



Using OLAP with Diseases Registry Warehouse for Clinical Decision Support

A. K. Hamoud¹, Dr. Taleb A.S. Obaid²

¹Computer Science Department, College of Sciences, University of Basrah, Iraq

²Computer Science Department, College of Sciences, University of Basrah, Iraq

¹ alaak7alaf@yahoo.com; ² taobaid@gmail.com

Abstract - Diseases registry databases in Iraqi hospitals hold huge information which can be used to support strategic clinical decisions. The clinicians and professionals often need tools to get the valuable information. This information cannot be got with normal functions used with operational databases. Using On-Line Analytical Processing (OLAP) gives insight and clear view on the information from different points. It makes it very easy and fast to view results as reports. The researchers used with clinical data warehouse to give decision makers on-time results to assist them in support their decisions. Two database sources used to construct and build data warehouse in order to use OLAP to view multidimensional data.

Keywords- OLAP; Clinical Data Warehouse, Decision Support Action

I. INTRODUCTION

Information access is of critical importance in the practice of the public health service. Having timely, accurate and readily available information is essential in monitoring the health of communities and populations. Having access to public health data is essential for determining the association of environmental exposures to diseases, as well as measuring the progress and the efficacy of interventions. Effective public health practice relies on the availability of public health data sources and assessment tools to convey information to investigators, practitioners, policy makers, and the general public [1].

For most of the last 3 decades, data warehouse use has been mainly in commercial, non-healthcare settings. Recently some healthcare organizations have built data warehouses that may have uses similar to those in the non-healthcare setting, but may also be quite dissimilar. In healthcare organizations these data warehouses collect data from patient care, population-based

databases, financial, claims, and administrative transactional systems, and then organize these data elements to support information retrieval, business intelligence, research, and decision making. Information management is critical to many customers within healthcare delivery systems, especially providers, management, and clinical operations. Such data warehouses have become the central information management platform for decision support in many organizations. Compared to data warehouses in non-healthcare settings, those in healthcare settings may focus more on privacy and security of protected health information, as well as compliance with federal and state regulations and organizational policies [2].

Most important effective tool which used with data warehouse is Online Analytical Processing (OLAP). OLAP enables analysts, managers, and executives to gain insight into data through fast, consistent, interactive access to a wide variety of possible views of information. OLAP transforms raw data so that it reflects the real dimensionality of the enterprise as understood by the user. While OLAP systems have the ability to answer "who?" and "what?" questions, it is their ability to answer "what if?" and "why?" that sets them apart from Data Warehouses. OLAP enables decision-making about future actions [3].

This paper aims to view the effectiveness of using OLAP with clinical Data warehouse in order to help the clinicians in order to support their decisions. The clinical data warehouse built based on medical records from Basra Health Department, Basra, Iraq. Three OLAP consoles designed and implemented using Microsoft Excel 2013 Pivot table, SQL server Integration Service (SSIS) 2012 and SQL Server Reporting Service (SSRS) 2012.

II. RELATED WORKS

In [4] M. C. Tremblay, R. Fuller, D. Berndt, J. Studnicki studied an implementation of an OLAP interface on the Comprehensive Assessment for Tracking Community Health (CATCH) data warehouse used by knowledge workers at a regional health planning agency in the State of Florida to understand the impact of this type of tool on how this tool decision makers is flexible to customize the selection, aggregation, and presentation of data. Consequently, they were able to perform in more of a consultative role to their clients, and improved their reputation in the community they serve. This research adds a new dimension to prior research in data warehousing by focusing on the decision support capabilities of OLAP.

In [5] Nadim W. Alkharouf, D. Curtis Jamison, and Benjamin F. Matthews used Analysis Services 2000 to construct an OLAP cube that was used to mine a time series experiment designed to identify genes associated with resistance of soybean to the soybean cyst nematode, a devastating pest of soybean. The data for these experiments is stored in the soybean genomics and microarray database Compared to traditional cluster analysis of gene expression data, OLAP was more effective and faster in finding biologically meaningful information.

In [6] S. Palaniappan and C. Ling present a prototype CDSS using OLAP with Data Mining (OLAP Mining) which combines the strengths of both. This approach provides a rich knowledge environment, not achievable by using OLAP alone or data mining alone. They found that their proposed system based on OLAP and Data Mining is powerful as it discovers hidden patterns in the data, enhances real-time indicators and discovers bottlenecks and improves information visualization.

Stolba and Min Tjoashowed in [7] integrated data warehousing, OLAP, and data mining techniques in the healthcare area. Decision support platform which supports decision making process of caregivers and clinical managers is built. They presented three case studies, which showed that a clinical data warehouse that facility evidence-based medicine is a reliable, powerful and user-friendly platform for strategic decision making.

