Data warehouse

What is a Data Warehouse?

[Barry Devlin]

A single, complete and consistent store of data obtained from a variety of different sources made available to end users in a what they can understand and use in a business context.

Inmon's definition of Data Warehouse: In 1993, the "father of data warehousing", Bill Inmon, gave this definition of a data warehouse as: A data warehouse is subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management's decision making process.

Data Warehouse Usage:-

1. Data warehouses and data marts are used in a wide range of applications.

2. Business executives use the data in data warehouses and data marts to perform data analysis and make strategic decisions.

3. In many areas, data warehouses are used as an integral part for enterprise management.

4. The data warehouse is mainly used for generating reports and answering predefined queries.

5. It is used to analyze summarized and detailed data, where the results are presented in the form of reports and charts.

6. Later, the data warehouse is used for strategic purposes, performing multidimensional analysis and sophisticated operations.

7. Finally, the data warehouse may be employed for knowledge discovery and strategic decision making using data mining tools.

8. In this context, the tools for data warehousing can he categorized into *access and retrieval tools, database reporting tools, data analysis tools, and data mining tools.*

Reasons for data Warehouse:

There are a few reasons why a data warehouse should exist:

a) You want to **integrate data across functions** or systems to provide a complete picture of the data subject e.g. customer orders, customer complaints, salespersons.

b) You **do not want to interfere with the fast performing transaction systems** by running large computer resource queries and reports whilst routine users and possibly customers are executing the essential business transactions.

c) You want to **reorganize the data to support fast reporting** and querying.

d) You want to **clean up the quality of the data** to give consistency and data integrity. Many systems do not have strict input validation and duplicates e.g. same customer entered more than once. Also there often different definitions for the same subject or entity within the business e.g. customer, client, prospect.

Need for Data Warehousing:

a) Industry has huge amount of operational data.

b) It is a platform for consolidated historical data for analysis.

c) It stores data of good quality so that knowledge worker can make correct / strategic decisions.

d) Better business intelligence for end-users.

e) Reduction in time to locate, access, and analyze information.

f) Consolidation of disparate information sources.

g) Strategic advantage over competitors.

h) Faster time-to-market for products and services.

i) Replacement of older, less-responsive decision support systems.

j) Reduction in demand on IS to generate reports.

Applications of Data Warehousing

12 Applications of Data Warehouse: Data Warehouses owing to their potential have deep-rooted applications in every industry which use historical data for prediction, statistical analysis, and decision making. Listed below are the applications of Data warehouses across innumerable industry backgrounds. In this article, we are going to discuss various **applications of data warehouse**.
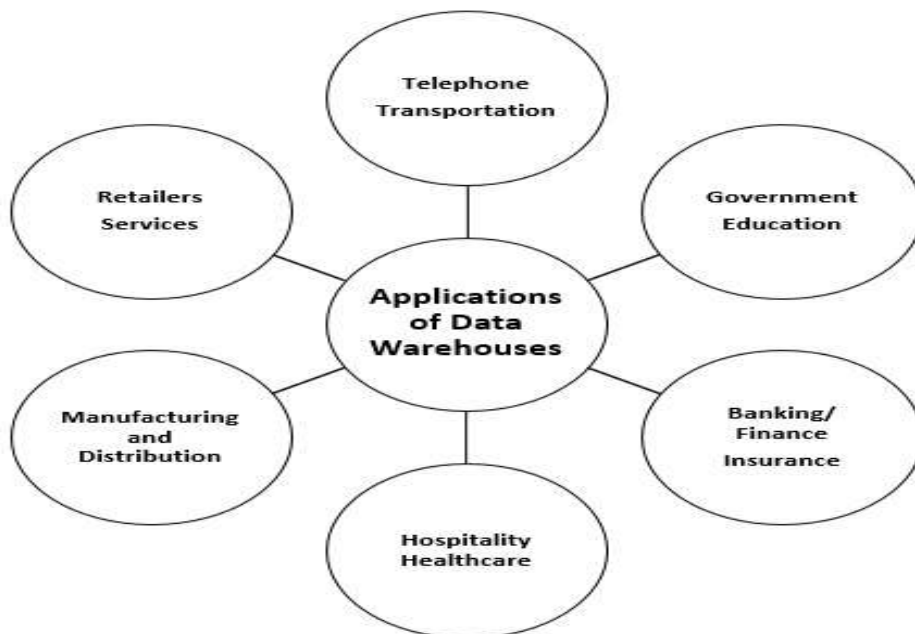
1.   **Banking Industry**

In the banking industry, concentration is given to risk management and policy reversal as well analyzing consumer data, market trends, government regulations and reports, and more importantly financial decision making.

Most banks also use warehouses to manage the resources available on deck in an effective manner. Certain banking sectors utilize them for market research, performance analysis of each product, interchange and exchange rates, and to develop marketing programs.

Analysis of card holder's transactions, spending patterns and merchant classification, all of which provide the bank with an opportunity to introduce special offers and lucrative deals based on cardholder activity. Apart from all these, there is also scope for co-branding.

