DOI: 10.24996/ijs.2021.SI.1.32

ISSN: 0067-2904

On-line Analytical Processing (OLAP) Operation for Outpatient Healthcare

Talib M. J. Al Taleb*، Sami Hasan ،Yaqoob Yousif Mahd
College of Information Engineering, Al-Nahrain University, Baghdad, Iraq.

Abstract:
This paper presents an architecture for the data warehouse of outpatient healthcare (DWOP) as a data repository collects data from two different sources (Databases of outpatient healthcare and Excel files from hospitals) and provides storage, functionality and responsiveness to queries to meet decision makers requirements.
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. "On-Line Analytical Processing (OLAP) is utilized for decision support to get interesting information" from the data warehouse with a rapid execution time. OLAP is considered one of Business Intelligence tools.

Keywords: Data Cubes, Data warehouse (Star schema), OLAP**.**

Introduction:

A data warehouse is a repository of variety data sources, organized under a unified schema at a single site in order to enable more effective strategic, tactical, and insights that are useful for guiding decision making [1]. And Kimball defines it as "A copy of transaction data specifically structured for query and analysis" [2].

A star schema model is adopted in this paper to represent a dimensional model because it is less complex, most widely used to develop data warehouses, and a more effective schema for handling simpler queries, fast aggregations and feeding cubes efficiently [3,4]. See Figure-1.

OLAP is a specialized analysis tools which is designed to provide excellent query performance, fast incremental updates of data sets, efficient management of summary data and rich analytic content, so it can be efficiently used for data presentation in data warehouse environment [5, 6]. OLAP is a category of software technology that provides the users many benefits to access the data and enables them to analyze multidimensional data interactively from different views of the data proportionate with the requirements of the business. [7]. Feature is essential of OLAP is "a multidimensional analysis"[8,9]. In other words, the ability to analyze metrics in different dimensions such as time, geography, disease, nationality, etc. [10, 11].

Data Cubes

Cube is a term that refers to multi-dimensional array of data "which are organized into multiple dimensions, see Figure-2. Each dimension contains multiple levels of abstractions defined by concept hierarchies. This organization provides the users a flexibility to view data from different perspectives, allowing interactive querying and analysis of the data at hand" [8, 2].

OLAP consists of four basic analytical operations [8]:

1. "Roll up: An operation for moving up the hierarchy level and grouping into larger units along a dimension. it is also called the drill up operation".
2. "Drill down: An operation for moving down the hierarchy level and stepping down the hierarchy".

*Email: talib_altalib@yahoo.com

3. “Slice and dice: The slice operation performs a selection on one dimension of the given cube and results in a sub cube. The dice operation defines a sub cube by performing a selection on two or more dimensions” of a cube [12].
4. “Pivot (rotate): Pivot (also called rotate) is a visualization operation that rotates the data axes in view in order to provide an alternative presentation of the data”. “By this way, the data can be rolled up, sliced, and diced as needed to handle the widest variety of questions that are relevant to a user’s area of interest”.

