EARTHQUAKE DATA STORAGE AND STATISTICS VISUALIZATION

Nadir M Junainah[1], Ali Morfeq[2], and Muhammed shafiqe[3]
[1] King Abdulaziz University, Department of Electrical and Computer Engineering, Faculty of Engineering, Jeddah, Saudi Arabia
njunainah0001@stu.kau.edu.sa
[2] King Abdulaziz University, Department of Electrical and Computer Engineering, Faculty of Engineering, Jeddah, Saudi Arabia
morfeq@kau.edu.sa
[3] King Abdulaziz University, Department of Electrical and Computer Engineering, Faculty of Engineering, Jeddah, Saudi Arabia
msmuhammad@kau.edu.sa

Abstract— *This paper focuses on the illustration and assessment of seismic hazard through the creation of an integrated database system that includes the Kingdom of Saudi Arabia's earthquakes. Each earthquake has a known magnitude, longitude, latitude and depth.*
We created another database that includes seismic divisions in the regions by using Ray-tracing approach and point in polygon algorithm, to know the location of each earthquake. We made a script to transfer these earthquakes data from the first database to the other database (etl-extract, transfer, load). Then we used Olap and in memory database techniques to work on earthquake statistics and put them on the map. Data storage operations and OLAP operations are implemented, along with data mining tools such as the aggregation to illustrate the time-location-magnitude and other aspects. Data extraction and data applications help to ascertain uncertainty about behavioral problems in seismic risk mitigation and disaster management.

Keywords**:** *earthquake, Data Warehouse, OLAP, data mining.*

I. INTRODUCTION

A simple definition of earthquake is: Earthquakes are sudden earth movements resulting from the sudden release of accumulated energy over a long period of time. Where earthquakes negatively affect the development of many countries, it causes the deterioration of the economy of countries and the destruction of the environment [1]. Given uncertainty of earthquake occurrence, we need to find related patterns to reduce the impact of earthquakes by accessing and handling large amounts of data using modern techniques to search and find data. This earthquake data analysis is very important for decision makers during disaster management, land use planning and earthquake risk mitigation. The first technique is a set of historical data on a specific field-oriented, integrated, and time-changing characteristic [2] that provides operations that can handle data.

Data mining is a process to identify valid, new, useful and ultimately understandable patterns of data [3]. The basic task of extracting data is to discover the rules; a set of relationships between variables can be created, allowing for useful analysis and description of the model. In this category, there are techniques such as correlation, factors, assembly rules, and bindings [3].

These rules can be used to predict some output variables. Sequences may be in time, in the future, to predict the occurrence of catastrophic events such as earthquakes [3].

The paper is organized as follows. Section 2 provides a background about ETL (Extract, Transform, and Load) and OLAP Technical. Section 3 discusses related work. Section 4 presents methodology using OLAP Technical and ETL, Finally, Section 5 discusses the result, and Section 6 summarizes our findings.

II. BACKGROUND

Seismic data constitute an ever-increasing set of earth science information for processing and analysis. Earth scientists and local or national management staff always work to collect these data for scientific or planning purposes to construct data extraction, transformation, load, processing and visualization system. Data aggregation is any process in which information is collected and presented in a simple summary form, for purposes such as statistical analysis. A common aggregation purpose is to get more information about particular groups based on specific variables [4].

In this article, we discuss some of these tools

A - ETL - Extract, Transform, Load

ETL: is short form of extract, transform, load, three database functions that are combined into one tool to pull data out of one database and place it into another database. A python script is written to perform this task.

Extract: is the process of reading data from a database. In this stage, the data is collected, often from multiple and different types of sources.

Transform: is the process of converting the extracted data from its previous form into the form it needs to be in so that it can be placed into another database. Transformation occurs by using rules or lookup tables or by combining the data with other data.

Load: is the process of writing the data into the target database.

How it Works

Data from one or more sources are extracted and then copied to the data warehouse. When dealing with large volumes of data and multiple source systems, the data is consolidated. ETL is used to migrate data from one database to another, and is often the specific process required to load data to and from data marts and data warehouses, but is a process that is also used to too large convert (transform) databases from one format or type to another[3]. A python script is written to perform this task. Figure 1 shows data transfer from-to ETL.