2.   **Finance Industry**

Similar to the applications seen in banking, mainly revolve around evaluation and trends of customer expenses which aids in maximizing the profits earned by their clients.

3.   **Consumer Goods Industry**

They are used for prediction of consumer trends, inventory management, market and advertising research. In-depth analysis of sales and production is also carried out. Apart from these, information is exchanged business partners and clientele.

4. Government and Education

The federal government utilizes the warehouses for research in compliance, whereas the state government uses it for services related to human resources like recruitment, and accounting like payroll management. The government uses data warehouses to maintain and analyze tax records, health policy records and their respective providers, and also their entire criminal law database is connected to the state's data warehouse. Criminal activity is predicted from the patterns and trends, results of the analysis of historical data associated with past criminals.

Universities use warehouses for extracting of information used for the proposal of research grants, understanding their student demographics, and human resource management. The entire financial department of most universities depends on data warehouses, inclusive of the Financial Aid department.

5. Healthcare

One of the most important sector which utilizes data warehouses is the Healthcare sector. All of their financial, clinical, and employee records are fed to warehouses as it helps them to strategize and predict outcomes, track and analyze their service feedback, generate patient reports, share data with tie-in insurance companies, medical aid services, etc.

6. Hospitality Industry

A major proportion of this industry is dominated by hotel and restaurant services, car rental services, and holiday home services. They utilize warehouse services to design and evaluate their advertising and promotion campaigns where they target customers based on their feedback and travel patterns.

7. Insurance

As the saying goes in the insurance services sector, "Insurance can never be bought, it can be only be sold", the warehouses are primarily used to analyze data patterns and customer trends, apart from maintaining records of already existing participants. The design of tailor-made customer offers and promotions is also possible through warehouses.

8. Manufacturing and Distribution Industry

This industry is one of the most important sources of income for any state. A manufacturing organization has to take several make-or-buy decisions which can influence the future of the sector, which is why they utilize high-end OLAP tools as a part of data warehouses to predict market changes, analyze current business trends, detect warning conditions, view marketing developments, and ultimately take better decisions.

They also use them for product shipment records, records of product portfolios, identify profitable product lines, analyze previous data and customer feedback to evaluate the weaker

product lines and eliminate them. For the distributions, the supply chain management of products operates through data warehouses.

9.   The Retailers

Retailers serve as middlemen between producers and consumers. It is important for them to maintain records of both the parties to ensure their existence in the market. They use warehouses to track items, their advertising promotions, and the consumers buying trends. They also analyze sales to determine fast selling and slow selling product lines and determine their shelf space through a process of elimination.

10.   Services Sector

Data warehouses find themselves to be of use in the service sector for maintenance of financial records, revenue patterns, customer profiling, resource management, and human resources.

11.   Telephone Industry

The telephone industry operates over both offline and online data burdening them with a lot of historical data which has to be consolidated and integrated. Apart from those operations, analysis of fixed assets, analysis of customer's calling patterns for sales representatives to push advertising campaigns, and tracking of customer queries, all require the facilities of a data warehouse.

12.   Transportation Industry

In the transportation industry, data warehouses record customer data enabling traders to experiment with target marketing where the marketing campaigns are designed by keeping customer requirements in mind.

The internal environment of the industry uses them to analyze customer feedback, performance, manage crews on board as well as analyze customer financial reports for pricing strategies.

Data Warehouse

Data warehouse is an information system that contains historical and commutative data from single or multiple sources. It simplifies reporting and analysis process of the organization. It is also a single version of truth for any company for decision making and forecasting.

Characteristics of Data warehouse

A data warehouse has following characteristics:

- Subject-Oriented
- Integrated
- Time-variant
- Non-volatile

(i).   Subject-Oriented

A data warehouse is subject oriented as it offers information regarding a theme instead of companies' ongoing operations. These subjects can be sales, marketing, distributions, etc.

A data warehouse never focuses on the ongoing operations. Instead, it put emphasis on modeling and analysis of data for **decision making**. It also provides a simple and concise view around the specific subject by excluding data which not helpful to support the decision process.

(ii).   Integrated

In Data Warehouse, integration means the establishment of a common unit of measure for all similar data from the dissimilar database. The data also needs to be stored in the Datawarehouse in common and universally acceptable manner.

A data warehouse is developed by integrating data from varied sources like a mainframe, relational databases, flat files, etc. Moreover, it must keep consistent naming conventions, format, and coding.

