Data Warehousing Methods and Techniques: A Survey of Literature to Identify Current Research Needs

Deepak Asrani
Department of Computer Science Engineering, Teerthanker Mahaveer University, Moradabad, U.P (India)

Renu Jain
Department of Computer Science Engineering, University Institute of Engineering & Technology, Kanpur, U.P (India)

Abstract- Data warehousing systems are backbone of decision making process for large business organizations. These systems provide strong platform for informed decision making in the current age of globalization and competitive environment. Effectiveness of data warehousing systems that are being developed is the result of continuous evolution and advancement in the technology. Though, immense research has been done in the past on this technology, still more research is required to further improve its quality and utilization. In this paper we have tried to present outcome of thorough reference of details of research efforts put up by various research groups in key areas of the technology like data warehouse modelling & design, ETL process designing, Business Intelligence (BI), data mining, data warehouse maintenance, and OLAP design. The objective of our study is to determine past sequence of advancements, current state of the art, and future course of research needs in this domain.

Keywords –Data warehousing, ETL, Decision making, OLAP, Data mining

I. INTRODUCTION

Data warehousing is a process of modelling, design and implementation of a data repository capable of serving complex business queries for varied decision making activities of the organization. Data warehousing systems are capable of serving all types of queries ranging from simple business reporting, online analytical processing (OLAP), or a complex data mining query. These systems store and maintain data for past several years which makes size of tables very large, thus results in slow query execution. Improvement in query execution time requires use of various techniques out of which materialized view maintenance is most effective and useful in which pre-computed aggregated information is stored in the data repository to save run time processing loads on the server and produce the response of query faster. Data warehousing systems get data from operational system of the organization and from some external data sources also. Design of both types of systems is different. While the operational systems are designed for operational efficiency of the organization's business process, the data warehousing systems are designed to support decision making activities of the organization. Differences in objectives of these systems, causes different design approaches. While operational system schema is designed to hold atomic data of all business transactions, data warehouse schema is designed to hold non atomic data at different levels of grains depending on organization's data warehousing needs. Collecting data from multiple sources requires cleaning for naming conflicts, data type mismatches, size differences, treatment to missing data values, removal of irrelevant data from the point of view of data analysis, and selection of relevant attributes for data analysis needs. These steps are key components of extraction transformation and loading (ETL) process. Figure 1 gives idea about data warehousing process and data mining application to extract knowledge useful in decision making of an organization.

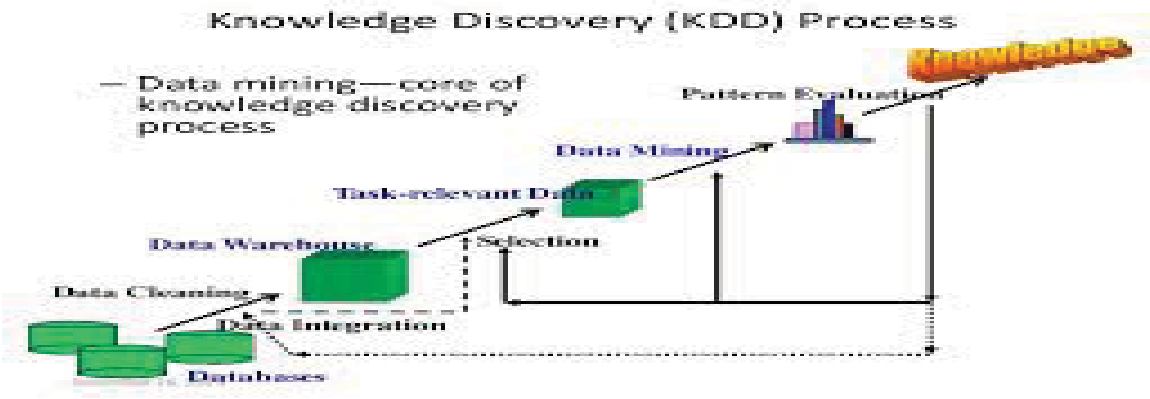

Fig-1 Data Warehousing and its Components

Quality of data warehousing systems depends on many aspects. In this paper we have tried to explore through research literature on multiple aspects of data warehousing to determine current state of the art, their effect, and further advancements expected by the industry. We have categorically covered research contributions in different areas of data warehousing.

Surveys and reports are means of creating awareness among user groups and researchers about various research initiatives that have been taken on different technical issues and help them in finding out the focus of research in a particular subject. It helps in determining the areas where sufficient research has not yet been done and there is much scope of further research.

