ADAMA SCIENCE and TECHNOLOGY UNIVERSITY

Research Title: Designing Prototype Data warehouse Decision Support system

BY: **Tagel Aboneh and Ejigu Tefera**

June, 2017

Adama, Ethiopia

Contents

Figures

Tables

Acronym

DW: Date warehouse

OLAP: Online Analytical Process

ROLAP: Relation Online Analytical Process

OLTP: On-line Transactional Process

SQL: Structure query language

SSIS: SQL Server Integration Services

SSAS: SQL Server Analysis Services

SSRS: SQL Server Reporting Services

SSMS: SQL-server management studio

DSA: Data Staging Area

OWB: Oracle Warehouse Builder

INARIS: Integrated National Agricultural Resources Information System

GIS: Geographic Information System

CDW: Central Data Warehouse

Abstract

A data warehouse is a “subject-oriented, integrated, time varying, non-volatile” collection of data that aimed at enabling managers and analysts to make better and faster decisions. Properly designed and integrated data is crucial to enhance the internal business processing and strategic decision making purpose. But, important data such as students, human resources, course, staff, schools or departments are recorded in an appropriate manner. As the result, managers and decision makers of the institutions always facing challenges in order to analyze, generate summarized report and make strategic decision in their respective level of management.

The main objective of the study is to design a prototype data warehouse system which provides an integrated platform for performing analytic and decision making process from academic data (students, courses, Instructors and academic departments integrated in Repository). Dimensional modeling approach is selected to structure and describe the data warehouse that comprises fact and dimensional tables. The proposed decision support data warehouse system is designed using Star schema dimensional modeling technique which efficient in query performance for large data sets and efficient navigation of data in multiple dimensions for analysis. Therefore, the proposed system is more users interactive and query process is very fast than OLTP. It successfully provides summarized and multidimensional analytical information for strategic decision makers.

Chapter One

1. Introduction

1.1. Background of the Study

Nowadays properly managing and utilizing decision support information systems will give competitive advantages for the Higher Educational Institutions. Because of its size or functions such Institutions often have a lot of information systems and subsystems which are crucial for their internal processing and operations. Examples of such subsystems include the student registration system, the accounting system, the course management system, the human resource management system, and many others. The fact that most Adama Science and Technology university's crucial data about students, instructors, curriculum, human resources, schools or departments are stored differently at schools, departments and individual's unit levels in fragmented manners. As the existing databases are created in different platforms, it causes system dissimilarity and lack of standard to manage the data properly. In addition these systems are existed in isolated and standalone machine that are employed for daily transactions and operations not for analytics [1].

Currently Adama Science and Technology University is working vigorously to cope up with the rapidly changing world. The university is also trying to implement the entirely required ICT infrastructures and facilities, that contribute for the enhancement of accessibility of network service and sharing educational resource within the University. Implementation of data warehouse management tools will decision maker by providing strategic information. However, Manager and decision makers of the institutions are always facing challenges to organize and compile fact or data to make a right decision in their respective level of management.

The competitive advantage from a new technology enforced managers to seek for a new way to increase their efficient and effective decision support to improve their decision-making processes. In this sense, the idea of data warehouse and data mining was born [2]. In fact, data warehousing is the process of collecting data from operational and functional databases, transforming, and then archiving them into special data repository called data warehouse with the goal of producing accurate and timely management information [3].

Therefore, the purpose of this proposed Data Warehouse system is to integrate heterogeneous data sources at corporate level. The data warehouse acts as a central repository and it contains the "single version of truth" for the organization that has been carefully constructed from data stored in disparate internal and external operational databases systems [4].

1.2. Research Questions:

The proposed research tried to address the following research questions:

- ✓ What type of data manipulation mechanism and methods does the university use to make strategic decision?
- ✓ What type of functional requirements do users need to get from the data warehouse system?
- ✓ How to plan and implement suitable data warehouse decision support system for ASTU registrar's system.

1.3. Statement of the Problem

The day to day operations of the schools, program and units in the university rely heavily upon the excel data and partly on operational system. The need of valuable information [5] for strategic decision necessitates the development of the data warehouse. In the existing system, operational data (students, courses, school, program and academic staff etc.) are organized in fragmented manner at different levels of management. Due to this problem, currently the academic data are stored and accessed in a manual and time consuming approach. As a result of fragmented databases, managers are struggling with the rise of increasingly complex and diverse information to make strategic decisions. Such system will not support to perform roll-up and drill-down operations in order to understand student enrollments by year, academic performance by course or any dimensional tables as we desired.

The main motivation for building data warehouse for the educational institute is from two sources, internal sources like inability of the current operational systems to provide required information and external sources like competitive advantage of data warehouse [6].

Moreover, the researchers have undertaken an observation to understand the current practices of the University registrar office how does they handle students' data and make decisions based on

the existing databases. After discussing with the admission and registration office, the following critical problems have been investigated.

1. The data is stored in different sources in fragmented manner.
2. The system will not provide analytical information for strategic decision
3. There is no integrated or centralized database system that helps managers to make multidimensional analysis of data warehouse by student, course, by department dimension as the time of need and fast reporting.
4. It very difficult generate pattern from existing historical data for prediction and future forecasting.

Therefore, the main purpose of this research is to investigate current system of information delivery process and to propose data warehouse system to provide timely, accurate and consistent information for decision makers. We can also test analytical performance of the proposed system using multidimensional analysis tools such as cube browsing and pivot analysis.

1.4. Objectives

1.4.1. General objective

The general objective of the study is to design a prototype Data warehouse decision support system in order to make effective and analytical decision. This kind of system will help to come up with robust decision from time variant and historical data.

1.4.2. Specific objectives

To achieve the general objective, the following specific objectives have been accomplished:

- ✓ Identify and gather user information requirements
- ✓ Requirements analysis and specifications
- ✓ Design data warehouse model using dimensional modeling techniques
- ✓ Design the data warehouse architecture and loading of data from various sources into an integrated data warehouse.
- ✓ Testing the proposed prototype decision support data warehouse system

1.5. Benefits and Beneficiaries of the Proposed System

Nowadays, many higher education institutions are starting to see the value of the integrated, standardized, clean and easy access to data for better decision making. With the usage of data warehousing, analysis, management of decision making process and other reports can be done in a simpler way [7]. In this aspect, the proposed student data warehouse has the following benefits for Adama Science and Technology University.

- ✓ It Provide a centralized source of information accessible across different academic units to quickly analyze problems and get satisfactory solutions. It used to get data quickly and easily to perform analysis (visualization, what-if). One can work with better information and make better analytical decisions based on data warehouse.
- ✓ The system can be used as a model to expand and deploy university corporate data warehouse that will integrate all university information system such as human resource system and properties administration system.
- ✓ The document will be used as reference material for post graduate students who is interested to conduct research in the area of data warehouse and data mining.

1.5.1. Beneficiaries

- ✓ Student admission and registration office and all associate registrar officers can use the warehouse to make analytical decision on student academic achievement.
- ✓ School deans and Department heads also can make strategic decision and prediction based on the centrally consolidated, time variant and subject oriented data warehouse.
- ✓ Students and Researchers will benefit for further research in the domain

Chapter Two

2. Literature Review

2.0. Introduction

Nowadays organizations are competing in the global market to gain competitive advantage over the others. Successfully supporting managerial decision-making is critically dependent upon the availability of integrated, high quality information organized and presented in a timely and easily understood manner [8]. One of the dominant areas is building of data warehouse system to support decision maker with strategic analytical information from central repository of historical data which provides an integrated platform to analyze historical data [9] [10]. In this chapter we will discuss different related literature of books, journals, proceeding and others sources of information. The reviewed material justify data warehouse state of the art, designing methodology, data warehouse architecture, development and deployment process, data warehouse application, related work and gap analysis.

2.1. State of the Art

Definition of data warehouse

Data warehousing can broadly defined as a collection of decision support technologies, aimed at enabling the knowledge worker (executive, manager, and analyst) to make better and faster decisions based on analytical information [11]. According to (Inmon, 1996) a data warehouse is a "subject oriented, integrated, nonvolatile and time variant collection of data in support of management's decisions". As the author described the fundamental reason for building a data warehouse is to improve the quality of information in the organization.

Data warehousing has come into being because the file structure of the large institution core business systems is unfavorable to information retrieval. The purpose of the data warehouse is to combine core business and data from other sources in a format that facilitates reporting and decision support. The major function of a data warehouse is to extract, load and translate data from different sources into one centralized large database [12]. In this case our main focuses of building Data warehouses are targeted to provide analytical information to support decision maker [13]. Historical, summarized and consolidated data is more important than detailed, individual records. Since data warehouses contain consolidated data, perhaps from several

operational databases, over potentially long periods of time, they tend to be orders of magnitude larger than operational databases; enterprise data warehouses are projected to be hundreds of gigabytes to terabytes in size. The workloads are query intensive with mostly ad hoc, complex queries that can access millions of records and perform a lot of scans, joins, and aggregates. Query throughput and response times are more important than transaction throughput [14].

Online transaction processing (OLTP) [15] systems are useful for addressing the operational data needs of a firm. However, they are not well suited for supporting decision-support queries or business questions that managers typically need to address. Such questions involve analytics including aggregation, drilldown, and slicing/dicing of data, which are best supported by online analytical processing (OLAP) [14] systems. Data warehouses support OLAP applications by storing and maintaining data in multidimensional format. Data in an OLAP warehouse is extracted and loaded from multiple OLTP data sources (including DB2, Oracle, IMS databases, and flat files) using Extract, Transfer, and Load (ETL) tools [16].

Typically, the data warehouse is maintained separately from the organization's operational databases. In addition, data warehouse contains granular corporate data. Data in the data warehouse is able to be used for many different purposes, including sitting and waiting for future requirements which are unknown today [17].

Data Warehousing Methodology

Data warehousing methodologies share a common set of tasks, including business requirements analysis, data design, architecture design, implementation, and deployment [16]. With data warehousing, you can provide a common data model for different interest areas regardless of data's source [18]. The researchers reviewed different types of data warehouse designing methodologies to select the appropriate one. The following figure shows the summaries of methodologies and each methodologies have their own unique characteristics and feature.

Attributes	NCR/Teradata Methodology	Oracle Methodology	IBM DB2 Methodology	Sybase Methodology	Microsoft SQL Server Methodology
Core Competency	Teradata DBMS (massively parallel DBMS)	Oracle DBMS	DB2 DBMS	Sybase DBMS	SQL Server DBMS
Requirements Modeling	Interview, JAD, Prioritization, templates, document analysis	Interview, Prioritization, subject areas	Interview, JAD	Interview	Interview document analysis
Data Modeling	ERD, relational schema	Dimensional model, Star schema	Dimensional model, Star schema	ERD, Star schema, Relational schema	Dimensional model, Star and Snowflake schemas
Support for Normalization/ Denormalization	Develops all relations as normalized, allows denormalization	Allows both	Allows both	More slanted toward denormalization	Allows both
Architecture Design Philosophy	Enterprise data warehouse	Data marts	Enterprise data warehouse and data marts	Data marts	Enterprise data warehouse and data marts
Implementation Strategy	Iterative	Dimensional Life Cycle	Iterative (prototyping)	Iterative (RAD)	Iterative
Metadata Management	Yes, uses a repository	Yes, uses Oracle Repository	Yes, uses a repository	Yes, uses a repository	Yes, uses Microsoft Repository
Query Design	Parallel query development	Allows parallel queries	Not reported	Not reported	Allows parallel queries
Scalability	Yes, to hundreds of Terabytes	Not reported	Yes	Not reported	Yes, to Terabytes
Change Management	Has post audit reviews, but not emphasized in the methodology	Not reported	Not reported	Has maintenance in the methodology	Not reported

Figure1: Data warehouse methodologies (adopted from [16])

During methodology selection phase the Business requirements analysis is used to elicit the business questions from the intended users of the data warehouse. Based on the users requirement the appropriate methodology is selected to design the prototypes data warehouse system.

Data warehouse system architecture

Architecture is a blueprint that allows communication, planning, maintenance, updating, and reuse of the system. It includes different components such as data design, technical design, and hardware and software infrastructure design. The data warehouse architecture [19] are designed on the basis of the specific requirements of a business [20]. A generalized model is depicted as follows: As data is transferred from an organization's operational databases to a staging area, from the staging area it transformed into a data warehouse and is set into conformed data marts

[19]. The copying of data is carried out by means of an ETL technology where data is extracted, transformed, and loaded. This is also represented with a schematic diagram below.

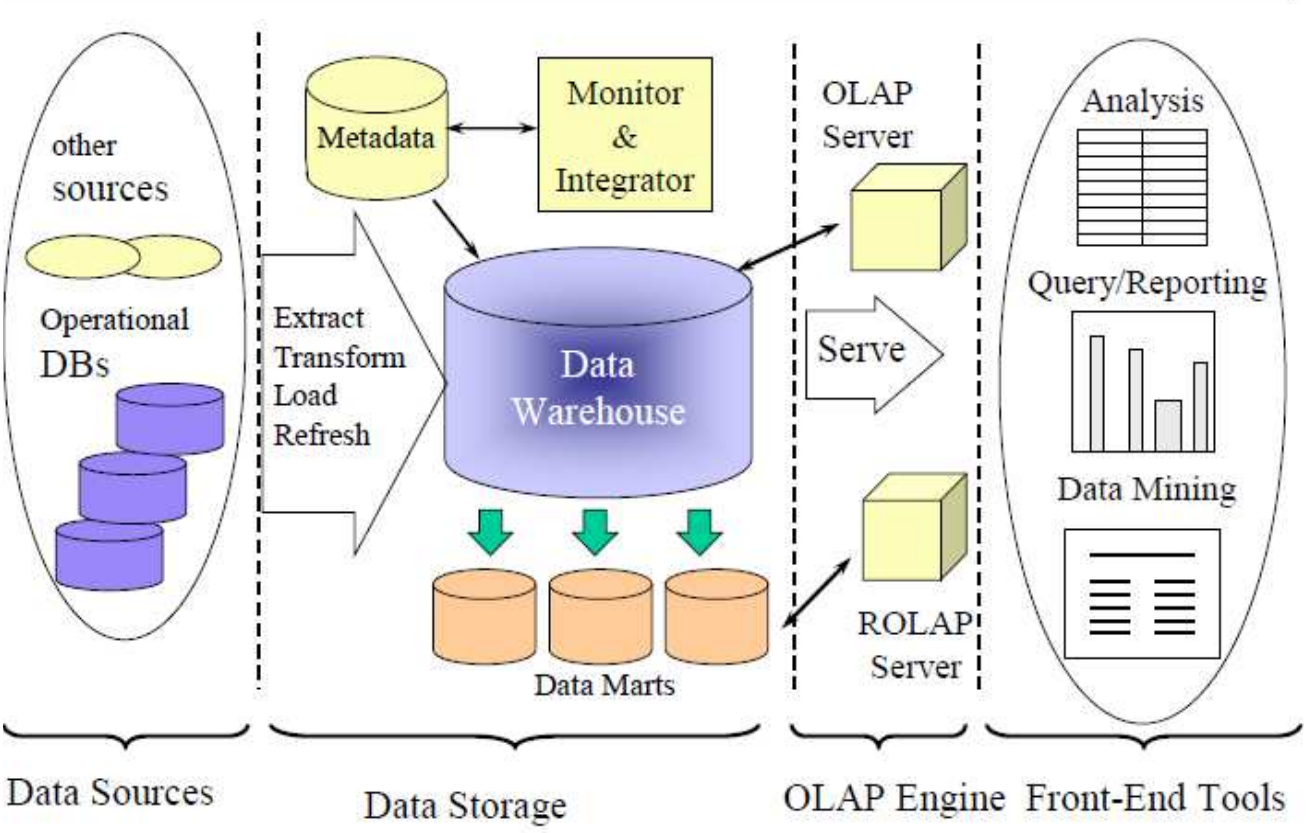


Figure 1: The data warehouse system architecture (adopt from [21]and [22])

2.5. The main components of data warehouse system

The above figure the main component of data warehouse components that are integrated each other to perform the required function and will be discussed as follows.