This integration helps in effective analysis of data. Consistency in naming conventions, attribute measures, encoding structure etc. have to be ensured. Consider the following example:
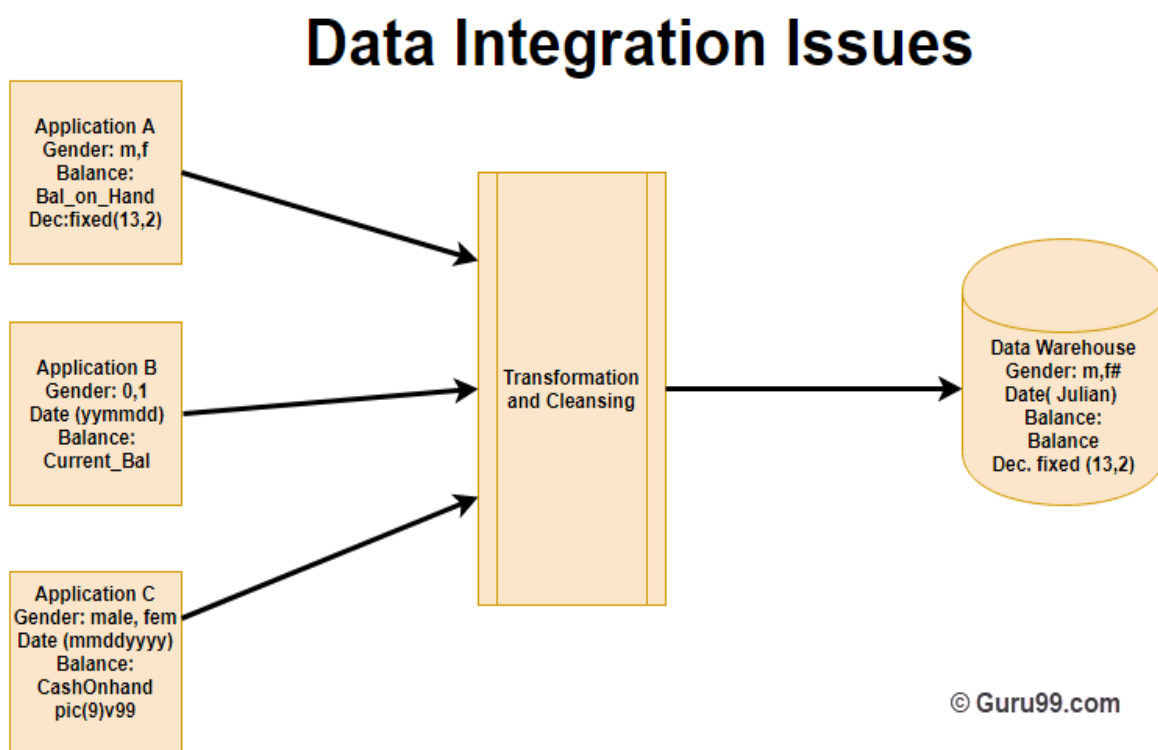
In the above example, there are three different application labeled A, B and C. Information stored in these applications are Gender, Date, and Balance. However, each application's data is stored different way.

- In Application A gender field store logical values like M or F
- In Application B gender field is a numerical value,
- In Application C application, gender field stored in the form of a character value.
- Same is the case with Date and balance

However, after transformation and cleaning process all this data is stored in common format in the Data Warehouse.

(iii).   Time-Variant

The time horizon for data warehouse is quite extensive compared with operational systems. The data collected in a data warehouse is recognized with a particular period and offers information from the historical point of view. It contains an element of time, explicitly or implicitly.

One such place where Datawarehouse data display time variance is in in the structure of the record key. Every primary key contained with the DW should have either implicitly or explicitly an element of time. Like the day, week month, etc.

Another aspect of time variance is that once data is inserted in the warehouse, it can't be updated or changed.

(iv). Non-volatile

Data warehouse is also non-volatile means the previous data is not erased when new data is entered in it.

Data is read-only and periodically refreshed. This also helps to analyze historical data and understand what & when happened. It does not require transaction process, recovery and concurrency control mechanisms.

Activities like delete, update, and insert which are performed in an operational application environment are omitted in Data warehouse environment. Only two types of data operations performed in the Data Warehousing are

1. Data loading
2. Data access

Here, are some major differences between Application and Data Warehouse

Operational Application	Data Warehouse
Complex program must be coded to make sure that data upgrade processes maintain high integrity of the final product.	This kind of issues does not happen because data update is not performed.
Data is placed in a normalized form to ensure minimal redundancy.	Data is not stored in normalized form.
Technology needed to support issues of transactions, data recovery, rollback, and resolution as its deadlock is quite complex.	It offers relative simplicity in technology.

Importance of Data wareHouse

- Data warehouse is an information system that contains historical and commutative data from single or multiple sources.
- A data warehouse is subject oriented as it offers information regarding subject instead of organization's ongoing operations.
- In Data Warehouse, integration means the establishment of a common unit of measure for all similar data from the different databases
- Data warehouse is also non-volatile means the previous data is not erased when new data is entered in it.
- A Datawarehouse is Time-variant as the data in a DW has high shelf life.
- There are 5 main components of a Datawarehouse. 1) Database 2) ETL Tools 3) Meta Data 4) Query Tools 5) DataMarts
- These are four main categories of query tools 1. Query and reporting, tools 2. Application Development tools, 3. Data mining tools 4. OLAP tools
- The data sourcing, transformation, and migration tools are used for performing all the conversions and summarizations.
- In the Data Warehouse Architecture, meta-data plays an important role as it specifies the source, usage, values, and features of data warehouse data.

