

Design and Implementation of Efficient Decision Support System Using Data Mart Architecture

1stRefed A. Jaleel

Information and Communication Engineering
Al-Nahrain University
Iraq
iraq_it_2010@yahoo.com

2nd Talib M.J. Abbas

Information and Communication Engineering
Al-Nahrain University
Iraq
talib_altalib@nahrainuniv.edu.iq

Abstract—This paper presents and discuss advantage about the design and implementation of Decision Support System by building data mart, which is a place for the storage and participation of data, which users can readily utilize to product improved decisions. Due to the problems within the systems of Information and Telecommunication Public Company obtaining the data in a successful method requires a harmonious potential Thus, This paper presents proposed framework of Decision Support System for Information and Telecommunication Public Company by which discusses the advantage about design, integration of the building data mart and what is the difference between it and relational database for-decision makers to obtain information that contains four layers (Interface layer determines data source, Extract Transformation Load Layer determines design process in SQL Server Integration service, Data mart layer determines data mart that conform Star Schema, Presentation layer determines the results from the system).

Keywords— Data Mart, Star schema, dimensional Model, ETL, Database of Document Management System.

I. INTRODUCTION

The data are usually very large, complex and not organized in a suitable format helpful to decision makers for extracting any meaningful information without help of database technicians. Moreover, strategic decision making requires analysis of data of several years. Existing databases and their formats are not designed to efficiently and directly provide such information to decision makers. In situations like this, a decision support system (DSS) using Data Warehousing (DW) or Data Mart (DM) technology can be useful to decision makers in making quick and critical decisions [1]

To understand what a data mart is, we must first know a little about DW. The concept of DW first occurred in the late 1980s in articles published by Bill Inman and others. Essentially, a DW contains a systematic approach to collecting, arranging, and saving data—generally from

internal production sources but also from external providers—to build a definitive source of information for analysis and decision support [2].

A data mart reflects the regulations and process specifications of each business unit within an organization. Each data mart is devoted to a particular business function or region. This subset of data may span across many or all of an enterprise's functional subject areas [2].

Many firms are wondering how extract data from data sources that in various formats to move it into data stores. As well as summarizing data from various IT systems with various data structures shows a big challenge. To meet these challenges there is a process called Extract Transformation Load (ETL) [3] by which various operational systems of the organization are cleansed, transformed, integrated and loaded into the data warehouse or data mart, data is then made available to the knowledge user for access using decision support tools [4].

II. MATERIALS AND METHODS

A. Use Case Diagram

The Use Case diagram considered for high-level demand analysis phase of a system for view, which used to gather functionalities and requirements of a system, and to identify external and internal actors interacting with the system, these actors are known as agents [5], it's such a bridge covers the gap between the simple user of the system and the software designer [6].

B. Entity Relationship Diagram [7][8]

Entity Relationship Diagram (ERD) is a main data-modeling tool, it helps arrange the data in your project into entities and illustrates the relationships between the entities. This process has proved to empower the analyst to result a robust database structure so that the data can be maintained and restored in a most efficient manner. ERD usually has three presentations, which are called: the conceptual data model, the logical data model, and the physical data model.

A conceptual data model is used to define the entities, which acts as the data of the problem scope, and their

relationships. It leaves details about the entities such as attributes data type and primary keys.

A logical data model is a refinement of the previous conceptual model. It details the domain entities and their relationships. It describes all the attributes that act each entity also, containing its unique identifier, the primary key and all the relationships between the entities and possibly including the keys identifying those relationships, the foreign keys.

A physical data model visually acts the structure of the data as executed by a particular class of Database Management System (DBMS). Therefore, entities are represented as tables, attributes are represented as table columns and have a given data type that can differentiate according to the selection DBMS, and the relationships between each table are identified through foreign keys. Unlike the previous models, this model tends to be platform-specific, because it reflects the database schema.

C. Data Mart

DM is a de-normalized database, using star model, with the function of storing historical data of a specific business department [9]. A DM is a constant physical store of operational and aggregated data statistically processed data that supports business people in making decisions depend primarily on analyses of past activities and results. A DM consist of a predefined subset of enterprise data arranged for rapid analysis and reporting [2]. The DM deal with all the benefits of ETL process, similarly as a DW does. The number of sources is likely to be less and the amount of data related with the DM is less than the warehouse given concentrated on a unique subject [10]. There are many reasons for creating DM shown in following table1.

