University Student Information System Development Using Data Warehouse

بناء نظام معلومات طلبة الجامعة باستخدام مستودعات البيانات

Prepared By

Nour Alhyari

Supervised by

Dr .Muzhir AL_Ani

This Thesis is submitted in partial fulfillment of the requirements for the Master's degree of science in Computer Science.
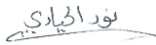
College of Computer Science and Informatics

Amman Arab University

(February, 2012)

I

I, the undersigned "Nour Hussein Alhyari " authorize hereby
Amman Arab University to provide copies of this thesis to libraries,
institutions and any other parties upon their request.

Name : Nour Hussein Alhiary

Signature:

نور الحياري

Date: 20 /2 / 2012

Resolution of the Examining Committee

This dissertation titled "Improving Of The University Educational System Based On Business Intelligence Technology" .Has been defended and approved on 25 / 1 / 2012.

Examining committee   signature

Dr. Rashfed A. Abass   R.A.S.Abbas

Dr. Akram Othman   e.m Othman

Dr Mahammed Nassar   M. Nassar

Dedication

To my parents for their support and endless effort.

To Belal for his encouragement and giving all the way .

To my dear sons Abed A lrahman and Omar.

To my sisters and brothers for all their support.

Acknowledgment

I 'm very grateful to God for all things I had done to finish my thesis ,and I would like to thank my supervisor Dr Muzhair AL-Ani for his guidance and help to complete this thesis, and I present a special thanks to Dr Alaa Alhamami and Dr Akram Almashaykhi for their encouragements . And I would like to thank the examining committee members.

List of Contents

List of Tables

List of Figures

List of Charts

List of Abbreviations

BI	Business Intelligence
CI	Competitive Intelligence
DSS	Decision Support Systems
DW	Data Warehouse
DSV	Data Source View
E-Discovery	Electronically Discovery
ER	Entity-Relation
ETL	Extract ,Transform, Load
GPA	Grade Point Average
KD	Knowledge Discovery
MAM	Motivation and Acceptance Model
MDX	Multi Dimensional Expressions
NN	Neural Network
OLAP	On Line Analytical Processing
OLTP	On Line Transactional Processing
SQL	Structured Query Language
TAM	Technology Acceptance Model

University Student Information System Development Using Data Warehouse

Abstract

The university environment is a Data-Driven institution and part of it is the student information system that had been collecting data regarding student absence, performance, and overall success. Data are also often collected concerning courses sections and courses categories, this data were collected using transactional databases, the huge amount of historically collected data were used for transactional purposes

but it never been a concern to analyze students data. This thesis developed a university student information system using a data warehouse for the purpose of analysis and forecasting. The developed system extracted the students data from two Jordanian private universities databases. In the extraction process the implemented system imported the data from its sources and transformed the data into unified representations, checked integrity constraints, sorted the data, and loaded the transformed data into a staging database which contains tables related to student academic achievement. After loading

the staging database the dimensional model is applied which uses the star schema design , the star schema of the university data warehouse consists of a single fact table and five dimensions that are Students dimension ,courses dimension, majors dimension, semesters dimension, and university dimension. Then the data warehouse dimensions and the fact table are loaded with students data, and then the results of the data warehouse were used for data cubing process in which the data can be viewed from multidimensional perspectives; finally the aggregated results from the cube are used for forecasting the students' academic achievement for the next semester.

Arabic summary

بناء نظام معلومات طلبة الجامعة باستخدام مستودعات البيانات

الملخص

الجامعة هي مؤسسة تعتمد على تدفق البيانات و جزء منها نظام معلومات الطلبة الذي يتم بوساطته جمع معلومات تتعلق بغياب الطلاب و علاماتهم و تحصيلهم الدراسي كما تُجمع المعلومات عن شعب المساقات الدراسية و أعدادها و أنواع المساقات الدراسية.و في الجامعات الأردنية كانت هذه المعلومات تجمع دائما لغايات التخزين في قواعد البيانات. و تشكل هذه القواعد مع مرور السنوات قواعد بيانات تحتوي على كمية هائلة من البيانات التاريخية التي يمكن الإفادة منها في التحليل و التنبؤ الإحصائي .

هذه الرسالة أنشأت مستودع بيانات لنظام معلومات الطلبة في الجامعة و تم إستخراج البيانات التي تمت دراستها من قواعد البيانات لنظام معلومات الطلبة في جامعتين من الجامعات الأردنية الخاصة .و بعد استخراج البيانات تم تحويلها في عملية توحيد تمثيل البيانات إلى بيانات ذات تمثيل متوافق و من ثم تحميلها في قاعدة بيانات مؤقتة ،ثم صُمم مستودع البيانات اعتمادا على النظام البُعدي الذي استخدم نموذج النجمة في تصميم مستودع البيانات الذي فيه تم الغاء جميع العلاقات الموجودة بين جداول قاعدة البيانات المؤقتة و تم إنشاء جدول لكل بُعد من أبعاد مستودع بيانات نظام معلومات الطلبة و هذه الأبعاد هي بُعد الطالب وبُعد المساقات الدراسية و بُعد الفصول الدراسية الذي يمثل بُعد الزمن و بُعد التخصصات الدراسية و بُعد الجامعة الذي يحدد مصدر البيانات من أي جامعة ، وتم

إنشاء جدول حقيقة يرتبط مع الأبعاد حيث يحتوي جدول الحقيقة على مجموعة من المفاتيح كل مفتاح يرتبط بمفتاح رئيس لبعد من الأبعاد و يحتوي جدول الحقيقة على القيم المراد قياسها و هي المعدل التراكمي للطالب و الشعب الدراسية للطالب و عدد مواد التخصص المسجلة في كل فصل دراسي للطالب .و من ثم تمت تعبئة أبعاد مستودع البيانات ثم تعبئة جدول الحقيقة ثم تم إنشاء معاينة لعرض بيانات المصدر تم ربطها مع المستودع الذي تم بناءه و في المعاينة تم تحديد الأبعاد التي دخلت في مكعب البيانات الذي فيه تم تحديد القيم المراد قياسها في المكعب لتحليلها و التي تقيس إنجازات الطلبة في الجامعتين في كل تخصص.و بعد إنشاء مكعب البيانات تم عرض البيانات فيه باستخدام تراكيب برمجية خاصة بأبعاد المكعب و التي تعرض البيانات الموجودة في المكعب من خلال كل بُعد و على أكثر من مستوى ثم تم استخدام نتائج المكعب للتنبؤ بالقيم المستقبلية للقيم التي تم قياسها باستخدام التحليل الإحصائي.

Chapter 1
Introduction

1.1. Background

Data Warehousing is one of the techniques that are used in Business Intelligence approach for business development .The data warehouse is related strongly with Business Intelligence because of its role in data storage and analysis. Its first definition was developed by Bill Inmon in 1980 he defined it as "a subject-oriented, integrated time variant and non volatile collection of data used in strategic decision making "[1].

Another definition for data warehouse is " a relational database that is designed for query and analysis rather than for transaction processing "[2].

Because of the importance of business intelligence as an approach based on data warehousing the business intelligence is introduced in this section.

The Business Intelligence (BI) term came into sight in 1989 and it gained an increased interest in the last 20 years because of its positive effect on organizational performance [3].

In 2003 (*Moss and Atre*) indicated that BI is not a product or a

1

system it is an architecture and a group of incorporated prepared applications of decision support processes, and it includes the databases and data warehouses that provide direct access to business data [4].

In 2004 *(Marshal et al*) defined BI as the gaining, evaluation and utilization of information associated to a business [5].

In 2009 (Chasalow) defined BI as "business information and business analyses within the context of key business processes that lead to decisions and actions and that result in improved business performance" [6].

1.2. University Student Information System

The University Student Information System (SIS) stores all the data that is related to the students' information which include: personal information, students' academic activities in the university, the courses of the students, and the students' grades. The student information system data are stored using transactional databases these databases are not used for analysis and forecasting, for this reason this thesis developed a data warehouse system to store the student data for analysis and forecasting. The developed system used two data sources that are two Jordanian private universities;

2

the thesis refers to these universities as *university1 and university 2* because the universities administration requested not to state their true names for purpose of privacy.

1.3. Statement of the Problem

This thesis developed a data warehouse system used to store the historical records of students data, by extracting the required data from two student information systems from two private Jordanian universities, The extracted data were then transformed through data integration to be loaded into a staging database and then the denormalization process for this staging database was performed, The dimensional model of the data warehouse was designed and the data warehouse were loaded. The resulted data from the data warehouse were used for data cubing and finally the resulted data from the data cubing are used in forecasting the values of three factors that are student cumulative GPA, sections frequency and major courses count that the students register in each semester.

1.4. Thesis Goals

The thesis goals are the following:

❖ Develop a data warehouse system using extract, transform and load processes from the student information systems databases of two private Jordanian universities.

❖ Study a group of factors that measure the universities performance including frequency of courses sections, the students Grade Point Average (GPA), and number of major courses that the students register in each semester.

1.5. Literature Review

Universities as many organizations adopted data warehousing in their systems because of its advantage of holding historical data that is used for analysis and data trending. A brief history about the use of Data warehousing in higher education   is listed here:

In 2011 (*Manjunath et al)* designed a university data warehouse using open source tools in order to reduce the cost of building a data warehouse system for the university environment[7].

In 2011 *(Singh et a)l* studied building a university data warehouse using a qualitative multi-site case study method in which the warehouse was designed based on enterprise-wide information needs [8].

4

In 2010 *(Piedade et al)* developed a student data warehouse system to monitor the students academic activities in the university system[9].

In 2010 *Rebbapragada et al* build a data warehouse that hold student performances records and used data mining techniques to predict the trends of university applicants in order to indicate admission offers for the applicants [10].

In 2008 *Mathews et al* implemented a data warehousing and business intelligence solution at McMaster University, the goal of the implementation

was to connect research funding and research financial databases. [11]

In 2008 *(Staškevičiūt, et al)* stated that for many years universities were considered as institutions that only provide services for the community. But the globalization processes force universities to move from services profile to market profile. So universities have to make extreme changes in their educational and administrative systems and overcome old traditional and centralized management techniques by adoption of Business Intelligence systems. The authors suggested that the intelligent university should be open to the society to construct new knowledge society [12].

5

In 2007 *(Raton)* discussed the benefits of the data warehouse solution that Florida Atlantic University applied for budgeting and human resources systems, the data warehouse enabled the university administration to turn the data into valuable knowledge that can be used for strategic planning and decision making processes [13].

In 2006 *Ballard* stated that Saint Louis University developed a data warehouse used for reporting, decision support, forecasting/trending, and data mining, the data warehouse loaded from different data sources in the university [14].

In 2004 (*Nelson)* stated that at De Montfort university   a static data warehouse were build for university students achievement reports ,the reports are downloaded into the data warehouse at the same time each year, enabling a comparison across sessions from these points onwards[15].

In 2001 (Bielec) researched the management techniques for data warehousing in higher education that were used in Drexel University while (Blitzblau and *Hanson*) discussed the project of building a data warehouse in Georgetown University in the same year [16,17].

This thesis developed a data warehouse for student information system using ETL processes by extracting the data from two private Jordanian universities this thesis is the first thesis in Jordan that focuses on the analysis of important factors that measure the university student information system performance.

