

A DATA MART APPROACH FOR A CENTRALIZED EGOVERNANCE DATA WAREHOUSE

V. V. SUBRAHMANYAM*, M. N. DOJA**

*School of Computer and Information Sciences, IGNOU, New Delhi, INDIA
**Department of Computer Engineering, Jamia Millia Islamia, New Delhi, INDIA

ABSTRACT

With many computer applications in place, large quantities of data have been collected over a period of years. Private organizations recognized that there is value in the historical data of their own organizations and have undertaken projects to build Data Warehouses (DW) to make the data accessible in a meaningful and timely manner through data mining and querying tools. But mostly in Government organizations, it is not so. DW and data mining technologies have extensive potential applications in the government – in various Central Government sectors such as Agriculture, Rural Development, Health and Energy etc.. We had selected this problem to design a feasible architecture in the context of Central Government of India. The basic aim of this paper is we had studied DW architectures implemented in private organisations and gave a thought to design a data mart approach architecture for a centralized eGovernance DW which covers all the major departments in the Central Government of India emphasizing the ways and means to select the subject-oriented areas for populating the data marts, implementation parameters, quality factors and at the end touched the issues like access and security involved in them. Also, we had covered presented a small case-study of a simple DW implemented in Andhra Pradesh State Government, India.

JEL CLASSIFICATION & KEYWORDS

O33 | O39 | Data Warehouse | Government | India Case Study

INTRODUCTION

Data warehousing has been in the commercial sector for a long time now. The first data warehouses emerged in the late-1980s and were called as "atomic data bases". But in the early 1990s data warehousing took off commercially with the advent of Extraction, transformation and loading (ETL) and Online Analytical Processing (OLAP). Soon, data warehousing blossomed into a full-fledged architecture known as the corporate information factory (the CIF). Research is in progress in many problem areas of DW (Widom, Jennifer Wisdom, 1995). There are several development trends going on in DW area as per the problem selection (V.V. Subrahmanyam, M.N. Doja, 2007).

Data warehousing spread across the business world like wildfire. Data warehousing started in the United States and eventually spread worldwide. Data warehouses are as common today in Malaysia as they are in Brazil, Australia, Europe and elsewhere.

A data warehouse (DW) is a huge data repository, which stores integrated information from various databases, for efficient querying and analysis. A data warehouse is a subject-oriented, integrated, time-varying, non-volatile collection of data in support of the management's decision-making process (Inmon, 2002). The information is extracted from heterogeneous sources as it is generated or updated. The information is then translated into a common data model

and integrated with existing data at the DW. Placing an adhoc query to the data warehouse whose data came from heterogeneous sources can retrieve complex information. The DW meant for eGovernance is a eGovernance-DW.

The key feature of DWs is that the tools to prepare reports are very user-friendly (web-based, using point-and-click technology). The DW can be valuable resource for all the users that need information to provide support for:

- Day-to-day operations
- Decision Support
- Strategic Planning
- Performance Management
- Compliance Reporting

From 1990's, they were very much accepted and are being used in private organizations very vividly. Today, there is an urgent need for data warehousing in government circles. As the volumes of data grow large and the need for new and innovative information becomes manifest, it becomes apparent that the organization or agency needs a data warehouse. But surprisingly, data warehouses have been slow to be adopted in the government circles (V.V. Subrahmanyam, M.N. Doja, 2007). The most fundamental reasons are:

- There is a significant difference in motivation for data warehousing in the commercial world and in the government. In the commercial world, the most fundamental motivations for data warehousing are to increase profit or increase market share protection. There are many other motivations for data warehousing in the commercial world, but these two motivations are the most basic and most visceral.
- Government agencies, on the other hand, try to optimize their resources while building a data warehouse to the benefit of the constituency they reach. They are not concerned with reducing the size of their department due to budgetary reasons and political power.

NEED AND BENEFITS OF A DW FOR EGOVERNANCE

Information is one of the valuable assets to any Government. Governments deal with enormous amount of data. When used properly, it can help planners and decision makers in making informed decisions leading to positive impact on targeted group of citizens (Beiber, 2008). However to use information to it's fullest potential, the planners and decision makers need instant access to relevant data in a properly summarized form. In spite of taking lots of initiative for computerization, the Government decision makers are currently having difficulty in obtaining meaningful information in a timely manner because they have to request and depend on IT staff for making special reports which often takes long time to generate. A DW can deliver strategic intelligence to the decision makers and provide an insight into the overall situation from the historical data. This greatly



A DATA MART APPROACH FOR A CENTRALIZED EGOVERNANCE DATA WAREHOUSE

facilitates decision-makers in taking micro level decisions in a timely manner without the need to depend on their IT staff. It permits several types of queries requiring complex analysis on data to be addressed by decision-makers (Chaudhuri and Dayal, Surajit Chaudhuri, Umeshwar Dayal, 1997). By organizing person and land-related data into a meaningful Information Warehouse, the Government decision makers can be empowered with a flexible tool that enables them to make informed policy decisions for citizen facilitation and accessing their impact over the intended section of the population.

Citizen facilitation is the core objective of any Government body. For facilitating the citizens of a state or a country, it is important to have the right information about the people and the places of the concerned territory. Hence, a DW built for eGovernance can typically have data related to person and the land. Such a DW can be beneficial to both the Government decision makers and citizens as well in the following manner:

How Decision-Makers Can Benefit?

- They do not have to deal with the heterogeneous and sporadic information generated by various state-level computerization projects as they can access current data with a high granularity from the DW.
- They can take micro-level decisions in a timely manner without the need to depend on their IT staff.
- Assimilated data which is otherwise scattered on different systems, in different departments about which a user might be unaware can be utilized directly from a DW.
- They can obtain easily decipherable and comprehensive information without the need to use sophisticated tools.
- They can perform extensive analysis of stored data to provide answers to the exhaustive queries to the administrative cadre. This helps them to formulate more effective strategies and policies for citizen facilitation.

How Citizens Can Benefit?

- They are the ultimate beneficiaries of the new policies formulated by the decision makers and policy planner's extensive analysis on person and land-related data.
- They can view frequently asked queries whose results will already be there in the database and will be immediately shown to the user saving the time required for processing.
- They can have easy access to the Government policies of the state.
- The web access to Information Warehouse enables them to access the public domain data from anywhere.
- The below is the case study of DW project for Multipurpose Household Survey (MPHS) of Andhra Pradesh Government.

A DATA WAREHOUSE PROJECT FOR EGOVERNANCE – A CASE STUDY

In India one of the state namely, Andhra Pradesh Government implemented a data warehouse of land and person data of 60 million population to enable well informed, timely and accurate policy decisions by the Government officials across various departments. Involved an outlay of Rs. 5 crores (US\$ 1 Million) to address the total State data. The Centre for Development of Advanced Computing (CDAC) in collaboration with the Andhra Pradesh

Technology Services (APTS) has developed a data warehouse for aiding the state level decision makers of Andhra Pradesh (AP) Government in their decision making process. The main objective of this effort is to organize the Multipurpose Household Survey (MPHS) data and the land records data of the AP Government into a meaningful information warehouse for enabling the decision makers in making informed decisions and accessing their impact over the intended section of the population.

The Microsoft Corporation India Private Limited helped implement the data warehousing solution, tailor-made to suit the needs of the State Government, and to streamline information gathering, analysis and application in the areas of Janmabhoomi programme, crop, treasury, land and rainfall data. A DW would help the State Government to collect, tabulate and mine the said data for effective decision making, better knowledge management as well as increased transparency. The solution is web-enabled for easy access and availability to officials. The Janmabhoomi data mart while the crops data mart will help officials to find information in seasonal irrigated areas for different crops up to the Mandal level. By allowing officials to predict rainfall on various areas in the State, the rainfall data mart is expected to significantly enhance planning for unforeseen circumstances. The treasury data mart was started with the primary objective of helping officials conduct sensitive analysis of State expenditure, such as fiscal deficit at given point of time. The data mart, approximately one giga byte big, covers data from 23 districts and includes 300 sub-treasuries and has consolidated data on all treasuries for one year.

The system installed in Andhra Pradesh Secretariat at Hyderabad is based on an 8-node PARAM 10000 configuration of C-DAC and provides a decision support capability to the state officials using industry standard tools and allowing analyses to be made on historical data with scalability and dynamism on data from Mandal to District to State levels. It also provides web-based access besides access on LAN set up within the Secretariat, through both thick and thin clients and kiosk with bilingual information. The data warehouse has enough potential to access the impact of various welfare schemes across the population of the state. The planners can design schemes focused on specific target groups and achieve high impact. The decision-makers can carry out analysis of population profile across the state in areas of economy, education, family units, shelter, etc. The warehouse can also be used for rural and urban development planning, agricultural yield and cropping patterns analysis and much more. These analyses will help in making decisions that are focused and the benefit of the government policies can reach the intended group. The various types and number of queries that can be handled by the data warehouse are limited only by the intelligence of the person using the data warehouse and the data fed to it. Let us see the architecture part of it:

- Since, Data warehousing solutions require high-end systems for storing, sending and analytical processing of a large volume of data. C-DAC's innovative PARAM 10000 architecture is an ideal platform for such solutions offered from desktops to a very high end computer systems.
- C-DAC has advented the OpenFrame Architecture for scalable and flexible High Performance Computing unifying the well known NOW (Network of Workstations), COW (Cluster of Workstations) and MPP (Massively Parallel processor) architectures. This architecture has been realized in C-DAC's new PARAM 10000 series supercomputers, which are scalable from the desktop to





A DATA MART APPROACH FOR A CENTRALIZED EGOVERNANCE DATA WAREHOUSE

teraflop range. The OpenFrame architecture of PARAM 10000 also realizes the server consolidation architecture required for building general-purpose High Performance Computing facilities.

- High Performance Secondary Storage of upto 1 Tera-byte capacity is based on SUN Enterprise Network Array A5000 of hot swappable FCAL (Fiber Channel Arbitrated Loop) disks supporting RAID (Redundant Array of Inexpensive Disks) levels 0, 0+1, 1, 5. A variety of industry standard tertiary storage systems can be interfaced to the PARAM system, depending on the usage requirements. The necessary support for Remote Site Mirroring for disaster recovery has been extended.
- They had used the industry standard products to deliver a reliable and lasting solution, and has therefore established a set of partnerships with the industry's best vendors like ORACLE and COGNOS.
- The advantages of the solution for decision support, data mining, statistical analysis, and ad-hoc reporting are even greater when the distribution of those applications can be accomplished using the web and therefore our solutions are web enabled.
- To provide support for Indian languages providing Data Warehousing solutions, they had incorporated own Multilingual Graphical Intelligence based Script Technology Solutions developed, if required.
- Some of the sample queries that can be handled by the system are, for example:
 - What is the percentage of people in different occupation - qualification-wise, religion-wise, age-group-wise?
 - How much is the unemployment in men or women versus age, area, and religion?
 - What is the growth rate of population region-wise versus resources food, shelter and education?
 - What is the percentage of land holding of people having income below certain level?
 - What is the crop-wise area and cultivation trend?

This solution is implanted to organize the MPHS data and the land records data of the AP Government into a meaningful information warehouse for enabling the decision makers in making informed decisions and accessing their impact over the intended section of the population.

Likewise, we should have a Data Warehouse which can accommodate all the departments in the state as well as at the central level. The subject oriented areas along with the architecture is proposed in the next section.

SELECTION OF SUBJECT-ORIENTED AREAS AND PROPOSED ARCHITECTURE

In 2002, Various types of approaches were discussed by Hackney, D in his paper. A large number of national data warehouses can be identified from the existing data resources within the Central Government Ministries. We can visualize various departments and ministries from the official website of India (India, NIC, 2004). Let us examine the potential subject areas on which data warehouse may be developed.

Population Census Data

The Registrar General and Census Commissioner of India decennially compiles the information of all the individuals, villages, population groups etc.. This information is wide ranging such as the individual-slip, a compilation of

www.researchjournals.co.uk

information of individual households, of which a database of 5% sample is maintained for analysis. A data warehouse can be built from this database upon which OLAP techniques can be applied. Data mining also can be performed from the analysis and knowledge discovery. We can perform multidimensional analysis of village level data in some sectors such as Education, Health and Infrastructure. There exists many other subject areas (eg. migration tables) within the census purview which may be amenable and appropriate on which work can be taken up in future.

Essential Commodities Prices

The Ministry of Food and Civil Supplies, Government of India, compiles daily data (on weekly basis) for about 300 observation centres in the entire country on the prices of essential commodities such as rice, edible oils, etc.. This data is compiled at the district level by the respective State Government agencies and transmitted online to Delhi for aggregation and storage. A data warehouse can be built for this data, and OLAP techniques can be applied for its analysis. A data mining and forecasting technique can be applied for advance forecasting of the actual prices of these essential commodities. The forecasting model can be strengthened for more accurate forecasting by taking into account the external factors such a rainfall, growth rate of population and inflation.

Agriculture Sector

The Agricultural Census performed by the Ministry of Agriculture, Government of India, compiles a large number of agricultural parameters at the national level. District-wise agricultural production, area and yield of crops is compiled; this can be built in to a DW for analysis, mining and forecasting. Statistics on consumption of fertilizers also can be turned into a data mart. Data on agricultural inputs such as seeds and fertilizers can also be effectively analyzed in a DW. Data from livestock census can be turned into a DW. Land use pattern statistics can also be analyzed in a data warehousing environment. Other data such as watershed details and also agricultural credit data can be effectively used for analysis by applying the technologies of OLAP and data mining. Thus there is substantial scope for application of data warehousing and data mining techniques in Agricultural sector (Abdullah, 2004).

Rural Development Sector

Data on individuals below poverty line can be built into a DW. The literacy status can be monitored in the rural area. Future plans can be formulated after mining the data and analyzing it. Drinking water census data can be effectively utilized by OLAP and data mining technologies. Monitoring and analysis of progress made on implementation of rural development programmes can also be made using OLAP and data mining techniques.

Health Sector

Community needs assessment data, immunization data, data from national programmes on controlling Swine Flu, Chicken Guinea, blindness, leprosy, malaria can all be used for data warehousing implementation, OLAP and data mining applications.

Planning

At the Planning Commission, DWs can be built for state plan data on all the sectors: human resources, health, labour, energy, education, trade and industry, five year plan etc. Monitoring and analysis of progress made on implementation of respective development programmes



A DATA MART APPROACH FOR A CENTRALIZED EGOVERNANCE DATA WAREHOUSE

various sectors can also be made using OLAP and data mining techniques.

Education Sector

The huge Educational Survey data has been converted into a DW. Various types of analytical queries and reports can be answered.

Commerce and Trade

Data bank on trade (imports and exports) can be analyzed and converted into a DW. World Price Monitoring systems can be made to perform better by using DW and data mining technologies. Provisional estimates of import and export also can be made more accurate using forecasting techniques.

Tourism Sector

Tourist arrival behavior and preferences, tourism products data, foreign exchange earnings data, and Hotels, travel and transportation data can be converted into a DW. Trends and patterns can be known using various types of analytical queries. Reports can be generated.

Revenue

Customs data, central excise data, and commercial taxes data can be converted into a DW. Various types of analytical queries and reports can be answered. Trends and patterns can be known using different querying tools.

The below shown Fig. 1 is the appropriate DW architecture for an eGovernance system with the said subject oriented areas.

Different data marts for separate departments, if built can be integrated into one DW for the government. In the government, the individual data marts are required to be maintained by the individual departments (or public sector organizations) and a central DW is required to be maintained by the ministry concerned for the sector concerned. A generic inter-sectoral DW is required to be maintained by a central body (as Planning Commission). Similarly, at the

State level, a generic inter-departmental DW can be built and maintained by a nodal agency, and detailed DWs can also be built and maintained at the district level by an appropriate agency. National Informatics Centre (NIC) may possibly play the role of the nodal agency at the Centre, State and District levels for developing and maintaining DWs in various sectors.

Various *Data Sources* are Document Management Systems, Customer Relationship Management (CRM), Enterprise Resource Planning (ERP), Web Service Applications, Enterprise Application Integration (EAI), Electronic Data Interchange (EDI), Groupware, government websites, various eGovernance applications of states and centre. Various data communication channels are mobile phones, digital TV, call-centres, Kiosks, PCs, Teleconferencing and Website.

The Data Marts (Subject Oriented DWs) should be populated from the transaction or other operational databases of the Government. Data from external sources may also fed. The next step is concerned with extracting data from multiple operational databases and from external sources; with cleansing, transforming and integrating this data for loading into the data warehouse server and of course, with periodically refreshing the warehouse. The four processes from extraction through loading are often referred to collectively as Data Staging.

Metadata, "data about data". It is useful to have a central information repository to tell users and the query tools what's in the DW, where to find it, who is authorized to access it, and what summaries have been pre-calculated.

The Data warehouse database itself. This database contains the detailed and summary data of the DW. Some people consider metadata to be part of the database as well.

Query Tools These usually include end-user interface for posing questions to the database, in a process called OLAP. They also include automated tools for uncovering patterns in the data, often referred to as Data Mining.

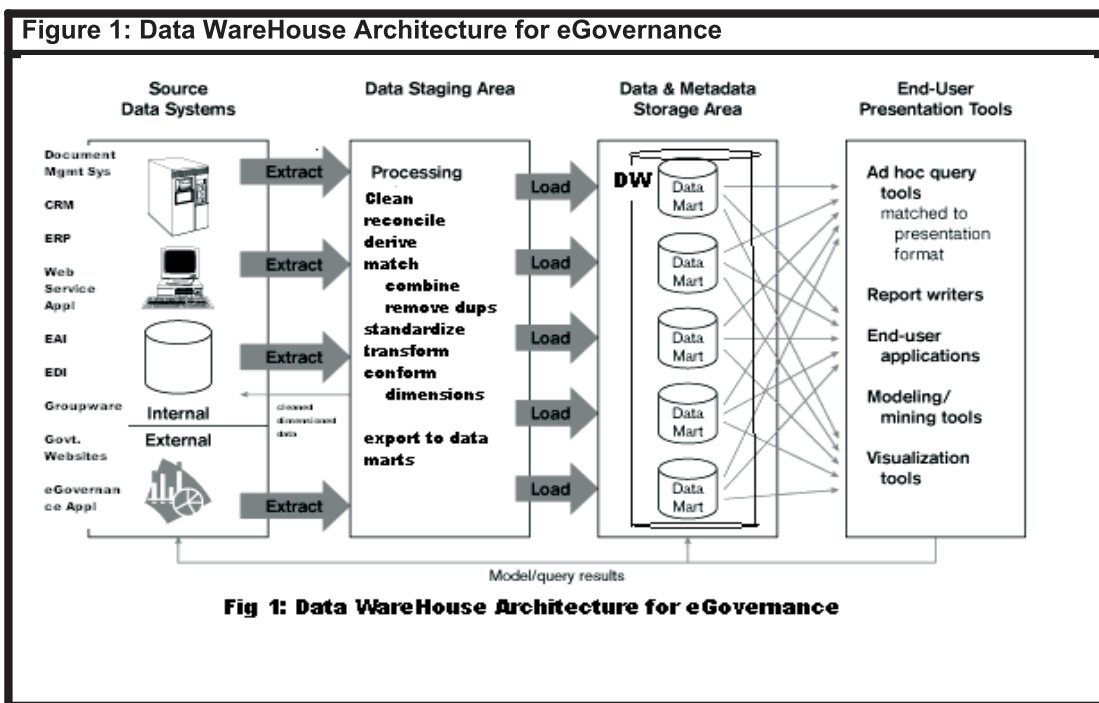


Fig 1: Data Warehouse Architecture for eGovernance



A DATA MART APPROACH FOR A CENTRALIZED EGOVERNANCE DATA WAREHOUSE

ACCESS PREMISSIONS TO THE DW

According to Subrahmanyam, Doja M.N (2009b), Government employees, Citizens, Businesses, other Government Departments, community members are to be given specific types of access to certain databases consistent with their position and job responsibilities. Access to the Data Marts / DW information is granted on a subject area basis. All system users must keep information obtained from system access confidential except as otherwise necessary to perform the task assigned. In all instances authorized system users and restricted access system users are responsible for having knowledge of and complying with all laws and Government policies relative to confidentiality. Access to the DW requires a signature to indicate acceptance of the eGovernance data access policies. Additionally, access must be approved by the appropriate Data Custodian for all the sectors-systems. Each subject area should have a data steward responsible for the approval of data access requests. Each data steward, along with governing policies, determines the access privileges on a case by case basis. Some of the eGovernance Data mart / DW subject areas are considered public information. This does not imply that approval is not required, rather that once approval is granted that all information within that subject area becomes available to you. Other subject areas within the eGovernance DW are more sensitive in nature, with privileges to that information governed by state and federal policies, as well as policies created and enforced by the Government. In all cases, when access is granted to the eGovernance DW, appropriate business use and transfer of information should be applied. Access privileges are systematically monitored and revoked on a frequent basis, to ensure a secure data access environment for the eGovernance. In order to request access, a person should complete a DW Access Form.

One of the benefits of the DW is allowing users to convert data into useful information. Often members of the same functional team or departments share similar reporting needs. The creation of user groups allows for the leveraging of group requirements with the development of reports. This allows Government constituents who use a common data mart to meet on a regular basis to promote an understanding of the groups' data and data utilization. The DW team will help coordinate the creation of a user group and will assist in the development of the team's requirements.

DATA QUALITY FACTORS IN DW

The objectives of Data Warehouse (DW) is to provide useful, accurate, relevant information for management decision making processes by integrating raw, unconnected data from both operational sources and stable data sets (V.V. Subrahmanyam, Doja M.N, 2008). Data quality is an important issue that should be accounted for starting with initial application design through implementation, maintenance and use (Berenguer, 2005). The quality of the reports is therefore only as good as the data entered into the operational system (William E. Winkler, 2004). As the DW continues to expand and add products to its suite, attention will inevitably be drawn to the quality of the data (Oliveira, 2005). The notion of data quality refers to the data's ability to confirm to requirements or be fit for use. Thus, data might be considered of high quality for one purpose, but of very poor quality for another. Typically, data quality attributes and objectives include:

- Accuracy - Data items may be valid but not necessarily accurate. Often it's necessary to cross check against other data items to ensure the data is accurate.

- Timely - Data items are available for reporting at (agreed) critical times during the processing cycle or at agreed snapshot dates.
- Relevant - Data items add some agreed value to the understanding of the aspect of the business activity being reported.
- Standardized - Where data items are measuring or categorizing the same attributes of real world objects, the same values are used to code or measure them, across all systems in the organization.
- Comparability - the same codes have the same meanings in all systems in which they are used.

CONCLUSION AND SUCCESS FACTORS

In 2002, in his contribution Watson highlighted various Data Warehousing failures elaborately in his case studies and findings. Governments deal with enormous amount of information and this itself is one of the valuable assets to any Government. The datawarehouse size depends upon the databases of the departments, we consider (Agosta, 2003). When used properly, it can help planners and decision makers in making informed decisions leading to positive impact on targeted group of citizens. However to use information to it's fullest potential, the planners and decision makers need instant access to relevant data in a properly summarized form. In spite of taking lots of initiative for computerization, the Government decision makers are currently having difficulty in obtaining meaningful information in a timely manner because they have to request and depend on IT staff for making special reports which often takes long time to generate. A DW can deliver strategic intelligence to the decision makers and provide an insight into the overall situation from the historical data. In order to support the Data mart / DW facility the authorities of the Government should:

- Appoint the technical staff for data administration activities, maintenance and for providing services to the staff for efficient use.
- A team should look after the selection of the centralized databases of various governmental departments to selected as the subject oriented areas for the DW (Harper, 2004).
- A Committee/task force may be appointed in order to look after the coordination activities and enforcing standards (William E. Winkler, 2004).
- The technical team should be able to ensure the data quality (Shankaranarayanan, 2006), which includes the intrinsic information quality, correctness, accuracy, consistency, completeness, contextual information quality, representational information quality, accessibility information quality, security and the ease of access to the users. If all these characteristics are in place, it gives the satisfaction to the user, who is uses it (Lei-da Chen , Khalid S. Soliman , En Mao , Mark N. Frolick, 2000).
- There should be an availability of the data dictionary with the technical team.
- Understanding of client's needs and a determination to meet those needs should be considered.
- Should have control on the size of the data in the Datawarehouse depending upon the problem and the computer application (Agosta, 2003).





REFERENCES

- Abdullah, The Case for an agri data warehouse: enabling analytical exploration of integrated agricultural data. In: Proceedings of the IASTED International Conference on Databases and Applications (DBA 2004).
- Agosta, Data warehouse size depends on the size of the business problem. DM Rev. v13 i16. 16-17, 2003.
- Bieber M., Data Warehousing in Government, DM Rev, 2008.
- Chaudhuri and Dayal, Surajit Chaudhuri , Umeshwar Dayal, An overview of data warehousing and OLAP technology, ACM SIGMOD Record, v.26 n.1, p.65-74, March 1997.
- Lei-da Chen , Khalid S. Soliman , En Mao , Mark N. Frolick, Measuring user satisfaction with data warehouses: an exploratory study, Information and Management, v.37 n.3, p.103-110, April 2000.
- Hackney, D., Architectures and Approaches for Successful Data Warehouses, Oracle White Paper, 2002.
- Harper, Data warehousing and the organization of governmental databases. IGI Publishing, 2004.
- India, N.I.C., 2004. Districts of India: A Gateway to Districts of India on the web, <http://www.districts.nic.in>, 2004.
- Inmon, William H. Inmon, Building the Data Warehouse, John Wiley & Sons, Inc., New York, NY, 2002.
- Kimball, Ralph , Margy Ross, The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, John Wiley & Sons, Inc., New York, NY, 2002.
- Oliveira, Taxonomy of data quality problems. In: Proceedings of the International Workshop on Data and Information Quality, 2005.
- Shankaranarayanan, The role of process metadata and data quality perceptions in decision making: an empirical framework and investigation. J. Info. Technol. Manage. v17 i1. 50-67, 2006.
- V.V. Subrahmanyam, M.N. Doja, Development Trends in the Field of Data Warehousing and OLAP, in the proceedings of Emerging Trends in Computer Science (ETCS-2007), MIET, Meerut, Pg Nos: 218-226, 2007.
- V.V. Subrahmanyam, M.N. Doja, A Survey of Conceptual Models for Data Warehouse Design, in the proceedings of International Conference on Data Management (ICDM 2008), Institute of Management Technology, Ghaziabad, UP. Pg Nos: 239 – 246, Mac Millan Advanced Research series, 2008.
- V.V. Subrahmanyam, M. N. Doja, An UML Based Approach for Designing the Conceptual Model of a Data Warehouse, in the proceedings of 3rd National Conference on Methods and Models in Computing (NCM2C-2008), Jawaharlal Nehru University (JNU), INDIA, Pg Nos: 3-11, published by Allied Publishers, 2009a.
- V.V. Subrahmanyam, M.N. Doja, Design Considerations for Building a Data Warehouse for an Open University System, in the proceedings of International Conference in Computing Technologies (ICONCT-09) Mepco Schlenk Engineering College, Sivakasi, Tamilnadu , INDIA, Pg.Nos: 401-406, 2009b.
- Watson, Data warehousing failures: case studies and findings, Data Warehousing. v4 i1. 44-55, 1999.
- Widom, Jennifer Wisdom, Research problems in data warehousing, Proceedings of the fourth international conference on Information and knowledge management, p.25-30, November 29-December 02, 1995, Baltimore, Maryland, United States, 1995.
- William E. Winkler, Methods for evaluating and creating data quality, Information Systems, v.29 n.7, p.531-550, October 2004.
- Reference website <http://www.home.nic.in>, 2010.

