iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2 **Issue:** V **Month of publication:** May 2014

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

Understanding How Crucial Hidden Value Discovery in Data Warehouse Is?

Ms. Anshika Goel[#1], Manish Kumar Singh[*2]

[1]*Asssistant Professor, Department of Computer Science, JIMS, GGS IP University, Delhi*

[2]*MCA Scholar, Department of computer Science, PDMCE, MDU, Rohtak, India*

Abstract*— Since the origin of data mining concepts, this area of research has proved to be quite promising and fastest evolving in computer science field. Due to the rapid growth in the demand of the speedy access and fast processing of data from a giant and highly integrated data repository like data warehouse, the data mining concept has been applied to other industry interest and research field, noticeably marketing and retailing, financial forecast, large business transactions, Customer Relationship Management (CRM), Educational Data Mining (EDM), science and engineering, and so forth. Though data mining concept has been furnished by many of the advancements made in its tools, techniques and methodologies to mine data from data warehouse over the past decade, the discovery of certain hidden facts and data from it is still a untouched concept that hasn't got so much considerations and contributions in the field of data mining research. However, the importance of discovering such hidden data has been felt and so been dealt in various fields today. This paper is an effort to through a light on this topic and to provide a mean to understand how crucial it is to discover such hidden facts & data in data warehouse.

***Keywords*— Hidden value discovery, CSV, WEKA, SPECs, OLAP, Data harnessing, CHAID**

I. INTRODUCTION

The amount of raw data generated and stored in corporate databases is growing at tremendous pace. Every day trillions of sales transactions and credit card purchases are recorded as raw data. This makes us 'data rich but information poor'. In today's competitive business environment, companies necessarily need to convert this raw data into significant information and knowledge to get an insight into their customers and markets needs and also to guide their marketing, investment, and management strategies. This presents tremendous opportunities for those who can unlock the information and knowledge hidden in the raw data, but also introduces new challenges.

This paper has three sections. The first section is Introduction itself that is devoted to the background detail of data mining and data warehouse concepts where the difference among the concepts of data, information and knowledge are first discussed, followed by the illustration of the concepts of data warehouse and data mining and finally eliciting what data mining is capable of. The second section is about understanding the theme of the paper that how crucial the discovery of hidden value in data warehouse is which involves detailed discussion of various tools, techniques, methodologies, model and architecture to trace, mine and present the hidden value in data warehouse; followed by the result presenting the significance of discovering the hidden value in data warehouse. In the final section, we have simply wrapped up the paper with a decisive conclusion and let the reader to carry out further study in this area.

A. Distinguishing Data, Information and Knowledge

Before moving ahead, it is important to understand the concept of data, information and knowledge. Though all these three concepts seem similar, they have peculiar differences among themselves.

1) *Data:* Data are any raw facts, figures, or text that can be stored and processed by a computer. Now days, organizations are accumulating infinite and exploding amounts of data in different formats and data repositories. This involves [1]:

- Operational (or transactional) data including inventory, sales, payroll, cost, accounting etc.
- nonoperational data, including macro economic data, industry sales and forecast data
- meta data (meaning data about data) including data dictionary definitions or logical database design.

2) *Information*: When data is organized, processed, structured or presented in a desirable form and patterns, associations, or relationships are extracted from it is called as information. In short, we are able to derive some meaning out of raw data. For example, analysis of sales data can yield information on which products are selling and when [2].

3) Knowledge: When we are able to further convert information into more meaningful historical patterns and future trends, it is known as knowledge. For example, summary information on sales can be analysed to provide knowledge of consumer buying behaviour. Thus, a manufacturer or retailer can determine which products are sold most and which products need more promotional efforts or which product will make most sales in future [3].

B. Concept of Data Warehouse and Data Mining

Data Warehouse is a giant and highly integrated repository of data which acts as a 'sea of data' into which data from various sources are stored. Data mining provides an invaluable mean to achieve speedy access and fast processing of data from data warehouse. These two concepts play a vital role in the knowledge building process of numerous business applications, industries and research field today.

1) *Data Warehouse:* Extraordinary advances in tools and techniques of capturing data, it's processing, data transmission, and storage capacity are compelling organizations to integrate their myriad data repositories like databases into *data warehouses* [4]. A data warehouse stores huge quantities of data by categories so it can be effortlessly managed, retrieved and interpreted. Data warehouses are ideal to maintain a central repository of all organizational data. But merely storing data in a data warehouse is not enough. This is because companies and other business houses are always in need of data for furnishing their knowledge related to customers, markets and products which could benefit them if extraction of meaningful patterns and trends could be made from data through data mining.

2) *Data Mining:* Data mining is a computer assisted process which take huge amount of raw data and processes out useful information and knowledge. It identifies valid, new, highly valuable and clear patterns in data. Data mining tools predict future trends behaviours, so that proactive, knowledge-driven decisions can be made [5]. They search the data warehouse for myriad significant information which could be missed by a normal expert which is not within their scope such as hidden facts, data or patterns, finding information related to the concept of making prediction, et cetera. Simple calculations and statistical measures are not considered data mining. The process must be partially or fully automated, having specialized computer algorithms (i.e., data mining algorithms) that search for patterns in the data.