The importance of data warehousing as a main component of business intelligence makes it important to list a brief literature of business intelligence in this section. Business intelligence stands for the decision support systems used in any business enterprise; these systems use the collected data from external and internal resources of the enterprise to make decisions and future plans for the enterprise to get more profits ,There are many published works related to BI some of them are listed in this section:

In 2010 *(Ivancenco, et al)* studied the relation between Information management with the performance management, and presented the main BI solutions that exist in Romania. They concluded that implementing BI systems in organizations facilitates management of data from external and internal resources by the use of BI instruments such as dashboards and different tools supported by BI software providers [18].

7

In 2010(*Işık*) examined the effect of BI capabilities on BI success in the

Case of different decision environments. The research results suggested that

All technological capabilities and the organizational capabilities had a role

In BI success, The results also indicated that the decision environment had a

Significant effect on quantitative data quality [19].

In 2010 (*Brannon*) discussed how BI tools can be used in E_discovery which refers to discovery of information stored electronically. The research explained the BI systems functionality and the importance of BI technologies in the process of e-discovery [20].

In 2009 (*Chasalow)* tested the characteristics of BI users and their organizations and presented models for those competences by presenting different cases for the study. He found that the skills of users of BI systems can be categorized into three types that are cognitive ability, non-cognitive ability, and knowledge. These skills can impact the ability of an organization in adoption of BI systems,

8

This study stated that the management approach and resources representation can affect the success of organizational adoption of BI systems[6].

In 2009 (*Janke*) identified the factors that determine the user acceptance of BI in commercial real estate by studying the Technology Acceptance Model (TAM) and modified versions of TAM that is Motivation and Acceptance Model (MAM) .The researcher reported that commercial real estate companies should invite BI systems user to interact with each other to share ideas about good practices for BI technologies in order to achieve high performance of BI systems [21].

In 2008 (*Peng*) concentrated on " the development of different data mining techniques to extract actionable patterns from various types of log data in computing and BI systems ". This research studied comprehensive cases on real applications. Peng improved new data mining techniques to extract actionable patterns that are used by domain experts to deploy these patterns in management of historical data and BI systems [22].

(*Konstantin, 2008*) investigated information technology intelligence which is a sub division of business intelligence and competitive

intelligence. He discovered that the information technology intelligence process includes assessment of data from external and internal resources of an enterprise. He proved by evidence that the final product of information technology intelligence helps consumers to take better decisions in any industrial organization [23].

In 2007 (*Celina et al*) described the process of development of BI systems considering the specifications of a BI system, they proposed a methodology of how BI can be constructed and implemented in an organization. The study stated that BI creation and BI consumption are two major stages in building and implementing BI in organizations [24].

In 2005 (*Robert K*) Presented the implementation of strategies commonly used in decision support systems to promote success to enable supply chain management. The implementation of these strategies is considered a demand for companies to compete in the globalized economy. The aim of this study was to indicate which strategies achieve success in decision support systems that is designed for supply chain management, the study reported that the users of decision support systems should have powerful analytical skills, they should have the ability to analyze, explain and deal with complex data [25].

10

In 2004 *B. Marshal et al* proposed a BI tool that is Electronic Business Intelligence portal (EBIZ port) which can be used for gathering information from different resources and analyzing them. This BI tool was made to deal with the huge amounts of data on the internet. This tool extracts , analyzes and uses data to extract intelligent information that benefits users. The EBIZ port helps users mainly in analyzing data by dynamic visualization of news of information technology industry [5].

In 2003 *Jermol, et al* discussed managing and implementing BI in a virtual enterprise; The authors chosen the European Virtual Enterprise (soleunet) as a case study. They found that Managing and implementing BI depend on appropriate business, work procedures, cognitive authority in collaborative work management, and network intelligence in the virtual enterprise. The authors mentioned that such a virtual enterprise can succeed in the market if it is supported by knowledge management. This knowledge management requires a systematic approach in collecting, analyzing and deploying and this can be achieved by the use of the suitable methods of knowledge management. The most important method is to build the' knowledge map of partnership's expertise '[26].

11

1.6. Thesis Contribution

The thesis developed a data warehouse for university students information system extracted the data from the databases of two Jordanian private universities. This thesis is the first study in Jordan that developed a data warehouse for university students' historical data by which new perspectives of the students' information system were studied by analyzing three factors that measure students performance using data cubing and forecasting techniques that resulted in new information which can help decision makers in the universities administration.

1.7 Thesis Organization

- *Chapter 2* will study data warehousing with explanation of its architecture and characteristics.
- *Chapter 3* is concerned with the developed system; it includes the system data warehouse design and the data cube design.
- *Chapter 4* includes the results of the data cubing and forecasting of the results.
- *Chapter 5* includes the conclusion and future work.

Chapter 2
Data Warehousing

2.1.Introduction

The data warehousing is a major component of business intelligence systems which are a combination of technologies and tools that are used to collect and analyze different information sources in an organization to achieve better business performance [27].

In 2001 *Lokken* identified business intelligence tools as a collection of basic tools that is responsible for obtaining data, cleaning it up, transforming it , restructuring it, and analyzing it using data warehousing and a group of business intelligence technologies to make better business decisions[28].

2.2 Business Intelligence Technologies

Business Intelligence technologies include data warehouses, data marts, On Line Analytical Processing (OLAP) servers, decision support systems, and extract transform load (ETL) tools query reporting and analysis tools , on-line analysis and exploration, visualization, data modeling and planning data mining tools, Portals, dashboards, and scorecards [28].

13

The main BI technologies that are included in this thesis are: data warehousing, OLAP, and data visualization.

2.3 Data Warehouse

Data warehouse forms a major component in BI systems . Its architecture and characteristics provide BI systems users with data views and accessibility techniques that are not available in classical database systems.

The first data warehouse definition was developed by Bill Inmon in 1980 he define it as "a subject-oriented, integrated time variant and non volatile collection of data used in strategic decision making "[1].

Other definition for data warehouse is "a relational database that is designed for query and analysis rather than for transaction processing "[2].

2.3.1 Data Warehouse Architecture

Data warehouse architecture is an approach that demonstrates the data structure in an organization.This approach includes the data transformation, analysis and data views for the users in the organization system [29]. The simple data warehouse architecture is viewed in figure 2.1.

14
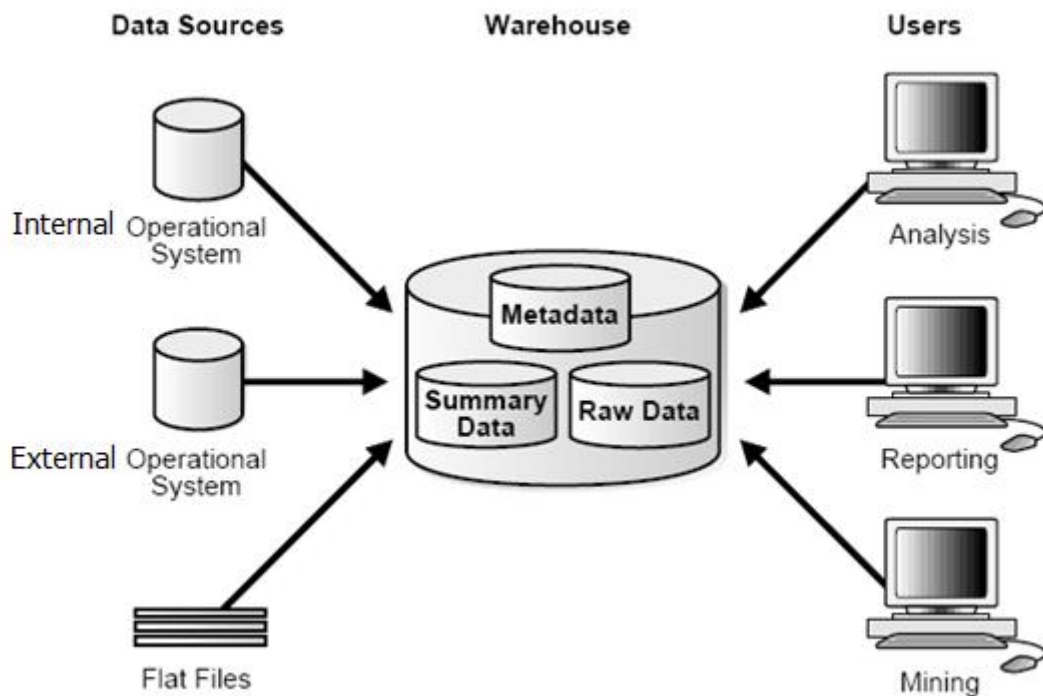
Figure 2.1 Simple Data Warehouse Architecture [30].

The data warehouse architecture consists of a group of layers that are:

- Operational Database / External Database Layer :

Operational systems handle data to provide a significant operational need by using structured methods for a relatively small number of distinct business transactions [31, 32].

- Information Access Layer :

The information access layer signifies the tools that the system uses on daily bases to view reports and other output forms.

- Data Access Layer: the data access layer permits the information access layer to deal with the operational layer. The structured query language (SQL) is the standard language used for data exchange.

- Data Directory (Metadata) Layer: the metadata layer includes data that describes the data in the data warehouse [33, 34].

- Process Management Layer: the process management layer allocates different jobs and processes according to a pre-defined schedule.

- Application Messaging Layer: the application messaging layer transfers, gathers, and supplies data to locations that require the data on the network.

- Data Warehouse Layer: the basic data in the data warehouse is located in the data warehouse layer which is used to create the logical views of the data staging layer.

- Data Staging Layer: data staging includes all of the processes required to manipulate data warehouse data and information access data from the whole system data bases (internal and external) [35,36].

16

2.3.2 Data Warehouse Characteristics

Bill Inmon, considered as the "father of the data warehouse", defined the data warehouse characteristics as follows [2,30] :

A) Subject-oriented

The goal of data warehouse design is analyzing data and creating useful information from this data, for this goal the databases of the data warehouse are planned to be build in such a manner that all data elements that are related to the same objects are put related together. For example to know more about students GPA data, a data warehouse can be built that concentrates on students. Using this warehouse, questions like *"which student has the highest GPA* can be answered [29].

B) Integrated

The data from different sources must be put in the same format by the data warehouse. The data warehouse must not have any data formatting problems such as naming conflicts [29,30] .

C) Time-variant

In the data warehouse all data related to an event in the past or at the mean time is required to achieve the data warehouse goal of business analysis and decision making [31].

D) Non-volatile

The data that was entered to the data warehouse must not modified because the

main goal of a data warehouse is analyzing previous events in the running

system of an organization [30].

2.3.3 Data Warehousing Methodologies

The methodologies of data warehousing include the following steps:

Business requirements analysis, data design, architecture design, implementation, and deployment [37,38].

- The business requirements analysis step includes specific strategies for choosing the business requirements such as interviews with the intended users of the data warehouse, and brain storming sessions.

- The data design step combines of data modeling and normalization, the most important methods for data modeling in data warehousing are Entity-Relational(ER) and Dimensional modeling. The ER method consists of the steps that are used to design online transactional processing databases which include designing ER conceptual diagram, converting ER diagram into a relational schema, then the normalization process is done for the relational schema.

- A dimensional model consists of a group of dimension tables and the fact table which is the most important table in the model [38].

