



Design of Data Warehouse for E-Commerce

Mohammad Sarfaraz

MBA IT, Symbiosis Institute of Compute Studies and Research, Atur Centre, Model Colony, Pune-16 Maharashtra India

Pravin Metkewar

Associate Professor, Symbiosis Institute of Compute Studies and Research, Atur Centre, Model Colony, Pune-16 Maharashtra India

ABSTRACT

E-commerce is one of the most rapidly expanding area in recent time, the everyday huge size of data is generated. E-commerce companies are collaborating with big IT companies for researching & designing the data warehouse for the E-commerce. Data Warehouse is core riggings for analysing data as data warehouse as it stores huge amount current and historical data and helps E-commerce companies for better decision making. In this paper we discuss requirement analysis design of data warehouse, we also discuss the dimension table for data warehouse, snowflake schema and star schema for the data warehouse will helpful in developing the data warehouse in E-commerce, and also discuss some problem and purposed a solution for that.

KEYWORDS : Data Warehouse, E-commerce, Snowflake schema, Dimension table, Star schema.

1. Introduction:

In recent time, the growth of the e-commerce is increased and with that, the need for storing data for the further analysis is also cultivated very rapidly. Data warehouse is a data depository which will store huge amount of data from various data sources. Electronic commerce offers for allotment of business information, conducting business transactions, and retaining business relationships by means of telecommunication systems. According to ASSOCHAM the e-commerce industry in India is likely to be worth \$38 billion by 2016, a 67% escalation over the \$23 billion earnings of 2015 and according to marketing research firm eMarketer e-commerce sales in India are projected to grow from \$38 billion in 2016 to \$55 billion in 2018 and same is happening across the world. Thus, business analysis of e-commerce will grow into a captivating trend for reasonable advantage.

A data warehouse collects organize and makes information out there for the aim of study so as to allow management the flexibility to access and analyse info regarding its business. this sort is information is often referred to as informational data. The systems accustomed work with informational information area unit said as the on-line analytical process.

OLTP systems that produce operational information, operational information focuses on transactional functions. This information is an element of the company infrastructure. it's elaborate, non-redundant and updateable.

OLAP is that the technology that allows shopper applications to with efficiency access information organized by information market and data warehouse. Information warehouse, offer an info organized for OLAP instead of OLTP, will solve OLTP issues.

The addition to e-commerce to the data warehouse conveys each complication and innovation to the project. E-commerce is already acting to unify once standalone group action process systems. These group action systems like the sales system, the promoting system, the inventory system, and therefore the shipments system all got to be accessible to every different for the e-commerce business to perform swimmingly over the web. Additionally, to typical business aspects like customers, sales, payments, and shipments, e-business currently have to analyse extra factors distinctive to the online situation. Another worry in design a data warehouse within the e-commerce atmosphere is once and the way we have a tendency to capture the info. several attention-grabbing items of knowledge may well be mechanically captured throughout the navigation of internet sites.

In this paper, we review data warehouse, design and some issues in building a data warehouse for e-commerce environments. This paper offers a short description of methods that talk about the data warehouse design and Implementation for E-business.

2. Data Warehouse Architecture

Bill Inmon defined data warehouse: "A (data) warehouse is a subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management's decision-making process."

Harjinder and Rao (1996) argue, "that data warehouse is a running process that agglutinates data from heterogeneous systems, including historic data and external data to attend the necessity of structured queries, analytical reports, and decision support."

Data warehouse is a huge depository serving as a centralized repository of all data generated by all departments and units of an outsized organization. Advanced data processing package is needed to extract important info from a data warehouse.

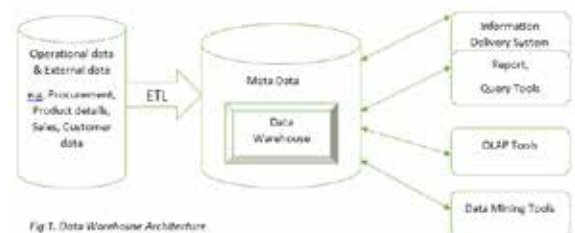


Fig 1. Data Warehouse Architecture

Fig 1. Data Warehouse Architecture

Data source

This represents the various data sources that feed knowledge into the data warehouse. the data sources are often of any format -- plain document, computer database, alternative sorts of info, Excel file, etc., will all act as a data source.