Enterprise Data warehousing

In computing, a data warehouse (DW or DWH), also known as an enterprise data warehouse (EDW), is a system used for reporting and data analysis. DWs are central repositories of integrated data from one or more disparate sources. They store current and historical data and are used for creating analytical reports for knowledge workers throughout the enterprise. Examples of reports could range from annual and quarterly comparisons and trends to detailed daily sales analyses.

The data stored in the warehouse is uploaded from the operational systems (such as marketing, sales, etc., shown in the figure to the right). The data may pass through an operational data store for additional operations before it is used in the DW for reporting.

A data warehouse is constructed by integrating data from multiple heterogeneous sources. It supports analytical reporting, structured and/or ad hoc queries and decision making. This tutorial adopts a step-by-step approach to explain all the necessary concepts of data warehousing.

The term "Data Warehouse" was first coined by Bill Inmon in 1990. According to Inmon, a data warehouse is a subject oriented, integrated, time-variant, and non-volatile collection of data. This data helps analysts to take informed decisions in an organization.

An operational database undergoes frequent changes on a daily basis on account of the transactions that take place. Suppose a business executive wants to analyze previous feedback on any data such as a product, a supplier, or any consumer data, then the executive will have no data available to analyze because the previous data has been updated due to transactions.

A data warehouses provides us generalized and consolidated data in multidimensional view. Along with generalized and consolidated view of data, a data warehouses also provides us Online Analytical Processing (OLAP) tools. These tools help us in interactive and effective analysis of data in a multidimensional space. This analysis results in data generalization and data mining.

Data mining functions such as association, clustering, classification, prediction can be integrated with OLAP operations to enhance the interactive mining of knowledge at multiple level of abstraction. That's why data warehouse has now become an important platform for data analysis and online analytical processing.

Understanding a Data Warehouse

- A data warehouse is a database, which is kept separate from the organization's operational database.
- There is no frequent updating done in a data warehouse.
- It possesses consolidated historical data, which helps the organization to analyze its business.
- A data warehouse helps executives to organize, understand, and use their data to take strategic decisions.
- Data warehouse systems help in the integration of diversity of application systems.
- A data warehouse system helps in consolidated historical data analysis.

Data Warehouse Features

The key features of a data warehouse are discussed below:

- **Subject Oriented** - A data warehouse is subject oriented because it provides information around a subject rather than the organization's ongoing operations. These subjects can be product, customers, suppliers, sales, revenue, etc. A data warehouse does not focus on the ongoing operations, rather it focuses on modelling and analysis of data for decision making.
- **Integrated** - A data warehouse is constructed by integrating data from heterogeneous sources such as relational databases, flat files, etc. This integration enhances the effective analysis of data.

- **Time Variant** - The data collected in a data warehouse is identified with a particular time period. The data in a data warehouse provides information from the historical point of view.
- **Non-volatile** - Non-volatile means the previous data is not erased when new data is added to it. A data warehouse is kept separate from the operational database and therefore frequent changes in operational database is not reflected in the data warehouse.

Note: A data warehouse does not require transaction processing, recovery, and concurrency controls, because it is physically stored and separate from the operational database.

Types of Data Warehouse

Information processing, analytical processing, and data mining are the three types of data warehouse applications that are discussed below:

- **Information Processing** - A data warehouse allows to process the data stored in it. The data can be processed by means of querying, basic statistical analysis, reporting using crosstabs, tables, charts, or graphs.
- **Analytical Processing** - A data warehouse supports analytical processing of the information stored in it. The data can be analyzed by means of basic OLAP operations, including slice-and-dice, drill down, drill up, and pivoting.
- **Data Mining** - Data mining supports knowledge discovery by finding hidden patterns and associations, constructing analytical models, performing classification and prediction. These mining results can be presented using the visualization tools.

Sr.No.	Data Warehouse (OLAP)	Operational Database(OLTP)
1	It involves historical processing of information.	It involves day-to-day processing.
2	OLAP systems are used by knowledge workers such as executives, managers, and analysts.	OLTP systems are used by clerks, DBAs, or database professionals.
3	It is used to analyze the business.	It is used to run the business.
4	It focuses on Information out.	It focuses on Data in.
5	It is based on Star Schema, Snowflake Schema, and Fact Constellation Schema.	It is based on Entity Relationship Model.
6	It focuses on Information out.	It is application oriented.
7	It contains historical data.	It contains current data.
8	It provides summarized and consolidated data.	It provides primitive and highly detailed data.
9	It provides summarized and multidimensional view of data.	It provides detailed and flat relational view of data.
10	The number of users is in hundreds.	The number of users is in thousands.
11	The number of records accessed is in millions.	The number of records accessed is in tens.
12	The database size is from 100GB to 100 TB.	The database size is from 100 MB to 100 GB.
13	These are highly flexible.	It provides high performance.