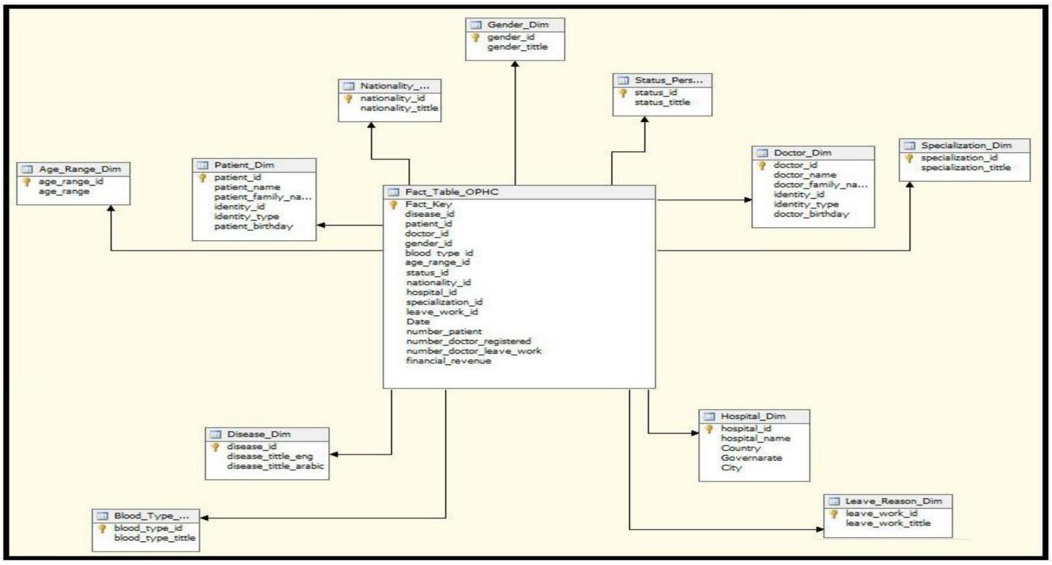

Figure 1-Star Schema

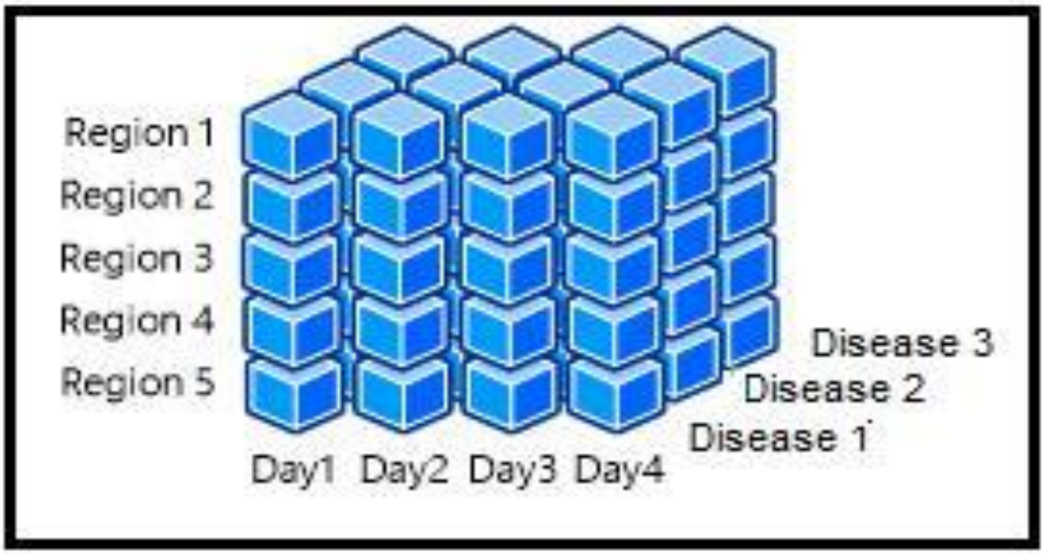

Figure 2-Cube

Cube Storage Modes

“Physical storage options affect the performance and storage requirements for cubes. The cube can be stored in a MOLAP (multidimensional OLAP) structure, a ROLAP (relational OLAP) database, or a HOLAP (hybrid OLAP) combination of multidimensional structure and relational database” [13, 14]. Table-1 is a summary comparison of the three cube storage options.

MOLAP: It is used a multidimensional structure in Analysis Service to store a copy of the base data and cube’s aggregations. It is suitable “for cubes with frequent use and the necessity for speedy query response due to optimize storage” and effectively multidimensional indexing. When source data changes, data in MOLAP storage should be periodically processed to incorporate those changes and make them available to users.

ROLAP: It is used “tables in the data warehouse relational database to contain a cube’s aggregations. In contrast to MOLAP storage, ROLAP does not store a copy of the base data, accessing the fact table when necessary to response queries”. However, ROLAP enables users to view data in real time and it is used for large data sets that are infrequently queried such as historical data from less recent previous years.

