Association for Information Systems

AIS Electronic Library (AISeL)

ICEB 2003 Proceedings

International Conference on Electronic Business (ICEB)

Winter 12-9-2003

Implementing a Three-Tier Data Warehouse

Ahmed A. El-Ragal

Adel A. El-Zoghabi

Mohamed M. Abd El-Salam

Frank Poole

Implementing a Three-Tier Data Warehouse

Ahmed A. El-Ragal
Assistant professor
MIS Department
College of Management & Technology
Arab Academy for Science, Technology, and Maritime Transport
P.O. Box 1029, Miami
Alexandria, Egypt
Tel + 203-5485473
Fax + 203-5566072
E-mail:
ahmed.elragal@aast.edu

Adel A. El-Zoghabi
Associate Professor
Department of Information Technology
Institute of Graduate Studies and Research
Alexandria University
P.O.Box 832, El-Shatby
Alexandria, EGYPT
Tel + 203-4295007
E-Mail:
zoghabi@yahoo.com

Mohamed M. Abd El-Salam
Engineer
MIS Department
College of Management & Technology
Arab Academy for Science, Technology, and Maritime Transport
P.O. Box 1029, Miami
Alexandria, Egypt
Tel + 203-5485473
Fax + 203-5566072
E-mail:
mmahmoud@mis.aast.edu

Frank Poole
Professor
School of Computing and Management Sciences,
Sheffield Hallam University
Pond Street, Sheffield
S1 1WB England
Tel + 440 114 225 4239
Fax
E-mail:
F.Poole@shu.ac.uk

Abstract

This paper aims at developing a 3-tier inter-universities data warehouse prototype for the Egyptian universities. The implementation scope is restricted to the student enrollment process. The bottom-up approach and the multi-tier (3-tier) client/server architecture were used to implement the proposed prototype. The implementation required a number of steps to be undertaken. First, a star schema data warehouse was built based on an operational database of a public university. Second, another star schema data warehouse was built based on an operational database of a private university. Third, data from the two warehouses were abstracted to formulate the inter-universities data warehouse (tier 3). Hence, a data cube based on the resulting inter-universities data warehouse was developed using an OLAP server (tier 2). The data cube is then accessed by a client tool (tier 1) for the purpose of query and analysis. Although two universities were used to implement the inter-universities data warehouse, this scenario could be applied with any number of universities in a quite similar way. This inter-universities prototype is scalable and flexible to retain more than two universities regardless of the size of the operational databases. The inter-universities prototype fosters the coordination between participating universities and supports the decision making process. The proposed system could be used to generate a variety of strategic reports.

1 Introduction

Data warehousing is gaining popularity as organizations realize the benefits of being able to perform sophisticated analyses of their data [31]. The quality of data and the variations in its form are the toughest obstacle to building good data warehouses [8]. Organizations moving toward the approach of adopting data warehouse should discuss and answer the following questions: What to put in that data warehouse? The principal task will be determining: what to extract from data sources, how to extract it, transform it, and invariably, clean it. Then, where will the data be stored? How the information is delivered to end-users? There are many different approaches, of course, and they need not be mutually exclusive [2].

2 Data warehouse (DW) definition

Several definitions have been proposed by [17], [34], [3], [24], [2], [4], [7], [1], [33], [13], [23]. Inmon and Hackathorn [17] defined

DW as a subject-oriented, integrated, time-variant, and non-volatile collection of data used in support of management's decision making process. Devlin [7] said that a DW is simply a single, complete, and consistent store of data obtained from a variety of sources and made available to end users in a way they can understand and use in a business context. Adamson and Venerable [1] defined it as a design aimed at getting data out; it is all about getting answers to business questions. Marakas [23] said that the data warehouse contains a copy of the transactions which should not updated or changed later by the data warehouse but could be updated by the transaction system from which the DW was originated. According to Marakas [23] and Inmon [16] there are four main characteristics that generally describe a DW. Those are Time-variant, Non-volatile, Subject-oriented, and Integrated.

