ISSN:0975-9646

Exploring the Applications of Data Warehouse in Public Life

Arvind Kumar
Computer Science Department
K. M. Government (P.G.) College Narwana (Jind)

***Abstract*— Data warehouse is the important means of preparing the peoples and government to face the challenges of the new millennium. Today data warehouse is used worldwide in various applications. Today due to increase in data complexity and manageability issues, data warehousing has attracted a great deal of interest in public life applications like business, finance, Inventory, Banking, healthcare and industries. As the importance of retrieving the information from knowledge-base cannot be denied, data warehousing is all about making the information available for decision making. Data warehouse is taken as the heart of the latest decision support systems (DSS). Due to the eagerness of data warehouse in public life, the need for the design and implementation of data warehouse in different applications is becoming crucial. Information from various operational data sources are integrated by data warehousing into a central repository to start the process of analysis and mining of integrated information and primarily used in strategic decision making by means of online analytical processing techniques (OLAP).**

Keywords—Data warehouse (DWH); Data warehouse applications; Online Analytical Processing (OLAP); Decision support systems (DSS);

I. INTRODUCTION

Operational and transactional systems are the new generation systems which are different from 1970's decision support systems (DSS) [1]. In order to complete the life cycle, DSS needs a Data Warehouse (DWH). A DWH pools the available data which is spread all over the organization, and makes a unify pool (like data structure) having the presence of similar and linked formats [2].

Data warehousing takes off in the 1980s as an answer to the very little or no availability of information propagated by online application systems, online applications were praised by a very limited domains of users, and integration was not there even [3]. Historical data kept by online applications are very little as they deposit their historical data for high performance in faster way. Thus organizations hold very little information as compared to data [3].

With the thirst and huge need for large blocks of information, DWH gain much importance and became an essential strategy component for medium and large organizations. Timely and accurately decision making at management level becomes difficult due to the incapability of traditional databases to handle increasing demands of online information access, retrieval, maintenance and update efficiently which greatly impacts every industry [5]. So companies start seeking the solution for all their problems and adopt DWH technology.

With sharp and harder competition, enterprises are targeting in availing fast and pinpoint information to have best decisions. Furthermore, with the thirst for huge chunks of information, enterprises' traditional DB (database) is off no use of smartly managing the increasing needs of online information update, access, maintenance, and retrieval. This lagging impressively effects the efficiently and effectively usage of internal data by the management to hold decision-making in time. As a result, to search for various ways and means to store, access, handle, and utilize the huge chunks of data in an effective manner, is the main concern of every business [5].

Organizations requires a database system for their daily decision making, with better adaptability, top flexibility, and best support. Considering the past decade, the educational (academia) side and the industry side, both have progressively plated different layouts to solve the problems and to present solution to craft an aforementioned system [5]. Adopting the data warehouse technology is one of the solutions to that. DWH was defined by Inmon [3, 4] as, ''pooling data from multiple separate sources to construct a main DWH". Proper data-analyzing tools can be used by different users to analyze and store required data.

Data Warehouse's purpose is to take large data from heterogeneous sources and furnish them in known formats that helps in understanding and for making smart decisions [6]. The Benefits linked to the DWH applications include the region of time saving, with the availability of clean and handful of information, tough and exact decisions making in accordance with the improvement of processes related to business and to help achieving strategic business objectives [2, 4, 5, 6].

Realizing the need after researching literature and for further exploring on this research article, taking in account the importance of the applications of DWH in public life and the shortfall of the factual research, we have all the concreate reason to explore the most applications of DWH in public life. In this paper we discussed different applications of DWH in public life along with available case studies. Its sections as follows; Section 2 presents DWH technology. Section 3 presents the applications of data warehousing in different domains. Section 4 provides a descriptive view of different case studies under the umbrella of government and business categories. Section 5 provides a brief usage analysis of Data Warehouse applications. Finally, conclusion is presented in Section 6.

II. DATA WAREHOUSE TECHNOLOGY

A data warehouse is a repository of information collected from multiple sources, stored under a unified schema, and that usually resides at a single site and it is subject-oriented, integrated, nonvolatile, and time-variant. Data warehouses are constructed via a process of data cleaning, data integration, data transformation, data loading, and periodic data refreshing. To facilitate decision making, the data in a data warehouse are organized around major subjects, such as customer, item, supplier, and activity. The data are stored to provide information from a historical perspective (such as from the past 5-10 years) and are typically summarized [8]. The primary goals of a data warehouse are the following:

- Make organization's information easily accessible.
- Present the organization's information consistently.
- Provide capacity to separate and combine data.
- Inclusion of tools set to query, analyze and present information.
- Publish used data.
- Serve as the foundation for business decision making.

Feature of Data Warehouse

1. **Subject-oriented:** A data warehouse is organized around major subjects such as customer, supplier, product and sales. Rather than concentrating on the day-to-day operation and transaction processing of an organization data- warehouse focuses on the modeling and analysis of data for decision makers.
2. **Integrated:** A data warehouse is usually constructed by integrating multiple heterogeneous sources, such as relational database, flat files and on-line transaction records. Data cleaning and data integration techniques are applied to ensure consistency in naming, conventions, encoding structures, attributes measures and so on.
3. **Time-variant:** Data are stored to provide information from a historical perspective (e.g., the past 5-10 years). Every key structure in the data warehouse contains, either implicitly or explicitly, an element of time[9].
4. **Nonvolatile:** Data in a data warehouse is never updated but used only for queries. Thus such data can only be loaded from other data bases such as the operational databases. End users who want to update the data must be the operational database, as only it can be changed, updated or deleted. This means that a data warehouse will always be filled with historical data .In sum, a data warehouse is a semantically consistent data store that serves as a physical implementation of a decision support data model and stores the information on which an enterprise needs to make strategic decision. A data warehouse is also often viewed as architecture, constructed by integrating data from multiple heterogeneous sources to support structured and/or ad hoc queries, analytical reporting and decision making.

Components of data warehouse

The components makes a complete environment of data warehouse. Each of these components performs specific function. As shown below in figure 1 there are four separate and distinct components of data warehouse:

1 Operational Source Systems These are the operational systems of record that collect data from multiple sources such as all applications, databases, computer system within the enterprise etc. The operational source systems are thought of as outside the data warehouse because we have no control over the content and format of the data in these operational systems. Because operational data is fragmented, inconsistent and not organized around subjects (such as customer, vendor and product), so it must be cleaned up to make it consistent and useful for decision making [9].

