ISSN 2286-4822
www.euacademic.org

EUROPEAN ACADEMIC RESEARCH
Vol. III, Issue 6/ September 2015

Impact Factor: 3.4546 (UIF)
DRJI Value: 5.9 (B+)

The Role of Data Warehouse during the Natural Disaster

ZIJADIN KRASNIQI
PhD. Candidate
Informatics, Mathematics, and Statistics
European University of Tirana
Albania

Abstract:

Data integration is the issue of data combination, which is found in different sources, as a result of the advances in information technology there are also Data Warehouses to provide users a unified view of these data, the issue of building systems for data integration is important in nowadays applications. Data integration becomes increasingly important in the cases of merging of systems with many agencies or the consolidation of applications within an agency in order to provide a unified view of assets data for the purpose of emergencies management. Taking into consideration the advances in Information Technology, the emergencies management requires modeling data from different sources, by coordinating activities between institutions and federal agencies. .

Today, more than ever, there is a great need to build an institutional system to coordinate all available resources and capabilities in order to be prepared to react in a coordinated and planned fashion in cases of natural disasters and other incidents. During the last years, the advances in a diversified technology of information and communication has proposed new alternatives to manage natural disasters in national and regional level. The importance of data analysis has increased significantly in recent years as organizations in all sectors are being required to improve their decision making processes in order to maintain their competitive advantage.

Key words: Data warehouse, Data integration, Data analysis, natural disasters, emergencies management.

Introduction

Traditional database systems do not satisfy the requirements of data analysis. They support the daily operations of an organization and their primary concern is to ensure fast access to data in the presence of multiple users, which necessitates transaction processing, concurrency control, and recovery techniques that guarantee data consistency. These systems are known as operational databases or online transaction processing (OLTP) systems. Typical operational databases contain detailed data, do not include historical data, and since they are usually highly normalized, they perform poorly when executing complex queries that need to join many relational tables together or to aggregate large volumes of data. Further, when users wish to analyze the behavior of an organization as a whole, data from various operational systems must be integrated. This can be a difficult task to accomplish because of the differences in data definition and content.

Data warehouses were proposed in order to better respond to the growing demands of decision-making users. A data warehouse is a collection of subject-oriented, integrated, nonvolatile, and time-varying data to support management decisions. (Malinowski 2008**)**

Concepts of Data Warehousing

The term data warehouse was coined by W. H. Inmon: "A warehouse is a subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management's decision making process".

In contrast to traditional On-Line Transaction Processing (OLTP) applications, decision support places some different requirements on database technology. Data warehouses provide storage, functionality, and responsiveness to queries that are beyond the capacity of OLTP databases. They contain historical, summarized, and consolidated data over potentially long periods of time. Their size can be from hundreds of gigabytes to terabytes. There is a great need to provide decision-makers at all levels of management with information at the desired level of detail, to support their decision-making. Apart from performing regular, predefined reporting activities, a number of parallel users are submitting ad hoc, complex queries. These queries require access to millions of records and cause numerous scan, join, and aggregate operations across the warehouse. As a result, query throughput and response times are main issues in multi-user decision-support systems. (Stolba 2007)

Data Warehouse Development Methodology

Bill Inmon and Ralph Kimball are two primary methodology leaders within the DWH field who have triggered the remarkable growth of DWH development in the last decade. Over the last ten to fifteen years, their methodologies have established a solid core of successfully deployed data warehouses.

According to Inmon, the data warehouse is constructed in an iterative fashion, from the enterprise data warehouse to the departmental databases, adopting a traditional relational database tool to serve most decision support needs Inmon stresses three basic premises for designing a data warehouse. The first is data normalization and metadata definition for both data and process modelling as typical in database management system development. The second is support to end users' decisions through arbitrary queries to the database. The third is data cleaning and unification in the data preparation stage (data staging) for transforming the operational data into data for 'informational' purposes. The development of the target DW includes a number of tasks beginning with a size/granularity analysis, which includes three levels of granularity: atomic, departmental, and individual DW. Then, the first subject area is selected, developed, and populated. An iterative approach is used to develop each successive departmental database utilizing codes and processes implemented for the first departmental DB to reduce development effort and time.