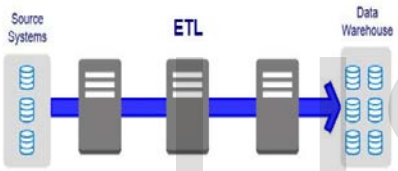

Fig. 1. Data from one or more sources is extracted and then copied to the data warehouse[3]

B- OLAP Technical

OLAP (Online Analytical Processing) is the technology behind many Business Intelligence (BI) applications. OLAP is a powerful technology for data discovery, including capabilities for limitless report viewing, complex analytical calculations, and predictive "what if" scenario (budget, forecast) planning[5].

OLAP overview

OLAP is a characterization of applications. It is not a database design technique. People often confuse OLAP with specific physical design techniques or data structure. This is a mistake. OLAP is a characterization of the application domain centered on slice-and dice and drill down analysis. Figure 2 presents OLAP workflow.

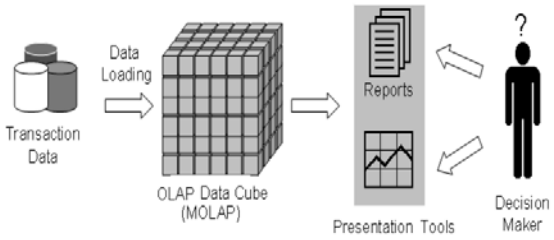

Fig. 2. How does OLAP pieces fit together[6]

How is OLAP Technology Used?

OLAP is an acronym for Online Analytical Processing. OLAP performs multidimensional analysis of business data and provides the capability for complex calculations, trend analysis, and sophisticated data modeling. It is the foundation for many kinds of business applications for Business Performance Management, Planning, Budgeting, Forecasting, Financial Reporting, Analysis, Simulation Models, Knowledge Discovery, and Data Warehouse Reporting. OLAP enables end-users to perform ad hoc analysis of data in multiple dimensions, thereby providing the insight and understanding they need for better decision making.

The relationship between DWH (Data Warehouse) or data mart & OLAP (Online Analytical Processing):

A data warehouse is a "subject-oriented, integrated, time-varying, non-volatile collection of data that is used primarily in organizational decision making. Typically, the data warehouse is maintained separately from the organization's operational databases and on different (and more powerful) servers. There are many reasons for doing this. The data warehouse supports on-line analytical processing (OLAP), the functional and performance requirements of which are quite different from those of the on-line transaction processing (OLTP) applications, traditionally supported by the operational data bases[7].

C- Characteristics of seismic waves

At all distances from the focus, rock layers and physical properties of surface soil affect the characteristics of the wave. , The mechanical properties of rocks, such as compression, stiffness, and density, play a role in the speed at which the waves pass and the shape and orbits of waves.

In most cases, flexible behavior occurs in earthquakes, but the strong shaking of surface soil from seismic waves that sometimes occur leads to inflexible behavior, including downward (ie, downward motion and out of non-uniform materials) and the removal of sandy soils[8].

D -Measurement of seismic waves

Seismograph: A device that records seismic waves caused by an earthquake, explosion or other phenomenon that vibrates the earth. Where seismic equipment is equipped with electromagnetic sensors that translate earth movements into electrical changes, then processed and recorded by analog or digital circuits of the devices. seismograph and seismometer systems are often used interchangeably. However, while both devices may detect and measure seismic waves, the seismograph has the ability to record phenomena. The log produced by a seismograph is called a seismogram on a screen or printed paper [9].

Accelerometer: a tool that measures the rate at which the velocity of an object is changing (ie, its acceleration) changes. Acceleration cannot be measured directly. The accelerometer measures the force exerted by the constraints placed on the reference mass to stabilize its position in an accelerated body. Acceleration is calculated using the relationship between the force of restriction and acceleration given by Newton's second law: force = mass × acceleration[10].

E -Visualizing 100 Years of Earthquakes

From the exploratory survey of earthquakes, we will notice that earthquakes seem to occur frequently at the end of the interval (beginning in the 1960s). This is because seismologists have become better at measuring and recording earthquakes, not because earthquakes have become more frequent.

