International Journal of u- and e- Service, Science and Technology
Vol.8, No.5 (2015), pp.299-308
http://dx.doi.org/10.14257/ijunesst.2015.8.5.27

Taxonomy of Different Data Warehouse Architectures & Need for Optimized Data Warehouse

[1,2]Muhammad Arif and [1]Ghulam Mujtaba

[1]*Faculty of Computer Science and Information Technology, University of Malaya 50603 Kuala Lumpur, Malaysia*
[2]*Computer Science Department, Comsats Institute of Information and Technology Islamabad Pakistan*

Abstract

Different architectures have been proposed for Data Warehouse. This paper gives a survey of the various architecture techniques used and justifies why there is still a need for optimized data warehouse architecture which overcomes the limitations or challenges of the existing techniques. Survey includes OLAM architecture (combination of Data mining and OLAP), complex Data Warehouse architecture, Real Time storage Architecture, Data Ware House Architecture through view point of information system and Model Driven approach for Data Warehouse. Ultimate goal of this paper is to propose an architecture which is fault tolerant, has less resource consumption and user friendly.

Keywords*: Information system, Model driven, Dataware House Architecture, Real time Storage, OLAM, OLAP.*

1. Introduction

In the modern era there is a high volume of data that need to be stored in a reliable way so that it can help the organization for timely access of critical data and decision making. A lot of techniques have been proposed with respect to data storage and its dependency. One of them is OLAP (Online Analytical Processing), it supports tactical and ROI (Return on Investment) applications. OLAP helps in:

Policy making: governing the goals related with data ware house.

Transformation: transforming raw data taken from the different resources to the accepted format of data ware house.

Metadata: It is the reusable information stored in a central location from where it provides the usage of common conventions.

Storage: Data is stored in a way that maximizes its availability and could be accessible most of the time.

Analysis: It analyzes the direction of the organization business by supporting end user queries, planning the data view point.

OLAP provides the fast and consistent query response, it stores gigabytes of data that is summarized and calculated after several analytics. The data structure of OLAP server is dimensional and hierarchical. It provides many user benefits that include insulation from SQL (Structured Query Language) and relational model and provides improved query performance. OLAP is easily deployable with small expense.

Along with OLAP, two more architecture techniques are hereby discussed, which are the pre-requisites for writing this paper. First one is OLAM (Online Analytical Mining) which is used for solving the two weaknesses of traditional data mining (DM), combining

ISSN: 2005-4246 IJUNESST

data warehouse technology, OLAP technology and DM technology, we put forward an OLAM system model, and this model has been researched and applied. In this architecture, the OLAP and the DM combines and user can participate to the whole analysis during the DM process, so data can be observed by different angles, the different levels and thus a more reasonable pattern can be generated to provide more supports for the user.

This approach combines the concrete realization of general data mining system and proposes the model of data mining system which is based on OLAM. This integrated architecture of the structure is divided into four layers, respectively are data storage, OLAP/OLAM data mining component, multi-dimensional database (multidimensional data view), user Graphical User Interface.

The other technique is real time storage architecture which is used for utilizing organizational data by transmitted from the operational and transactional environment to a dimensional or normal database. Variety of platforms is an issue regarding the structure, but also variety of data types such as non-structural and text format files should be considered. In this way, for raising customer satisfaction and quick and accurate services exploiting updated data storages is necessary. Although the need for real time data storages is increasing day by day in commercial systems and it has been emphasized in many researches, whereas lack of updated data is still considered as their major weakness. The architecture focuses on two issues, (i) Extract–Transformation–Load operations (ETL) and (ii) Combination of different data sources.

For real-time data storage, routinely and continuous data updating is needed and once when changing from source systems. This case is opposite to ETL operations central core in classic data storages architecture. So, we intend ETL operations accomplished in push form originated from source systems once stabilized change be made. Hence, we use a technology which is called triggers that causes the data integration in source systems. In other words, data storage source created is an operational data store (ODS). Data has been integrated in operational stored (ODS) data source. When we have edit, Insert and delete operations, the corresponding triggers in source systems transfer changed data to operational data storage source .XML format is used in the approach in which operational data along with triggers is being stored in the tables of the storages resources. XML form data is continuously processed by converter and after that the data is added into Fact and Dimension tables placed in dimensional data storage resources. Operational data store (ODS) can be removed for reduction of layers of data between producer and consumer but we lose benefits of operational data store resources. Three different formats of stored data in three locations can be accessed by user application.