ETL

This is where information gains its "intelligence", as logic is applied to rework the info from a transactional nature to a methodical nature. This layer is additionally where data cleansing happens. The ETL style section is commonly the foremost long innovate a data warehousing project, and an ETL tool is commonly employed in this layer.

Metadata

This is where info regarding knowledge keep within the data warehouse system is stored. A logical information model would be an associate example of one thing that is within the metadata layer. A metadata tool is commonly accustomed manage information.

Data Mining

It's the method of discovering important new correlations, patterns,

and trends by mining enormous amounts of information keep in the data warehouse so that this can help the organization in better performance

3. Dimension table, Snowflake and Star Schema for e-commerce

The diagram signifies specific characteristics inside one dimension. The diagram additionally models varied hierarchies among attributes and properties like slowly ever-changing attributes and cardinality. The basic ideas of dimensional modelling include A fact, dimensions, and measures. A fact may be the group of connected data things, consisting of measures and context data. It usually represents business things or business transactions. A dimension may be an assortment of knowledge that describes one business dimension. Dimensions confirm the discourse background for the facts; they're the parameters over that we would like to perform OLAP. Alive may be a numeric attribute of a reality, representing the performance or behaviour of the business relative to the size.



Fig 2. Dimension Table for Product details.

In this dimension table which was purposed by Il-Yeol Song and Kelly LeVan-Shultz of product details will store all the details related to the Product e-commerce companies are trying to sell. This table will store the data like Brand, Manufacture, Product no.& description, product physical dimension weight etc. Similar tables will be there for likes of customers, Seller, Bill/Order etc.

The primary characteristic of a dimensional model may be a set of elaborate business facts enclosed by multiple dimensions that describe those facts. once realised in an exceedingly info, the schema for a dimensional model contains a central truth table and multiple dimension tables. A dimensional model might turn out a star schema or a snowflake schema.

Star Schema is that the basic structure for a dimensional model. it's one giant truth table and a group of smaller dimensions' tables organized in an exceedingly radial pattern around the central table. The snowflake schema is that the results of mouldering one or additional of the size. The many-to-one associations among traits of a dimension will separate new dimension tables, forming a hierarchy. Xudong Chen proposed the star schema which has basic star schema for e-Commerce for sale contains, customer table, Manufacturing table, Product table, Seller table and Time&Date table are a dimensional table.

Both star and snowflake schemas area unit dimensional models; the distinction is in their physical implementations. Snowflake schemas support simple dimension maintenance as a result of they're additionally normalized. Star schemas area unit easier for direct user access and sometimes support less complicated and additional economical queries.



Figure3. Physical Star Schema for E-Commerce Sales

This Star schema of E-commerce sales will have a Sales Fact will have all the Facts related to Sale, the tie will have different field related to it. And all the other dimension table will have all related filed to these dimension table



Fig.4. Snowflake schema

4.Problem Data quality

Data quality is a big challenge in data warehouse, as various data of e-commerce like product name, seller name, price, product specification, product description and related information can have redundant and inconsistent data and to maintain this data is big challenge.

Performance

Performance is another challenge in data warehouse for The E-Commerce the data mining becomes a very slow process as millions of records with hundreds of columns are stored every second and to retrieve these certain data from these millions of records takes time and therefore performance becomes big issue data warehouse

Data cleansing and processing

Data cleansing and processing is another challenge in data warehouse for e-commerce even as the data collected in digital form but there can be some data related issues like a rose missing data, mistakes, different types of data type in same column, corruption of data etc. for this data cleansing and processing is required for better performance and efficiency there is a certain need of algorithm to solve these problems.