Data mining derives its name from the similarities between searching for valuable information in a large database and mining a mountain for an invaluable and lusty ore. They both demand any of the two things- either shifting through a considerable amount of material, or a probing it with quite diligent as well as intelligent approach in order to locate the value where it resides.

C. What Data Mining is capable of?

Data mining is used main by companies which are into retail, finance, marketing etc. That is companies having strong consumer focus. Data mining enables companies to balance the internal factors such as price with external factors such as competition and other economic demographics. Therefore sales, customer satisfaction and profits can be increased considerably [6]. Data mining uses pattern recognition and statistical techniques to extract trends, patterns, relationships, facts, exceptions and anomalies from extremely large data warehouses which cannot be identified by any other method

or technology. It enables them to extract minute important information from massive data and take complex decisions.

Specific uses of data mining include [7]:

1) *Market segmentation*: the common traits of customers who buy the same products.
2) *Churn* - Predict the loyalty of the customers to see if they are likely buy from a competitor.
3) *Market basket analysis* - Understand what products or services are commonly purchased together.
4) *Direct marketing* - Identify which section of customers to be focused for direct marketing (eg. E-mail) to obtain the highest response rate.
5) *Interactive marketing* – to identify which section of customers is to be targeted for indirect marketing (eg. Web sight).
6) *Trend analysis* - Reveal the difference between a typical customer this month and last.
7) *Fraud detection* - Identify fraudulent transactions.

II. UNDERSTANDING HOW CRUCIAL HIDEN VALUE DISCOVERY IN DATA WAREHOUSE IS?

Data mining is quite crucial in this age of cut-throat competition where the data is the "lifeline of any kind of modern business scenario". This becomes quite significant when large gigabytes of scanned date is to be stored and managed through virtual warehouse, rightly known as Data Warehouse, which demands speedy processing and thunderbolt like data accessing. Nobody can condemn that these tasks need a shear amount of data shifting and probing of "iron & steel" senses to make sure the exact location of data value. This is quite evident that data mining strategies have larger than life potentials to offer myriad opportunities to business ventures even at the difficult circumstances that couldn't be provided by any other technology. Most importantly, when it comes to manage complex databases of comparatively tough quality and gigabyte like size, data mining strategies and technologies become quite handy. This is because of the following reasons:

- *Data mining has sound and quite effective techniques that could make prediction of myriad kinds of trends, information, events and behaviours***.**

  This is quite beneficial and easy to achieve that gives speedy responses to any sort of query in fractions of seconds. It was a daunting task few years back when the analysis used to be done manually. It is quite important to understand that the business problems that involves quick prediction always demands the large reservoir of past as well as various kinds of data that could be latter matched, analysed and processed to make quick and sound decision, prediction and forecast of any sort. The famous business fields where such data mining strategy could work well includes targeted marketing, risk management, resource allocation, forecasting bankruptcy and other kinds of fault identification.

- *Data mining has number of tools and techniques that can help in exploring the previously hidden knowledge***.**

  This can be done by traversing through each database in a data warehouse and doing identification of such hidden patterns in one turn. This capability of data mining is quite beneficial in discovering patterns of retail sales while purchasing unrelated products together that are otherwise difficult to achieve; identifying problems related to pattern discovery while detecting transactions of fraudulent kind made through e-Banking; and in detecting various anomalies occurring while making data entry.

- One of the major benefit of using data mining tools and techniques is that they provide automation functionality on any of the existing platforms of hardware and software. Further, such automation can be applied over any of the system by upgrading the existing platform which is very beneficial for developing new and innovative products.

- As we have discussed in the beginning of this section that data mining tools and techniques provide speedy processing in quick span of time. This is boon for users demanding large number of models to be experimented automatically so as to understand data

which are complex and difficult to locate. This capability of data mining concept can be fruitfully extended to be applicable over parallel processing systems for providing high performance and performing complex database analysis in a quick span of time. Such analysis can be performed over any kind of databases of any level depth or breadth.

A. *Discovering Hidden Values- A Discussion*

Data mining provides a large number of tools and techniques for discovering hidden values from a data warehouse

1) *Data Mining Tools*: Data mining tools are basically software tools that are helpful in importing and exporting data value and model of various kinds in swift way [8]. The first step in any data mining process is to generate and host data from various sources like databases, data marts, web data or at many times using software corresponding to the measurement devices. The matter of fact is that almost all of the data mining tools used today comprise of high performance visualization techniques that help in presenting results of data mining in quite easy and readable form. Such tools are basically beneficial for business application, applied research, forecasting, detailed analysis, decision making and so forth. Moreover, the interactive methods provided by the data mining tools prove to be a great leap in explorative kind of data analysis. It is to be noted that Oracle or similar kind of software interfaces to databases that support SQL (Structured Query Language) standard have become handy while importing data in business applications.