There are fewer large earthquakes at centers of propagation extending from compressive compression zones. There is also a wide area of seismic activity stretching from Italy and Greece through Asia Minor, Iran, Central Asia and the Himalayas to China. This is a huge pressure zone caused by the African, Indian and other small plates that collide with Eurasia.[11]. Figure 3 shows the distributions of the earthquakes epicenters in and around the kingdom, it shows that the earthquake activity is concentrated along the Red Sea rift, Gulf of Aqaba, Al-Ayis city and around Jazan.

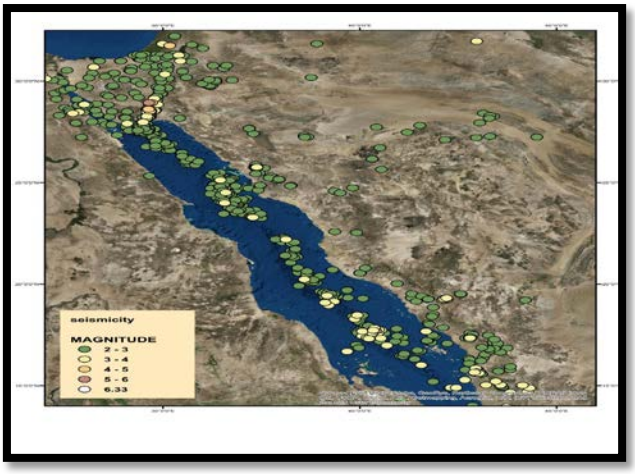

Fig. 3. Siesmic Activity

F - Al-Ais city

Al-Ais city is one of the cities belonging to Madinah in the Kingdom of Saudi Arabia, especially in the Kingdom of Saudi Arabia, with an estimated area of about twenty nine thousand square kilometers, and its population is estimated at about sixty-six thousand people based on the statistics for 2006. Rising from the sea surface about 1200 meters, a warm summer and cold in winter.

Jazan City: Jazan city is located at the south Western part of Kingdom Saudi Arabia. It is considered as one of the biggest cities in the kingdom. The Saudi government paid a great attention to the southern part of the kingdom and especially Jazan. The Saudi government intends to construct a new industrial city in Jazan and also, many urban and industrial projects are being built in Jazan. Jazan is considered as one of the mining centers in the kingdom.

III. RELATED WORK

In [12] this paper, presents various techniques to visualize time varying, 3D earthquake modeling datasets. The team worked described three different programs for a complete visualization of seismic data. First, the team created animations using the parallel terrain display system, RIVA, to overlap the displacement surface mapping scheme created by GeoFEST and Virtual California at the top of LandSAT imagery and digital terrain. Second, the team envisions natural stress and shear stress on the rusty parts produced by GeoFEST interactively using the ParVox Parallel Network Size Viewer. Thirdly, the team used an interactive imaging and zooming system, MSLT, to visualize faulty surfaces, simulated stress and slippages on damaged surfaces beneath a realistic terrain image.

In [13] This paper focuses on the prediction of seismic hazards and also for the prevention of earthquake damage through the use of a decision-making system. It implements data storage and OLAP operations, along with data mining tools such as pairing rules, decision trees and aggregation to predict aspects such as location, time of year, magnitude of earthquake, etc.

In [14] this research, the research team proposed a new SDMMS structure, then they described its function and identified potential benefits. By providing extensive examples, the paper discussed spatial-temporal concepts necessary for modelling seismic activity, efficient storage and retrieval of seismic data, KDD and DW techniques to handle a large number of observations Available seismic analysis and visualization techniques that enable the user to fully utilize SDMMS capabilities by connecting them to KDD operations

IV. METHODOLOGY

This paper presents a methodology using polygon analysis for seismic data sets, creating summaries of polygonal meta collections. For polygon generation, our work uses a specific algorithm called point in polygon algorithm [15] to generate polygons from continuous density functions or functions as described in [1][2].

We have made two databases

- The first database includes the earthquakes data in Saudi Arabia that observed during specific periods. The data of each earthquake were recorded in terms of earthquakes size, longitude, latitude and error rate.
- The second database includes subsets. Each subset represents the seismic activity in each division of the Kingdom of Saudi Arabia. We transferred the earthquake data from the production database to the second database via etl (extract, transfer, load). Then we used Qlick sense program to make earthquake statistics and put them on the map using a mixture of OLAP and memory database. The data selection is performed using point in polygon algorithm. The point in polygon algorithm is performed using the following steps [15]. The implemented system is shown in figure 4.

