

Data Warehouse Implementation of Examination Databases

Er. Majid Zaman¹, Dr. S. M. K. Quadri², Er. Muheet Ahmed Butt³

Directorate of Information Technology & Support Systems, University of Kashmir, Srinagar, J&K, India
Head & Director, PG Department of Computer Science, University of Kashmir, Srinagar, Srinagar, J&K, India
Directorate of Information Technology & Support Systems, University of Kashmir, Srinagar, J&K, India

Abstract: A data warehouse is a relational database that is designed for query and analysis rather than for transaction processing. It usually contains historical data derived from transaction data, but it can include data from other sources. In addition to a relational database, a data warehouse environment includes an extraction, transportation, transformation, and loading (ETL) solution, an online analytical processing (OLAP) engine, client analysis tools, and other applications that manage the process of gathering data and delivering it to business users [2][10]. This was the case at University of Kashmir (UOK) Examination Department, where such a project brought together various attributes of Examination System which included Conduct, Secrecy, Transit, Tabulation, Accounts and other related data sources in to an integrated data warehouse.

Keywords: Data Warehouse, OLAP, ETL, Examinations.

1.0 Introduction

To remain competitive in today's academic climate, any University needs a foundation of quality data that too in examination system where the data need to be more precise and consistent. Organizations of higher education need this capability as much as Fortune 500 companies. Organizations, both governmental and business, have to manage large amount of information stored in some form of databases or files [1][9]. One of the main problems to deal with information managing is the weak interoperability between various databases and information systems. Especially this problem is serious when we want to organize collaboration between the information systems of various departments within the organization. Solution is data-warehousing, the idea of drawing data from several different (heterogeneous) data sources on different platforms and computers. Building a data warehouse is extremely complex and takes commitment from both the information technology department and the business analysts of the organization. It takes planning hard work, dedication, and time to create a relational database that delivers the right data to the right user. University of Kashmir's Examinations data warehouse (UOKEDW) is not a panacea for every data problem, but it is a very good start toward a permanent solution.

2.0 Towards Computerization

University of Kashmir Examination Department declares 450 evaluation and 450 re-evaluation results of around 3.5 Lakh enrolled students appearing in both under and Post Graduate Examinations of around 200 affiliated colleges. University of Kashmir, like any academic organization in India started realizing need and importance of computerization in late 1990 and early 2002, however it was not those days that we were provided with RAD tools. In house computerization was done in FoxBASE, dbase and SCO Unix as operating system, three teams were set up initially to computerize following systems,

- Tabulation
- Transit
- Secrecy
- Certificates
- Accounts
- And Conduct

As year passed on, information explosion within the department was at its peak, it was not much long before university administration realized that in order to manage such a crucial data of the Examination Department and provide friendly access to information, new teams needs to be setup with enhanced budget be provided in order to store/manage data to meet the changing demands.

Newly constituted teams were given decent budget along with freedom to choose the tools required for development, within no time university was riding on with the success of Information Technology and almost all the areas of the university administration were computerized-n autonomous solutions within single organization each working fine but with no possibility of integration.

3.0 UOK Examinations Data Warehouse Development

Development of UOK Examination data warehouse started in the in around 2004 as a client/server project. The tools which were mainly used in the first phase of development were Microsoft SQL Server 2000 and Windows Server 2005 workstation. Every software development carried out on these platforms was done in

home. While getting the warehouse server in place, the software development team provided data access software for efficient usage of the warehouse in implementation of both read and write operations [3]. Although many of the access tools were in their adolescence at the time, accessing data was much easier with these graphical user interface (GUI) tools than with the fourth-generation tools then in use.

University of Kashmir formed a development team of two Project leaders with ten Master of Computer Applications final year students and Computing Assistants from the data administration and Examination Automation Centre to build the data warehouse. The team selected a representative group of business analysts to serve as pilot users to test the warehouse and access software. During the next few months, the team built a “student” warehouse model based on over 150 questions, which the pilot users considered difficult or critical to answer using current information resources [10][11].

During 2006, many of the original data warehouse team members shifted back to their regular duties, leaving a core of six fulltime equivalent employees working on the project. That core has remained intact, receiving additional help from UOK’s institutional research office and many of the business analysts who are regular users of the warehouse. Also, the data administration department initiated a formal program to train users on the warehouse. To date, there are over 250 trained warehouse users which have been trained while carrying out training programs for various employees. The major goal is to train 400 odd employees, approximately 10 percent of UOK’s employees.

