International Journal of Multidisciplinary Research and Development
Online ISSN: 2349-4182, Print ISSN: 2349-5979
Impact Factor: RJIF 5.72
www.allsubjectjournal.com
Volume 4; Issue 12; December 2017; Page No. 117-126

Analytics – Life Cycle

Subash Thota
Data Architect, Data Architect, Washington DC, USA

Abstract
As an IT professional, we work on computer application as an analyst, programmer, designer, developer, Database administrator or project manager. In these roles, IT professional is involved in design, implementation, and maintenance of systems that support day-to-day business operations. As an enterprise grows, hundreds of computer applications are needed to support the various business processes. These applications are efficient in gathering, storing and processing all the data required to perform the daily operations of an enterprise successfully.
This existing setup of operational support to an enterprise couldn't cater strategic information management and Decision support system which can combat increased business competition and complexity. Thus, developed the notion of Data Warehouse, which is a centralized solution to enterprise's strategic information need that also acts as Decision Support system. This article gives an overview of all the process involved in building Data Warehouse. Data Warehouse is a multi-step process that begins with an analysis of the legacy data and culminates in the loading and reconciliation of data into the new applications.

Keywords: data warehouse, analytics, data integration, enterprise data

Introduction
The term "Data Warehouse" was first coined by Bill Inmon in 1990. According to Inmon, a data warehouse is a subject oriented, time-variant, an integrated, and non-volatile collection of data. This data helps analysts to make informed decisions in an organization. A data warehouse provides us generalized and consolidated data in the multi-dimensional view. Along with the generalized and unified view of data, a data warehouses also gives us Online Analytical Processing (OLAP) tools. These tools help us in an interactive and efficient analysis of data in a multidimensional space. This study 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 levels of abstraction. That's why data warehouse has now become an important platform for data analysis and online analytical processing. The type of information needed for strategic decision making is different from operational system. This leads to needing a new environment for providing strategic information for analysis, discerning trends and monitoring performance.
The desired features of this new environment are:

- Database designed for analytic tasks
- Data from multiple application/source
- Read-intensive data usage
- Direct interaction with the system by the users without IT assistance
- Content updates periodically and stable
- Historical and current data
- Ability for user to run query and get result online
- Ability for user to initiate reports

Requirement Phase
Requirement phase is beginning the phase of the project; requirement phase states that the number and a brief description of systems from which data will be migrated, what kind of data available in those systems, the overall quality of data, to which system the data should be migrated. Business requirements for the data help determine what data to migrate. These requirements can be derived from agreements, objectives, and scope. It contains the in the scope of the project along with stakeholders. The Proper analysis gives a clear understanding of scope to begin, and will not be completed until tested against clearly identified project constraints. During this Phase, below activities are performed

- Legacy system understanding to baseline the scope
- Acceptance Criteria and sign-off
- Legacy data extraction format discussion and finalization
- Understanding target data structure
- Legacy data analysis & Profiling to identify the data gaps and anomalies
- Deliverables and Ownership

Design Phase
During the design phase, i.e., after Analysis phase, High-level Approach and Low-level design components are created. Following are the contents of the design phase.

- Approach for end to end solution
- Master and transactional data handling approach
- Approach for historical data handling
- Approach for data cleansing
- Data Reconciliation and Error handling approach
- Target Load and validation approach
- Cut-over approach

- Software and Hardware requirements for end to end solution
- Data model changes between Legacy and target.
- Data type changes between Legacy and target
- Missing/Un-available values for target mandatory columns
- Business Process specific changes or implementing some Business Rules
- Changes in List of Values between Legacy and target
- Target attribute that needs to be derived from more than one columns of Legacy
- Target Specific unique identifier requirement

The detailed design phase comprises of the following key activities:

- Data Profiling - Data profiling is an analysis of the candidate data sources for a data warehouse to clarify the structure, content, relationships and derivation rules of the data. The benefits of data profiling are shortened the implementation cycle of significant projects, to improve data quality, and improve understanding of data for the users.
- Source to Target attribute mapping
- Staging area design
- Technical design
- Audit, Rejection and Error handling

Development Phase

After completing the Design phase, Development team works components as listed below:

- ETL mappings for Cleansing and Transformation
- Building Fallout Framework
- Building Reconciliation Scripts
- All these components are Unit Tested before Dry Runs.

Actual coding and unit testing will be done in the construction phase. During the development phase, the structures which are like target system should be created. Data from different legacy systems will be extracted from a staging area & source data will be consolidated in the staging area. The Construction and Unit Testing (CUT) phase include the development of mappings in ETL tool depend on Source to Target Mapping sheet.

Transformation Rules will be applied to required mappings and loaded the data into target structures. Reconciliation programs will also be developed during this phase. Unit test cases (UTC) to be written for validating source systems, source definitions, target definitions and to check the connection strings, validation of data types and ports, etc.

