Using Data Marts Approach in Iraqi Educational Institution process

Alla Talal Yassin

[1]Iraqi Commission for Computers and Informatics/ Baghdad, Iraq

allatalal@yahoo.com , www.allatalal.com

Abstract: Worldwide, educational institution has channeled big investments towards building information systems. This paper deals with data marts ideas, which is a database, or collection of databases, designed to help managers make strategic decisions about their business. In this paper we suggest a project named IC-DM for the educational institution which has access to two different data marts that user will need to decide which is the most appropriate for the project that they are working on. Data from multiple sources, and multiple levels was linked, or merged into a data mart.

Keywords: about four key words separated by commas**.**

الخلاصة: تعتبر المؤسسة التعليمية الأستثمار الاكبر انتشاراً حول العالم نحو بناء انظمة المعلوماتية. يقدم البحث تفاصيل بافكار Data marts والتي يمكن القول بأنها تلك القواعد من البيانات او ماهو منخب والمصمم لمساعدة المدراء لاتخاذ استراتيجية القرار في اعمالهم. في هذا البحث تم اقتراح مشروع سمي ب IC-DM للمؤسسة التعليمية والتي لها امكانية الوصول الى اثنين من تجمعات البيانات المستخدمة المختلفة التي سوف تحدد ماهو انسب للمشروع والذي يتم ربط البيانات من مصادر ومستويات متعددة، او دمجها في data mart.

1. Introduction

Many educational institutions accumulate huge volumes of data every year. Educational institutions where data are annually gathered from learning departments are no exception to this phenomenon. Generally, the data that is structured in the data mart does come from the data warehouse. Since the data in the data warehouse is structured in a normalized pattern it must be reshaped to create the data mart. A data mart can be less expensive than implementing a data warehouse, thus making it more practical for the small business [1].

Student degree has their data mart, courses have theirs, and general information has theirs, and so on. And the data mart for student degree only faintly resembles anyone else's data mart. The data mart is typically housed in multidimensional technology which is great for flexibility of analysis but is not optimal for large amounts of data. The Data found in data marts is highly indexed.

2. Statement of The Problem

The current educational institution information system does not support some features for data analysis. The system equally offers limited capability to integrate all necessary data from different sources that are useful in generating sector performance indicators. This implies that knowledge workers in educational institution find it difficult to perform multi-dimensional analysis for policy decision support.

3. General Objective

To develop data marts that will enable educational institution to utilize existing data to effectively manage, control and direct education policy programs using better information systems.

4. Literature Review

Haiyan and Herstein (2003) suggest that successful management of today's education systems requires effective policy-making and system monitoring through data and information [3]. Thomas and Carolyn (2005) defined data mart as subset of a data warehouse that supports the requirements of a particular department or business function. The education management information system (EMIS) is a sub-system of an education system whose aim is to collect, store, process, analyze and disseminate information[4]. Cassidy (2005) describes an Education Management Information System (EMIS) as a system for the collection, integration, processing, maintenance and dissemination of data and information to support decision making, policy-analysis and formulation, planning, monitoring and management at all levels of an education system [5].

Data marts have many goals like Integrate multiple data sources into single repository, Apply agreed-upon business rules for consistent interpretation of data, Provide access to commonly requested data and Track historical data for point-in-time comparisons [6, 7].

5. Data Marts Approaches

There seems to be two approaches to the construction of a data mart [6].

• First, is to create the data warehouse and then create the data mart.

• The second approach to building data marts is to skip over the data warehouse and begin with the building of several data marts that will eventually join together to build the data warehouse.

6. IC-DM Project Methodology

Many educational institutions have built multiple data marts to serve different departments. Often, these marts have been created with a narrow departmental scope, typically focusing on a single application or subject area, such as student service, employer service, operations, or finance. Data marts are easier to build and simple than an integrated data warehouse (IDW) that is used to get data out to users and, therefore, tend to multiply quickly and become more costly to maintain and operate.

