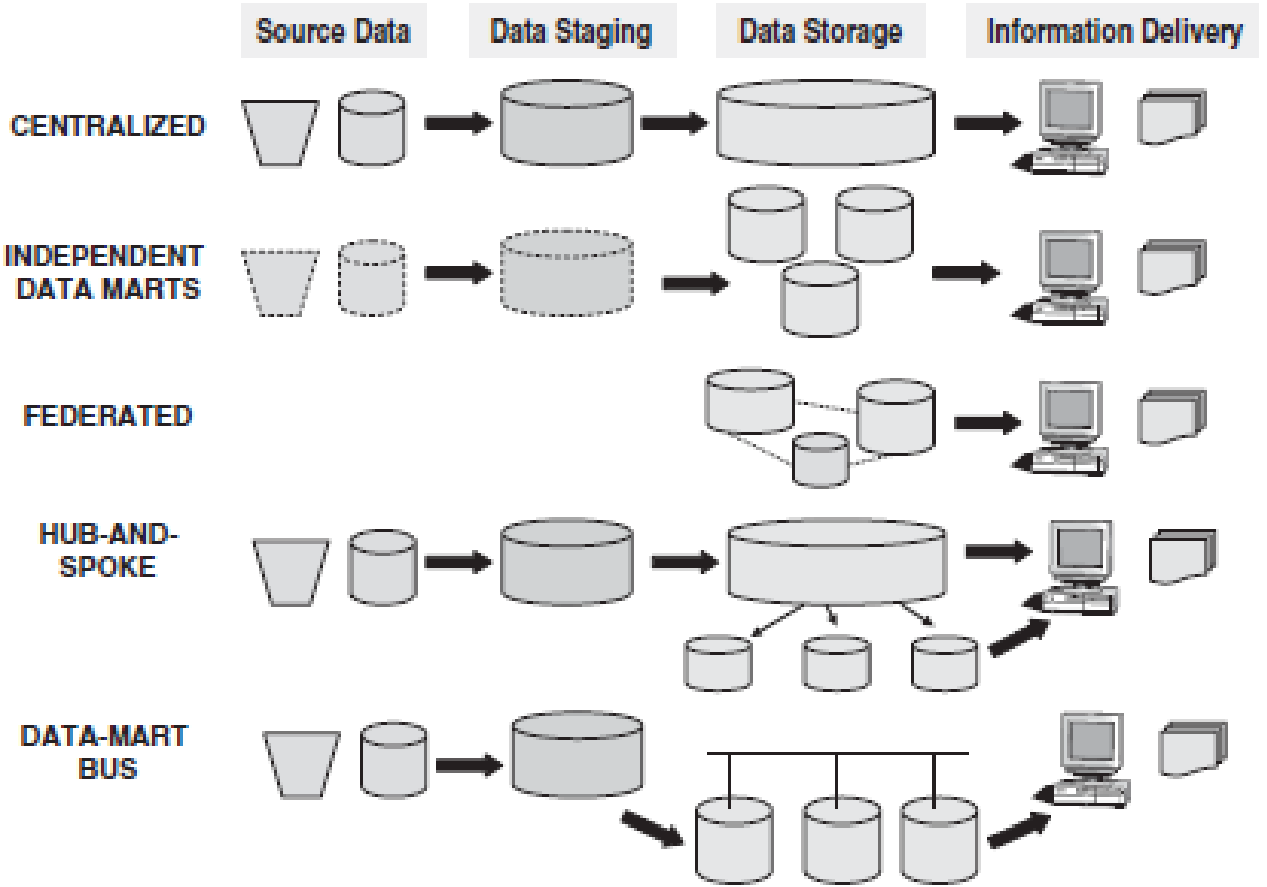
- Three-tier data

Types of Data warehouse Architecture

- Three-tier data



Types of Data warehouse Architecture



Types of Data warehouse Architecture

Architecture Type	Description
Single-Tier	- Infrequently practiced approach. - Goal: Minimize data redundancy by storing a minimal amount of data. - Disadvantage: Lacks a component separating analytical and transactional processing.
Two-Tier	- Includes a staging area between data sources and the data warehouse layer. - Ensures all data loaded is cleansed and in the appropriate format.
Three-Tier	- Widely used architecture for data warehouse systems. - Bottom tier: Database of the warehouse where cleansed and transformed data is loaded. - Middle tier: Application layer providing an abstracted view of the database. - Top tier: User access and interaction with the data, representing the front-end client layer. - Utilizes reporting tools, query, analysis, or data mining tools.
Cloud-Based	- Relatively new compared to legacy options. - Accesses data warehouses through the cloud. - Various cloud-based data warehouse options with different architectures. - Provides benefits of integrating, analyzing, and acting on data from different sources. • **Up-front costs:** Traditional on-premises data warehouses involve pricey up-front expenses for different components. Cloud architecture accessed through the cloud eliminates these up-front costs. • **Ongoing costs:** Cloud offers a low, pay-as-you-go model, avoiding upgrade and maintenance costs faced by businesses with on-prem data warehouses. • **Speed:** Cloud-based data warehouse architecture is substantially speedier than on-premises options, partly due to the use of ELT — which is an uncommon process for on-premises counterparts. • **Flexibility:** Cloud data warehouses are designed to account for the variety of formats and structures found in big data. Traditional relational options are designed simply to integrate similarly structured data. • **Scale:** The elastic resources of the cloud make it ideal for the scale required of big datasets. Additionally, cloud-based data warehousing options can also scale down as needed, which is difficult to do with other approaches.