Some of the noticeable examples of software tools

that facilitate import and export of data are:

- CSV (Comma Separated Value) format that is basically used by some of the tools of non- data mining that provides an improvised standard for exporting Excel of text files.
- WEKA is a file format of Attribute-Relation kind of file development that is used by the software tools applicable over text or binary files.
- PMML is a XML-based standard used by IBM and at latest by SAS that helps to import and export of components (of existing system models) in other systems and processes [9].
- OLEDB (Object Linking and Embedding Database) is an Application Programming Interface (API) standard developed by Microsoft for quick access of myriad kind of data stored and maintained at different location. There is a collection of interfaces provided by OLEDB that help in exploiting various COM (Component Object Model) functionalities inside any text or binary files

We can categorize the software mining tools depending upon the type of mining and operation performed by them. At present, many researchers have categorized the data mining software tools into the following types:

a. **SPECs (Specialties)**: They comprise of artificial neural network methods that are easy to apply . Some of the famous SPECs tools are Bayesia Lab (for Bayesian network0, MagnumOpus (for association analysis), CART (for decision trees), Neuroshell and Wizrule.

b. **DMS (Data Mining suites)**: They are much alike to SPECs but this is ideal only for data mining purposes which provide significant functionalities, noticeably time series, feature tables and text mining. They are applicable ont only to business applications but can also be applied to field concerned to determining business solutions, achieving import & export of various models, reporting fault and supporting myriad platforms. Important examples of DMS tools are IBM SPSS, KXEN, STATISTICA, GhostMiner, TIBCO spotfire and SAS Miner.

c. **MATs (Mathematical Packages)** : They comprise of a set of extendable and reliable algorithms along with a wide collection of visualization routines that provide significant data mining functionalities including time series, feature tables and image import/export. However, they are not entirely applicable for data mining purposes. Rather, they are employed for developing algorithms for

carrying out tasks related to applied research. They allow the development of algorithms only in the form of research prototypes (RES) and extension types (EXT). They are available either as commercial tools such as R-PLUS, METLAB or as open-source tools such as Kepler, R.

d. **BIs (Business Intelligence Packages)**: They provide some of the basic data mining functionalities. They employ statistical methods for solving out business related problems. Further, they have the capability of providing powerful database coupling by employing client/server architecture. The BIs tools are, in general, available as commercial tools including IBM Cognos 8, IBM's DB2 Data Warehouse, PolyVista and Teradata Database.

e. **LIBs**: They comprise of data mining libraries that provide a wide range of methods & functionalities for carrying out data mining tasks. LIBs library facilitate the user to embed various kinds of tool in other software and processes by providing a huge collection of APIs. However, they lack GUI and other interactive features. The most significant characteristic of LIBs is its platform-independence capability and they are written in C++ or Java which make them popular among the C++ and Java programmers. They are widely available both as commercial tools (example:- XELOPES that is based on C++, Java, C# and Neurofusion that is C++ based) and open-source tools (example:- MLC++ that is C++ based, WEKA that is Java based and LIBSVM that is based on C++ and Java).

f. **INTs (Integration Packages)**: They provide a large collection of extendable open-source algorithms that could either be available as standalone tools such as KNIME (GUI-version of KEEL, WEKA and TANGRA) or as larger extension package such as MAT type which comprise of tools including METLAB's PR tools, Gait-CAD and R's RWEKA.

g. **SOLs (Solutions)**: They comprise of a huge repository of customizable tools that are well suited to challenging application fields, noticeably image processing (that employs tools such as ITK & ImageJ), image analysis in microscopy, text mining (that employ tool such as GATE) and gene expression profiles' mining that provides significant benefits such as reliable evaluation measures, sound technique of domain specific feature extensions and input & visualization formats.

h. **EXT**: They are special kind of data mining tools that basically provide a huge collection of smaller and simpler add-ons for other existing tools including METLAB, Excel, R, et cetera with a limited amount of functionalities. They are available both as open-source or commercial tools. Typical examples of EXT tools are XLMiner and ForecasterXL.

i. **RES**: They are one the most famous data mining tools that provide a considerable collection of innovative and new genre of algorithms with limited GUI support as well as limited functionality and are devoid of automation facility. They are preferred by the users of applied research and algorithm development fields. Some of the most fascinated RES tools are Himalaya (it provides the swashbuckling capability of mining minimum amount of frequent items sets), Pegasus (it provides he facility of graph mining) and GIFT (it is a crucial tool that helps in retrieving content-base image.

2) Data Mining Techniques: The real task of data mining in data warehouse could be either automatic or it could be semi-automatic that involves sound analysis of huge amount of data which are extracted using number of techniques. The most common techniques of extracting data in data warehouse could be any of the following:

- *Cluster analysis* that involves extraction of previously unknown patterns in the form of groups of data records;
- *Anomaly detection* that involves extraction of previously unknown patterns in the form of unusual records; and
- *Association rule mining* that involves extraction of previously unknown pattern in the form of dependencies.