  The dimensions express the main objects in the data for example student, teacher, department and faculty. Members represent single data values for example the faculty of science, attributes are "properties of a member"[32].

  The fact table consists of a group of attributes that in most cases have numeric

  Values.The key of the fact table is multi-attribute, while the primary key of dimension tables is single attribute (that can be replaced) in each dimension table, the primary key matches one attribute in the fact table. The dimensional model can be represented in a special type of schema that is called *star join schema* [31].

- The data warehouse architecture contains a group of planned activities such as communication, maintenance and reuse. The data warehouse architecture design can be divided into two categories that are enterprise wide data warehouse design and data mart design [37].

  T. Ariyachandra et al(2008) studied five data warehouse architectures that are independent data marts, bus, hub and spoke, centralized, and federated:

19

Independent data marts are the first place in which the data necessary for the process of decision support is stored. These marts are designed to store data separately from other data stores in the system.

The bus architecture has distinct data marts that are required for business processes in which all the data marts are designed using the same dimensions.

The *hub and spoke architecture* starts with the step of indicating requirements and analyzing these requirements at the organization level.

The *centralized architecture* is the same as the hub and spoke but it has no dependent data marts [39].

The federated data ware house architecture supports the environments in which the data that is used for the process of decision making is in different parts of the system. In this architecture, the data is integrated using specific techniques such as shared keys.

Data warehouse implementation consists of data sourcing, data staging, and build up applications that are decision support oriented.

The data warehouse implementation steps rely on the following two main considerations: data quality management and metadata management [40].

The data warehouse deployment step concentrates on the integration of the designed data warehouse in the required system. Data warehouse modification and maintenance are to satisfy user needs.

2.4 On Line Analytical Processing

Online Analytical Processing (OLAP) is a database model that provides manipulation and processing for multidimensional data. The OLAP executes different functions on stored data of the data warehouse to view the data in multi-representations to simplify the process of decision making. An example of multidimensional representations of databases is the star schema which is used in the implemented system. The OLAP derives its data from different On Line Transactional Processing (OLTP) data sources using the Extract, Transform and Load (ETL) tools .The OLTP performs a specific operations on the data base system that supports the current state of the system. The OLTP stores and modifies the day-to-day transactions [41,42].

2.5 Data Cubes

The data cube is a multidimensional matrix that is used by OLAP queries for data storage .It can be considered as extensions of 2-D tables of the DW tables. The cube is so valuable because the data can be indexed in different ways using the cube dimensions.

The OLAP queries limit the data by arrangements of dimensions and then perform statistical functions on particular attributes of the data. These statistical functions called data aggregation .The OLAP systems store the aggregated data in multidimensional databases called cubes. [43, 44, 45]

The OLAP cubes can be visualized as expansion of a two – dimensional array . The OLAP queries manipulate the data stored in the data cube and visualize it into different hierarchies. The main types of OLAP queries are:

• Roll_up: The Roll_up query produces an aggregation operation through the cube dimensions.

• Drill-down: The Drill_down query enables the system users to mine down through a specific dimension.

• Slice and Dice: The slice and dice produce selection and projection through the dimensions of the data cube. The slice query produces a sub cube by selection over one cube dimension.

- The dice query produces a sub cube over the three dimensions of the data cube [48].

- Pivot: The pivot query produces a hierarchy of the data cube that include transition or swapping of the cube dimensions.

2.6 Data Visualization

Data visualization is the process of viewing data and presenting it in different visual computerized forms. The data visualization process include three major steps that are data acquisition, data selection and data presentation. The data acquisition includes exchange process between the system and decision makers to explore the system after this step.The needed data are selected then it is converted to a visual form. The data visualization could be static or interactive. In interactive visualization system users can choose the representation method of data while the static data visualization has a restricted method for presenting data [51,52].

Chapter 3
Development

3.1 Introduction

This chapter provides a description of the development of a data warehouse solution for student academic achievement data; the developed system is a solution for applying data warehousing tools in the university organization.

The built system studied the frequency of the courses sections, students GPA, and studied the of number of major courses that the students register in each semester to measure the university performance .Previous works had studied the effect of student absence on their academic achievements as a main factor but the factors in this thesis are studied for the first time.

The university environment is a Data-Driven (DD) institution that had for some time been collecting mountains of data regarding student absence, performance, and overall success. Data are also often collected concerning courses sections and courses categories. The collected data were used in transactional databases only but it has never been a concern to analyze student behavior in choosing their sections and their courses categories. The implemented system used the data stored in the data warehouse for data cubing process

24

in which the data can be viewed from multidimensional perspectives; finally the aggregated results from the cube are used for forecasting the students' academic achievement for the next semester.

The system architecture is shown in figure 3.1.
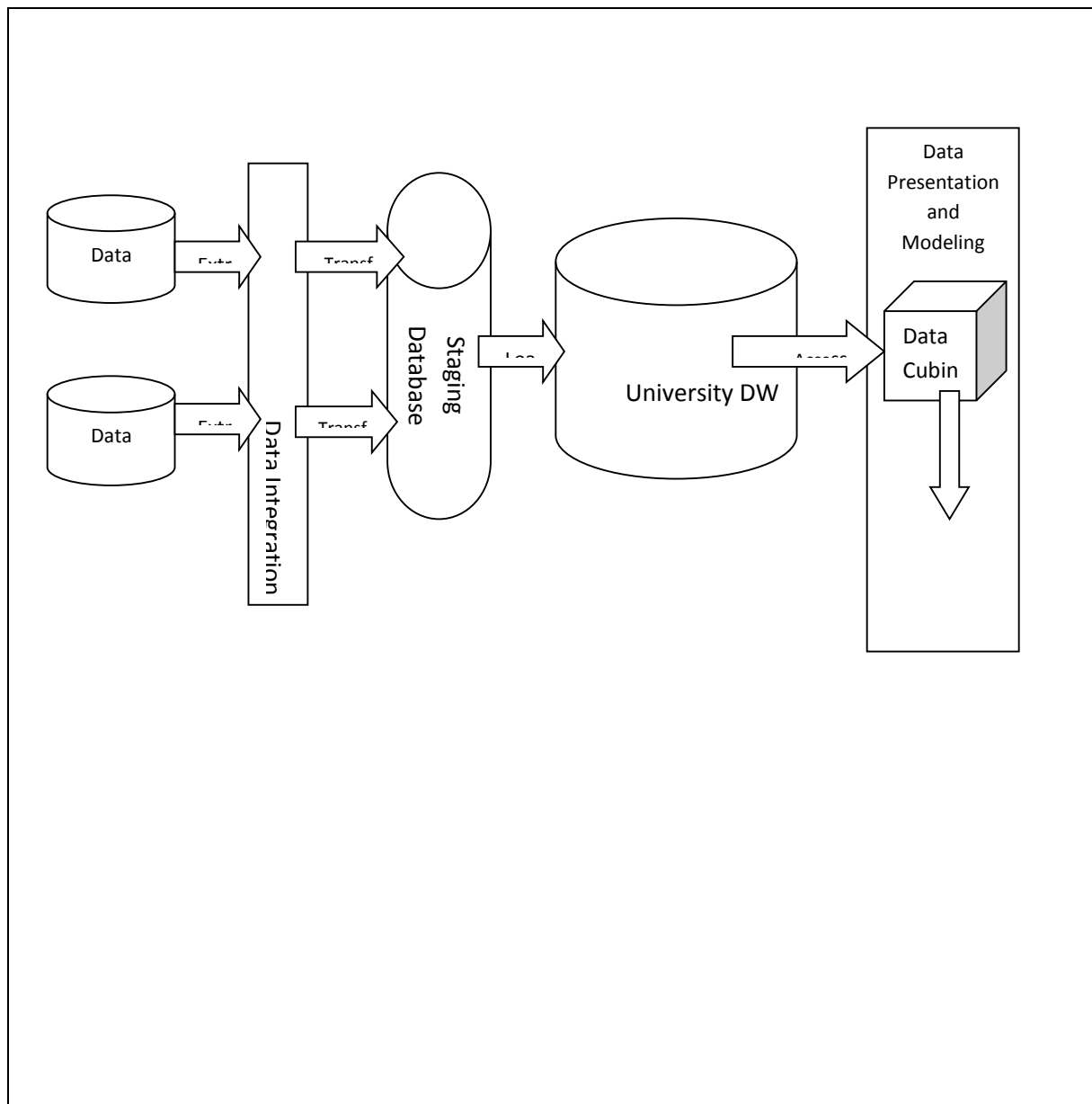
Figure 3.1 The Implemented System Architecture.

3.2 Extraction of Data

The developed system collected the required data from two Jordanian universities. Each university database contains a large set of tables and a group of systems that include financial system, library system, human resources system and student information system (SIS). The developed system is concerned with the data stored in the SIS. The number of records from each university is listed in table 3.1.

Table 3.1 Number of Records from Data Sources.

Data Source	Number of Records
University 1	41801
University 2	20498
Total Number of Records	62399

3.3 The Data Warehouse Design

The developed system presented the processes of converting university transactional database into data warehouse for the purpose of Data cubing, analyzing and forecasting. The constructed data warehouse (DW) is built by extracting data from source data

26

bases, verifying and transforming them, and then loading a new data base which becomes the developed system DW. The processes of extraction, transformation and loading, are done periodically after the academic semester is finished, provides historical layers of data, which are snapshots of the academic achievement of the students from different sources.

3.4 Transforming the Data.

In the extraction process the implemented system imported the data from its sources. In the transformation process the data into unified representations, checked integrity constraints, sorted the data, and loaded the transformed data into a staging database which contains tables related to student academic achievement .The tables are listed in table 3.2.

Table 3.2 The Temporary Database Tables Description.

Table name	Description
SIS _Academic Records	Holds students information about semester GPA and cumulative GPA.

27

SIS_Courses Categories	Holds the records that indicate the course type.
SIS_Course_Status	Indicate if any course is in the semester schedule.
SIS_Courses	Holds all information about all courses.
SIS_Depts.	University departments' records.
SIS_Faculties	University faculties records.
SIS_Majors	University majors' records.
SIS_Section Times	Indicate the courses sections time in each semester.
SIS_Semesters	Indicate the academic semester's information.
SIS_Student Activities	Holds the records for student registered courses in each semester.

SIS_Student cCourses	Holds the grades for each student in passed courses.
SIS_Student Status	Indicate if any student is registered in a certain semester.
SIS_Students	Holds personal information for students.
SIS_Study Semester	Indicate the semester codes for the academic year.
SIS_Times	Records for time sections.
SIS_Time Table	Give codes for section times.

3.5 Loading the Data Warehouse

After loading the staging database, the dimensional model is applied. This model used the star schema design as in figure 3. The star schema of the university DW consists of a single fact table and four dimensions. The Dimensions of the DW are:

- Students dimension.

- Courses dimension.

- Majors dimension.

- Semesters dimension.

