ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 7, Issue, 08, pp.19127-19133, August, 2015

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

REVIEWARTICLE

STUDY OF THE RELATIONSHIP BETWEEN THE WEB AND DATA WAREHOUSES (DATA TRANSFER BETWEEN THE WEB AND DATA WAREHOUSES)

***Atheer HadiIssa Alrammahi**

IRAQ- DIWANIYAH- College of Veterinary Medicine, University of Al-Qadisiyah

ARTICLE INFO

Article History:
Received 15th May, 2015
Received in revised form
18th June, 2015
Accepted 03rd July, 2015
Published online 21st August, 2015

Key words:
Click stream data,
Data warehouse,
e-business environment,
ODS,
Web environment.

ABSTRACT

We have seen that the Web environment supported by the data warehouse a variety of ways, the link or surface to transfer data from the Web to the data warehouse is somewhat simple. Web data is contained in the records, the records manager nourish software information on flow, so the software is director of the amendments and summarize the data and re-organized. And pass the data manager software to the data warehouse. The surface or interface to transfer data from the data warehouse to the Web where a little complicated, that the data transmitted from the data warehouse to be operational data store (ODS) in the portfolio recording the ODS. If the ODS becomes the only point of contact between the Web and data warehouse environment for the purposes of the flow of data from the data warehouse to the Web. The reason for this is simple: that the ODS is able to ensure that operations at the Internet address system and quickly, and that are essential to the efficient processing Web. In addition, the data warehouse is a secure place for the huge amounts of data that can be downloaded from the Web environment and stored. Data Warehouse also provides a central point so that the company can integrate data and integrate with data coming from one or more of the Web sites to a single common source.

Citation: Atheer HadiIssa Alrammahi, 2015. "Study of the relationship between the web and data warehouses (data transfer between the web and data warehouses)", *International Journal of Current Research*, 7, (8), 19127-19133.

INTRODUCTION

One of the technologies most widely discussed is the Internet and its associated environment, the World Wide Web. Web technology has broad popular support among entrepreneurs and technicians alike. The web environment is owned and managed by the corporation. It may be outsourced. But in most cases, the Web is a normal part of computer operations, and is often used as a center for the integration of business systems. When it is outsourced, it becomes much more difficult to capture, retrieve and integrate data from the web with processing companies. An interaction occurs when the Web environment creates a transaction to be executed an order from a customer, for example. The transaction is formatted and sent to corporate systems, where it is processed like any other order. In this regard, the Web is not just another source of transactions falling within the business. However, the web interacts with other corporate systems as well, through the collection of Web activity log. The Web is raising the number of employees, the average of use and difficulty of its applications.

Corresponding author: Atheer HadiIssa Alrammahi,
College of Veterinary Medicine, University of Al-Qadisiyah

The use of this medium as an interface to information systems (IS) of the organization and its exercise is familiar. With the enlargement of knowledge and potential of users, the need to understand and assemble the difficulty of users is increasing quickly. In addition to adapting the system, there are additional profits connected with monitoring the use of an IS.It backs up the estimation of the system next to its original specifications and goals.

In contrast, permitting the expansion of personalization strategies, supports boost system performance, helps marketing decisions, aids in the detection of business opportunities that might otherwise go unnoticed and can help increase system security (Lawrence Steve and Lee Giles, 1999). Over the last twenty years, the benefit to analyze the data has increased considerably, because the competitive advantages that information can provide the decision-making process. Today, one of the keys to survival in the business world is to be able to analyze, plan and respond to changing economic circumstances as soon as possible. Several organizations have billions of bytes of data, but they suffer from multiple problems that make it difficult to leverage data: the data are spread across various computer systems, data from different sources are inconsistent, the data be found too late, etc. to resolve these problems, new concepts and tools have evolved into a new information