2.5.1. Operational sources system

The operational sources system is mainly concerned about processing performance and availability. Generally, the source system maintains a small amount of historical data. The queries designed against source systems are narrow. On the other hand, one-record-at-a-time queries which operate as part of the normal transaction flow and act according to the demands on the operational system.

2.5.2. A Staging Area Component

The primary reason for the existence of a staging area is to ensure that all needed data is consolidated before it can be integrated into the main components of a Data Warehouse. In an active business, there exist many limitations in the hardware, network resource as well as differences in business cycles and data processing cycles which make it a challenge to extract all the data from the databases simultaneously. After extracting the data to the staging area many alterations such as cleansing the data (correcting misspelling, resolving domain conflicts, dealing with missing elements or parsing into standard format), combining data from multiple sources, duplicate data, and assigning warehouse keys take place.

2.5.3. Data presentation

The data presentation area is the place where data is organized, stored, and made available to the user. In addition, the presentation area is the place where business communities see data and gain access using data access tools.

2.5.4. Data access tools

The data access tools element is the final of the data warehouse. This element provides many capabilities for the business users to control the presentation area for analytical decision-making. Generally, the data access tools can act as simple query tool or can be complex as a data mining application.

2.5.5. The Data Warehouse Server

From the staging area by means of ETL, the data is then integrated with the various internal and external operational databases of the organization which operate across the globe. This leads to a humongous collection of detailed data. For example, the data of every sale ever recorded by a business would be convoluted which enables it to be statistically analyzed very efficiently. With such abundance of data, the organization's reviewers would not access the Data Warehouse server directly. They access only the various front-end OLAP tools that analyze subject-oriented data and represent it as Data Marts.

2.5.6. Front-end Data marts

The ETL technology allows an operation of transferring data from the warehouse to a data mart is done. Extracted data is represented on one or several Data Marts which enables it to be

accessed by the organizations reviewers. The Data Marts often showcase a multi-dimensional view of extracted data with the help of front-end Data Warehousing OLAP Tools will be used to visualize the analyzed data or information

ETL Technology

Business intelligence (BI) has gained wide recognition in different industries in order to analyze large volumes of data. One important component of BI is the Extract, Transform, and Load (ETL) process of data from sources data base to its data warehouse. It describes the gathering of data from various sources (extract), its modification to match a desired state (transformation) and its import into a database or data warehouse (load) [23]. According to some industry experts approximately 60-80 percent of a data warehousing project effort is spent on the ETL process alone [24].

Extraction-Transformation-Loading (ETL) tools are specialized tools that deal with data warehouse homogeneity, cleaning and loading problems. ETL and Data Cleaning tools are estimated to cost at least one third of the effort and expenses in the budget of the data warehouse. First, the data is extracted [25] from the source data stores that can be On-Line Transaction Processing (OLTP) or legacy systems, files under any format, web pages, and various kinds of documents (e.g., spreadsheets and text documents) or even data coming in a streaming fashion. After this phase, the extracted data is propagated to a special-purpose area of the warehouse, called the Data Staging Area (DSA) [21], where their transformation, homogenization, and cleansing takes place. The most frequently used transformations include filters and checks to ensure that the data propagated to the warehouse respect business rules and integrity constraints, as well as schema transformations that ensure that data fit the target data warehouse schema. Finally, the data is loaded to the central data warehouse (DW) and all its counterparts (e.g., data marts and views) [26].

Nowadays, business necessities and demands require near real-time data warehouse refreshment and significant attention is drawn to this kind of technological advancement [26]. The design, development and deployment of ETL processes, which is currently, performed in an ad-hoc, in house fashion, needs modeling, design and methodological foundations. The most important components during the design and deployment phase in a data warehousing is the design flow of data from the source relations towards the target data warehouse relations, this flow is provided

by the ETL tools. Extraction-Transformation-Loading (ETL) tools are pieces of software responsible for the extraction of data from several sources, their cleansing, customization and insertion into a data warehouse [26]. There are currently many commercial tools available in the market e.g. Oracle Warehouse Builder (OWB), IBM Information Server (Data stage) 9.1, SAS Data Integration Studio 4.21 SAS Institute, SQL Server Integration Services (SSIS) 10 Microsoft, Dataflow Manager 6.5 Pitney Bowes Business Insight, Clover ETL 3.0.1 Javlin, DB2 Warehouse Edition 9.1 IBM, Pentaho Data Integration 4.1 Pentaho.

Data Warehouse characteristics

The four keywords that define data warehouse are, subject-oriented, integrated, time-variant, and nonvolatile, distinguish DW from other data repository systems, such as relational database systems, transaction processing systems, and file systems.

Subject-oriented: DW is organized around major subjects, such as customer, supplier, product, and sales. Rather than concentrating on the day-to-day operations and transaction processing of an organization, a DW focuses on the modeling and analysis of data for decision makers. Hence, DW typically provide a simple and concise view around particular subject issues by excluding data that are not useful in the decision support process.

Integrated: DW is usually constructed by integrating multiple heterogeneous sources, such as relational databases, flat files, and on-line transaction records.

Time-variant: Data are stored to provide information from a historical perspective (e.g., the past 5-10 years). Every key structure in the DW contains, either implicitly or explicitly, an element of time.

Nonvolatile: DW is always a physically separate store of data transformed from the application data found in the operational environment. Due to this separation, a DW does not require transaction processing, recovery, and concurrency control mechanisms.

2.3. The Difference between Data Warehouse and OLTP Data Bases

The data warehouse and the online transaction processing OLTP data base are both relational databases. However, the objectives of both these databases are different. The OLTP database records transactions in real time and aims to automate clerical data entry processes of a business

entity [27]. Addition, modification and deletion of data in the OLTP database is essential and the semantics of the application used in the front end impact on the organization of the data in the database. The data warehouse on the other hand does not cater to real time operational requirements of the enterprise. It is more a storehouse of current and historical data and may also contain data extracted from external data sources. The following table summarized the main difference between the two data bases [15].

Table 1: The difference between DW and OLTP

Data warehouse database	**OLTP database**
Designed for analysis of business measures by categories and attributes	Designed for real time business operations.
Optimized for bulk loads and large, complex, unpredictable queries that access many rows per table.	Optimized for a common set of transactions, usually adding or retrieving a single row at a time per table.
Loaded with consistent, valid data; requires no real time validation	Optimized for validation of incoming data during transactions; uses validation data tables.
Supports few concurrent users relative to OLTP	Supports thousands of concurrent users.

2.2. Application of Data warehouse

Data warehousing technologies have been successfully deployed in many industries [28] manufacturing (for order shipment and customer support), retail (for user profiling and inventory management), financial services [29] (for claims analysis, risk analysis, credit card analysis, and fraud detection), transportation (for fleet management), telecommunications(for call analysis and fraud detection), utilities (for power usage analysis), educational system [9] and healthcare (for outcomes analysis) [14].

Currently data warehouses have being implement for governmental and non-governmental institution for different purpose such business, education, property and resources administration, market analysis and prediction, historical and large record management, decision support system and etc. The next section will discuss some of the major application areas of data warehouses:

2.2.1. Application of Data warehouse in E-government

One of the application areas of data warehouse is E-government allows governments to serve citizens in a short time, effective, and cost efficient method. E-government can provide four types of services, which are Government-to-Citizen (G2C), Government-to-Business (G2B), Government-to-Employee (G2E), and Government-to-Government (G2G) [30].

China have started first E-Government program in the late 1980s, in which the governments both at central and local levels built up office automation (OA) systems and established an intranet, subsequently the Central Government of China had formally launched five Golden Projects (Golden Bridge Project, Golden Customs Project, Golden Card Project, Golden Tax Project and Government online Project) aimed at building E-Government in China ever since 1990s. A data warehouse has been defined as a collection of data in support of management decisions which is: subject oriented, integrated, nonvolatile, time variant. The data warehouse has now been more generally seen as a strategy to bring heterogeneous data together under a common conceptual and technical umbrella and to make the data available for new operation or decision support application [31].

2.2.2. Applications in Agriculture

A NATP Mission Mode Project Integrated National Agricultural Resources Information System (INARIS) was undertaken at IASRI. In this project, a state of art Central Data Warehouse (CDW) of agricultural resources of the country is developed. This is probably the first attempt of data warehousing of agricultural resources in the world. This will provide systematic and periodic information to research scientists, planners, decision makers and developmental agencies in the form of On-line Analytical Processing (OLAP) decision support system. In addition, the system also provides the facility of spatial analysis of the data through web based using functionalities of Geographic Information System (GIS) [28].

2.2.3. Application of data warehouse in education

Universities are encountering growing demands by policymakers and communities who are challenging for valuable information [32] about student achievement and university system accountability. The top manager of the universities required to measure annual progress for every single student and their academic status [33]. Using the integrated data as well as data warehousing and online analytical processing application procedures, automatically or semi-

automatically provide suggestions to improve teaching and learning process for teachers as well as the students [34]. An operational database in which the data concerning students, professors, courses, curriculum are being stored. The database includes the appropriate data in the department to operate efficiently and effectively. However, the operational database suffers from the lack of recording past data [35]. Therefore, every time the database describes the current status of the information concerning the department. That has as consequence, that during the execution of an update or deletion process, the previous information is vanished and there is no appropriate action in order to retrieve it. The proposed system constitutes an integrated platform for a thorough analysis of department's past data. Analysis of data could be achieved with OLAP operations. Star-schema is used for dimensions modeling in the proposed system. OLAP server used as well in order to process the dimensions, to construct data cubes and to execute the OLAP operations. Finally the software package Microsoft SharePoint Server 2007 is used as presentation server. The researcher used statistical framework that enables the longitudinal study of the students' performance in the department and particularly facilitates the search for factors that may affect their performance [36].

2.7. Challenges of Building Data Warehouse

Building data warehouse will give competitive advantage for decision makers, but building such system will have numbers of challenges. All data warehousing projects do not pose same challenges and not all of them are complex but they are different. The following issues are identified from the problem statement and literature review as the main challenge in data warehouse designing and implementing process.

- ✓ Ensuring acceptable data quality: data are duplicated across systems and disparate data sources add to data inconsistency. All these issues lead to data quality challenges. Resolving these issues and conflict become due to limited knowledge of business users outside the scope of their own systems.
- ✓ Ensuring acceptable performance: most of the time designer and users often forget about the performance when they first conceive the plan to implement a data warehouse. Achieving the performance objectives is not easy. Performance is directly dependent on the complexity of the system which, in turn, depends on the design.

- ✓ Testing the data warehouse: testing in data warehouse is real challenges. Because of high dependencies, regression testing requires lot of planning. Making the data available for testing for certain component may not be possible as fresh data loading often changes the surrogate keys of the dimension table thereby breaking the referential integrity of the data. Thus continuing fresh testing along regression testing becomes impossible.
- ✓ Reconciliation of data: The process of ensuring correctness and consistency of data in a data warehouse. Unlike testing, which is predominantly a part of software development life cycle, reconciliation is a continuous process that needs to be carried out even after the development is over.
- ✓ Cost of development: developing data warehouse from scratch is time consuming and very expensive.

Related works

Several researches have been conducted to implement data warehouse in different sectors to gain competitive advantage. In this section, we present a brief discussion about some relevant approaches.

Christopher A. [37] designed Prototyping an Academic Data Warehouse: Case for a Public University in Kenya. According to the researcher the main objective of this study was to determine the gaps between top decision makers and IT personnel in accessing, analyzing and reporting data. The authors defined that business intelligence technologies are reporting, online analytical processing, analytics, data mining, process mining, business performance management, text mining, and predictive analytics. They indicated organizations challenged when implementing Data Warehouse technologies. This includes management (management commitment and support, project management, user involvement and participation, skills) and technological (selection of DW architecture, creation of the enterprise schema, data integration and scalability, data quality, design of human-computer interfaces, data mining, security and privacy risks, and networks and telecommunication). Research Design (A survey was conducted to ascertain the gaps between top decision makers in the college and IT department), Survey Analysis (An analysis was done to determine the understanding and perception of the college users) and A star schema (contains a central fact table surrounded by many dimensions tables) methodologies are employed in this research. Data Transformation Services (DTS) tool has been used to extract data from various sources. The study successfully implemented a data warehouse database composed of a single multidimensional cube based on subject area: Fee Payment. This

database brought together data from multiple sources providing principals and decision makers of the college greater insight into the college financial performance.

Chapter Three

3. Research Methodology

A Methodology in software development describes the expected evolution and management of the system development process. Data warehousing methodologies share a common set of tasks, including business requirements analysis, data design, architecture design, implementation, and deployment. The proposed decision support data warehouse system is build based on a spiral model paradigm of the software engineering, in which prototype approach of software development is applied.