Components of a data warehouse

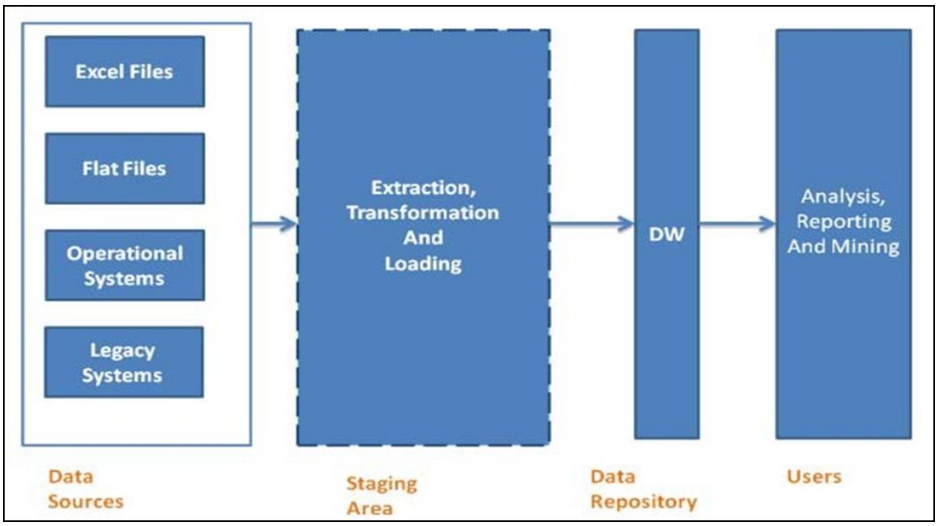


Fig 1

Data Sources

A flat file database stores data in a normal text format. Contrary to a relational database where the data is stored in the form of tables, in a flat file database, the data stored does not have a folders or paths related to them. No manipulations are performed on the data. Delimiters are used in flat files to separate the data columns.

Excel spreadsheets are regularly used in data warehousing operations. They are impressive, low-priced, and flexible tools that many decision-makers find convenient to use. Excel also provides graphing features that allow the end-user to present the required data in chart and graph formats. These formats can be easily integrated into MS Word and PowerPoint presentations.

Operational systems of a business contain the day to day transactions of the data at a low-level. For example, the sales data, HR data, marketing data are used as input sources for a data warehouse.

Legacy systems are the applications of the yesteryear. They mirror the requirements of a business that might be twenty to twenty-five-year-old. They are used till date since over the years these systems have captured the business knowledge and rules that are exceptionally difficult to translate to a new platform/application.

Staging Area

The first part of the staging area is the most challenging process of extraction. Depending on how accurately the data is extracted the subsequent operations succeed or fail. The source systems might be complicated or poorly documented due to which the process becomes more difficult. The data may be extracted not only once but also periodically when changes occur on the source side.

The second stage is the transformation where the data is converted from one format to another. Since data often exists in different locations and formats across the enterprises, data conversion is mandatory to ensure that data from one application is comprehensible to other applications and databases.

The third stage is the loading where the extracted and transformed data is loaded into a data mart or a data warehouse depending on the business. The populated data is used for presentation applications by the end users.

Data Repository

The data is loaded into a data warehouse in the form of facts and dimensions. This is the data warehouse for business community (remember, they can't see data staging area). It is Series of integrated data marts.

- Data must be stored and accessed in dimensional schemas
- No normalization (3NF) should be used
- Dimensional schemas are simple and intuitive for business users; Normalized schemas are difficult to grasp by them
- Data must be atomic (at lower granularity)
- Not only summarized – they don't allow for arbitrary, complex queries
- Data marts must be built on dimensions and facts that are confirmed
- Otherwise, data marts are stovepipes
- Conformation leads to bus architecture – data marts can cooperate

Users

The loaded data is accessed for reporting, analysis, and mining. The reporting tools like Business Objects and Cognos are used by users to generate reports. The data is also used for predicting trends. Ad hoc, complex queries are targeted to a small percentage of business users.

Data Warehouse Features

The key features of a data warehouse are discussed below:

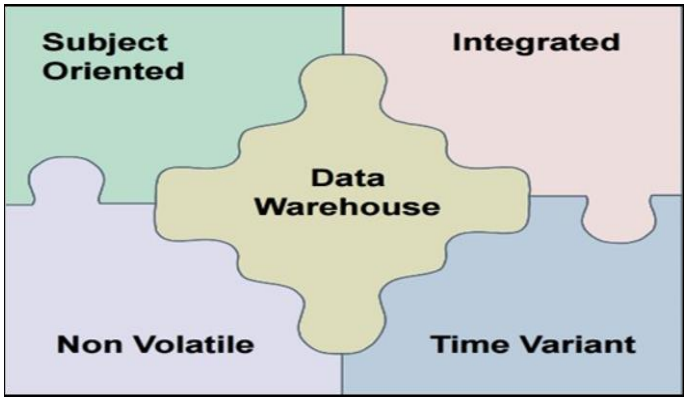
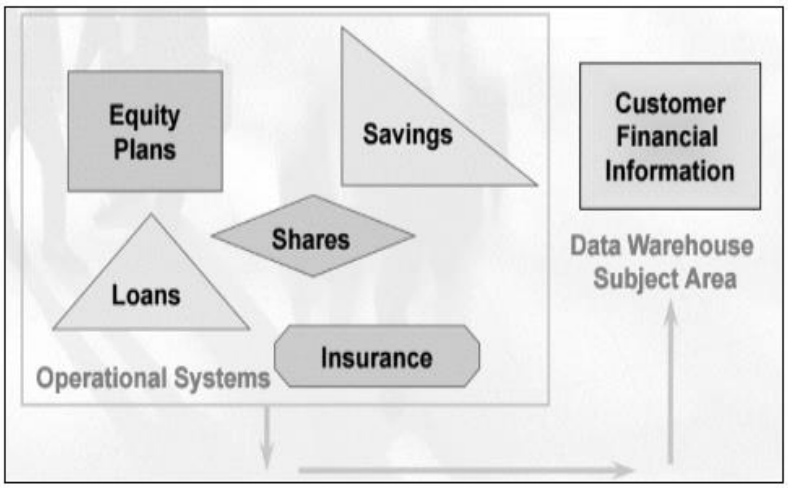


Fig 2

Subject-Oriented Data

In every industry, datasets are organized around individual applications to support operational systems. These individuals' datasets must provide data for the specific application to perform specific functions efficiently. In contrast, in Data Warehouse, data is stored by subjects, not by applications. These are the business-critical subjects for the enterprise, e.g., for a manufacturing company, sales, shipments, and inventory are critical business subjects while for a retail store, sales at the check-out counter are a critical subject.



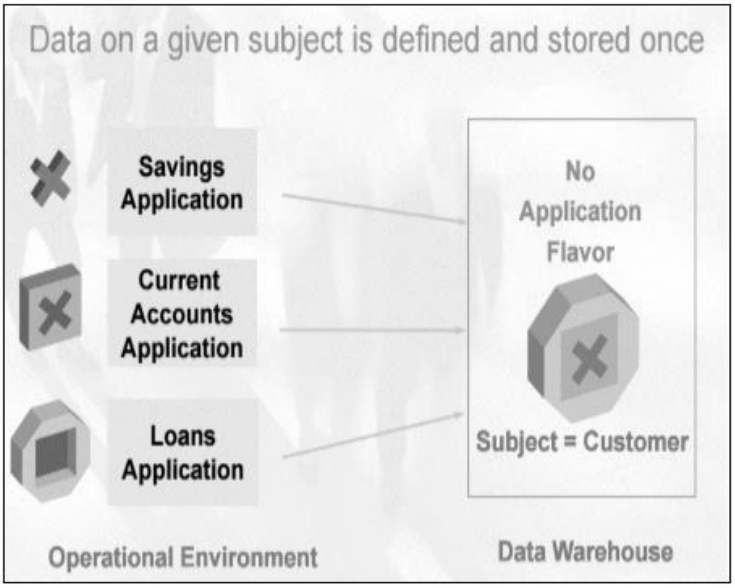
- Organized around major subjects, such as customer product sales customer, product, sales.
- Focusing on the modeling and analysis of data for decision makers, not on daily operations or transaction processing transaction processing.
- Provide a simple and concise view around subject issues by excluding data that are not useful in the decision support process.

Fig 3

Integrated Data

To have proper decision making, data from all various applications need to be pulled out to common area. The data in the data warehouse comes from several operational systems. Source data are in different databases, files, and data segments. These are disparate applications, so the operational

platforms and operating systems could be different. The file layouts, character code representations, and field naming conventions all could be different. Before the data from various disparate sources can be usefully stored in a data warehouse, we must remove the inconsistencies. We must standardize the various data elements and make sure of the meanings of data names in each source application. Before moving the data into the data warehouse, we must go through a process of transformation, consolidation, and integration of the source data.



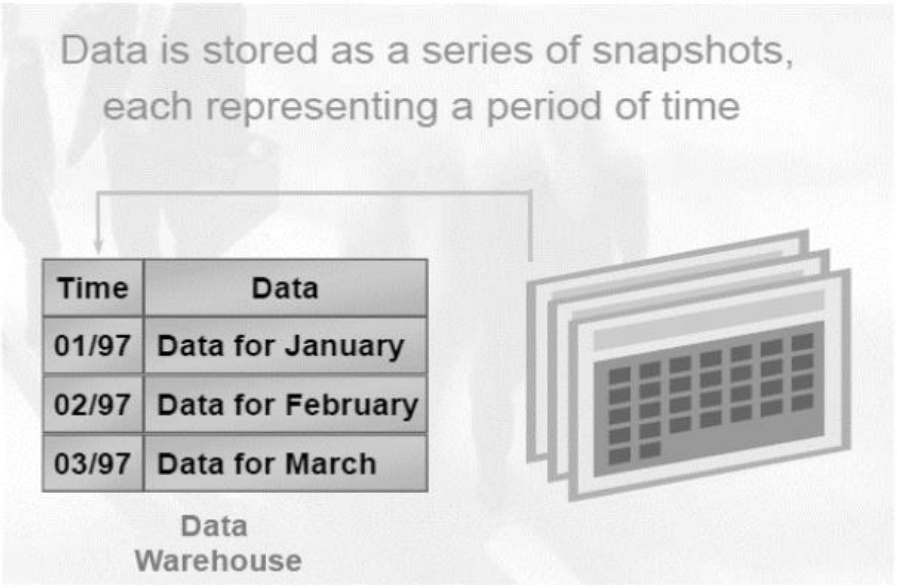
- Constructed by integrating multiple, heterogeneous data sources–relational databases, flat files, on-line transaction records
- One set of consistent, accurate quality information
- Standardization
  - Naming convention
  - Coding structures
  - Data attributes
  - Measures
- Data cleaning and data integration techniques are applied.
  - Ensure consistency in naming conventions, encoding structures attribute measures, etc. encoding structures, attribute measures, etc. Among different data sources
  - When data is moved to the warehouse, it is converted