All above discussed techniques of data mining in data warehouse require the sound and fine database techniques. One of such technique is *spatial indices* that provide the extracted patterns a summarized view of input data that could further be used in deep rooted analysis of patterns or in other associated challenging fields of data mining including predictive analytics and machine learning.

In the past decade, the researchers of data mining field have employed various techniques to mine hidden data value in data warehouse that are applicable over relatively small amount of data which aid in locating and building knowledge from the smallest of the smallest data residing in the data warehouse. Some of these robust and sound techniques include:

a. **Genetic algorithms** which are data mining techniques that help in achieving optimization of the mined data that strongly favor the concept of evolution of extracted hidden value in data warehouse. This could be achieved effectively by using various concepts of genetic algorithms, noticeably natural selection, mutation, and genetic combination.

b. **Artificial neural networks** which are one of the significant techniques of data mining that are effective in providing a pool of non-linear kind of predictive model. Such a pool of predictive models is obtained through innovative methods such as training and building biological neural network like resembling structures.

c. **Rule induction** which is a sound data mining technique that is famous for employing if-then rules that boost the process of hidden value mining in data warehouse on the basis of statistical analysis. This technique has proven quite effective over the last decade in building KDD systems and knowledge acquisition.

d. **Decisions trees** which represent hierarchical model of hidden data value extraction in data warehousing employ the technique of building tree like data structures to generate a large set of decisions that could further be used to "mint" knowledge-building *rules*. Some of the noticeable decision tree techniques are CHAID (Chi Square Automatic Interaction Detection) and CART (Classification and Regression Trees).

e. **Nearest neighbour method** which is quite beneficial and one of the most innovative hidden data value extraction in data warehouse that employ the technique of class combination of K records that help in classifying each record contained in a dataset. Thus this method is also knows as K-nearest neighbour technique.

At present, after the emerging size and competition in modern business world, the data mining concept has evolved at a sky-reaching level by extending its reach to data warehouse and OLAP server techniques that have penned the new innovative and reliable business processing and access standard that would contribute to the business field and KDD for next 100 years. This is evident from the following facts that enunciate the myriad capabilities and functionalities offered by the OLAP server technique [10]:

- OLAP that stands for Online Analytical Processing constitute a significant data mining technique, especially for hidden value in data warehouse, that boost up the process of organizing a giant and complex database which favors speedy data processing access [10].
- OLAP is basically optimized for providing the benefits of querying and reporting about the mined data, rather than just performing the task of transactions' processing.
- OLAP constitute a huge repository of historical data that are optimized into fine data structures for carrying out swashbuckling analysis.
- OLAP has one major benefit that its data are stored in multi-dimensional database where each attribute (such as product, time, month, object, region et cetera) are evaluated as a distinct 'dimension' to provide fast processing and access of data. In contrast, relational database provide a conventional two dimensional database that has limited usage and knowledge building.
- OLAP data are hierarchically organized that are stored in data repositories emulating cubes rather than tables

3) Categorization of Hidden Value Mining Tasks in Data Warehouse: As compared to data mining tasks, the extraction of hidden value in data warehouse is quite cumbersome and a tough challenge to be followed. However, the phases of locating and extracting normal data and hidden value in data warehouse are not uncommon. Both kind of mining process involves the following two undoubtedly significant phases:

a. Pre Mining phase tasks: The tasks belonging to this kind of mining phase are performed before the enunciation of data mining process which include the tasks of *data cleaning* and data *integration.*

Data cleaning is the process of removing noisy and inconsistent data.

Data integration is the process of obtaining data from various sources to a single location s per some common format.

b. **Post Mining phase tasks:** The tasks belonging to this kind of mining phase are performed after the enunciation of data mining process which include tasks of *pattern evaluation* and *knowledge presentation.*

Pattern evaluation is the process of recognizing the desired patterns that could fit to represent knowledge.

Data integration is the process of presenting the discovered rules obtained through visualization and various other knowledge representation methodologies.

4) *Modelling of Hidden Value Mining Tasks in Data Warehouse*: A model lays the foundation to carry out mining of hidden value in data warehouse in smooth and frequent manner. It is a mean to perform a set of actions in one time frame with certain known facts which could be applicable latter in another time frame with unknown facts. In other words, a model represents a set of facts and ideas describing the past events

We can understand the above discussion on a Model through this example. Suppose, we are living in any part of the world other than Agra, India and we want to spend some qualitative time to watch an IPL (Indian Premiere League) cricket match between Mumbai Indians (MI) and Chennai Super Kings (CSK) at Agra. So the first thing that we want to do before visiting the venue is to Google Agra to know the fact where does it located, search its route map and finally locate the venue. Since Agra is a common place on the world map whose path has been defined on the Google through the events and tours made by the people in past that become the source of visiting Agra for the people in the present and future era. So in this example the map acts as a model to visit Agra in order to reach the venue and enjoy the cricket.