In this project Prototyping is chosen as suitable approach for the proposed decision support data warehouse design. This approach is more amenable to a phased development which focuses on selected dimensional subjects, which are much smaller in scope and complexity than the requirements for an enterprise wide data warehouse of the university.

The main advantages of prototyping, with particular reference to a data warehouse project, can be summarized as follows:

- Prototypes help designers to validate requirements, because they allow users to evaluate designers' proposals by trying them out, rather than interpreting design documents. This is particularly crucial to enable a better understanding of hierarchies by users.
- Prototypes are especially valuable to improve the design of reports and analysis applications, due to their interactive nature.
- Prototypes can be used to advance testing to the early phases of design, thus reducing the impact of error corrections. For instance, an early loading test can be effectively coupled with a preliminary functional test of front end applications to check for correct data balancing.
- Prototypes can be used to evaluate the feasibility of alternative solutions during logical design of multidimensional schema and during ETL design.

Moreover, building the data warehouse starts with identifying the key business processes and the key business questions that the data warehouse is needed to answer. The main dimensions (including departments, courses, students, and academic staff) of the data

warehouse are identified to build decision support data warehouse system. Meanwhile the source data of the identified dimensions are captured, analyzed and documented.

3.1. Development Phases

The proposed decision support system follows Iterative (Spiral) approach to develop the first prototype decision support data warehouse. One of the advantages of spiral (iteration) methodology allows the possibility of revisiting any activities to correct errors/incorporate new requirements. So that some activities may be repeated till the entire expected output is produced. To develop the proposed decision support data warehouse system, the following main activities are executed efficiently as presented in figure 2 below.

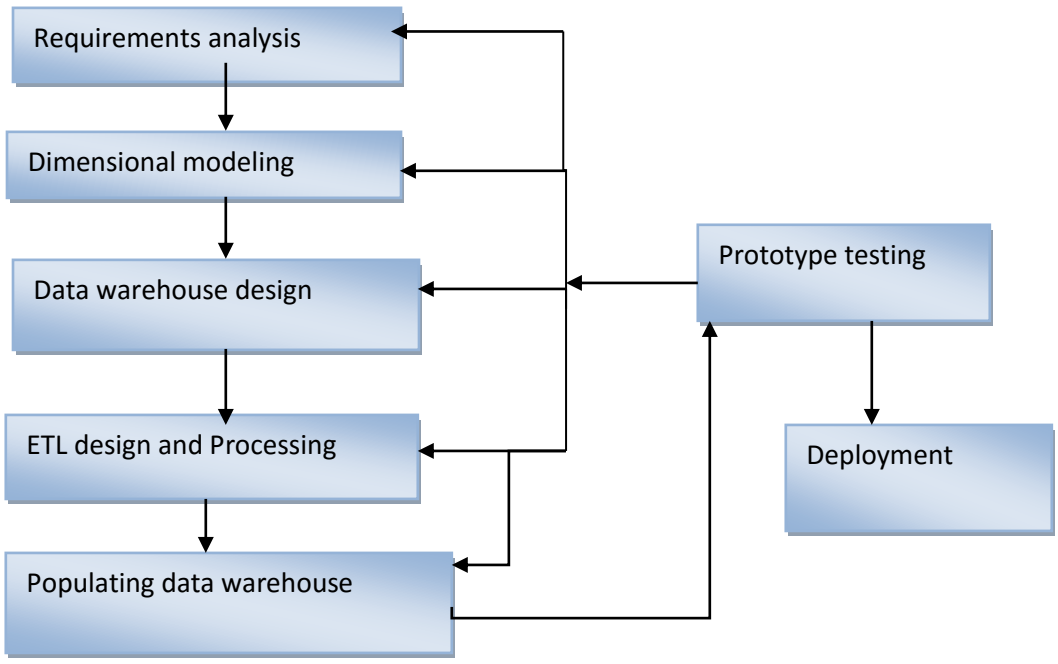


Figure 2: Decision support data warehouse development phases

3.2. Requirements Gathering

Various types of requirements elicitation strategies has been used in practice, ranging from standard systems development life-cycle techniques such as interviews and observation of the working environment. In this study the user requirement are gathered through interviewing the concerned office and observation techniques are also employed.

Interview

An interview has made with stakeholders to elicit or validate needs and requirements of stakeholders. To gather functional requirements ASTU registrar staff participated in the interview. Initially the purpose and function of the research introduced for the registrar co-deans, he direct our request to data encoder for detail discussion how the business process going on. The discussion also includes challenging issues such as data quality, data sources uniformity, absences of corporate data management system, existed system also described by encoder. Finally she provide us all the raw data as per our request. During the discussion we tried to identify to gap and the challenges on the existing system. In particular three key people from administrative staffs and associate registrar officer were interviewed about their requirements and the current systems. During the interview functionalities required by the stakeholders have been identified.

Observation

To understand the user interactions, business processes and components of the system and observation of current system has been done. The existing system was developed using Microsoft access database management system. The database serves for academic grade recording and generating grade reports.

Document analysis

In addition to interviewing the stakeholder and observation of working flows, relevant document like curriculum of different department have been analyzed. Coursed and school dimension table are generated from this document. The other sources of data gathering is the university academic affair office to collect staff (instructor) data which is important to correlate with department and course ratio.

3.3. Requirements analysis

Requirements analysis is an important phase of data warehouse design to identify which information is relevant to the decision making process by considering the user needs or the actual availability of data in the operational sources. Hence requirements analysis is applied to identify and determine information requirements of users, how are these requirements

transformed and organized. It is also helpful to analysis the functional requirements that will define the functionalities that the data warehouse system will provide users to accomplish their tasks, thereby satisfying the business requirements. According to the interview and observation with domain experts the following problems are specified.

Problem one

The current system does not support school deans and registrar officers to generate reports that show:

- The number of students succeed in each semester in each department
- How many of them are failed in each department?

Problem two:

The due to its transaction operation, current system unable to support the university decision makers:

- To analyzing the number of academic staff and their profile in each department according to their qualification and specialization
- To analyze the total number of active students attending their study at each academic year.
- To evaluate the performance of student in each academic year by departments, gender of students.

In addition to functional requirements, the main requirement of the study is information requirements. So that entity types, the attributes of each entity type and the number of entity sets are identified during the interview phase. As the objective of the study is to design a data warehouse that supports the managers and school deans and registrar officers to make strategic and timely decision, the main entities integrated in the system are student, academic staff, departments and courses offered in each department as well as in the university. The appropriate dimensions and keys, descriptive attribute are identified based on the scope of the proposed decision support data warehouse system.

The data structure of the dimensions and fact tables is described in the following table as fact and dimension tables.

Table 2: Dimensional table's description

Fact Table	Fact Table keys	Dimension Tables	Dimension Attributes
AcademicFacttablea	StudentKey, DeptKey, CourseKey, InstructorKey,	Student dimension	StudentKey,Name, Gender, IDNO, Region, EntryYear
		Department dimension	DeptKey, DeptName, Address
		Course dimension	Coursekey, CourseCode, Title, CreditHrs, Prerequisite, Compulsory
		Instructor dimension	InstructorKey, Gender, HireDate, AcademicRank, Specialization
		Time dimension	Timekey, year, semester

3.4. Data Warehouse Modeling

Once the requirements are captured, a data warehouse model is created based on those requirements. The analysed objects and business process are represented logically using dimensional modelling technique.

3.4.1. Dimensional Model

Dimensional modeling is powerful in representing the requirements of the business user in the context of database tables. The dimensional data design model used in data warehouse is much more effective for querying than the relational model used in OLTP systems.

Multidimensional modeling is a technique for conceptualizing and visualizing data models as a set of measures that are described by common aspects of the business. It is especially useful for summarizing and rearranging the data and presenting views of the data to support data analysis. The multidimensional conceptual view of data is characterized by representing data as if placed in n-dimensional space, allowing us to easily understand and analysed data in terms of facts (the subjects of analysis) and dimensions showing the different points of view where a subject can be analysed from.

Dimensional modeling uses three basic concepts: measures, facts, and dimensions. A dimensional model is also commonly called a star schema. The star schema is the dimensional model based on a fact table at the centre and its associated dimensions. In this model, the fact table consisting of numeric measurements and foreign key columns joined to a set of dimensional tables filled with descriptive attributes.

To design the star schema for the data warehouse, the measures and dimensions are identified from domain expert. The identified data is specified using dimensional modeling technique. Star schema is employed to design the academic data warehouse. In a star schema, related dimensions are grouped as columns in dimension tables, and the facts are stored as columns in a fact table. The star schema has two types of tables, fact tables at the center of the star and dimension tables at the points of the star as indicated in the figure 4 below. Therefore, the dimension tables of academicDW include student, course, instructor, and department. Thus the dimensional models are defined by a Surrogate key as the Primary Key.

The fact table of academicDW is designed to join all dimensional tables to facilitate the analysis operation.

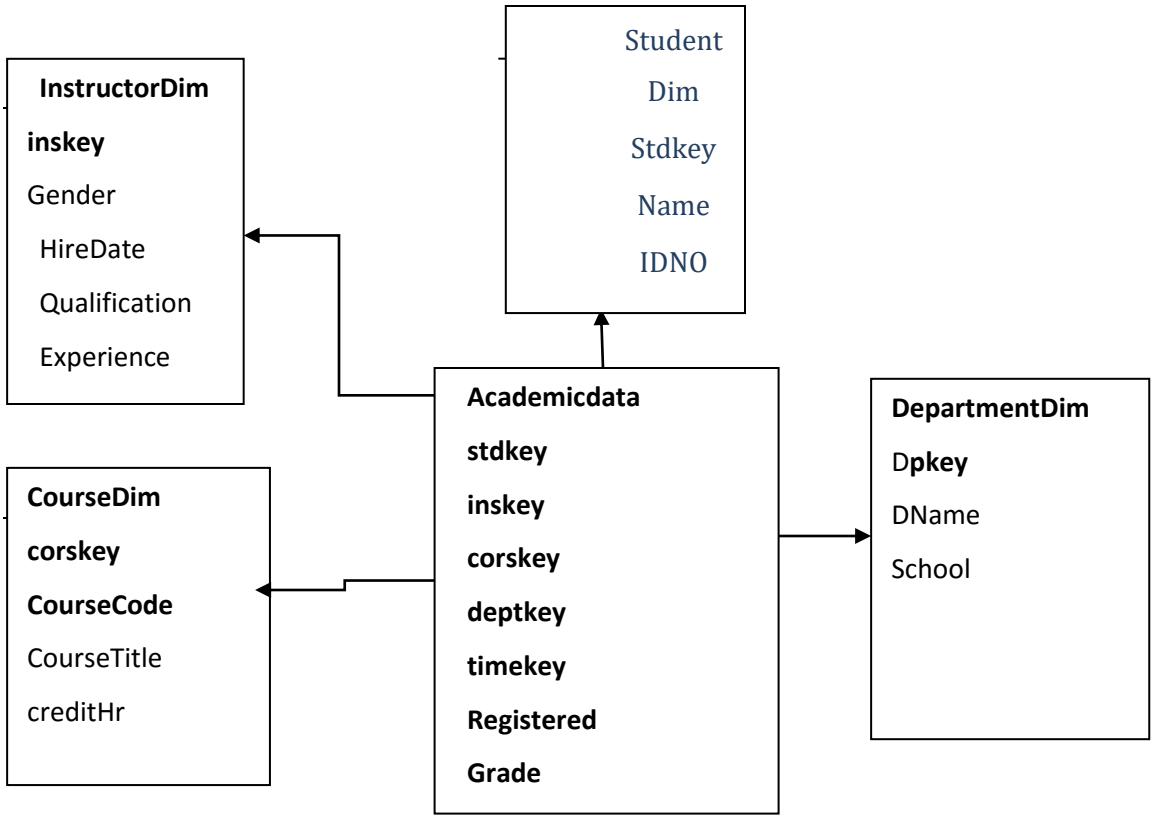


Figure 3: Data warehouse star schema

Chapter Four

4. Data warehouse Design

Designing the data warehouse structure is different from designing the operational systems. The operational systems consist of simple pre-defined queries. On the other hand, in data warehousing environments queries join with more tables and more computation time and informality. The overall architecture of the proposed decision support data warehouse is designed having the following main components. The components are the source data, the ETL processing, data warehouse, OLAP and the Reporting environments. The architecture shows where the data is extracted, how processed task is performed, how the sources data is populated on fact table and make use the data warehouse for analytic decision making and reporting.

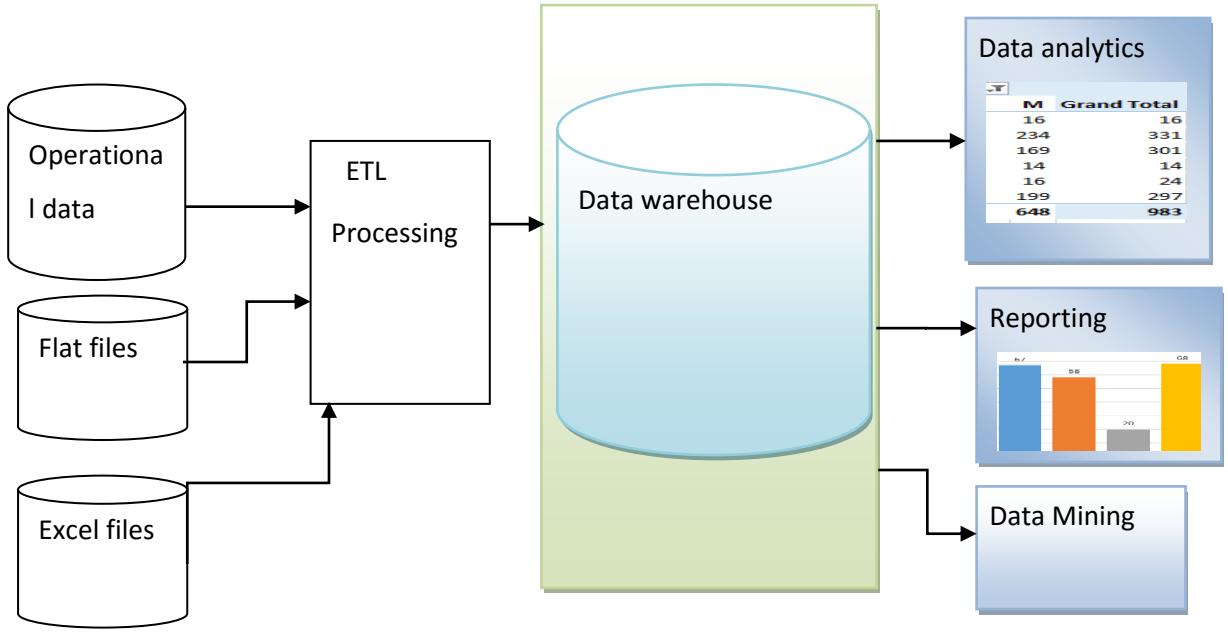


Figure 4: Proposed data warehouse system Architecture

The main components of the proposed data warehouse system discussed as follows:

Data sources

A data source component represents all of the sources from which the raw data originate. The data source components feed into the data acquisition component, which evaluates the integrity of quality of data. When programming the data acquisition component, rules must be defined in order to ensure the acquisition process occurs properly. The acquisition process involves cleansing (manual cleansing such as filling missing row and column, removing duplicated data), enhancement, restructuring (standardizing the data as per system requirement), integration, and

aggregation of source data steps necessary in order to generate useful, accurate data. The required data for the decision support data warehouse system is selected and structured from different sources. The sources data for this prototype decision support data warehouse system are found in different formats (i.e. Microsoft word text files and excel format).