Planning Stages of Data warehousing

The key steps in developing a data warehouse can be summarized as follows:

1. Project initiation
2. Requirements analysis
3. Design (architecture, databases and applications)
4. Construction (selecting and installing tools, developing data feeds and building reports)
5. Deployment (release & training)
6. Maintenance
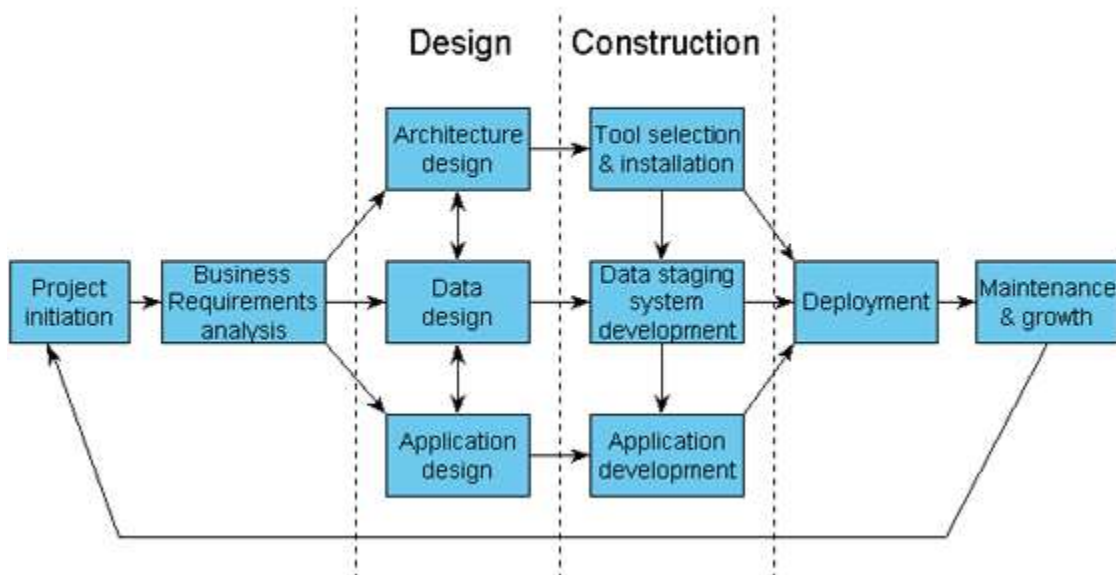
It is advisable to conduct a pilot exercise before embarking on a full-scale development effort. This will include most of the above steps, and provides an opportunity to:

- understand new concepts and processes, and identify potential problems;
- make more realistic plans and manage expectations;
- evaluate alternative tools;
- demonstrate benefits and gain management commitment.

Testing should be an integral part of construction, not a separate step in the development process.

1. Project initiation
No data warehousing project should commence without:

- a clear statement of business objectives and scope;
- a sound business case, including measurable benefits;
- an outline project plan, including estimated costs, timescales and resource requirements;
- high level executive backing, including a commitment to provide the necessary resources;

A small team is usually set up to prepare and present a suitable project initiation document. This is normally a joint effort between business and IT managers. If the organization has limited data warehousing experience, it is useful to obtain

external advice at this stage. If the project goes ahead, the project plan and business case should be reviewed at each stage.

It is widely regarded as good practise to develop a data warehouse in small, manageable phases (see pitfalls). Thus the analysis, design, construction and deployment steps will be repeated in cycles.

It is generally a good tactic to provide something that is not already available during the first phase, as this will help to stimulate real interest. This could be new data or enhanced functionality. It is also better to start with something relatively easy, which the warehousing team can deliver whilst still learning the ropes.

See project management techniques for more information on relevant methodologies and useful references.

2. Requirements analysis

Establishing a broad view of the business' requirements should always be the first step. The understanding gained will guide everything that follows, and the details can be filled in for each phase in turn.

Collecting requirements typically involves 4 principal activities:

- Interviewing a number of potential users to find out what they do, the information they need and how they analyse it in order to make decisions. It is often helpful to analyse some of the reports they currently use.
- Interviewing information systems specialists to find out what data are available in potential source systems, and how they are organised.
- Analysing the requirements to establish those that are feasible given available data.
- Running facilitated workshops that bring representative users and IT staff together to build consensus about what is needed, what is feasible and where to start.

3. Design

The goal of the design process is to define the warehouse components that will need to be built. The architecture, data and application designs are all inter-related, and are normally produced in parallel.

(i). Architecture design

The warehouse architecture describes all the hardware and software components that form the data warehousing environment and explains:

- how the components will work together;
- where they are located (geographically and on what platform);
- who uses them;
- who will build and maintain them.

The architecture needs to be considered at the outset, as this provides a framework for the selection of tools and the detailed design of individual components during the first and subsequent phases of development.

(ii).   Data design

This step determines the structure of the primary data stores used in the warehouse environment, based on the outcome of the requirements analysis. It is best to produce a broad outline quickly, and then break the detailed design into phases, each of which usually progresses from logical to physical:

The logical design determines what data are stored in the main data warehouse and any associated functional data marts. There are a number of data modelling techniques that can be used to help.

Once the logical design is established, the next step is to define the physical characteristics of individual data stores (including aggregates) and any associated indexes required to optimize performance (see database optimization).

The data design is critical to further progress, in that it defines the target for the data feeds and provides the source data for all reporting and analysis applications.