When we talk about mining hidden value in data warehouse, we tend to build a mining model in somewhat as discussed above on the bases of certain 'functionalities' specifying the kind of patterns discovered while performing the data mining tasks as discussed in the previous section.

In general, the mining model to extract hidden value in data warehouse could be devised in somewhat similar to mining model to extract the normal data in data warehouse which basically depends upon certain functionalities.

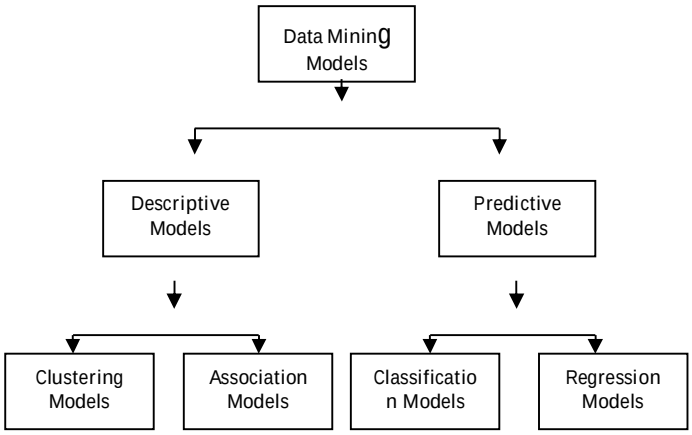


Fig: 1 Mining model of extracting hidden value in data warehouse

a. Descriptive Model: This model of mining hidden value characterizes the general

properties of the data in the database which involve s the inclusion of patterns in existing data and is generally associated with creating meaningful subgroups. This model is further bifurcated into two sub-models, namely:

Clustering model which clumps similar things together to create meaningful patterns where events or people included into the groups are known as *clusters* which play the role of reducing data complexity. More precisely, the objects in this model are clustered or grouped based on the principles of maximizing the intra-class similarity and minimizing the inter-class similarity.

Association model which determine the "degree of likeliness" of how frequently two or more objects associate together at a given time to create meaningful patterns. This model exploits two parameters to do so which are namely- *confidence* and *support*. In addition to this, this model makes rich use of association rule which comprise a single attribute or predicate that often repeats.

b. Predictive Model: This model of mining hidden value characterizes the act of performing inferences on the current data in order to make predictions. Such prediction or forecasting is made to determine the external value of a specific attribute. In order to do so, the predictive model is also bifurcated into two sub-models, namely;

Classification model which makes use of class membership (or classifier)to forecast the value of an attribute. For instance: we can use the classifier such as 'safe' or 'risky' for the act of skiing while we can use the classifier such as 'high' or 'low' to define altitudes. When it comes to make classification of hidden value in data warehouse, it can be achieved in two steps, chiefly- *learning*, followed by *classification.*

Regression model which makes use of *machine learning* techniques to fit an equation to a dataset [11]. For instance, linear regression is the simplest form of regression model which employs the famous formula of a straight line (y=m*x + b) to predict the value of y based upon the given value of x by determining the appropriate values of m and b. The advanced techniques, such as multiple regression, allows the use of more than one input variable to fit the more complex models such as a quadratic equation.

5) Building of Architecture of Hidden Value Mining System in Data Warehouse: The mining model to extract the hidden value in data warehouse when analysed thoroughly, we can simply formulate the architecture for the system which can mine hidden data in quiet smooth manner. However, such architecture is complex in which diverse components have their own significant role to play. These components are grouped under different stages or phases of mining, including:

a. **Data harnessing stage**: In this stage of the architecture, various components come into play to collect hidden data from data warehouse. OLTP, acronym for Online Transaction Processing System, constitute this stage which uncovers the hidden values that usually go missed while current processing of data and uses the concept of luminous path to trace and locate the data. The data thus located is cleaned, transformed, loaded and refreshed to move it to next stage.

b. **Data presentation stage**: In this stage, the data passed from the data harnessing stage is then transformed into a presentable stage so that it can be used for further study and analysis. Here, the sophisticated components like OLAP and Data marts come into force for not only serve the purpose of data presentation but also prepare data to be used by considering various dimensions.

- OLAP defines the logical design of data so that data can easily be navigated through the sub-data repositories of data warehouse and helps in modelling decision based data. Moreover, it provides the capability to users to achieve speedy manipulation and visualization of data in the face of multidimensional view. There are two OLAP servers to achieve such capabilities in different scenarios:

  ROLAP servers: they are extendable form of relational DBMS where the data is kept in relational database by making the use of schema technique such as star-join kind of schema that supports SQL extensions (operators including cross-join and cube) and possesses index structure(generally bitmap and join indexes)
  .
  MOLAP servers: they are multidimensional form of DBMS where the data are kept in n-dimensional array that involves direct access to array kind of data structure, providing brilliant indexing properties with certain limitations like bad storage implementation while working with sparse data.
- Data marts define the departmental kind of data that implies storage, access and presentation of all data in the face of dimensional models with drill-across techniques that facilitate to tie data marts with bus architecture of data warehouse. This is significant in the situation like working with fewer volumes of data with limited data sources where the fast roll-out and a simpler data cleansing technique are required.

c. Data access stage: In this stage, several tools are employed to provide end user access to data presented in the data presentation stage. The sophisticated tools, noticeably OLAP tools, DSS, BI apps, Query/Reporting tools and myriad mining tools are employed to access significant data from hidden value of data warehouse.