ETL Technology

ETL is a data integration function that involves extracting data from outside sources (operational systems), transforming it to fit business needs, and ultimately loading it into a data warehouse. To solve the problem, companies use extract, transform and load (ETL) technology, which includes reading data from its source, cleaning it up and formatting it uniformly, and then writing it to the target repository to be exploited.

ETL tools aggregates, consolidates, cleanses and finally validates the data so it can be used effectively for intelligent decision. The use of ETL tool increases the productivity associated with the complexities of load balancing, logging, distribution of data, scalability of system and interfaces. It is because of the ETL tools that large bytes of data (as big as gigabyte) are accessed at a time.

Extract, Transform, Load; three database functions that are combined into one tool that automates the process to pull data out of one database and place it into another database. The details ETL process are described as follows:

1. Extract

The goal of the extraction phase is to convert the data into a single format which is appropriate for transformation processing. Each separate system may also use a different data organization/format. Most of the time the data in source system is very complex, thus determining which data is relevant is very difficult. Designing and creating extraction processes is a very time consuming programming effort. During data extraction phase the main challenges are data quality, missing data attribute and un proper organization of data. To keep the data up to date in data warehouse, data has to be extracted several times in a periodic manner.

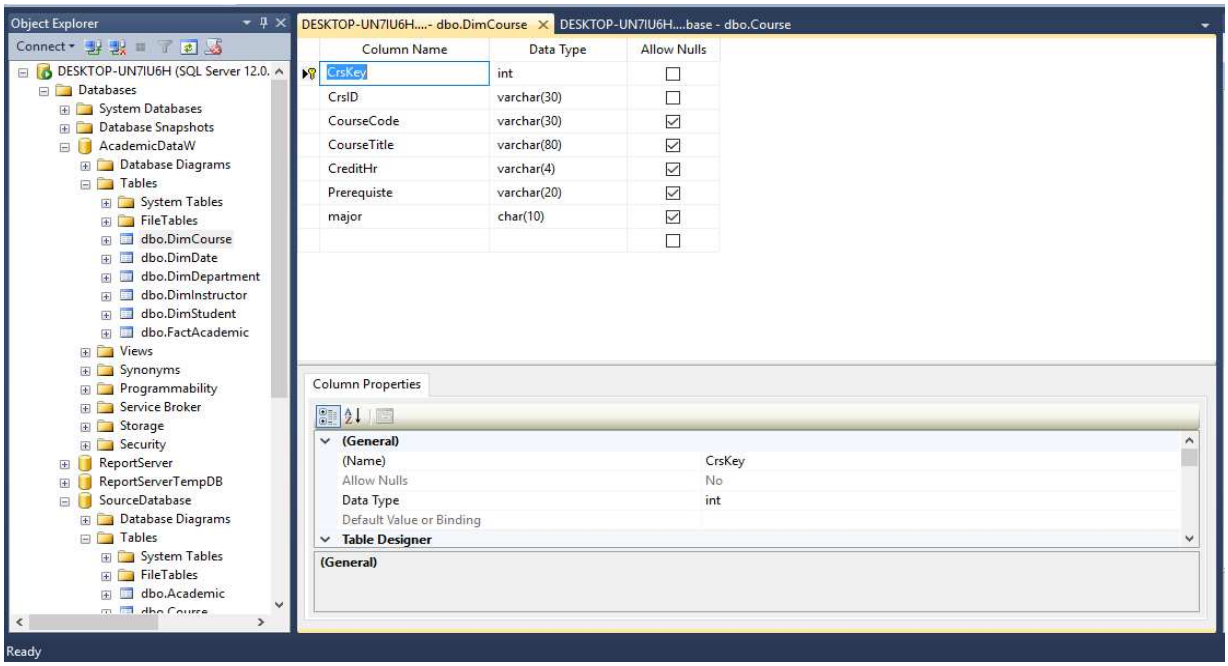
2. Transform:

The transform stage applies a series of rules or functions to the extracted data from the source to derive the data for loading into the end target. Some data sources will require very little or even no manipulation of data.

The source data for course dimension is extracted from Microsoft word document and converted into spreadsheet files for structuring and easy loading into the target data warehouse dimensions. The hire date column of instructor data was created in different format. Example 09/6/2004 and 01-12-1997 values are existed in one column. Hence these types of records are converted into same format that SQL server data type will allow to be loaded in the destination data warehouse. The following transformation has been done successfully:

- ✓ Selecting only certain columns to load,
- ✓ Translating coded values (e.g., if the source system stores 1 for male and 2 for female, but the warehouse stores M for male and F for female);
- ✓ Encoding free form values (e.g., mapping "Male" to "1");
- ✓ Deriving a new calculated value (e.g., count total or aggregate/registered numbers of student);
- ✓ Sorting;
- ✓ Joining data from multiple sources (e.g., lookup, merge) and duplicating the data;
- ✓ Aggregation like count
- ✓ Generating surrogate key values for dimension table;
- ✓ Transposing or pivoting (turning multiple columns into multiple rows or vice versa);
- ✓ Splitting a column into multiple columns;
- ✓ Lookup and validate the relevant data from tables or referential files for slowly changing dimensions;

The following figure show data conversation process before loading in to the data warehouses. In this step we performed determination of data types, determination of data size to minimize the truncation problem, changes of keys and finally allowing or not allowing the null values. If the types and size of data differ between the sources database and data warehouse, it will create fatal error. This conversion processes are done for all dimension and fact table.



3. Loading Process

The final step in the ETL process is loading data into a temporary data store where it is cleaned up and made consistent. Consistency checks are executed only when all the data sources have been loaded into the temporary data store. Before the data actually loaded in to the data warehouse performing optimization task will enhance query efficiency.

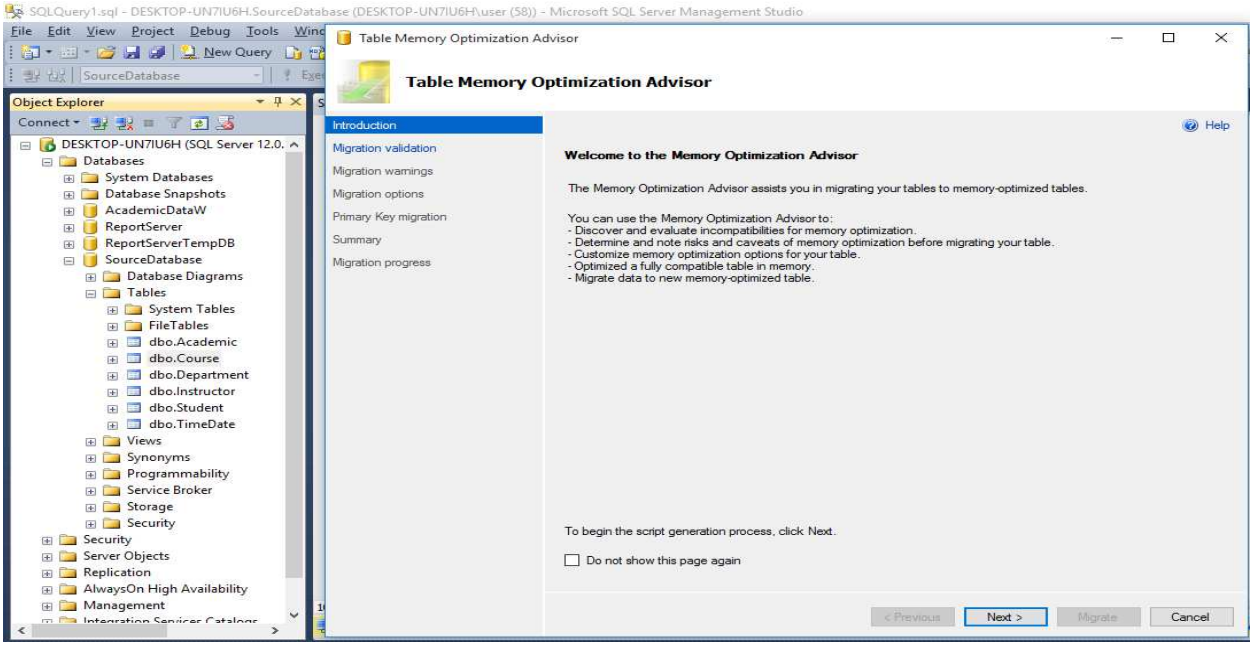


Figure 5: Table memory optimization

Table memory optimization is one important aspect to utilize scarce storage resources of the server. Memory issues are rise when the size of data sources are increased exponentially. Therefore, proper utilization is not an option but is mandatory unless you compromise it.

4.2. Data warehouse Implementation phase

4.2.1. Implementation Tools

To implement the proposed prototype data warehouse decision support system Microsoft SQL Server 2012 have been used. This software have many robust feature for business intelligence application domain. The application has the relational database management system that is capable of storing all the data required for the data warehouse designing. It has the functionality that can extract data from different sources and consolidate it into one single location for better analysis. The following Microsoft SQL Server tools are used to build the proposed decision support data warehouse system.

1. **SQL Management Studio:**

   Includes both graphical and scripting tools for the managing of all the components within the Microsoft SQL Server 2012. SQL Management Studio scripting tool is used to create all dimension and fact tables.

2. **SQL Server Integration Services:**

   Provide a platform for data integration, extracting, transformation, and loading. Implementation of ETL processes is enabled by easy-to-use graphical editor. In ETL integration stage the extracted and transformed sources data is loaded into the data warehouse; the source data are (excel files of student data and plain text files like curriculum or staff data). SQL server 2012 integration service is used to design the ETL process and mapping of the source data with the designed dimension tables.

3. **SQL Server Analysis Services:**

   This platform allows to perform the tasks online analytical processing depending on the built data warehouse. OLAP framework and Analysis services allow data warehouse designer in the implementation and processing of multidimensional data structures. The software evaluate all dimension to provide analytical information for the decision maker. In addition, this platform provide feature of integrating data warehouse with data mining technology for purposes of trends/pattern discovering in current data, analytical services offer functionality for designing and implementation of such models from the given data.

SQL Analysis Services (SSAS) includes a group of capabilities for OLAP as well as Data Mining; the method of extracting and finding patterns from available data.

4. **SQL Server Reporting Services** this package offer all functionality for report creation from several data sources, its publishing, and delivery to users.

4.3. Data Warehouse Logical Design

Logical design is the phase in which the conceptual dimensional model of the data warehouse is translated into a logical schema according to the target logical dimensional database, considering the expected workload and integrity constraints and querying performances of the data warehouse. It takes the conceptual schema described in chapter three and we created a corresponding logical schema on the chosen logical model. While nowadays most of the DW systems are based on the relational logical model (ROLAP).

After the logical design defined properly the source data have been load into the destination data warehouse, all dimensional schema and fact tables are created using Microsoft SQL Management Studio on the database server. Relationships between fact and dimension tables are defined by specifying foreign keys integrity constraints.

Dimension Tables

A dimension is a structure, often composed of one or more hierarchies, that categorizes data. They are normally descriptive, textual values. Several distinct dimensions, combined with facts, enable users to answer business questions. In this study the dimension tables defined for the proposed decision support system are CourseDim, DepartmentDim, StudentDim and InstructorDim. Each dimension tables, describing a different aspect (subject) of the fact and joined to the fact table through foreign key.

To maintain the data warehouse integrity, a fact table consists of the foreign keys to all the dimension tables in the star schema and facts that are numerical business measurements for aggregation. Measures are the values to be aggregated when queries group rows together. The primary keys of the dimension table are auto generated surrogate keys to uniquely identify a record in a dimension table and also defined in the fact table as foreign key that helps to join dimension tables to fact table. A fact table is the primary table in a dimensional model where foreign keys and measure columns are stored. Surrogate keys can be derived from the existing natural keys or it can be a simple ETL generated integer number provides the means to

maintain data warehouse information. One of the simple way to improve performance of queries processing is to use surrogate keys. In this research we have been used ETL generated surrogate keys. Figure 3.3 shows SQL scripts to create the destination dimension schema.