Fig 4

Time-Variant Data

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. A data warehouse, because of the very nature of its purpose, must contain historical data, not just current values. Data is stored as snapshots over past and current periods. 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



- The time horizon for the data warehouse is significantly longer than that of operational systems

– Operational database: current value data

– Data warehouse data: provide information from a historical perspective (e.g., past 5-10 years)

- Every key structure in the data warehouse

– Contains an element of time, explicitly or implicitly

– But the key of operational data may or may not contain "time element."

Fig 5

Non-Volatile Data

Data from the operational systems are moved into the data warehouse at specific intervals. Depending on the requirements of the business, these data movements take place twice a day, once a day, once a week, or once in two weeks. The business transactions update the operational system

databases in real time. We add, change, or delete data from an operational system as each transaction happens but do not usually update the data in the data warehouse. We do not delete the data in the data warehouse in real time. Once the data is captured in the data warehouse, we do not run individual transactions to change the data there. The data in a data warehouse is not as volatile as the data in an operational database.

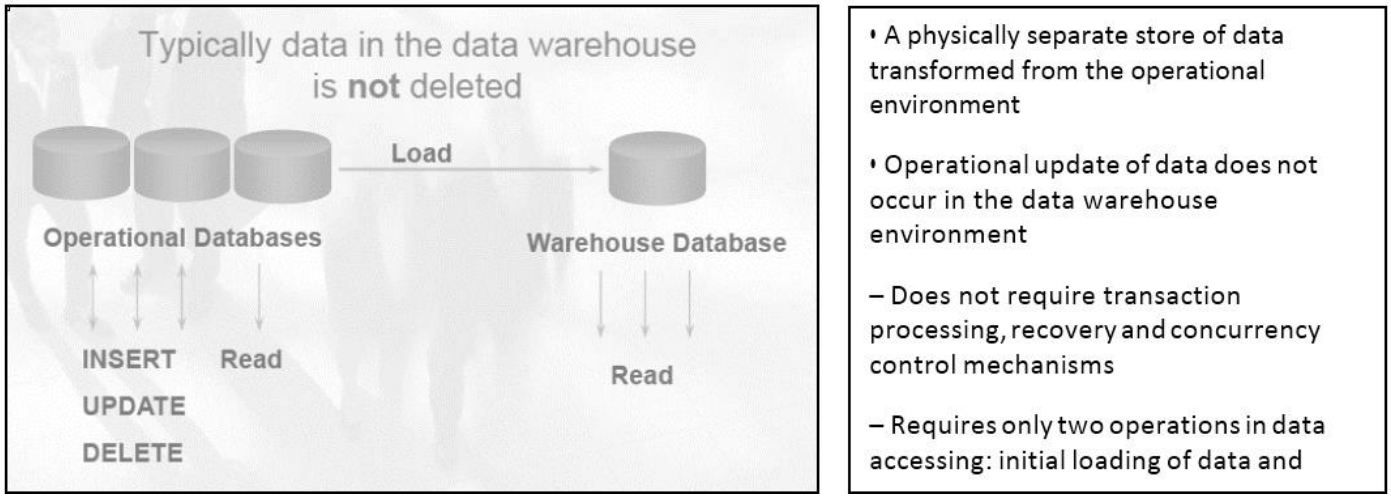


Fig 6

Integration of Databases

For the integration of data from disparate, homogeneous sources, data warehouse follows two different approaches:

1. Query-driven approach.
2. Update driven approach

Query-driven approach

The query driven approach is the traditional approach of integration of heterogeneous databases. It is the approach that is used for the construction of wrappers and integrators on top of multiple disparate databases. These integrators are called mediators as well. The process of query driver approach to data integration follows three major steps. Once the query has been issued to the side of the client, this query is translated into the queries that are fit for the individual heterogeneous site that is involved in this step. This translation is done by a metadata dictionary. The queries are then mapped and then sent to the local query processor, and the results thus received from the heterogeneous sites are integrated. This helps in the formation of a set of global answers.