- Access to the operational data store (ODS) which has current data in normal form.
- Access to dimensional data store source (DDS) for analysis and complex reporting.
- Access to advanced analytical tools along with compatibility of multi-dimensional databases (MDB) data format.

2. Literature Review

A. Research and Application of OLAM Architecture

OLAM [1] is a technique which combines features of both OLAP and DATA MINING which enables decision-making people and analysis people to observe the data from different angles and levels. OLAM features include processing a great deal of computing and analysis comparison. It is a multi-dimensional model which deeply excavates data. It is subject oriented technology which forms new level of knowledge. With the www technology combination, OLAM can be helpful for network information resources departments for management of huge data in complex form.

This integrated architecture divides the entire structure into four layers, respectively are data storage, OLAP/OLAM data mining component, multi-dimensional database (multidimensional data view), user GUI (Graphical User Interface). GUI layer is used as a way of communication with the system i.e. OLAM engine and OLAP engine. Selection the mining algorithms, as well as reception of feedback information from OLAM engine and OLAP engine is also performed on the same layer. Multi-dimensional Data View is the combination of two different data organization. One type is multidimensional database and other kind is relational database. Data mining component mainly comprises of the four parts *OLAM engine*, *OLAP engine*, *DM engine* and *GUI*. Data mining component is mainly responsible for mining of data and gives the feedback in shape of file to user. The OLAM analysis mining is more efficient then OLAP, because more data mining tasks can be performed by it which includes , prediction, time series analysis, clustering, concept description, classification, and association mining. This approach can be applied to different areas where we want to excavate info and analyze production data from different dimensions, different levels and in this way some new rules can be originated from that data which can be very useful.

B. An Architecture Framework for Complex Data Warehouse

As many Decision Support System (DSS) applications need to process data that is not in numerical or symbolic form, it can be multisource, multimedia, multimodal, multi structure. Management and analyses of complex data involves a lot of different issues regarding their storage, structure and processing, and key element in that processes is metadata. For storing complex data XML is used.

Two architecture approaches for data warehouse are being discussed in the paper. One is main family of architectures is data driven band based on a classical, centralized data ware house where data are the main focus. The second architectures provide solutions based on virtual warehouse, which are process driven and where metadata is the key element. Mediator wrapper approaches are used for these solutions for exploiting distributed data sources. Integration of heterogeneous data from various sources can be facilitated in warehouse with the aid of XML, This framework consists of the data warehouse kernel, source type, data sources, and a metadata and knowledge base layer that includes three sub modules related to three management processes.

Processes used for managing a data warehouse are the ETL and integration process which is used for giving the input to the warehouse from different data sources, the administration and monitoring process which is used for management of knowledge and metadata, the analysis and usage process which is used to run user queries, produces reports, supports OLAP, builds data cubes *etc*. These processes are exploiting and updating the metadata and knowledge base.

According to author that technique can be very useful in industries like medicine, banking and geography.

C. *Model Driven Architecture Approach for Data Warehouse*

For efficient decision making, entrepreneur requires more information from information Technology areas. Model Driven Architecture (MDA) and Data Warehousing are the solutions for that problem. For the creation of scripts for automatic generation and loads DW based databases on the OLTP (Online Transaction Processing) MDA cartridge is used by a MDA tool. So whenever change occurs in model, it automatically updates the scripts without involvement of manual work.

The MDA creates a separation between logical parts of the application from implementation technology part. The point of view by MDA is independent computation, specific platform and independent platform. UML is adapted as a modeling pattern in the architecture. The extraction and loading is carried out by any of the two choices i.e.

through programs written in procedural languages or through particular software. The metadata standard CWM is enables the integration of intelligent e-business, data warehouse systems and business systems in heterogeneous environments. According to the AndroMDA applications architecture, three coding layers are used i.e. presentation, business and data access.