- OLAP tools: are designed in wide range to assist user in carrying out tasks pertaining to different scenarios, like:

  Drill –across makes use of confronted joins & dimensions to move data between myriad schemas of star-join;

  Drill-up is used to maximize the aggregation;

  Drill-down is used to minimize the aggregation level;

  Ranking is used to sort data in meaningful order;

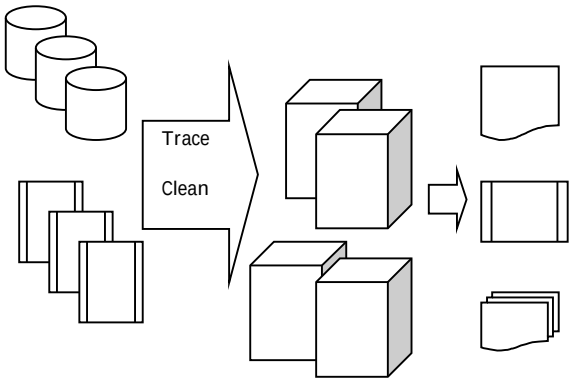
  Pivoting is used to interchange rows with columns or vice versa; and

  Slicing & dicing is used to analyse data in database through different views.

- DSS (Decision Support System): assists user to build decision by working upon certain facts and

figures using machine learning technique.

- BI (Business Intelligence) apps: are designed to assist user in achieving various strategic, operational and analytical tasks over the presented hidden value in data warehouse.
- Query/Reporting tools: assist user to access data from data warehouse by making simple queries. If the requested data or information does not exist in the data repository, an error is reported by displaying suitable error messages.



Data Harnessing Stage (i) Data Presentation Stag (ii) Access Stage (iii)

Fig: 2: Architecture of Hidden Value Mining System

B. *Result*

Data warehouse and data mining have brought new trends in computing environment and information technology, especially to the fields that demands large-scale processing and analysis of data. Today, myriad range of companies and business houses have incorporated numerous advanced and effective data mining applications to build knowledge base through the extraction, cleaning, loading and refreshing of hidden value in data warehouse. This is basically done to boost the performance and reputation of industries that are information-intensive and that cater for maintaining sound relationship with customers. There are two factors that have led the data mining of hidden value in data warehouse a huge success. These are:

- The deployment of a giant and fully-integrated data warehouse that could hold a huge collection of data relating to different sources, type and genre.
- A thorough understanding of various concepts and processes of business that is to be worked upon by various data mining tools, techniques and methodology. So it is important to make sure the field (example: direct mail marketing, customer prospecting, financial services, campaign management, et cetera) where the data mining concept is to be applied.

A data warehouse is a repository of subjectively selected and suitable operational data, which can successfully answer any ad-hoc, complex, statistical, analytical queries. It is a collection of databases, data tables, and mechanisms to access the data on a single subject [12].

Data mining has changed the discipline of information technology in past few decades while envisaging the properties of tools, methods and technology needed in the acquisition, organization, analysis, use and dissemination of information [13]. Data mining can be used to achieve myriad tasks. It can be used to make two types of knowledge discovery namely:

1). Supervised knowledge discovery: it works on the pre-classified data where each data element is made to be associated with a unique label so that the class to which data element belongs can easily be determined. It is done through classification mining model where the output is known in advance.

2). UN-SUPERVISED KNOWLEDGE DISCOVERY: it does not make use of pre-classified data elements to knowledge discovery. Rather, it forms groups on the basis of common characteristics by the method s of clustering mining model. It is noteworthy to know that unsupervised learning is used when output is not known in advance.

When we talk about how significant the mining of hidden value in data warehouse is, we simply want to list the fields where the hidden values can be well tracked, cleaned, organized and presented to form the knowledge base of the particular field. These hidden value could be anything that are overlooked while accessing easy-to- locate data from data warehouse which can play a vital role in knowledge discovery and decision building system. Some of the noticeable fields where this discovery of hidden value in data warehouse could play a crucial role are:

- **Web data mining:** Web data is the collection of data from the web/internet which is used in the context of web personalization. A web data is broadly divided into four categories [14][15]- content data (used in appropriate structure to the end-users in various forms including text, images, or structured data), structure data (representing the manner in which content is organized which could either be entities used within a web page such as HTML or XML tags or entities used to link web pages in a website such as hyperlinks), usage data (constitute the data used while accessing a website, prominently the visitor's IP address, time and date of access, plus web access log), and user profile data (represent the information regarding the website's user consisting of demographic information associated to a website, plus the information associated with users' preferences & interests). It all depends upon a web crawler how effectively it accesses a data. If any of the data is missed to be crawled then it could lead to ineffective search of information over the web. To counter this unexpected problem, today most of the search engines use the concept of mining hidden value of data warehouse by making use of luminous path to track such data which go missing while crawling.