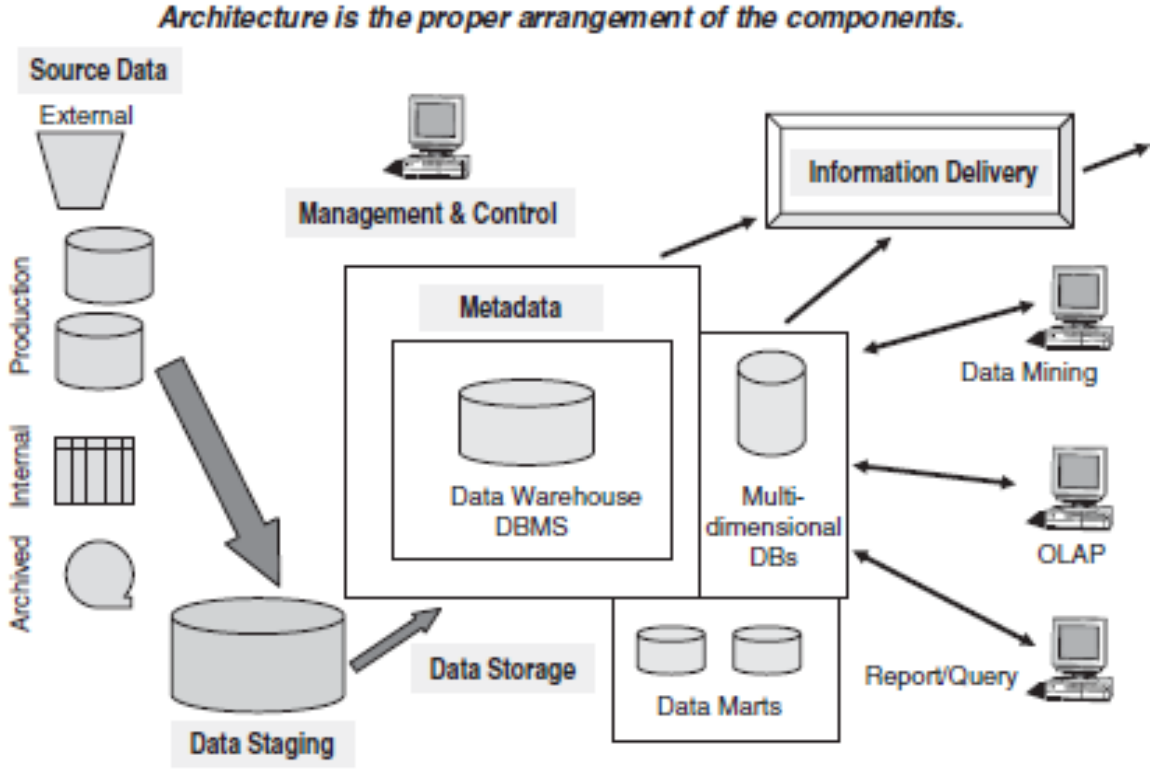
Types of Data warehouse Architecture

Architectural Type	Description
Centralized Data Warehouse	- Considers enterprise-level information requirements. - Establishes an overall infrastructure. - Stores atomic level normalized data in the third normal form, occasionally including summarized data. - Queries and applications access normalized data in the central data warehouse. - No separate data marts exist.
Independent data Marts	- Evolves in companies where organizational units develop their own data marts for specific purposes. - Each data mart serves a particular organizational unit. - Data marts are independent and do not provide a single version of the truth. - Likely to have inconsistent data definitions and standards across different data marts, hindering data analysis across them. - Variances in data make analysis across related subjects difficult.
Federated	- Suited for companies with existing decision-support structures like operational systems, extracted datasets, and primitive data marts. - Integrates data physically or logically through shared key fields, global metadata, distributed queries, etc. - No single overall data warehouse.
Hub-and-Spoke	- Follows the Inmon Corporate Information Factory approach. - Includes an enterprise-wide data warehouse with atomic data in the third normal form. - Features dependent data marts obtaining data from the centralized data warehouse. - Centralized data warehouse serves as the hub, feeding data to data marts on the spokes. - Dependent data marts serve various purposes: departmental analytical needs, specialized queries, data mining, etc. - Each dependent data mart may have different data structures based on individual requirements. - Queries are directed mostly to dependent data marts, although the centralized data warehouse may also be queried.