Major Category	**Sub Category**
Modeling and Design	Conceptual, Logical, and Physical Design
	ETL Process Design
	Architecture Design
	Modeling and Design Standards
Data Warehouse Accessing Technologies	Query Processing
	OLAP Design Tools and Techniques
	Reporting Tools and Techniques
	Data Mining Operations and Techniques
Data Warehouse Maintenance	Materialized View Maintenance
	Meta Data Management
	Data Warehouse Security
	Index Maintenance

Table-1 Key categories and research areas in the field of data warehousing

Key research areas includes data warehouse security, data quality analysis, data visualization, ETL process, data distribution, business intelligence, web warehouses, data warehouse for xml data, stream data biomedical data and spatial data in different structured formats. Table 1 gives over view of research categorization in the field of data warehousing.

The organization of this paper is as follows: Section 2 covers analysis of research contributions in the field of modelling and design. Section 3 gives details of research efforts in the field of data warehouse accessing technologies. In section 4 details of research in data warehouse maintenance have been presented. Section 5 concludes the paper.

II. Data Warehouse Modelling and Design

Modeling and design of a data warehouse refers to analysis and diagrammatic representation of data storage structure, data transfer process from operational system, Infrastructure required for implementation and its architecture on the basis of requirement analysis of organization's data warehousing needs. Though, relational model is a standard for description and storage of operational systems structure, but it could also be used for modeling and design of data warehousing systems. However, the modeling and design of data warehousing system usually follows other standard of schema representation called Dimensional model. In Dimensional modeling data is

organized in the form of fact tables holding business measurements surrounded by various parameters of analysis referred as dimension tables. Data warehouse schema representation using this approach is called star schema. Sometimes, other extensions of star schema having normalized dimension's hierarchy or multiple fact tables are used, that are referred as Snow Flake schema and fact constellation schema respectively. Dimensional modeling is popular for Data warehousing because of its high readability and simplicity which is effective for presentation of the data warehouse design to business managers of the organization. Many research initiatives in the direction of modeling and design have been taken. Some of them have been presented here.

[1] Presented results of study on various Data warehouse designing approaches. Authors have emphasized upon the importance of analysis of operational system for data warehousing solution development. They also proposed a generalized framework for conceptual design of Data Warehouse based on object oriented approach using UML. Authors have presented Data Warehouse designing process in the form of a sequence of steps, their inputs and output. The steps in the proposed process are Analysis of operational system, requirement specification, conceptual design, logical design, and physical design.

[2] Proposed use of UML for design of Data warehouse giving four UML profiles namely multidimensional modeling profile, Data mapping profile, ETL profile, and Database Deployment profile.

[3] Proposed a graphical model for conceptual design of data warehouse and gave a semi automated methodology to build it. Author in [4] has given a generalized framework for data warehouse design.

The Framework for data warehouse design proposed by [1] is divided into two levels namely Requirements level and Design level. At requirements level, key component is integrator that integrates the collected requirements. Both levels have a metadata repository to maintain details. In the design layer UML class diagram is constructed that eventually helps in design star and snowflake schema.

[5] is a report on eighth ACM international workshop on Data ware housing and OLAP that has covered research contributions in the area of querying OLAP databases, data warehouse models, data warehouse design, and query processing and view maintenance.

[6] Applied software pattern techniques in developing divisional models. They have provided a meta model to help designers in specifying dimensional design pattern and apply them for designing dimensional models. The effectiveness of proposed technique were experimentally evaluated and was formed highly effective.

[7] Raised issue of representing spatial data which needs complex structures. The authors have proposed a spatial multidimensional model for representing complex spatial objects and multidimensional paradigm to handle geographical data.

[8] Points to higher failure rate of data warehousing projects and claims that requirement analysis is usually overloaded in real world data warehouse projects. Authors have proposed a goal oriented approach for requirement analysis. Organizational modeling and decisional modeling is done centered on stake holders and decision makes respectively.

[9] Points out the issue of non existence of a standard method for designing the whole data warehouse. Authors have presented a novel approach to align the whole data warehousing project called model driven Architecture and multi Dimension.

[10] Worked on assessment of Linux file system performance for data warehousing work loads. Experiments were conducted for different I/O modes, file fragmentations, and storage methods. The author found that anticipatory I/O scheduler is not good where direct I/O gives good performance as it bypasses buffer cache.