Fig. 4. Our design

1. we run a semi-infinite ray horizontally (increasing x, fixed y) out from the test point, then we count how many edges it crosses. At each crossing, the ray switches between inside and outside. This is called the Jordan Curve Theorem.
2. Through the top of the head, we treated the beam case to go head properly by careful selection of inequality

This pretends to shift the ray infinitesimally down so that it either clearly intersects, or clearly doesn't touch. Since this is merely a conceptual, infinitesimal, shift, it never creates an intersection that didn't exist before, and never destroys an intersection that clearly existed before. We followed the procedures as mentioned in figure 5 .

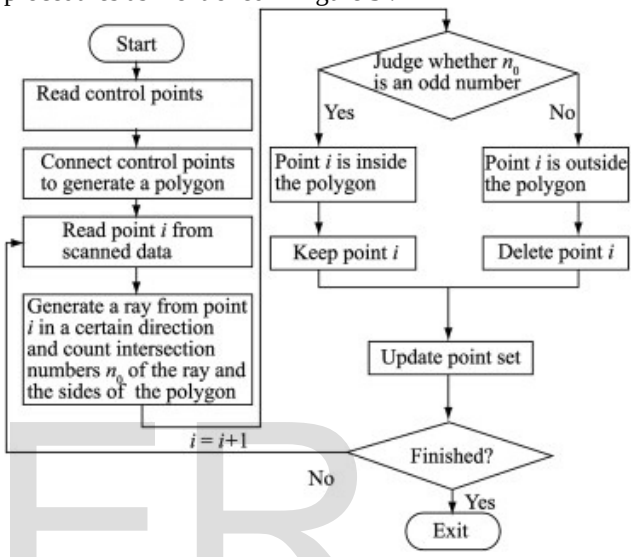

Fig. 5. UML for Ray casting algorithm[16].

We have shown statistics for a number of seismic shakes by using three dimensions (time, energy release of the quake, and the region)

V. RESULT

Data warehouses and OLAP systems help to interactively analyze huge amounts of data. These data, extracted from transactional databases, frequently contains spatial information that is useful for the decision-making process [17]. In this process, the information collected is used to create a data warehouse using SQL server technology. Then, the data warehouse gives us the ability to add data according to defined hierarchies, in order to apply data mining techniques

A. Load Manager Process

The load management process is also known as ETL system which stands for Extraction, Transformation, and Load consists of:

Extraction: Once the data are obtained from the Saudi National Seismological Network (SNSN) operated by the Saudi Geological Survey (SGS), it is analyzed efficiently. As a part of the extraction process, a dataset is generated to be used in the data mart creation, avoiding duplication in order to ensure data integrity. Figure 5 shows that the maximum number of earthquakes occurred in the southern region were in 2014. Figure 6 shows an example of the implemented data set for multiple years.

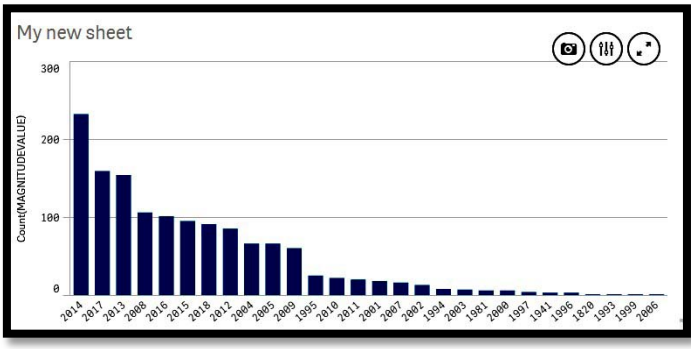

Fig. 6. Annual number of earthquakes in the south region

ORGIN_YEAR

AREANA...	ORGIN_YE...	RANGE	MAGNITUDEV...	Count(MAGNITUDE...	ORGIN_MO...
Totals				**1370**	
south	2018	>1	1.08	1	Jan 2018
south	2018	>1	1.09	1	Jan 2018
south	2018	>1	1.11	1	Jan 2018
south	2018	>1	1.15	1	Jan 2018
south	2018	>1	1.20	1	Jan 2018
south	2018	>1	1.27	1	Jan 2018
south	2018	>1	1.28	1	Jan 2018
south	2018	>1	1.44	1	Jan 2018
south	2018	>1	1.51	1	Jan 2018
south	2018	>1	1.67	1	Jan 2018
south	2018	>1	1.75	1	Jan 2018
south	2018	>1	1.76	1	Jan 2018
south	2018	>1	1.82	1	Jan 2018