On the other hand, according to Kimball, a requirement-driven method is used; this begins with project planning for assessing readiness, scoping, and justification. Kimball's approach is a bottom-up one, which places the business requirement analysis on the top of its development phases. Such a requirement analysis triggers the life cycle and forms a foundation of three parallel tracks, which deal with technology, data, and analytic application. Kimball's architecture forms the enterprise data warehouse by the bus architecture that collects data marts based on conformed dimensions and facts. Each data mart represents a business process in the organization by means of dimensional modelling (e.g. star schema). Designing dimensional models involves a four-step process:

- business process selection starting with the process with the most impact,
- grain declaration which serves to decide what level of detail the data warehouse will contain,
- dimensions selection, and
- Facts identification.

A physical design is developed next, giving attention to aggregation and indexing strategies. For functionality purposes, a consistent analytic framework that satisfies the user's analytic requirements is established. Please refer to table which simplify the main differences between the Inmon and Kimball approaches.

Table: Inmon vs. Kimball approach

	Inmon	**Kimball**
Approach name	Waterfall, ClF, Hub and Spoke	Iterative Data Mart Bus Architecture with Linked Dimensions
Business requirements	Top-down Data-driven	Bottom-up Requirements-driven
DW model	ER model	Multidimensional model
Development approach	Iterative approach used to develop departmental DB from enterprise DW	collects data marts based on conformed dimensions and fact to form enterprise DW
Primary audience	IT professionals	End users

Meanwhile, the continuous growth in DWH adoptions has increased the awareness of researchers and developers to propose and develop new methodologies that compromise e and trade-off between Inmon and Kimball or even adopt new techniques and/or architectures. (Arwa 2011)

Source of Main OCEMA Data

Operative Center of Emergencies Management Agency (OCEMA) disposes a great data volume. It can be one of the sectors, which saves more information than any other type of industry, taking into consideration the functions that performs and possesses. (Juliana 2014) So, such an Operative Center has many information resources, but there are a few categories which we will discuss in the following:

Data Analysis in the Hydro-Meteorological Institution - The Case of Floods

The actual status of the hydrometric network in Kosovo is not enough to provide a full and reliable prediction system for floods (and droughts). Only 50% of monitoring stations are operative. There should be made some investments in the network and supportive personnel to improve its reliability and usefulness, especially regarding the capability to broadcast and read data in real time. Based on the gathering of hydro-meteorological data, it can be useful only if:

- Data is gathered regularly
- Data is reliable and verified
- Data processing is done in the useful informational form
- Information is published in real time from meteorological stations

If the hydrometric network if fully or only partially automated (so that the rainfall amount or the river levels can be reported to the center in real time, this creates a very powerful information system and warning system which can be used to monitor the potential conditions for floods. We may conclude mentioning that regularly providing hydro-meteorological data in an easy approachable manner and easily understandable format is an important component of good preparedness in case of floods. A critical element within the floods management is the continuous observation and the progress assessment of a flood disaster. Actually, in order to find out the starting of a flood, there must be continuously monitored special and important variables of waters and ponds levels, not only when floods are expected. The rainfalls monitoring will be useful in the short-term prediction of floods only if the registered data will go to the main control center in "real time". If we take into consideration the meteorological stations data in Peja and Ferizaj in 2012 - the average monthly values of meteorological elements and phenomena are represented in the table below

Table: Data from meteorological stations in Peja, Prishtina and Ferizaj

2012	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	m.years.
Tmax .	2.7	0.2	14.1	17.2	21.1	28.9	32.1	32.6	22.6	18.1	16	6.4	17.7
Tmin.	-4.9	-7.5	1.3	5.1	9.8	14.2	17.5	16.3	11.8	7.9	5.3	0.6	6.4
Tavg.	-1.5	4.1	7.6	11.8	16.3	23.1	25.3	26.0	16.8	11.6	8.5	4.1	12.8
Humidity %	89	86	83	79	73	72	59	53	72.1	86	83	87.3	76.8
Atm.pre	943.6	948.3	940.2	946.3	951.2	956.8	961.7	963.1	958.3	955.4	952.8	953.2	952.5
Wind/ms.	1.2	1.8	1.6	1.9	1.3	0.8	1.2	0.9	0.8	1.4	1.3	1.1	1.3
Rainfall	121.6	41	13.2	58.4	96.3	5.8	40.6	0.3	24.7	34.5	65.4	70.3	572.1