Hristovsk, Rogac, Markota [8] described the possibilities of using data warehousing and OLAP technologies in public health care in general and then their own experience with these technologies gained during the implementation of a data

warehouse of outpatient data at the national level. Such data warehouse serves as a basis for advanced decision support systems based on statistical, OLAP or data mining methods. They used OLAP to enable interactive exploration and analysis of the data. They found out that data warehousing and OLAP are suitable for the domain of public health and that they enable new analytical possibilities in addition to the traditional statistical approaches.

In [9] H. N. Khraibet, A. Mousa and M. S. Abu Bakar explained how to design and develop Intelligent Iraqi Health System (IIHS) by using (OLAP) model to help decision makers in the Iraqi Ministry of Health. Computer System Usability Questionnaire was used to evaluate the prototype of the IIHS. Finally they provided a guide for Business Intelligence developers and decision makers for Ministry of Health in Iraq.

III. DATA WAREHOUSE

Data warehousing is a collection of techniques to produce a single repository with unified schema to hold all the data from heterogeneous sources. It includes all operations of importing data from operational sources, cleaning them and transforming them in order to load them into data warehouse. It also holds operations of updating the data warehouse by refreshing it in order to load the new data so it can be viewed as multidimensional view by the users, analysts and report viewers.

It can be considered as intermediate stage between analyst and source data which holds all the operations of converting historical data into knowledge, as depicted in figure (1). The most common definition for data Warehouse is “subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management’s decision-making process” [10].

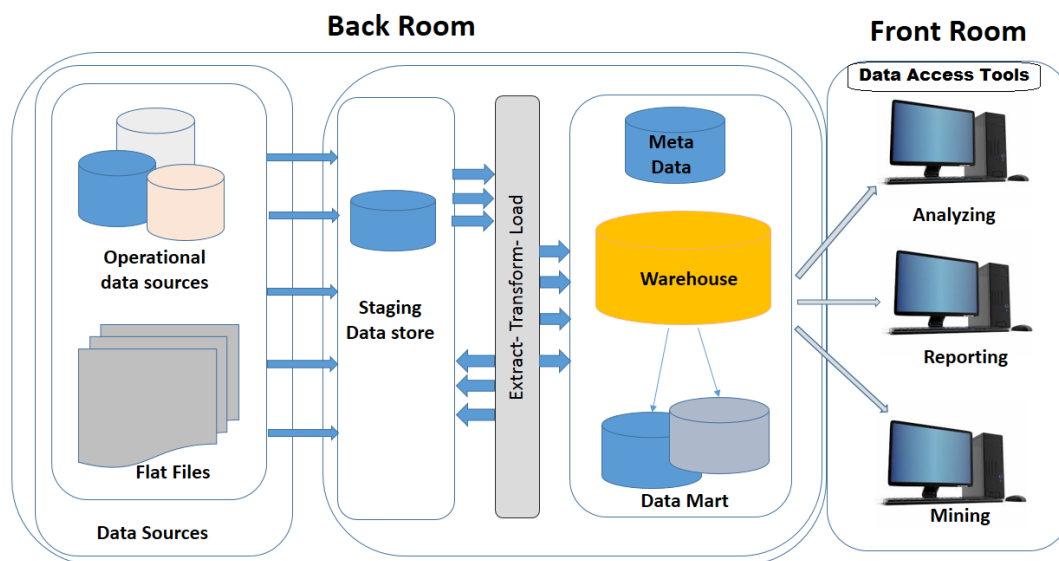


Figure 1. Data Warehouse

Data can now be stored in many different types of databases. One type of database architecture that has recently emerged is data warehouse, which is a repository of multiple heterogeneous data sources, organized under a unified schema at a single site in order to facilitate management decision-making. Data warehouse technology includes data cleaning, data integrating, and on-

line analytical processing (OLAP), analysis techniques with functionalities such as summarization, consolidation and aggregation, as well as the ability to view information from different angles [11].

Whereas a data warehouse can be at the enterprise level and serve the needs of the whole organization, a data mart can be at the department level and serve a small group of people. A data mart strives to analyze a particular area; a data warehouse can comprise a few or many data marts [12].

IV. ONLINE ANALYTICAL PROCESSING (OLAP)

The result of consolidating different kinds of databases in single warehouse forces us to use a different tool in query which differs completely in structure and result from the traditional tools used in relational databases. Since there are many tools used with data warehouse for analyzing but the most familiar tools used with data warehouse in query is OLAP. OLAP is a term that describes a technology that uses a multi-dimensional view of aggregate data to provide quick access to strategic information for the purposes of advanced analysis (Codd *et al.*, 1995). OLAP enables users to gain a deeper understanding and knowledge about various aspects of their corporate data through a fast, consistent, interactive access to a wide variety of possible views of the data. OLAP allows the user to view corporate data in such a way that it is a better model of the true dimensionality of the enterprise [13].