Fig. 7. Earhquakes with magnitude larger than 1 and smaller than 2 occurred in the south region in 2018

Figure 7 shows the earthquake activity in the southern part of the kingdom. And figure 8 shows the obtained statisitics for Al-Ayis city, it shows that the maximum activity was in 2009. 2009 acitivty was associated with the volcanic activity in Harrat Lunayyir. Figure 9 shows an example of the implemented data set of Al-Ayis city.

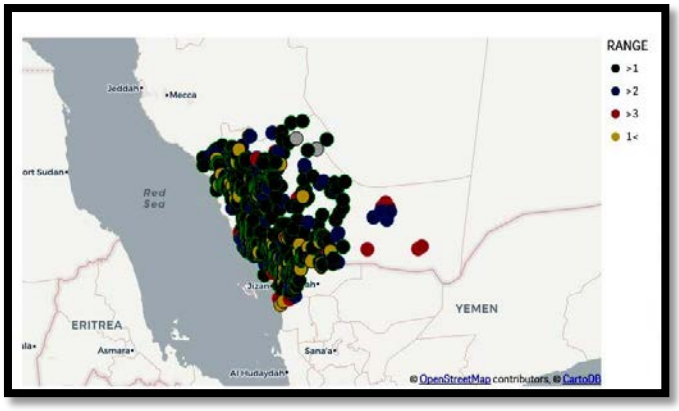

Fig. 8. Seismic activity in the southren part of saudi arabia

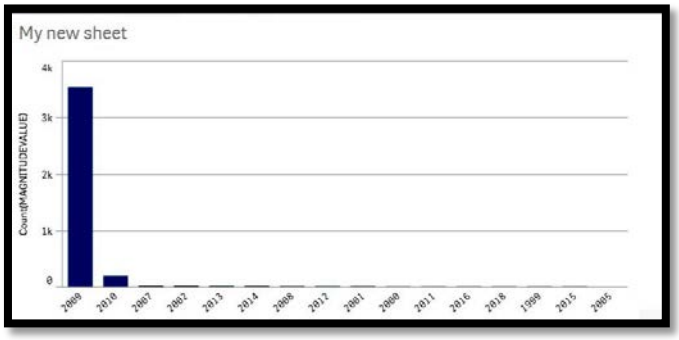

Fig. 9. Annual number of earthquakes in Al-Ayis region

ORGIN_YEAR

AREANA...	ORGIN_YE...	RANGE	MAGNITUDEV...	Count(MAGNITUDE...	ORGIN_MO...
Totals				**3807**	
LNY	2018	>2	2.61	1	May 2018
LNY	2018	>2	2.68	1	Jun 2018
LNY	2018	>2	2.11	1	Jul 2018
LNY	2016	>2	2.05	1	Jan 2016
LNY	2016	>2	2.06	1	Jan 2016
LNY	2016	>2	2.01	1	May 2016
LNY	2015	>2	2.05	1	May 2015
LNY	2015	>2	2.01	1	Aug 2015
LNY	2014	>2	2.38	1	Apr 2014
LNY	2014	>2	2.42	1	Apr 2014
LNY	2014	>2	2.46	1	May 2014
LNY	2014	>2	2.14	1	Aug 2014
LNY	2014	>2	2.45	1	Aug 2014

Fig. 10. Earthquakes with magnitude larger than 2 and smaller than 3 occurred in Al-Ayis region