4.0 Examination Warehouse Architecture

UOK’s Examination data warehouse resides in a client/ server environment. UOKEDW extracts data from loaded on to the Unix Server Majorly kept for Secure Data Entry Process for marks entry and loads it into a Microsoft Windows server running an MS SQL as RDBMS. UOKEDW server is a IBM Xenon Server with 8 GB of memory and two processors, running the Windows 2008 Server operating system. Users connect through Ethernet to the warehouse over UOK Campus Area Network backbone via Transmission Control Protocol/Internet Protocol (TCP/IP). The suggested GUI data access has was first implemented in Visual Basic 6.0 now transformed to Visual Basic .NET 2008., which runs identically on the Windows. Microsoft Access® is another tool used mostly for data migration from one database Server to other. The process of using GUI tools to build structured query language (SQL) requests and bring the results back to a client machine. With client/server architecture, once the data are in the workstation, users “own” the data, cutting and pasting at

will into their favorite software (e.g., spreadsheet, word processor, graphic tools) [4].

5.0 Examination Data Warehouse implementation Phases

The following steps were under taken for designing of UOKEDW

5.1 Requirement Gathering: The first thing that the constituted project team was engaged in gathering requirements from the various employees working in the examination system. Because end users were typically not familiar with the data warehousing process or concept, requirement gathering was implemented using one-to-one meetings or as Joint Application Development (JAD) sessions, where multiple stake holders in the examination wing were interacted with so that the requirement analysis done in a proper manner.

5.2 Physical Environment Setup: Once the requirements gathering were somewhat clear, it became necessary to set up the physical servers and databases. At a minimum, it was necessary to set up a development environment and a production environment. It was not enough to simply have different physical environments set up [5]. The different processes many data warehousing projects where there were three environments: Development, Testing, and Production (such as ETL, OLAP Cube, and reporting) also need to be set up properly for each environment. The primary goal of this phase was to identify what constitutes as a success for this particular phase of the data warehouse project.

5.3 Data Modeling: This was a very important step in the data warehousing project. Indeed, it was fair to say that the foundation of the data warehousing system is the data model [7]. A good data model will allow the data warehousing system to grow easily, as well as allowing for good performance. In UOKEDW project, the logical data model was built based on user requirements, and then it is translated into the physical data model.

5.4 ETL (Extraction, Transformation, Loading) process typically took the longest to develop the UOKE’s data warehouse implementation cycle as there were many heterogeneous data bases involved in extraction transformation and loading process. The reason for this was that it took time to get the source data, understand the necessary columns, understand the business rules, and understand the logical and physical data models before ETL would have been successfully carried out.

5.5 OLAP Tube Design: The OLAP cube was derived from the Requirement Gathering phase [6]. The users

working in the Examination Wing had some idea on what they want, but it was difficult for them to specify the exact report / analysis they wanted to see and analyse. When this was the case, it is usually a good idea to include enough information so that they feel like they have gained something through the data warehouse, but not so much that it stretched the data warehouse scope by a mile. Hence front end development became an important part of a data warehousing initiative of UOK.

5.6 Front End Options: The front-end options ranged from an internal front-end development using scripting languages such as VB, VB .NET, ASP, PHP, to off-the-shelf products such as Crystal Reports, to the more high-level products such as Actuate. When choosing vendor tools, it was made sure that it could be easily customized to suit the business of examination, especially the possible changes to the reporting requirements of the Examination System. Possible changes included not just the difference in report layout and report content, but also included possible changes in the back-end structure.

5.7 Report Specification: Report specification typically came directly from the requirements phase [8]. To the end user/employee working in the examination system, the only direct touch point he or she had with the data warehousing system is the reports they see and analyse. So, report development, although not as time consuming as some of the other steps such as ETL and data modeling, nevertheless play a very important role in determining the success of the data warehousing project.

5.8 Query Processing – In this the OLAP reports or reports were made to run directly against the RDBMS often exceeded the time limit, and it was hence ideal for the data warehousing team to invest some time to tune the query, especially the most popularly ones.

6.0 Security and privacy

Security and safeguarding privacy are major concerns when building a data warehouse and when it came to Examinations sensitive data the applicability of security and privacy was at prime focus. Security in a database means protecting data against unauthorized disclosure, alteration, or destruction. Granting select (authorization to read only) access to tables or views achieves a certain level of security in a warehouse [12]. At UOKEDW, read-only access to the data warehouse was provided at the database level. This procedure followed an open access policy for employees approved by UOKEDW administration in 2008. This policy was based on the notion that giving employees access to data and holding them accountable is better for the organization than withholding the data.