6.1 The purpose

The purpose of a data mart is to provide easy access to frequently needed data that helps to facilitate decision making. In this paper we Design a project named (IC-DM) which has access to two different data marts each of these data marts has strengths and weaknesses and the user will need to decide which is the most appropriate for the project that they are working on. Data from multiple sources, multiple years, and multiple levels was linked, integrated, or merged into a data mart. Figure (1) shows the schema for the traditional approach for DM-1, while Figure (2) shows the schema for the data marts approach.
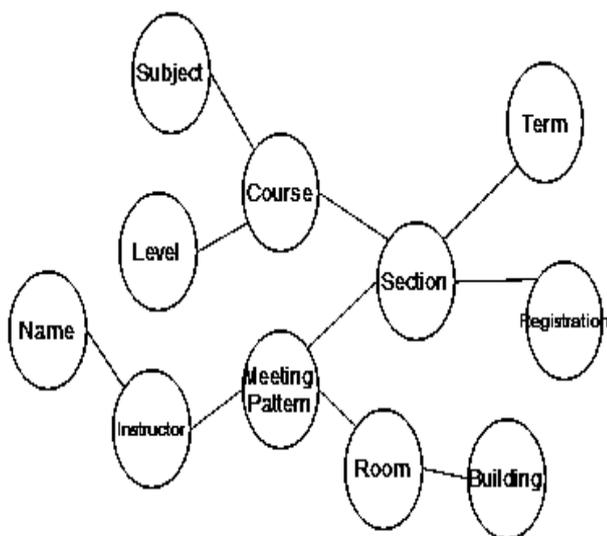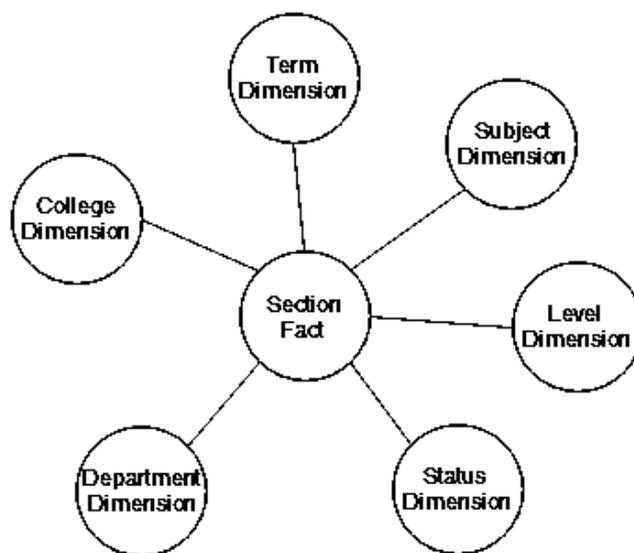

Figure 1: Traditional Approach for DM-1

Figure 2: Data Mats Approach for DM-1

6.2 The process

In our suggested Project (IC_DM), the first data marts (DM-1) is only available to the IC-Chief Office (IC-CO) Logging in requires a network user name and password (i.e. user name and password used to login to your computer).

(DM-1) is primarily used to provide section detail information by Student code. The data marts provide value for specific user and departmental needs and the data they hold cannot be used in other areas of the IC Departments. The second data mart (DM-2) is available to the workers.

6.3 The content

The data mart provides information about students, courses, student services, outcomes, faculty and staff. The emphasis of a data mart is to answer the questions of administrators, educators, and students. Figure (3) illustrate the data marts features. Because the data mart is aimed at supplying information to a wide variety of users, the easy-to-use interface and query explanations insure the data are easily accessed and processed.
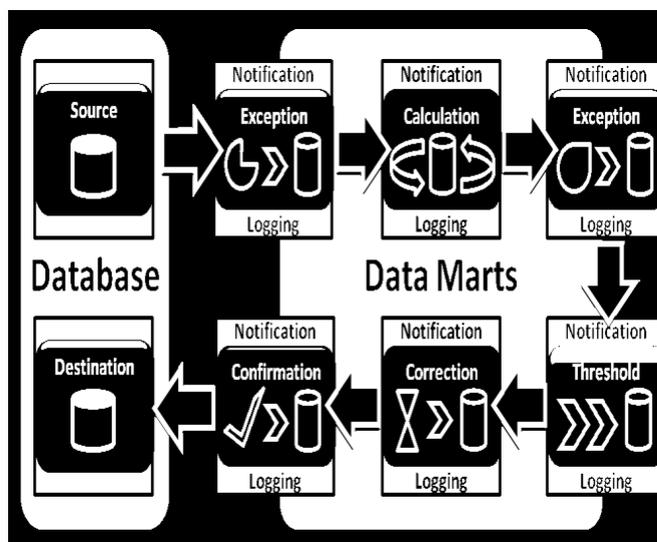

