Original Research Articles

Researchers

Sweety Patel

Department of Computer Science, Fairleigh Dickinson University, NJ-07666, USA

Email-

sweetu83patel@yahoo.com

What is Data Warehouse?

Abstract:

Data warehouse is a database of unique data structure that allows relativity quick and easy performance of complex quires over large amount of data. A classical production information system is preliminarily adapted to input data. It allows the company to be operational and run smoothly and that means mostly data entity stored data in the production system require the data to be administered on those data should be able to extract useful information from large amount of data. The administrative structure of the company should be able to extract useful information from the large amount of the data, which will be used as planning and decision making for the stored data. The mode provided by the data warehouse is faster & easier access to information to view and analyse large amounts of data in which time measures the reach in seconds or minutes.

Keywords: Data Warehouse.

Introduction:

An organized system of enterprise data which is derived from multiple data sources, and it is designed for decision making as its primary goal in the organization.

As per the Bill Inmon, Father on Data Warehouse:

It has characteristics as below.

It is integrated.

It is time variant.

It is non-volatile.

Purpose of the Data Warehouse

It is work as foundation for decision making process, as for taking considering organization (data related) we preliminary focuses on the DW. Fig. 1 shows a process view of the Data Warehouse.

It makes information easy to accessible as we can generate reports, like Operational & Enterprise report from the data warehouse. DW is not only serves analytics, data warehouse is not only serve reporting and analytics as data warehouse can be used as for operational reason like a contact center

executive looking at customer single view, while doing up sell or cross-cell to customer.

DW does not provide services

With Business Intelligence(BI) only With time being , DW as being to platform which can supping metadata and master data management initiatives, and not only what only we traditionally know as BI. DW is not mandatory to achieve some of the objectives like repetition on key performance indicators to repotting for enterprising. Data mining, by picking up data directly from DW is also done by the ERP in DW.

Keeping analysis / Reporting and Production Separate

If we run a query on production system, we find that it will look all tables and take most of the resources, as it will assessing a lot of data doing a lot of calculations. This may result in halt in production system. If hundreds of such queries run same time with this resource, then how much resource allocation is done because of this how much work is done slowing make system efficiency low as with performance.

Reporting and analysis work typically access data across the database takes whereas production work typically access specific customer product or channel record at any period of time. That's why it is important to have information generation work to be done from offline platform, here Data Warehouse. Production system becomes separate from the new production system for making out reports and analysis.

Integration of Information System from Multiple System

IT IS WORK AS SINGLE POINT OF SOURCE OF INFORMATION.

 Example: If we consider ATM system then all different system required to integrate and from them only data is produced. This will make query extremely slow and it require working on intermediate table to retrieve data to represent on report for analysis. Moreover it is not reliable result production system as it needs to be synchronizing in the database with particular point of time. Sometimes it synchronizes with end of the day by running batch runs.
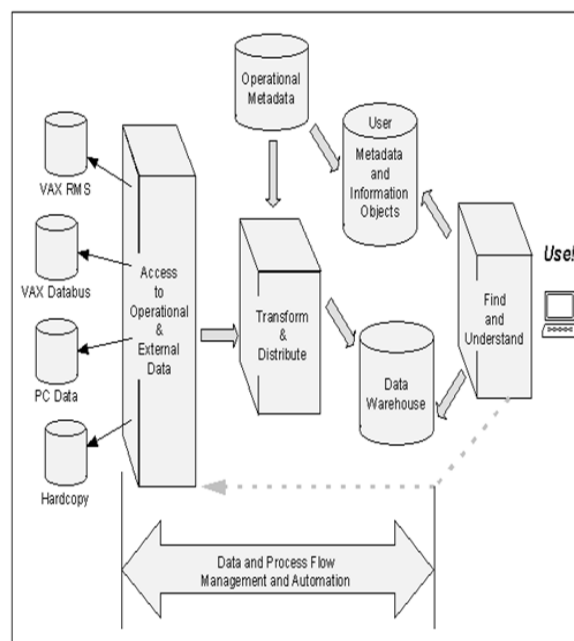

Fig.1 Data Warehousing - A Process View

Data consistency and quality is the data warehouse purpose. Organization runs with heterogeneous system may lead to data integration for analysis purpose each of this system contain different format of data representation & that's why we need to put them with consistent format. By bringing the data from those disparate sources at a common place, anybody can effectively may undertake to bring the uniformity across the whole organization.