3 DW architectures

The data warehouse application is a client/server application. The role of the client in all the variations generally includes managing the user interaction and formatting and reporting data retrieved from the data warehouse [8]. There are two basic architectures for building a data warehouse. The two-tier data warehouse architecture consists of front-end client components and back-end server components. The two-tier architecture utilizes existing legacy systems as database servers for the data warehouse (tier 2), while LAN-based workstations perform all complex calculations and summaries that go beyond the capabilities of the data warehouse's DBMS (tier 1) [8]. The two-tier architecture is not scalable and does not support large numbers of on-line end users. It encourages the development of *fat-client* front-end systems, in which excessive application processing is allocated to desktop workstations [26]. The multi-tier (*three-tier*) data warehouse architecture or as sometimes called the *thin-client* model, handles the scalability and flexibility problems through the use of an application server between the data warehouse server and the client. The application server manages the interaction with the data warehouse, performs data filtering, summarization, and aggregation, and sends the results to the client. A variation in the three-tier architecture is the use of a multi-dimensional database (MDD) server that stores the data in a special structure designed to facilitate dimensional analysis. The client remains responsible for managing the user interaction and formatting and reporting resulting data, but requires less processing power due to the reduced calculations and database server interaction [9], [8], [3].

4 The star schema architecture

The best way to build the DW database is by using the Star Schema structure (sometimes referred to as multi-dimensional data modeling). The Star Schema captures the measurements of importance to the business and the parameters by which the business measurements are broken down. It is a direct reflection of how business processes happen. The measurements are referred to as FACTS, whilst the parameters by which a measurement can be viewed are called DIMENSIONS [30], [1], [18].

5 OLAP

OLAP is a common use of a data warehouse that involves real-time access and analysis of multi-dimensional information [32]. OLAP always involves interactive querying of data, following a thread of analysis through multiple passes [27], [29], [14]. There are three major architectures for OLAP systems [6]:

1) Multi-dimensional databases OLAP, (MOLAP). MOLAP uses multi-dimensional databases to provide analyses.
2) Relational OLAP, (ROLAP). ROLAP uses data directly from relational databases using either a star or snowflake database schema.
3) Hybrid OLAP, (HOLAP). In this type, some summarized or aggregated data are stored in a multi-dimensional database,

while other data, such as detailed transaction data, is stored in related relational databases. The user can navigate the multi-dimensional database analyzing the summary (aggregate) data, but if the user requires data that is not stored in the multi-dimensional database, the system automatically generates SQL to retrieve the data from the back-end relational database and then loads that data into the multi-dimensional database, which is referred to as drill-through [6].

6 Inter-organizations data warehouse

An inter-organizations data warehouse can be established between two or more organizations whenever they agree to share data. The separate organizations have separate technologies, IT staffs, data elements, and administrative routines. What ties these diverse entities together is a strong desire to share data for the purpose of producing coordinated overall results. An inter-organizations DW does not come out because a group of organizations just want to share data. Significant planning and arrangement must take place before an inter-organizations DW can go live. All organizations participating in the DW must accept a centralized agreement. The members of the DW agree to the following [19]:

1) To define and use conformed dimensions. For instance, all members will use the ITEM dimension in a retail chain. All must agree on the definition of attributes in this dimension so that a compatible set of values from separate dimension tables can be combined together.
2) To define and use conformed facts. For instance, all members must agree on the definition of the "revenue" measure in the DW so that separate revenue measurements coming from separate fact tables can be combined mathematically. Whenever conformed dimensions and conformed facts are arranged correctly, the separate organizational members of the DW are quite free to develop their databases independently.

7 Foundations and significance of research

This research is based on recommendations made by many scholars including: [5], [10], [12]. These recommendations suggest the establishment of an Inter-Organizations Data Warehouse. Inter-organizations data warehouse can be established between two or more organizations for the purpose of producing coordinated overall results, whenever they agree to: share data, use conformed dimensions, and use conformed facts. Inter-organizations data warehouse can help member organizations to enhance their business understanding and accordingly the decision quality [19].

8 Research objectives

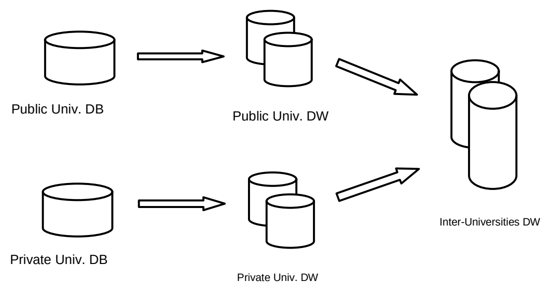
1) Developing an Inter-Universities Data Warehouse prototype for the Egyptian Universities;
2) Support the decision makers by generating some reports based on the Inter-Universities Data Warehouse.