(iii).   Application design

The application design describes the reports and analyses required by a particular group of users, and usually specifies:

- a number of template report layouts;
- how and when these reports will be delivered to users;
- the functional requirements for the user interface.

There may be one or more applications associated with each data mart or phase of development.

4. Construction

Warehouse components are usually developed iteratively and in parallel. That said, the most efficient sequence to begin construction is probably as follows:

[1].   Tool selection & installation

Selecting tools is best carried out as part of a pilot exercise, using a sample of real data. This allows the development team to assess how well competing tools handle problems specific to their organization, and to test system performance before committing to purchase.

The most important choices are the:

- ETL tool
- Database(s) for the warehouse (usually relational) and marts (often multi-dimensional)
- Reporting and analysis tools

Clearly these need to be compatible, and it is worth checking reference sites to make sure they work well together.

It pays to define standards and configure the development, testing and production environments as soon as tools are installed, rather than waiting until development is well underway. Most vendors are willing to provide assistance with these steps, and this is normally well worth the investment.

[2].   Data staging system

This comprises the physical warehouse database, data feeds and any associated data marts and aggregates. The following steps are typical:

- Create target tables in the central warehouse database;
- Request initial and regular extracts from source systems;
- Write procedures to transform extract data ready for loading (optionally creating interim tables in a data staging area);
- Write procedures to load initial data into the warehouse (using a bulk loader);
- Create and populate any data marts;
- Write procedure to load regular updates into the warehouse;
- Develop special procedures for a once-off bulk load of historic data;
- Write validation/exception handling procedures;
- Write archiving & backup procedures;
- Create a provisional set of aggregates;
- Automate all regular procedures;
- Document the whole process.

However thorough the design process, problems with the real data are bound to surface at this stage. Substantial time should be allowed to resolve any issues that arise, establish appropriate data cleansing procedures (preferably within the source systems environment) and to validate all data before they are released for live use.

[3].   Application development

This step can begin once a sample or initial extract has been loaded, but it is usually best to leave the bulk of application development until the underlying data mart (or part of the central warehouse) and associated meta-data (especially object names) are stable.

It is a good idea to involve users in the development of reports and analytic applications, preferably through prototyping, but at least by asking them to carry out acceptance testing. Most modern business intelligence tools do not require programming, so it is possible for non-IT staff to build some of their own reports as well.

5. Deployment

It is too often assumed that the first version of a data warehouse can be rolled out in a matter of weeks, simply by showing all the users how to use the new reporting tools.

In practice, training needs to cover not just the basic use of the tools, but also the data that have been made available, and, more significantly perhaps, the new business processes or different ways of working that are intended. This training usually works best if delivered on a one-to-one basis.

As well as training, planning for deployment needs to cover:

- Installing and configuring desktop PCs - any hardware upgrades or amendments to the 'standard build' need to be organized well in advance;
- Implementing appropriate security measures - to control access to applications and data;

- Setting up a support organization to deal with questions about the tools, the applications and the data. However thoroughly the data were checked and documented prior to publication, users are likely to spot anomalies requiring investigation and to need assistance interpreting the results they obtain from the warehouse and reconciling these with existing reports;
- Providing more advanced tool training later, when users are ready, and assisting potential power users to develop their first few reports.

If the first users find errors and inconsistencies in the data, don't feel comfortable with the tool or can't be bothered to learn how to use it properly, or won't accept new procedures and responsibilities, all the time spent building the warehouse may ultimately be wasted. The following guidelines will help to reduce these risks:

- Do not start deployment until the data are ready (available and validated) and the tools and update procedures have been tested;
- Use a small, representative group to try out the finished system before rolling out, including users with a range of abilities and attitudes;
- Do not grant system access to users until they have been trained.

6. Maintenance

A data warehouse is not like an OLTP system: development is never finished, but follows an iterative cycle (analyse – build – deploy). Also, once live, a warehousing environment requires substantial effort to keep running. Thus the development team should not anticipate handing over and moving on to other projects, but to spend half of their time on support and maintenance.

The most important activities are:

- Monitoring the realisation of expected benefits;
- Providing ongoing support to users (see deployment);
- Training new staff;
- Assisting with the identification and cleansing of dirty data;
- Maintaining both feeds & meta-data as source systems change over time;
- Tuning the warehouse for maximum performance (this includes managing indexes and aggregates according to actual usage);
- Purging dormant data;
- Recording successes and using these to continuously market the warehouse.

In addition, mechanisms need to be established to manage growth, in particular the prioritisation of requested enhancements, which often require the addition of further data sources.

Various Data Warehouse Design Approaches: Top-Down and Bottom-Up

Data Warehouse design approaches are very important aspect of building data warehouse. Selection of right data warehouse design could save lot of time and project cost.

There are two different Data Warehouse Design Approaches normally followed when designing a Data Warehouse solution and based on the requirements of your project you can choose which one suits your particular

scenario. These methodologies are a result of research from Bill Inmon and Ralph Kimball.