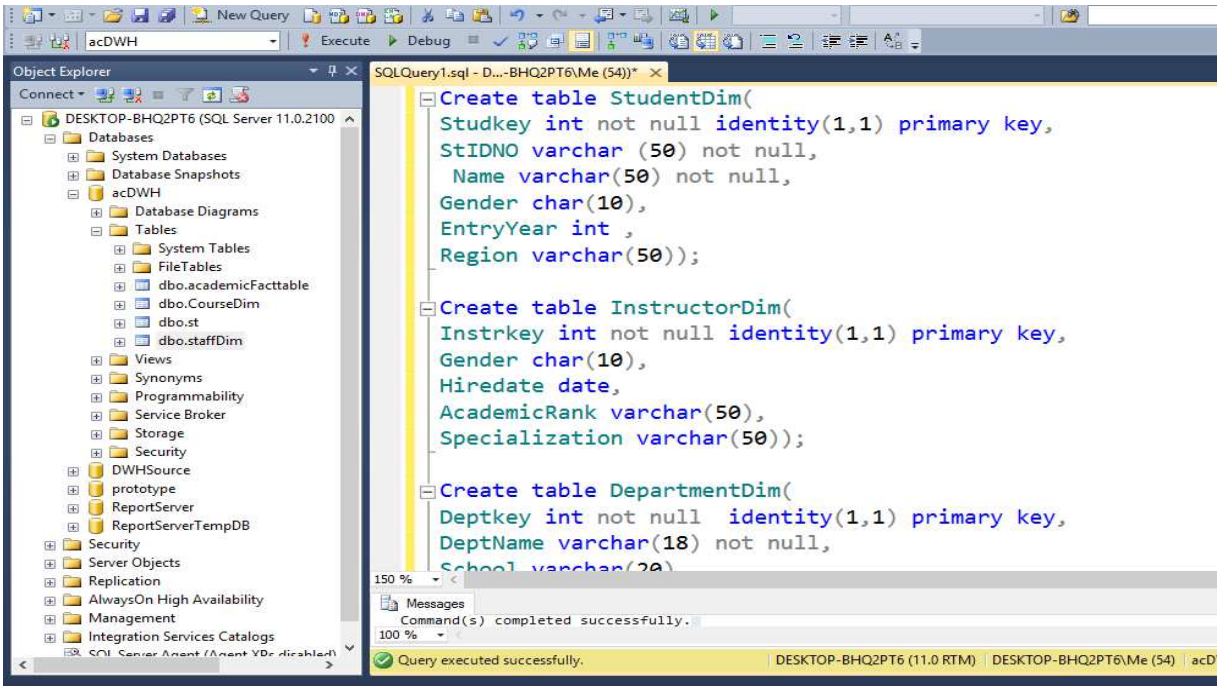


Figure 6: SQL script to define dimension tables

The next step of building the data warehouse is the fact tables integrate dimension keys and relevant measures within the table. We have single measure in the fact table to count the total aggregate of registered student, but it is possible to increase the numbers of measures. The proposed system has one fact table linked with each dimensional table to make analysis. In figure 5 above we define SQL Server scripts to create academic fact table.

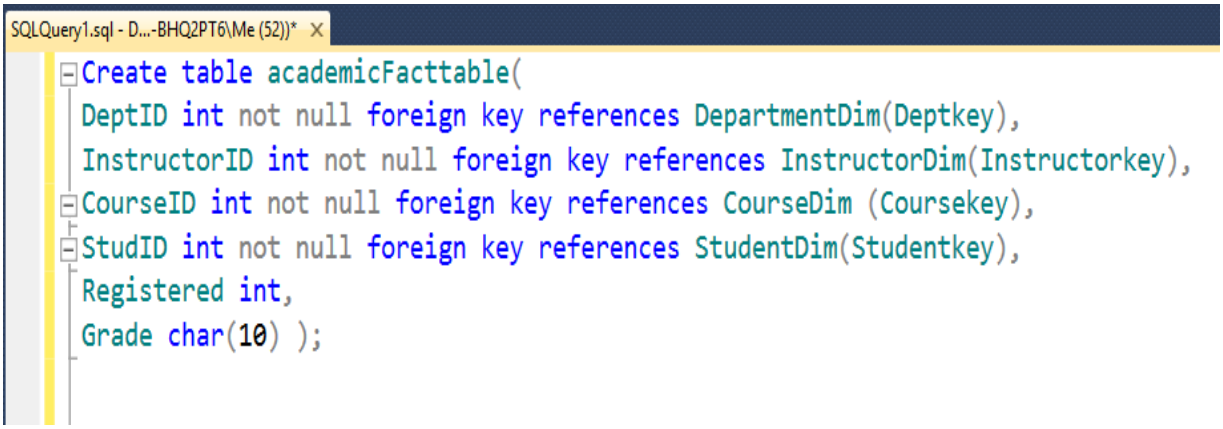


Figure 7: Dimension and fact tables' integration

The successful executions of the above fact table SQL Server script sources code, it will generate the proposed star schema as it depicted in figure 7 below. The SQL data definition code guarantees integrity of the data warehouse. Designing a relationship between the fact table and the dimension tables used to perform aggregate and OLAP operations (drill across, drill down rollup) analysis in all dimensions or some selected columns.

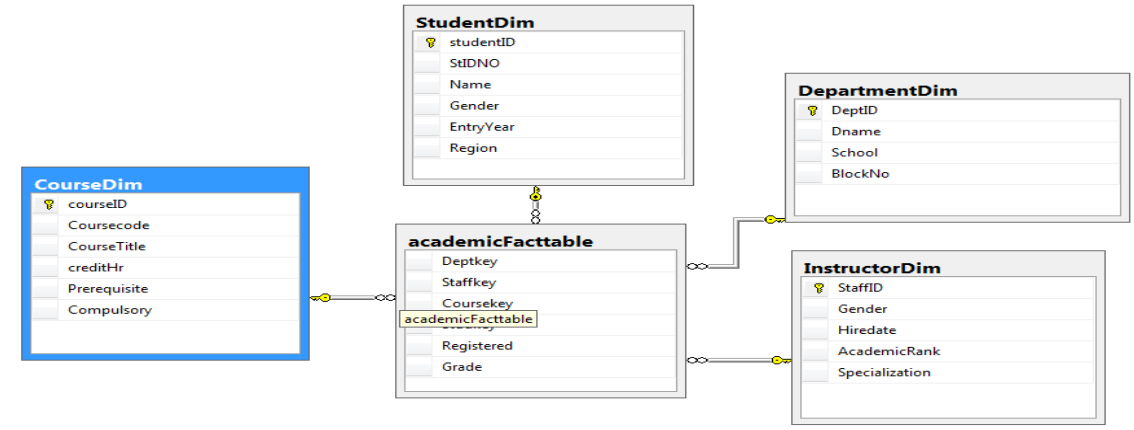


Figure 8: Diagram of data warehouse star schema

One of the most important parts of the data warehouse is the extracting, converting data types and column sizes and loading of data from the operational transactional databases to the data warehouse. In mapping function, column size and data type is converted into the same size and data type as per server configuration requirement in order to load the source data into the target destination. At this step we created empty sources database for both dimension and fact table. Once this step completed the next step is the major task of data warehouse designing process that is SQL Server Integration Service (SSIS).

An SSIS package was created and executed for loading the source data into the data warehouse. The two major task in SSIS are working on control flow and data flow tasks.

1. Control flow task: SSIS tasks are the foundation of the Control Flow in SSIS. When you are on the Control Flow design surface in SSDT, the SSIS Toolbox is populated with a set of task components that can be snapped together to represent a workflow for your package. To add a task to a flow, click and drag it from the SSIS Toolbox onto the design surface.
2. The Data Flow task encapsulates the data flow engine that moves data between sources and destinations, and provides the functionality for transforming, cleaning, and

modifying data as it is moved. The Data Flow task is where most of the work of an extract, transform, and load (ETL) process occurs. We takes data from the specified source data, map source data and the destination and then loads it into the specified destination dimensions. This Step by step process have been performed for all dimension table. Figure 8 presents the process of checking whether the source and the destination data connected by configuring the OLE DB connection manager to load the required data into the destination.

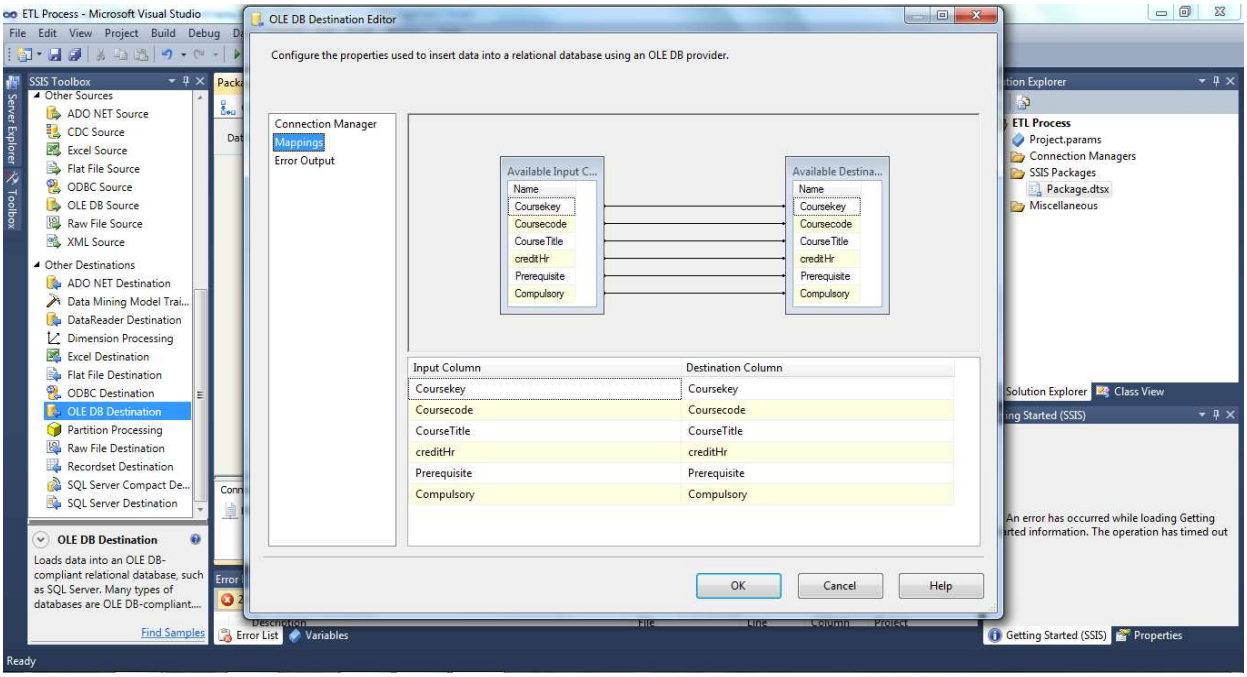


Figure 9: Mapping of the source and destination data

After checking whether the source and destination are mapped correctly, the ETL project is executed. If the connection manager correctly defined and configured then the execution result moves the records from data source into the destination data warehouse dimension. Figure 9 presents the follow of SSIS package execution completed successfully.

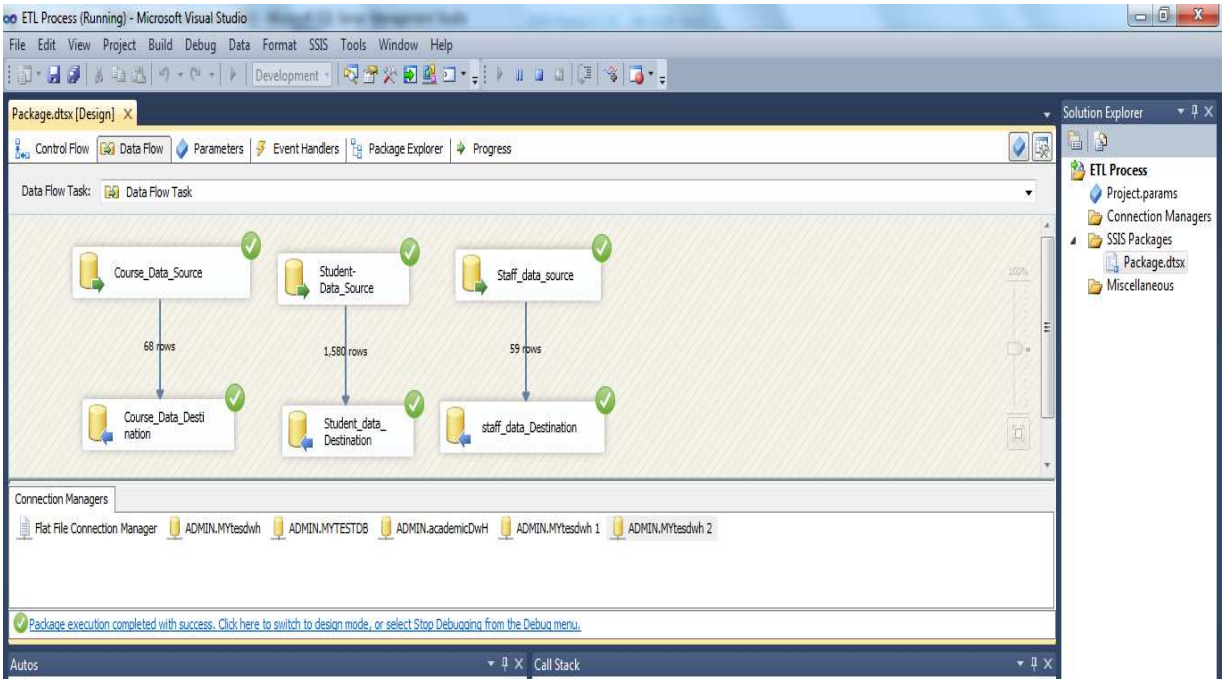
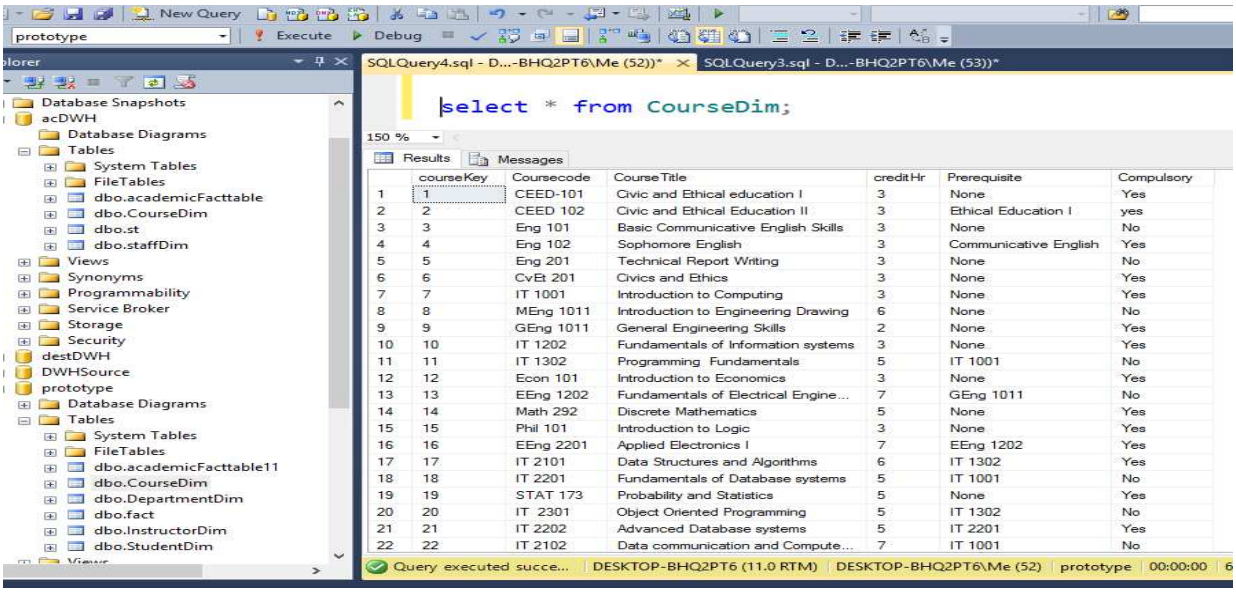


