Designing and Implementation Data Warehouse for Support Iraqi Government

Mervat Mustafa Raouf
Computer science at AL-Mansour University College
Republic of Iraq Ministry of Higher Education and Scientific
Research Iraqi Commission for Computers and Informatics
Informatics Institute
Iraq, Baghdad

Dr. Saad Hameed
Assistant Professor, Computer science at AL-Mansour
University College
Iraqi committee for computer and informatics
Iraq, Baghdad

Abstract:- Recently, Iraq has suffered from disasters, wars and emergencies, In order to ensure adequate security for individuals in the future, the government and decision-makers must find ways todevelop strategies to manage disasters and problems and deal with them and provide the necessary to avoid the largest amount of losses in the events expected to occur again.It is achieved by leveraging the worldwide growth and technological progress and growing the amount of data generated by many organizations with expertise in coping with sensitive situations and issues.In this paper two systems were presented and design:The first is the compilation of data sets from civil society organizations that have always been involved in addressing global crisis,problems previously gathered on the International platform, which is Humanitarian Data Exchange(HDX). As for the second system, it is a data warehousing system designed and implemented to support the Iraqi government by extracting and transferring the databases in the first system relevant to the Iraqi state and carrying out a data clustering process using four types ofclustering algorithms: k-mean, k-medoids, DBSCAN, OPTICS. The results of which will be displayed using a tool containing a map of Iraq showing the location-point of data, other issues will be displayed in the form of reports.

I. INTRODUCTION

Confrontation with disasters and crises in all aspects such as planning, anticipation, and handling depending of the event they ariseit puts governments and their decision-makers on the shoulders of ensuring adequate security for individuals, businesses, and other means. Therefore, they need to develop a detailed disaster management strategy and problems. This is accomplished by leveraging universal evolution and technological progress, and by rescue the quantities of data produced by the many organizations with expertise in dealing with critical situations and issues. Besides, these systems must be designed to challenge, find solutions, and take advantage of the accumulation of preceding data. They also need to develop systems to question, find answers, and take advantage of data backlog. This data, therefore, requires a data warehouse to store and

access, even after a thousand years have passed, to take advantage of the solutions of previous crises and the old data related to them. A data warehouse is a focal repository that collects, cleans, ensures quality, and releases data from different sources and formats only when itis suitable for user consumption via the ETL process (extract, transform, and upload)[1][2]. The data warehouse a very detailed data and data summary level; it is capable of manipulating existing and historic data and generating analytical reports. In this paper we based onRalph Kimball approach. The buildof DW-system is done by SQL Server software [3]. The data sources that chosen is provided to the data warehouse system are data source system that consisting of three database systems.The first one containing data from previous crises and problems, the second one having sources for theIraqi state, and the third one containing different and varied data and the basis for these data is the HDX Humanitarian Platform. Additionally, the clustering process was used within the data warehouse on raster data and displayed on Iraq's map in an illustrated manner.

This paper as illustrated in Figure (1), aims to develop a data warehouse architecture to support the Iraqi digital government.
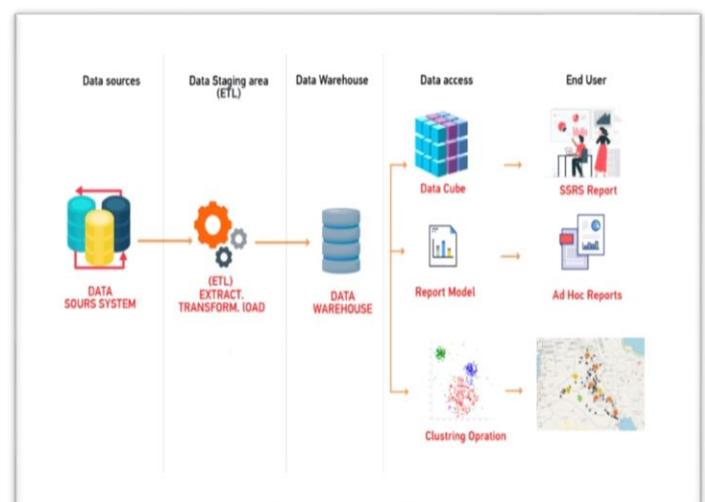