OLAP is a technology designed to provide high performance for query-business case. OLAP is designed to operate efficiently with the data organized according to common dimensional model used in the data warehouse. OLAP organizes data warehouse in multidimensional cubes based on this dimensional model, and then processing these cubes to provide maximum performance for query-that multiply the data in different ways. The key indicator of a successful OLAP application is its ability to provide information as needed, i.e., its ability to provide "just-in-time" information for effective decision-making. Just-in-time information is computed data that usually reflects complex relationships and is often calculated on the fly. Analyzing and modeling complex relationships are practical only if response times are consistently short. A truly flexible data model ensures that OLAP systems can respond to changing business requirements as needed for effective decision making [3, 14].

V. OLAP TYPES

There are different types of OLAP systems depend on type of data and platform which used to apply OLAP functionalities.

The major types are:

- a. *MOLAP* is a "multi-dimensional online analytical processing". 'MOLAP' is the 'classic' form of OLAP and is sometimes referred to as just OLAP. MOLAP stores this data in an optimized multi-dimensional array storage, rather than in a relational database. MOLAP tools generally utilize a pre-calculated data set referred to as a data cube. The data cube contains all the possible answers to a given range of questions. MOLAP tools have a very fast response time and the ability to quickly write back data into the data set [13, 15, 16].
- b. *ROLAP* works directly with relational databases. The base data and the dimension tables are stored as relational tables and new tables are created to hold the aggregated information. This methodology relies on manipulating the data stored in the relational database to give the appearance of traditional OLAP's slicing and dicing functionality. In essence, each action of slicing and dicing is equivalent to adding a "WHERE" clause in the SQL statement. ROLAP tools do not use pre-calculated data cubes. ROLAP also has the ability to drill down to the lowest level of detail in the database [13, 17, 16].
- c. "Hybrid OLAP (*HOLAP*) a database will divide data between relational and specialized storage. HOLAP database will use relational tables to hold the larger quantities of detailed data, and use specialized storage for at least some aspects of the smaller quantities of more-aggregate or less-detailed data. HOLAP addresses the shortcomings of MOLAP and ROLAP by

combining the capabilities of both approaches. HOLAP tools can utilize both pre-calculated cubes and relational data sources [18, 16].

VI. Data Warehouse with OLAP

Data warehousing developed, despite the presence of operational databases due to following reasons:

- An operational database is designed and tuned from known tasks and workloads, such as indexing using primary keys, searching for particular records and optimizing 'canned queries'. As DW queries are often complex, they involve the computation of large groups of data at summarized levels and may require the use of special data organization, access and implementation methods based on multidimensional views. Processing OLAP queries in operational databases would substantially degrade the performance of operational tasks.
- An operational database supports the concurrent processing of multiple transactions which mean the data can be altered and edited while using OLAP and data warehousing need read only property to access and view data.
- Decision support systems deal with historical data which need both data warehousing and OLAP to be processed, the operational systems deal with present data.
- Decision support needs consolidation (such as aggregation and summarization) of data from heterogeneous sources; and operational databases contain only detailed raw data [11].

VII. OLAP WITH HEALTHCARE

The Electronic Health Record (EHR) or the Electronic Patient Record (EPR) is on-line transaction processing (OLTP) system that enables on-line inserting and updating of given health care services and documentation, medical results tracking, and real-time deciding support. Because it contains information details about daily activities, such OLTP system has great OLAP capabilities in medical, financial, and administrative area. Health Care employers have understood lately the benefits of those systems, and have been beginning to show curtain interest in data analysis which would have helped them to easily achieve answers to number of every day situation questions. Unfortunately, the most part of classical OLTP EHR systems has not suitable support for OLAP Systems [19].

A data warehouse project was initiated in 1998 to automate much of the community assessment process, although the interpretation of the many health status indicators is still a matter of art. Today, the data warehouse integrates fine-grained event data such as vital statistics (birth and death records), hospital discharge data, and free-standing clinic data, along with several more detailed disease registries [4].

The combination of the ability to compare a number of variables efficiently and the ability to change the aggregation detail through the drill-down process creates the power in OLAP. Because OLAP data structures can be easily updated with new data, they can be used as basic dashboards for continuous measurement of system performance before and after process changes have been implemented. With little effort, the continuous OLAP reporting data can be converted to statistical process control charts for the management of operational change [20].