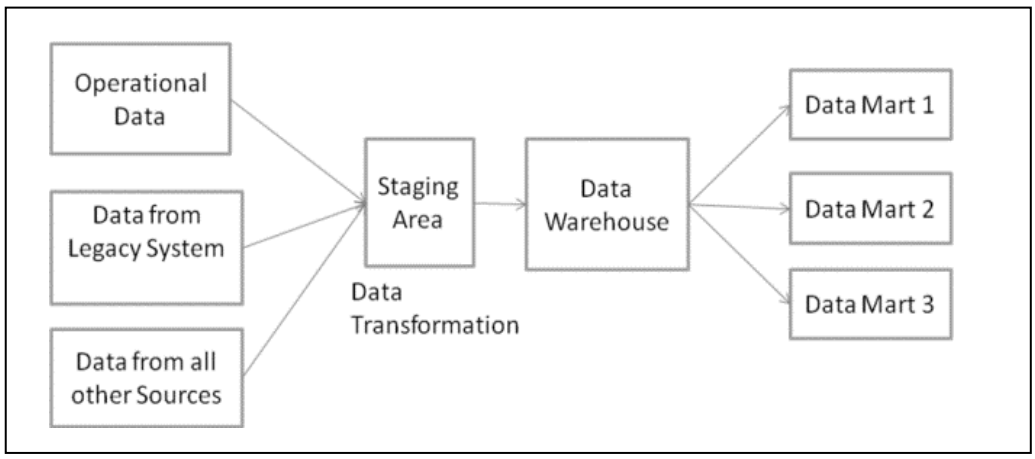


Fig 7

Though a very useful approach, this approach, however, has its own set of disadvantages as well. The processes of filtration and integration that are involved in the query driven approach are very complex. Though helpful, the approach may not always prove to be efficient. Besides, in case of frequent queries, the approach is also very expensive. In fact, when it comes to queries that are in requirement of aggregations, the approach can prove to be even more expensive.

The advantages of this approach are:

- A truly corporate effort, an enterprise view of data
- Inherently architected—not a union of disparate data marts
- Single, central storage of data about the content
- Centralized rules and control

- May see quick results if implemented with iterations

The disadvantages are:

- Takes longer to build even with an iterative method
- High exposure/risk of failure
- Needs high level of cross-functional skills
- High outlay without proof of concept

Update-driven approach

This is an alternative to the traditional approach. In update-driven approach, the information from multiple heterogeneous sources are integrated in advance and are stored in a warehouse. This information is available for direct querying and analysis. Here, 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 with data marts, and the data warehouse is built.

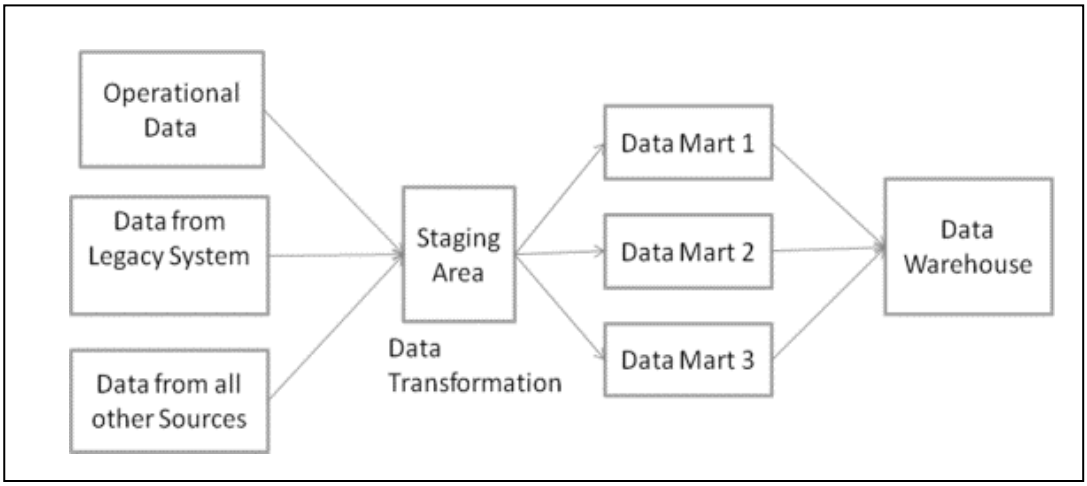


Fig 8

The advantages of this approach are:

- Faster and easier implementation of manageable pieces
- Favorable return on investment and proof of concept
- Less risk of failure
- Inherently incremental; can schedule important data marts first
- Allows project team to learn and grow

The disadvantages are:

- Each data mart has its narrow view of data
- Permeates redundant data in every data mart
- Perpetuates inconsistent and irreconcilable data
- Proliferates unmanageable interfaces

Process flow in data warehouse

Four major functions need to be performed for getting the data ready for Data Warehouse, which are extracting the data, transforming the data, and then loading the data into the data warehouse storage, managing queries and directing them to the appropriate data sources.