Figure 10: SSIS process to populate data warehouse

Using SQL Select statement is used to check whether the course destination dimension is properly populated with records. From the figure above you can see that how many rows of data are transferred from the source to destination. In addition, the green thick arrow shows successful of executions of packages. On the other hands, figure 10 shows the loaded sources data base on the destination dimension. You can check the loaded data with simple select SQL statement.



	courseKey	Coursecode	CourseTitle	creditHr	Prerequisite	Compulsory
1	1	CEED-101	Civic and Ethical education I	3	None	Yes
2	2	CEED 102	Civic and Ethical Education II	3	Ethical Education I	yes
3	3	Eng 101	Basic Communicative English Skills	3	None	No
4	4	Eng 102	Sophomore English	3	Communicative English	Yes
5	5	Eng 201	Technical Report Writing	3	None	No
6	6	CvEt 201	Civics and Ethics	3	None	Yes
7	7	IT 1001	Introduction to Computing	3	None	Yes
8	8	MEng 1011	Introduction to Engineering Drawing	6	None	No
9	9	GEng 1011	General Engineering Skills	2	None	Yes
10	10	IT 1202	Fundamentals of Information systems	3	None	Yes
11	11	IT 1302	Programming Fundamentals	5	IT 1001	No
12	12	Econ 101	Introduction to Economics	3	None	Yes
13	13	EEng 1202	Fundamentals of Electrical Engine...	7	GEng 1011	No
14	14	Math 292	Discrete Mathematics	5	None	Yes
15	15	Phil 101	Introduction to Logic	3	None	Yes
16	16	EEng 2201	Applied Electronics I	7	EEng 1202	Yes
17	17	IT 2101	Data Structures and Algorithms	6	IT 1302	Yes
18	18	IT 2201	Fundamentals of Database systems	5	IT 1001	No
19	19	STAT 173	Probability and Statistics	5	None	Yes
20	20	IT 2301	Object Oriented Programming	5	IT 1302	No
21	21	IT 2202	Advanced Database systems	5	IT 2201	Yes
22	22	IT 2102	Data communication and Compute...	7	IT 1001	No

Figure 11: SQL viewing of the data from data warehouse course dimension

After the destination dimensions successfully populated, the remaining task in SSIS is population data from dimensional table into fact table in the data warehouse system. To populate the dimension table into fact table, first we from data flow task use the lookup and perform connection manager to map the source data base with destination data warehouse. The other important task at this stage is configuring the sources, and column size to maintain truncation errors. Hence surrogate key column values of all dimension tables are extracted and loaded into fact table of the data warehouse. The following figure shows the successful population of dimension table into data warehouse fact table. This is also the end of SSIS in data warehouse designing implementation process.

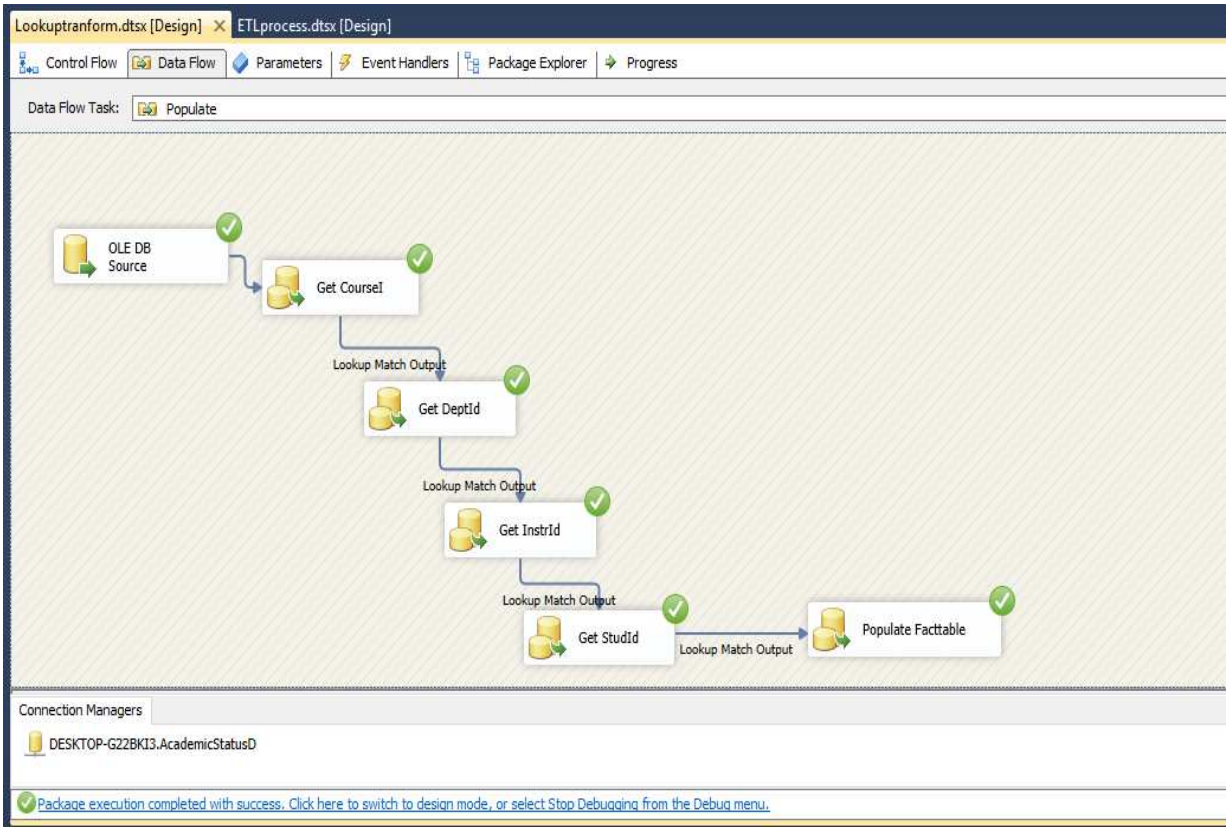


Figure 12: Populating of fact table

4.4. On-line Analytical Processing

After the SQL Server Integration Service successfully completed, the next important step is performing the analytical process. This is the stage where we prepare our data warehouse to perform analytical output by referencing the existing dimension tables. OLAP (On-line Analytical Processing) one of the tools in SQL Server that used for ease information analysis and navigation all dimension in data warehouse in order to extract relevant knowledge of the organization. An analysis services project is used to design, deploy and process OLAP dimensions. Data for analysis purpose is extracted or transformed from populated dimension and fact table in data warehouse. When you successfully create OLAP project then it will automatically generate data source, data sources view, cubes, dimension and mining structure. This is important step to configure all OLAP components step by step to get required analytical output. Miss configure of some of the components will cause fatal error, so proper handling of the server is important.

4.4.1. Deploying the Cube and processing

In deploying the data cube for dimensional analysis, the first task is to create a data source. The data source contains the information that Analysis Services uses to connect to the prototype data warehouse source database. It contains the data provider name, the server and database name, and the authentication credentials that Analysis Services will use. In addition to the data source, data source view is a logical representation of dimensions and measures in fact table used by Analysis Services objects is also created. Figure 12 shows the design of data cube.

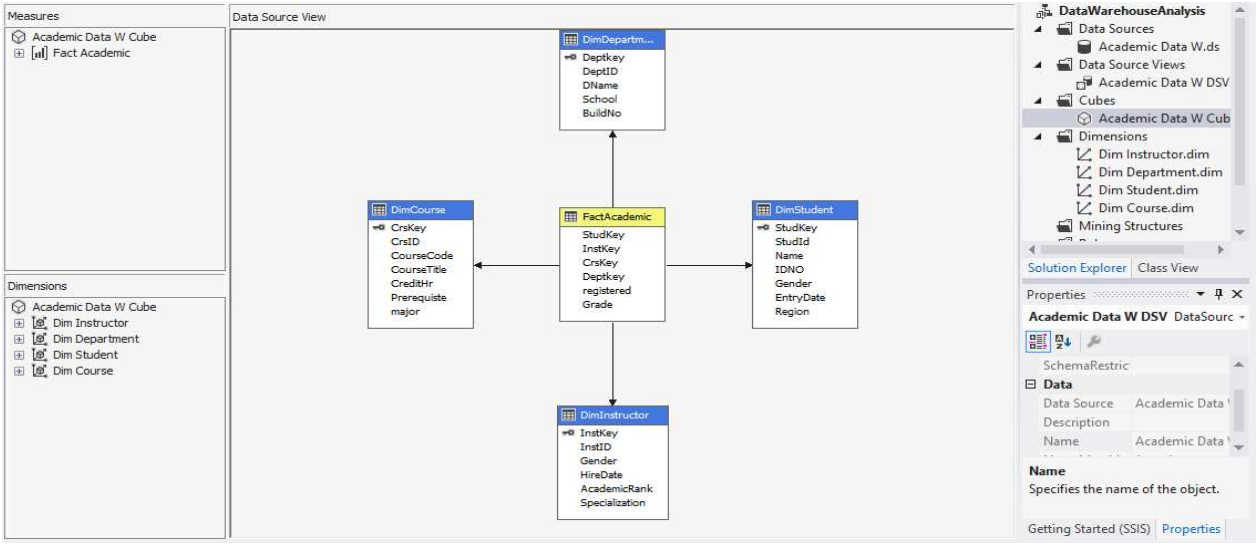


Figure 13: Designing data warehouse cube

4.4.2. OLAP and Data mining integration

The other potential features of data warehouse is the capability of integrating with data mining technology to discover a new knowledge from existing data. A mining structure is a data structure that represents discovered knowledge based on analysis of OLAP or relational data. A mining model is can be used to make predictions it supported by the data mining techniques used to create the mining decision tree model. But, it is possible to test other data mining algorithm to evaluate its performance and efficiency. The Microsoft decision tree algorithm is one of the classification algorithm that supports the prediction of both discrete and continuous attributes. The following figure shows the step by step integration process of data warehouse and data mining.

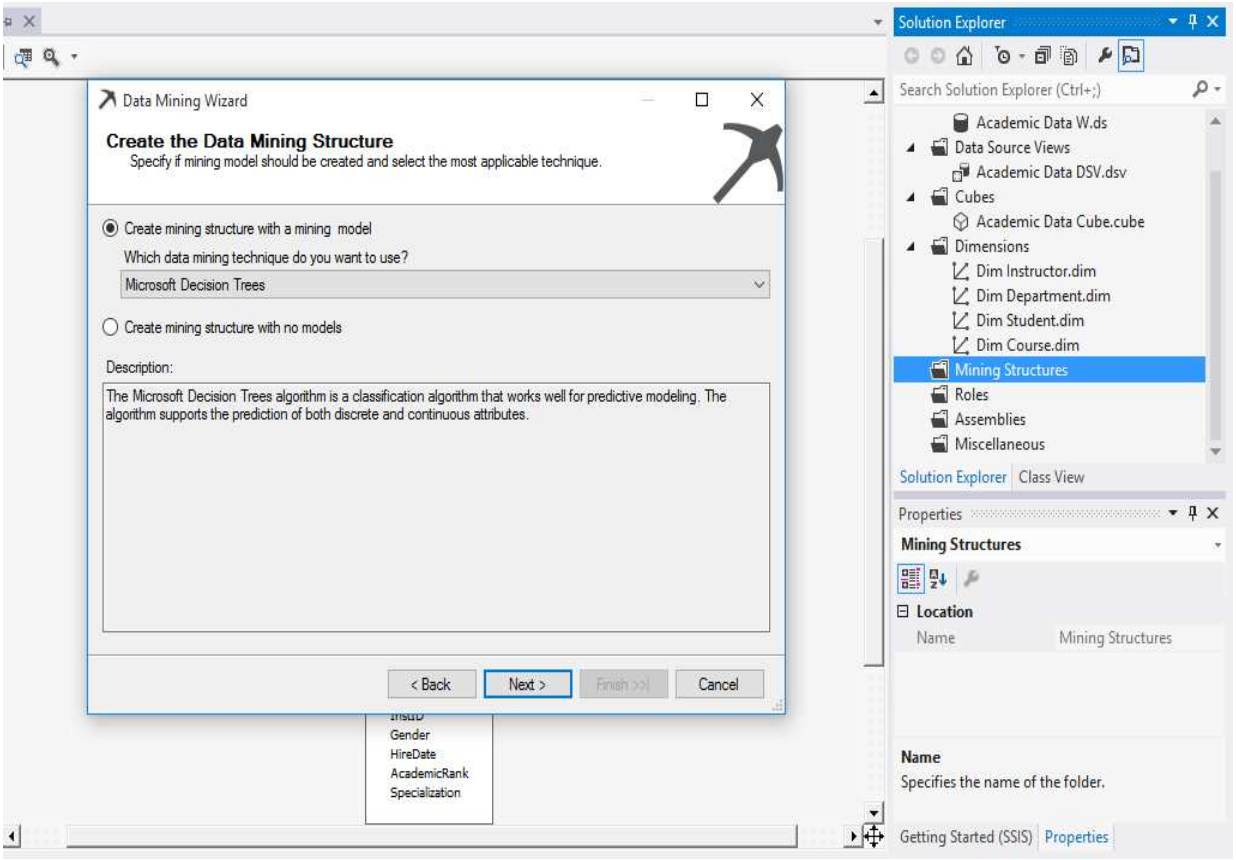


Figure 14: creating data mining structure

After determining the training data set, the next step is to specify for model testing. The input data from reserved data warehouse randomly split into a training set and a test. Based on the percentage of data for testing and maximum number of cases in the test data set you provided. The next figure show the process of selecting those keys to specify the training data