9 Methodology

The adopted methodology consists of the following components together with their respective mechanisms:

1) The bottom-up approach for DW development will be used, and
2) A multi-tier (three-tier) client/server architecture will be employed.

The bottom-up approach for DW development will be used as follows. First, a Star Schema Data Warehouse is built on the operational database of a Public University. Second, another Star Schema Data Warehouse is built on the operational database of a Private University. Third, data from the two warehouses is abstracted to form the Inter-Universities DW, as shown in Fig. 1. The multi-tier (three-tier) client/server architecture is deployed as follows. A *data cube* based on



the resulting Inter-Universities DW (tier 3) will be developed and run by an OLAP server (tier 2). The data cube can be accessed by the client's tool (tier 1) for query and analysis as needed.Fig.1. Bottom-Up Approach for Building Inter-Universities DW

9.1 Data sources

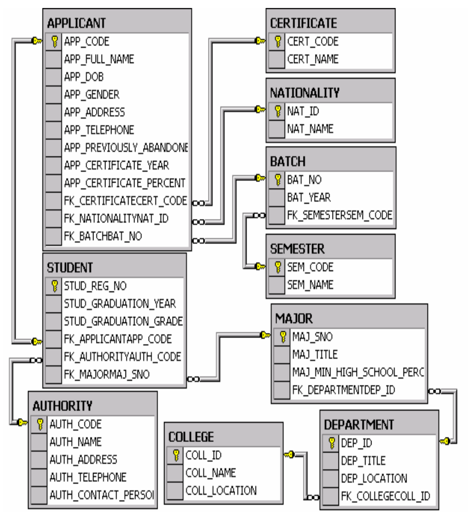


Fig. 2. ER Diagram of the 'Public University' Database

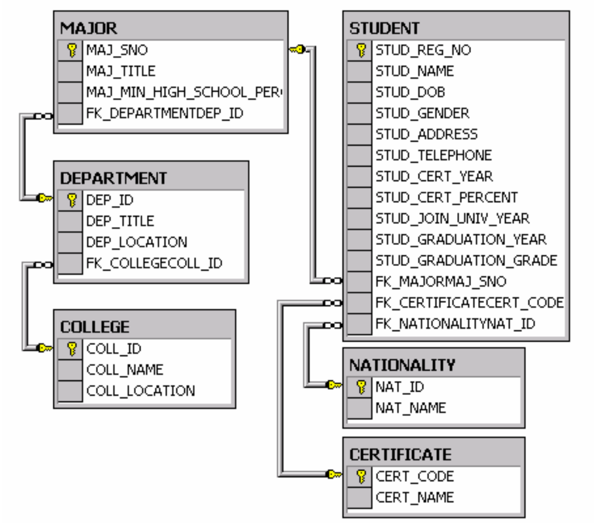


Fig. 3. ER Diagram of the 'Private University' Database

The following sections present a description of the proposed operational databases of the 'Public University' and the 'Private University'. This is done within the scope of student enrollment.

9.2 The public university database

This data source represents any of the Public Universities. An ER diagram of the Public University database, within the scope of student enrollment, is shown in Fig. 2. This database is a downscale of the Admission and Registration database in [12]. The database consists of the following entities: STUDENT, CERTIFICATE, NATIONALITY, MAJOR, DEPARTMENT, and COLLEGE.