Types of Data warehouse Architecture

Architectural Type	Description
Data-Mart Bus	- Follows the Kimball conformed supermarts approach. - Begins with analyzing requirements for a specific business subject. - Builds the first data mart (supermart) using business dimensions and metrics. - Business dimensions are shared among future data marts. - By conforming dimensions among data marts, logically integrated supermarts provide an enterprise view of the data. - Data marts contain atomic data organized as a dimensional data model. - Result of adopting an enhanced bottom-up approach to data warehouse development.

Components of Data warehouse Architecture



Components of Data warehouse Architecture

Source Data component

Category	Brief Content
Production Data	- Originates from various operational systems of the enterprise, including financial, manufacturing, supply chain, and customer relationship management systems. - Data selection based on data warehouse information requirements. - Encounters data format variations and resides on different hardware platforms. - Queries in operational systems are narrow and predictable, focusing on specific instances of business objects. - Lack of data conformance among operational systems leads to disparities. - Challenge is to standardize, transform, and integrate disparate data for storage in the data warehouse.
Internal Data	- Consists of users' "private" spreadsheets, documents, customer profiles, and departmental databases within the organization. - Detailed customer profiles important for one-to-one business with significant customer contributions. - This is held in private files cannot be ignored; collective judgment call needed on inclusion in the data warehouse. - Adds complexity to data transformation and integration processes before storage in the data warehouse.
Archived Data	- Periodically archived from operational systems based on organizational circumstances. - Frequency and portions of data archived determined by organizational needs; archived data often from old legacy systems. - Various archiving methods exist, including staged archival methods involving separate databases and tape cartridges. - Historical data essential for analysis over time, useful for discerning patterns and analyzing trends.
External Data	- Executives rely on external sources for a high percentage of the information they use. - This includes industry statistics, market share data, and standard financial indicators. - It aids in spotting industry trends and comparing performance against other organizations. - Usually, external data does not conform to internal formats, requiring conversion methods and organization of data transmissions. - Accommodating variations in data transmission methods is necessary.

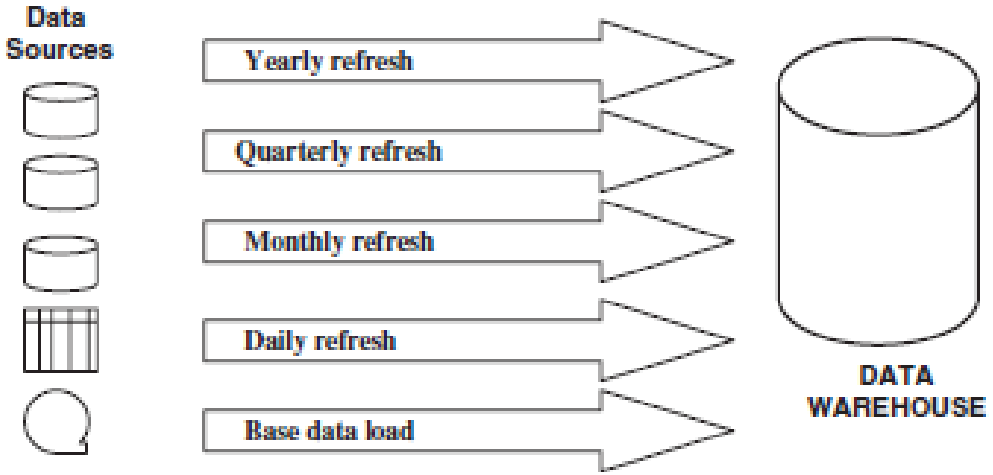
Components of Data warehouse Architecture

Data Staging Component

- After extracted data from various operational systems and from external sources, prepare the data for storing in the data warehouse.
- The extracted data coming from several sources needs to be changed, converted, and made ready in a format that is suitable to be stored for querying and analysis.
- Three major functions(extract, transform, and load) need to be performed in a staging area for getting the data ready.
- Data staging provides a place and an area with a set of functions to clean, change, combine, convert, deduplicate, and prepare source data for storage and use in the data warehouse.
- In a data warehouse data is subject-oriented and is pulled from many source operational systems
- The literal staging area would be literal copies of source system content but in a more convenient environment (such as an ODBC accessible relational database).
- This literal staging area then acts as a surrogate for the source systems.