- **Pharmaceutical company:** It uses the knowledge base (build from the hidden value in data warehouse), consisting of latest sale made and their corresponding results to furnish targeting of physicians with high value in order to forecast the marketing strategies that could have larger impact on the society. To avoid any issue of data go unread/unprocessed, such company has a team to perform activities in order to trace the hidden value in data warehouse. The recent research made in this field has a lot more to offer the sales marketing business that provides a huge collection of live plus hidden data related to past and current events made in the sales field that could be well analysed to deliver effective and impactful decisions for selling or marketing of the product.

- **Educational Data Mining (EDM):** It is one of the emerging and fast growing innovative field of research which contribute immensely to the collection of huge amount of student data from web logs to provide a much more semantically rich data contained in the student models [16]. EDM describes a field of research associated with the application of data mining, machine learning and statistics to information generated from educational settings (example universities and intelligent tutoring systems) [17]. In this field also, the best possible strategies are build to trace any unprocessed or hidden data in data warehouse by developing methods for exploring such data at a high level, by considering multiple levels of meaningful hierarchy, so as to discover new findings about the way the people learn in the situation demanding such setting [18]. While doing so, EDM has much more to contribute to theory of learning envisaged by researchers with reference to educational psychology and the learning sciences [19]. This field has much more to contribute to learning analytics where the EDM and learning analytics could be compared and contrasted [20] that could be achieved by processing huge amount of data in data warehouse, plus by seeking the way of mining hidden value in data warehouse.
- **Transportation Company**: It can use the huge collection of data in a data warehouse- both processed as well as hidden value to enforce best possible strategies for defining greater dimensions to its services' access. In addition to this, it can analyse data to provide better prospects of maintaining relationship with customer and providing them good experience. The company has well skilled workforce to trace all kind of data using best possible measures so no data can go unread or hidden while processing so as to deliver quality services to the customers.

- **E-banking**: The data processed from the hidden value of data warehouse can be used effectively to handle to locate the transaction made by new users of the E-banking services so that the necessary information that could be sometimes overlooked that could lead to major transaction failure. Such failure can be avoided if t he E-Banking system has all information supplied to it. And this is leveraged by the concept of using the hidden value in data warehouse.

- **Movers and Packers**: The analysis of all the data mine in data warehouse can be significant to improvise the sales activities made to retailers. Just imagine what happen if any of the data go unnoticed while billing or selling made by such a company to retailers.. May be it could be disastrous for the company. The bet possible alternative is provided by the discovery of hidden value in data warehouse

All above fields' instances are enough to explain the significance of discovering of hidden value in data warehouse through data mining applications. There are many more fields and examples exist here to discuss but the time and pages would go limited to carry out such discussion.

III. Conclusions

There have been made a lot of research and so numerous papers have been written envisaging the concept of data warehouse and data mining but a limited work has been carried out to discuss the issue and the significance of discovery of hidden value in data warehouse. We have tried our best to make a deep discussion eliciting the issue of hidden value go unprocessed in data warehouse and have written in detail how crucial could it be to discover such data in our research paper. We hope our effort would find a suitable place while making research and study of discovery o hidden value in data warehouse.

Acknowledgment

We are very obliged to be very thankful to our family and friends who have cherished and so motivated us in writing this paper. Moreover, we are extremely thankful for all those authors and researchers whose research papers and their work have guided us diligently to carry out our work

References

[1] Mishra, A.*"A Survey on the Research Challengesfor Data Mining"*.

[2] http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm

[3] http://www.thearling.com/text/dmwhite/dmwhite.htm

[4] Jaideep Srivastava, Prasanna Desikan, Vipin Kumar "Web Mining – Accomplishments & Future Directions"

[5] Ben G. Weber and Michael Mateas "A Data Mining Approach to Strategy Prediction"

[6] Data Mining: Introductory And Advanced Topics By Margaret H Dunham, Pearson education

[7] Jiawei Han and Jing Gao University of Illinois "Research Challenges for Data Mining in Science and Engineering" at Urbana-Champaign.

[8]]. Mikut, R. and Reischi, "Data Mining Tools". January/February 2011, Volume-00.

[9] Peckter, R. "What's PMML and What's new in PMML 4.0". ACM SIGKDD Exploration, Newsletter 2009.

[10] www.office.mirosoft.com/en-in/excel-help/overview-of-online-analytical-processing-olap-HP010177437.aspx

[11] Chapel, M. www.databases.about.com/od/datamining/g/regression.htm.