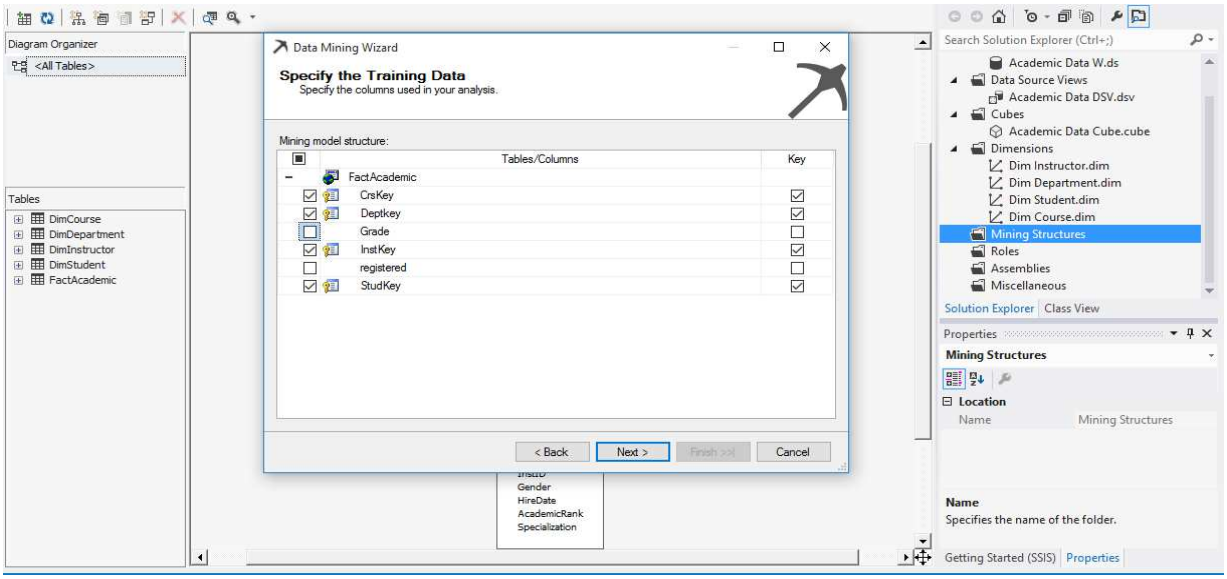


Figure 15: Specifying the training data set

After you properly configure specification of training and testing data the next step selecting the dimension tables key as indicated in the above table.

Now this stage where we select dimension key and column mapping to show the successful implementation of integration process.

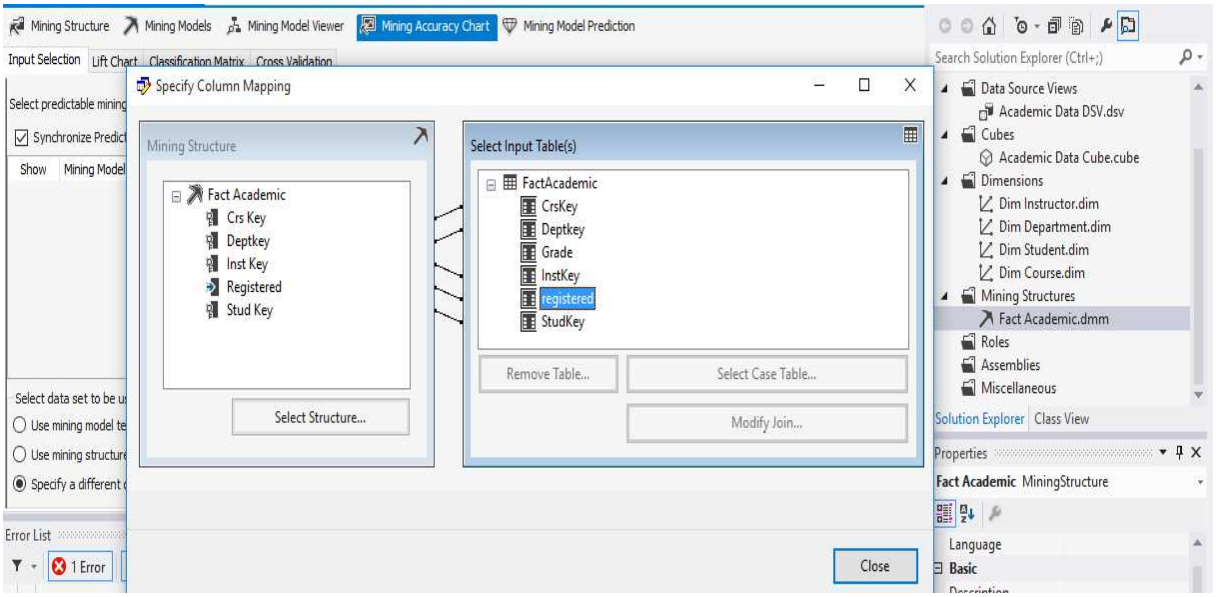


Figure 16: Mapping column

The main challenges of the above integration process is the structure selection and the nature of training and testing data set. If the data is not as per your system requirement then it will generate error**.**

Therefore, proper selection of algorithm, model, training and testing data set and finally mapping column is important tasks in the integration process. In this research our main target to designing prototype data warehouse decision support system to provide analytical information for decision maker. Therefore, next result and discussion section only focus on the OLAP analytics using cube browse.

Chapter Five

5. Result and discussion

In this section we will have detail discuss on the analytical outputs of the proposed prototype data warehouse system.

5.1. Browse Cube Data

When the cube is one of the main tools in SQL server data warehouse designing that is used as user interface to display analytical results on the screen. Results will be displayed by referring all relevant dimension in the data warehouse. It is one of the easiest and fastest tools for the developers which is available in business intelligence development studio. The following figure presents the of cube browsing sample output students registered in department by region and gender.

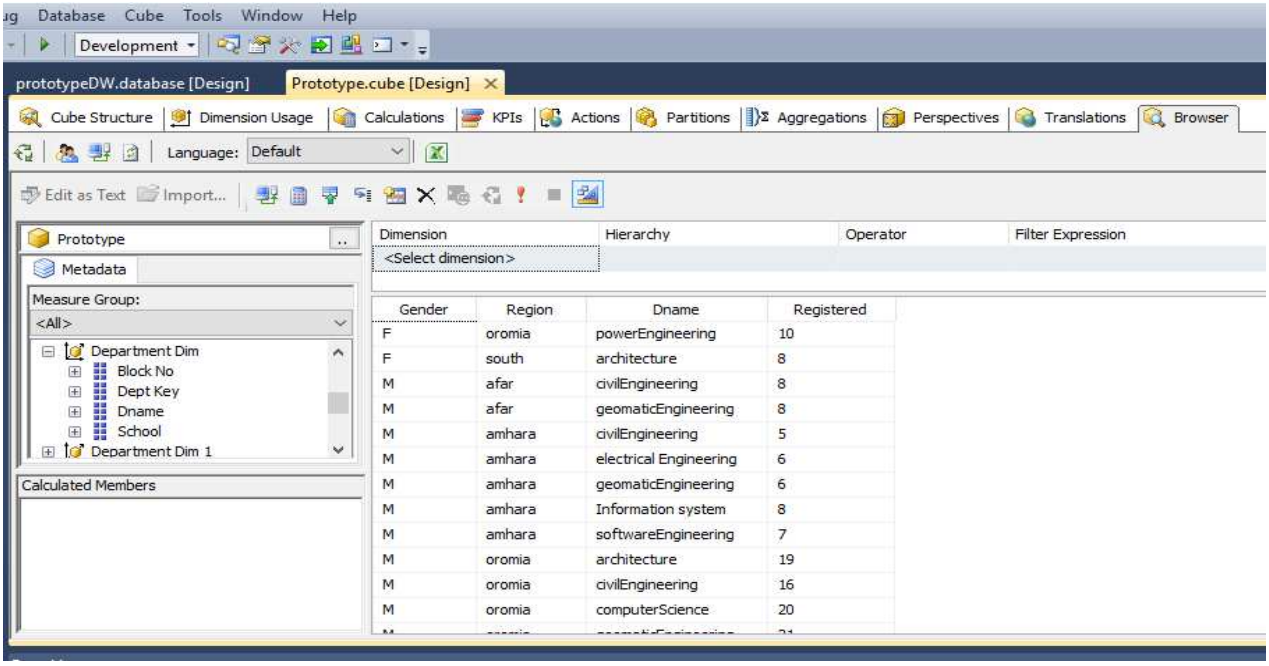


Figure 17: Data warehouse dimensional analytics using cube browsing

Figure 13 shows the analysis data warehouse to identify the total number of active student/students attending their study in each department based entry year. Therefore, to display the above result OLAP cube browsing refer department dimension, from fact table it select the total numbers of student registered in the each department, gender and region of the student is important attribute or column to see the participation of student in the university.

Another important task in browsing analytical output from data warehouse is the integration SQL Server with excel pivot to get the same result as Cube browsing tool. Pivoting is one of the powerful applications that allow one to navigate and build OLAP reports in a web browser. The following figure presents the grade distribution of students.

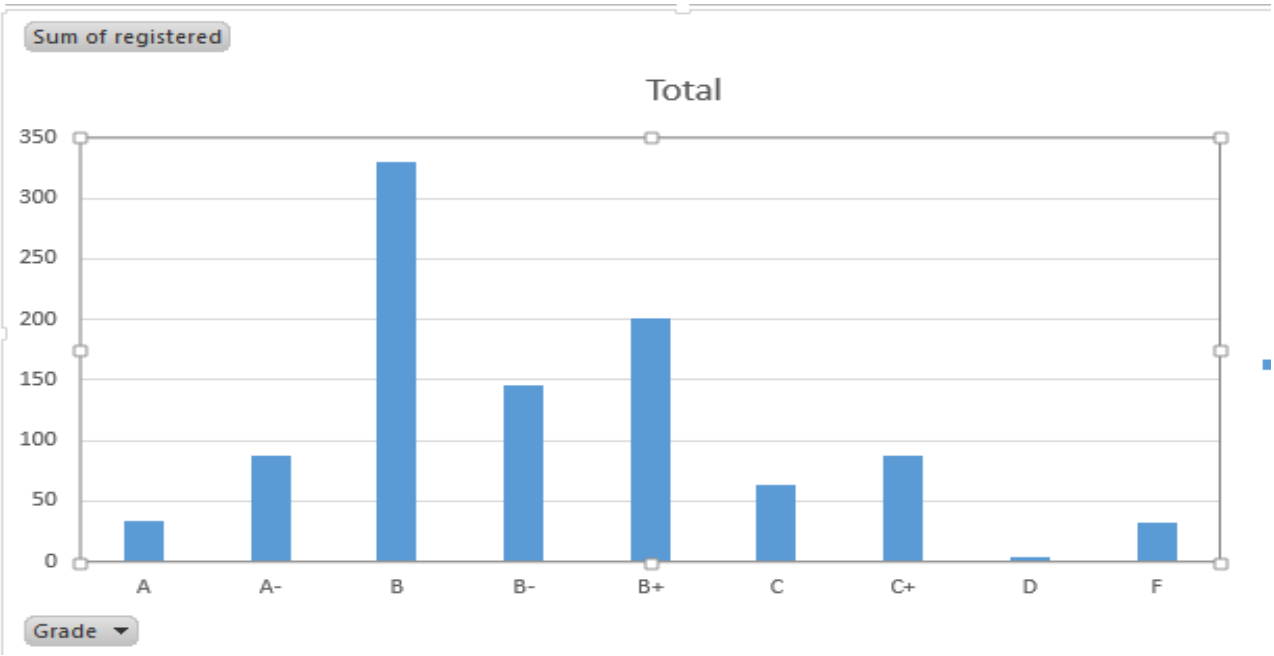


Figure 18: Average Grade distribution

The above figure shows the distribution of student and their academic status. Top managers are interested to see the general pattern of student status rather than the daily data transaction. This patterns will help then to ask strategic question and to provide strategic respond for the overall academic activities. What is the majorities of student academic performance in certain semester or academic year? What measure should be taken for those students who scored grade D and F.? What is source cause for low academic performance? What measures need to take to increase their number of high achiever? How was the facility, infrastructure, and service provision to enhance the quality of education?

This is some of the important analytical information used as input to make strategic decision. On the other hands, analytical input will reduce out time and effort to process transactional data. The following figure shows, the distribution of student per department.

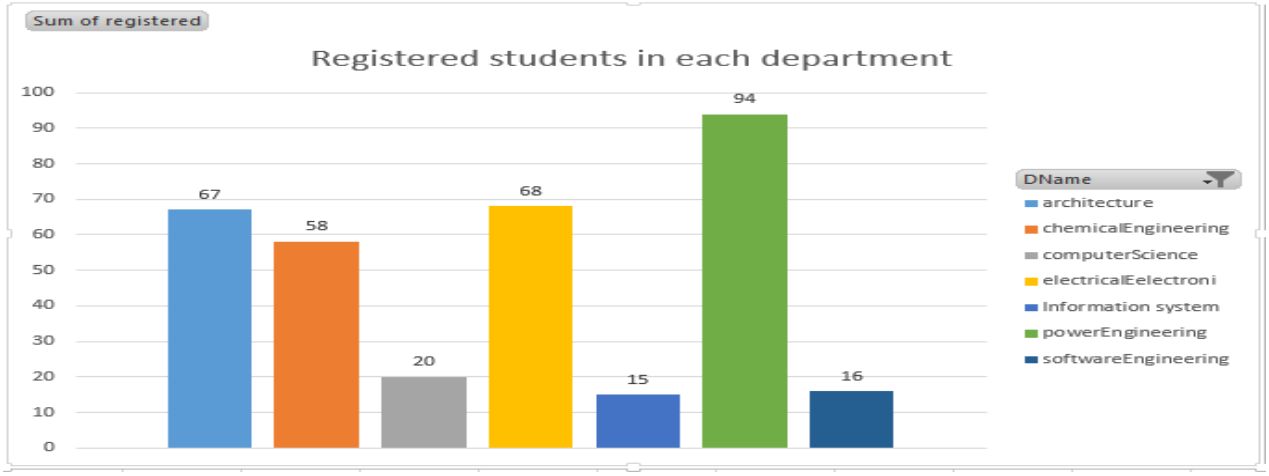


Figure 19: Distribution of student per department

To get the above output we referred department dimension table and from measure group we select total numbers of registered student column. The two dimension give us critical insight for decision makers. Why student highly interested in a certain department? How was capacity of that department to accommodate large numbers of student? How can we balance the burden of our academic staff? Do we have enough resources and facility or laboratory for the student? What about other department with minimum numbers of student? How can we utilize the existing manpower to its maximum level?