Components of Data warehouse Architecture

Data Staging Component



Components of Data warehouse Architecture

Data Staging Component

Data Extraction:

Data Sources	Relational databases, legacy networks, hierarchical models, flat files, spreadsheets, local datasets
Techniques	Employing appropriate tools, in-house programs
Complexity	May become complex due to diverse formats and sources

Data Transformation:

Tasks	Cleaning, standardization, conflict resolution, combining data, purging unnecessary data, summarization
Challenges	Dealing with synonyms, homonyms, and disparate sources; ongoing adaptation to changes from source systems
Standardization	Ensuring consistent data types, field lengths, and semantics across sources

Data Loading:

Initial Loading	Moving large volumes of data into warehouse storage during construction
Ongoing Loading	Continuously extracting changes from source data, transforming, and feeding incremental revisions
Time Consumption	Initial loading consumes substantial time; ongoing loading is continual

Components of Data warehouse Architecture

Data storage Component

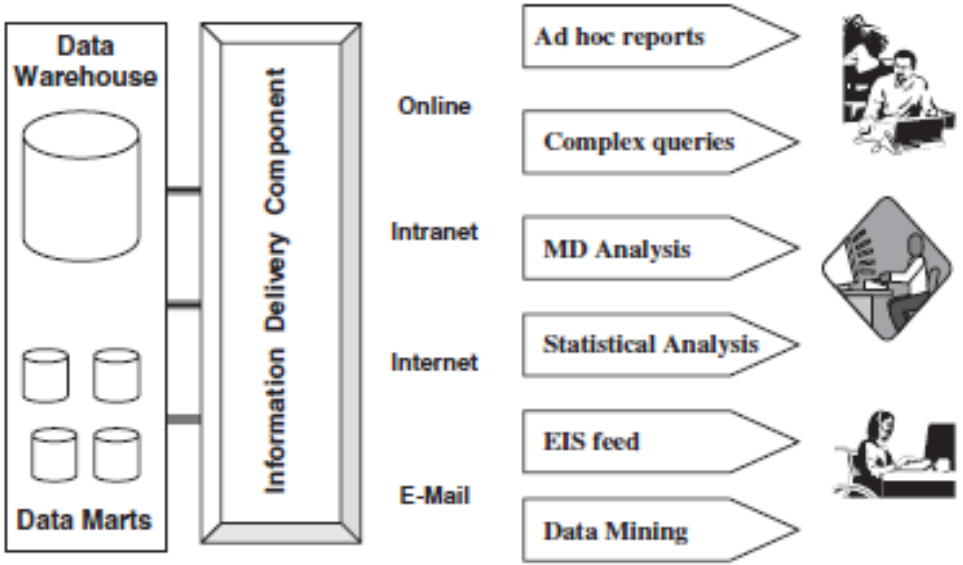
- The data warehouse uses a storage repository, separate from the operational systems handling day-to-day operations, which primarily use current data in online transaction processing applications.
- Also, these data repositories contain the data structured in highly normalized formats for fast and efficient processing.
- In databases supporting operational systems, the updates to data happen as transactions occur.
- These transactions hit the databases in a random fashion.
- When analysts use the data in the data warehouse for analysis, the data is stable and that it represents snapshots at specified periods.
- The data warehouses are “read-only” data repositories.
- Generally, the database in datawarehouse must be open.
- Data warehouses employ relational database management systems and also employ multidimensional database management systems.
- Data extracted from the datawarehouse storage is aggregated in many ways and the summary data is kept in the multidimensional databases (MDDBs).
- Such multidimensional database systems are usually proprietary products.

Components of Data warehouse Architecture

Information Delivery Component

- To provide information to the wide community of Datawarehouse users, the information delivery component includes different methods of information delivery.
- Ad hoc reports are predefined reports primarily meant for novice and casual users.
- Provision for complex queries, multidimensional (MD) analysis, and statistical analysis cater to the needs of the business analysts and power users.
- Information fed into executive information systems (EIS) is meant for senior executives and high-level managers.
- Some data warehouses also provide data to data-mining applications.
- Data-mining applications are knowledge discovery systems where the mining algorithms help you discover trends and patterns from the usage of your data.
- Most commonly, you provide for online queries and reports.
- The users will enter their requests online and will receive the results online.
- You may set up delivery of scheduled reports through e-mail or you may make adequate use of your organization's intranet for information delivery.
- Recently, information delivery over the Internet has been gaining ground.

Components of Data warehouse Architecture



Components of Data warehouse Architecture

Metadata Component

- Metadata is like the data dictionary or the data catalog in a database management system.
- In the data dictionary, information about the logical data structures, files and addresses, indexes are maintained
- The data dictionary contains data about the data in the database.