[12] www.mysafaribooksonline,com/book/information_technology-and-softwaredevelopment/9788131760291/databasefundamentals/cho2005

[13] Cunnigham, S. and Frank, E. 1999. "Proceedings of the Sixth International Conference on Neural Information Processing". Market Basket Analysis of Library Circulation Data, pp-825-830

[14] Srivastava, J., Cooley, R., Deshpande , M. and Tan, P. N. January, 2000."Web Usage Mining: Discovery and Application of Usage Patterns from web data". SGKDD Explorations. Issue-2, pp-12-23.

[15] Eirinaki, M. "Web Mining: A Roadmap". pp.1-57.

[16] Merceron, A. and Yacef, K. "Educational Data Mining: A Case Study".

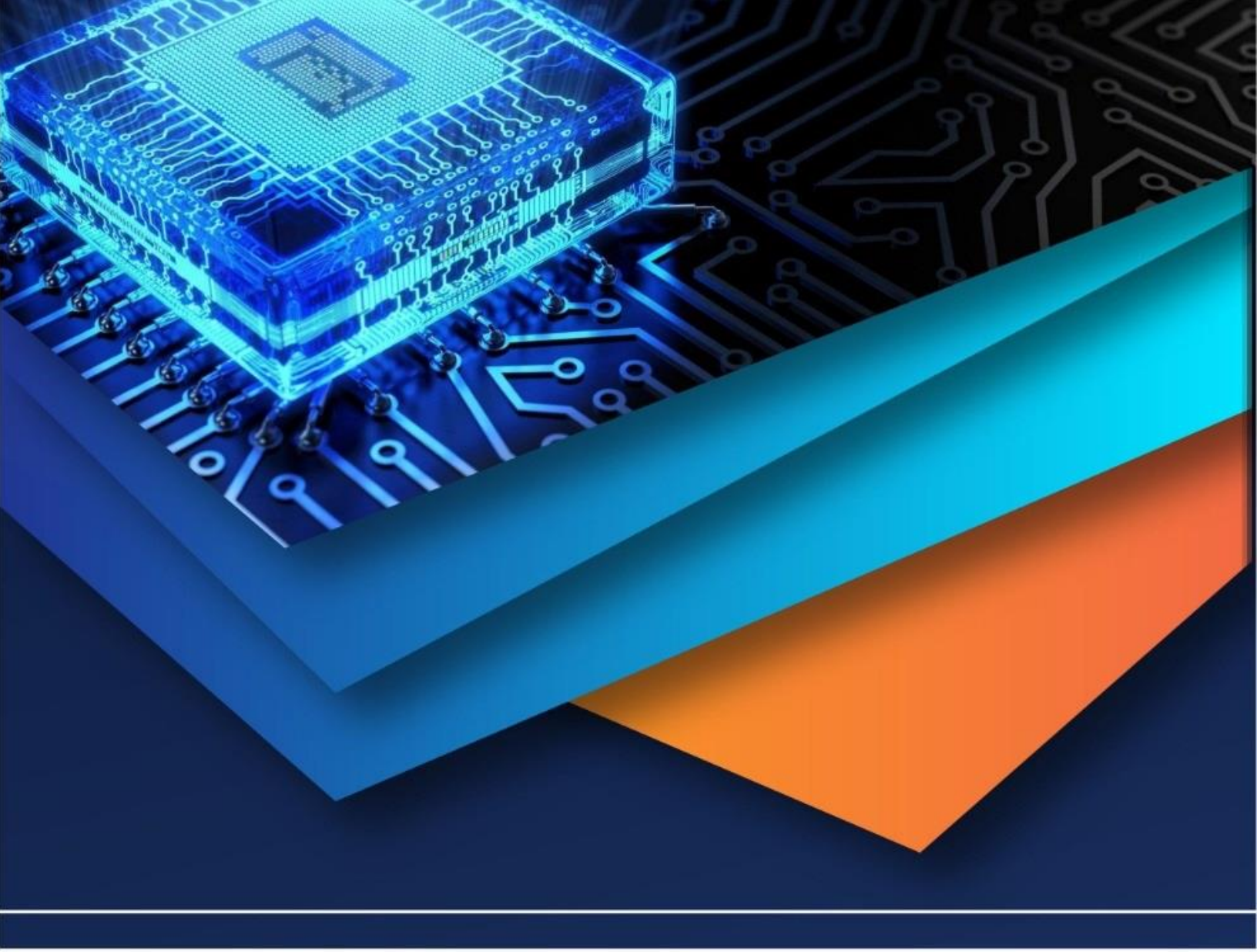
[17] www.en.m.wikipedia.org/wiki/Educational_data_mining

[18] "Educational data Mining.org", 2-13

[19] Baker, R. 2010. "Data Mining for Education". International Encyclopedia of Education. 6th Edition, Vol-7, pp-112-118, oxford, U.K, Elsevier.

written in detail how crucial could it be to discover such data in our research paper. We hope our effort would find a suitable place while making research and study of discovery of hidden value in data warehouse.

[20] Siemens, G. and Baker, R. "Learning analytics and Educational Data Mining: Toward Comunication and Collaboration". Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, pp-252-254



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY