

A Data Warehouse Design for Dangerous Pathogen Monitoring

Amadou DICKO ^{a,1}, Seydou Golo BARRO ^{b,d}, Yaya TRAORE^c and Pascal STACCINI^d

^a National Institute of Public Health, Burkina Faso

^b Nazi BONI University, Burkina Faso

^c Joseph Ki-ZERBO University, Burkina Faso

^d RETINES Lab - IRIS department - Medical Faculty of Nice University, France

Abstract. Emerging diseases are a major public health problem as illustrated by the current coronavirus disease (COVID-19) pandemic. To make the right decisions, public health departments need a decision-making system. In Africa few IT systems have been put in place to help managers of public health in the analysis of their multidisciplinary data. The majority of digital health solutions are operational databases, as well, focused on surveillance activities that do not include the laboratory component. This paper describes the design model and implementation of data warehouse for dangerous pathogen monitoring in a laboratories network. Talend data integration is used to extract data in Excel sheets, transform it and load it into a MySQL database.

Keywords. data warehouse, laboratory, pathogen, Talend

Introduction

Emerging diseases are a major public health problem as illustrated by the current coronavirus disease (COVID-19) pandemic which has caused a major health crisis and quarantined half of planet population [1]. Facing the emergence of new pathogens, the rapid spread of epidemics and bacterial resistance, monitoring and surveillance of diseases are becoming particularly important and indispensable [2]. In Burkina Faso, public health managers are subject to a need for monitoring and prevention of emerging infectious diseases as they have become increasingly worrying, given the increase in their number of incidences in the country [3].

The National public health institute aims to ensure health monitoring, to carry out public health study and research. It periodically publishes reports and statistics relating to the phenomenon of epidemic spread affecting different age categories of the population and different geographic areas. To date, few IT systems have been put in place to help managers of public health in the analysis of their multi-disciplinary data. Burkina Faso has an epidemiological surveillance system based on the national network of laboratories located at three ministries (Health, Agriculture and Scientific Research). The activities of detection, diagnosis, scientific research and storage of samples of potentially dangerous

¹ Corresponding Author, Amadou Dicko, National Institute of Public Health, Burkina Faso; E-mail: dicko22@gmail.com

strains carried out within them, are susceptible to risks of exposure to microorganisms as well for laboratories workers, the community and the environment. These risks can be accidental or deliberate and require the implementation of biosecurity and biosecurity measures. Therefore, each laboratory should implement good practices in terms of biosafety and biosecurity which are key requirements of the International health regulation [4]. All of these good practices must be converted in order to follow the traceability. Setting in place of an inventory system for dangerous pathogens in Burkina Faso responds also a strong recommendation from the joint external evaluation World health organisation conducted in December 2017. In fact, this evaluation revealed major gaps in terms of the management of pathogens and showed that the efforts made were largely insufficient. By consequent, type, risk level, volume, storage security level of pathogens at the national level are not known. Also, within the laboratories, pathogen traceability is not ensured. All of these elements constitute a security breach, do not allow real time knowledge of stored pathogens and endanger the health of the entire population.

The general objective of this work is to improve the management of pathogens in medical biology laboratories of Burkina Faso by setting up a data warehouse.

|Background

A data warehouse is different from a regular database which is used to run the business on a current basis and support everyday transactions and processes. Data warehouse is defined as an subject-oriented, integrated, time-variant, and non-volatile collection of data primarily used in strategic decision making by means of on- line analytical processing (OLAP) techniques [5]. The construction of the data warehouse through the ETL processes and an interactive dashboard according to the key performance indicators defined by the business decision-makers [6].

Currently, data relating to pathogens stored in the country laboratories are collected in excel sheets. About twenty laboratories have this file that they fill out and send to the central laboratory on a monthly basis. This process does not allow to do a consolidated analysis of activity at the national level. The current system is not designed with a view to supporting decision-making. It is therefore important to rethink a more automated, simpler and more practical solution by implementing an application that facilitates data integration and produces more quickly dashboards.