[1]. Bill Inmon – Top-down Data Warehouse Design Approach

"Bill Inmon" is sometimes also referred to as the "father of data warehousing"; his design methodology is based on a top-down approach. In the top-down approach, the data warehouse is designed first and then data mart are built on top of data warehouse.
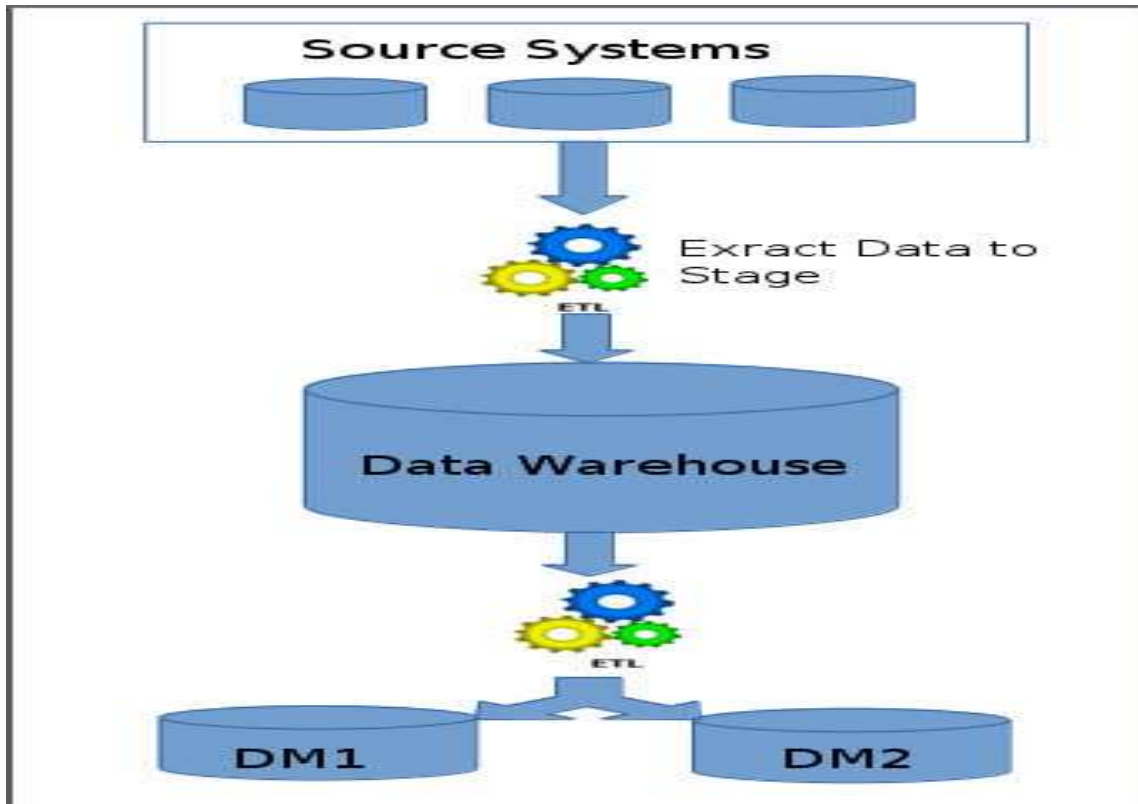
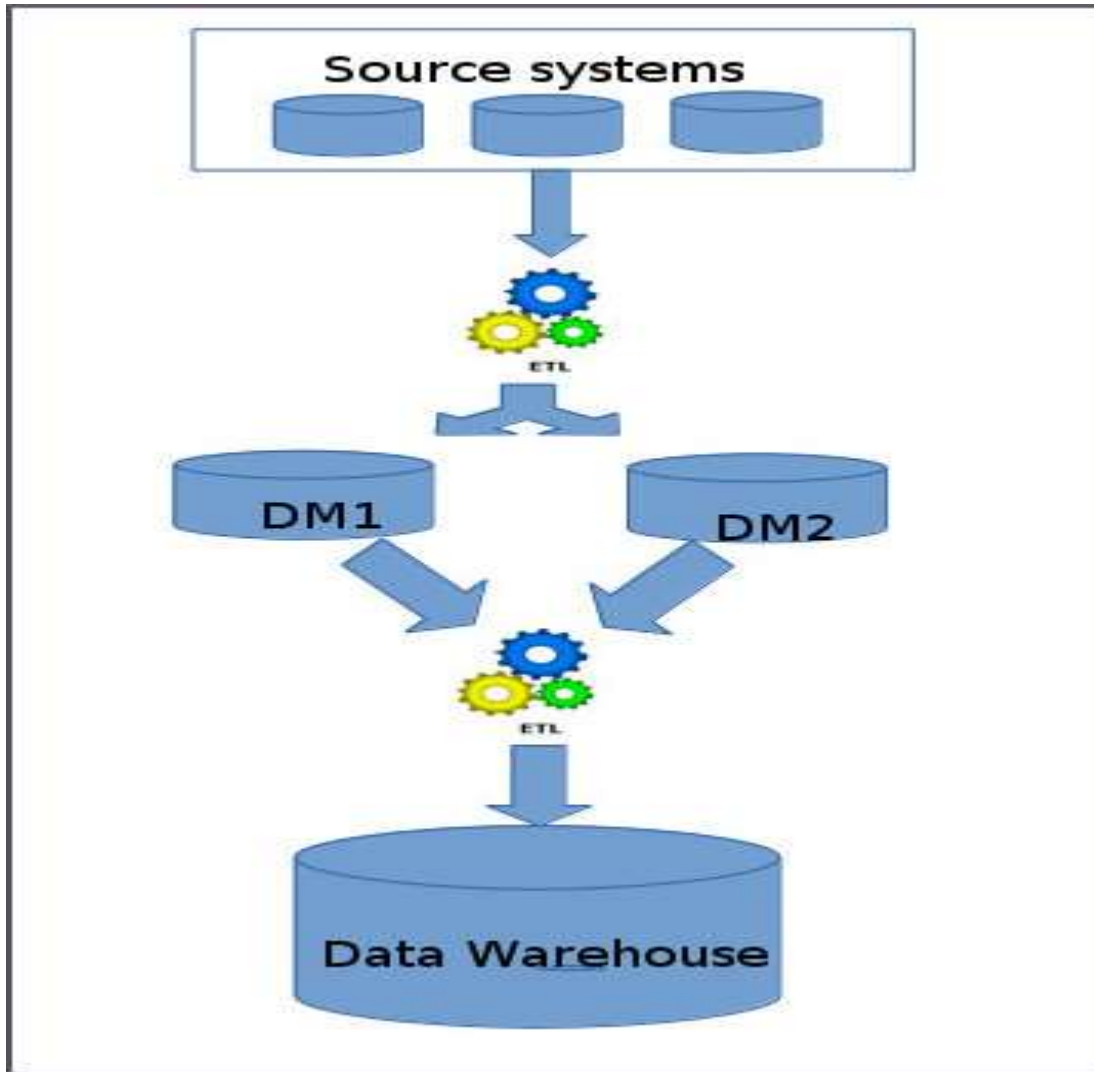The above image depicts how the top-down approach works.