Management and Control Component

- This coordinates the services and activities within the data warehouse.
- This component controls the data transformation and the data transfer into the data warehouse storage.
- It moderates the information delivery to the users.
- It works with the database management systems and enables data to be properly stored in the repositories.
- It monitors the movement of data into the staging area and from there into the data warehouse storage itself.
- This interacts with metadata component to perform management and control functions.
- As the metadata component contains information about the data warehouse itself, the metadata is the source of information for the management module.

Data marts

- A data mart is a subset of a data warehouse focused on a particular line of business, department, or subject area.
- Data marts make specific data available to a defined group of users, which allows those users to quickly access critical insights without wasting time searching through an entire data warehouse.
- Primary purpose of a data mart is to isolate—or partition—a smaller set of data from a whole to provide easier data access for the end consumers.
- Data mart is for a specific company department and normally a subset of an enterprise-wide data warehouse.
- Data marts improve query speed with a smaller, more specialized set of data.

Classification of data mart

- Dependent data marts are partitioned segments within an enterprise data warehouse. This top-down approach begins with the storage of all business data in one central location. The newly created data marts extract a defined subset of the primary data whenever required for analysis.
- Independent data marts act as a standalone system that doesn't rely on a data warehouse. Analysts can extract data on a particular subject or business process from internal or external data sources, process it, and then store it in a data mart repository until the team needs it.
- Hybrid data marts combine data from existing data warehouses and other operational sources. This unified approach leverages the speed and user-friendly interface of a top-down approach and also offers the enterprise-level integration of the independent method.

Planning and project management

Data warehouse readiness assessment

The readiness assessment report is serving the following purposes:

† Lower the risks of big surprises occurring during implementation

† Provide a proactive approach to problem resolution

† Reassess corporate commitment

† Review and reidentify project scope and size

† Identify critical success factors

† Restate user expectations

† Ascertain training needs

Gathering the business requirement

The users of the data warehouse as follows:

- **Senior executives (including the sponsors)-**Executives will give you a sense of direction and scope for your data warehouse. They are the ones closely involved in the focused area.
- **Key departmental managers-** report to the executives in focus.
- **Business analysts-**prepare reports and analyses for the executives and managers.
- **Operational system database administrators (DBAs)-** give information about the data sources for the warehouse.

Requirements need to gather:

- Data elements: fact classes, dimensions
- Recording of data in terms of time
- Data extracts from source systems
- Business rules: attributes, ranges, domains, operational records

3 techniques adopted for obtaining information from groups of people:

(1) interviews, one-on-one or in small groups;
(2) joint application development (JAD) sessions;
(3) questionnaires.

Gathering the business requirement

Interviews

† Two or three persons at a time
† Easy to schedule
† Good approach when details are intricate
† Some users are comfortable only with one-on-one interviews
† Need good preparation to be effective
† Always conduct pre-interview research
† Establish objectives for each interview
† Decide on the question types
† Also encourage users to prepare for the interview

Group Sessions

† Groups of 20 or fewer persons at a time
† Use only after getting a baseline understanding of the requirements
† Not good for initial data gathering
† Use when free flow of ideas is essential
† Useful for confirming requirements
† Efficient when users are scattered across locations
† Need to be very well organized

Questionnaires

† Can gather lots of requirements quickly
† Useful when people to be questioned are widely dispersed
† Good in exploration phase to get overall reactions
† May be used for people whose work schedule is too tight for interviews
† However, questionnaires do not permit interactive responses like interviews

Gathering the business requirement

Types of Questions

Open-Ended Questions.

- Benefits: They put interviewees at ease, allow insights into values and beliefs, provides exposure to interviewees' vocabulary, opens up opportunities for more questioning, and are interesting and spontaneous.
- Drawbacks are that they could result in too much unnecessary detail, the risk of losing control in the interview, they may take too much time, not proportional to the information gathered.

Closed Questions.

- These allow limited responses to interviewees.
- Some closed questions are bipolar in the sense that these look for "Yes or No" type answers.
- Closed questions enable you to save time and get to the point quickly and easily.
- Closed questions allow for interviews to be compared, provide control over the interview and the ability to cover a lot of ground quickly, and are likely to gather only the relevant information.
- Drawbacks are the inability to get rich details, less chance for building trust and rapport between interviewer and interviewee, and they may become boring and dull.

Probes:

- These are really follow-up questions.
- Probes may be used after open-ended or closed questions.
- The intention would be to go beyond the initial questions and answers.
- Probes are useful in drawing out an interviewee's point of view.

Gathering the business requirement

Arrangement of Questions

Pyramid Structure.