Oracle database is chosen as a database and the scripting language is Oracle PL/SQL. First step is creation of DW database through SQL scripting commands allowed by cartridge. The tables to be generated are defined by Fact and Dimension stereotypes. The second step is the generation of a script for loading data in the DW. This script extracts and record information in the particular tables of the data warehouse without manual working.

D. Data Warehousing and OLAP for Decision Support

OLAP and Data warehousing has gained the great support with respect to decision support from different vendors and marketing technologies. Data is increasing day by day and it is important for vendors to progress and it is possible only if they take decisions on data, they have to analyze like where the selling growth is high and where there is downfall. On the other way around OLTP applications have repeated functionalities they only care about transaction process.

Data warehouse gather terabytes of data from different sources and made them available to decision maker for analysis purposes. ROLAP (Relational OLAP) is a type of OLAP with relational database management system, similarly MOLAP servers may be used that support multidimensional data. Different techniques are there that can be implemented with Data warehouse with respect to the data to perform efficiently and help to improve decision support environment.

E. Comparison of data warehouse development methodologies case study of the process Warehouse

Data Warehouse helps the organizations to redesign the structure that needs improvement. It always provides its best output that will be shown afterwards in the decision policy. Data Warehouse is data driven, requirements are considered at end when data warehouse is populated with data and effects of the data are coming into play. Kimball proposes a four steps approach; Westerman has its main focus on implementing business strategy [2]. The process warehouse is an analytical database that aims to improve business process, it has five perspectives in order to represent fundamental process concept.

The Wal-Mart methodology is being tested on the prototype and it provides the significance to decrease their process cycle. The goal-driven development methodology is a foundation for decision support for all organizations [2]. For data mining and data exploration data-driven development methodology is being considered

Table 1. Explain Author Name, Type, Tool and Challenged Handled by Data Warehouse Architecture

Reference	Author name	Published date	conference / journal paper	DWH Type	Tool name	Challenge Handled
[1]	Fushan Wang	24-05-10	Conference paper	OLAM	SQL Server 2000/ OLAP & OLAM algo	To achieve high performance in OLAP and to enhance Data mining efficiency simultaneously

[2]	Jerome Darmount, Omar Boussaid	04-08-08	Journal Paper IJWET	Complex DWH	- -	The need for an DWH architecture to support Complex data.
[3]	Maria M Dias , Andre Luis A Menolli	12-12-08	Conference Paper (CIMCA)	Info System Architecture for DWH	Oracle 9i OLAP Tool	To transform Large volume of data into useful Information.
[4]	Omid Yektamanesh , Jafar Habibi, Hamed Ahmedi	2010	Conference Paper (CIMCSim)	Real Time Architecture for DWH	ArchStudio	Performing realtime data storage in DWH
[5]	Lucia Abrunhosa, Beatriz Helena, Jano Mereira	2010	Conference Paper (ICAS)	Model Driven Architecture for DWH	AndroMDA	Creation of automatic scripts for loading data in DWH
[6]	Surajit Chaudary, Umeshwar Dayal	1997	ACM SIGMOD	Enterprise DWH	---	Decision support
[7]	Robert Winter	2001	Conference Paper	Enterprise DWH	---	Possible extension of data warehousing approach to application integration.
[8]	Mohammad Rifaie, Keivan Kianmehr, Rida Alhajj, Mick J. Ridley	2008	IEEE	Enterprise DWH	Enterprise standard tools	Data Modeling and principles to be considered in modeling DWH
[9]	Sabina Redzanovic, Panagiotis Chountas, Thierry Chaussalet, Farid Fouladinejad, Mohamed Tadjer	2011	24th International Symposium	Complex DWH	ePACT	Integrating prescribing data from different GP practices, equipped with adhoc OLAP utility.
[10]	Joshua Nealon, Wenny Rahayu, Eric Pardede	2009	Conference Paper	Clinical DWH	--	Increase the performance of OLAP queries on clinical DWH

Table 2. Explain the Experimental Setup, Related Architecture and Future Ideas in Data Warehouse Architecture