Table 1-Three cube storage comparison

Feature	MOLAP	ROLAP	HOLAP
Base data storage	Cube	Relational table	Relational table
Aggregation storage	Cube	Relational table	Cube
Query performance	Fastest	Slowest	Fast
Storage consumption	High	Low	Medium
Maintenance	High	Low	Medium

Comparison between OLAP and OLTP

The databases that a business uses to store all its transactions and records are called online transaction processing (OLTP). This system covers daily operations of an institution. However, it could not designed for analysis. On the contrary, OLAP serves users or knowledge workers in the role of data analysis and decision making [15]. Table-2 is a summary comparison between features of OLAP and OLTP.

Table 2-“Comparison between OLTP and OLAP"

Feature	OLTP	OLAP
“Characteristic”	“Operational Processing”	“Informational processing”
Data Base Design	ER Model	Dimensional data model
“Function”	“Day-to-day operation”	“Long-term informational requirements, decision support”
Data	Current; guaranteed up to date	historical
View of data	“Detailed, flat relational”	“Summarized, multidimensional”
“Unit of work”	“Short, simple transaction”	“Complex query”
“Access”	“Read/write”	“Mostly read”
Help people	Normal users to carry out activities	Managers to make decision
Speed with huge data	Slower	Faster

OLAP Operation

A data warehouse is a place to store data with a design that makes analysing data easier, and OLAP is a method to analyse data as well as to provide self-service business intelligence capabilities to decision makers [16]. Microsoft SQL Server - Analysis Services software (SSAS) was used to build the number of OLAP Cubes that meet the main decision makers' queries. These cubes were stored in SSAS and MOLAP storage was adopted. Microsoft SQL Server - Report Services software (SSRS) and Excel Power Pivot targeted these cubes in SSAS and used for visualization of data from multiple perspectives [17]. Create Multiple and smaller cubes that can be much more user friendly and lead to faster query performance than one monster cube, especially if fact tables have very different dimensionality. Increasing the number of attributes, measures and dimensions in the cube space will start costing performance, it does not add to the storage cost, but it does hurt formula engine performance. Also multiple and smaller cubes are easier and less complex to scale out in case of OLAP distributed. Many cubes were created to improve the performance of the main queries that decision makers are asking about them constantly, which are:

a. Disease_View.cube: the creation of this cube is to answer queries concerning the number of infections for each disease according to governorate, hospital, gender, age range through definite period of time. Therefore, this Cube has been created which takes its data from the fact table

(Fact_Table_OPHC), and it uses the measure (number_patient) and five dimensions (Disease_Dim, Gender_Dim, Age_Range_Dim, Hospital_Dim, Time_Dim), see Fig. (3).

b. Doctor_View.cube: through this cube, decision makers could reach to know the doctors' number who are continuing their jobs according to governorate, hospital, gender and specialization. Therefore, this Cube has been created which takes its data from the fact table (Fact_Table_OPHC), and it uses the measure (number_doctor_continuous_work) and four dimensions (Doctor_Dim, Gender_Dim, Hospital_Dim, Specialization_Dim).

c. Revenue_View.cube: this cube has been created to show the financial revenues from various hospitals according to governorate and hospital through definite period of time.

d. Blood_Disease.cube: some diseases are related to the blood such as hemolytic anemia, decision makers require to know the type of blood for patients with these diseases and this Cube has been created for.

e. Nationality.cube: this cube gives responses about the nationality of patients with specific diseases or the number of patients who visited the healthcare centers.

f. Doctor_See_Patient.cube: through this cube, the rate of patients seen by a doctor could be known.

g. Left_Doctor.cube: this cube is used to get answers about the number of doctors left their works and why.