VIII. MODEL

The constructed system consists of data warehouse and OLAP consoles. The source databses of our data warehouse are designed and created by the developer depended on the form structure used by the employees in Basrah hospitals. Some fields are added to show the effectivence of adding more information to the form related to patients. The added fields are dependable by whole clinical instiutes like boold type, RH Factor. Two types of databases for diseases registry are built, the first one is relational sql database stored by SQL Server. The second database is flatfile type which is Comma Separated Values (CSV file). The construction of the proposed system is done by following two main steps which are:

A. Design Data Warehouse

The strategy of designing proposed data warehouse is shown in figure (2).

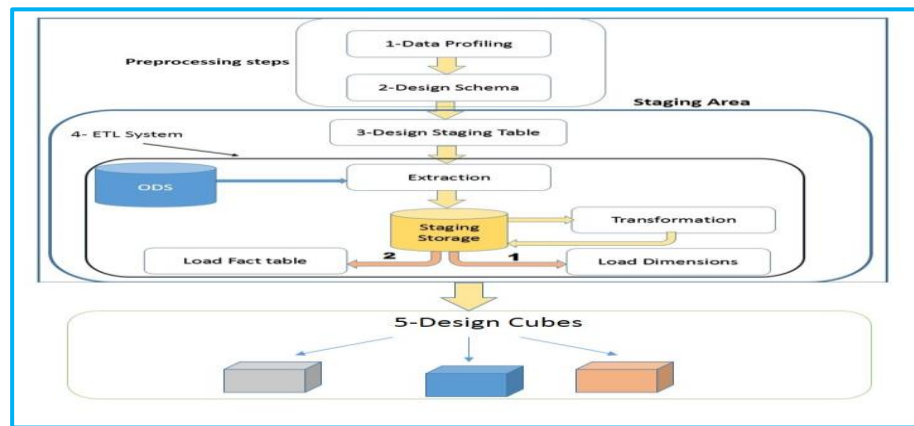


Figure 2. Data Warehouse Designing Strategy

The figure 2, show many steps followed to build data warehouse. Data profiling is a necessary operation since its results are so important to find candidate keys and to pre-analyse the data in the Operational Data Store (ODS). After analysing and finding candidate keys, the step of designing our data warehouse schema. The researcher chose star schema as architecture for our data warehouse for many reasons listed below. Finally the operation of designing ETL systems for each database and consolidating them under single schema is done to produce data warehouse.

1. **Data Profiling:** It is one of the important process which should be done before any step to data warehouse. Its results affect the whole operation of designing and implementing data warehouse and the whole performance of the system. Data profiling gives abundant information about the candidate keys to be in the dimensions and fact table, ratio of null records, columns values patterns and percentage of each value in each column. Generally, it provides analysis of operational source databases of your data warehouse.
2. **Design Schema:** Schema is the shape which we use to arrange the tables. Star schema is chosen as a structure for our disease data warehouse for the following reasons:
 1. Results from data profiling process led us to choose this architecture.
 2. The schema is easy to design, configure, and implement.
 3. We can adapt more dimensions in the future.
 4. Query response will be so fast.
 5. It can be understood by the analyst rather than the normal user, see figure 3.