Reference	Experimental environment	Related architecture / work	Future idea
[1]	Analysis of large amount of historical data of student electing course	OLAM Architecture	It can be used for banking , security, telecommunication and other fields of data mining.
[2]	Experimental data related to Medicine , Banking and Geography	Complex DWH Architecture	With this architecture one can store any format of complex data in DWH.
[3]	Decision Support Systems	DWH Architecture through Information System Architecture	Can be used to facilitate Business Intelligence for an enterprise.
[4]	Archiplago tool	Real time Data storage Architecture	As it is still immature so a lot of working is required in this domain to accomplish the task
[5]	UML and MDA	Model Driven Architecture for DWH	Future idea is to make the approach plateform independent.
[6]	Decision Support Systems	ETL Architecture	Providing better decision support environment
[7]	Insurance and banking companies	Application Architecture	Propose reference models for important parts of transactional layer of information logistics and collaboration with partner companies
[8]	Any organization	Enterprise data Infrastructure	Development of efficient robust enterprise data warehouse architecture
[9]	Primary care prescribing in NHS therapeutic processes.	Architecture measuring local prescribing activity	Decision making process in medicines management and prescribing combined with other health care services
[10]	Clinical environment	Windowing Data Structure Architecture	Determining efficient and practical data structure combination, for table, column and popular window collection

Table 3. Explain the Techniques, Modifications, Evaluation and Motivations of Data Warehouse Architecture

Reference	OLAP technique name	DWH scope	Novel idea or modification	Technique name	Evaluation	SQL/ ORACLE	Motivation
[1]	Any OLAP technique can be used in this system	It can be used in banking, security *etc*	Novel idea	OLAM	Good	SQL	-
[2]	Any OLAP technique can be used in this system	It can be used to support any format of data	Novel Idea	Complex data Warehouse Architecture Framework	Good	-	-

[3]	Any OLAP technique can be used in this system	It can Facilitate business intelligence	Modification	Information System Architecture	Good	Oracle 9i	-
[4]	Any OLAP technique can be used in this system	It can perform real time data storage in DWH	Novel Idea	Real Time Data Storage Architecture	Good	-	Reducing copy operation between source and destination
[5]	ROLAP	It can automatically load data into DWH from OLTP.	Novel Idea	Model Driven Architecture for DWH	Excellent	Oracle /PL SQL	-
[6]	ROLAP	DWH using query processing	Novel Idea	ROLAP	Fair	SQL	-
[7]	MOLAP	Collaborating different business companies	Novel Idea	Multi-Channel management	Good	SQL	Collaboration with partner companies is a good foundation for addressing the issue
[8]	Any OLAP technique	Elimination of redundancy	Novel idea	Data transformation, mapping, scrubbing or summarization must occur prior to placing data in EDW	Good	SQL/ ORACLE	Reducing the redundancy to its minimum.
[9]	MOLAP	Decision making process in medicine management	Modification	Data from ETL passed through content management	Good	SQL	Improved decision making support
[10]	Any OLAP technique	WDSA as a DWH so that whole DWH can be represented inside WDSA	Modification	Utilizing windowing for effective OLAP optimization	Excellent	SQL (RDBMS)	Reduce the search space without incurring much processing overhead

Table 4. Explain the Schema, Approach, Application, Analysis and ETL Process of the Data Warehouse

Reference	Gui Schema Yes/no	Data warehouse approach	Data ware house type	Application type	Analysis type	Layered approach	Tool	ETL Yes/no
[1]	Yes	OLAM	Financial / Telecommunication *etc*	Financial / Telecom / Educational	-	No	SQL Server 2000	Yes

[2]	No	Complex DWH	Can be used with any DWH Type	Medicine, Banking and Geography	-	No	-	Yes
[3]	No	Info System Architecture	Business / Financial/ Intelligence	Business Intelligence	-	Yes	Delphi 7.0 Oracle	Yes
[4]	No	Real Time Data storage arch	Compatible with all types of DWHs	Compatible with all types of DWHs	-	Yes	ArchStudio	Yes
[5]	Yes	Model Driven Approach	Compatible with all types of DWHs	Compatible with all types of DWHs	-	Yes	AndroMDA	Yes
[6]	No	ROLAP	Compatible with all types of DWHs	Compatible with all types of DWHs	-	No	-	Yes
[7]	No	Dimensional approach	Business / Financial/ Intelligence	Business Intelligence	-	Yes	-	Yes
[8]	No	EDW	Compatible with all types of DWHs	Compatible with all types of DWHs	Data Analysis	Yes	-	Yes
[9]	Yes	Dimensional approach	Compatible with all types of DWHs	Compatible with all types of DWHs	Decsion making on data	Yes	ePACT	Yes
[10]	Yes	Dimensional approach	All DWH with WDSA architecture	Compatible with all types of DWHs with WDSA	Data mining	No	--	Yes