- This is an inductive method of arranging the questions.
- You begin with very specific closed questions and then expand the topics with open-ended questions.
- This structure is useful when the interviewee needs to warm up to the topics being discussed.
- Use this structure when general views of the topics are to be extracted at the end.

Funnel Structure.

- This is a deductive method.
- Begin with general open-ended questions and then narrow the topics with specific, closed questions.
- This structure is useful when the interviewee is emotional about the topics under discussion.
- Use this structure when gradual levels of details are needed at the end.

Diamond-Shaped Structure.

- In this case, you warm up the interview with specific closed questions.
- You then proceed towards broad, general, open-ended questions.
- Finally you narrow the interview and achieve closure with specific closed questions.
- Usually, this structure is better than the other two.
- However, this structure may lengthen the interview.

Gathering the business requirement

Interview Techniques

† Select and train the project team members conducting the interviews
† Assign specific roles for each team member (lead interviewer/scribe)
† Prepare a list of users to be interviewed and prepare a broad schedule
† List your expectations from each set of interviews
† Complete pre-interview research
† Prepare interview questionnaires
† Prepare the users for the interviews
† Conduct a kick-off meeting of all users to be interviewed

Pre-interview research is important for the success of the interviews.

† History and current structure of the business unit
† Number of employees and their roles and responsibilities
† Locations of the users
† Primary purpose of the business unit in the enterprise
† Relationship of the business unit to the strategic initiatives of the enterprise
† Secondary purposes of the business unit
† Relationship of the business unit to other units and to outside organizations
† Contribution of the business unit to corporate revenues and costs
† The company's market
† Competition in the market

Gathering the business requirement

Adapting the JAD Methodology

- If you can gather a lot of baseline data up front from different sources, group sessions may be a good substitute for individual interviews.
- In this method, you can get several interested users to meet in group sessions.
- Overall, this method could result in fewer group sessions than individual interview sessions.
- The overall time for requirements gathering may prove to be less and, therefore, shorten the project.
- Also, group sessions may be more effective if the users are dispersed in remote locations.
- Joint application development (JAD) techniques were successfully utilized to gather requirements for operational systems in the 1980s.
- Users of computer systems had grown to be more computer-savvy and their direct participation in the development of applications proved to be very useful.
- As the name implies, JAD is a joint process, with all the concerned groups getting together for a well-defined purpose.
- It is a methodology for developing computer applications jointly by the users and the IT professionals in a well-structured manner.
- JAD centers around discussion workshops lasting a certain number of days under the direction of a facilitator.
- Under suitable condiions, the JAD approach may be adapted for building a data warehouse.

Gathering the business requirement

JAD consists of a five-phased approach:

Project definition

- Complete high-level interviews
- Conduct management interviews
- Prepare management definition guide

Research

- Become familiar with the business area and systems
- Document user information requirements
- Document business processes
- Gather preliminary information
- Prepare agenda for the sessions

Preparation

- Create working document from previous phase
- Train the scribes
- Prepare visual aids
- Conduct presession meetings
- Set up a venue for the sessions
- Prepare a checklist for objectives

JAD sessions

- Open with a review of the agenda and purpose
- Review assumptions
- Review data requirements
- Review business metrics and dimensions
- Discuss dimension hierarchies and roll-ups
- Resolve all open issues
- Close sessions with lists of action items

Final document

- Convert the working document
- Map the gathered information
- List all data sources
- Identify all business metrics
- List all business dimensions and hierarchies
- Assemble and edit the document
- Conduct review sessions
- Get final approvals
- Establish procedure to change requirements

Gathering the business requirement

Using Questionnaires

Type and Choice of Questions.

- You may use both open-ended and closed questions in a questionnaire.
- Choice of language is important.
- Use the language of the respondents, not cryptic technical jargon.
- Be specific, not vague. Keep questions short and precise.
- Avoid objectionable, politically incorrect language.
- Avoid talking down to the respondents.
- Target the questions to the appropriate respondent group.

Questionnaire Design.

- The order of the questions is important.
- Start the questionnaire with less controversial, highly important questions.
- Cluster questions with similar content.
- The design must be inviting and pleasing.
- Allow ample white space.
- Provide sufficient space for responses.
- Make it easy to mark or indicate responses while using scales.
- Maintain a consistent style.

Gathering the business requirement

Using Questionnaires

Application of Scales.

- Questionnaires usually contain nominal and interval scales.
- These make it easy to respond.
- Nominal scales are used to classify things.
- Interval scales are used for quantitative analysis.

Administering Questionnaires.

- Carefully decide on who gets the questionnaire.
- Ensure that there are no omissions.
- Some ways of administering the questionnaires include at an initial group session, through personal delivery and later collection, self administration by respondents, by mail to respondent locations, and electronically via e-mail.