Extract and Load Process

This function must deal with numerous data sources. Appropriate techniques are employed for each data source. Source data may be from different source machines in diverse data formats. Part of the source data may be in relational database systems. Some data may be on other legacy network and hierarchical data models. Many data sources may still be in flat files. Also, there could be data from spreadsheets and local departmental data sets.

Controlling the process

Controlling the process involves determining when to start data extraction and the consistency check on data. Controlling process ensures that the tools, the logic modules, and the programs are executed in correct sequence and at the correct time.

When to Initiate Extract

Data needs to be in a consistent state when it is extracted, i.e., the data warehouse should represent a single, consistent version of the information to the user. For example, in a customer profiling data warehouse in the telecommunication sector, it is illogical to merge the list of customers at 8 pm on Wednesday from a customer database with the customer subscription events up to 8 pm on Tuesday. This would mean that we are finding the customers for whom there are no associated subscriptions.

Loading the data

After extracting the data, it is loaded into a temporary data store where it is cleaned up and made consistently.

Transform Process

There are various processes that are performed in this step.

Cleaning Process

Firstly, data extracted from each source is cleaned. Cleaning may just be a correction of misspellings, or may include resolution of conflicts between state codes and zip codes in the source data, or may deal with providing default values for

missing data elements, or elimination of duplicates when we bring in the same data from multiple source systems. It can be done by making the data consistent:

- Within itself.
- With other data within the same data source.
- With the data in other source systems.
- With the existing data present in the warehouse

Standardize and Transform Data

Standardization of data elements forms a large part of data transformation. Standardization of data types and field lengths for same data elements retrieved from the various sources is done in this part along with Semantic standardization which is another major task. Data transformation involves many forms of combining pieces of data from the different sources. On the other hand, data transformation also involves purging source data that is not useful and separating outsources records into new combinations. Sorting and merging of data takes place on a large scale in the data staging area. Transforming involves converting the source data into a structure. Structuring the data increases the query performance and decreases the operational cost. The data contained in a data warehouse must be transformed to support performance requirements and control the ongoing operational costs.

Partition the data

It will optimize the hardware performance and simplify the management of data warehouse. Here we partition each fact table into multiple separate partitions.

Aggregation

Aggregation is required to speed up common queries. Aggregation relies on the fact that most common queries will analyze a subset or an aggregation of the detailed data. At the end of data, transformation function data received is a collection of integrated data which is cleaned, standardized, and summarized.

Backup and Archive the data

To recover the data in the event of data loss, software failure, or hardware failure, it is necessary to keep regular backups. Archiving involves removing the old data from the system in a format that allow it to be quickly restored whenever required.

Data Loading and Query management process

Once all the data has been cleansed and transformed into a structure consistent with the data warehouse requirements, data is ready for loading into the data warehouse. Two distinct groups of tasks form the data loading function. When the design and construction of the data warehouse get completed and go live for the first time gets scheduled, initial loading of the data into the data warehouse storage is done. The initial load moves large volumes of data using up substantial amounts of time. As the data warehouse starts functioning, the process continues to extract the changes to the source data, transform the data revisions, and feed the incremental data revisions on an ongoing basis.

This process performs the following functions:

- Manages the queries.
- Helps speed up the execution time of queries.
- Directs the queries to their most effective data sources.
- Ensures that all the system sources are used most effectively.
- Monitors actual query profiles.

The information generated in this process is used by the warehouse management process to determine which aggregations to generate. This process does not operate during the regular load of information into a data warehouse.

Data warehouse architecture

Different data warehousing systems have different structures. Some may have an ODS (operational data store), while some may have multiple data marts. Some may have a small number of data sources, while some may have dozens of data sources. Given this, it is far more reasonable to present the different layers of a data warehouse architecture rather than discussing the specifics of any one system. From a high perspective, the data warehouse architecture can be represented as a block diagram with five main components:

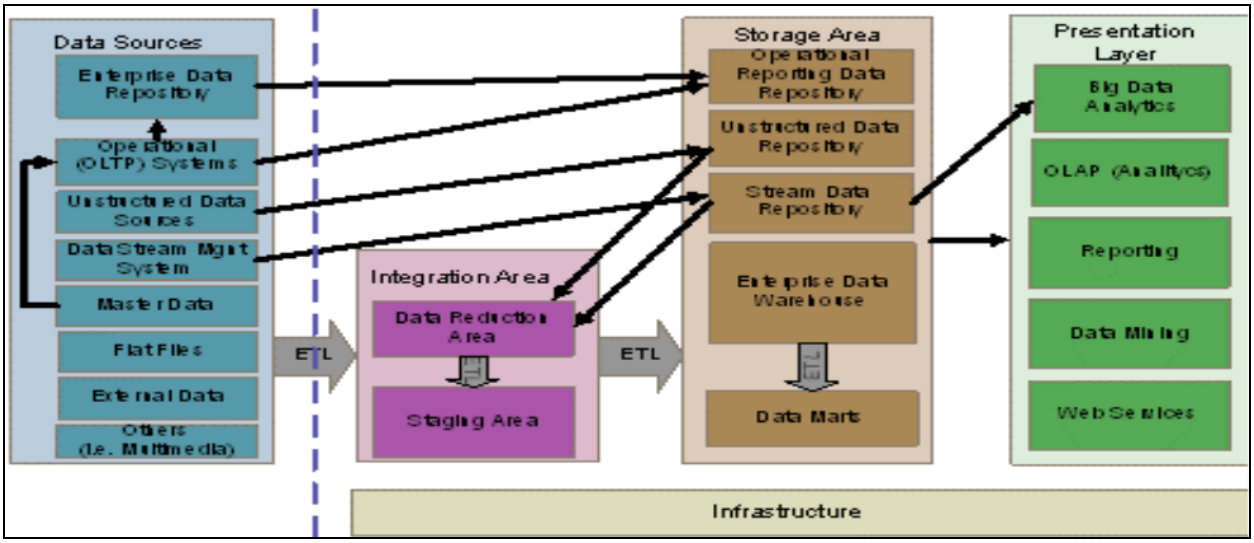


Fig 9

- The data sources
- The integration area
- The storage area
- The presentation layer and
- The hardware Infrastructure

The data flow from the data sources to the integration area, storage area, and presentation layer. The blocks in the figure are sized to allow clarity; neither, the size of a block or its position, represent importance order.

The Data Sources

The Data Sources are not part of Data Warehouse Architecture, but of the Enterprise Architecture. The Data Sources are all the containers of information in the enterprise. This data can be structured (i.e., relational databases, Excel spreadsheets, etc.), unstructured (i.e., Word documents, text files, flat files, etc.), big data repositories (i.e., sensors readings records, website logs) or multimedia (i.e., videos, images, voice records, etc.).

Two important components are the Master Data Repository and the Enterprise Data Repository. If a Master Data Repository has been implemented, the main data warehouse dimensions can be directly derived from this component. If the Enterprise Data Repository has been implemented, we can use it as the main source for populating the Operational Reporting Data Repository and to obtain the new data to be transferred to the Staging Area. One of the common limitations in using the Enterprise Data Repository for analytics is that, often, does not contain temporary data information (timestamps of the events). We use the back-end tools and utilities to feed data into the bottom tier. Even it the bulk of the data preparation occurs in the Integration Area, ETL (Extract Transform – Load) processes are necessary at every step-in data flow.

The Integration Area

The integration area is where the data, originated from disparate sources, is linked, transformed (is needed) and structured in a suitable format to be stored in the Enterprise Data Warehouse. This area has two main components, the Data Reduction Area and the Staging Area. The Data Reduction Area is where we implement different techniques to extract insight from the Big Data Repository. These insights are summarized, transformed into relational structures and move to the staging area before becoming part of the Enterprise Data Warehouse. The staging is where we made the final preparation of the new data to become confirmed with the Enterprise Data Warehouse structures and semantic. The data in the Integration Area is highly volatile; most of it is only deltas that will be added to the Enterprise Data Warehouse. Once moved to the Enterprise Data Warehouse, this data can be discarded to give place to new data from the Data Sources. We have the OLAP Server that can be implemented in either of the following ways.

- By Relational OLAP (ROLAP), which is an extended relational database management system. The ROLAP maps the operations on multidimensional data to standard relational operations.
- By Multidimensional OLAP (MOLAP) model, which directly implements the multidimensional data and operations.

The Storage Area

The storage area is where the persistent data reside. Except for the Operational Reporting Data Repository, the other repositories in this area grow constantly, new data is added, but no old data is deleted, this area contains the enterprise memory. The Operational Reporting Data Repository is composed of a federation of replicated databases from the Operational Systems databases.

The function of this component is to provide the data to produce the operational reports of the enterprise; it exists to avoid disruption of the operation of the Operational Systems due to resources required to produce the operational reports. The Operational Data Repository only contains data for a limited period. The Unstructured Data Repository is probably the biggest storage area; it contains different types of documents. The Stream Data Repository contains data from Stream Data Sources, like real-time sensors. There are high volumes of data arriving at high frequency, but we reduce the space used for this kind of data by identifying patterns, variations and tendencies and instead of saving the raw data, we save the new processed results. The Enterprise Data Warehouse is a mix of normalized and de-normalized data structures that contain the memories of the enterprise; this is implemented using a relational database(s).

The Data Marts are de-normalized data structured that has been pre-processed and structured to serve as the high-performance source for the Business Intelligence and Decision Support Systems.

The Presentation Layer

Business intelligence Application and Decision Support Systems are not included in the Presentation Layer because they are mainly compound of the component included in this section, making them implicitly part of the Presentation Layer. The presentation layer is the front end of the Data Warehouse; it is composing of all the tools required to obtain insight from the data stored in the Storage Area of the Data Warehouse Architecture, from simple reporting tools to complex data mining tools. This layer holds the query tools and reporting tools, analysis tools and data mining tools.