9.3 The private university database

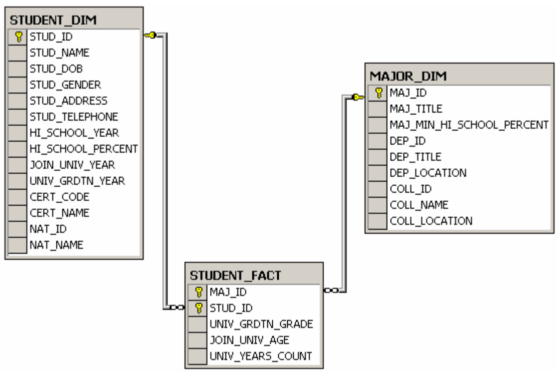
This data source represents any of the Private Universities. An ER diagram of the Private University database, within the scope of student enrollment, is shown in Fig. 3. This database is a downscale of the Admission and Registration database in [12]. The database consists of the following entities: APPLICANT, BATCH, SEMESTER, STUDENT, AUTHORITY, CERTIFICATE, NATIONALITY, MAJOR, DEPARTMENT, and COLLEGE.

9.4 Design of the star schema DW

In the following sections, the steps of designing the Star Schema DW for each university, based on its operational database, are presented. The design steps of a Star Schema DW include determining the grain of the fact table, the dimensional tables, and the measured facts. Then, the design of the Inter-Universities Data Warehouse is presented.

9.5 Design of the public university star schema DW

The Public University Star Schema DW is based on the Public University operational database, presented earlier. Since the proposed Star Schema DW concerns with the process of



student enrollment, the grain of the fact table has been chosen to be the individual *student*. Two dimensions were used: the *Major* (named as MAJOR_DIM), and the *Student* (named as STUDENT_DIM), as shown next.

- **MAJOR_DIM.** It was built by de-normalizing the tables: MAJOR, DEPARTMENT, and COLLEGE.
- **STUDENT_DIM.** It was built by de-normalizing the tables: STUDENT, CERTIFICATE, and NATIONALITY.

The following measured facts of a student were chosen: *university graduation grade*, *age at joining the university*, and *count of years in university*. An ER diagram of the Public University Star Schema structure, within the scope of student enrollment, is shown in Fig.4.

Fig. 4. ER Diagram of the ‘Public University’ Star Schema

9.6 Design of the private university star schema DW

The Private University Star Schema DW is based on the Private University operational database, presented earlier. Since the proposed Star Schema DW concerns with the process of *student enrollment*, the grain of the fact table has been chosen to be the individual *student*. Three dimensions were used: the *Applicant* (named as APP_DIM), the *Major* (named as MAJOR_DIM), and the *Student* (named as STUDENT_DIM), as shown next.

- **APP_DIM.** It was built by de-normalizing the tables: APPLICANT, CERTIFICATE, NATIONALITY, and BATCH.
- **MAJOR_DIM.** It was built by de-normalizing the tables: MAJOR, DEPARTMENT, and COLLEGE.
- **STUDENT_DIM.** It was built by de-normalizing the tables: STUDENT, and AUTHORITY.

The following measured facts of a student were chosen: *university graduation grade*, *age at joining the university*, and *count of years in university*. An ER diagram of the Private University Star Schema structure, within the scope of student enrollment, is shown in Fig. 5.

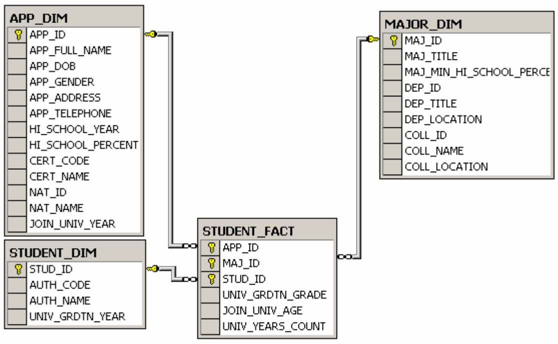


Fig. 5. ER Diagram of the ‘Private University’ Star Schema

10 Design of the inter-universities star schema DW

The Inter-Universities Star Schema DW is based on the two data warehouses: the Public University DW and the Private University DW. By abstracting data of the conformed dimensions and facts from the two data warehouses, the Inter-Universities DW can be built. The conformed dimensions that have been chosen are: the *major* and the *student*. Not all attributes of the two dimensions have been taken. The conformed facts of a student that have been chosen are: *university graduation grade* and *count of years in university* (also, not all of them have been taken). An ER diagram of the Inter-Universities Star Schema structure, within the scope of student enrollment, is shown in Fig. 6.