TABLE 1. Purposes for building data marts [11].

Simple of creation	It's simple to create data marts as its given to a specific subject
Time Response	response time for end user has improved
Cost Lower	Cost lower than implementing a whole DW
Improvement	Easy to improve over time
Plainness	Focus on a unique subject
Domain	Simple to understand as domain is limited

D. Dimantional Data Model

Dimensional data model commonly used in data warehousing systems as star schema, snowflake schema and fact constellation/galaxy schema. This is various from the third normal form, usually used for On Line Transactional Processing (OLTP) systems. As can imagine, the same data would be saved variously in a dimensional model from a

third normal form model [12]. The benefits of dimensional modelling method summarized as follow [13]:

- Better data presentation and navigation.
- low-cost and easy maintenance.
- Better performance.

E. Problem Statement and the Motivation

The decision makers in Information and Telecommunication Public Company (ITPC) are actually based on analyzing raw data on demand, which costs a lot of time, stress and may not constantly be precise since it is manually analyzed by individuals. The decision makers (business users) poverty the capability to access data easily when needed, to handle this shortcoming, to deal with the issues of data inconsistency and time loss, there have been motivations to implement data mart, to solve the integrating issue of data and to guarantee that data is found in the time it is prerequisites. The most important motivations to implement data mart are efficient data report platform, technology reliability, the less time of query and data format integration. It is integrated repository of data in order to standardize data analysis and enable simplified usage patterns.

F. Architecture of the Proposed Framework

The structure of the proposed DSS is explained in Fig.1. The Interface layer contains database of document management system, database was made in SQL Server, making for it use case diagrams and ERD models. In ETL layer, cleaning database form dirty data in SQL Server Integrated Service (SSIS) and then loading it to data mart. The Data mart layer contains star schema.

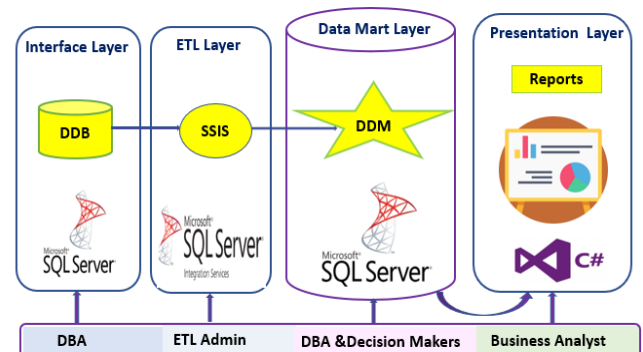


Fig.1. Architecture of Efficient Decision Support System.

G. Interface Layer

This Layer is considered as a system analysis phase, it contains Document Database for archiving, Information about the database obtained by necessary methods such as Introspection and Interviews, analyzed system by using use case diagram to show the interaction between users and the system, and build procedures and system that will obtain it in an efficient method.

A Simplified process of the system obvious in this layer, In Document Management System (DMS), the actors represent employees (sender user, recipient user); external user subject to requests in document form; application administrator performs application of running repair operations. Use case diagram at the document level shown in fig.2. consists of seven use cases: create document, verify document existence, archiving status, save new version, modify document, archive document, and delete document that addresses all possible cases in which the actors can interact with the DMS.

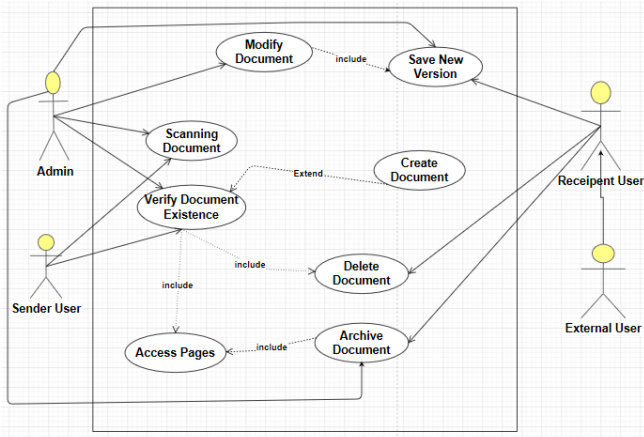


Fig.2. Use Case Diagram of DMS.