Below are the steps that are involved in top-down approach:

- Data is extracted from the various source systems. The extracts are loaded and validated in the stage area. Validation is required to make sure the extracted data is accurate and correct. You can use the ETL tools or approach to extract and push to the data warehouse.

- Data is extracted from the data warehouse in regular basis in stage area. At this step, you will apply various aggregation, summerization techniques on extracted data and loaded back to the data warehouse.

- Once the aggregation and summerization is completed, various data marts extract that data and apply the some more transformation to make the data structure as defined by the data marts.

[2]. Ralph Kimball – Bottom-up Data Warehouse Design Approach

Ralph Kimball is a renowned author on the subject of data warehousing. His data warehouse design approach is called dimensional modelling or the Kimball methodology. This methodology follows the bottom-up approach.

As per this method, data marts are first created to provide the reporting and analytics capability for specific business process, later with these data marts enterprise data warehouse is created.

The above image depicts how the bottom-up approach works.

Basically, Kimball model reverses the Inmon model i.e. Data marts are directly loaded with the data from the source systems and then ETL process is used to load in to Data Warehouse. The above image depicts how the top-down approach works.

Below are the steps that are involved in bottom-up approach:

- The data flow in the bottom up approach starts from extraction of data from various source system into the stage area where it is processed and loaded into the data marts that are handling specific business process.

- After data marts are refreshed the current data is once again extracted in stage area and transformations are applied to create data into the data mart structure. The data is the extracted from Data Mart to the staging area is aggregated, summarized and so on loaded into EDW and then made available for the end user for analysis and enables critical business decisions.

Data Warehouse Design Approaches

Data warehouse design is one of the key technique in building the data warehouse. Choosing a right data warehouse design can save the project time and cost. Basically there are two data warehouse design approaches are popular.

1. **Bottom-Up Design:**

   In the bottom-up design approach, the data marts are created first to provide reporting capability. A data mart addresses a single business area such as sales, Finance etc. These data marts are then integrated to build a complete data warehouse. The integration of data marts is

implemented using data warehouse bus architecture. In the bus architecture, a dimension is shared between facts in two or more data marts. These dimensions are called conformed dimensions. These conformed dimensions are integrated from data marts and then data warehouse is built.

Advantages of bottom-up design are:
- This model contains consistent data marts and these data marts can be delivered quickly.
- As the data marts are created first, reports can be generated quickly.
- The data warehouse can be extended easily to accommodate new business units. It is just creating new data marts and then integrating with other data marts.

Disadvantages of bottom-up design are:
- The positions of the data warehouse and the data marts are reversed in the bottom-up approach design.

2. Top-Down Design:

In the top-down design approach the, data warehouse is built first. The data marts are then created from the data warehouse.

Advantages of top-down design are:
- Provides consistent dimensional views of data across data marts, as all data marts are loaded from the data warehouse.
- This approach is robust against business changes. Creating a new data mart from the data warehouse is very easy.

Disadvantages of top-down design are:
- This methodology is inflexible to changing departmental needs during implementation phase.
- It represents a very large project and the cost of implementing the project is significant.