Fig 1:- Structure of DW System

II. DESIGN THE DATA WAREHOUSE ETL PROCEDURE

This procedure was designed to utilize the Visual Studio-BI-IS (Business Intelligence - Integration Service) software [3] [4]. The ETL procedure was designed from the first source system (the previous problem database), the second source (the resource database for the Iraqi government), and the third source (the diversified database). Figure (2) refers to a single database ETL process. These rules were selected and established because they contain numerous useful data to support the Iraqi government during crises, including the Coronavirus (COVID19)crisis, post-ISIS terroristoccupation, and the effects of violence that occur in several regions within Iraq. Besides that, these database systems illustrate many of the data collected by humanitarian organizations and NGOs that support humanity.
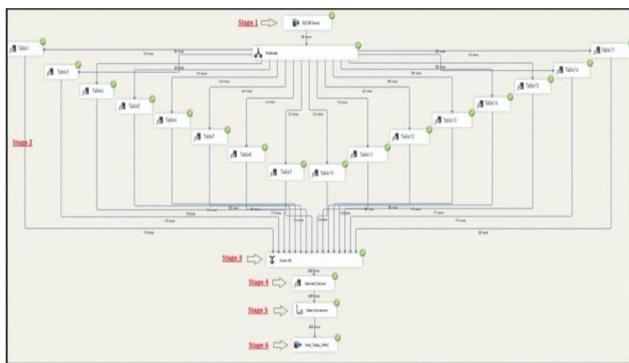

Fig (2):- Single Database ETL Process

III. DIMENSIONAL DATA MODE (DDM)

A dimensional model (DDM) [5][6] is developed in the initial phase of designing the data warehouse. DDM is the most logically styling technicality that attempts to exhibiting the data model in an attainable and axiomatic standard frame to prepare a basis for a developed analysis of the data. It should be intended to confront the requirements of big organisms, which must cure the requirements of decision-makers and contain easily accessible info. [7]. In every dDDM, a fact table is the prim table. It includes business mensuration and more one foreign keysrelated to dimensional tables. Adimension is one or many tables which relate to the table offacts. Every dimension contains adjectival textualinfo., and it has a primary key, which be of service to the fundamentals for referential completeness with the fact table to which it is attached. Most dimension tables implicate some attributes containing data collected from various source [5].

IV. STAR SCHEMA OF DW SYSTEM

As shown in Figure (3) Star is the simplest type of schema. which contains a table of facts, rest in the middle, and is linked with other tables of dimensions like a star. Each of tables has the primary key in the fact table, which is related to a foreign key. The reasons of star schema that makingadopted: lower complex, facilities to understand, used most widely for DWandmore effective scheme for

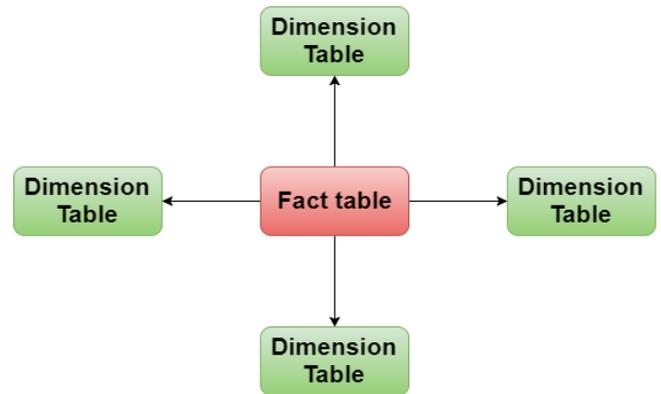handling simpler queries, quick aggregations, and feeding cubes (efficient OLAP cubes)[8][3].

Fig (4):- Star Schema

V. ACCESS DATA IN OLAP METHOD

A data warehouse (DW) is a data placed with a designing that facilitates data analysis, and OLAP is a method for analyzing multidimensional data, in addition to providing smart services, financial services, financial services, intelligent solutions, self-solutions, insights. Fast, consistent, and interactive data are accessing a range of different perspectives. "Multidimensional analysis" is the primary platform for OLAP. That is, the capability to analyze metrics in various dimensions, such as timing, genus, nationalism, etc [10][11]. OLAP can create multi-dimensional balances from OLAP cubes. OLAP cubes are data structures that solve a data warehouse problem. Generate an OLAP out of three fundamental analytical processes: roll up,dig away and slice and dice cut.