High response time is required - production database are turned to expected transaction load. Even if the query is run on the offline database warehouse may take lot of time to retrieve back of data as production line and once needed to decrease a time in query teaching then indexing normalization all these aspects must be considered while designing data warehousing.

***High response* time - normalized data vs. dimensional modeling** - production / source system database are typically normalized to enable integrity and non-redundancy of data. This type of design is fine for the transaction, which involves few records of data at a single point of time. However for large analysis and mining queries the response time is normalized database will be slow as there are many joins on that data. So, dimensional modeling is required to put a data into the system effectively

Data Warehouse Provide an Adaptive and Flexible Source Information

It is easier for the user to define the production work and functionalities they want but it is difficult for the developer to make out analysis for the particular data from the databases. Analysis needs to keep a record of changed data at a particular time point & data warehouse has capability to adapt it quickly to the changing requirements.

Establish the foundation for the decision making support decision process of analysis, data mining, forecasting, decision modeling etc. By having common platforms for the data which make out of decision procedure easy and fast to get a response back from the data.

Logical Versus Physical Design in Data Warehouse

Once organization has decided to build a data warehouse we have to define business requirements and agreed for the scope of the application and created a conceptual design, not it is time to translate user requirements to system deliverable and for doing so, we create the logical and physical design for the Data Warehouse.

We need to define as:

The specific data content
Relationship must within & between groups of data.
Data warehouse must be supported by system environment.
Transformation of data required.
The frequency with which data is refreshed

Logical design of data warehouse

Logical design is more conceptual. Logical design deals with the logical relationship between the objects for logical design of data warehouse is represented by ER modeling technique. ER modeling involves identifying the entities is called important objects and attributes (properties about objects) and the relationship among them. An entity is a chunk of information which maps to a table in the database while the part of entity is attribute which maps to a column of table into the database.

A unique identifier can be used to make sure the data is consistent. Logical data modeling is the exercise to document and define the relationship between the data elements it involves many things like:

Indenting entities from the given set of business environments.eg: customers, orders identifying how specific instants of each entity are different from other instances like logical keys.eg Customer's_id, order_number

Grouping the different attribute that refers to same entity

Finally documenting or making business rules (relationships) between the entities. (E.g. one order must be placed by one customer & customer may place one or more orders). Consideration like ,logical data modeling does not represents any physical data module which shows like how data is stored and does not attempt to anticipate or correct any performance issues that may arise during implementation.

Physical design of data warehouse

Physical design deals with the effective way of storing and retrieving the data from the data warehouse. In the physical design the logical design needs to be converted into a description of a physical database structure using proper mapping of logical database to physical database.

Physical design involves creation of the database objects like tables, columns, indexes, primary key, foreign keys, views, sequences etc. Task such as how data will be stored all related business rules to manipulate the data are stored into the physical data modelling. Consideration for how the data needs to be accessed combined joins and the performance characteristics of the intended development environment will be documented. Decision about where to place the data (same database, different databases or different servers) as well as participating archival purging plans, has to be done within the constraints of the business requirements.

Conclusion

A data warehouse is a relational data base that is designed for query and analysis rather than for transaction processing. It usually contains historical data derived from transaction data but it can include data from other sources. Transaction workload is separated from analysis workload and enables organization to consolidated data from several sources. In addition to relational database a data warehouse environmental include an extraction transportation transformation and loading solution (ETL), an online analytical processing (OLAP) engine analyst is tool for client and other applications that manage the process or gathering data & developing it to business users.

References

[1] Nong Ye, The Handbook of Data Mining (Lawrence Erlbaum Associates, Mahwah, NJ. Publication, 2003).

[2] Jiawei Han and Micheline Kamber, Data Mining:Concepts and Techniques (Morgan Kaufmann Publishers, University of Illinois at Urbana-Champaign).

[3] Bharat Bhushan Agarwal and Sumit Prakash Taval, Data Mining and Data Warehousing (Laxmi Publications, New Delhi - 110002, India).

[4] Ralph Kimball,Joe Caserta, Data Warehouse. ETL Toolkit. Practical Techniques for. Extracting, Cleaning,. Conforming, and. Delivering Data (Wisely Publication,Inc).

Author Details:

Sweety Patel
Department of Computer Science, Fairleigh Dickinson University, USA