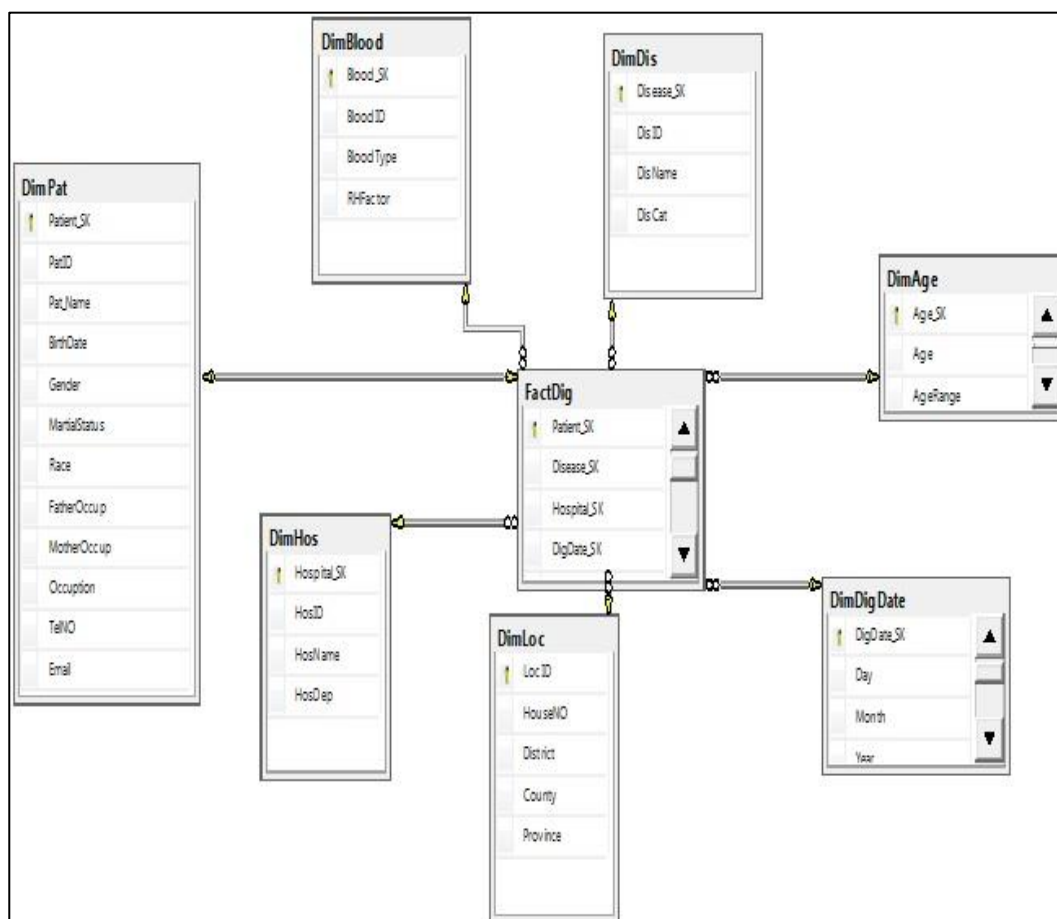


Figure 3. Data Warehouse Schema

3. *Design Staging Table*: the component of staging area will be explained. Staging area consists of staging table and ETL system. It is the intermediate component between Operational Data Store (ODS) and Data Warehouse. The operations of extraction data from data sources, transforming, cleaning and integrating them into single schema constraints and finally loading data into dimensions and fact tables are done in this area. It is just like the kitchen in the restaurant, the customer (clinician) should not access it, but the cooker (designer) receives and cooks the meals.
4. *ETL System*: The ETL system is configured depending on Sql Server Integration Service 2012. Two different ETL systems are built depending on database sources. The database kind, fields and data types of database determine the ETL tools used in building ETL system. Each ETL system holds different tools and strategy to follow in extracting, transforming and loading data from source data store to Data Warehouse.
5. *Design Cubes*: There are many reasons for creating cubes. They include:
 - Giving users access to the data they need to analyze most often.
 - Providing data in a form that matches the collective view of the data by a group of users in a department or business function.
 - Improving end-user response time due to the reduction in the volume of data to be accessed.
 - Appropriately providing structured data as dictated by the requirements of end-user access tools such as Online Analytical Processing (OLAP) and data mining tools, which may require internal database structures. In practice, these tools often create their data mart designed to support their specific functionality.
 - Normally data marts use less tasks such as data cleansing, loading, transformation, and integration are far easier, and hence implementing and setting up a data mart is simpler than establishing a corporate data warehouse.
 - The cost of implementing data marts is normally less than that required to establish a data warehouse.

- The potential users of a data mart are more clearly defined and can be more easily targeted to obtain support for a data mart project rather than a corporate data warehouse project [16].

B. OLAP

Many cubes are built based on dimension tables from the constructed data warehouse. These cubes are different in their construction depend on the intended results and reports. The result cubes are depended as a base for the report consoles, OLAP and KPIs. Figure (4) showed the cube DWHealthcare_1 which holds nearly almost dimensions of the data warehouse. The construction of this cube is done by using SQL Server Integration Service (SSIS) 2012. The figure shows the chosen dimensions.

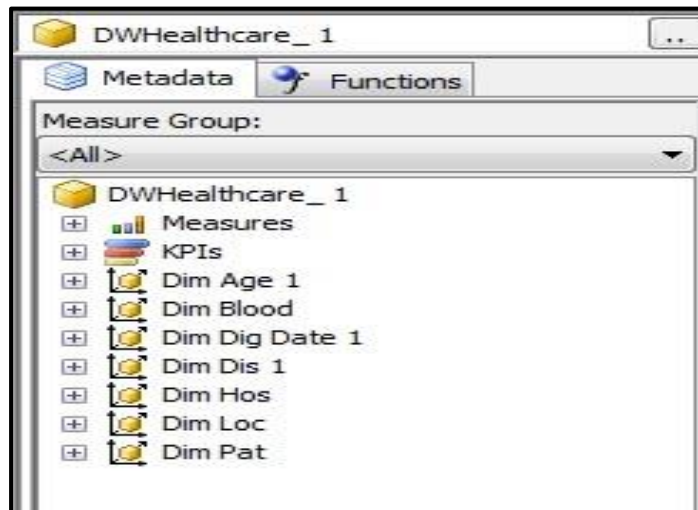


Figure 4. DWHealthcare_1 Cube

Three different consoles used to show how to use the operations of OLAP. The first one by using SSIS. SSIS offers the user the facility to view the result immediately by choosing the dimensions and measurements and the result will be displayed as a table, see figure 5.