Figure 3: Data Mats features

In IC-DM depend on features like exceptions which are intended for data filtering, both prior and after data mart calculations. The system allows data excepting in certain period (holidays, emergency) by specified parameters, for example (Departments; Regions; Network elements). The second features are correction subsystem has designed for editing data mart calculation results. It is used by operating departments and allows editing values of calculated indicators; exclusion of erroneous data and manual adding of indicators. The subsystem is designed for confirmation of data mart calculation results by the operating department. It designed for verification of conformance of calculated values to the set threshold conditions. Data marts grouping subsystem is designed for the unification of data marts in order to simplify operations of the same type on them, for example: blocking, launch of execution, refreshment/recalculation. Figure (4) shows the Data marts steps starting with planning then developing and finally the deployment step. Other Subsystem is designed for distributed execution of preliminary calculations and aggregation of raw data on remote servers. Distributed execution this allows reducing central server load and transmitted data volume. And subsystem is designed to inform user about the events which took place. User can receive notifications in the many ways (E-mail notifications; SMS notifications and notifications within user interface).
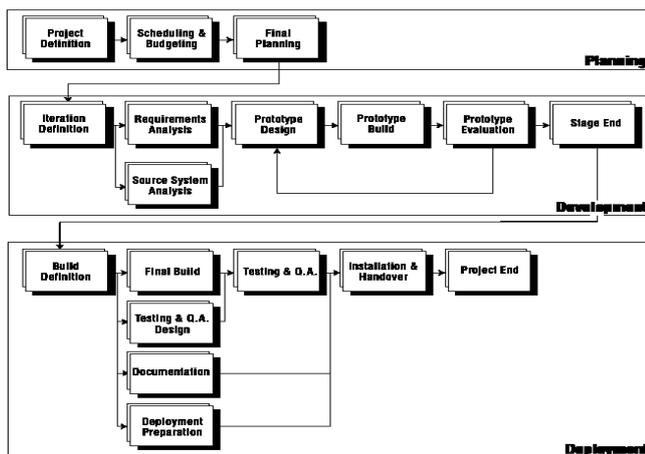

Figure 4: The data mart planning, developing and deployment

6.4 Data Mart Structure

The data we used for IC-DM can be divided into four categories: measures, dimensions[1], attributes, hierarchies. Figure (5) shows the four categories. The following list presents some DM-1 example:

 Star Schema: related fact and dimension tables for each subject area: (Student Enrollment, Course Enrollment, Registration, Degrees).

 Views: combine two or more fact and dimension tables (Enforce college/department security, Combine data from multiple stars).

Facts: subject area metrics (Section capacity, Actual enrollment, Generated hours).

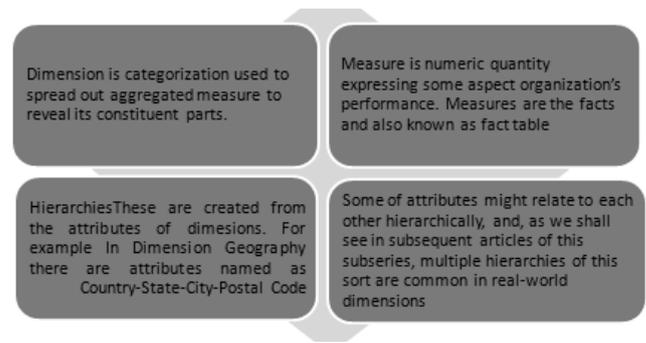**Dimensions**: subject area qualifiers (Term, Campus, College).

Figure 5: The data mart structure categories

In the practice there are some limitations that will be found in data marts which were: First, there is the extreme redundancy of information. Many data marts require some the same basic data, and when the data mart is not connected with a data warehouse this data must be restructured into each new data mart. Secondly, there is a significant level of inconsistency in the results of DSS from different data marts that are not integrated with a data warehouse. Third, these data marts must be built larger than if they were built upon a data warehouse, therefore, having higher cost. Finally, the redundancy and inconsistency will grow larger with each data mart built and eventually the cost of building data marts without a data warehouse will exceed the costs of using the classical approach to building data marts. Thus, the classical approach tends to be more widely practiced approach to building and structuring data marts.

7. Steps in Building A Data Mart

7.1 Identify Business Drivers
Before implementing any Data Mart, it is imperative to have a strong business driver for the project. A business driver answers the questions, what is the business need and Return on Investment (ROI) that fuels the need for the effort?