[11] Explored through features of data warehousing solutions and worked to find out the features that must be considered in a data warehousing system. Author has also elaborated in detail OLAP and its two approaches named ROLAP and MOLAP and have tried to set guidelines to select between these two in various contexts. Comparison of both has been done on various useful parameters. The study is supported by experiments. Author has emphasized upon benefits of the OLAP analysis for improving business process and draw utility from it. Several real life industry examples have been quoted as evidence. Author has defined usefulness of OLAP solutions in terms of perceived usefulness (PU) and perceived ease of use (PEU) of the implementation in terms of these two features. With the help of experiments the author has shown that ROLAP is highly effective in all aspects except visualization where MOLAP is a better option.

In [12] authors have done a comprehensive comparison of different data warehousing methodologies. As all data warehousing projects are not same and thus cannot be given same treatment. Different techniques and methodologies need to be applied for different projects. No experiments are conducted. Study has been conducted on different data warehousing products available in the market. The companies whose products have been considered for comparison has been categorized in three types namely core technology vendors (DBMS vendors), Infrastructure vendors, and data modelling companies. Authors have exclusively covered all steps of development

cycle, methodologies, tools, and products available to present a fair idea about all the components and to decide how and which option to choose from. Authors have also considered cases where organization's businesses change or expand over time and have tried to find out what changes are required in data warehousing implementation or maintenance has been discussed. The authors have found that very few vendors have incorporated change management in their methodology.

In [13] authors have focussed on the issues related to warehousing for specific business process of an organization. Problems have been identified and generic solution has been proposed supported by experiments. Authors have addressed three major challenges faced during data warehousing for business process data i. developing ad hoc, process-specific solutions for warehousing and reporting on process data is not a sustainable model ii. Abstracting process data is a second challenge iii. Business process automation and analysis/reporting software are co developed. Authors have also suggested solutions for these issues. Authors have suggested that prototyping could be applied in data warehousing to get an early feedback of analysts/managers and saves huge investments. Specific contributions of the authors could be summarized as follows: Authors have analyzed and classified analysis requirements for process data warehousing. They have provided a configurable warehouse model that can satisfy complex reporting needs for virtually any process, also taking into account performance constraints. They have shown that how to abstract from low-level data about executed processes to higher-level views of the same process, suitable for reporting purposes. They have described how to ETL process data, and in particular how to semi-automatically maintain ETL procedures in the wake of changes in the source applications. Then they have shown how the solution can be quickly prototyped using an emulation environment to get early feedback from users. On the basis of study and experimentations the authors have found that all of the reporting requirements can be classified along the following categories of metrics: *Process metrics*, *Resource metrics, and Business data metrics.*

In [14] authors have integrated two technologies named data warehousing and GIS. They have proposed a logical multi dimensional model which is suitable for modeling spatial data warehouse implementation based on object relational data base management system. They have taken implementation of data warehousing solution for agriculture crops and have shown execution of multi dimensional queries over spatial data warehouse and evaluated their performance. The authors have proposed a logical spatial query optimizer and developed a prototype named *Map- Warehouse*, with the objective to validate the proposed ideas.

In [15] authors have worked towards data warehouse design on the basis of analyzing and developing taxonomy of the domain. They have taken Customer Relationship Management as their domain of study. Authors have mentioned the role of CRM taxomomy in designing data warehouse and evaluating data schema for CRM analysis. In CRM many aspects are analyzed like costomer analysis, profitability analysis, sales and marketing analysis, product analysis, They have shown that Effective CRM analyses require a detailed data warehouse model that can support various CRM analyses and deep understanding on CRM-related business questions. They have presented a taxonomy of CRM analysis categories which includes CRM strategies, CRM category analyses, CRM business questions, their potential uses, and key performance indicators (KPIs) for those analysis types. The approach proposed by the authors could be used for other domains also where the mentioned questions will play the same role in designing data warehouse. Authors have identified and mentioned that there are no agreed upon standardized rules for how to design a data warehouse to support CRM and how to effectively use CRM technologies. Thus, the ultimate long-term purpose of their study was to systematically examine CRM factors that affect design decisions for CRM data warehouses and to build a taxonomy of CRM.

[16] have presented an approach to automate the multidimensional design of Data Warehouses based on semi-automated method that tried to find business-related multidimensional concepts from a domain ontology representing different data sources of a given business domain.

[17] Presented a method which semi-automatically generates star schemas from an Entity Relationship Diagram, by analyzing both the semantics and the structure of the ERD. The novel features of the method named SAMSTAR were (a) the use of the notion of Connection Topology Value (CTV) for identifying fact and dimension candidates and (b) the use of Annotated Dimensional Design Patterns (A DDPs) as well as WordNet to extend the list of dimensions.