SSAS allows the user to view the result immediately by select the dimensions and measures and the result will be shown as a table. Figure-4 illustrates an example about how to use the operations of OLAP

Cube (Disease_Indexed_View.cube) by using SSAS, it shows the number of infections for different diseases in Baghdad through January 2000.

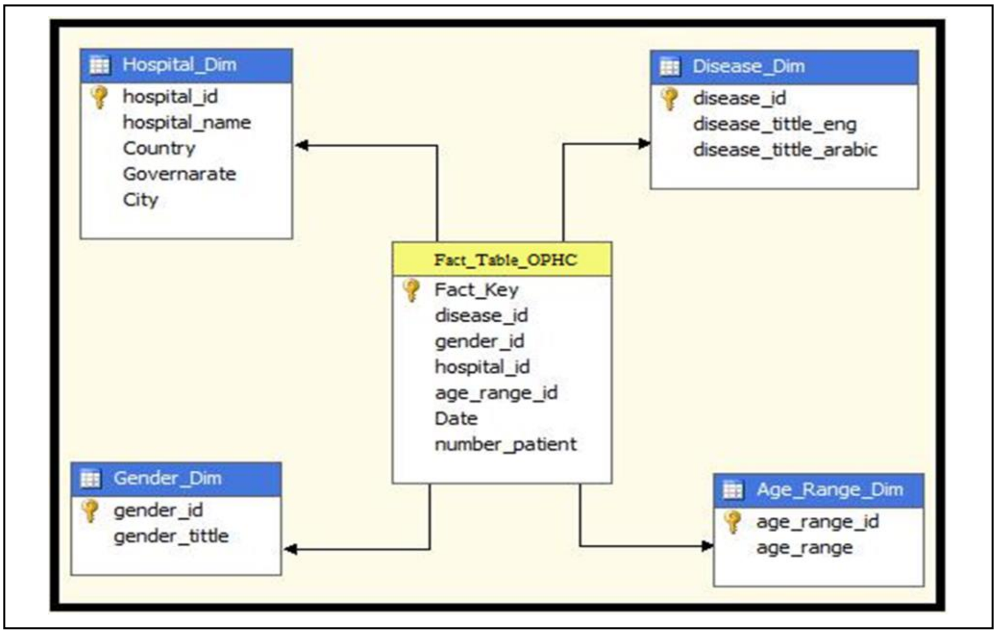

Figure 3-OLAP schema (Disease_ View.cube)

Figure-5 illustrates an example about how to utilize roll up operation, the dimension (Time_Dim) is

selected and zoom out of it to see a summarized level about the number of patients with cholera in the year.

Figure-6 illustrates an example about how to utilize drill down operation, the dimension (Time_Dim) is selected and zoom in to see a higher level of detail about the number of patient with cholera in each month of the year.

Figure-7 shows the slice operation, one dimension (Hospital_Dim) is selected. This figure displays the number of infections for all diseases registered in hospitals in Baghdad and AL-Basrah governorates in January 2000 that represents a slice out of the data cube.

Figure-8 shows the dice operation. This figure shows the number of infections for a limited number of diseases, the date and hospitals dimensions include the same range as previous Figure-7, this allows analyst to reach particular values of multiple dimensions.