This permit the data to be rolled up, sliced, and diced as prerequisite to handle the enormous variety of issues relevant to a user's area of service [5][12].The cube can be stored in a multidimensional architecture of the MOLAP (OLAP), a ROLAP (OLAP Relational) database, or a multifaceted structure and relational database cluster of the HOLAP (Hybrid OLAP). Visual Studio (Business Intelligence - Analysis Service) has been certified for retailing specific OLAP cubes that meet the decision-makers' inquiries[12].

VI. CLUSTERING OPERATION

One of the vital means in dealing with this data is to collect it into a group of categories or groups where the data clustering technique is in which logically similar information is stored together. Here lies the importance of trying to revise and classify this vast amount of data and compile it so that it performs the essential purpose of extracting hidden information or categorizing data according to its relationships with each other to benefit from it for technical purposes. Clustering involves creating groups of similar and dissimilar matter. The problem with clustering is to find set of similar matter in the data, and the similarity between matters is measured using the similarity function. Cluster is especially useful to improve retrieval and support browsing and that done by increase the efficiency in the database system from

reduce the disk access times. In clustering, matters with similar properties are collected in one class, and single disk access makes the entire class available. Clustering algorithms can be used in many fields, such as marketing project, biology, libraries, insurance system, city planning project, earthquakes, and others. A cluster is a procedure that is often without a supervisor and is done by extracting groups that share a particular thing or are similar to a particular thing of their own so that each cluster is different from the other cluster[9]. In this paper, four clustering algorithms(k-means, k-medoids, DBSCAN, OPTICS) are used due to the difference in data types and the corresponding clustering types. Four types of algorithms have been used to cluster the data according to the cluster's appropriate methods and this is due to the difference and diversity of the repository data.

VII. USER INTERFACE

This is the latest phase in the operation of constructing theIraqi Data warehouse system; interfaces were designed through this phase, allowing decision-makers to reach the required reports easily and also display the geodata on the Iraqi map to show the position of data on the real world. The interface pages are designed using the software programming "ASP.NET 4.5, C #, HTML, CSS" languages. Figure(5) shows the DW system's main interface and login. This page is designed to enable decision-makers to get on the reports page and delivering the data before and after clustering. Using SQL Server Report Services [3]. Figure (6) show the display of data, Figure (7) refers to a report conceived.
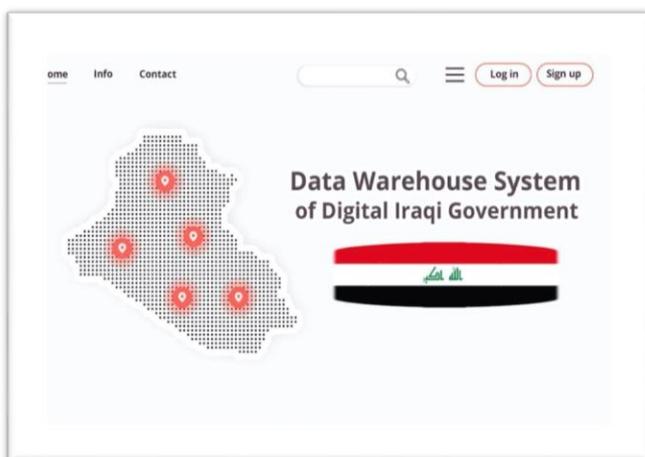