In [18] four organizations that have implemented data warehousing solution and were using them efficiently were studied and based on their experiences ten success factors of data warehousing projects were identified and have been presented in the form of data warehousing commandments. Major features included business-driven data warehousing initiative, Executive sponsorship and commitment, Funding commitment, Project team with access to cross-functional project management and implementation experience, Attention to source data quality, flexible enterprise data model, Data stewardship, long term plan for automated data extraction methods / tools, Knowledge of DW compatibility with existing systems, and Hardware/software proof of concept.

In [19] with the help of a very basic example considered by the authors for simplicity, they have explained the significance of implementing a data warehousing system separated from an operational system because the need and expectations from both the systems is different and a single schema cannot serve both type of requirements. Analytical operations can not run efficiently on an operational system. Subsequently, authors have shown possible design options for modeling data warehousing solutions and their pros and cons over one another. Enterprise data warehouse (EDW), data marts (DM), for each department, common ETL, or separate ETL.
In [20] authors have taken a case study of health care application where RFID tags were used to identify the objects and resulted in high volume of data. Data warehousing implementation for such a high volume data was very complex. In this paper, we show how to apply traditional Data Warehousing techniques to model these massive amounts of RFID Data. In short, we describe how to construct an RFID Warehouse so that important query analyses can be performed very efficiently. We also show how to process a continuous stream of RFID Data to answer real time queries using Sliding Window techniques. By the help of using synthetic Datasets, we conclude that querying over Data Warehouse is much faster than traditional Relational DBMS. We also find that the aforesaid performance improvement is expected to be much higher as the size of the Dataset increases. In this Paper, we have shown how to apply two well-recognized techniques - Data Warehousing and Sliding Window Protocol for answering both statistical and real-time queries using RFID in an Assistive environment.

III. Data Warehouse Accessing Technologies

Data Warehouse Accessing tools & techniques are used for data retrieval from data repository for decision making. Data Warehouse accessing is normally divided into i. Standard data reporting ii. On Line Analytical processing (OLAP), to dimensionally analyze business data iii. Data mining operations to extract knowledge
OLAP key concerns are efficient slicing, dicing, drill-down, rollup and pivoting. Quality of OLAP data cube depends on inclusion of required dimensions in the OLAP design, useful pre-computed aggregations for efficient drill down & roll up operations, availability of a standard application interface for easy choice selection by the business users and navigation from one level to another.
In data mining solutions, facility to perform all standard data mining operations, be it the classification clustering, associations, discrimination or other important patterns generation from the existing business data should be available.

[21] Worked on issues of representing OLAP databases and their query language. The author has designed a framework based on functional symbols and query constraints in the form of polymorphic functions.
[22] Raised issue of poor visualization of results of data warehouse accessing by users. The authors have found that most interesting information is generally not shown correctly due to poor visualization capabilities or implementation. To solve such issues the author has proposed a framework in which user is able to specify its information preferences and visualization constraints based on device used for display of results.
[23] Presented a unique concept of querying data warehouse schema as well as related documents to obtain comprehensive details of information required. For this, the authors have proposed the R-cube that specifies relevance of each fact in a query and defines the related documents that provide information on selected facts. Thus, the users can query a data warehouse and obtain further information from related documents.
[24] Raised issue of traditional business intelligence architecture not supporting real time BI needs. The author has addressed this problem by proposing a BI architecture named SARESA, with the aim to provide real time analytics that enables proactive responses to business situations for effective business management. SARESA the author has proposed the sense and response loops and service oriented architecture that is able to detect different business situations and exceptions, perform complex analytical tasks and shows gap between current situation and desired management goals.
[25] Worked on issues of approximate answering of query for OLAP applications. The author claims scalability of query techniques and the accuracy of answers are limitation of state of the art protocols in OLAP-Author has worked on limitations related to accuracy of queries by giving a statistical framework.
Hadoop is and open source implementation of MapReduce Framework for performing parallel analysis and data mining. In [26] authors have discussed performance-oriented query execution strategies for data warehouse queries considering join and aggregation operations. Authors have compared their experimental results with a standard commercial parallel database and an open-source MapReduce implementation featuring a SQL interface (Hive). They have shown that HadoopDB successfully competes with other systems.
In [27] authors have developed a data warehousing engine to deal with peta bytes of data from internet applications to provide deep insight from such a big volume of data. The engine was developed at yahoo. The engine is based on

massive parallel processing architecture and column architecture. The proposed architecture is highly scalable, flexible, specialized for analytic operations and lower administrative and hardware costs. The authors have presented both hardware and software architecture of data warehouse engine.