Age Range	Blood Type	RH Factor	Dis Name	Count Diseases
2	O	+	ViralDis...	3
2	O	+	ViralEnc...	12
2	O	+	Whoopi...	4
2	O	+	WormIn...	24
3	A	+	Altyvoid...	8
3	A	+	Amoebiasis	59
3	A	+	BloodPol...	6
3	A	+	Brucellosis	2
3	A	+	Chicken...	50
3	A	+	Cholera	6
3	A	+	ClinicalS...	5
3	A	+	Croup	3
3	A	+	Diarrhea...	151

Figure 5. SSIS OLAP

The figure shows the number of infections for each disease according to age range and blood type. By using SSIS, user can navigate and view result easily and very fast by just dropping dimensions and measurements.

Second console designed by using Microsoft Excel 2013 Pivot Table. After connecting Excel Pivot Table to SQL Server Analytical Server which holds the designed cube, user can easily view the result as chart, see figure 6.

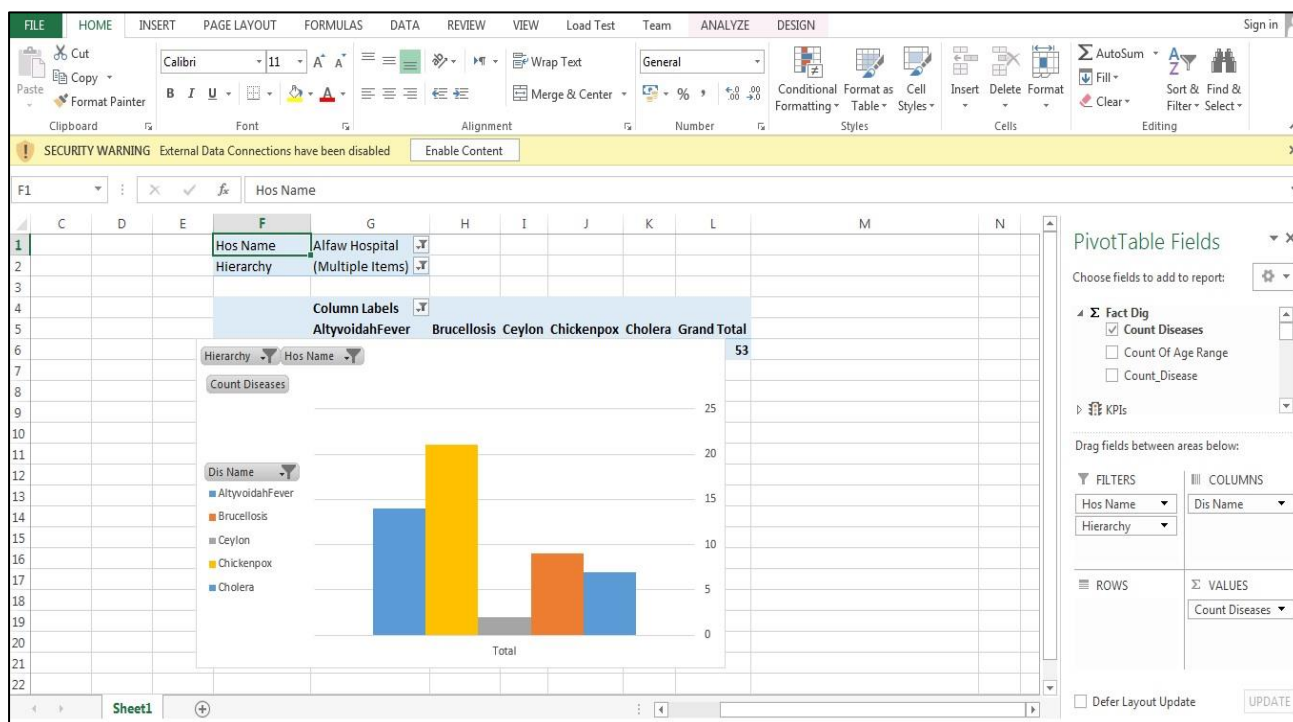


Figure 6. OLAP Excel Console

The figure shows the number of infections for each disease according to hospital. This console provides flexibility in filtering columns and rows to get more informations. User can choose columns and filter them by any selected values based on specific measurements in order to view reports. This console is very easy to use and so simple to understand the results obtained.