2 Data Staging Area The data staging area is everything between the operational source systems and the data presentation area (see fig 1). The data staging area of data warehouse is both a storage area and a set of process commonly referred to as extract-transform-load (ETL). Extraction is the first step in the process of getting data into the data warehouse environment. Extracting means reading and understanding the source data and coping the data needed for the data warehouse into the staging area for further manipulation.Once the data is extracted to staging area there are numerous transformations such as cleansing the data, combining from multiple sources, deduplicating data and assigning warehouse keys etc.

Hence by applying these transformations the raw operational data is transformed into a warehouse deliverable fit for user queries. In data staging area simple activities of sorting and sequential processing are also done. The final step of the ETL process is the loading of data. Loading in the data warehouse usually takes the form of flat files, relational tables, quality assured dimensional tables.

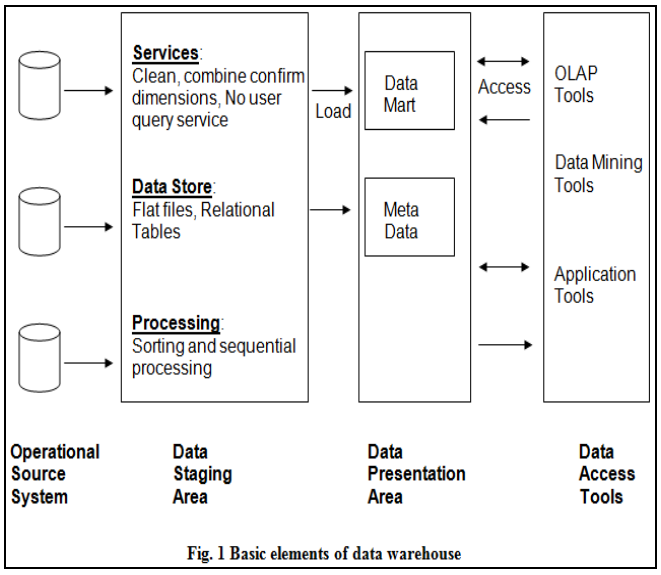

Fig. 1 Basic elements of data warehouse

3 Data Presentation Area The data presentation area is where data is organized, stored, and made available for direct querying by user, report writers, and other analytical applications (i.e. data access tools). The presentation area is a series of integrated data marts containing many fact tables

and decision tables. The dimensional modeling is the most viable technique for delivering data to data warehouse users for making databases simple and understandable.
The dimensional modeling is quite different from third normal form (3NF) modeling is a design technique that removes data redundancies. Hence dimensional model contains the same information as a normalized model but packages the data in a format whose design goals are user understandability, query performance and resilience to change.
In presentation area data mart may contain detailed, atomic data. All the data marts must be build using common dimensions and facts. Without shared, conformed dimensions and facts, a data mart is a stand alone. Each data mart may contain several fact tables, each with 5 to 15 dimension tables. If the design has been done correctly many of these dimension tables will be shared from fact table to fact table.
The data be presented, stored and accessed in dimensional schema. If the presentation area is based on relational database, then these dimensionally modeled tables are referred to as star schema. If the presentation area is based on multidimensional database or Online Analytic Processing (OLAP) technology then the data is stored in cubes.
Dimensional modeling is applicable to both relational and multidimensional databases. Both have common logical design with recognizable dimensions but their physical implementation differs [7].

4 Data Access Models The principal purpose of data warehouse is to provide information to business users for decision making. These users interact with data warehouse using these data access tools. Many of these tools require an information specialist or a domain expert, who can analyze the information and can interact with the data warehouse environment in order to reach meaningful conclusions. There are following four category of data access tools: Query and Reporting Tools, Application Tools, OLAP Tools, Data Mining Tools

1 **Query and Reporting Tools** Query and reporting tools are simple and built-in data access tools. This category of tools can be further subdivided into two groups: reporting tools and managed query tools.

Reporting Tools These tools can be further subdivided into two groups: Production Reporting Tools and Report Writers. Production reporting tools let companies to generate regular operational reports. Report writers are the tools designed for end users.

Managed Query Tools These tools makes the query task easier for end user by eliminating complexity of SQL and database structures and inserting the meta layer between end user and database. Meta layer is a software that provides subject oriented view of data base. Hence these tools are designed for ease to use and can automatically generate SQL statements and accept SQL statements.

2 Applications Tools These application tools provides more facilities to end users than the built-in query and reporting tools. These tools provide complex set of queries and data models. So the end user must be SQL and data modelling expert. Example of such application tools is Power Builder from Power Soft, Forte from Forte Software etc [8].

3 OLAP Tools Online analytical processing (OLAP) tools are based on the concept of multidimensional databases. OLAP tools are capable of analyzing online a large number of past transactions and large number of records and summarize them. This type of data is usually multidimensional in nature i.e. the data is organized in a multidimensional model and supported by a multidimensional database. The OLAP tools can be of three types: Multidimensional OLAP tools (MOLAP), Relational OLAP tools (ROLAP), Hybrid OLAP tools (HOLAP). Examples of some OLAP tools are Meta Cube from Informix, Oracle Express, Cognos Power Play etc.

III APPLICATIONS OF DATA WAREHOUSE IN PUBLIC LIFE

Importance of DWH cannot be denied due to its benefits because decisions at management level will no longer need to be taken on the limited and inaccurate data and it also helps the companies to avoid different challenges. So it becomes the need of every individual company to implement data warehouse. It is estimated that by 2020 around 200% more devices will join the Internet and share data. DWH strongly depends upon devices and inter linked data. The more interlinked devices are, the more powerful and useful DWH. According to the forecast by many organization [25, 26] by 2016 around 6.4 billion connected peers will join the room globally, an increase of 30% from 2015. Cisco and other research agencies [25, 26] think that approximately 20 - 50 billion devices will be connected by 2020[25,26].
Other side of the picture is that cost will increase too. If we talk about spending on hardware, the applications related to consumer will hit to $546 billion by the end of 2016; apart from that the usage of connected items in the organization will be somewhere around $868 billion by the end of 2016 [25, 26].
As we can see that different levels are defined. These levels are associated with the hierarchy such that first level is the core component. The first level is always be a central DWH (core system(s), hardware system(s)). Furthermore, 2nd level is associated with one of the world's top domains (Root level, business and Government). The reason behind selecting Business and Government as top
Figure 2 shows the cycle of public life applications of data warehouse in different fields and how they are interrelated according to user preference. As we can see that different levels are defined. These levels are associated with the hierarchy such that first level is the core component. The first level is always be a central DWH (core system(s), hardware system(s)). Furthermore, 2nd level is associated with one of the world's top domains (Root level, business and Government).

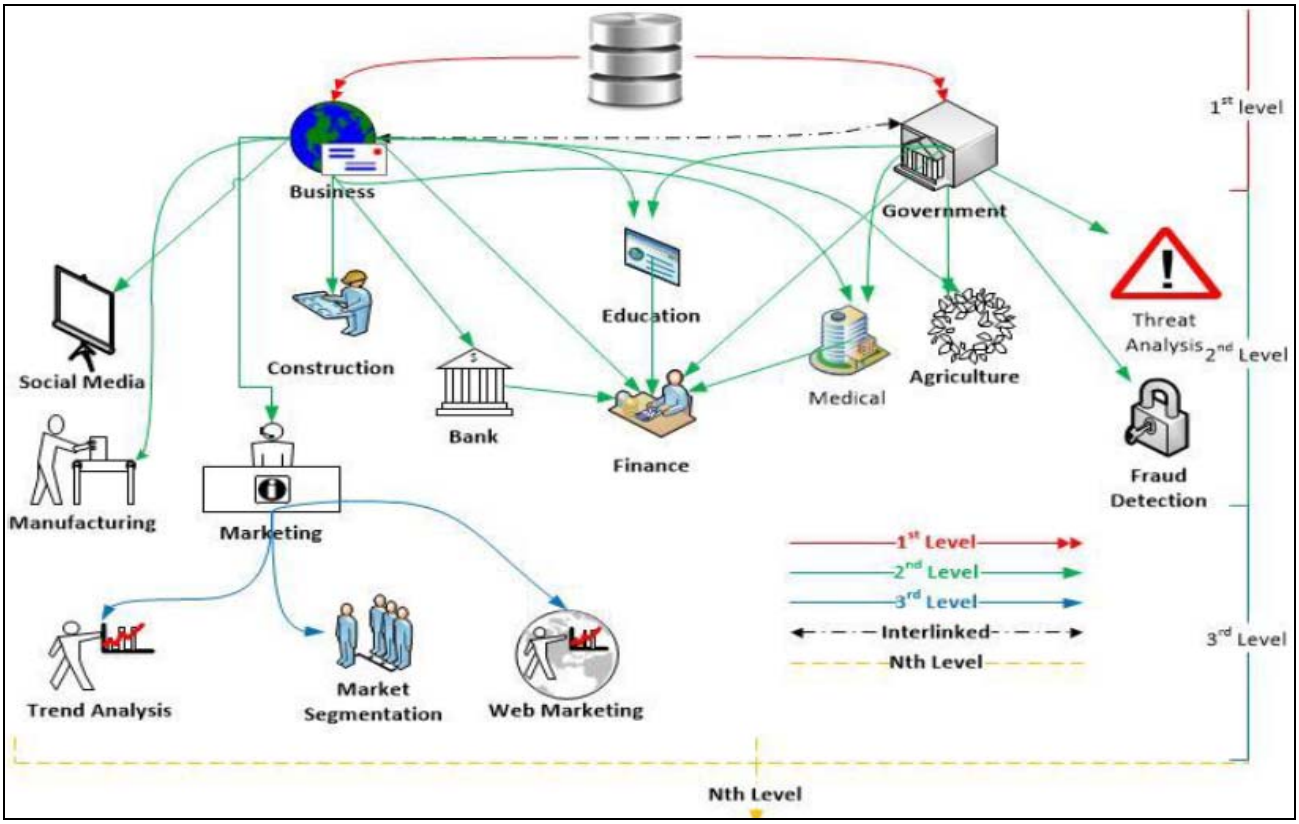

Fig 2 Applications of Data Warehouse in Public Life

The reason behind selecting Business and Government as top of hierarchy is a handful of literature, and all other domains are encapsulated under them. With the presence of 2nd level all other sublevel gets populated. The 2nd level serves as the only pillar that supports all other domains. 2nd level is said to be a specific level. 3rd level domains are the more general than specific. The Nth level is the most general level that holds all minor to major domains

A. Buisness

Improvement related to decision making and increasing organizational performances are the basic reasons to adopt DWH in business [27]. Business holds a key location in applications of data warehouse. All other private and semi-private organizations come under its umbrella.

In DWH, for easiness a single repository is used to store data, which is extracted from different databases. This data repository provides forecasting which helps the business personals and business managers. This complete cycle is used to help in identifying the requirements for business and to draft a plan for business [28] as shown in fig 2.

1) Social media websites

Social media is a great example of data warehousing. Social media industry is emerging and so is the need to implement DWH in it. A number of features from Facebook, Twitter and other social media sites are also based on analyzing large data sets [29]. It gathers all data like groups, likes, friends, location mapping etc. and stores it in a single central repository. Although all this information is stored in separated databases but the most relevant and significant information is stored in a central aggregated database [28].

2) Construction (material based industries)

Data warehouse approach in construction industry seems to be efficient in decision making as it provides construction managers the complete internal and external knowledge about available data so that they can measure and monitor the construction performance.

Application of DWH in construction industry clearly shows that construction bosses can smartly judge the stock remaining, inventory related trend linked to the materials, the amount and quantity of each material and also the price of all materials [30, 56]. It would also be helpful in reasonable resource allocation to fulfill the required services, maintenance and operation of the systems, allocation of financial budgets, effective managing of investment related long term plans and identification of potential risks [31].

3) Manufacturing Industry

DWH plays a vital role in daily house to industrial hold things. Manufacturing industry includes product and process design, scheduling, planning, production, maintenance and huge investments in equipment, manpower and heavy machinery. In this scenario, decisions taken will have wide-ranging effects in terms of profitability and long-term strategic issues. Many industries are trying to convert themselves and many should adopt DWH technology rather than traditional decision making so that a warehouse gathers, standardizes and stores data from various applications for improvement in processes and increasing its efficiency as analyzing the data in separate applications is time-consuming. At this stage, some

transaction processing systems, which are updated timely, are often hired to propagate the routine business of manufacturing and construction companies [56, 57].

4) Marketing

Every business is not successful without proper marketing and marketing is not successful without knowing the latest trends and demands. Relationship marketing is a new terminology linked with how different businesses handle their customers and the relationships in between that are assets for them and how they can be improved for long-term profitability. DWH in marketing is used to examine the patterns of customer's behavior and use this customer information for implementing relationship marketing. They play a vital role in identifying and targeting the profitable customers [32].

a) Trend Analysis

It is a technique that is used to predict future outcomes from historical results or information. Different medium to large scale enterprises are converting to this. In trend analysis, DWH can be used to examine the behaviors of the customer by using historical records over consecutive months.

b) Web Marketing

Web is a hub of billions of devices and around 20 – 50 billion devices till 2020. It refers to a category of advertising that includes any marketing activity conducted online. Facebook, google, and many major to minor such like sites uses web marketing and are relying on latest updated data warehouse.

c) Market Segmentation

Behavior identification is the top most priority of any organization. Market segmentation is the identification of the customer's behavior and common characteristics related to the purchases made against that product of related company. Many organizations are focusing on integrating data warehouse to get best behavior analysis.

5) Banking

The banking industry is categorized as one of the highest information demanding industry in the business world. With the advancement in information technology sector, the role of business intelligence (BI) increases with great number in the process of banking operations [54]. The increased business speed and growing competition has shown the need of banking intelligence dramatically. Bank intelligence is the ability to gather, manage, and analyze a large amount of data on bank customers, products, operations, services, suppliers, partners and all the transactions. As data increases, it becomes difficult to collect, handle and transform it into useful knowledge and DWH solves this problem. Many data warehouse flavors are designed for the support of banking industry.

6) Education

DWH in education field is becoming popular day by day. Use of DWH in educational field presents several potential benefits in making appropriate decisions and for evaluating data in time which is the basic target of DWH process. DWH provides an integrated and total view of an institute [33]. Most of the related departments use data warehouse as a source of information about faculty and students. DWH helps the students in getting their results and notes from a web enabled database quickly through a student portal and last but not the least it helps in decision making by providing current and historical information of the institute. On a large scale, a DWH can integrate the information of different institutes into a single central repository for analysis and strategic decision making.

7) Finance

With the advancement in technology, especially IT industry has opened the doors to the new ways of handling business considering financial systems. Government and Business domain holds equal part in finance. Financial systems may include banks, post offices, insurance companies, income tax and all other tax departments etc. Implementation of data warehouse in financial industry has several benefits e.g. it can maintain transparency in account opening and transactions. Similarly, government can take decisions against any financial crises. These systems are intelligent enough to spot the defaulters and may act according to the situation. As data warehousing is maintained in this scenario so efficient decision making process can easily be performed. These data warehouses in finance applications can also be used for the analyzation and to have forecasting of different aspects of business, stock and bond performance analysis [34, 58, 60].

8) Inventory

It can be defined as storage of items and availability of items. In any departmental store whose monthly turn over is more than 1000000/- it deals with hundreds of items. It has good turn over and reputation in city. Firm employed 10 or more employees. One person maintains the account of firm. Seven staff members are for counter sale and three for outstation work time. The account gets information from counter salesman such as daily sales, cash slips, information purchase all goods from wholesaler firm adopt a simple policy for ordering new items when stock goes down to P-items (called reorder point) and will order Q more items (called reorder quantity) from wholesaler. The firm has to balance the reorder print and reorder quantity for properly running the store.

Thus the firm build a project for data warehouse in which huge amount of information (such as daily sale, purchase, cash, credit, back sales, bills, reorder cost, reorder point, reorder quantity of hundred of items) is stored on the daily basis. After a long run of firm it may be needed for owner of firm to analyze the total investment and total profit from firm in monthly, yearly basis. This analysis can be done by applying data mining tools on data stored in data warehouse.

9) Telecommunication

A large number of wireless telecommunication companies is interested in data warehouse technology which already has collected raw material from multiple sources concerned with calls incoming or outgoing, technical data, customer billing data. The data warehouse team will format raw data

for consistency in data warehouse and check the data for validity and quality Assurance. Then this data will be fitted into data model of data warehouse. The data mining tools and techniques are applied on this data stored in data warehouse for telecommunication companies and extracting the useful and hidden information to take future decisions.

10) CRM (Customer Relationship Management)
CRM is the process that manages that interaction between company and its customer. The interacting with customers is not simple as it has been in past. The way in which companies interact with their customers has changed. A customer is continue business is no longer granted. As a result companies have found that they need to understand their customers better. Company need to know their customers preference not only for products but also for style, service etc. They need to manage the relationship with each and every customer and make each as profitable customer. Those companies that are successful will find increased revenue at lower cost of sales and work time. The methodology that makes this possible is called customer relationship management (CRM).
Many organizations have collected and stored a huge amount of data about their customers, suppliers and business partners. However inability to discover valuable information hidden in the data prevents them to take benefits from this large database. The business desire is to extract valid, previously unknown information from this large data base and use it for profit. To fulfill these goals the organization needs data ware- house to manage this large amount of data and then apply the data mining techniques to extract the useful information for decision making [7].

11) Science and Engineering
Advances in technology are making massive data sets common in many scientific disciplines. To find useful information in these data sets, scientists and engineers are turning to data mining techniques.
In recent years, data mining has been widely used in area of science and engineering, such as bioinformatics, genetics, medicine, education and electrical power engineering ,medical imaging, astronomy, chemistry, remote sensing, and physics. In the area of study on human genetics, the important goal is to understand the mapping relationship between the inter-individual variation in human DNA sequences and variability in disease susceptibility. In many terms, it is to find out how the changes in an individual's DNA sequence affect the risk of developing common diseases such as cancer. This is very important to help improve the diagnosis, prevention and treatment of the diseases. The data mining technique that is used to perform this task is known as multifactor dimensionality reduction.

12) Space Applications
Space Applications is the application of data warehouse techniques to spatial data. It is called Spatial data mining.
Spatial data mining is the application of data mining techniques to spatial data. Spatial data mining follows along the same functions in data mining, with the end objective to find patterns in geography. So far, data mining and Geographic Information Systems (GIS) have existed as two separate technologies, each with its own methods, traditions and approaches to visualization and data analysis. Particularly, most contemporary GIS have only very basic spatial analysis functionality. The immense explosion in geographically referenced data occasioned by developments in IT, digital mapping, remote sensing, and the global diffusion of GIS emphasizes the importance of developing data driven inductive approaches to geographical analysis and modeling.
Data mining, which is the partially automated search for hidden patterns in large databases, offers great potential benefits for applied GIS-based decision-making.

B. Government
Amongst the two major sub-divisions of DWH industry, government holds equal division. Government can use data warehousing technique in different fields e.g. for searching terrorist profile and threat assessments, in agriculture, in educational industry, in financing department, medical departments and for fraud detection. The telecommunication industry and Banking industry holds many issues related to user frauds. Figure 2 shows application of data warehouse in government departments.

1)NIC
The general information service terminal of national informatics center (GISTNIC) data warehouse for Tamil Nadu was implemented in 1998. In this data warehouse the information was collected on village level such sector as education, health, rainfall, data on individuals below poverty line (BPL Survey) etc. This GISTNIC data warehouse is a web enabled data warehouse which provides information on national issues ranging across diverse subject like food, agriculture and latest updates in science and technology. This information was collected to fullfill the needs of politicians, economists and to all the citizens. The web enabled SAS software is used as data mining tool which provide online information to these decisions makers in the government sector [9].

2) *World Bank :* The world bank collects and maintains huge amount of data of economic and development of all the countries of world. For the purpose of monitoring of various world bank's project in all the countries of world the world bank started collecting data and analysing it. For more than hundred developing countries the world banks also captured data on their economics and financial credit [9]. The world bank build data warehouse and analyse it using OLAP tools which is not possible manually.

3) Medical
Medical sector is emerging as the highest DWH implementer industry. In health-care, data quality and demand for quality medical services has become increasingly important [55, 59]. Due to the intricacy and variety of medical cum clinical data, the adoption of data warehouses by health care was slow as compared to other fields. Over the past few years it was reported that the usage

of DWH increased by the administrative and clinical areas. Data warehouses can help in improving the care of specific patients. These health-care institutions are adopting data warehousing for strategic decision making as a decision supporting tool. It provides the tools for acquiring medical data, for extracting the relevant information from that data and finally making this knowledge available to all the concerned persons. Administrative data in data warehouse can help in providing the information about skilled staff needed for a particular treatment and this information further used for the treatment scheduling and to help supporting medical personals in human resources area [36].

4) Data warehouse for Ministry of Rural Development:

The Ministry of Rural Development (MRD) has been regularly reviewing the data on Rurals. The analysis of this large amount of data become difficult. But the data warehouse technology can make it easy to analyse which includes all analysis variables into consideration. Rural development data warehouse can be further developed by collecting more detail information about the infrastructure of the districts. Thus rural development data warehouse aims at making the rural infrastructure data's beneficial for the society. It also provides ready to use data to the people at Ministry of Rural Development.

1) To reduce the amount of resources – time and manpower- spent on managing the volumes and variety of databases handled by Ministry of Rural Development.

2) To provide powerful decision making tools in the hands of end-users (at Ministry of Rural Development) in order to facilitate better decision making for rural development programs.

5) Fraud and Threat detection

Governments are playing their part to detect any threat and fraud caused by ill-minded people. Unfortunately, almost no specific data warehouse implementation that is known is available. Data warehouse access to governments are there, but they need a data warehouse system that is linked with every corner so that threats and terrorists will be monitored

.

IV. CASE STUDIES

In this section few case studies are discussed. As discussed earlier data warehouse world is a blend of two parts i.e. business side and government side. Both sides have their own further divisions and any other increment will be added under them. A graphical view is presented in the Figure 3, which is related to the contribution made by business and government domains to DWH. It is clearly observed by the survey that 80% of Business and 20% Government related organizations are contributing in the progress of data warehouse.

A. Business

DWH in business is now emerging like a hurricane. Around 80% of data warehouse implementation is captured by business. Following are few case studies related to business implementation of data warehouse.

1) Finance

Financial services company (FSC) is considered to be the leading marketer of investment besides banking for products. They implemented DWH named as VISION. The user of VISION consists of financial and marketing analysts, managers. It was developed with substantial business and technical goals that can gave a factual and precise picture of best customers of banks and also about most important products [27].

2) Medical

This case study is based on generation of evidence-based guidelines performed by University Health Network (Toronto) which clearly showed that it is authentic, influential and user-friendly to have a DWH related to clinic for best strategic decision making. Without this IT support, it would not be imaginable to look for evidence-based medicine as it is difficult for clinicians to gather data for a specific disease [36].

3) Banking

Their research problem is based on the factors that banking industry should consider before and during the adoption of DWH technology. Their results revealed the number of banks in Taiwan that adopted this technology and also the architectures that these banks implemented [5].

4) Manufacturing

Large Manufacturing Company (LMC) is making its way to top for production of home related appliances. LMC implemented data warehouse technology as there is a great need to improve the technical infrastructure of the company. Before this, data was scattered in different formats throughout the company and this makes normal and basic functioning difficult for business units. This warehouse provides support to marketing, manufacturing and logistic applications by providing data to dependent data marts [27].

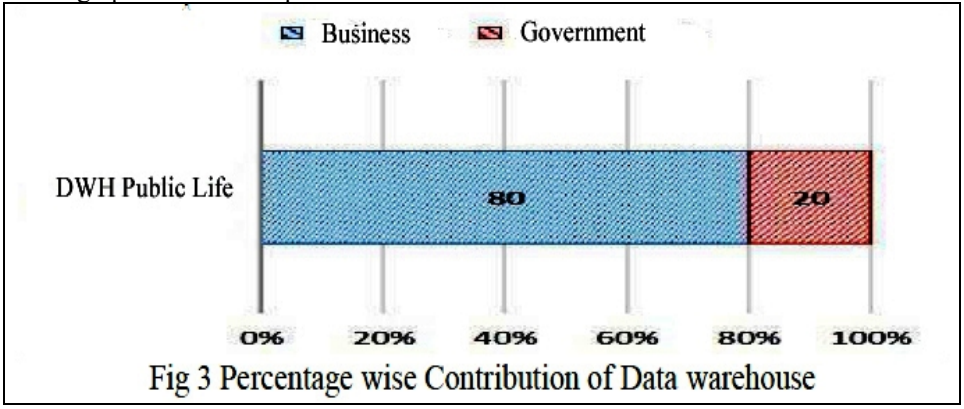

Fig 3 Percentage wise Contribution of Data warehouse

B. Government

Data warehouse in government plays a vital and critical role. Around 20-35% of data warehouse industry is captured by government. Many developing countries are now transferring to the use of data warehouse. Few case studies related to government and usages of data warehouse are as follows.

1)NIC

The general information service terminal of national informatics center (GISTNIC) data warehouse for Tamil Nadu was implemented in 1998. In this data warehouse the information was collected on village level such sector as education, health, rainfall, data on individuals below poverty line (BPL Survey) etc. This GISTNIC data warehouse is a web enabled data warehouse which provides information on national issues ranging across diverse subject like food, agriculture and latest updates in science and technology. This information was collected to fullfill the needs of politicians, economists and to all the citizens. The web enabled SAS software is used as data mining tool which provide online information to these decisions makers in the government sector [9].

2) Medical

In Utah and Idaho, Intermountain Healthcare implemented EDW. This healthcare system operates 22 hospitals, 179 clinics, physician offices. This case study is about venous thrombosis patients. Datasets consists of: records of Inpatients, columns of outpatient, financial data linked to or from patient's accounts, data from laboratories related to clinics for the process of imaging and surgery [35] etc. Their DWH is updated each night that includes: Large Metadata Repository, Security and auditing infrastructure and Master Reference Data. By using latest information from data warehouse patients with high risk are identified and their reports were sent at every hospital or clinic [35].

3) Finance

Internal Revenue Service is the agency of U.S. that is responsible for tax collection and tax laws imposition. They implemented data warehouse CRIS as there is no way to recoup entity with convinced attribute and perform some analysis on these marked entities. This implemented DWH consisted of five domains: business entity, tax returns entity, related to taxpayer transactions entity, peoples' income sources entity and tax payments details entity [27].

V. ANALYSIS AND RESULTS

In this section we will see the areas, cross domains and usage of data warehouse around the world and the graphical view of inter related data effecting data warehouse.

A. Comparison of different cross domain areas affecting data warehouse

Table 1 shows the comparison of different cross domain areas and their interlinked data.

B. Graphical representation of Survey

Following graph shows percentage captured by different areas in DWH around the world. As we can see from the Figure 4, medical holds top position in using DWH technology.

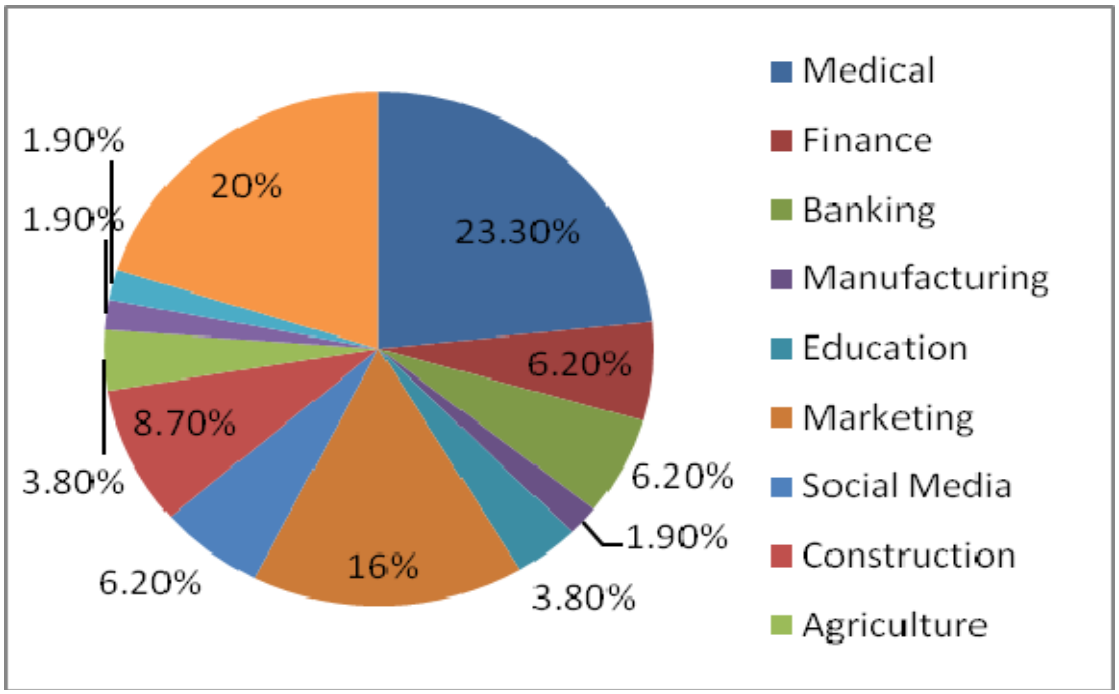

Fig 4 Percentage Distribution of DWH in Public Life

TABLE I. COMPARISON OF DIFFERENT CROSS DOMAIN AREAS AFFECTING DATA WAREHOUSE

Domain	Area of Usage	Percentage used	Cross domain	Reference No.
Medical	Hospitals, Clinics, Physician offices	23.30%	Government/Business	[37,38,40,39,35,41,42,36,28,12]
Finance	Tax departments	6.20%	Government/Business	[34,45,28,27]
Banking	Baking industry all around the world	6.20%	Business	[28,5,46]
Manufacturing	Home appliances	1.90%	Business	[27]
Education	Schools, colleges, universities	3.80%	Government/Business	[33]
Marketing	Customer relationship management, trend analysis and information system	16%	Business	[28, 47, 48, 49, 50,32, 52]
Social Media	Facebook, Twitter, others.	6.20%	Business	[29]
Construction	Infrastructure management	8.70%	Business	[43,44,31]
Agriculture	Agricultural production department	3.80%	Government/Business	[6]
Fraud Detection	Airports, Crime Agencies	1.90%	Government	[37,38,40,39,35,41,42,36,28,12]
Threat Analysis	Airports, Crime Agencies	1.90%	Government	[34,45,28,27]
Others	Others	20%	All	[28,5,46]