In [28] Authors have presented report on ACM international workshop on data warehousing and OLAP. Authors have categorized data warehousing as warehouse design and OLAP. Key research concerns identified are in the area of design, management and uses of data warehousing solutions. Authors have emphasized upon the need of software and tools to improve data management and analysis on large amount of data and information accumulated in corporate databases. Paper categorization has been done in four types named as design, physical data organization, data warehouse processing, and spatiotemporal data warehouses & data mining.

[29] Worked on data cube compression and query approximation. They presented a probabilistic model for data cube compression and query approximation addressing the problem of automatically analyzing large multidimensional tables to get a concise and compact representation of data, identify patterns and provide approximate answers to queries. They analyzed the potential of a probabilistic modelling technique, called non-negative multi-way array factorization, for approximating aggregate and multidimensional values. The proposed technique compared favorably to the log-linear modelling cube compression technique.

In [30] authors have worked on a data warehouse environment that could offer near real time data warehousing and fast query execution based on the concept of Workload Balancing by Election (WINE), which allows users to express their individual demands on the Quality of Service and the Quality of Data, respectively. WINE applied this information in order to balance and prioritize over both queries and update transactions according to user needs. A simulation study showed that the proposed algorithm outperforms competitor baseline algorithms over the entire spectrum of workloads and user requirements.

[31] Pointed out limitations of relational database technology for data warehousing applications and have discussed many alternate technologies suitable for data warehousing and have given their comparison. Authors have identified several research issues and have given their usefulness and limitations. They have compared database and Map Reduce from the point of view of data warehouse implementation. Map Reduce is open source and data mining algorithm could be implemented easily. Database offers data mining and OLAP features available but their customization is difficult. Authors have compared pros and cons of each technology for data warehousing and have identified research issues, considering practical aspects like ease of use, programming flexibility and cost as well as technical aspects like data modelling, storage, hardware, scalability, query processing, fault tolerance and data mining.

IV. Data Warehouse Maintenance

Data Warehouse maintenance covers execution of routine Extraction Transformation and Loading (ETL) operations as per schedule, security management, index management, OLAP structure management, index management, materialized view maintenance, usage analysis and corresponding changes in the Data Warehouse repository like dynamic aggregate view (materialize view) maintenance.

[32] Worked on the problem of indexing and pointed out that with the existence of attribute hierarchies resulting in multi granularity. The authors have presented a hierarchy and caching framework that supports manipulation of hierarchies in a parallel ROLAP environment. The authors could experimentally demonstrate that little overhead is required to handle arbitrarily hierarchical queries.

[33] Pointed out the problem of multiple view maintenance. The author has proposed method of efficient incremental maintenance plain for the given view. The author has experimentally shown the efficiency of the proposed method is quiet high.

[34] Presented details of major achievement in the conceptual DW modeling and pointed out problem that still need research attention, one of them being DW schema versioning issues.

In [35] authors have worked on integration of relational database, object oriented database and XML database and have given a framework of integration. The work of authors is based on experiments where they have demonstrated ETL process in detail, maintenance of data warehouse with the help of AutoMed meta data which is a system for expressing data transformation and integration processes in heterogeneous database environments. Authors have discussed the use of AutoMed metadata in data warehousing environments and have shown how AutoMed metadata can be used to express the data schemas, and the data cleansing, transformation, and integration processes. They have also shown how this metadata can then be used for populating the data warehouse, incrementally maintaining the warehouse data after data source updates, and tracing the lineage of warehouse data.

[36] Compared five probabilistic techniques for aggregate view size estimation including Stochastic Probabilistic Counting and LOGLOG Probabilistic Counting. The experiments showed that only Generalized Counting, Gibbons-Tirthapura, and Adaptive Counting provide universally tight estimates irrespective of the size of the view; of those, only Adaptive Counting remains constantly fast as the memory budget is increased.
[37] Worked on compression of bitmap indexes in data warehouses using the compression technique, called Run-Length Huffman (RLH), which was based on both run-length encoding and Huffman encoding. RLH was implemented and experimentally compared to the Word Aligned Hybrid (WAH) bitmap compression technique. The authors have experimentally shown that RLH offers shorter query response times than WAH for certain cardinalities of indexed attributes. Moreover, bitmaps compressed with RLH are smaller than bitmaps compressed with WAH.
[38] Presented techniques for deciding the physical implementation of ETL workflows. They dealt with the problem of determining the best possible physical implementation of an ETL workflow.
[39] Considered the problem of efficient computation of view subsets. They argued that given the enormous size of the fact table in a star schema, virtually all current systems augment the primary fact table with a small number of focused summary tables. They have presented suite of greedy algorithms for the construction of such view subsets. Experimental results demonstrated cost savings of between 20 and 70% relative to the naive alternative algorithms, depending upon the degree of materialization required.
[40] Proposes a novel way of automatically developing data warehouse configuration in rule-based CRM systems. Selecting views for materialization in a data warehouse is one of the important decision-making tasks in its design. Given a set of campaign rules expressing marketing strategies, the proposed method generates data warehouse configuration (including database schema and indexing constraint) that can satisfy all the input rules. Our method begins on the premise that data warehouse configuration can be reversibly extracted from marketing campaign rules. This method includes algorithms for database schema generation, indexing constraint generation, schema normalization for removing data redundancies, and OLAP (On-Line Analytic Processing) query generation. we should be able to produce data warehouse configuration satisfying those rules. Data warehouse configuration consists of data warehouse schema, indexing constraint, and predefined analysis (or OLAP1) queries. Once data warehouse schema is generated, indexing constraint and analysis queries can be easily extracted. Authors have designed a system for automatic data warehousing configuration on the basis of the given CRM rule. The designed system would be able to design schema, indexing constraints, and analysis OLAP queries.