7.2 Identify Sponsorship
Data marts in general require approval at a much lower level of the corporate sponsorship chain than a Data Warehouse, for example at the manager level. Data Warehouses require a very high level of sponsorship, usually CEO and/or board level. On the other hand, the higher the sponsorship level is for a data mart, the more likely the project will become successful.

7.3 Conduct Readiness Assessment/Site Check
In order to have a working knowledge of the systems, data, and people who will be working on the data mart, it is essential to spend the time up front completing a readiness assessment. It is important to have people on the design team who understand the source system. At IC_MD , it was necessary to have someone who understood the billing system, so we could understand the calculations going into the data mart, and why.

---

[1] Dimensions include time (or date), customer, product, geography, lab type, campus, patient, promotions, gender (and other demographics), and so forth. Each dimension is associated with the facts / measures to which it relates via the linkages / joins between the table(s) housing the dimension (the dimension table) and the fact table.

7.4 Survey User Needs

The next vital step in creating a data mart, before the design phase, is to answer some basic questions concerning the customer that will drive our design. Since a data mart is based on a single subject area, it is important to clearly define a subject that is measurable.

7.5 Install database Data Mart Suite

The Data Mart Suite installation is very straightforward. It installs the data mart database, DMDB, as well as the enterprise manager database, EMDB. It sizes and tunes the database to a sample size data mart, which should be customized to the specific environment. It does not install the Data Mart Designer product by default and must be installed during a custom install.

7.6 Design Target System for Data Mart

Once all of the information is gathered, it is time to start designing the target database.

7.7 Data Mapping, Extraction, Transformation and Scrubbing

Once the star schema is designed, it is time to load the target system. Create a base view, a pictorial look at the database tables, for the target system.

7.8 Build Aggregation

Summary tables are a very important consideration for data mart design; they are built in order to reduce the amount of processing time to gather data that is frequently used from the detailed fact table.

7.9 Build Oracle Discoverer End User Layer

Once all of the Data Mart tables are built and registered, it is time to work with the Discoverer Administration product and create the End User Layer. The end user layer will be customized to fit the types of reports that the customer needs to see.

7.10 Create Discoverer Workbooks For Users

If there are users who have not used the Discoverer product before and are not familiar with creating workbooks, it is a good idea to create workbooks and have users run them, instead of having them create the reports from scratch.

7.11 Create Recurring Download Procedures

As a last development step, we scheduled download procedures for source to target data loads. It is essential to decide the frequency of downloads for the data mart [6].

8. Conclusion

The Data Mart is a low cost bundling tool which used properly leads to a successful Data Mart roll-out. This paper shows that data preparation is a key success factor for analytical projects. In order to be responsive to informatics requirements for planning, policy analysis, monitoring and evaluation in education sector, it is imperative that knowledge workers and decision makers have simple tools of analysis that can facilitate such processes. We found that, when we worked on the IC_DM suggested project methodology which is a data mart suite, could perform by

any educational institution. The Data Mart Suite is a powerful design and build product suite which has great promise for the future. As illustrated in literature review Data marts achieve a technique called dimensional data modeling, improve end user response time and enable users to analyze data trends. The Data mats diagnostic benefits make the difficulties in data access, consolidation, and cleansing in this this approach possible acceptance.

References

[1] Cassidy.T (2005), Education management information system in Latin America and the C C Caribbean: Lessons and challenges. Retrieved May 5, 2007.

[2] Connolly .T & Begg C.(2005), Database systems- A practical Approach to design, Implementation, n and Management. London: Pearson.

[3] Carrizo .L, Sauvageot.C and Bella.N (2003), Information tools for the preparation and monitoring of e education plans, UNESCO, Paris France. Retrieved April 3, 2008 from,http://unesdoc.unesco.org/images/0013/001323/132 306e.pdf

[4] Haiyan H. and Herstein J (2003), 'Education Management Information System (EMIS): Integrated D a data and Information Systems and Their Implications in Educational Management', Paper P r e presented at the Annual Conference of Comparative and International Education Society New r Orleans, LA(USA) Retrieved May23,2008from,http://www.infodev.org/en/Publication. 151.html.

[5] www.teradata.com

[6] Nepomnjashiy A. (2002), OLAP and Data Warehousing - Data in Jail. Retrieved March 19, 2008 from http://www.dbasupport.com/oracle/ora9i/olapjail.shtml

[7] http://dssresources.com/cases/redlandgenstar/html