3. Future Work

Many organizations use Business Intelligence (BI) technology to make better decisions for the improvement of business. Data ware House is one of the key factor / foundation for Business Intelligence. The prerequisites for Business Intelligence is that information should be in updated form, can be easily mapped into knowledge / Intelligence and can be in any format (complex data types).

The other important challenge for the researchers is to introduce budget estimation for such a state of the art Data ware House Architecture which can facilitate Business Intelligence.

4. Conclusion

The main theme of this paper is the comparison of different Data ware House Architectures. Each architecture comes with its own unique benefits but there is still need for a state of the art architecture which can support all features. Data warehouse that does not appear to always be up to date tend to lose the confidence of company, on the other hand Data ware House which cannot store complex data is also useless sometimes. Additionally DWH which cannot be a good source for DSS (Decision Support Systems) is also another drawback. The conclusion for this paper is to draw the attention of the researchers to design such a model for Data ware House which can accommodate all such benefits.

References

[1] F. Wang, “Research and application of OLAM architecture”, InFuture Computer and Communication (ICFCC), 2010 2nd International Conference, **(2010)**.

[2] J. Darmont, O. Boussaid, J. C. Ralaivao and K. Aouiche, “An architecture framework for complex data warehouses”, **(2007)**.

[3] M. M. Dias, T. C. Tait, A. L. A. Menolli and R. C. Pacheco, “Data warehouse architecture through viewpoint of information system architecture”, Computational Intelligence for Modelling Control & Automation, International Conference, **(2008)**.

[4] J. Nealon, W. Rahayu and E. Pardede, "Improving clinical data warehouse performance via a windowing data structure architecture", Computational Science and Its Applications, International Conference, **(2009)**.
[5] O. Yektamanesh, J. Habibi, H. Ahmadi and G. V. Shanjani, "Towards an Architecture for Real-Time Data Storage", InComputational Intelligence, Modelling and Simulation (CIMSiM), 2010 Second International Conference, **(2010)**.
[6] L. A. Fernandes, B. H. Neto, V. Fagundes, G. Zimbrão, J. M. de Souza and R. Salvador, "Model-Driven Architecture Approach for Data Warehouse", Autonomic and Autonomous Systems (ICAS), 2010 Sixth International Conference, **(2010)**.
[7] S. Redzanovic, P. Chountas, T. Chaussalet, F. Fouladinejad and M. Tadjer, "Data warehousing based architecture for the reporting of the NHS primary care prescribing", Computer-Based Medical Systems (CBMS), 24th International Symposium, **(2011)**.
[8] M. Rifaie, K. Kianmehr, R. Alhajj and M. J. Ridley, "Data warehouse architecture and design", Information Reuse and Integration, IRI 2008. IEEE International Conference, **(2008)**.
[9] S. Redzanovic, P. Chounta, T. Chaussalet, F. Fouladinejad and M. Tadjer, "Data warehousing based architecture for the reporting of the NHS primary care prescribing", Computer-Based Medical Systems, (CBMS), 24th International Symposium, **(2011)**.
[10] J. Nealon,, W. Rahayuand E. Pardede, "Improving clinical data warehouse performance via a windowing data structure architecture. InComputational Science and Its Applications", ICCSA'09. International Conference, **(2009)**.

Author

Muhammad Arif, he is a PhD student at Faculty of CS and IT, University of Malaya. Currently he is working on Medical image Processing. His research interests include image processing, E learning, Artificial intelligence and data mining. He joined UM as a Bright Spark Scholar in September 2013 for the period of 3 years. Before this he completed masters and bachelor degrees in Pakistan. He received his BS degree in Computer Science from University of Sargodha, Pakistan in 2011. He obtained his MS degree in Computer Science from COMSATS Islamabad 2013 Pakistan.