IV.CONCLUSION

Data warehousing is a widely accepted technology for decision making applications for large business organizations. Insight view of the organization's performance in all aspects of business is a crucial requirement in order to enable the organization to identify those aspects of their business where improvements are required for further growth of the organization. Data warehousing applications provide these features to business users of the organizations. To have highly efficient data warehousing solutions, many aspects of this technology needed research initiatives that have been put up by the research community. Total forty research papers have been referred. Major contributions are seen in the area of data warehouse modeling & design, data warehouse accessing technologies and data warehouse maintenance. From the review of literature it is clear that in the beginning major focus was on modeling and design and by now this aspect of data warehousing has matured enough. Though, the work in data accessing techniques has also been done but with the ever increasing volume of data, we need to work further in the direction of data accessing and visualization. Concepts of data maintenance have also been explored by the researchers in the past but with the evolution of unstructured data in huge amounts, newer techniques need to be explored and devised for managing such data. Technologies like Big data, Map reduce need to be taken into main stream of data warehousing as they support data reduction and parallel processing feature that gives both way improvement in the processing. Work towards standardization of data warehouse solution development methodology is required. Adaptation of data warehousing technology has been done only by the large sized business enterprises due to high costs and proprietary methods of implementation. A standard framework needs to be developed so that organizations of small to mid size can also easily embrace this technology.

REFERENCES

[1] Jindal, A., & Taneja, S. (2012) "Comparative Study of Data Warehouse Design Approaches: A Survey", International Journal of Database Management Systems (DBMS) Vol. 4, No. 1, February 2012 p33-44.

[2] Juan Trujilio, E.Medina and S.Lujan Mora (2002) ,"A Web Oriented Approach to manage Multidimensional Models through XML Schemas and XSLT '' , EDBT 2002 Workshops, LNCS 2490, pp. 29–44, 2002. Springer-Verlag Berlin Heidelberg.

[3] Stefano Rizzi, Matteo Golfarelli. (1998) "A Methodological Framework for Data Warehouse Design", DOLAP 98 Washington DC USA.Copyright ACM 1999 l-581 13-120-8/98/l 1...$5.00.

[4] Stefano Rizzi, Matteo Golfarelli,D.Maio (1998) "The Dimensional Fact Model:A Conceptual Model for Data Warehouses." International Journal of Cooperative Information Systems (IJC IS),7(2- 3):215-247.

[5] Trujillo, J., & Song, Y., "A Report on the Eighth ACM International Workshop on Data Warehousing and OLAP (DOLAP'05)" SIGMOD Record, Vol. 35, No. 4, December 2006 p48-50.

[6] M.E. Jones and I-Y. Song, "Dimensional Modeling: Identifying, Classifying, and Applying Patterns ", Eighth ACM International Workshop on Data Warehousing and OLAP, ISBN: 1-59593-162-7, ACM Press, NY, 2005.

[7] S. Bimonte *et al.* "Towards a Spatial Multidimensional Model", Eighth ACM International Workshop on Data Warehousing and OLAP, ISBN: 1-59593-162-7, ACM Press, NY, 2005.

[8] P. Giorgini *et al.,* "Goal-Oriented Requirement Analysis for Data Warehouse Design", Eighth ACM International Workshop on Data Warehousing and OLAP, ISBN: 1-59593-162-7, ACM Press, NY, 2005.

[9] J-N. Mazón *et al.,* "Applying MDA to the Development of Data Warehouses", Eighth ACM International Workshop on Data Warehousing and OLAP, ISBN: 1-59593-162-7, ACM Press, NY, 2005.