VI. DISCUSSION AND CONCLUSION

This research survey describes the applications of data warehouse in various domains including government and non-government organizations. Our analysis is based on the literature review and case studies provided in this survey. The analysis of this study shows that the non-governmental organizations use data warehouse technology much more than the government organizations. The governments mostly use data warehouse for controlling the crime and fraud. Non-governmental organizations mostly use DWH for data analysis, prediction and making decisions. Case studies describe the importance of data warehouse in four domains; Healthcare, Banking, Finance and Manufacturing. The details of these case studies and their use of data warehouse have been discussed in the Section 4. The analysis of the Table 1 shows that data warehouse is being used in many application domains. The Figure 4 clearly depicts the areas that are using data warehouse. It shows that medical and marketing areas are using data warehouse much more than the other domains, whereas manufacturing, agriculture, education, and government sector are rarely using data warehouse.

The analysis shows that data warehouse technology have been adopted in business as well as in government organizations for managing their huge data and for decision making. Still many organizations have not gone for the adoption of DWH technology. Either they do not realize its importance or there may be difficulties in its adoption. The reasons for ignoring the importance of implementing DWH technology have been discussed in literature that include quite large investment in terms of capital, more time utilization, looking for intangible benefits are difficult, the last but not the least problems holding with recent data management systems' infrastructure etc.

REFERENCES

[1] T. Ariyachandra, H. J. Watson, "Key organizational factors in data warehouse architecture selection", Decision Support Systems 49 (2010) 200–212.

[2] T. R. Sahama, P. R. Croll, "A Data Warehouse Architecture for Clinical Data Warehousing", in Roddick, J. F. and Warren, J. R., Eds. Proceedings Australasian Workshop on Health Knowledge Management and Discovery (HKMD 2007) CRPIT, 68, pages pp. 227-232, Ballarat, Victoria.

[3] W.H. Inmon., "DW 2.0 Architecture for the Next Generation of Data Warehousing", DM Review, Apr 2006, Vol. 16 Issue 4, p.8-25.

[4] W.H. Inmon, "Building the Data Warehouse", Third Edition, York: John Wiley & Sons, 2002.

[5] Hwang, Hsin-Ginn, et al. "Critical factors influencing the adoption of data warehouse technology: a study of the banking industry in Taiwan." *Decision Support Systems* 37.1 (2004): 1-21.

[6] Nilakanta, Sree, Kevin Scheibe, and Anil Rai. "Dimensional issues in agricultural data warehouse designs." *Computers and electronics in agriculture* 60.2 (2008): 263-278.

[7] Alex Berson, Stephen Smith, Kurt Thearling "Building Data Mining Applications for CRM",2001.

[8] C.S.R.Prabhu, "Datawarehouse Concepts, Techniques, Products and Applications",2nd Edition 2004.

[9] Margy Ross, Ralph Kimball, "The Data warehouse Toolkit", 2nd Edition 2002.

[10] J.V.D. Hoven, Data warehousing: bringing it all together, Information Systems Management (1998 Spring) 92 – 96.

[11] R. Kimball, The Data Warehouse Toolkit, Wiley, New York, 1996.

[12] R.M.T. Lu, K.A. Mazouz, A conceptual model of data warehousing for medical device manufacturers, Proc. of the 22nd Annual EMBS International Conference 2000 (July).

[13] D. Powell, To outsource or not to outsource? Netwo rking Management (1993) 56 – 59.

[14] Y. Yao, H. He, Data warehousing and the Internet's impact on ERP, IT Professional (2000 March) 37–41.

[15] Rob, P., Coronel, C., 2006. Database Systems: Design, Implementation, and Management. Course Technology.

[16] Sen, A., Sinha, A.P., 2005. A comparison of data warehousing methodologies. Commun. ACM 48 (3), 79–84

[17] Kimball, R., 2002. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling. John Wiley & Sons, Inc.

[18] Alsquor, M., Matouk, K., Owoc, M. L., A survey of data warehouse architectures:preliminary results. Proceedings of the Federated Conference on Computer Science*and Information Systems*,Wroclaw,2012,Sivut 1121-1126.

[19] Hackney, D., 2002. Architectures and Approaches for Successful Data Warehouses, Oracle White Paper.

[20] CHAKIR, Aziza, Hicham MEDROMI, and Adil SAYOUTI. "Actions for data warehouse success." Editorial Preface 4.8 (2013).

[21] Chaudhuri, S., Dayal, U., 1997. An Overview of Data Warehousing and OLAP Technology. ACM SIGMOD Record 26 (1), 64–74.

[22] Thakur, Garima, and Anjana Gosain. "A Comprehensive Analysis of Materialized Views in a Data Warehouse Environment." IJACSA) International Journal of Advanced Computer Science and Applications 2.5 (2011).

[23] Watson, H.J., Haley, B.J., 1997. Data warehousing: a framework and survey of current practices. J. Data Warehousing 2 (1), 10–17.

[24] Watson, H.J., Gerard, J.G., Gonzalez, L.E., Haywood, M.E., Fenton, D., 1999. Data warehousing failures: case studies and findings. J. Data Warehousing 4 (1), 44–55

[25] www.informationweek.com/mobile/mobile-devices/gartner-21-billion-iot-devices-to-invade-by-2020/d/d-id/1323081

[26] www.gartner.com/newsroom/id/3165317

[27] Watson, Hugh J., Dale L. Goodhue, and Barbara H. Wixom. "The benefits of data warehousing: why some organizations realize exceptional payoffs." *Information & Management* 39.6 (2002): 491-502.

[28] Joseph, Madhuri V. "Significance of Data Warehousing and Data Mining in Business Applications." *International Journal of Soft Computing and Engineering (IJSCE) ISSN* (2013): 2231-2307.

[29] Thusoo, Ashish, et al. "Data warehousing and analytics infrastructure at facebook." *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*. ACM, 2010.