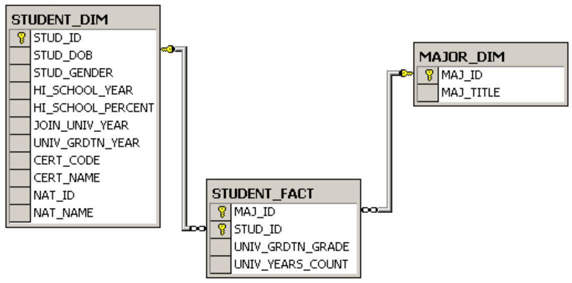


Fig. 6. ER Diagram of the ‘Inter-Universities’ Star Schema

11 Implementation tools

As mentioned earlier, the bottom-up approach was used for building the Inter-Universities DW, on which a data cube was developed using an OLAP server, to be accessed by the end user. This is referred to as multi-tier (3-tier) client/server architecture. Next, the tools used to implement each tier are presented. The back-end tier is the Inter-Universities DW, and was built using Microsoft® SQL Server™ 2000. The middle tier is the data cube, and was developed and managed by Microsoft® SQL Server™ 2000 Analysis Services. The front-end tier is the end user (client), and uses Microsoft® Excel™ 2000 to access the data cube and generate the required reports. Microsoft® Excel™ 2000 provides a client-based PivotTable® Service component which receives meta data with data from the server in response to a query. PivotTable Service shares much of the same functionality as the server, enabling it to bring the server's multidimensional calculation engine, caching features, and query management directly to the client computer. This client/server data management model optimizes performance and minimizes network traffic [25].

12 Building the star schema data warehouses

Building the back-end tier—the ‘Inter-Universities’ Star Schema DW— is based on both the ‘Public University’ Star Schema DW and the ‘Private University’ Star Schema DW. The Star Schema Data Warehouses were built using the tools delivered with Microsoft® SQL Server™ 2000 as will be shown. Building each Star Schema Data Warehouse includes two main steps: loading the Star Schema dimensions, and loading the Star Schema fact tables. To load one or more of the Star Schema dimensions and fact tables, it is needed to build a Data Transformation Services (DTS) Package using the DTS Designer delivered with Microsoft® SQL Server™ 2000. The DTS Designer is a graphically oriented tool for building a number of connections and a number of data transformation flows (also known as data pumps) to run between the connections. The DTS Package also serves for mapping operational data model to data warehouse data model. Before creating the required DTS Packages, it is needed to create the Star Schema database structure for each DW and all the tables in it including keys and other constrains.

13 Building the inter-universities star schema data warehouse

The required DTS Package was created with three connections. The first connection to the Public University star schema, the second connection to the Private University star schema, and the third connection to the Inter-Universities star schema.

14 Developing the data cube

The data cube (or as sometimes called OLAP cube) was developed and managed by Microsoft® SQL Server™ 2000 Analysis Services, based on the Inter-Universities DW built above. The MOLAP storage mode (or as sometimes called multidimensional structure) was chosen for creating the proposed data cube. Microsoft® SQL Server™ 2000 Analysis Services system provides PivotTable® Service, which is OLE DB compliant provider. The data cube built above can be accessed for data query and analysis by Microsoft® Excel™ 2000 through Microsoft® OLE DB Provider for OLAP interface. Microsoft® Excel™ 2000 provides the ability to build PivotTable and PivotChart reports based on this interface. Thus, a wide variety of reports can be generated to fulfill the requirements of various user communities.

15 Sample reports

The following sections present sample reports that are generated as a result of using a random data sample of 1900 records. The data sample contains some personal information about the students joining the university, and some other information about admission, majors, graduation grades, etc.

15.1 Student' high school scores Vs. average university graduation grades

Fig. 7 shows the average university graduation grades of the students against their high school scores in different majors for the university-members of the Inter-Universities DW.

15.2 University demand curve

This report shows the demand curve for education in the universities represented by the number of students joining the universities each year. This report can help the decision maker to arrange the required resources to fulfill the potential demand either on the major level or the university level.

15.3 Graduate students count Vs. University years count

This report shows how many students spent how many years in the university for the different majors. The report helps to show the distribution of graduate students by their university year's count.