Gov_Health	year- month										
Baghdad	January 2000	Other bacterial intestinal infections									
	Gender Id	Age Range Id									
	female	male									Grand Total
		less than 1 year	1-4	5-9	10-14	15-19	20-44	45-65	greater than 65 years	Total	
Disease Tittle Eng	Number Patient	Number Patient	Number Patient	Number Patient	Number Patient	Number Patient	Number Patient	Number Patient	Number Patient	Number Patient	Number Patient
Alzheimer	44								25	25	69
Anthrax				22						22	22
Acute Flaccid Paralysis	6	11	23	3	25					62	68
Acute hepatitis A	15					24	33			57	72
Arenaviral hemorrhagic fever	27		12		3					15	42
Chickenpox	21	3			21	9				33	54
Cholera	66	11	12		13	9		15		60	126
Diphtheria	52	15			13					28	80
Echinococcosis	47							6		6	53
Leprosy	12			28	21					49	61
Measles	14		37	39						76	90
Mumps	52				13		29	33	9	84	136
Other bacterial intestinal inf...	33	12	16		9					37	70
Pediculosis and phthiriasis	17		21	9						30	47
Poliomyelitis	35	30	17							47	82
Respiratory tuberculosis	58		9	13	13	9				44	102
Rubella Suspicion	43			3		17	17			37	80
Scabies	33		18	15	15					48	81
Scarlet fever	25			13	13					26	51
Schistosomiasis [bilharziasis]	44		19	20	10					49	93
Typhoid Fever	45			6			11			17	62
Viral meningitis	27		10							10	37
Whooping cough	45	10				27	15	9		61	106
Grand Total	761	92	194	171	169	95	108	63	34	926	1687

Figure 4-SSAS OLAP (Disease_ View.cube)

Disease Tittle Eng			
Cholera			
	Governarate	Hospital Id	
	Al Basrah	Baghdad	Grand Total
Year	Number Patient	Number Patient	Number Patient
Calendar 1989	262	368	630
Calendar 1990	241	353	594
Calendar 1991	242	367	609
Calendar 1992	219	312	531
Calendar 1993	205	325	530
Calendar 1994	236	344	580
Calendar 1995	187	297	484
Calendar 1996	219	312	531
Calendar 1997	242	367	609
Calendar 1998	241	353	594
Calendar 1999	262	368	630
Calendar 2000	205	325	530
Calendar 2001	236	344	580
Calendar 2002	187	297	484
Calendar 2003	219	312	531
Calendar 2004	242	367	609
Calendar 2005	241	353	594
Calendar 2006	262	368	630
Grand Total	4148	6132	10280

Figure 5-Roll up operation

Disease Tittle Eng: Cholera

Year	Half Year	Quarter	Month	Al Basrah (Number Patient)	Baghdad (Number Patient)	Grand Total (Number Patient)
⊞ Calendar 1989				262	368	630
⊞ Calendar 1990				241	353	594
⊞ Calendar 1991				242	367	609
⊞ Calendar 1992				219	312	531
⊞ Calendar 1993				205	325	530
⊞ Calendar 1994				236	344	580
⊞ Calendar 1995				187	297	484
⊞ Calendar 1996				219	312	531
⊞ Calendar 1997				242	367	609
⊞ Calendar 1998				241	353	594
⊞ Calendar 1999				262	368	630
⊟ Calendar 2000	⊟ Semester	⊟ Quarter 1, 2000	January 2000	80	126	206
			Total	80	126	206
		⊞ Quarter 2, 2000		66	101	167
		Total		146	227	373
	⊞ Semester 2, 2000			59	98	157
	Total			205	325	530
⊞ Calendar 2001				236	344	580
⊞ Calendar 2002				187	297	484
⊞ Calendar 2003				219	312	531
⊞ Calendar 2004				242	367	609
⊞ Calendar 2005				241	353	594
⊞ Calendar 2006				262	368	630
Grand Total				4148	6132	10280

Figure 6-Drill down operation

year- month: January 2000

Governarate	Hospital Id	Number Patient
⊟ Al Basrah	AL-Basrah	545
	Al-Fayhaa	340
	Umm Qasr	441
	Total	1326
⊟ Baghdad	AL-Nahrain	450
	AL-Yarmouk	546
	Medical City	691
	Total	1687
Grand Total		3013

Figure 7-Slice operation Ultimately, successful OLAP supports healthcare institutions to compete in providing best healthy services that help to identify and prioritize problems, develop and evaluate policies and actions, and organize health service delivery.