[30] Chowdhury, Rajdeep, et al. "Implementation of Central Dogma Based Cryptographic Algorithm in Data Warehouse Architecture for Performance Enhancement." International Journal of Advanced Computer Science & Applications 1.6: 29-34.

[31] Park, Taeil, and Hyoungkwan Kim. "A data warehouse-based decision support system for sewer infrastructure management." *Automation in Construction* 30 (2013): 37-49.

[32] Ryals, Lynette, and Adrian Payne. "Customer relationship management in financial services: towards information-enabled relationship marketing." Journal of strategic marketing 9.1 (2001): 3-27.

[33] Goyal, Monika, and Rajan Vohra. "Applications of data mining in higher education." International journal of computer science 9.2 (2012): 113.

[34] Bhedi, Vaibhav R., Shrinivas P. Deshpande, and Ujwal A. Lanjewar. "Data Warehouse Architecture for Financial Institutes to Become Robust Integrated Core Financial System using BUID." International Journal of Advanced Research in Computer and Communication Engineering 3.3 (2014): 2278

[35] Evans, R. Scott, James F. Lloyd, and Lee A. Pierce. "Clinical use of an enterprise data warehouse." AMIA Annual Symposium Proceedings. Vol. 2012. American Medical Informatics Association, 2012.

[36] Stolba, Nevena, and A. Min Tjoa. "The relevance of data warehousing and data mining in the field of evidence-based medicine to support healthcare decision making." International Journal of Computer Systems Science and Engineering 3.3 (2006): 143-148.

[37] Schubart, Jane R., and Jonathan S. Einbinder. "Evaluation of a data warehouse in an academic health sciences center." International journal of medical informatics 60.3 (2000): 319-333.

[38] Liu, Baoyan, et al. "Data processing and analysis in real□world traditional Chinese medicine clinical data: challenges and approaches." Statistics in medicine 31.7 (2012): 653-660.

[39] Leitheiser, Robert L. "Data quality in health care data warehouse environments." System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on. IEEE, 2001.

[40] Yoo, Sooyoung, et al. "Electronically implemented clinical indicators based on a data warehouse in a tertiary hospital: its clinical benefit and effectiveness." International journal of medical informatics 83.7 (2014): 507-516.

[41] Prokosch, Hans-Ulrich, and T. Ganslandt. "Perspectives for medical informatics." Methods Inf Med 48.1 (2009): 38-44.

[42] Adlassnig, Klaus-Peter, et al. "Fuzziness in healthcare-associated infection monitoring and surveillance." Norbert Wiener in the 21st Century (21CW), 2014 IEEE Conference on. IEEE, 2014.

[43] Chong, Heap Yih, Rosli Mohamad Zin, and Siong Choy Chong. "Employing data warehousing for contract administration: e-dispute resolution prototype." Journal of Construction Engineering and Management 139.6 (2012): 611-619.

[44] Chau, Kwok-Wing, et al. "Application of data warehouse and decision support system in construction management." Automation in construction 12.2 (2003): 213-224.

[45] Chen, Wenzhe. "The Application of Data Warehouse Technology in Modern Finance." 2015 International Conference on Advances in Mechanical Engineering and Industrial Informatics. Atlantis Press, 2015.

[46] Lin, Zhonglin, et al. "Banking intelligence: application of data warehouse in bank operations." Service Operations and Logistics, and Informatics, 2008. IEEE/SOLI 2008. IEEE International Conference on. Vol. 1. IEEE, 2008.

[47] Shaw, Michael J., et al. "Knowledge management and data mining for marketing." Decision support systems 31.1 (2001): 127-137.

[48] Ngai, Eric WT, Li Xiu, and Dorothy CK Chau. "Application of data mining techniques in customer relationship management: A literature review and classification." Expert systems with applications 36.2 (2009): 2592-2602.

[49] Nedeva, Veselina Ivanova. "ANALYSIS OF MARKETING INFORMATION SYSTEMS AND CONCEPTION OF AN INTEGRATED MARKETING INFORMATION SYSTEM." International Journal of Computing 3.2 (2014): 127-133.

[50] Payton, Fay, and Debra Zahay. "Why doesn't marketing use the corporate data warehouse? The role of trust and quality in adoption of data-warehousing technology for CRM applications." Journal of Business & Industrial Marketing 20.4/5 (2005): 237-244.

[51] Thomas, Davenport, et al. "Data to Knowledge to Results, Building an Analytic Capability." California Management Review 43.2 (2001).

[52] Cunningham, Colleen, Il-Yeol Song, and Peter P. Chen. "Data warehouse design to support customer relationship management analyses." Proceedings of the 7th ACM international workshop on Data warehousing and OLAP. ACM, 2004.

[53] Watson, Hugh J., Celia Fuller, and Thilini Ariyachandra. "Data warehouse governance: best practices at Blue Cross and Blue Shield of North Carolina." Decision Support Systems 38.3 (2004): 435-450.

[54] Sarkar, Anirban. "Data Warehouse Requirements Analysis Framework: Business-Object Based Approach." International Journal 3 (2012).

[55] Diana, Nova Eka, and Aan Kardiana. "Comprehensive Centralized-Data Warehouse for Managing Malaria Cases." International Journal of Advanced Computer Science & Applications 1.6: 40-46.

[56] N.L. Sarda, Temporal issues in data warehouse systems, Database Applications in Non-Traditional Environments '99, TheProceedings of the 1999 International Symposium on Database Application in Non-traditional Environments (DANTE '99), IEEE Computer Society, Los Alamitos, 1999, pp. 27–34

[57] J.-B. Yang, N.-J. Yau, Application of case-based reasoning in construction engineering and management, Proceedings of the Third Congress held in conjunction with A/E/C Systems 1996, Computing in Civil Engineering, American Society of Civil Engineers, New York, 1996, pp. 663–669.

[58] K.W. Chau, Y. Cao, M. Anson, J.P. Zhang, Application of data warehouse and decision support system in construction management, Automation in Construction 12 (2) (2002) 213–224.

[59] J. Vanegas, P. Chinowsky, Computing in Civil Engineering, American Society of Civil Engineers, New York, 1996.

[60] J. Dyche, e-Data Turning Data into Information with Data Warehousing, Addison-Wesley, Reading, 2000