DATA MINING AND DATA WAREHOUSING

JADHAV S.D. AND SHINDE S.R.

Computer Science Engineering, Mahatma Gandhi Mission College of Engineering, Nanded- 431602, MS, India.
*Corresponding Author: Email– snl_jdhv@yahoo.in, shindeshwetar@gmail.com

Abstract- One may claim that the exponential growth in the amount of data provides great opportunities for data mining. In many real world applications, the number of sources over which this information is fragmented grows at an even faster rate, resulting in barriers to wide-spread application of data mining. A data warehouse is designed especially for decision support queries.
Data warehousing is the process of extracting and transforming operational data into informational data and loading it into a central data store or warehouse. The idea behind data mining, then is the " non trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in India"
Data mining is concerned with the analysis of data and the use of software technique for finding patterns and regularities in sets of data. Data mining potential can be enhanced if the appropriate data has been collected and stored in data warehouse
Keywords- data warehouse, data mining, software.

Citation: Jadhav S.D. and Shinde S.R. (2012) Data Mining and Data Warehousing. World Research Journal of Computer Architecture, ISSN: 2278-8514 & E-ISSN: 2278-8522, Volume 1, Issue 1, pp.-16-18.

Introduction
What is Data Warehouse?
A data warehouse in its simplest perception, is in more than a collection of the key pieces of information used to manage the and direct business for the most Popular outcome. A large amount the right information is the key to survival in today's competitive environment. And this kind of information can be available only if there's a totally integrated enterprise data warehouse.
A data warehouse is repository of integrated information, available for queries and analysis. For such a repository, data and information extracted from heterogeneous resources and consolidated in a single source. This makes it much easier and efficient to query the data.
There are two fundamentally different types of information systems in enterprises: operational systems and informational systems
Operational systems run daily enterprises information like ERP (enterprises resource planning). Information systems analyze the data make decision on how enterprise will be operate, not only information systems have different focus from operational ones, they often have a different scope altogether.
There are some specific rules that govern the basic warehouse, namely that such a structure should be:

- Time dependent
- Non-volatile
- Subject oriented
- Integrated

Need for Data Warehouse
- To summarize the large volumes of data.
- To integrate data's from different sources.
- Make decision makers to access past data.
- Enable people to make informed decision.

Users
From the definition we can infer that the data warehouse users are as follows:

- This person's job involves drawing conclusions from, and making decision Based on large masses of data.
- This person doesn't want to get involved with finding and organizing the Data for this purpose.
- This person also doesn't want to access a database highly technical fashion.

Structure Of Data Warehouse

Data warehousing is one of the hottest industry trends for good reason. The structure of a data warehouse consist as follows.

- Physical data warehouse
- Logical data warehouse
- Data marts

Physical Data Marts- in which all the data for the data warehouse are stored, along with meta data and processing for scrubbing, organizing, packing and processing detail the data.

Logical Marts- also contain as physical database but does not contain actual data. Instead it contains the information necessary to access the data wherever they reside.

Data Mart- is subset of an enterprise wide data warehouse, which potentially supports an enterprise element.

Data Warehouse-Arcitecture

The architecture of an information system refers to the way its pieces are laid out, what types of tasks allocated to each piece of hoe pieces interaction with each other and how they interact with outside world. The architecture of data warehouse is shown in fig. 1.
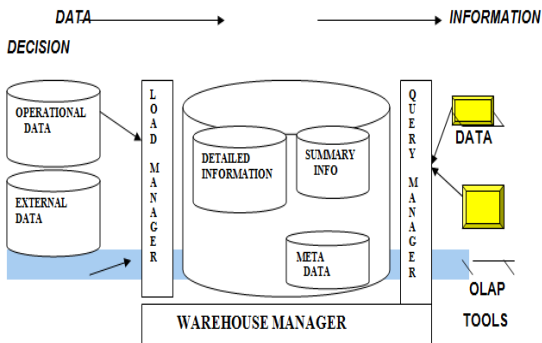

Fig. 1- Data Warehouse Architecture

The architecture consist of following components

- Load Manager
- Warehouse manager
- Query manager

Each component has some specific process.

Load Manager

It is constructed using a combination of off-the- shelf tools, spoke coding,
C programs and shell scripts. It extracts the data from the source systems. It first loads the extracted data from source systems. It performs simple transformation into a structure similar to the one in the data warehouse.