And another dimension will be added in the data source view which is the University dimension.
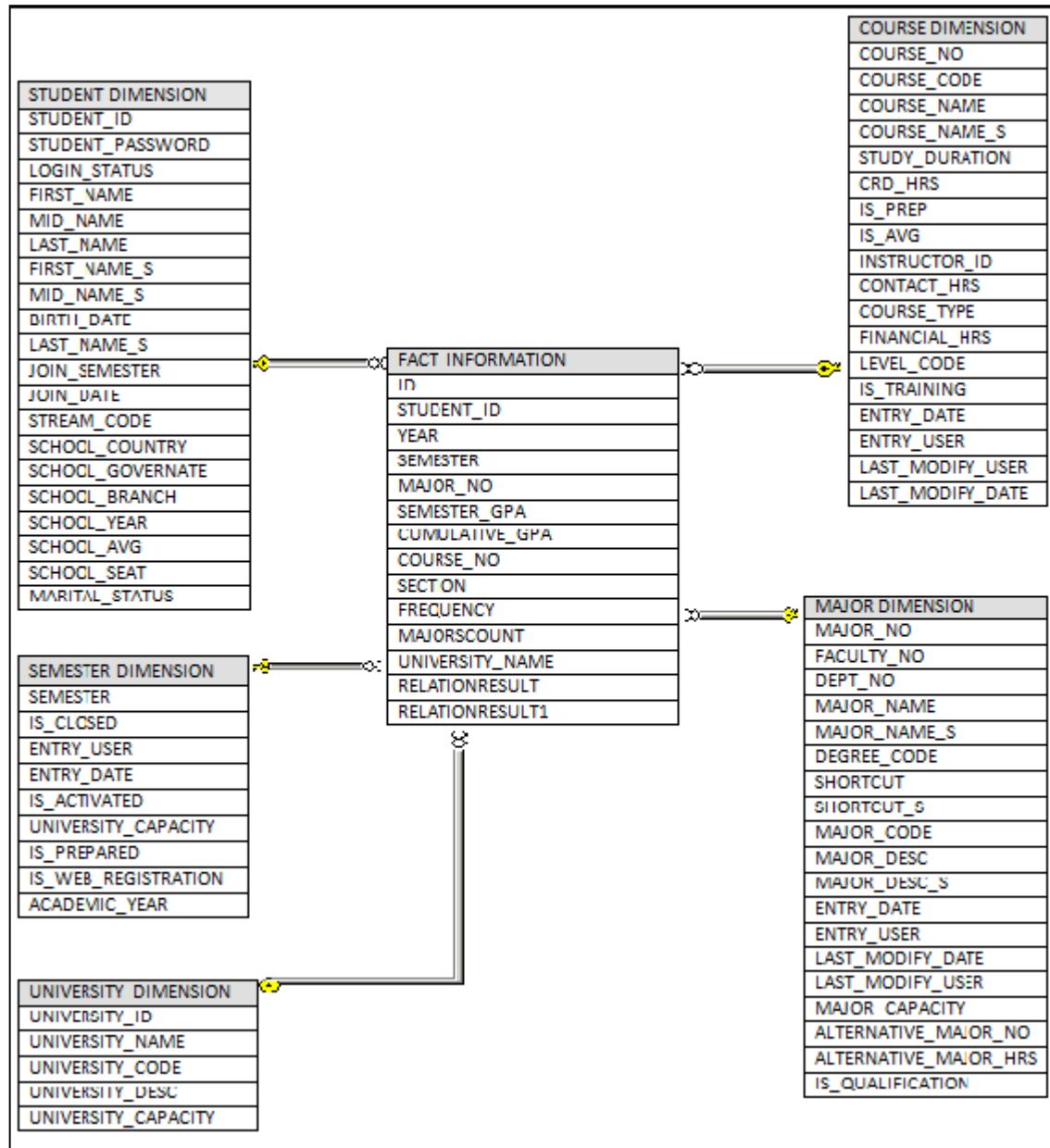
Figure 3.2 Star schema of the university DW.

The student data was designed to hold all the data related to the student , also the course dimension was designed to hold all the data related to the courses and the major dimension holds all the data related to majors while the semesters dimension holds all the data that are related to academic semesters. The university dimension holds the data that indicates students' records from each university. The dimensions attributes that are inserted into the data warehouse are not all the listed attributes in the star schema design, only the attributes that are related for the studied factors are inserted. The student dimension is designed to hold the student's personal information and it is related only to the fact table. The Student dimension data was loaded using the code in figure 3.3. The student attributes that are needed are extracted from the staging database and these attributes are selected from the staging database and then inserted into the data warehouse. The inserted attributes are: (Student_ID, Student Name, Join Date, School Year, and School Average).

```
use Test

Insert Into [UOP_DW].[dbo].[Students]

       ([STUDENT_ID]

       ,[STUDENT_NAME]

       ,[JOINDATE]

       ,[SCHOOLYEAR]

       ,[SCHOOL_AVG])

Select

STUDENT_ID,

IsNull(FIRST_NAME_S,'') + ' ' + IsNull(MID_NAME_S, '') + ' ' +

IsNull(LAST_NAME_S, '') AS 'STD_NAME',

IsNull(JOIN_DATE, '1/1/1900'),

IsNull(SCHOOL_YEAR, '1900'),

IsNull(SCHOOL_AVG, 0)

from SIS_STUDENTS
```

Figure 3.3 Insert the Student Dimension in the Data Warehouse

The courses dimension is designed to hold the course's information and it is related only to the fact table. The courses dimension data was loaded using the code in figure 3.3. The courses attributes that are needed are extracted from the staging database by selection then they are inserted into the data warehouse. The courses attributes are: (Course Number, and Course Name).

```
Use Test

Insert Into [UOP_DW].[dbo].[Courses]

(COURSE_NO,

 COURSE_NAME)

Select

COURSE_NO,

IsNull(COURSE_NAME_S, '')

From SIS_COURSES
```

Figure 3.4 Insert the Courses Dimension in the Data Warehouse.

The major dimension is designed to hold all majors information and it is related only to the fact table. The majors dimension data was loaded using the code in figure 3.5 the majors' attributes that are needed are extracted from the staging database by selection, then they are inserted into the data warehouse. The majors attributes are: (Major Number, and Major Name, and Major Description).

33

```
Use Test

   Insert Into [UOP_DW].[dbo].[Majors]

   (MAJOR_NO,

    MAJOR_NAME,

    MAJOR_DESCRIPTION)
Select

   MAJOR_NO,

   MAJOR_NAME_S,

   MAJOR_DESC_S

   From SIS_MAJORS
```

Figure 3.5 Insert the Majors Dimension into the Data Warehouse.

The university dimension holds records that indicate the data source and the semesters dimension holds the values of academic years and semesters. After loading all the dimension tables, the fact table is loaded using the data from the dimensions. The fact table of the data warehouse is Fact Information table.

The fact table attributes represent surrogate key for each dimension table and it contains two aggregated values. The first is section frequency value for a certain student in a certain academic semester. The other one is count of major courses for the same student in the same academic semester. The values of the attributes in the fact table firs are inserted using the code in figure 3.6.

The keys from other tables are inserted then the sections frequency and major courses count are calculated using two separate functions called in the insertion code of the fact table.

```
Use Test
Insert Into [UOP_DW].[dbo].[FactInformation]
        ([STUDENT_ID]
        ,[YEAR]
        ,[SEMESTER]
        ,[MAJOR_NO]
        ,[SEMESTER_GPA]
        ,[CUMULATIVE_GPA]
        ,[COURSE_NO]
        ,[SECTION]
        ,[FREQUENCY]
        ,[MAJORSCOUNT]
        ,[UNIVERSITY_NAME])
Select   STUDENT_ID,
       Left(SEMESTER, 4) As 'Year',
       RIGHT(SEMESTER, 1) As 'Semester',
        MAJOR_NO,
        Sum(IsNull(SEMESTER_GPA, 0)) As SEMESTER_GPA,
        Sum(IsNull(CUM_GPA, 0)) As CUM_GPA,
0 As 'COURSE_NO',
0 As 'SECTION',
```

```
IsNull(dbo.fn_CourseFrequency(SEMESTER, STUDENT_ID), 0) As 'Frequency',

IsNull(dbo.fn_MajorsCount(SEMESTER, STUDENT_ID), 0) As 'MajorsCount',

' University 1' As 'UNIVERSITY_NAME'

From SIS_ACADEMIC_RECORDS

Where Cast(Left(SEMESTER, 4) as numeric) > 2005

And Cast(LEFT(Student_ID, 4) As Numeric) <= 2010

And STATUS_CODE = 1

Group By  STUDENT_ID,

     SEMESTER,

     MAJOR_NO
```

Figure 3.6 Insert the Fact Table attributes.

The sections frequency was calculated using the function in figure 3.7; the section frequency is calculated from the staging database by counting the number of records for a certain course for each student. The section frequency reflects the number of registered students in different courses sections.

```
USE [Test]   GO

SET ANSI_NULLS ON   GO

SET QUOTED_IDENTIFIER ON GO

ALTER FUNCTION [dbo].[fn_CourseFrequency]

(      @SEMESTER As Numeric,

     @STUDENT_ID As Numeric

     --@COURSE_NO As Numeric,

     --@SECTION As Numeric   )
```

```
RETURNS Numeric

AS

BEGIN

        DECLARE @Frequency Numeric

/*      Select @Frequency = COUNT(*)

        From  SIS_SECTION_TIMES

        Group  By  SEMESTER--,  COURSE_NO,  SECTION,  ACTIVITY_CODE,

TIME_CODE  HAVING  SEMESTER = @SEMESTER  And Stud  */

        Select @Frequency = COUNT(*)

        From SIS_STUDENT_COURSES

        Where STUDENT_ID = @STUDENT_ID

        And SEMESTER = @SEMESTER

        --And COURSE_NO = @COURSE_NO

        --And SECTION = @SECTION

        RETURN @Frequency

END
```

Figure 3.7 Calculating the Sections Frequency.

The count of major courses for each student in the same academic

semester was calculated using the function code in figure 3.8. The

major courses count is calculated by counting number of records for

the courses that have the value of major as its category.

```
USE [Test]    GO

SET ANSI_NULLS ON    GO

SET QUOTED_IDENTIFIER ON   GO

ALTER FUNCTION [dbo].[fn_MajorsCount]

(   @SEMESTER As Numeric,

       @STUDENT_ID As Numeric

--       @COURSE_NO As Numeric )

RETURNS Numeric

BEGIN    DECLARE @MajorsCount Numeric

      Select @MajorsCount = COUNT(*)

  From SIS_STUDENT_COURSES

      Where CATEGORY_CODE = 3 -- Specialization Requirement

      And SEMESTER = @SEMESTER

      And STUDENT_ID = @STUDENT_ID

      --And COURSE_NO = @COURSE_NO

Group By   STUDENT_ID, SEMESTER, COURSE_NO

      RETURN @MajorsCount

END
```

Figure 3.8 Calculating the Major Courses Count.