Project principles

- **Sponsorship**. No data warehouse project succeeds without strong and committed executive sponsorship.
- **Project Manager**. It is a serious mistake to have a project manager who is more technology oriented than user-oriented and business-oriented.
- **New Paradigm**. Data warehousing is new for most companies; innovative project management methods are essential to deal with the unexpected challenges.
- **Team Roles**. Team roles are not to be assigned arbitrarily; the roles must reflect the needs of each individual data warehouse project.
- **Data Quality**. Three critical aspects of data in the data warehouse are: quality, quality,and quality.
- **User Requirements**. Although obvious, user requirements alone form the driving force of every task on the project schedule.
- **Building for Growth**. Number of users and number of queries increase very quickly after deployment; data warehouses not built for growth will crumble swiftly.
- **Project Politics**. The first data warehouse project in a company poses challenges and threats to users at different levels; trying to handle project politics is like walking the proverbial tightrope, to be trodden with extreme caution.
- **Realistic Expectations**. It is easy to promise the world in the first data warehouse project; setting expectations at the right and attainable levels is the best course.
- **Dimensional Data Modeling**. A well-designed dimensional data model is a required foundation and blueprint.
- **External Data**. A data warehouse does not live by internal data alone; data from relevant external sources is a necessary ingredient.
- **Training**. Data warehouse user tools are different and new. If the users do not know how to use the tools, they will not use the data warehouse. An unused data warehouse is a failed data warehouse.

Selecting the operating system

- **Scalability**. Again, scalability is first on the list because this is one common feature of every data warehouse. Data warehouses grow, and they grow very fast. Along with the hardware and database software, the operating system must be able to support the increase in the number of users and applications.
- **Security**. When multiple client workstations access the server, the operating system must be able to protect each client and associated resources. The operating system must provide each client with a secure environment.
- **Reliability**. The operating system must be able to protect the environment from application malfunctions.
- **Availability**. This is a corollary to reliability. The computing environment must continue to be available after abnormal application terminations.
- **Preemptive Multitasking**. The server hardware must be able to balance the allocation of time and resources among the multiple tasks. Also, the operating system must be able to let a higher priority task preempt or interrupt another task as and when needed.
- **Use a Multithreaded Approach**. The operating system must be able to service multiple requests concurrently by distributing threads to multiple processors in a multiprocessor hardware configuration. This feature is very important because multiprocessor configurations are architectures of choice in a data warehouse environment.
- **Memory Protection**. Again, in a data warehouse environment, large numbers of queries are common. That means that multiple queries will be executing concurrently. A memory protection feature in an operating system prevents one task from violating the memory space of another.

Selecting the database software

- Query governor—to anticipate and abort runaway queries
- Query optimizer—to parse and optimize user queries
- Query management—to balance the execution of different types of queries
- Load utility—for high-performance data loading, recovery, and restart
- Metadata management—with an active data catalog or dictionary
- Scalability—in terms of both number of users and data volumes
- Extensibility—having hybrid extensions to OLAP databases
- Portability—across platforms
- Query tool Application Program Interfaces (APIs)—for tools from leading vendors
- Administration—providing support for all DBA functions

Project team

- A data warehouse project is like other software projects in that it is human-intensive.
- It takes several trained and especially skilled persons to form the project team.
- Organizing the project team for a data warehouse project has to do with matching diverse roles with proper skills and levels of experience.

Organizing the Project Team

- Organizing a project team involves putting the right person in the right job.
- If organizing and putting together a team to work on an OLTP system development, you know that the required skills set is of a reasonable size and is manageable.

Executive Sponsor	Data Provision Specialist
Project Director	Business Analyst
Project Manager	System Administrator
User Representative Manager	Data Migration Specialist
Data Warehouse Administrator	Data Grooming Specialist
Organizational Change Manager	Data Mart Leader
Database Administrator	Infrastructure Specialist
Metadata Manager	Power User
Business Requirements Analyst	Training Leader
Data Warehouse Architect	Technical Writer
Data Acquisition Developer	Tools Specialist
Data Access Developer	Vendor Relations Specialist
Data Quality Analyst	Web Master
Data Warehouse Tester	Data Modeler
Maintenance Developer	Security Architect

Project team

Roles and Responsibilities

- Project team roles are designated to perform one or more related tasks. In many data warehouse projects, the team roles are synonymous with the job titles given to the team members.
- In the OLTP system project, you will find the job titles of project manager, business analyst, systems analyst, programmer, data analyst, database administrator, and so on. However, datawarehouse projects are not yet standardized as far as job tiles go.
- There still is an element of experimentation and exploration.
- Data warehousing authors and practitioners tend to classify roles or job titles in various ways.
- They first produce broad classifications and then include individual job titles within these classifications.