The Hardware Infrastructure

The hardware infrastructure includes all the required equipment's (i.e., Servers, Clusters, storage devices, etc.) necessary to implement all the other Data Warehouse components.

Phases of Data Warehouse

Table 1

Phase 1 - Data Assessment		
Key Activities	**Key Participating Groups**	**Deliverables/Outputs**
• Identification of data sources • Run system extracts and queries • Conduct user interviews and awareness programs on data flow process • Review scope and validation strategy • Create work plan and milestone dates	• Data leads • End / Business users • Program sponsors	• Scope document • Strategy document • Work breakdown structure with milestone dates
Phase 2 - Data Cleansing		
Key Activities	**Key Participating Groups**	**Deliverables/Outputs**
• Identify data cleansing needs and expectations • Create data prep worksheets • Clean up source data in current system • Format unstructured data in other systems • Run extracts and queries to determine data quality • Create metrics to capture data volume, peak hours and off-peak hours	• Data team • Client Information Support team	• Cleaned/changed source data that increases the success of automated data conversion • Control metrics and dashboards
Phase 3 - Test Extract and Load		
Key Activities	**Key Participating Groups**	**Deliverables/Outputs**
• Create/verify data element mappings • Run data extracts from current systems • Create tables, scripts, jobs to automate the extraction • Address additional data clean-up issues • Execute application specific customizations • Run mock migrations • Load extracts into the new system using ETL tools or SQL loader with bulk loading functions • Conduct internal data validation checks including business rules and referential integrity checks • Report exceptions to client team • Perform data validation	• Data team • Client Information Support team • DBA team	• Extracts from source system • Data modules, jobs, scripts • Application loaded with converted data • Exceptions, alerts, and error handling control points
Phase 4 - Final Extract and Load		
Key Activities	**Key Participating Groups**	**Deliverables/Outputs**
• Run final extracts from the current system(s) • Execute specific customizations on target database • Execute application specific customizations • Run pilot loads • Load extracts into the new system using ETL tools or SQL loader with bulk loading functions • Conduct internal data validation checks including business rules and referential integrity checks • Report exceptions to client team • Perform data validation	• Data team • Client Information Support team • DBA team	• Data Extracts from source system • Data modules, jobs, scripts • Exceptions, alerts, and error handling control points
Phase 5 - Validation		
Key Activities	**Key Participating Groups**	**Deliverables/Outputs**
• Prepare validation reports and data movement metrics • Review validation reports and metrics • Record count of verifications on the new system • Reconcile any exceptions or unexpected variations on the data. • Sign off on validation	• Data team • Client Information Support team • Business users	• Signed-off validation document
	Phase 6 - Post Activities	
Key Activities	**Key Participating Groups**	**Deliverables/Outputs**
• Complete data reports and cross-reference	• Data team	• Exception reports, cross-reference

files/manuals • Data correctness/quality reports • Target system reports and its correctness • Infrastructure capacity report and dashboards • Sign off	• Client Information Support team • Business users • Business sponsor	files/manuals • Infrastructure dashboards • Signed-off project closure document

Conclusion

This document focused the fundamentals that would benefit the reader to understand the notion of Data Warehouse, Data Marts, Data Warehouses Architectures and ETL processes along with the need and importance of Data Warehousing. For enterprises to keep up the pace with cut to cut competition and grow simultaneously, building a Data Warehouse is only the viable solution. As the size of the databases grows, the estimates of what constitutes a huge database continue to grow. The hardware and software that are available today do not allow to keep a significant amount of data online. The record contains textual information and some multimedia data. Multimedia data cannot be easily manipulated as text data. Searching the multimedia data is not an easy task, whereas textual information can be retrieved by the relational software available today. Apart from size planning, it is complex to build and run data warehouse systems that are ever increasing in size. As the number of users increases, the size of the data warehouse also increases. These users will also require accessing the system with the growth of the Internet; there is a requirement for users to access data online.

Hence the future shape of the data warehouse will be very different from what is being created today.

References

1. https://www.tutorialspoint.com/dwh/dwh_future_aspects.htm
2. http://research.ijcaonline.org/volume30/number6/pxc3875088.pdf
3. http://naeemgik.blogspot.com/2014/
4. http://wideskills.com/data-warehousing/introduction-to-datawarehousing
5. https://www.ismll.uni-hildesheim.de/lehre/bi-12s/script/bi-00-Lection3_4.pdf
6. https://www.coursehero.com/file/p1oeqiq/Naming-conventions-could-be-different-attributes-for-data-items-could-be/
7. https://kctbsanalytics.wordpress.com/2016/08/15/integrating-heterogeneous-databases/
8. https://www.coursehero.com/file/p5qga34s/Standardization-of-data-elements-forms-a-large-part-of-data-transformation-You/
9. https://www.etl-tools.com/datawarehousing/loading-data-into-the-data-warehouse.html
10. https://www.coursehero.com/file/19843518/data-mining-and-warehousing/
11. https://arbime.wordpress.com/2012/08/17/the-data-warehouse-architecture-2/