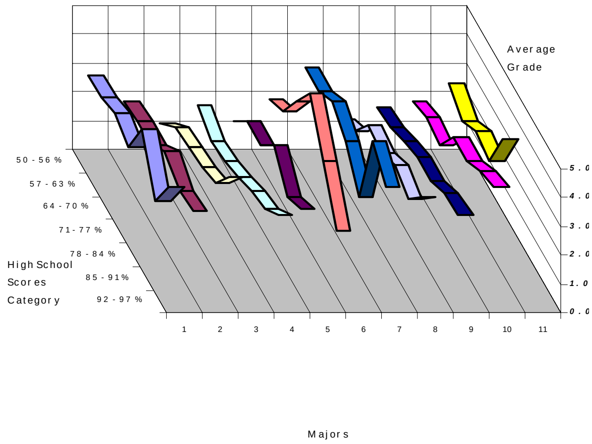


Fig. 7. Students' High School Scores vs. Average University Graduation Grades

16 Conclusions

In this paper the bottom-up approach and the multi-tier (3-tier) client/server architecture were used to develop the Inter-Universities DW. The back-end tier (the Star Schema DW) was built using Microsoft® SQL Server™ 2000 and tools delivered with it. All transformations have been done without changing the operational databases at all. The middle tier (the OLAP data cube) was developed and managed by Microsoft® SQL Server™ 2000 Analysis Services. Analysis Services system provides PivotTable® Service, which is OLE DB compliant provider, to be used by applications from diverse vendors to retrieve data from the server and present it to the user. The front-end tier (end user) uses Microsoft® Excel™ 2000 to access the data cube and generate the required reports. Microsoft® Excel™ 2000 provides a client-based PivotTable® Service component which shares much of the same functionality as the server.

Two universities were used to show how to apply this scenario, but it could be applied with any number of universities in a quite similar way, no matter how big is the size of the operational database since the conformed dimensions and facts have been defined. The proposed system can be used to generate a wide variety of reports showing overall and detailed information about the evolution of education in universities.

References

[1] C. Adamson and M. Venerable, *Data Warehouse Design Solutions*, John Wiley and Sons, New York, 1998.

[2] R. Barquin, "A Data Warehousing Manifesto," *Planning And Designing The Data Warehouse*, R. Barquin and H. Edelstein, ed., Prentice Hall, New Jersey, 1997, pp. 3-16.

[3] A. Berson, *Client/Server Architecture*, McGraw-Hill, New York, 1996.

[4] A. Berson and S. Smith, *Data Warehousing, Data Mining, and OLAP*, McGraw-Hill, New York, 1997.

[5] B. Cooper, H. Watson, B. Wixom, and D.L. Goodhue, "Data Warehousing Supports Corporate Strategy At First American Corporation,"*MIS Quarterly*, vol. 24, no 4, 2000, pp. 547-567.

[6] R.S. Craig, J.A. Vivona, and D. Berkovitch, *Microsoft Data Warehousing*, McGraw-Hill, New York, 1999.

[7] B. Devlin, *Data Warehouse: From Architecture To Implementation*, Addison-Wesley Longman, Reading, Mass., 1997.

[8] H. Edelstein, "An Introduction to Data Warehousing," *Planning And Designing The Data Warehouse*, R. Barquin and H. Edelstein, ed., Prentice Hall, New Jersey, 1997, pp. 31-50.

[9] R. Elmasri and S.B. Navathe, *Fundamentals of Database Systems*, Addison Wesley Longman, Mass., 2000.

[10] A. El-Ragal and T. Mangles, "Knowledge Discovery in Database Techniques Applied to Students Recruitment Systems in Universities," *Proc. Conf. IAIM 2000*, Brisbane, Australia, 8-10 Dec. 2000, pp. 7-20.

[11] A. El-Ragal and Y. Thabet, *"Query Optimization for OLAP Applications"*, *Proc. Conf. ICEB 2002*, Taiwan, 10-13 Dec. 2002, pp. 337-339.

[12] A. El-Ragal, T. Mangles, and I. Chaston, "Developing Star Schema Structures: A Practical Study Applied to the Egyptian Universities," *Proc. Conf. BIT World 2001*, AUC, Cairo, Egypt, 4-6 June 2001, pp. 128.