[10] Wong, P. W., Hendrickson, Y. R., Rizvi, H., & Pratt, S., "Performance Evaluation of Linux File Systems for Data Warehousing Workloads INFOSCALE '06", Proceedings of the First International Conference on Scalable Information Systems, May 29-June 1 2006, Hong Kong © 2006 ACM 1-59593-428-06/05...$5.00 P43-50.

[11] Gorla, N., "Features to Consider in a Data Warehousing System", COMMUNICATIONS OF THE ACM November 2003/Vol. 46, No. 11 P111-115.

[12] Sen, A., & Sinha, A. P., " A Comparison of Data Warehousing Methodologies", COMMUNICATIONS OF THE ACM March 2005/Vol. 48, No. 3 P79-84.

[13] Casati, F., Castellanos M. , Dayal, U. , & Salazar, N., "A Generic solution for Warehousing Business Process Data", September 23-28, 2007, Vienna, Austria. Copyright 2007 VLDB Endowment, ACM 978-1-59593-649-3/07/09. P1128-1137.

[14]]Sampaio, M. C., Desousa, A. G., Desouza, C. B., " Towards a Logical Multidimensional Model for Spatial Data Warehousing and OLAP *DOLAP'06"*, November 10, 2006, Arlington, Virginia,USA.Copyright 2006 ACM 1-59593-530-4/06/0011...$5.00. P83-90.

[15] Cunningham, C., & Song, Y., "A Taxonomy of Customer Relationship Management Analyses for Data Warehousing 2007", Australian Computer Society, Inc. Twenty-Sixth *International Conference on Conceptual Modeling ER 2007 Tutorials, Posters, Panels and Industrial Contributions, Auckland, New Zealand*. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 83. John Grundy, Sven Hartmann, Alberto H. F. Laender, Leszek Maciaszek and John F. Roddick,Eds. Reproduction for academic, not-for profit purposes permitted provided this text is included. P97-102.

[16] O. Romero and A. Abello., "Automating multidimensional design from ontologies, "DOLAP 2007, ACM 10th International Workshop on Data Warehousing and OLAP", ISBN 978-1-59593-827-5, ACM Press, 2007., pp. 1–8, 2007.

[17] I.-Y. Song, R. Khare, and B. Dai, "SAMSTAR: a semi-automated lexical method for generating star schemas from an entity-relationship diagram", "DOLAP 2007, ACM 10th International Workshop on Data Warehousing and OLAP", ISBN 978-1-59593-827-5, ACM Press, 2007., pp. 9–16, 2007.

[18] Sammon, D., & Finnegan, P., "The Ten Commandments of Data Warehousing The DATA BASE for Advances in Information Systems" - Fall 2000 (Vol. 31, No. 4) p82-91.

[19] Jukic, N., "Modeling Strategies and Alternatives for Data Warehousing Projects Communication of the ACM", April 2006/Vol. 49, No. 4 p-83-90.

[20] Ferdous, S., Fegaras, L., Makedon, F., Applying Data Warehousing Technique in Pervasive Assistive Environment *PETRA'10*, June 23–25, 2010, Samos, Greece.Copyright © 2010 ACM ISBN 978-1-4503-0071-1/10/06... $10.00 P33-39.

[21] K. Q. Pu "Modeling, Querying and Reasoning about OLAP Databases: a Functional Approach", Eighth ACM International Workshop on Data Warehousing and OLAP, ISBN: 1-59593-162-7, ACM Press, NY, 2005.

[22] L. Bellatrach *et al.* "A Personalization Framework for OLAP Queries", Eighth ACM International Workshop on Data Warehousing and OLAP, ISBN: 1-59593-162-7, ACM Press, NY, 2005.

[23] J. M. Pérez *et al.* "A Relevance-Extended Multidimensional Model for a Data Warehouse Contextualized with Documents", Eighth ACM International Workshop on Data Warehousing and OLAP, ISBN: 1-59593-162-7, ACM Press, NY, 2005.

[24] A. Simitsis. "Mapping Conceptual to Logical Models for ETL Processes ", Eighth ACM International Workshop on Data Warehousing and OLAP", ISBN: 1-59593-162-7, ACM Press, NY, 2005.

[25] A. Cuzzocrea. "Providing Probabilistically-Bounded Approximate Answers to Non-Holistic Aggregate Range Queries in OLAP", Eighth ACM International Workshop on Data Warehousing and OLAP, ISBN: 1-59593-162-7, ACM Press, NY, 2005.