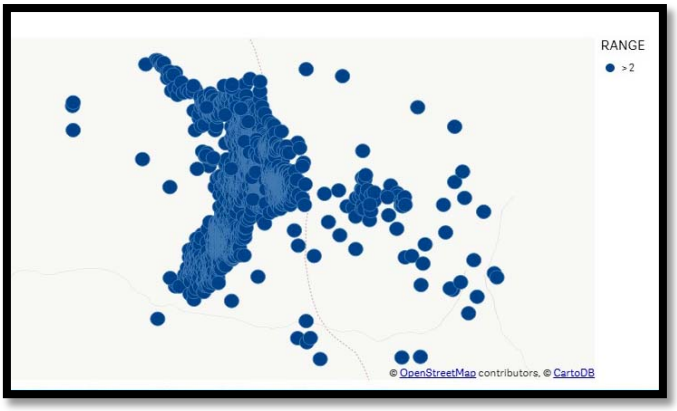

Fig. 11. Seismic activity in Al-ays part of saudi arabia with magnitude larger than 2 and smller than 3

VI. CONCLUSION

In this paper, first, we have created a database in which data recorded the earthquakes observed in Saudi Arabia, in terms of Earthquake data, magnitude, latitude and longitude, the error rate. Next, we have created a second database that included seismic data and we have divided it into regions by using (a polygon algorithm). Then we have used (ETL-extract, transfer, load,) to transferred seismic data from the production database to the second database. after that, we have used OLAP to make earthquake statistics and present them on the map. Finally, We have shown statistics for the number of earthquakes by using three dimensions (time and force of the quake and the region) using the program Quick Sense. and have recorded some tremors in a given date and a certain strength and location of Al-Ais region and the southern part of the Saudi Arabia.

REFERENCES

[1] Rinsurongkawong, V. Chen, C.S., Eick, C. F., Twa, M.,"Analyzing change in spatial data by utilizing polygon , models," in Proc. of International Conference on Computing for Geospatial Research & Application, 2010

[2] Chen, C.S., Rinsurongkawong, V., Eick, C.F., Twa, M., "Change analysis in spatial data by combining contouring algorithms with supervised density functions, in Proc. Of Asia-Pacific Conference on Knowledge Discovery and Data Mining, 2009.

[3] Webopedia , https://www.webopedia.com/TERM/E/ETL.html , last visited : 1/7/2018

[4] https://searchsqlserver.techtarget.com/definition/data-aggregation, last visited : 5/7/2018

[5] http://olap.com , http://olap.com/olap-definition/ , last visited : 6/7/2018

[6]Mohieldin Mahmoud , Alameen Eltoum , Ramadan Faraj, Using Online Analytical Processing (OLAP) in Data Warehousing, access online at : https://www.ijsr.net/archive/v4i6/SUB156082.pdf , last visited : 1/10/2018

[7] Young-Rae Cho,Data Warehouse and OLAP Operations, access at : https://pdfs.semanticscholar.org/presentation/d14a/8022aec27592b74a227270ebe49ab29ca4a4.pdf, Department of Computer Science Baylor University, last visited : 6/7/2018

[8] Britannica, https://www.britannica.com/science/earthquake-geology/Properties-of-seismic-waves , last visited : 10/7/2018

[9] Britannica, https://www.britannica.com/science/seismograph, last visited : 6/7/2018.

[10] Britannica, https://www.britannica.com/technology/accelerometer , last visited : 6/7/2018

[11] geovisualist , https://geovisualist.com/2015/06/22/visualizing-100-years-of-earthquakes/

[12] P. Peggy Li & Herbert Siegel , Visualization of Earthquake Simulation Data, Jet Propulsion Laboratory 4800 Oak Grove Drive Pasadena, CA 91109

[13] B. Priego, H. Pineda, J. Somodevilla, decision support system for seismic risks, Benemérita Universidad Autónoma de Puebla, August 2012

[14] Patras, Greece, A Seismic Data Management and Mining System, Computer Technology Institute

[15] Franklin.R. PNPOLY - Point Inclusion in Polygon Test, 2006, access at : http://www. ecse. rpi. edu/Homepages/wrf/Research/Short_Notes/pnpoly. Html

[16] Hu, C., Zhou, Y.-h., Zhao, C.-j., Pan, Z.-g., Slope excavation quality assessment and excavated volume calculation in hydraulic projects based on laser scanning technology, Water Science and Engineering (2015).

[17] Chen, C.S., Rinsurongkawong, V., Eick, C.F., Twa, M., "Change analysis in spatial data by combining contouring algorithms with supervised density functions, in Proc. Of Asia-Pacific Conference on Knowledge Discovery and Data Mining, 2009.