Although many RDBMSs support column level security [13][14], UOKEDW has not implemented this feature, primarily due to the high cost of administering user access. In traditional operating systems, tasks or screens control access, meaning users only have access to a single record or instance of data (e.g., verifying admission status of a student). In a data warehouse, users have access to a table or set of tables in a subject area, which means access goes beyond retrieving single records to retrieving groups of records.

At UOKEDW, the Assistant Registrar is the trustee of the student Registration database, Assistant Controller Secrecy is the trustee of Secrecy Database, Assistant Controller Conduct is the Trustee of Conduct Database, and so forth. In these databases the write/read access is given at various levels to the employees depending on the nature of work they have in examination system. To obtain name and address information, the data trustee grants access to the person database. The user's business need determines whether access is granted. Given the large number of records in UOKE's data warehouse, placing name and address in a separate database achieves a certain level of privacy.

7.0 Conclusion

The future of UOKE's data warehouse is becoming more clear. Initially, the warehouse served as a resource for accessing information from legacy systems. Eventually, the warehouse will serve as a telescope into UOKEDW's distributed data stores. Some of these data will reside in the data warehouse, while other elements will be "viewed" from the RDBMSs where the data reside. UOKEDW foresees a time when the telescope extends beyond UOKEDW to other organizations with common goals, such as the neighboring Maricopa County Community College District. The real power of the warehouse will be actualized in years to come. The data warehouse fills an important data administration role in a client/server environment. As distributed application developers move further away from the central computing core, the data elements in the warehouse ensure the integrity of the organization's enterprise data.

The bottom line is that data warehousing is here to stay. Warehousing gives organizations the opportunity to "get their feet wet" in client/ server technology, distributed solutions, and RDBMS. This is essential for any future mission critical application, making the data warehouse a low-risk, high-return investment.

8.0 References

- [1] Jorge Bernardino, Pedro Furtado, Henrique Madeira, "A Cost Effective Approach for Very Large Data Warehouses", Proceedings of the International Database Engineering and Applications Symposium, 2002

- [2] Kimball, Ralph, "The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses", John Wiley & Sons, Inc, 1996.
- [3] Krishna. "Principles of Curriculum Design and Revision: A Case Study in Implementing Computing Curricula CC2001". ITiCSE'05, June 27–29, 2005
- [4] Larry, Greenfield, LGI Systems Inc., "The Data Warehousing Information Center," 1997 pp <http://pwp.starnetinc.com/larryg/index.html>.
- [5] M. A. R. Kortnik, D. L. Moody. "From Entities to Stars, Snowflakes, Clusters, Constellations and Galaxies: A Methodology for Data Warehouse Design." 18th. International Conference on Conceptual Modelling. Industrial Track Proceedings. ER'99.
- [6] W. J. Labio, Y. Zhuge, J. N. Wiener, H. Gupta, H. Garcia-Molina, J. Widom. Stanford University. "The WHIPS Prototype for Data Warehouse Creation and Maintenance". SIGMOD 1997
- [7] Marotta. "A transformations based approach for designing Data Warehouses Internal Report." InCo. Universidad de la República, Montevideo, Uruguay. 1999.
- [8] D. Theodoratos, T. Sellis (DWQ project). "Designing Data Warehouses." DKE '99
- [9] F. Bancilhon, N. Spyrtatos, "Update Semantics of Relational Views," ACM TODS 6, 1981, 557–575.
- [10] A. Bauer, H. Günzel, eds., "Data Warehouse Systeme — Architektur, Entwicklung, Anwendung, dpunkt.verlag", 2001.
- [11] J. Lechtenböcker, G. Vossen, "On the Computation of Relational View Complements," Proc. 21st PODS 2002, 142–149.
- [12] J. Lechtenböcker, G. Vossen, "Multidimensional Normal Forms for Data Warehouse Design," 2002, to appear in Information Systems, Elsevier Science.
- [13] W. Lehner, J. Albrecht, H. Wedekind, "Normal Forms for Multidimensional Databases," Proc. 10th SSDBM 1998, 63–72.
- [14] H. Lenz, A. Shoshani, "Summarizability in OLAP and statistical databases," Proc. 9th SSDBM 1997, 132–143.