3.6 Creating Data Source View

A Data Source View (DSV) holds all the metadata from the available

data sources. It is useful for filtering the tables, views, and queries

needed for the data cubing. When tables are brought into a DSV,

the relationships that have been established on the DW will be

carried over and will be visible in the DSV. The DSV was created
first by indicating the university DW as its data source then the tables
of the DSV were selected as in figure 3.9.By moving the tables to
the DSV, the relations of the DW are moved to the DSV.
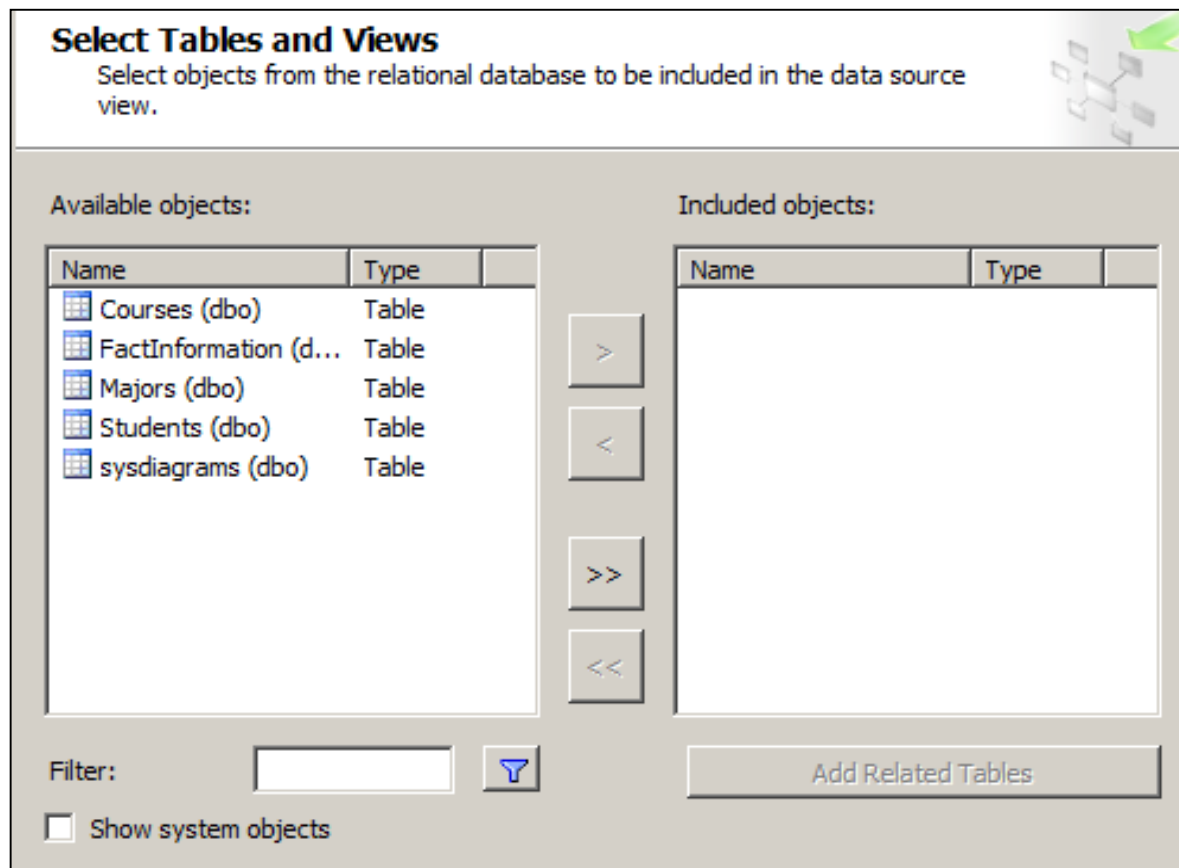
Figure 3.9 Selecting Tables of the DSV.

3.7 Building the Data Cube

After creating the DSV the Data cube is created. It is a
multidimensional matrix that allows the implemented system users
to explore and analyze the collected data from three different
dimensions. These dimensions are:

www.manaraa.com

- University Dimension.

- Student Dimension.

- Semesters Dimension which represents the time dimension.

- Majors Dimension.

Creating the cube started by choosing the DSV that the cube will explore the data from. Next the cube dimensions are selected as in figure 3.10.

Select Measure Group Tables
Select a data source view or diagram and then select the tables that will be used for measure groups.

Data source view:
UOP DW

Measure group tables: Suggest
☐ University
☐ FactInformation
☐ Majors
☐ Students
☐ Semesters
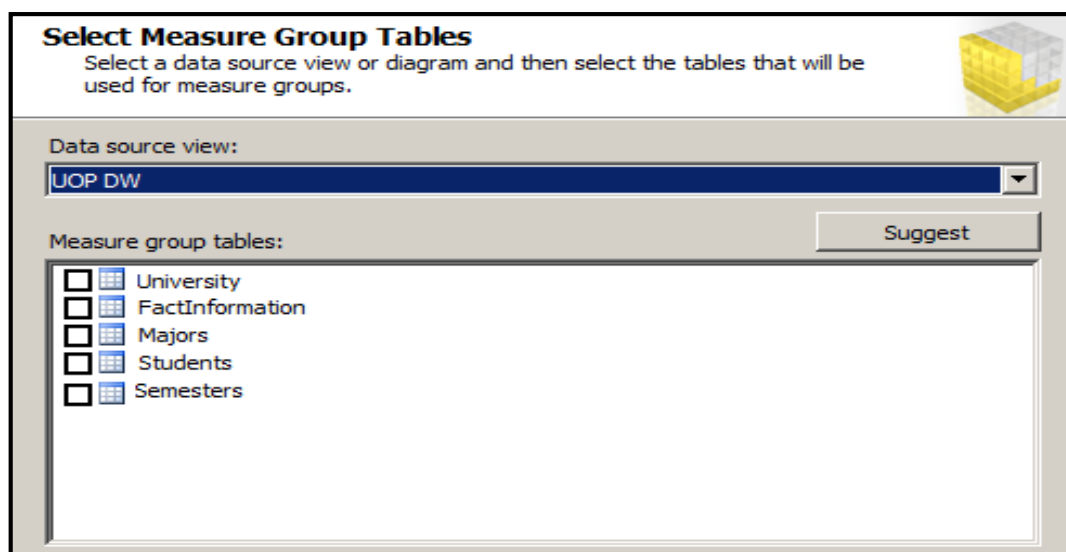
Figure 3.10 Selections of Cube Dimensions.

After the cube dimensions are selected, the measures to be used in data analysis and exploration are selected from the FactInformation table of the university DW as shown in figure 3.11.

الـمنـــارة للاستشارات

www.manaraa.com
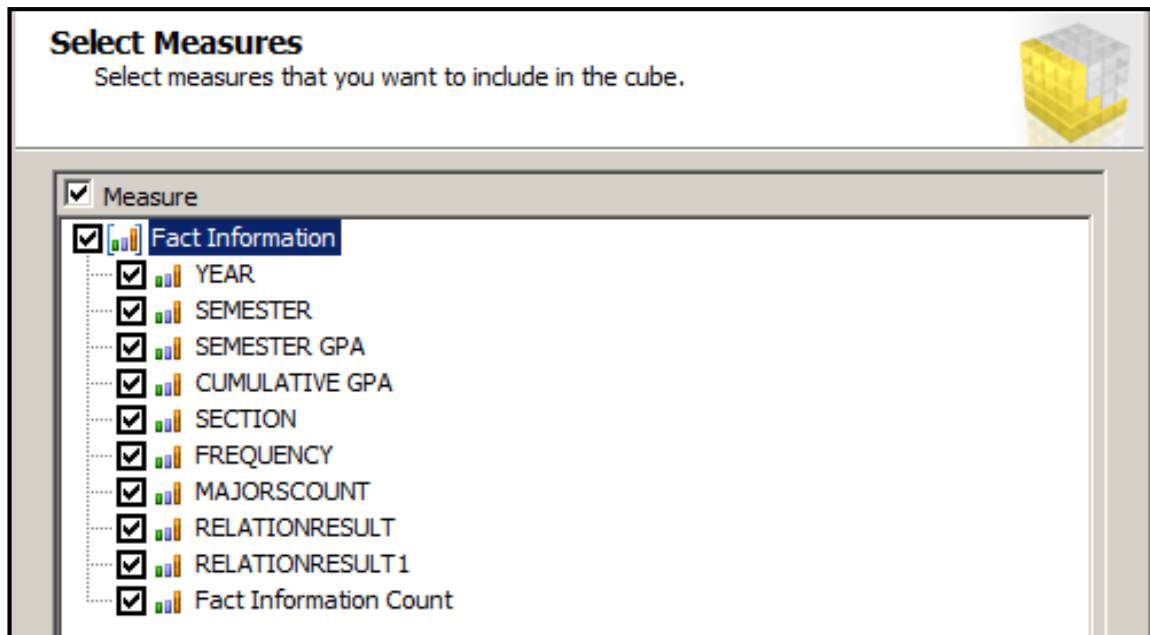
Figure 3.11 Selecting Measures of the Cube.

The measures to be used in the university cube are section
frequency, registered major courses count and cumulative GPA.
The final step is generating the university cube from the created
DSV as shown in figure 3.12.
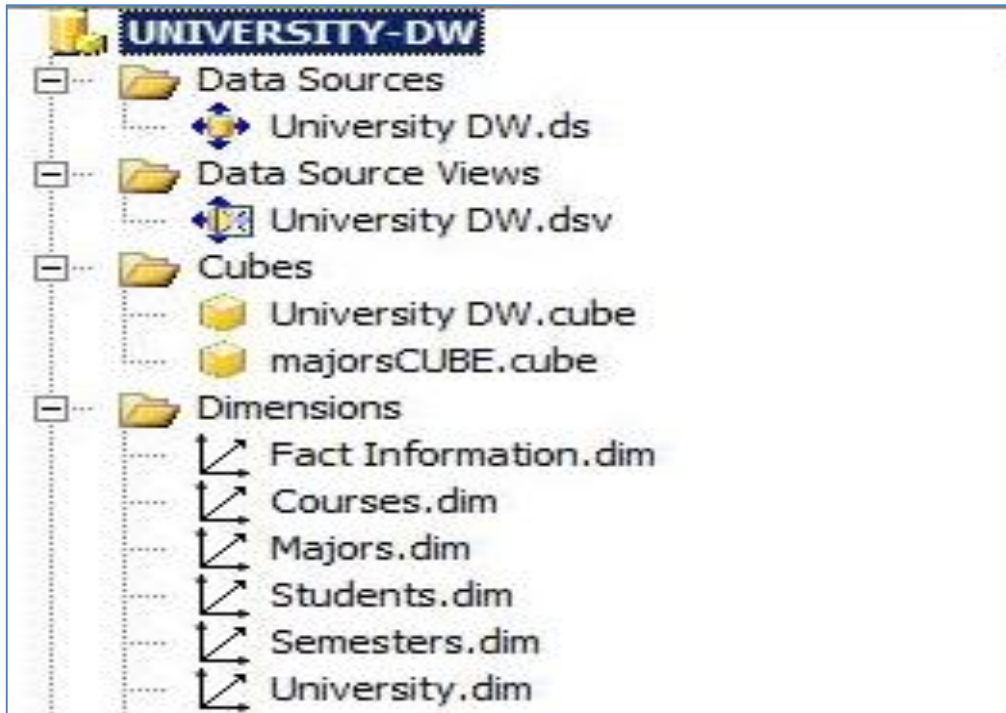
Figure 3.12 the generated cube.