[13] L. Feng and T. Dillon, "Using Fuzzy Linguistic Representations to Provide Explanatory Semantics for Data Warehouses," *IEEE Transaction on Knowledge and Data Engineering*, vol.15, no.1, 2003, pp.86-102.

[14] V. Harinarayan, A. Rajaraman, and J.D. Ullman, "Implementing Data Cubes Efficiently," *Proc. Conf. ACM SIGMOD*, Montreal, Canada, 1996, pp. 205-216.

[15] M. Humphries, M. Hawkins, and M.C. Dy, *Data Warehousing Architecture and Implementation*, Prentice Hall PTR, New Jersey, 1999.

[16] W. Inmon, *Building the Data Warehouse*, QED publishing Group, 1993.

[17] W. Inmon and R.D. Hackathorn, *Using the Data Warehouse*, John Wiley and Sons, Inc., New York, 1994.

[18] R. Kimball, *The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses*, John Wiley and Sons, Toronto, 1996.

[19] R. Kimball, "Joint Effort," *Intelligent Enterprise*, vol. 4, no 8, 2001, pp. 42.

[20] E.J. Levin, "Developing a Data Warehousing Strategy," *Planning And Designing The Data Warehouse*, R. Barquin and H. Edelstein, ed., Prentice Hall, New Jersey, 1997, pp. 53-89.

[21] G. Livingston and B. Rumsby, "Database Design for The Data Warehouses: The Basic Requirements," *Planning And Designing The Data Warehouse*, R. Barquin and H. Edelstein, ed., Prentice Hall, New Jersey, 1997, pp. 179-198

[22] Mallach and G. Efrem, *Decision Support and Data Warehouse Systems*, McGraw-Hill, New York, 2000.

[23] G.M. Marakas, *Modern Data Warehousing, Mining, and*

Visualization: Core Concepts, Prentice-Hall, N.J., 2003
[24] R. Mattison, *Data Warehousing and Data Mining for Telecommunications*, Artech House, Norwood, Mass., 1997.
[25] Microsoft, *SQL Server 2000 Books Online*, Microsoft Corporation, 2000.
[26] P. Mimno, "Data Warehousing Architectures," *Planning And Designing The Data Warehouse*, R. Barquin and H. Edelstein, ed., Prentice Hall, New Jersey, 1997, pp. 159-177.
[27] T.P. Nadeau and T.J. Teorey, **"**A Pareto Model for OLAP View Size Estimation," *Proc. IBM Centers for Advanced Studies Conference* **(**CASCON 2001), Toronto, Canada, 5-7 Nov. 2001.
[28] A. Paller, "A Roadmap To Data Warehousing," *Planning And Designing The Data Warehouse*, R. Barquin and H. Edelstein, ed., Prentice Hall, New Jersey, 1997, pp. 17-29.
[29] N. Raden, "Choosing the Right OLAP Technology," *Planning And Designing The Data Warehouse*, R. Barquin and H. Edelstein, ed., Prentice Hall, New Jersey, 1997, pp. 199-224.
[30] J. Sorensen and K. Alnor, "Creating Data Warehouse Using SQL Server," *Proc. Int'l Workshop Design And Management of Data Warehouses* (DMDW 99), Heidelberg, 1999, pp. 10-1:10-9.
[31] J. Srivastava and P. Chen, "Warehouse Creation-A Potential Roadblock to Data Warehousing," *IEEE Transactions on Knowledge and Data Engineering*, vol. 11, no 1, 1999, pp. 118-126.
[32] Y. Taha, A. Helal, and K. Ahmed, "Data Warehousing: Usage, Architecture, And Research Issues," *ISMM Microcomputer Application Journal*, vol. 16, no 2, 1997, pp. 70-80.
[33] E. Turban and J. Aronson, *Decision Support Systems and Intelligent Systems*, Prentice Hall, New Jersey, 1998.
[34] J. Widom, "Research Problems In Data Warehousing," *Proc. 4th Int'l Conf. Information and Knowledge Management* (CIKM 95), ACM, 1995, pp. 25-30.
[35] B.H. Wixom and H.J. Watson, "An Empirical Investigation Of The Factors Affecting Data Warehousing Success," *MIS Quarterly*, vol. 25, no 1, 2001, pp. 17-41.