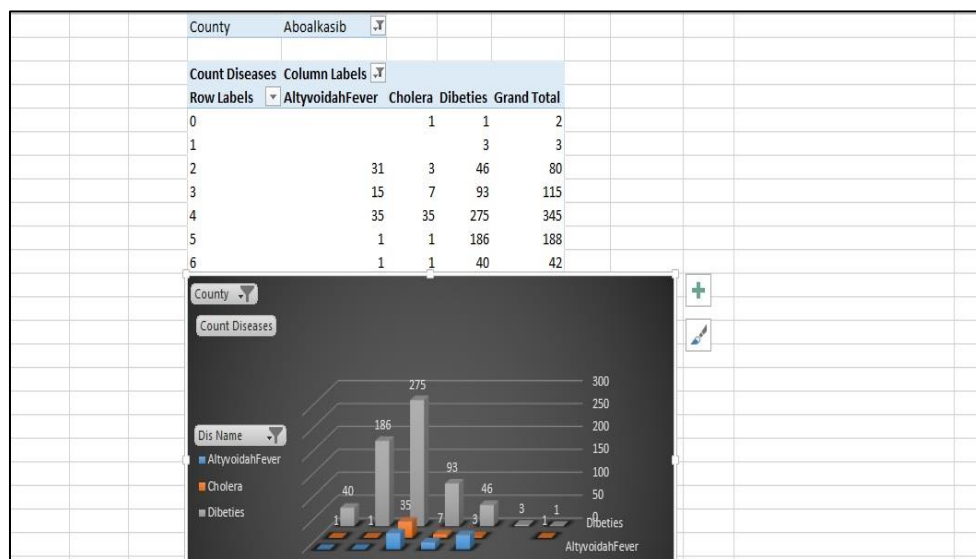
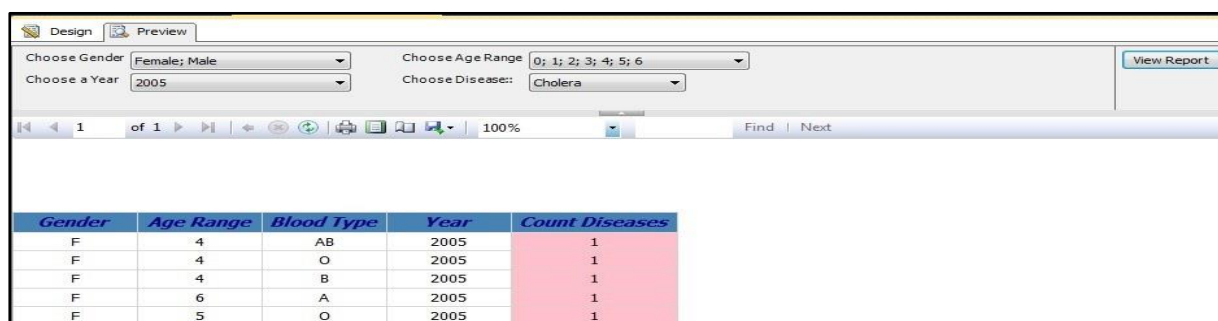


Figure 7. Excel OLAP for (Cholera, Diabetes and Altyvoid Fever)

Next figure 7 shows the result as 3-D chart representing infections for Aboalkaseeb County in Basra. The infections are (Cholera, Diabetes and AltyvoidFever). The result chart shows the total number of infections according to age range. The table in the figure shows the number for each category. User can choose more infections to compare with or to show the category which have more infections.

The next console designed by using SQL Server Reporting Service (SSRS) 2012. In this console, the chosen dimensions (Gender, Year, Age Range and Disease) will decide the viewed results. The results will be shown in different ways starting from table and ending with 3-D chart.



The screenshot shows an SSRS report interface. At the top, there are filters for 'Choose Gender' (Female; Male), 'Choose a Year' (2005), 'Choose Age Range' (0; 1; 2; 3; 4; 5; 6), and 'Choose Disease' (Cholera). A 'View Report' button is on the right. Below the filters is a table with the following data:

Gender	Age Range	Blood Type	Year	Count Diseases
F	4	AB	2005	1
F	4	O	2005	1
F	4	B	2005	1
F	6	A	2005	1
F	5	O	2005	1

Figure 8. OLAP Report

Based on user choices, the result table displayed count of infections for Cholera disease according to chosen genders, Age Ranges Blood Type in year 2005. It is flexible Console which can be accessed by URL.