The implemented Data cube is a multidimensional extension of 2-D tables of the DW tables, and it can be viewed as in figure 3.13. The data in the cube can be accessed using The Multidimensional Expressions (MDX) language; the MDX provides a rich and powerful syntax for the retrieval and manipulation of multidimensional data stored in the cube. The cube diagram illustrates that the intersection of multidimensional members which are the values of dimensions attributes creates cells from which the data is obtained to identify and extract such data, whether it is a single cell or a block of cells,

MDX uses a reference system called tuples. Tuples list dimensions and members to identify individual cells as well as larger sections of cells in the cube.
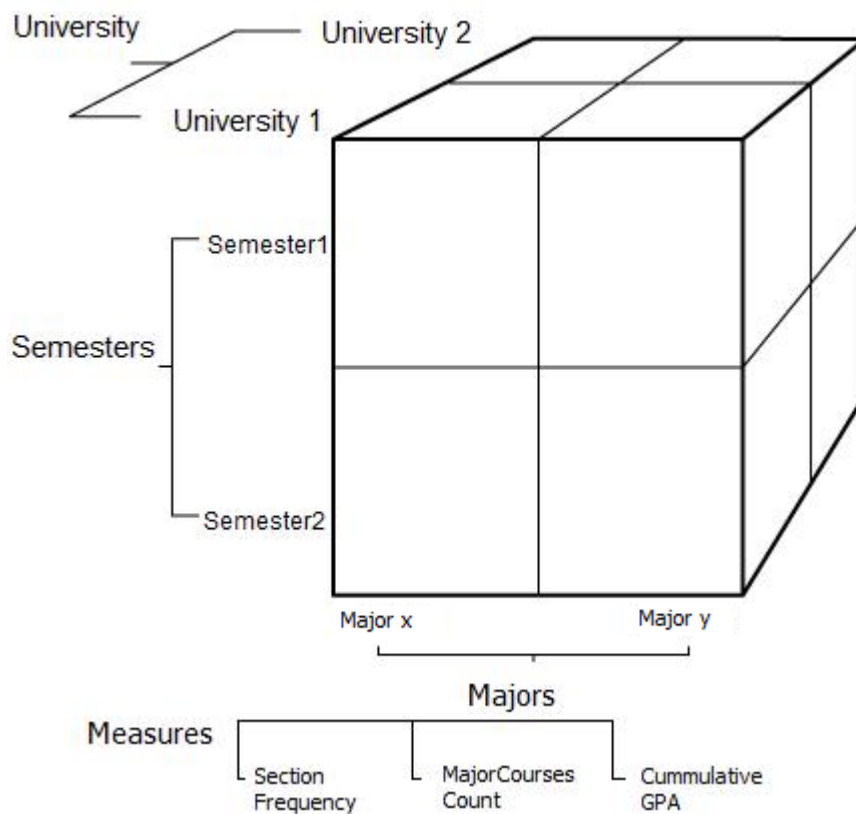
Figure 3.13 The University Cube.

Because each cell is an intersection of all the dimensions of the cube, tuples can uniquely identify every cell in the cube. For the purposes of reference, measures in a cube are treated as a private dimension, named Measures, in the cube itself. For example, in diagram 3.13, the following tuple identifies a cell in which the value is 9060:

(Majors.[MajorName].Pharmacy,Semesters.[Year].2006,University1,

Measures.SectionFrequency)

The previous value states the sections frequency for the registered students in Pharmacy College in university 1 in the year 2006.

An example of MDX query is the following query:

SELECT

{ [University].[UNIVERSITY NAME].Members } ON COLUMNS,

{ [Semesters].[YEAR].Members } ON ROWS

FROM University-CUBE

The result of this query is shown in table 3.3; the query result gives

the aggregated values of cumulative GPA for the two universities.

Table 3.3 Result of MDX Query.

year	University1 Cumulative GPA	University2 Cumulative GPA
2006	18449.99	*NULL*
2007	19230.63	*NULL*
2008	19380.84	4480.53
2009	20134.61	10257.68
2010	20440.17	15448.7

In the listed years in the table these values represent the summation of all students GPA in a given year and it is clear that the null values refer to the years in which the data is missing in university 2. The cumulative GPA results are increasing through the five years from 2006 until 2010 which is a good pointer for the improvement of university academic achievement.

Chapter 4
Results exploration and forecasting

4.1 Cube Results Exploration

The cube can be explored by viewing its results through different perspectives of its dimensions. The most important values are the values of its measures.   These measures are first the cumulative GPA which reflects the student's academic achievements, the frequency which reflects the registration of students for courses sections, and the Major count.

Now the cube is explored by viewing the values of its aggregated measures which is a roll up operation through the cube dimensions as in table 4.1.

Table 4.1 Aggregated Measures

CUMULATIVE GPA	FREQUENCY	MAJORSCOUNT
131487.219999998	609814	48394

The values that resulted from the operations are the aggregated sums of the three factors for the two universities these factors are cumulative GPA for students and sections frequency for students registered in the two universities in the time range between 2006 and 2010.these values seem to be of un clear result for the cube ,because in the cube data certain dimensions must be queried with certain conditions in order to have a meaningful data.

These values must be viewed from a certain dimension in order to have a clear explanation, for that the cube is viewed by slicing it based on university dimension so the results are be clear as in table 4.2.

Table 4.2 Cube Measures with University Dimension.

University 1			University 2		
CUMULATIVE GPA	FREQUENCY	MAJORSCOUNT	CUMULATIVE GPA	FREQUENCY	MAJORSCOUNT
89691.919999999	399102	43752	41795.2999999993	210712	4642

The previous values represent the summation of the studied factors for each university, and it is shown that these values still have a none obvious meaning, and by adding the year attribute from the semesters dimension it will have the following perspective as in table 4.3.

Table 4.3 The Cube Measures with Semesters Dimension.

	University 1			University 2		
YEAR	CUMULATIVE GPA	FREQUENCY	MAJORSCOUNT	CUMULATIVE GPA	FREQUENCY	MAJORSCOUNT
2006	18449.99	101094	6608	*NULL*	*NULL*	*NULL*
2007	19230.63	96818	10194	*NULL*	*NULL*	*NULL*
2008	19380.84	102442	13534	4480.53	30424	*NULL*
2009	20134.61	109284	14772	10257.68	62354	*NULL*
2010	20440.17	106410	16050	15448.7	82988	*NULL*

In the previous table the cell values represent the summation of each factor in a certain year; the aggregated cumulative GPA was

47

increasing from the years 2006 until 2010 which is a good indicator that the students' numbers are increased in the two universities, and for clearer projection of the cube data the time dimension is sliced again using the semester attribute as in table 4.4.

Table 4.4 Projection of Cube Data over Time Dimension.

YEAR	SEMESTER	UNIVERSITY NAME ▾ University 1 CUMULATIVE GPA	FREQUENCY	MAJORSCOUNT
⊟ 2006	1	9352.04000000001	54214	2972
	2	9097.95000000001	46880	3636
	3	6426.93	17462	2208
	Total	24876.92	118556	8816
⊟ 2007	1	9802.03000000002	49172	4708
	2	9428.6	47646	5486
	3	6913.80000000001	18208	4066
	Total	26144.43	115026	14260

By slicing the cube dimensions it is clear that the given results are explored more accurately, and the movement of the views of the cube represents different perspectives of the data which enable the cube viewers to understand the data characteristics. In the previous table the cell values represent the summation of the studied factors over academic semesters within a given year which enable the cube viewer to compare the results of the students in any given year. For example in the year 2006 the cumulative GPA was decreasing in the second semester than it was in the first semester and in the third semester it gives a more decreased value because the number of registered students was decreased.

Now the cube data can be viewed from different perspective though slicing the cube through majors' dimension as in table 4.5.

Table 4.5 Cube Slice through Majors Dimension.

YEAR	MAJOR	University 1		
		CUMULATIVE GPA	FREQUENCY	MAJORSCOUNT
2006	Accounting	2374.21	10964	866
	Arabic Language & Literature	360.81	1470	90
	Architecture	672.72	2398	272
	Business Administration	1546.83	8658	720
	Chemistry	740.070000000001	3364	132
	Child Education	689.41	3070	234
	Computer Information Systems	363.17	1756	98
	Computer Science	898.919999999999	5670	246
	English language and Literature	437.38	1888	116
	English language/translation	1191.4	4690	412
	English Language/Translation	191.64	286	100
	Financial and Banking Sciences	1419.71	7730	622
	Graphic Design	1252.36	5188	296
	Interior Design	535.4	1930	176
	Journalism	1548.13	6484	670
	Management Information System	2362.45	12146	824
	Marketing	1616.77	8700	756
	Nutrition	363.09	2316	100
	Pharmaceutical Sciences	73.65	188	100
	Pharmacy	3666.5	19172	1462
	Software Engineering	1283.64	5326	338
	Teacher Class	1288.66	5162	386

Table 4.5 results enable the cube viewers to explore its data from major perspective; the cells values present the summation of the studied factors for each major in university1 in the year 2006, by these values a comparison between majors' students can be explored and the majors that have the highest values for cumulative GPA give indicator that it have the highest number of registered students or have the highest values of students cumulative GPA

which in both cases states the major level of performance. And to
indicate this result more each student results for the studied factors
can be shown individually as in table 4.6.

Table 4.6 Slice and Dice through Student Dimension.

| | | UNIVERSITY NAME ▼ | |
| | | University 2 | |
YEAR ▼	STUDENT ID ▼	CUMULATIVE GPA	FREQUENCY
	200810086	2.18	16
	200820191	2.18	10
	200810302	2.2	26
	200810786	2.2	26
	200810925	2.22	16
	200830050	2.22	12
	200810224	2.24	10
	200810494	2.25	18
	200810502	2.25	18
	200810745	2.25	30
	200810769	2.25	24

Table 4.6 is a cube view from the student dimension through a slice
and dice operation, and the results indicate each student academic
performance which enables the cube viewer to make comparison
between the results in the different academic semesters for a certain
student, also comparisons between the same major students can be
made in order to evaluate their performance ,these comparisons

can be made also between different majors and by this the university performance can be evaluated through the different perspectives of the cube.

4.2 Forecasting of the Cube Results

The forecasting is done using exponential smoothing model; the forecasting is done for university1 records results only, table 4.7 shows the forecasting statistics for cumulative GPA, and chart 4.1 shows the   values of cumulative GPA and forecasted values.

Table 4.7 Forecasting Statistics for Cumulative GPA.