technology known Data Warehousing. Data Warehouse Projects (DW) are costly: they often need several years to properly implement and require millions of dollars of hardware, software and consulting services. Data Warehouse Projects (DW) are expensive: they often require several years to properly implement and need millions of dollars of hardware, software and advisory services. Sales DW and associated products continue year after year, more and more. The market for DW tool has reached 7.9 billion in 2003 and grew by 11 percent this year, more than three times the growth rate of the previous year (Hui-Huang Hsu, 2008). In the meantime, according to (Inmon, 2002), the online analytical processing (OLAP) market increased from $ 1 billion in 1996 to $ 4.3 billion in 2004 and demonstrated valued at 15.7 percent growth 2004.Well as much progress has been made in the field of DW, there now is no standard method or data model for the design of DW. Moreover, several reports indicate that about 40-50% of projects fail DW (Hui-Huang Hsu, 2008 and Franconi, 2002). Consequently, a new method based on DW standards can help develop DW. This article aims is to study the relationship between the Web and data warehouses by knowing way transmission of data between the Web and data warehouses and vice versa, as well as by supporting the e-business environment.

Data warehouse and Web

Internet and the environment associated with a more extensive areas of technology is. The website www embraced by Wall Street, which is considered the base of the new economy. The technology Web (www) enjoys broad popular support among businessmen and specialists, and despite the lack of clarity in the relationship between the Web and data warehouse, but there is a very strong relationship between the company's data warehouse and Web sites. In fact, the data warehousing believes successful operations of the business environment based on the Web. The owned and managed by the company's Web environment. In some cases, the Web environment from outside the company, but in most cases, the Web is a natural part of the computer operations, often used as a hub for the integration of business systems. If the Web environment from outside the company become more difficult to control, and recovery, and also to integrate web data with the company's operations.

In this study, we focus on the interaction of Web environment with the company's systems is done in two basic methods: First: The first reaction occurs when Web-based network and business operations that need to be implemented at the request of the consumer. For example: commercial transactions that attend and sent to the company's systems, where treated much like any other request. In this regard, the Web is simply another source of business operations and access to the business, but the Web to interact with systems company in another way and that by pooling activities in the Web log.

The Figure 1 shows the image of Web activities and the location of these activities in the Web log. The log contains the so-called (click stream) the flow of data from the Web, and every time it is clicked by the Internet user to go to a different location called click stream records (Inmon, 2005).

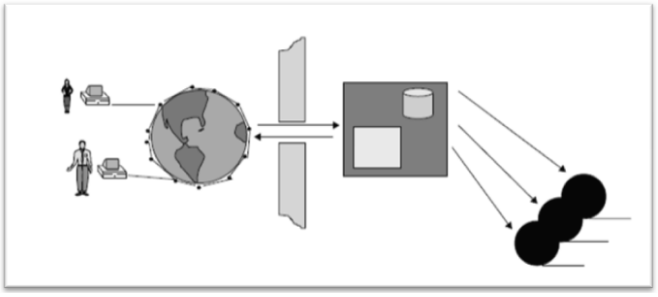

Fig. 1. The activity of the Web environment is spun off into Web logs in records (Inmon, 2005)

The user of various products of the company, which records what the user is looking for and what they are buying and what the user which the user thinks of purchase. All of this is classified and collected in the data warehouse, And that each of them has equal importance. Even things that do not are looking for does not buy can be identified by the word "click stream data" which is key to understanding the current and the Internet user. By understanding the desire of the Internet user, business analyst can understand firsthand how producing and advertising and promoting, in order to receive by the public, they are more specific and more powerful than before, but the technology requires a strong interaction occurs is tacky, There are some obstacles to understand the data that comes from the business environment: (Inmon *et al*., 2008). For example: the Web to generate data to be a very low level of detail, in fact, the decline of this level is not suitable for analysis or to be the entrance to the data warehouse. In order to make useful data warehouse for analysis and data flow, the data record that must be read and be filtered. The figure (2) "click stream data" record shows that it passes through software called (Director granulated) before it enters the data warehouse environment.

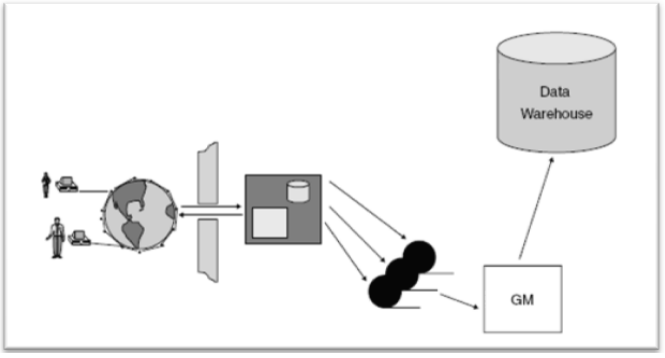

Fig. 2. Data passes through the Granularity Manager before entering the data warehouse (Inmon *et al*., 2008)

Many of the treatment occur in granulated Director, which read "click stream data" and works as follows:

1-Edit external data.
2-External recording of a single multi happen, and the registration of click stream record.
3-Edit incorrect external data.
4-Remember that to be a single Web environment, and a private key data that needs to be used in integration with other company data conversion.
5-Summarize data.
6-Data collection.

Approximately 90% of the raw data flow to be rejected or summarized and passing through the granular director (Inmon, 2005). The way in which the data back to the Web environment are not directly or completely explicit. The data warehouse simply do not pass data directly. To understand why there is little direct operations of the data in the data warehouse, it is important to understand why you need a Web environment to the data warehouse data in the first place. The Web environment that you need this kind of data, which is in the data warehouse in which the combined company's information be (Combined).For example: Suppose that a Web site devoted to selling clothes, now assume that the business analyst has decided that it is appropriate for the consumer to become a clothing and other consumer goods, such as garden tools, sports equipment, travel goods, jewelry. The analyst may decide to start promoting special luxury women's clothing and jewelry, but where can the analyst to find data on women consumers who bought jewelry in the past? He or she is naturally looking at the data warehouse where historical information about consumers be present.

In another example: Suppose that a Web site devoted to selling cars, the analyst wants to know who had bought the car brand that the company be sold, where historical information about this selection exist? Of course in the data warehouse (Abello *et al*., 2000). The data warehouse provides the base historical information integrated, and be available to Business Analyst, said that the link between the data warehouse and the web shown in Figure (3), where the data out of the data warehouse and shows (ODS), so that they are available for direct operations from the Web.At first look it seems secluded (ODS) is located between the data warehouse and Web, there are some good reasons for this situation.

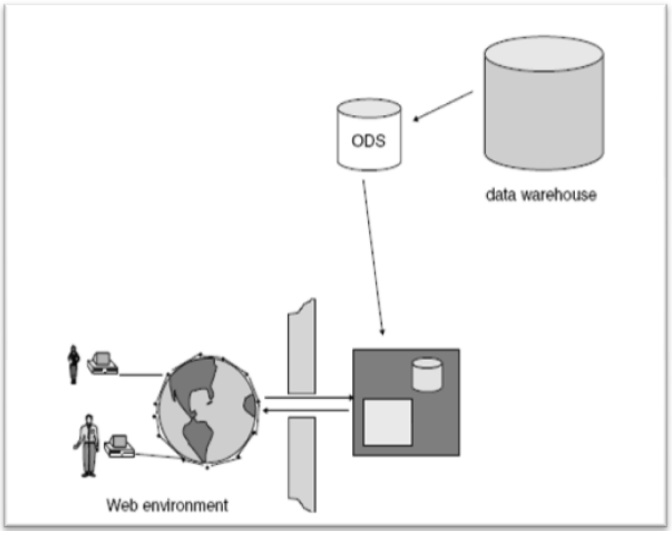

Fig. 3. Data is passed to the ODS before it goes to the Web. (Abello *et al*., 2000)

The (ODS) is a hybrid structure so that has some of the qualities and attributes of the data other than the operating system warehouse. ODS system contains an integrated database that can support decision support DSS systems, the ODS supports high-performance commercial transactions operations, the last property system ODS make him something of value for the network Web, when providing Web site ODS system, know that the Web environment will reach post on the subject certain parts of the second. This speed in responding to make it possible for network Web to accomplish transactions accurately time. If the web is connected directly with the data warehouse, it is possible to take a few minutes to reach the answer from the data warehouse, and in the world of the Internet users are more sensitive to response time and therefore could be this is not acceptable and unacceptable, more clearly that the data is not designed warehouse to support time response on the Internet (Abello *et al*., 2000). In any case, the ODS system is designed for this purpose. Therefore, the direct access to the Web environment is through the ODS as it seems in Figure (4).At first look it seems that there is a lot of excess data between the data warehouse and ODS after all this is a feeder of ODS data warehouse. But in fact, there is very little overlap between the data warehouse and ODS. The data repository containing detailed data on commercial transactions. While ODS contains a simplified overview of the data (or the so-called side-data). To understand the differences between Peep Simplified (side) for data and detailed for commercial transactions is shown in Figure (5).

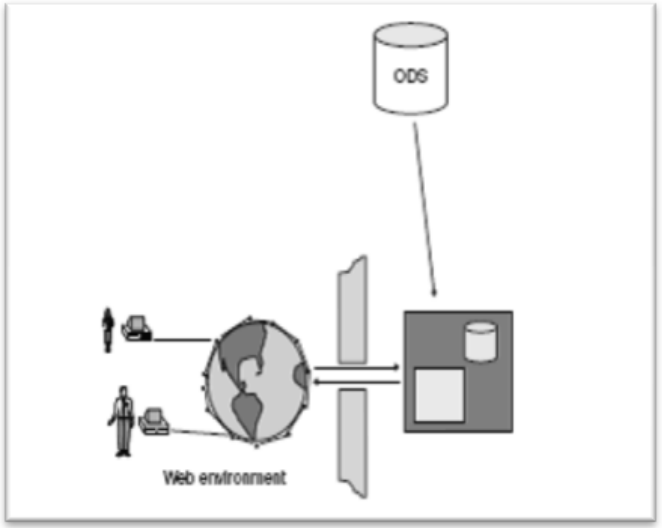

Fig. 4. The ODS provides fast response time (Inmon *et al*., 2008)

The data warehouse contains all the data patterns of transactions on the previous business transactions between the consumer and the company. The detailed business transaction data includes information on the following:

1. Search for bicycles for men.
2. Search red swimwear for women.
3. Purchases of clothes women blue swimsuit.
4. Search for tight clothing.

The data warehouse maintains a detailed log, interaction with consumer and business transactions carried out by the consumer with the business, regardless of the source interactions. The interaction gets over the Web, and through the catalog request when buying from retail store and so on. Typically it records the reaction time and place of interaction and the nature of the business transactions recorded in the data warehouse.

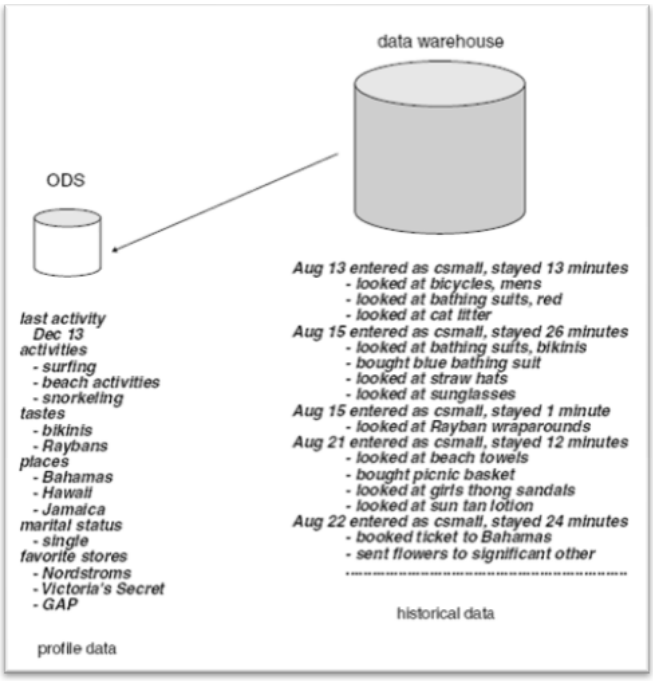

Fig. 5. The ODS and the data warehouse hold different kinds of data (Gray, 2005)

In addition, the data repository containing historical data, and commercial transactions in the data warehouse will return to business analyst and think that the benefit for a year or two years or whatever length of time it feels the benefit. The historical data reactants contain raw transaction data without explanation or interpretation of the data. On the other hand, the (ODS) is full of explanatory or interpretive data, if the data from the data warehouse to analyze and transform data summarized (side) (Gray, 2005)

Registration summarized data in the ODS

Figure (6) Summary record shows its form based on reading all the historical and integrated data in the data warehouse. The Archive Summary (lateral) contains all kinds of information, which formed as a result of reading and interpretation of commercial transaction data. For example, if the consumer is displayed in Figure (6) The manual data archive consumer in the following shows:

1. Someone who loves the beach, enjoy sailing, sunbathing, snorkeling.
2. Likely to travel to places such as Hawaii, Jamaica and the Bahamas.
3. Single.
4. Fond of shopping is likely to travel to duplicate and many places like Nord Storm (US chain stores for a comprehensive well-being).

In other words, it is likely that the consumer has the tendency and the tendency of the phenomenon to be in the data archive, see Figure (7) Note that the consumer may not go to Hawaii, Netherlands It predicts that the consumer may love to go there. And to create summarized data from commercial transactions data, it has to be a certain amount of analysis operations. The Figure (6) shows the reading of transactional data in order to address the summarized data. In Figure (6) integrated historical transactions detailed data read and analyzed to produce summarized data, the analysis works on a regular basis, dependent on the data and change the purpose of the business from behind the integrated analysis rate, and repeat analysis and update sequential data summarized perhaps it urges often once a day or seldom happens once a year, there is a wide changes in repetition. The analytical, both program hermeneutical analysis and predictive, dependent on the last consumer activities and any other information analytical program can be obtained. The program represents the analytical information to produce a perception or a personal plan for the consumer, the plan foresees the truth or perception.

Model specific factual information:

1. The date of the last deal with the consumer.
2. The nature of the final deal.
3. Recent purchase size.

There is other information that is not close to reality, predictive aspect of the analysis includes the following information "for example":

- Whether the consumer luxuriant.
- Consumer sex.
- Consumer age.
- Whether the consumer travel frequently.
- Where the consumer likes to travel.

The condensed consolidated archive contains a map for the consumer which is available in the ODS in the same moment.

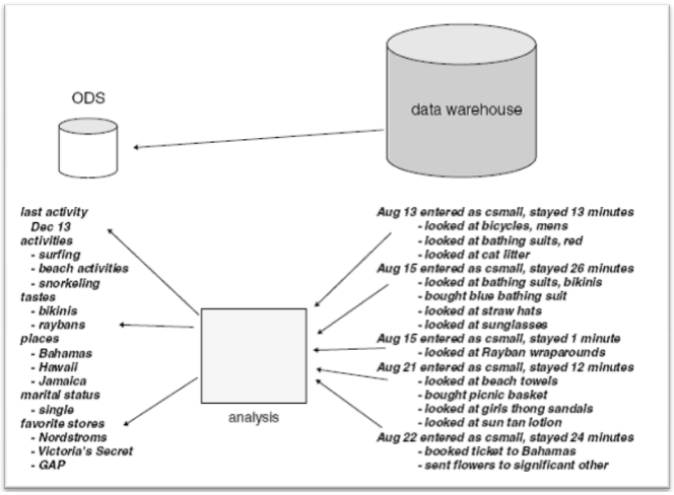

Fig. 6. Periodically the detailed historical data is read, analyzed, and loaded into the format required for the ODS. (Gray, 2005)

At that moment, the Web environment provides a time limit excellent and integrated response, and the image of interpretive consumer base made, and of course other information as well as the consumer are available in the data warehouse for information and ODS, pretty typical, shareholder stakes information, and product information and sales information and similarly also Available in a web analyst. Securing a good response time and re-analysis of the data is not the only roles played by the data warehouse environment in the web support,

the other role is the large amounts of data management. Web address create very large amounts of information even when the Web site director granulated used to maximize the efficiency, still exploding mountains of data. The first motivation for many Web designers to store data on the web environment it is the same, but very quickly the web become like the swamp, nothing works accurately, data becomes complex or tangle in everything at the entrance to the question and query, in storage, in catalogs, and in monitoring, and anywhere. The Figure (7) shows that the data periodically devote themselves in the data from the Web repository, and return periodically discharge from the data warehouse environment to flow (Abello *et al*., 2000).

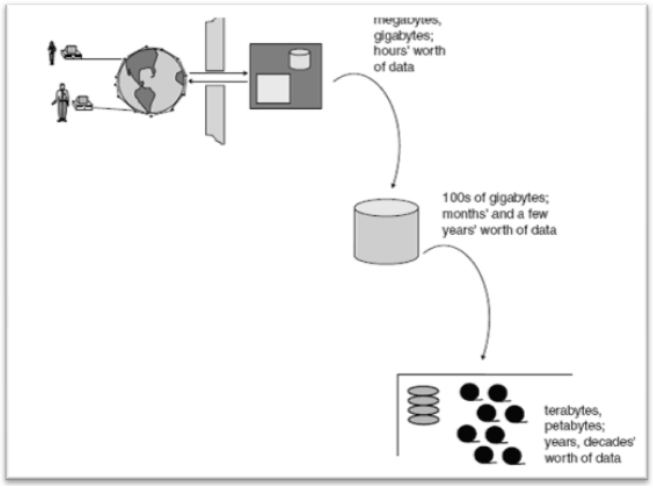

Fig. 7. Volume of data cascade down from the web to the data warehouse to alternative storage (Abello *et al*., 2000)

The director granulated "software" takes care of storing the data from the Web to the data warehouse on a daily basis or even hours, depending on the rate and amount of Web traffic data from a data warehouse stored monthly or a quarterly basis to the vast storage environment. Indeed, there is no management of the amount of data at any level of construction levels. Typically Web environment that have a financial value for the day data, while the data warehouse can get financial value annually. Typically vast storage components they stuck more than a decade of financial data.

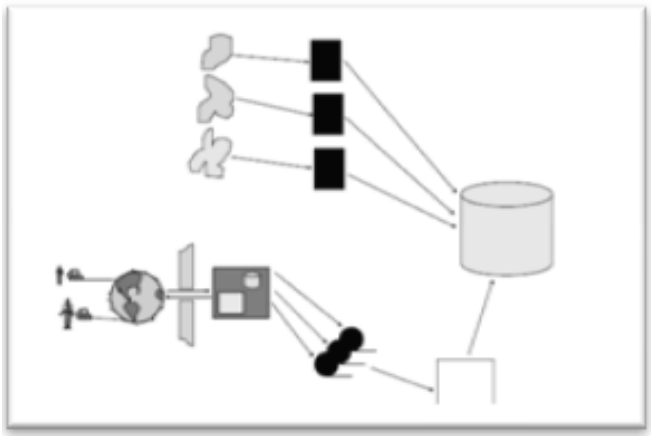

Fig. 8. The data warehouse is where web data is integrated with other data from the corporation (Thomas *et al*., 2009)

The data warehouse supports Web environment with data integrity, Figure (8) shows that the natural operating systems provide data warehouse counterproductive fed data so that becomes available in an integrated manner. The data come from outside granulated "software" director of the merge with previously integrated business data in the data warehouse. Indeed, the data warehouse becomes the only place where you can get a complete picture of all business data from the Web, from other systems, from anywhere. Another important area of the data warehouse is the ability to support multiple Web sites for large companies, multiple Web sites be real life, and their support will be essential to integrate and take the data from each site.

Support e-business environment

The e-business environment is the ultimate environment that supports by the data warehouse. The figure (9) refers to the support of the data warehouse environment for the Web. The common thing between the Web and data warehouse environment that is both simple and complicated at the same time. It is simple from the standpoint that the data transmitted from the data warehouse back and forth to the web environment, it is complex in terms of the movement of anything that last less than a simple shape.

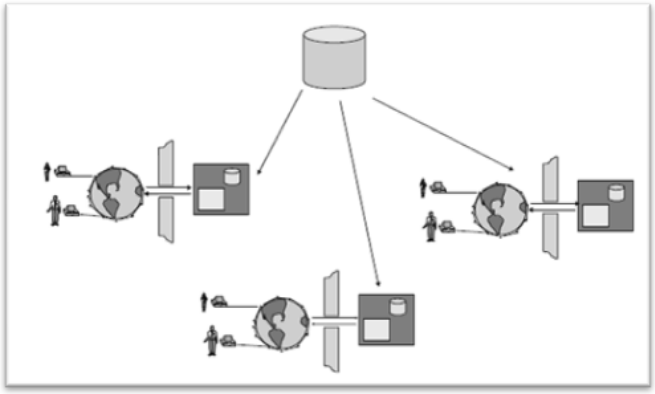

Fig. 9. The data warehouse can service more than one e business (Thomas *et al*., 2009)

RESULTS AND DISCUSSION

After studying the concepts of Web and its relationship to data warehouses and the handling of them come to the following conclusions, which describes a method for transferring data from the Web to data warehouses and vice versa.

Transmission of data from the Web to the data warehouse

The data in the Web environment combine very well and very low level of detail, and also a low level for use in a data warehouse, so the way the data from the Web environment to the data warehouse should be a policeman and that the level of detail should be high. The variety of things have been accomplished in Web environment before they become useful in a data warehouse are as follows:

- Strange data deleted.

Such as the case of adding data to each other.

- Data non-consecutive "attention".
- Data liberation "adjusted".
- Data continues.
- Data transformation "change".

Web-based data usually come through Web logs, which are formed in the web environment and by Experience There are about 90% reduction of the data that make up such as data from web explain: that the data coming from the web traverse the software, called "granulated director" in several modalities, director of software close to the "ETL" software in the movement of data from legacy to data warehouse environment. Data coming to the Web environment comes primarily from the process stream flows, which occur in the web environment. The data flow process is a good process to tell what might happen in the Web-based user sessions. To actually be useful, the data flow must be connected to the other prevailing trend of the data, which pass through the company's systems "in parallel."It is only when integrated with the company's data flow data consists There is a whole benefit of Web information we found (Duncan *et al*., 1999).

Transmission of data from the data warehouse to the Web

Web environment very sensitive to response time you cannot wait more than Millie second or two seconds when you need the information, if I had to make the web environment to wait more than that, the performance will weaken, "or at least". The Web OLTP environment similar to instant transaction processing environment. At least as far as it is sensitive to response time. That for these reasons that we have mentioned there is no broker or directly between the surface of the data warehouse environment and the Web, instead mediator between the environments that pass through the operational data stock in the company, so that resides in an environment similar to the data warehouse (Duncan *et al*., 1999). The operational data inventory is designed to secure the response time in milliseconds, and that the data warehouse is not the case, if the data passes through the data warehouse to the ODS. The data in the ODS waiting every time requests to enter the Web environment, then make Web requests and gets the information it needs quickly and also regular, If the ODS has a portfolio of different information for the data warehouse, which contains historical information detailed Moreover, the ODS contains extensive information about the company and correct, pass the data from the ODS to the web environment, it can be used any number of ways, and that the data can be used to show web users that occurs with dialogue, personal character or a direct dialogue (Thomas Connolly & Carolyn Begg., 2002)

Support Web site

The data warehouse provides a Web-based e-business environment and many of them are important capabilities:

1. The ability to absorb vast amounts of data, if the data warehouse provides storage of surplus machine such as alternative AC, next to the line and storage time Web data pass through the software manager, the data warehouse equipped to address effectively unlimited amounts of data. Data transmitted quickly from the Web environment to the data warehouse, and data sizes problem of Web environment is not a hindrance to performance or capability in a web environment.
2. Entrance to data integration, Web data itself abstract and somewhat inappropriate, but when the problem data in the Web merge data different company, the mixture becomes very effective and powerful, the Web data that are placed in the data warehouse able to be integrated, so the information many useful form.
3. The ODS Web operations and be able to have a good performance, and achieves good performance of the data warehouse.

The significant environment which secured data warehouse environment for web-based e-business, data warehouse and provides the backdrop for infrastructure needed by the Web in order to be successful (Thomas Connolly & Carolyn Begg., 2002).

Conclusions

Data ware houses contribute to a range of benefits:

- The web-based data pass "or subject to" strict / conversion / lose / practice to clean. Before it would be appropriate to put them or included in the data warehouse.
- The coming of the ODS or data warehouse data can be used to create or secure web designer requirements.
- Effective data ware houses in sectors characterized by a large heavily in contact with customers such as banks, insurance companies and commercial centers considered it helps to compile historical data from the first contact with the organization process which allows them to analyze behaviors purchasing customers and identify lucrative categories to the min order to design programs to help them strengthen customer base.
- the transmission of data from the Web to the data warehouse process includes the following steps:
  - Web data gathered in the log.
  - It handles data record across (Director granulated).
  - Director granulated filtered data is sent to the data warehouse.
- Reliance on the Internet as a source of information for customers on the one hand, and as a means of communicating information about the organization at a low cost to the client as well as it helps to provide individual relationship with each of the client organization

REFERENCES

Abello, A., Samos, J., Saltor, F., 2000. A Data Warehouse Multidimensional Data Models Classification, Technical Report, Pp: 22-25.

Duncan, K., J. Thomann, D. Wells, W. McKnight, D. Marco, M. Haisten, D. Wells, 1999. Proceedings of the Fourth Annual Implementation Conference, TDWI.

Franconi, E., 2002. Introduction to Data Warehousing, Lecture Notes, http://www.inf.unibz.it/~franconi/teaching/2002/cs636/2.

Gray, P. 2005. Marketing designs about information systems (John Wiley & Sons, second edition.

Hui-Huang Hsu, 2008. Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications, pp: 93-102.

Inmon, W. H. 2002. "Building the Data Warehouse (John Wiley, 3th Edition.

Inmon, W.H. 2005. Building the Data Warehouse (Chichester, John Wiley & Sons, fourth edition.

Inmon, W.H., D. Strauss & G. Neushloss, 2008. The Architecture for the next generation of Data Warehousing (Derek Strauss, GeniaNeushloss, first edition.

Lawrence Steve; Lee Giles, C., 1999. "Accessibility of information on the web" New York, NY, USA, ACM, pp:16-27.

Thomas Connolly & Carolyn Begg., 2002. "Database Systems, 3th Edition", Addison- Wesley.

Thomas, C. Hammergren, Alan. R. Simon, 2009. Data warehouse for dummies (John Wiley & Sons, second edition.

****