IX. CONCLUSION AND DISSCUSSION

The success factor for any organization is to consolidate their related data in single platform and use this platform to watch their progress to achieve their goals. Data Warehouse is the most important technology which can be used as platform to unify department's data of organization. Iraqi Institutes often used different data storage to register their medical data. We highly recommending Iraqi Institutes to use this project to assist clinicians to view in-time result related to disease cases. Since OLAP is the most effective tool used with Data Warehouse, so it can be used to answering dimensional queries and showed result in fast and effective way. OLAP can be considered as tool of Business Intelligence (BI) since it provides effective possibilities to view multidimensional data from different points of view.

REFERENCES

- [1] P. JINPON, et al., "Business Intelligence and its Applications in the Public Healthcare System," *Walailak Journal of Science and Technology (WJST)*, vol. 8, pp. 97-110, 2011.
- [2] Elliott, Thomas E., et al. "Data Warehouse Governance Programs in Healthcare Settings: A Literature Review and a Call to Action." *eGEMs (Generating Evidence & Methods to improve patient outcomes)* 1.1 (2013): 15.
- [3] O. Council, "Apb-1 olap benchmark, release ii," <http://www.olapcouncil.org/research/bmarkly.htm>, 1998.
- [4] M. C. Tremblay, et al., "Doing more with more information: Changing healthcare planning with OLAP tools," *Decision Support Systems*, vol. 43, pp. 1305-1320, 2007.
- [5] N. W. Alkharouf, et al., "Online analytical processing (OLAP): a fast and effective data mining tool for gene expression databases," *BioMed Research International*, vol. 2005, pp. 181-188, 2005.
- [6] S. Palaniappan and C. Ling, "Clinical decision support using OLAP with data mining," *International Journal of Computer Science and Network Security*, vol. 8, pp. 290-296, 2008.

- [7] The Relevance of Data Warehousing and Data Mining in the Field of Evidence-Based Medicine to Support Healthcare Decision Making, Nevena Stolba and A Min Tjoa, Institute of Software Technology and Interactive Systems, Vienna University of Technology, Vienna, Austria, December 24, 2005.
- [8] Using Data Warehousing and OLAP in Public Health Care, Dimitar Hristovsk, Mija Rogac, Mladen Markota. Institute of Biomedical Informatics, Medical Faculty, University of Ljubljana, Slovenia Public Health Institute of the Republic of Slovenia, 2000.
- [9] Intelligent Iraqi Health System (IIHS) Using OnLine Analytical Process (OLAP) Model Hayder Naser Khraibet, Ayad Hameed Mousa, and Muhamad Shahbani Abu Bakar, Proceedings of the 4th International Conference on Computing and Informatics, ICOCI 2013 28-30 August, 2013 Sarawak, Malaysia. Universiti Utara Malaysia (<http://www.uum.edu.my>).
- [10] Building the Data Warehouse Third Edition, W. H. Inmon, 2002, W.H. Inmon, Inc., ISBN: 0-471-08130-2.
- [11] G. S. Reddy, et al., "DATA WAREHOUSING, DATA MINING, OLAP AND OLTP TECHNOLOGIES ARE ESSENTIAL ELEMENTS TO SUPPORT DECISION-MAKING PROCESS IN INDUSTRIES," International Journal on Computer Science & Engineering, 2010.
- [12] R. Verma and J. Harper, "Life cycle of a data warehousing project in healthcare," Journal of healthcare information management, vol. 15, pp. 107-118, 2001.
- [13] Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals. Paulraj Ponniah, 2001 John Wiley & Sons, Inc. ISBNs: 0-471-41254-6 (Hardback); 0-471-22162-7 (Electronic).
- [14] Kimball, R. (1996) The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses, New York: John Wiley & Sons.
- [15] <https://en.wikipedia.org/wiki/MOLAP>
- [16] Database Systems: A Practical Approach to Design, Implementation, and Management, Fourth edition, Pearson Education Limited 2005.
- [17] <https://en.wikipedia.org/wiki/ROLAP>
- [18] <https://en.wikipedia.org/wiki/HOLAP>
- [19] S. Pešić, et al., "Benefits of Using OLAP Versus RDBMS for Data Analyses in Health Care Information Systems," FACULTY OF ELECTRICAL ENGINEERING UNIVERSITY OF BANJA LUKA, p. 56, 2009
- [20] B. D. Gordon and B. R. Asplin, "Using online analytical processing to manage emergency department operations," Academic emergency medicine, vol. 11, pp. 1206-1212, 2004