University 1			
Year	Semester	Cumulative GPA	Forecasted GPA
2006	semester1	9352.04	
2006	semester2	9097.95	9352.04
2007	semester1	9802.03	9352.04
2007	semester2	9428.6	9174.177
2008	semester1	9783.61	9613.6741
2008	semester2	9597.23	9484.12223
2009	semester1	10226.46	9693.763669
2009	semester2	9908.15	9626.190101
2010	semester1	10334.61	10046.37903
2010	semseter2	10105.56	9949.618709
2011	semester1		**10219.11261**
2011	semester2		**10239.62578**
2012	semester1		**10395.26656**
2012	semester2		**10456.31801**
2013	semester1		**10551.27487**

The values for the cumulative GPA for university 1 are increasing through the first and second semester in the years from 2006 until

51

2010 for university 1 which is a good indicator of the university academic achievement of its students, when forecasting by using exponential smoothing method is calculated for the second five semesters after semester 2 in year 2010 it is predicted to increase indicating that the number of registered students in the university are going to be increased.
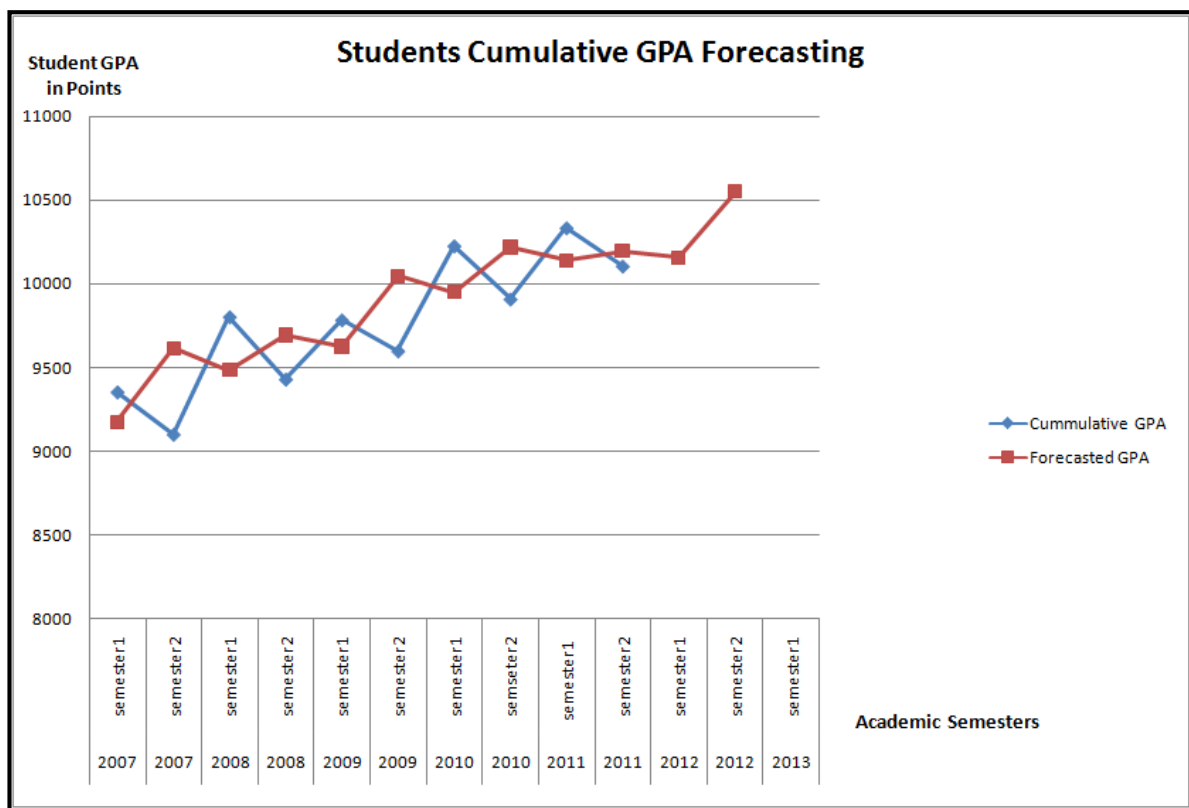

Chart 4.1 Forecasting Cumulative GPA.

The chart results present the lines of the cumulative GPA using small squares for the points and the forecasted cumulative GPA using the big squares for the points ,the forecasted line indicates that the value of the cumulative GPA are going to be increased in

the first semester of the year 2011 which indicates that the expected number of registered students will be increased. That is a good indicator for the university performance evaluation. And to evaluate the university performance more accurately the statistics for forecasting section frequency are shown in table 4.8

Table 4.8 Forecasting Statistics for Section Frequency

University 1			
Year	Semester	Section Frequency	Forecasted Frequency
2006	semester1	54214	
2006	semester2	46880	54214
2007	semester1	49172	54214
2007	semester2	47646	49080.2
2008	semester1	51502	49144.46
2008	semester2	50940	48095.538
2009	semester1	56272	50480.0614
2009	semester2	53012	50802.01842
2010	semester1	55414	54631.00553
2010	semseter2	50996	53497.70166
2011	semester1		54839.1105
2011	semester2		52148.93315
2012	semester1		54032.05729
2012	semester2		52713.87039
2013	semester1		53636.60122

The results in table 4.8 shows the results of forecasting by using exponential smoothing method when calculated for the second five semesters after semester 2 in year 2010 it is predicted to be increased in the next semester which indicates that the number of

registered students in the university is going to be increased in this semester. But the prediction values for later semesters indicates a decrease in section frequency which is not a good indicator for university performance, the forecasting results are shown in chart 4.2.
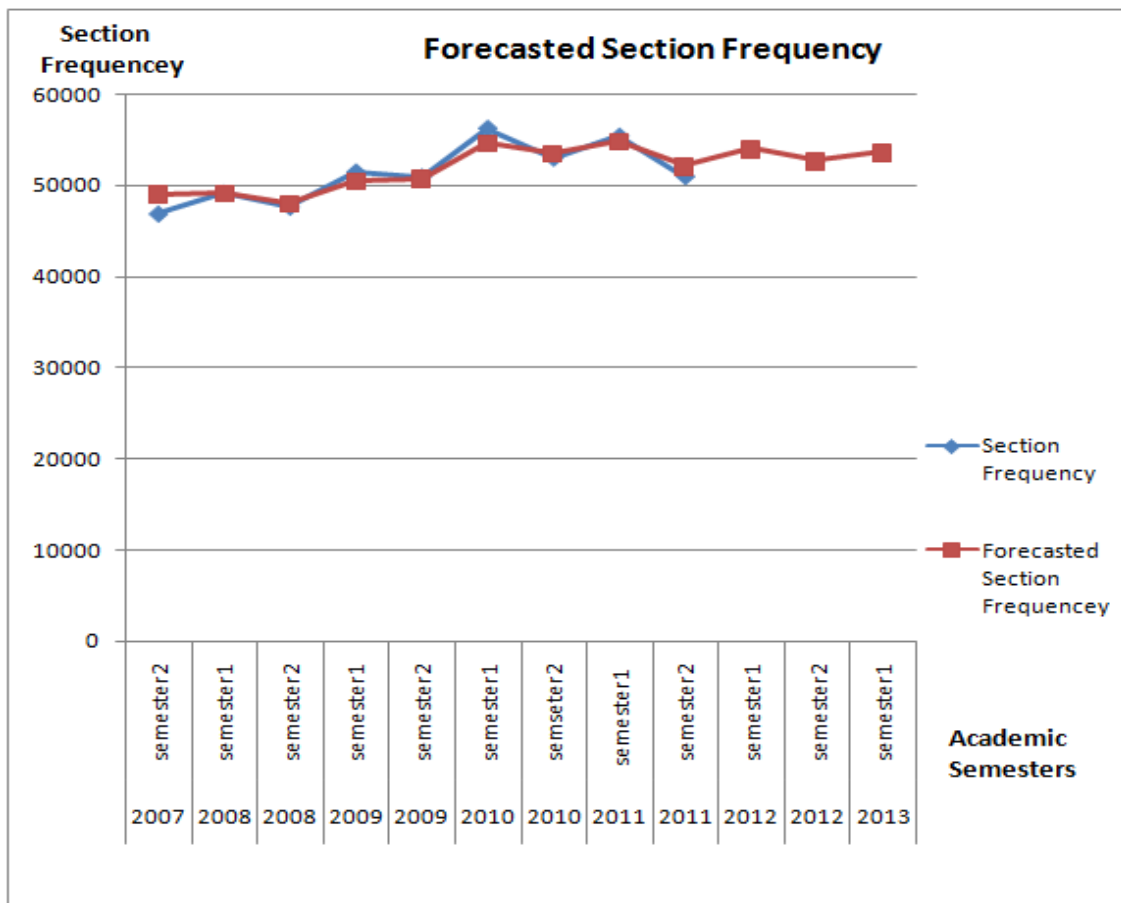

Chart 4.2 Forecasting Section Frequency.

Chart4.2 presents the line of section frequency using small squares for points and the line of forecasted section frequency using big squares for points, the forecasted line for section frequency indicates that the value of section frequency is predicted to be increased in the first semester of the year 2011 which is a good

المنارة للاستشارات

www.manaraa.com

indicator for the increasing number for registered students in the university, and for more accurate exploration of the university performance the statistics of the major course count in university1 are viewed in table 4.9 .

Table 4.9 Forecasting Results for Major Courses Count.

University 1			
Year	Semester	Major Courses Count	Forecasted Major courses count
2006	semester1	2972	
2006	semester2	3636	2972
2007	semester1	4708	2972
2007	semester2	5486	3436.8
2008	semester1	6600	4326.64
2008	semester2	6934	5138.192
2009	semester1	7222	6161.4576
2009	semester2	7550	6702.23728
2010	semester1	7850	7066.071184
2010	semseter2	8200	7404.821355
2011	semester1		7716.446407
2011	semester2		8054.933922
2012	semester1		8296.480177
2012	semester2		8127.397783
2013	semester1		8252.069405

Table 4.9 results show the improvement of major courses count for the years range from 2006 until 2010 which are increasing and that is a good indicator for the university performance as shown in chart 4.3, the chart used the line of small squares for points of the major courses count and used the line of big squares for the points of forecasted major courses count, the forecasted results of the major courses count are predicted to increase.
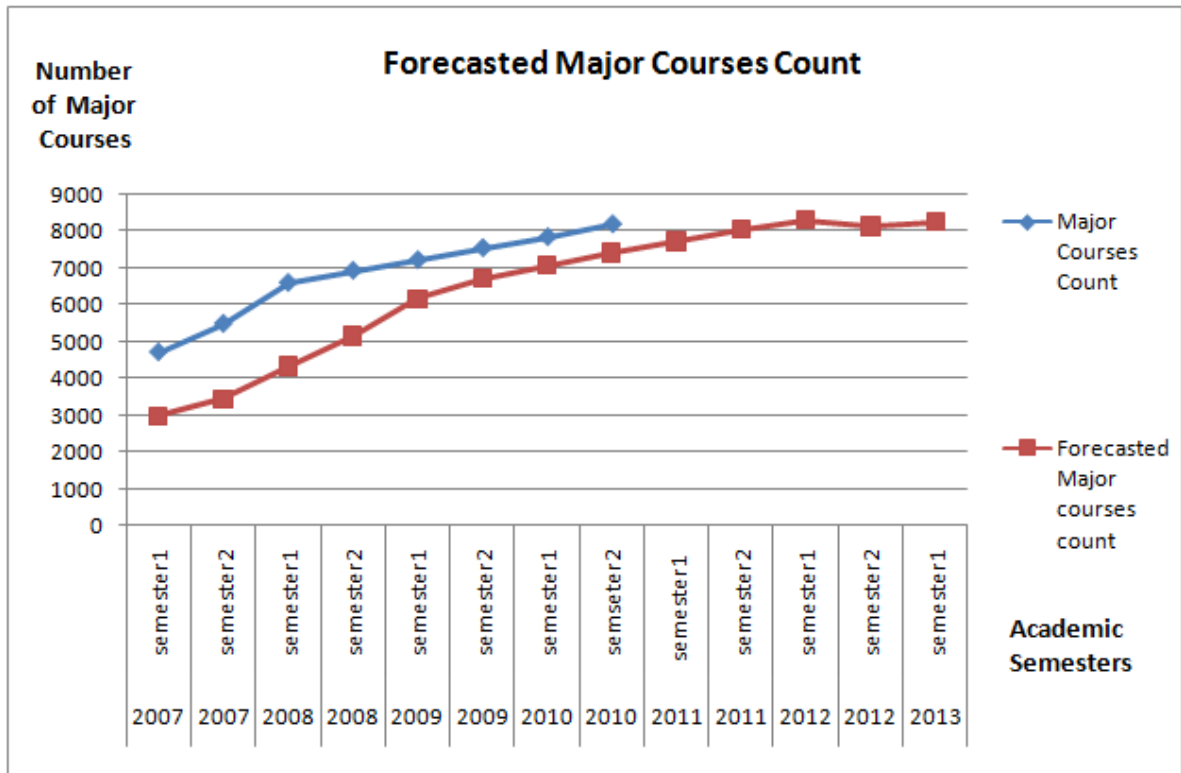
55

Chart 4.3 Forecasting results for Major Courses Count.