In order to implement DSS effectively, it is essential to have outstanding knowledge regarding ERD models for data source to define a realistic presentation of the data in the database and see the data precisely, The ERD models as mentioned defining the Conceptual model, the Logical model, and Physical model. In this paper only the physical model will be offered as shown in Fig.3, which drawn in Visual Paradigm software.

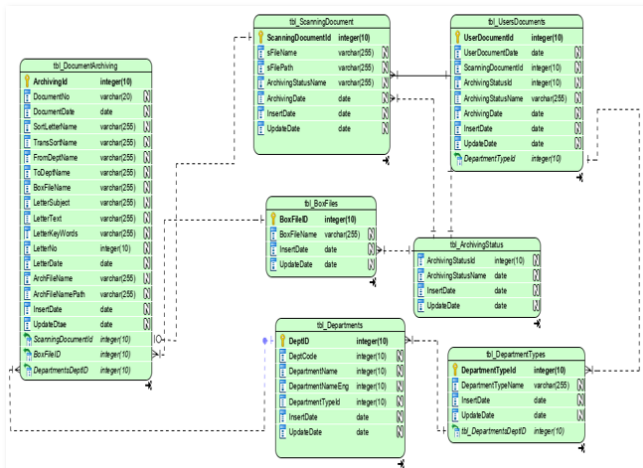


Fig.3. Physical Data Model for DMS.

H. Data Mart Layer

The physical database design structure known as a star schema has a particular affinity for the DM, a star schema made up of fact table and dimensions. The fact table represents the types of data that occur in great volume. The

dimension tables represent smaller tables that pre-joined to the data in the fact table by means of a foreign key relationship, because a star schema generally entails some physical de-normalization, there is the implication that star schemas are optimal for only one type and class of processing [14].

The star technique sometimes wrong applied to the DW because the DW will finish up being optimal for one group of users at the expense of everyone else [15]. The star schema obtains its name from its occurrence [13]. In this paper building star schema for DMS in SQL server 2017 R2 according to requirement that has three measures which are average document archiving, average scanning document and number of user document. For building Document Data Mart (DDM), used the star schema and add granularity as possible in order to enhance analysis. The physical design of the schema offers the relationship between the dimension and fact tables, it's determined the actual structure of database.

The structure of dimension tables for the data mart is as follows; Date, Box Files, Doc. Archiving, Scanning Documents, Departments, and User Documents as shown in Fig.4.

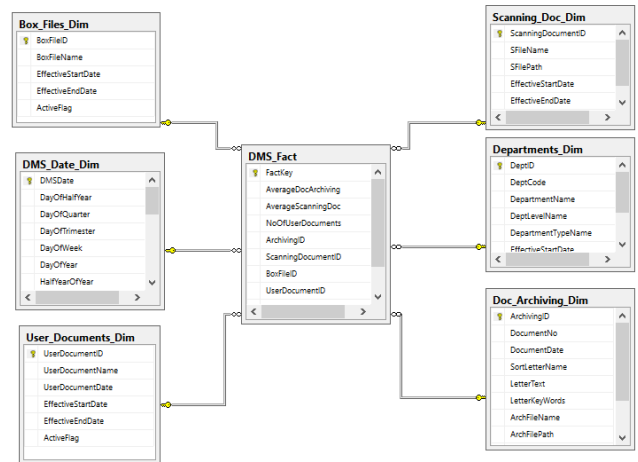


Fig.4. Physical Design of DMS Data Mart.

I. ETL Layer

In this layer the process of ETL for populating DMS data mart was designed by using (Visual Studio-Business Intelligence - Integration Service) software [16]. Three packages of SQL Server Integration Services (SSIS) are created. Where ETL process was designed for the source (Database) system. ETL process to loading data in data mart of DMS contains many packages in SSIS. The first and second package for populating dimension tables and third package for populating fact table of DMS data mart.