We should give clear direction for such important question to balance distribution and proper utilization of resources (human and materials). Having this input different stakeholder or concerned unit take their own part to bring the desire change.

Another important concern for top management is the numbers of skilled man power in different specialization to assure the quality of education. This kinds of analytical information is important to designing long and short term strategic plan to upgrade our academic staff profile. Therefore, the following figure shows the distribution of academic staff per specialization.

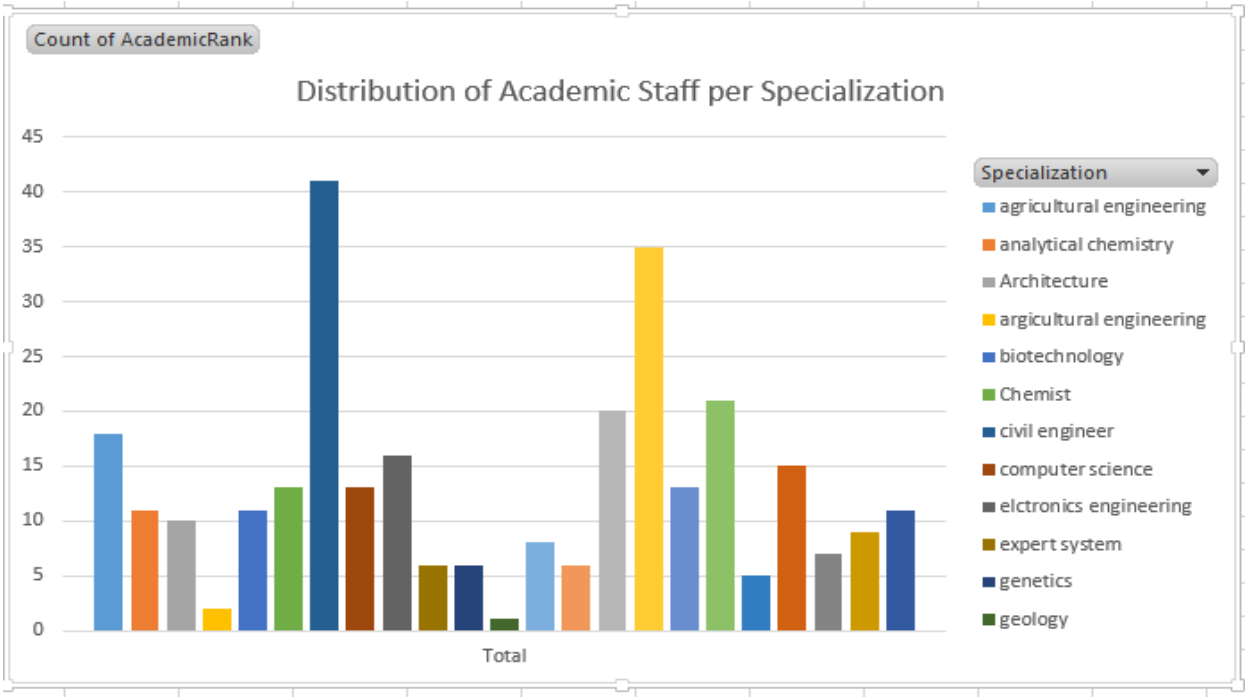


Figure 20: Distribution of academic staff

The next dimension we interested is to see the participation of student across the nation. We selected region, gender and registered total numbers of student. How was the representativeness of student for science and Technology University? How can the University encourage student in order to join on hard science. Which region is least participated/interested in science and technology oriented university. What kinds of strategic plan we need to design to attract student. What kinds of student recruitment promotion is low cost and high profit for the University.

Table 3: Calculating numbers of students in region and gender wise

	Gender		
Region	F	M	Grand Total
afar		16	16
amhara	97	234	331
oromia	132	169	301
somalia		14	14
south	8	16	24
tigray	98	199	297
Grand Total	335	648	983

In addition, it is very important know the participation of student from each nation and nationalities to gain competitive on student diversity. Obviously student diversity will add values by sharing culture and practices the student in their respective region. It also gives us a direction to make some action to increase the participation of student from all regions.

"Due some challenges we create artificial column like region that is why the above figure shows unreal distribution result such as student from Oromia is less than student from Amhara region. The truth is this sample data for prototype designing only. We have made the same thing on student grade score to show the status of student across the school, the reason is we are not to get such data from the concerned office"

Data warehouse prototype evaluation

There are different software testing strategies to evaluate the performance the system. For the purpose this research we selected the following performance testing strategies to evaluate the proposed prototype data warehouse decision support system:

A. **Data completeness testing:**

This Testing evaluate that data quality is undoubtedly at the core of data warehouse testing. Testing data quality mainly entails an accurate check on the correctness of the data loaded by ETL procedures and accessed by front-end tools. By correcting missing and incomplete data we can enhance the performance of the system.

B. Query processing speed testing:

To design dimension modelling we used de-normalized star-schema in which all the dimension tables are connect with central fact table. Query processing faster than snow fleck schema because analysis is on the basis of relevant dimension. The other performance enhancement procedure is the use of Surrogate keys that can be derived from the existing natural keys or it can be a simple ETL generated integer number provides the means to maintain data warehouse information. One simple way improve performance of queries is to use surrogate keys. For this project ETL generated surrogate keys are used.

C. Reporting feature testing:

This testing approach evaluate the systems accessibility by end-users to analyze data; typically, capability of analyzing the cube from different dimension and modifying the dimension until we get the required report. From reporting feature aspect the proposed data warehouse is efficient to generate feature oriented analytical information for decision maker.

D. OLTP and OLAP testing:

Finally we evaluate system from online transaction process and online analytical process point of views. OLTP provide daily transaction output data or figures by referring specific table which is not relevant for strategic decision. On the other hands, OLAP technology provides user and data scalability, performance, read/write capabilities and calculation functionality, it meets all the requirements of a data warehouse. The OLAP technology option supports collaboration throughout the business management cycle of reporting, analysis, what-if modeling and planning. Using the OLAP technology data warehouse analytics will be made from n-dimension depending on the user requirement.

5.2. Conclusion and recommendation

Nowadays every organization has created and processed large amounts of data. The challenge for organization is to transform the data into knowledge for competitive advantages. Although organizations records and produces huge data, this data is not utilized beyond consumption of day today reports. However, data warehouse allows origination to transform and integrate the data in correct manner and format to understand the patterns within customer experience and make analytical decisions.

The objective of this research is to develop a prototype decision support data warehouse to verify the application of data warehouse technology specifically the capability of online analytical processing (OLAP) technology for multidimensional analysis using dimensional modeling approach. Spiral model paradigm of the software engineering approaches employed to design the prototype data warehouse system.

As the most challenging task in data warehouse building the requirement analysis, data conversion, data quality and integration activities are demonstrated to implement the prototype decision support data warehouse with selected dimensions. These selected dimensions are populated with data and OLAP cube is designed and multidimensional analysis is performed using Cube browsing and pivot table. The result of the study indicated that the proposed data warehouse system provide analytical information efficiently depending users requirement (relevant dimension).

In general the application of data warehouse for Adama Science and Technology University as higher institution could be provide immense contribution to generate standard reports for strategic decision and it would give almost instantaneous outcome for the concerned Office. Data warehouse business intelligence allow decision makers to intelligently query and analyze information in very efficient manner.

Recommendation

In this research a prototype decision support data warehouse is developed for Adama Science and Technology University. It possible to implement real corporate data warehouse system by integrating all business process and operational data it is also possible to apply data mining algorithms to discover a new pattern and knowledge in addition to analytical processing.

References

[1] Resenceles, Data Warehouse Project White Paper, 206.

[2] Zehra KAMIŞLI, "Database Management Systems: A Case Study of Faculty of Open Education," *The Turkish Online Journal of Educational Technology –,* 2004.

[3] D. Mankad, "The study of Data Warehouse Design and Usage," *International journal of scientific and research publication,* pp. 2250 - 3153, 2013.

[4] A. Sarkar, "Data Warehouse Requirements Analysis Framework: Business-Object Based Approach," *International Journal of Advanced Computer Science and Applications,* 2012.

[5] M. T. N. A. R. S. Hegadi, "Realistic Analysis of Data Warehousing and Data Mining Application in Education Domain," *International Journal of Machine Learning and Computing,* 2012.

[6] J. Mckendrick, a new dimension to data warehousing: 2011 ioug data warehousing survey, Unisphere Research, 2011.

[7] N. Jing, "Data Warehouse Design and Optimization for Drilling Engineering," *The Open Petroleum Engineering Journal,* pp. 124-129, 2012.

[8] D. M. M. Hamad, "Knowledge Driven Decision Support System Based on Knowledge Warehouse and Data Mining for Market Managemen," *International Journal of Application or Innovation in Engineering & Management ,* vol. 3, no. 1, 2014.

[9] A. A. A. A. O. Akintola K.G., building data warehousing and data mining from course management systems: A Case Study of Federal University of Technology (FUTA) Course Management Information Systems, 2011.

[10] T. P. PLT, data warehouse tutorial, 2014.

[11] S. G. A. A. Vavouras, "Data Warehousing: Concepts and Mechanisms," 1999.

[12] R. M. A. B. B. Claudia Choi, "Combination of a data warehouse concept with web services for the establishment of the Pseudomonas systems biology database SYSTOMONAS," *Journal of Integrative Bioinformatics,* 2007.

[13] F. D. T. Carlo DELL'AQUILA, "An Academic Data Warehouse," *Proceedings of the 7th WSEAS International Conference on Applied Informatics and Communications,* 2007.

[14] S. Chaudhuri, "An Overview of datawarehousing and olaptechnology," *Hewlett-Packard Labs, paloalto,* 2005.

[15] G. R. A. M. P. C. Rao, "data warehousing, data mining, olap and oltp technologies are essential elements to support decision-making process in industries," *International Journal on Computer Science and Engineering,* 2010.

[16] A. S. A. A. P. Sinha, "A Comparison of Data Warehousing Methodologies," 2005.

[17] K. I. Mohammed, "data warehouse design and implementation based on quality requirements," *International Journal of Advances in Engineering & Technology,* 2014.

[18] S. Singh, "data warehouse and its methods," *Journal of Global Research in Computer Science,* vol. Volume 2, no. No. 5, 2011.

[19] P. Muley, "“Exploring the Scope of Data Warehouse and Business “Exploring the Scope of Data Warehouse and Business," *IOSR Journal of*

Business and Management, vol. 18, no. 7, p. 59, 2016.

[20] J. D. A. M. Sarrab, "Three Tier level Data Warehouse Architecture for Ghanaian Petroleum Industry," *International Journal of Database Management System,* 2012.

[21] P. Sharma, "Advanced Applications of Data Warehousing Using 3-tier Architecture," *DESIDOC Journal of Library & Information Technology,* pp. 61-66, 2009.

[22] T. Horvli, "Data Warehouse Presentation," 2004.

[23] N. Anand, "Application of ETL Tools in Business Intelligence," *International Journal of Scientific and Research Publications,* vol. 2, no. 11, 2012.

[24] V. Gour, "Improve Performance of Extract, Transform and Load (ETL) in Data Warehouse," *International Journal on Computer Science and Engineering,* vol. 2, no. 3, pp. 786-789, 2010.

[25] P. S. Shivtare, "Data Warehouse with Data integration problem and solution," *IOSR Journal of Computer Engineering (IOSR),* pp. 67 - 71, 2015.

[26] Amanpartap Singh Pall, "ETL Tools in Enterprise Data Warehouse," *International Journal of Advance Foundation And Research In Science & Engineering (IJAFRSE) ,* vol. 1, 2015.

[27] A. Parekh, "Introduction on Data Warehouse with OLTP and OLAP," *International Journal Of Engineering And Computer Scienc,* p. 2569, 2013.

[28] A. Rai, "data warehouse and its applications in agriculture," *Indian Agricultural Statistics Research Institute ,* 2005.

[29] P. Ofori Boateng, "data warehousing," *Business Intelligence Journal,* 2012.

[30] V. R. RAO, "A Framework for e-Government Data Mining Applications (egdma) for Effective Citizen Services," *International Journal of Computer Science and Information Technology Research ,* pp. Pp: (209-225), 2014.

[31] X. Hu, "Data Warehouse Technology and Application in Data Centre Design for E-government," 2010.

[32] A. A. A. A. E. Al, "a framework for educational data warehouse (edw) architecture using business intelligence (bi) technologies," *Journal of Theoretical and Applied Information Technology,* 2014.

[33] M. T. N, "Design and Analysis of DWH and BI in Education Domain," *IJCSI International Journal of Computer Science Issues,* 2011.

[34] S. Mirabedini, "The Role of Data warehousing in Educational Data Analysis," *Journal of Novel Applied Sciences,* pp. 1439-1445, 2014.

[35] S. Reddy, "data warehousing, data mining, olap and oltp technologies are essential elements to support decision-making process in industries," *International Journal on Computer Science and Engineering,* 2010.

[36] N. M. A. N. L. A. Nikolaos Dimokas, "A Prototype System for Educational Data Warehousing and Mining," *Work supported by National Project ΕΠΕΑΕΚ,* 2007.

[37] C. A, "Prototyping an Academic Data Warehouse: Case for a Public University in Kenya," *British Journal of Applied Science & Technology,* pp. 550 - 557, 2015.

[38] M. VELICANU, "Building a Data Warehouse step by step," *Informatica Economica,* 2007.

Appendix A

Sample unstructured interview questionnaire for data encoders

Personal information:

Your name_____

Your position_____

Your responsibility in the unit_____

Research question:

1. What is the sources of your student data?
2. Do you have and central data base to manage student data?
3. What kinds data management system currently the office uses
4. How you manage the quality of data and the means of quality management
5. What the procedure on update student's data?
6. How you process the data to generate report for decision makers.
7. What is the challenges of using transactional data for strategic decision
8. Do you any system to generate pattern from the existing historical data