† Staffing for initial development, staffing for testing, staffing for ongoing maintenance, staffing for data warehouse management
† IT and end-users, then subclassifications within each of the two broad classifications,followed by further subclassifications
† Front office roles, back office roles
† Coaches, regular lineup, special teams
† Management, development, support
† Administration, data acquisition, data storage, information delivery

Project team

Skills and Experience Levels

Executive Sponsor
Direction, support, arbitration.

Project Manager
Assignments, monitoring, control.

User Liaison Manager
Coordination with user groups.

Lead Architect
Architecture design.

Infrastructure Specialist
Infrastructure design/construction.

Business Analyst
Requirements definition.

Data Modeler
Relational and dimensional modeling.

Data Warehouse Administrator
DBA functions.

Data Transformation Specialist
Data extraction, integration, transformation.

Quality Assurance Analyst
Quality control for warehouse data.

Testing Coordinator
Program, system, tools testing.

End-User Applications Specialist
Confirmation of data meanings/relationships.

Development Programmer
In-house programs and scripts.

Lead Trainer
Coordination of User and Team training.

- To fit into the roles and discharge the responsibilities, the selected persons must have the right abilities.
- They should possess suitable skills and the proper work experience.
- So you must produce a list of skills and experience required for the various roles.
- It is not easy to find IT professionals to fill all the roles established for your data warehouse.
- OLTP systems are ubiquitous.
- All IT professionals have assumed some role or the other in an OLTP system project.
- This is not the case with data warehouse projects.
- Not too many professionals have direct hands-on experience in the development of data warehouses.
- Outstanding skills and abilities are in short supply.

Project team

User Participation

Executive Sponsor
Senior level executive, in-depth knowledge of the business, enthusiasm and ability to moderate and arbitrate as necessary.

Project Manager
People skills, project management experience, business and user oriented, ability to be practical and effective.

User Liaison Manager
People skills, respected in user community, organization skills, team player, knowledge of systems from user viewpoint.

Lead Architect
Analytical skills, ability to see the big picture, expertise in interfaces, knowledge of data warehouse concepts.

Infrastructure Specialist
Specialist in hardware, operating systems, computing platforms, experience as operations staff.

Business Analyst
Analytical skills, ability to interact with users, sufficient industry experience as analyst.

Data Modeler
Expertise in relational and dimensional modeling with case tools, experience as data analyst.

Data Warehouse Administrator
Expert in physical database design and implementation, experience as relational DBA, MDDBMS experience a plus.

Data Transformation Specialist
Knowledge of data structures, in-depth knowledge of source systems, experience as analyst.

Quality Assurance Analyst
Knowledge of data quality techniques, knowledge of source systems data, experience as analyst.

Testing Coordinator
Familiarity with testing methods and standards, use of testing tools, knowledge of some data warehouse information delivery tools, experience as programmer/analyst.

End-User Applications Specialist
In-depth knowledge of source applications.

Development Programmer
Programming and analysis skills, experience as programmer in selected language and DBMS.

Lead Trainer
Training skills, experience in IT/User training, coordination and organization skills.

- In a OLTP application, the users interact with the system through GUI screens.
- They use the screens for data input and for retrieving information.
- The users receive any additional information through reports produced by the system at periodic intervals.
- If the users need special reports, they must get IT involved to write ad hoc programs that are not part of the regular application.

Project team

User Participation

Project Planning

Provide goals, objectives, expectations, business information during preliminary survey; grant active top management support; initiate project as executive sponsor.

Requirements Definition

Actively participate in meetings for defining requirements; identify all source systems; define metrics for measuring business success, and business dimensions for analysis; define information needed from data warehouse.

Design

Review dimensional data model, data extraction and transformation design; provide anticipated usage for database sizing; review architectural design and metadata; participate in tool selection; review information delivery design.

Construction

Actively participate in user acceptance testing; test information delivery tools; validate data extraction and transformation functions; confirm data quality; test usage of metadata; benchmark query functions; test OLAP functions; participate in application documentation.

Deployment

Verify audit trails and confirm initial data load; match deliverables against stated expectations; arrange and participate in user training; provide final acceptance.

Maintenance

Provide input for enhancements; test and accept enhancements.

Few team roles that users can assume to participate in the development:

- Project sponsor-responsible for supporting the project effort all the way (must be an executive)
- User department liaison representatives—help IT to coordinate meetings and review sessions and ensure active participation by the user departments
- Subject area experts—provide guidance in the requirements of the users in specific subject areas and clarify semantic meanings of business terms used in the enterprise
- Data review specialists—review the data models prepared by IT; confirm the data elements and data relationships
- Information delivery consultants-examine & test information delivery tools; assist in tool selection
- User support technicians—act as the first-level, front-line support for the users in their respective departments