Fig (5):- The Login page of DW System

VIII. CONCLUSION

This paper proposes the design and implementation of the DW system to support the digital Iraqi government. It gathers and stores data from various sources in one repository and solves the issues facing decision-makers in the Iraqi State. Advanced research tools are also used, generated from periodic information, and showing geo point datainterfaces. The future work of this project is to implement the DW system with the use of data mining [13][14][15].
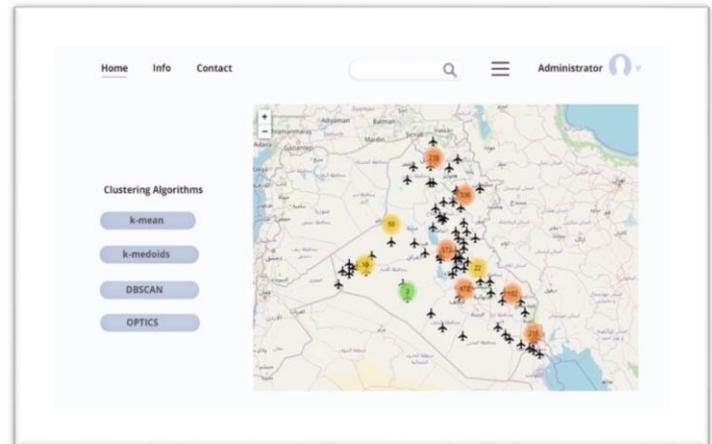

Fig (6):- The Display of Data on Map

حالات وفيات المدنيين الموثقة الناجمة عن العنف

Fig (7):- The Repot page

REFERENCES

[1]. LekhaNarra, Tony Sahama, Peta Stapleton, "Clinical Data Warehousing A Business Analytics approach for managing health data", Proceedings of the 8th Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2015), Pages 101 – 104, Sydney, Australia, January 2015.

[2]. Ralph Kimball, Margy Ross, "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling", Third Edition, John Wiley & Sons, Inc., Indiana, United States of America, 2013.

[3]. R. Mistry, W. Misner, "Introducing Microsoft SQL Server 2012", First Edition, Microsoft Press A Division of Microsoft Corporation, United States of America, 2012.

[4]. Warren Thornthwaite, "Implementing a Microsoft SQL Server Parallel Data Warehouse Using the Kimball Approach", Copyright Microsoft Corporation, United States of America, June 2011.

[5]. Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy, Bob Becker, "The Data Warehouse Lifecycle Toolkit: Practical Techniques for Building Data Warehouse and Business Intelligence Systems", Second Edition John Wiley & Sons, Inc., 2008.

[6]. K. Laudon, J. Laudon, "Management Information Systems: Manage the Digital Firm", Twelfth Edition, Pearson Education, Inc., United States of America, 2012.

[7]. MS. ALPA R. PATEL; PROF. (DR.) JAYESH M. PATEL, "DATA MODELING TECHNIQUES FOR DATA WAREHOUSE", International Journal of Multidisciplinary Research, Vol.2, Issue 2, Pages 240-246, February 2012.

[8]. Rajib Dutta, "Healthcare Data Warehouse System Architecture for Influenza (FLU) Diseases", Computer Science & Information Technology Journal, Vol. 3, No. 2, Pages 77-89, March 2013.

[9]. R. Xu and D. Wunsch, "Survey of Clustering Algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, May 2005.[Online].Available: http://dx.doi.org/10.1109/TNN.2005.845141

[10]. Dr. WalidQassimQwaider, "Apply On-Line Analytical Processing (OLAP) With Data Mining for Clinical Decision Support", International Journal of Managing Information Technology (IJMIT) Vol.4, No.1, Pages 25 - 37, February 2012.

[11]. N.Colossi, W.Malloy, B.Reinwald, "relational extensions for OLAP", IBM systems journal, Vol. 41, No. 4, Pages 714-731, 2002.

[12]. PaulrajPonniah, "Data Warehousing Fundamentals", John Wiley & Sons, Inc. 2001.

[13]. Sami Hasan, "Rapidly-Fabricated Architectures of Parallel Multidimension Algorithms", Lambert academic publishing, 2017.

[14]. Sami Hasan, "Performance-Aware Architectures for Parallel 4D Color fMRI Filtering Algorithm: A Complete Performance Indices Package," in IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 7, pp. 2116- 2129, July 1 2016.

[15]. Sami K. Hasan, "FPGA Implementations for Parallel Multidimensional Filtering Algorithms", PhD thesis, Agriculture and Engineering Newcastle University, 2013.