The first package for loading Users Document, Scanning Documents and BoxFiles dimension tables contain three steps as follows:

- Step1:** Extract data from source tables (Box files, User Documents and Scanning Document) by make manager connection to DMS database.
- Step2:** Make Lookup transformation to source tables for mapping between input columns and columns in dimension tables.
- Step3:** Loading data to dimension tables like users documents, box files and scanning documents by make manager connection to DMS data mart. See Fig.3.

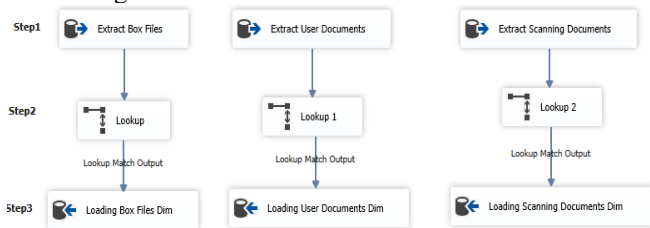


Fig.5. ETL for Box Files, User Doc. and Scan Doc. Dimensions.

The second package for loading Departments and Documents Archiving dimension tables contains four steps as follows:

- Step1:** Extract data from source tables (Departments and Documents Archiving) by make manager connection to DMS database.
- Step2:** Using union all transformation to unit data, which needed in dimension tables.
- Step3:** Using lookup transformation for mapping between available columns and dimension columns.
- Step4:** Loading data to dimension tables Departments and documents archiving by make manager connection to DMS data mart. See Fig. 6.

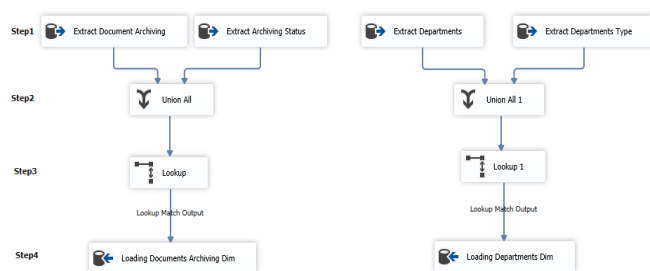


Fig.6. ETL for Doc. Archiving and Depart. Dimensions.

The third package of loading fact table of DMS has seven steps as follows:

- step1:** Prepare source data from source tables (Box Files, Document Arciving, Scanning Document, Departments, Users Documents) which having dimension keys and measures.
- Step2:** Acquire the dimension surrogate keys for any related dimension (Box Files ID, Document Archiving ID, Scanning Document ID, Departments ID, Users Document ID) by using Lookup Transformation and

then mapping between Alternate key of sources tables and to the Lookup Columns.

- Step3:** For the DMS Date, a Lookup not used since the Date Key is an intelligent key meaning the key is an integer value depend on the date itself. Thus, use a Derived column to compute the Date Key for the DMS fact and in order to put derived measures.
- Step4:** Sort transformation is used to records of the source and extract the data of the fact into the data flow before the Merge Join.
- Step5:** A Merge Join used for mapping records of source input to the records of fact table.
- Step6:** Using Union All for unit all stages in the previous step.
- Step 7:** Mapping between the columns of the database with columns of the fact table for data mart, and then insert data in this fact table. See Fig.7.

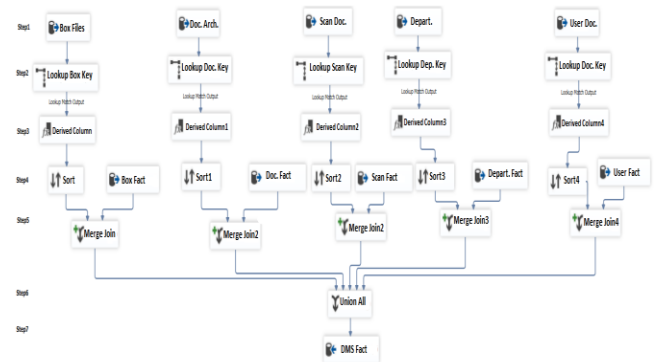


Fig.7. ETL for DMS Fact.

III. RESULTS AND DISSCUSSION

A. Presebtation Layer

The Presentation Layer is the last part of the architecture and it's represented as a Graphical User Interface (GUI) which designed in the windows application environment using C# as the programming language. The decision makers in department of DMS have his own user name and password to enter the data mart, after logging to the system the member is re-directed to the reports page. Fig.8 and Fig.9 shows login page, reports pages respectively.



Fig.8. Login Page of Data Mart for DMS.



Fig. 9. Report Pages of DMS Data Mart.

There are many reports available to document department such as (document, departments archive) and many more detailed. Complex reports that is made possible doing from the data mart but the previous system that represent in ERD was not able to produce them in the same performance of data mart since it was a normal database for example, some of the requirements of the document department is knowing which department in company archive their documents more than once, the report about this requirement shown in Fig.10 ,this considered a benefit in data mart report which enable employees to query about summarized data in less time and another example for report about number of documents archiving for many departments in company through a certain period as shown in Fig. 11.

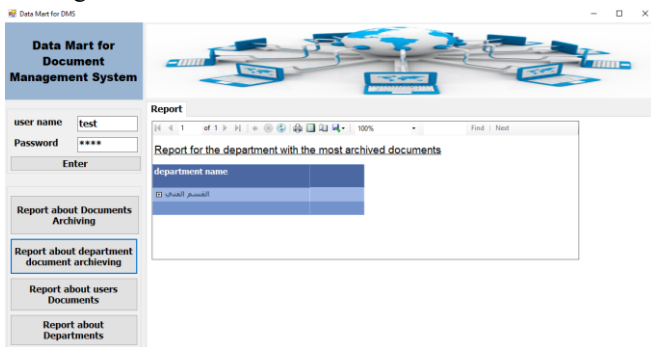


Fig.10. Report about department document Archiving.

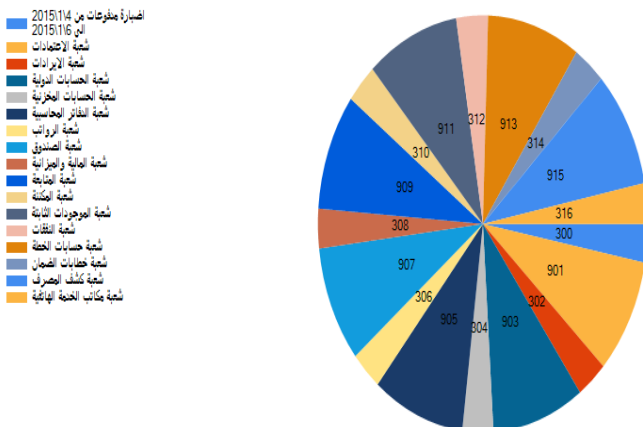


Fig. 11. Report about number of documents Archiving.

B. Data Mart Performance

In order to test the differences in response time of queries between database and data mart, building a group of testing, the same query is used, direction in Document Database, against data mart. In this section taking set of queries, select from database of DMS. Taking three select statement of queries (Select Top 100, Select Top 1000 and Select all records). After consulting all the results in table2, it is enjoyable to offer a big hole in time between querying the databases against the data mart of Document. As shown in the following Fig.12.

TABLE 2. Execution Time of the Query for DDM.

Query	ERD	Star Schema
Select Top 100	7.9 M	0.2 M
Select Top 1000	18.9 M	1.7M
Select *	42 M	10 M

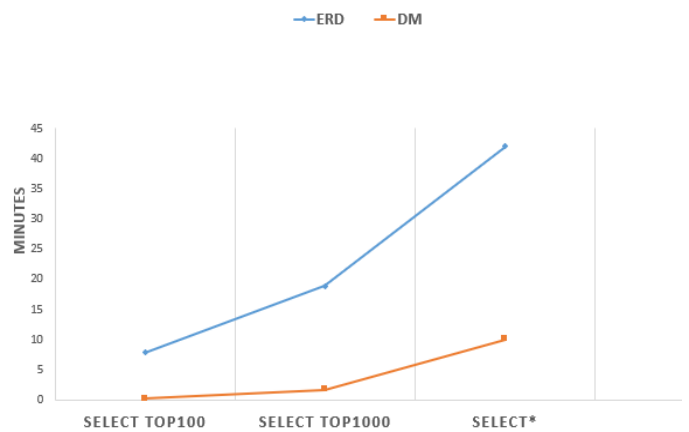


Fig.12. Graphical Testing Set for DMS Data Mart.

IV. CONCLUSIONS

Efficient DSS has been built for DMS that collects data from normalized database of DMS and stores in a repository of data mart, and solved the problems faced by decision maker, and improves speed to answer queries. It was easy to see how populate star schema by an active ETL tool from database. With the implementation, looking into executing the same query versus the same data but in various system. When comparing the results of executing the queries, it is obvious in every instance that the data return from the organized star schema dramatically quicker than from the ERD.

ACKNOWLEDGMENT

First and foremost, praise and thanks to Allah who facilitated this paper. I would like to express my appreciation and great thanks to ITPC and my supervisor Prof. Dr. Talib M. J. Abbas, for his valuable instructions, patience and support throughout writing this paper.

REFERENCES

- [1] A. Yosef, Y. Rahal, F. Ammareen and K. Joumaa, Development of a Decision Support System using Data Warehousing (Site section), A project submitted to Informatics Department in partial fulfillment of the requirement for the degree of bachelor, 24 December 2013.
- [2] N. Malhotra, "Implementation of data marts in data warehouse.", International Journal of Advance Research, ideas and innovations in Technology, Volume1, Issue 2, 2015.
- [3] O. Azeroual, G. Saake and M. Abuosba, "ETL best practices for data quality checks in RIS databases", Journal Informatics, March 2019.
- [4] P. Bhadresh and Dr S. Shah. "Proposed local data mart approach for data warehouse architecture." International Journal of Emerging Technology and Advanced Engineering, Volume 4, Issue 2, 2014.
- [5] K. El Miloudi and A. Ettouhami, "A Multiview formal model of use case diagrams using Z notation: towards improving functional requirements quality," Hindawi Journal of Engineering, 2018.
- [6] I. Essebaa and S. Chantit, "Tool support to automate transformations from SBVR to UML use case diagram", In Proceedings of the 13th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE), pp. 525-532, 2018.
- [7] M. M. Khudhair, A Data Warehouse for Higher Education Student Information System to Support Decision Making, A Thesis Submitted to the Informatics Institute for Postgraduate Studies at the Iraqi Commission for Computers and Informatics in a Partial Fulfillment of the Requirements for the Degree of Master of Science in Software Engineering, 2017.
- [8] A. Ribeiro and A. Silva, "Data modeling and data analytics: a survey from a big data perspective", Journal of Software Engineering and Applications, vol. 8, no. 12, pp. 617, 2015.
- [8] F. C. Velásquez, Business intelligence and organizational culture: case study in a higher education company in Perú, Research project presented in partial satisfaction of the requirements to obtain the degree of Master in Business Administration, 2018.
- [9] P. Maurya, "Conceptual study on data marts – a building block of data warehouse", International Journal of Computational Engineering Research (IJCER), ISSN 2250 – 3005 Volume 8, Issue 4, April 2018.
- [10] R. Chhabra and P. Pahw, "Data mart designing and integration approaches", International Journal of Computer Science and Mobile Computing, Vol.3 Issue.4, April- 2014.
- [11] A Hadi Alramahi, Designing a variety of data warehouse schemas for meta-search, A Thesis Submitted to the Collage of Computer Science and communication in the University of Al Qadisiyah in Partial Fulfillment of the Requirements for the Degree of Master, 2016.
- [13] B. W. Shbair, Design and Implementation of a Data Warehouse Using Dynamic Materialized Views Selection Model: The Islamic University of Gaza as a Case Study, A thesis submitted in partial fulfillment of the requirements for the degree of Master of Computer Engineering, August 2017.
- [14] W. H. Inman, Claudia Imhoff, and Ryan Sousa, Corporate Information Factory, Second Edition, John Wiley & Sons, ,2008.
- [15] W. H. Inman, Building the Data Warehouse", Third Edition, John Wiley & Sons, Inc.,2005.
- [16] R. Mistry and W. Misner, Introducing Microsoft SQL Server 2012, First Edition, Microsoft Press A Division of Microsoft Corporation, United States of America, 2012.