The results from table 4.9 of forecasting show improvement of the university performance from the year 2006 until the year 2010. But this improved numbers are increased in small rates. For that the university needs to review its strategies in evaluation the student performance, and it needs to monitor the student success rates in their major courses to indicate if they are proceeding in rates that will increase the university evaluation measures.

Chapter 5
Conclusion and Future Work

5.1 Discussion

This thesis developed a data warehouse system for university student information system. The main goal of the data warehouse development was for analysis. The data warehouse was designed because it is easier and faster to execute queries in a data warehouse design rather than in a transactional database system. Then the data warehouse is used for analysis and data cubing in which the data was viewed from different dimensions according to specified measures and finally the resulted data from the cube were used in forecasting.

5.2 Conclusion

This research studied the student information system in two private Jordanian universities data in order to show that it can be analyzed not only stored in its databases the thesis ended with these conclusions.

- The development of a data warehouse for student information system

in the university environment depends on historical data of the students' records and it requires a direct connection between the

transactional database of university student information system with a data warehouse that is designed to hold the transactions of students activities and it must be loaded in regular basis at the end of each academic semester so comparisons can be done for students achievements in each semester.

- The integration of data from different universities into one data warehouse requires the sharing of data between different universities so students performance can be compared between different universities.

- A main obstacle in this thesis was in data collection because there is missing data in students' historical records; the main cause of missing data is the ignorance of the employees who were responsible for the records storage.

- The data cubing results indicates that the studied universities should evaluate the progress of its students through their registration semesters not only through the cumulative GPA but using other techniques to solve the students' problems.

5.3 Future Work

The developed system could be extended to focus into the following points:

- Analysis of the developed data cube can be further improved to find the data trends in students different activities using data mining algorithms.

- The developed system could be extended to study the students trends in registration in different courses.

- The developed system could study the courses offering for certain students groups according to their needs.

References

1. Claudia Imhoff et al , "Mastering Data Warehouse Design: Relational and Dimensional Techniques" , John Wiley & Sons   2003.

2. Oracle," Database Data Warehousing Guide 10g Release 1", Report (10.1) December 2003.

3. Hsinchun Chen,"Business and Market Intelligence 2.0", IEEE Intelligent Systems Journal February 2010.

4. Larissa T Moss, Shaku Atre,"Business Intelligence Roadmap: The Complete Project Life Cycle for Decision-Support Applications", Addison –Wesley USA 2003 1st Edition.

5. Byron Marshal et al, "Ebizport: Collecting and Analyzing Business Intelligence Information", Journal of the American Society/ For Information Science and Technology, Vol 55(10) Pp 873-891, August 2004.

6. Lewis C Chasalow, "a Model of Organizational Competencies for Business Intelligence Success", Dissertation, Virginia Commonwealth University, May 2009.

7. Manjunath et al," Design and Analysis of DWH and BI in Education Domain ", International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011.

8. Singh et al," The Analytical Data Warehouse: A sustainable approach for Empowering institutional Decision Making" , International Journal of Engineering Science and Technology (IJEST) Vol. 3 No. 7 July 2011.

9. Piedade et al," Business Intelligence in Higher Education", CISTI 2010.

10. Rebbapragada et al ,"Data Mining and Revenue Management Methodologies In College Admissions", April 2010 | vol. 53 | no. 4 | communications of the ACM.

11. http://www.mcmaster.ca accessed 30 January 2012.

12. Inga Staškevičiūtė et al, "The Intelligent University's Conceptual Model", Engineering Economics Journal, Vol .No 4(59) Pp 53-58, 2008.

13. http://www.informationbuilders.com accesses 30 January 2012.

14. http://www.slu.edu/appc/enterprise_data_warehouse.html accessed 29 January 2012.

15. Report by Suzanne Nelson http://www.dmu.ac.uk accessed 29 January 2012.

16. Bielec. J," When Universities Act like Businesses", University Business Journal, Vol 5(9), P. 9. 2002.

17. Blitzblau.R et al,"Transforming Georgetown through Technology ", Educause Quarterly Journal, Pp. 40-50. November 2001.

18. Veronica Ivancenco et al ," Efficient Information Management – Essential Factor for High Performance Management" , Journal of Metalurgia International Vol. Xv (2010) Special Issue No. 4 Pp 51-54.

19. Öykü Işık "Business Intelligence Success: an Empirical Evaluation of the Role of BI Capabilities and the Decision Environment", Dissertation University of North Texas, August 2010.

20. Nadia Brannon, "Business Intelligence and E-Discovery", Intellectual Property & Technology Law Journal Volume 22 . Number 7 . July 2010

21. Scott M. Janke," Business Intelligence within Commercial Real Estate: an Application of the Motivation and Acceptance Model", Dissertation, Capella University May 2009.

22. Wei Peng,"Actionable Pattern Analysis in Computing and Business Intelligence Systems", Dissertation, Florida International University, 2008.

23. Konstantin Taskov, "Organizational Factors Contributing to an Effective Information Technology Intelligence System", Dissertation University of North Texas December 2008.

24. Celina M. Olszak et al, "Approach to Building and Implementing Business Intelligence Systems", Journal of Information, Knowledge, And Management Volume 2 Pp 135-147, 2007.

25. Robert K. Joyce,"Successful Implementation of Decision Support Systems for Supply Chain Management within Manufacturing Companies", Dissertation, Nova Southeastern University 2005.

26. Mitja Jermol et al, "Managing Business Intelligence in a Virtual Enterprise: a Case Study and Knowledge Management Lessons Learned", Journal of Intelligent & Fuzzy Systems,Vol 4 Pp 121-136 , 2003.

27. K .Rajesh," Business Intelligence for Enterprises", The ICFAI Journal of Information Technology March 2008.

28. Lokken. B, "Business Intelligence: an Intelligent Move or Not?" Http://Businessintelligence.Ittoolbox.Com/Pub/Ao031202.Pdf Accessed on 15/1/2011.

29. Sheila R. Garner,"Data Warehouse Implementation Strategies: A Mixed Method Analysis of Critical Success Factors ", Dissertation Capella University August 2007.

30. Orr.K, "Data Warehousing Technology", Retrieved From Http://Www.Kenorr.Com. Accessed March 2011.

31. Kimball. R & Ross. M," The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling", Wiley & Sons 2nd Edition. New York. (2002).

32. Imhoff, C. et al, "Building the Customer-Centric Enterprise: Data Warehousing Techniques for Supporting Customer Relationship Management", Wiley & Sons. New York (2001).

33. Todman.C," The Conceptual Model. In (Ed.), Designing A Data Warehouse: Supporting Customer Relationship Management", Prentice Hall. (1st Ed., Pp. 60-73). New York   (2001).

34. Lee, H et al,"a Metadata Oriented Architecture for Building Data Warehouse", Journal of Database Management, 12(4), 15-25. (2005).

35. Harold Palmer,"a Data Warehouse Methodology and Model For Student Data In Higher Education", Dissertation, Nova Southeastern University, 2006.

36. Balkan. L, et al," Data Standardization Means Quality Information From Distributed Systems", Cause/Effect Journal, Pp. 33-41. (1992, Winter).

62

37. Inmon, W.H.," Building the Data Warehouse", 3rd Edition. Wiley, New York, 2002.

38. Aouiche. K, Jouve. P, & Darmont. J," Advances in Databases and Information Systems". Springe Journal , Berlin/Heidelberg. (2006).

39. Thilini Ariyachandra, et al,"Which Data Warehouse Architecture is Best?", Communications Of The ACM conference Vol. 51 | No. 10 October 2008.

40. Batini, C., et al ," Conceptual Database Design: An Enity-Relationship Approach", Benjamin/Cummings report, Redwood City, Ca,1992.

41. Michael W. Martin ,"Implementing Real-Time OLAP With Multidimensional Dynamic Clustering ", Dissertation, North Carolina State University 2006.

42. Mahesh H Panchal ," Online Analytical Mining: Architecture and Challenges ,Journal Of Systems Management, Vol. Viii, No. 2, 2010.

43. Zohreh Asgharzadeh Talebi ," Exact and Inexact Methods for Solving the View and Index Selection Problem for OLAP Performance Improvement Dissertation North Carolina State University 2010.

44. Kevin .S.Beyer," Precomputation Techniques for Interactive Data Exploration ", Dissertation University of Wisconsin, 2002.

45. Arun Sen. et al ,"Using a Common Set of Attributes to Determine Which Methodology to Use In a Particular Data Warehousing Project, a Comparison of Data Warehousing Methodologies ", Communications of the ACM conference March 2005/Vol. 48, No. 3.

46. Ahmed taleb ,"An efficient framework for the manipulation of OLAP hierarchics ",Thesis ,Concordia University Montereal Canada, March 2007.

47. Boyong Liang,"Compressing data cube in parallel OLAP systems", Thesis Carleton university Ottawa Canada, April 2005.

48. Michael.W.Martin,"Implementing Real-Time OLAP with Multidimensional Dynamic Clustering", Dissertation North Carolina State University 2006.

49. Zhenshan Guo,"Partial Aggregation and Query Processing of Olap Cubes", Thesis Simon Fraser University Spring 2009.

50. S. Dehuri et al,"Parallel Processing of OLAP Queries Using a Cluster of Workstations", International Journal of Information Technology & Decision Making   Vol. 6, No. 2 (2007) 279–299.

51. William Dilla et al ,"Interactive Data Visualization: New Directions for Accounting Information Systems ",Journal of Information Systems American Accounting Association Vol. 24, No. 2 ,Fall 2010, Pp. 1–37.

52. Stephen Few, "Fundamental Differences In Analytical Tools Exploratory, Custom, Or Customizable ",Http://Www.Panopticon.Com/Showroom/Literature.Htm Accessed 8-3-2011.

53. Umeshwar Dayal et al,"   Data Integration Flows for Business Intelligence ", Proceedings of the 12th International Conference On Extending Database Technology: Advances in Database Technology   ACM New York, USA 2009.

54. Accelerated Edge Incorporated," How to Select Business Intelligence Query Tool ", Report, September 2004.

بناء نظام معلومات طلبة الجامعة باستخدام مستودعات البيانات

إعداد
نور حسين الحياري

إشراف
الدكتور مزهر العاني

قدمت هذه الرسالة لإستكمال متطلبات الحصول على درجة الماجستير في
علم الحاسوب

جامعة عمان العربية

كلية العلوم الحاسوبية و المعلوماتية

(2012 ، شباط)