[26] Pawlikowski, K. B., Abadi, D. J., Silberschatz, A., & Paulson, E., "Efficient Processing of Data Warehousing Queries in a Split Execution Environment", P1165-1176 SIGMOD'11, June 12–16, 2011, Athens, Greece. Copyright 2011 ACM 978-1-4503-0661-4/11/06 ...$10.00.

[27] Ahuja, M., Chen, C. C., Gottapu, R., Hallmann, J., Hasan W., Johnson, R., Kozyrczak, M., Pabbati R. , Pandit, N., Pokuri, S., & Uppala, K.,." Peta-Scale Data Warehousing at Yahoo! *SIGMOD'09"*, June 29–July 2, 2009, Providence, Rhode Island, USA. Copyright 2009 ACM 978-1-60558-551-2/09/06...$5.00. P855-861.

[28] Pedersen, T.B, & Song, Y., "A Report on the Tenth ACM International Workshop on Data Warehousing and OLAP (DOLAP'07)", SIGMOD Record, Vol. 37, No.1, March 2008.

[29] R. Missaoui, C. Goutte, A. K. Choupo, and A. Boujenoui. "A probabilistic model for data cube compression and query approximation, DOLAP 2007", ACM 10th International Workshop on Data Warehousing and OLAP, ISBN 978-1-59593-827-5, ACM Press, 2007., pp. 33–40, 2007.

[30] M. Thiele, U. Fischer, and W. Lehner., "Partition-based workload scheduling in living data warehouse environments. DOLAP 2007", ACM 10th International Workshop on Data Warehousing and OLAP, ISBN 978-1-59593-827-5, ACM Press, 2007., pp. 57–64, 2007.

[31] Ordonez, C., Song, Y., Carlos, G. A., "Relational versus Non-Relational Database Systems for Data Warehousing *DOLAP'10",* October 30, 2010, Toronto, Ontario, Canada. ACM 978-1-4503-0383-5/10/10. P67-68.

[32] F. Dehne *et al*. "Parallel Querying of ROLAP Cubes in the Presence of Hierarchies", Eighth ACM International Workshop on Data Warehousing and OLAP, ISBN: 1-59593-162-7, ACM Press, NY, 2005.

[33] K.Y. Lee and M.H. Kim. "Optimizing the Incremental Maintenance of Multiple Join Views", Eighth ACM International Workshop on Data Warehousing and OLAP, ISBN: 1-59593-162-7, ACM Press, NY, 2005.

[34] Jens Lechtenborger. "Issues in Data Warehouse Modeling", Eighth ACM International Workshop on Data Warehousing and OLAP, ISBN: 1-59593-162-7, ACM Press, NY, 2005.

[35] Fan, H., & Poulovassilis, A., "Using AutoMed Metadata in Data Warehousing Environments *DOLAP'03",* November 7, 2003, New Orleans, Louisiana, USA. Copyright 2003 ACM 1-58113-727-3/03/0011 ...$5.00. P86-93.

[36] K. Aouiche and D. Lemire., "A comparison of five probabilistic view-size estimation techniques in OLAP. DOLAP 2007", ACM 10th International Workshop on Data Warehousing and OLAP, ISBN 978-1-59593-827-5, ACM Press, 2007., pp. 17–24, 2007.

[37] M. Stabno and R. Wrembel. "RLH: bitmap compression technique based on run-length and huffman encoding. DOLAP 2007", ACM 10th International Workshop on Data Warehousing and OLAP, ISBN 978-1-59593-827-5, ACM Press, 2007., pp. 41–48, 2007.

[38] V. Tziovara, P. Vassiliadis, and A. Simitsis., "Deciding the physical implementation of ETL workflows. DOLAP 2007", ACM 10th International Workshop on Data Warehousing and OLAP, ISBN 978-1-59593-827-5, ACM Press, 2007., pp. 49–56, 2007.

[39] F. K. H. A. Dehne, T. Eavis, and A. Rau-Chaplin., " Efficient computation of view subsets. DOLAP 2007, ACM 10th International Workshop on Data Warehousing and OLAP, ISBN 978-1-59593-827-5, ACM Press, 2007., pp. 65–72, 2007.

[40] Kim, H. J., Lee, T. H., Lee, S. G., Chun, J. G., "Automated Data Warehousing for Rule-based CRM Systems", 2003, Australian Computer Society, Inc. Fourteenth Australasian Database Conference(ADC2003), Adelaide, Australia. Conferences in Research and Practice in Information Technology, Vol. 17. Xiaofang Zhou and Klaus-Dieter Schewe, Ed.