year- month: January 2000

Governarate	Hospital Id	Disease Tittle Eng	Number Patient
⊟ Al Basrah	⊞ AL-Basrah		180
	⊞ Al-Fayhaa		143
	⊞ Umm Qasr		132
	Total		455
⊟ Baghdad	⊟ AL-Nahrain	Alzheimer	25
		Cholera	42
		Measles	23
		Mumps	34
		Rubella Suspicion	24
		Typhoid Fever	15
		Total	163
	⊞ AL-Yarmouk		174
	⊞ Medical City		226
	Total		563
Grand Total			1018

Figure 8-Dice operation

Conclusion:

This paper presents the relation between OLAP and data warehouse which contains huge amount of structured data, and how does OLAP play a great role to support health environment especially for every institutions that collect large amounts of data by giving decision makers an efficient way to access these healthcare data to support their decision making in higher quality and more timely than those made traditionally.

OLAP allows users to investigate the data both in aggregated and detailed form. It enables the user to move from one level of aggregation to another easily, in an intuitive fashion.

References:

1. Al Taleb TM, Hasan S. and Mahdi YY. **2018**. Enhance the Performance of a Data Warehouse by Indexed View. *Journal of Fundamental and Applied Sciences*. 2018; **10**(6S): 211-5.
2. Kimball R. and Ross M. **2013**. *The data warehouse toolkit: The definitive guide to dimensional modeling*. John Wiley & Sons; 2013 Ju .l 1
3. Narra, Lekha, Sahama, Tony, & Stapleton, Peta. **2015**. Clinical data warehousing: a business analytics approach for managing health data. In Maeder, A & Warren, J (Eds.) Proceedings of the 8th Australasian Workshop on Health Informatics and Knowledge Management [Conferences in Research and Practice in Information Technology (CRPIT), Volume 164]. Australian Computer Society, Australia, pp. 101-104.
4. Kimball R., Ross M., Thornthwaite W., Mundy J. and Becker B. **2008**. *The data warehouse lifecycle toolkit*. John Wiley & Sons; 2008 Jan 10.
5. Qwaider WQ. **2000**. Apply on-line analytical processing (OLAP) with data mining for clinical decision support*International Journal of Managing Information Technology* .,:(1)**4** 37-25.
6. Train OL. and Jacobson R. **2000**.*The Microsoft SQL Server 2000 Analysis Services Step by Step*. Microsoft Press; 1999 Dec 1.
7. Singh K., Upadhyaya S. and Cantt M. **2011**. Constructing the olap cube from relational databases/flat files. *International Journal of Computer* .(2)**2** ,*Trend and Technology*
8. Colossi N, Malloy W. and Reinwald B. **2002**. Relational extensions for OLAP. *IBM Systems Journal*. 2002; **41**(4): 714-31.
9. Ponniah P. **2004**. *Data warehousing fundamentals: a comprehensive guide for IT professionals*. John Wiley & Sons; 2004 Apr 7.
10. Tariq A. and Rafi K. **2012**. Intelligent Decision Support Systems-A Framework. *In Information and Knowledge Management*, 2012, **2**(6): 12-20.
11. Zhang Q. and Zeng F. **2011**. Data warehouse and OLAP technology applied in the financial analysis of listed company. Computer Science2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC).
12. Eavis T., Dimitrov G., Dimitrov I., Cueva D., Lopez A. and Taleb A. **2010**. Parallel OLAP with the Sidera server. *Future Generation Computer Systems*. 2010 Feb 1; **26**(2):259-66.
13. Nolan S, Huguelet T. **2000**. *Microsoft SQL server 7.0 data warehousing training kit*. Microsoft Prees. USA. 2000..
14. Mailvaganam H. Designing OLAP Solutions
15. Parekh A.**2013**. Introduction on Data Warehouse with OLTP and OLAP. *International Journal Of Engineering And Computer Science*. 2013; 2569.
16. Patel A. and Patel J. **2012**. Data modeling techniques for data warehouse. *International Journal of Multidisciplinary Research*. 2012 Feb; **2**(2): 240-246.
17. Misner S. and Mistry R. **2012**. Introducing Microsoft SQL Server.