Rainfall, snowfall levels in high mountains, river flows, underground water levels and levels of accumulative potentials of lakes are all included in this data. All this variables must be continuously monitored and reported from the center to OCEMA which is responsible for the national monitoring of floods. (Brian 2011) The quality and reliability of the data are very important. Incorrect or incomplete rainfall data are more harmful than the absence of data. Nothing can be predicted, planned or managed without the data. From the Kosovo Hydro-Meteorological Institution until now we do not have a database which we can refer to in real time and have access to data updates from the Hydro-Meteorological Institution.

Analysis from Kosovo Seismologic Institution - KSI

Most of the data from Kosovo Seismic Institution - KSI, are archived in forms of reports, tables, drafts and maps. Their utilization is done with traditional methods. Means of communication are manual, there is no continuous and updated flow gathering, assessment and administration. It is impossible for users to provide new and reliable information in real time about the issues in the seismic field where they are interested in. We have mainly monthly data called "Monthly Newsfeed" from the Seismic Institution, which are represented in the table below.

Table: Kosovo Seismic Institution - KSI [Google 2014]

Date	Time	Place	M	Lat	Long	Depth
10.03.2010	13:38:03.1 13:44:41.8 2:13:20 PM.9 3:32:43 PM.2	Kosovo Kosovo Kosovo Kosovo	5.2 1.8 1.9 2.2	42.76 42.75 42.69 42.65	20.62 20.63 20.58 20.95	5.22km 5.00km 18.83km Unknown
11.03.2010	4:33:54 AM.9 8:59:19 AM.4	Kosovo	1.5 3.3	42.73 42.74	20.58 20.60	16.06km Unknown
12.03.2010	2:28:16 PM.1	Kosovo	1.3	42.47	20.67	Unknown
13.03.2010	1:54:00 AM.3 1:53:44 PM.5	Kosovo	1.8 1.5	42.72 42.76	20.57 20.61	6.28 km 15.28km
15.03.2010	9:37:40 PM.5	Kosovo	2.1	42.44	20.90	14.04km
18.03.2010	2:48:10 AM.9	Kosovo	1.6	42.71	20.60	9.99 km
26.03.2010	7:46:10 AM.8 7:46:33 AM.8	Kosovo	1.8 1.8	42.71 42.71	20.58 20.57	17.18km 17.10km

These data reports from KSI are made in monthly basis, and actually the need for information is necessary to be in real time. At the same time, we must say that the territory of Kosovo is affected many times with heavy consequences by earthquakes in neighboring countries, such as earthquakes in Albania, Montenegro, Macedonia, etc.

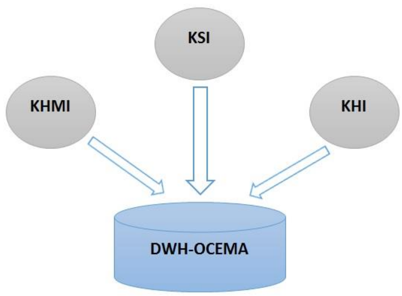
Analysis of Data from Kosovo Health Institution - KHI

The health informative system is mainly manual, on paper and most patient's data and treatments are recorded in paper forms which are gathered in specific institutions and then it is entered in electronic format from operators in hospitals and main centers of family health care. Personnel patients data (the so-called "midnight statistics"), are mainly entered in Excel. However, since 2002, data for activities, possibility of sicknesses and deaths in the primary, secondary and tertiary care are increasingly being entered in the main database "Access-Master Database" Until now, the Access-Master Database is implemented in all QKMFs, in 7 hospitals (5 regional hospitals and 2 city hospitals) and in QKUK. However, data is reported with six months or more delay, the analysis of the same data give different results and not all health institutions are reporting. [Google 2010]

Data Warehouse System

This part of the study represents the theoretical side of building the Data Warehouse which is very welcomed for the decision taking processes in business, telecommunication and health care fields. But it also has a great impact in natural disasters management, and this actually needs real time data. Data Warehouse systems, from the developing point of view, are very evolutionary and dynamic, they never stop transforming following the requirements. The Data Warehouse system is a system that gathers information from different heterogeneous sources, as in our case data from: **Kosovo Hydro-Meteorological Institution**, **Kosovo Seismic Institution and Kosovo Health Institution.** (Kimball 2008) What we mentioned above are considered internal resources, since OCEMA has also external resources, such as data from national security services. The Fig. below shows resources used to build DWH, which helps later the analysis and reporting of decision making process in OCEMA.

Fig: Diagram of passing information to DWH-OCEMA



Based on these systems, OCEMA is interested in the gathering and integration of information, this makes it necessary for a data integration generated from the platforms of KSI, KHI, KHMI.Integration of data is the process of combining data from different sources and providing the user a unified view about this data. This process is submerged in a series of business situations (when two similar companies must merge their databases) and scientific (combination of different seismologic, medical, meteorological, fire and emergency researches).The problem of data integration becomes more critical while the amount of data that must be shared is increased. By the term 'data investigation' we mean the process of inserting in a data system, which come from multiple heterogeneous sources. Escalation is an important issue in data assimilation since information flow might be very high during a critical situation. The area of data integration or information integration has made a great progress in theoretical foundations and in development of algorithms and means.

Data integration is the issue of data combination, which are found in different sources; as a result of advancements in information technology there is also Data Warehouse to provide users a unified view of these data, the problem of building systems for data integration is very important in nowadays applications.

By taking into consideration the advancements in Information Technology, the management of emergencies requires modeling data from different sources by coordinating activities between institutions and governmental agencies.

Today it exists more than ever the need to build an institutional system to coordinate all available resources and capabilities to prepare to react in a coordinated and planned fashion in situations of natural disasters and other incidents. During the last few years, an advancement of a diversified information technology and communication has proposed new alternatives to manage natural disasters in national and regional level. Information technology has become a basis for the development of many information systems today. In most of life areas it has become very difficult to maintain communications, coordination or international communications without the broad usage of information technology.(Krasniqi 2015)

Conclusions

Nations are not safe from natural and man-made disasters despite the tremendous efforts put to mitigate disasters. The traditional ways are not sufficient enough to handle disasters. Therefore, new ways must be explored in order to mitigate disasters. On the basis of these facts this article has presented a comprehensive overview of many issues, problems, challenges and threats reshaping disaster management. In order to address these issues, Firstly, DWH technology eliminates data redundancy and inconsistency by achieving the paperless sharing of data throughout an organization and increasing the confidence in data quality. Secondly, a DWH improves the efficiency of operational tasks as it has the promise of enabling organizations to generate cost-effective reports and ad-hoc analyses for improved decision making. Thirdly, a DWH better understands and meets the needs of customers, maximizes campaign results, builds loyalty and, ultimately, improves bottom-line results.

REFERENCES

Arwa Abdullah J (2011): Transitioning a Clinical Unit to a DataWarehouse,

Brian Faulkner (2011): Korniza Kosovare për menaxhim të rrezikut nga përmbytjet,

Kosovo Hydro-Meteorological Institution-KHMI, Monthly bulletins 2014

Kosovo Seismic Institution – KSI Monthly bulletins 2014

Juliana L (2014): The study of data warehouse systems and the construction of a reporting model based on business intelligence technology, Tirana, Albania

Malinowski E, Zim´anyi E (2008):Advanced DataWarehouse Design, Springer

Stolba N (2007): Towards a Sustainable Data Warehouse Approach for Evidence-Based Healthcare ,Wien.

Ralph Kimball (2008) : The Data Warehouse Lifecycle Toolkit, Second Edition

Krasniqi Z (2014):Application of information technology in emergency and natural disaster management”2014, Albania

Google.2010.“Strategy for health information system in Kosovo 2010 – 2020” http://www.kryeministriks.net/repository/docs/STRATEGJIA__PER__SISTEMIN_E_INFORMIMIT_SHENDETESOR_NE_KOSOVE_2010_-_2020.pdf