Lack of visualisation tools integrated with data warehouse

There are lack of visualisation tools RBI tools which are integrated with data warehouse itself which help business user to get detailed information about user and their purchasing pattern which will have the business user for better understanding and taking decisions according to that.

5.Proposed Solution

To resolve these problem, I propose a solution by using AI in data warehouse. Artificial Intelligence will be integrated seamlessly across the whole data warehouse to e-commerce.

In data warehouse architecture the AI component will be integrated with the across the whole data warehouse so AI can be integrated seamlessly across the warehouse. It will help to perform different process with less effort.

Before ETL Process the AI check the data and if data is correct and relevant to that than it will store the data in the warehouse and if the missing or form is different the AI will clean and transform the data and the AI will store in the data warehouse according to the relevancy of data.

For example: if in one of the record the "Gender" of the customer is omitted in the database, then AI will commences working then Artificial Intelligence will look for the "Name" of that records, through its Intelligence it will predicts its gender according to that and put the

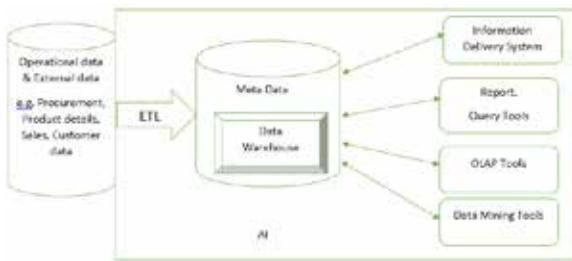


Fig 5. Proposed Data Warehouse Architecture

AI also help the user during processing and extraction of data. As if the user wants to extract certain data user only need to enter the scenarios and AI will select the desired attribute and select the data according to that scenario. This will make data warehouse more useable to the generic user without much technical knowledge.

For example: User will enter scenario like seller wants to know which product is selling best in fourth quarter. The with AI system will look into the scenario and then it will sell the required attributes and take the records from the desired scenario like in this case AI will take the product name, product id, sells no., month etc. will be selected with top selling product in last three months.

6. Conclusion

In this paper, I reviewed the look of data warehouses for e-commerce setting. we have a tendency to stated demand analysis, style problems. I even have presented dimension table structures, a base star schema, and a snowflake schema for e-commerce environments, and also discuss some problem of data warehouse for e commerce and proposed Artificial Intelligence as solution to resolve these difficult.

The fundamentals to create data warehousing is data design. The business users distinguish what data they have and the way they require to use it. target the users, confirm what data is required, find sources for the information, and organize the information during a dimensional model that represents the business desires. Dimensional modelling is that the groundwork for data warehouse design. A dimensional model could help create a star schema or a snowflake schema for e-commerce. The representation of a dimensional model contains a central truth table and multiple dimension tables.

REFERENCES

1. Berson, A., S.J. Smith, Data Warehousing, Data Mining, and OLAP, McGraw-Hill, 1997
2. Adamson, C., and Venerable, M., Data Warehouse Design Solutions, John Wiley, 1998.
3. Data Mining in E-Commerce: A CRM Platform, Lipsa Sadath
4. Data Warehousing and E-Commerce Author: William J. Lewis, Publisher: Prentice Hall, 2001 , ISBN 0130911542
5. Lessons and Challenges from Mining Retail E-Commerce Data : RON KOHAVI, LLEW MA-SON, RAJESH PAREKH, ZIJIAN ZHENG.
6. E-Business Data Warehouse Design and Implementation: Xudong Chen
7. Data Warehouse Design for E-Commerce Environment: Il-Yeol Song and Kelly Le-Van-Shultz
8. Data Warehouse Design Methodologies for Electronic Commerce: Dinh Thuan Nguyen and Tung Hoan
9. Chaudhuri, S., U.Dayal, "Database Technology for Decision Support Systems", Computer, Vol 34, No12, pp48-65, 2001.
10. Pedersen, T.B., C.S. Jensen, "Multidimensional Database Technology", Computer, Vol 34, No12, pp40-46, 2001.