Warehouse Manager

- It is constructed using a combination of third party systems management software, bespoke code, C programs and shell scripts.
- Support warehouse management process, such as transforming data, backup and archives into data warehouse.

Query Manager

- It is constructed using a combination of user access tools, specialist data warehousing monitoring tools, native database facilities, bespoke coding, C programs and shell scripts.
- Direct queries to appropriate table.
- Schedule the execution of user queries.

Partition Algorithm to Discover all Requirement Sets from the Data Warehousing using the Data Mining

Introduction Data Mining

Data mining or knowledge discovery in data bases is the nontrivial extraction of implicit, previously unknown and potentially useful information from the data. This encompasses a number of technical approaches, such as clustering, data summarization, finding dependency networks, classification analyzing changes, and detecting anomalies. Data mining search for the relationship and global patterns that exists in large databases byt are hidden among of data, such as the relationship between patient data and medical diagnosis. The relationship represents valuable knowledge about the databases, and objects in the database, it the database is a faithful mirror of the real word registered by the database. If refers to using a variety of techniques to identify nuggets of information or decision making knowledge in the database and extracting these in such a way that they can be put to use in areas such as decision support, prediction, forecasting and estimation. In particular, finding associations between items in a database of customer transaction. Market basket analysis technique used to group items together. A rule may contain more than one, item in the antecedent and the consequent of the rule. In this paper. we concentrate on finding association, but with different slant (i.e.) by using partition algorithm. In the next section, we review the basis concepts of association rule.

Partition Algorithm

Partition algorithm is based on the observation on the frequent sets are normally very few in number compared to the set of all item sets. The partition algorithm uses two scans of databases to discover all frequent sets by scanning the database once. This set is super set of all frequent item sets i.e it may contain false positives. The algorithm executes in two phases. In the first phase, the partition algorithm logically divides the database into a number of non-overlapping partitions. The partitions are considered one at a time and all frequent item sets for that partition are generated. Partition algorithm as follows.

P = Partition-database(T); n = Number of partitions
For I = 1 to n begin   //Phase 1
read-in-partition(Ti in P)
L1=generate a1 frequent items set of T using a priori method in main memory
End

```
For (k=2; LIK = 1,2,.......,n,k++) do begin        //Merge Phase
CGK = U I =I n LIK end
For I =1 to n do begin
read_in_partition(T1 in P)                         //Phase 2
for all candidates C CG compuate S(C ) Ti end
LG = { C CG/ S ( C ) T1 >= }
Answer = LG
```

Advantages

- Data warehouse are free from the restrictions of the transactional environment. There is an increased efficiency in query processing.
- Artificial intelligence techniques, which may include genetic algorithm And neural networks, are used classification and are employed to discover knowledge from the data warehouse that may be unexpected or Difficult to specify queries.

Applications

Data warehouse application include:
- Sales and marketing analysis across all industries.
- Inventory turn and product tracking in manufacturing.
- Category management, vendor analysis, and marketing, program effectiveness analysis in retail
- Profitability analysis or risk assessment in banking.
- Claims analysis or fraud detection in insurance.

Data mining has many and varied fields of applications such as:
- Retail/Marketing
- Banking
- Medicine
- Transportation
- Insurance and Health Care

Conclusion

Data warehousing provides the means to change raw data into information for making effective business decision – the emphasis on information, not data. The data warehouse is the hub for decision support data. Comprehensive data warehouse that integrate operational data with customer, supplier, and market information have resulted in an explosion of information. Completion requires timely and sophisticated analysis on an integrated view of the data.

Data mining tool can enhance inference process. Speed up design cycle, but con not be substitute for statistical and domain expertise. Data mining allows for the creation of a self learning organization.

So the future of data warehouse lies in their accessibility from the internet. Successful implementation of a data warehouse and data mining requires a high performance; scalable combination of hardware and software which can integrate easily within existing system, so customer can use data warehouse to improve their decision-making-and their competitive advantage

A good data warehouse provides the RIGHT data…to the RIGHT PEOPLE… at the RIGHT time… RIGHT now! While data warehousing organizes data for business analysis, internet has emerged as the standard for information sharing.

References

[1] Arun K. Pujari, *Data mining technologies.*
[2] *Data warehousing, Data mining and OLAP.*
[3] *Berson & Smith, Mc-Graw Hill.*
[4] Bhavani Thuraisingam, *Data mining techniques, tools and trends.*
[5] Elmasri, *Data Base Systems-Tata Mc-Graw Hill.*