There are several steps in setting up a data warehouse. The first consists in analyzing data sources and identifying user needs. The second is the organization of data within the warehouse. Finally, the third consists in establishing various interrogation tools (of analysis, of data or query) [7]. This article covers the first two steps.

|Business model

Figure 1 shows the business model used for the implementation of the solution. The flat files are received monthly from laboratories and put in a container. Jobs are launched in the Talend tool in order to extract, transform and load the data into MySQL database.

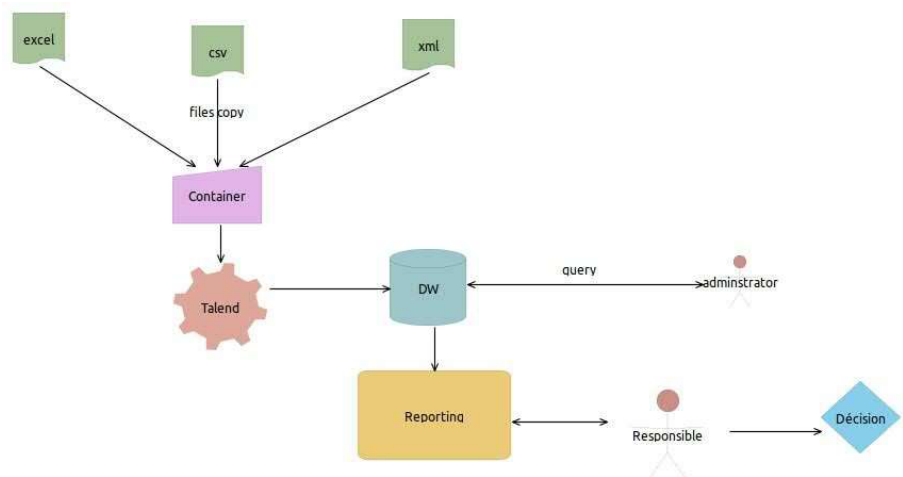


Figure 1. Business model

Conceptual schema

The proposed data warehouse is a star schema type design with one center fact table and three dimensions [8]. Figure 2 reveals the basic star schema conceptual diagram of the proposed data warehouse.

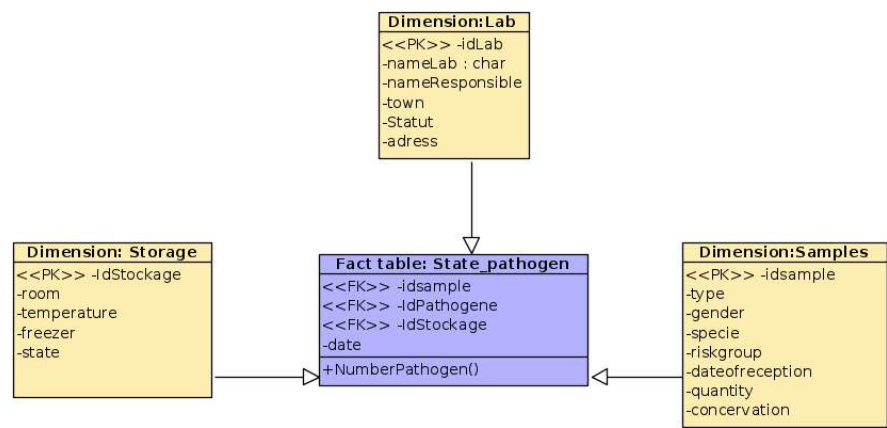


Figure 2. Star Schema of the data warehouse

Implementation: Loading and Indexing

The proposed data warehouse design is implemented under MySQL. It contains four refined distinct relations interrelated together. Figure 3 shows the logical diagram of the data warehouse implemented under MySQL. Now, as all data were successfully cleaned and transformed, they are exported from their buffer zone in which they were temporary

stored to the new informational database or data warehouse just built.

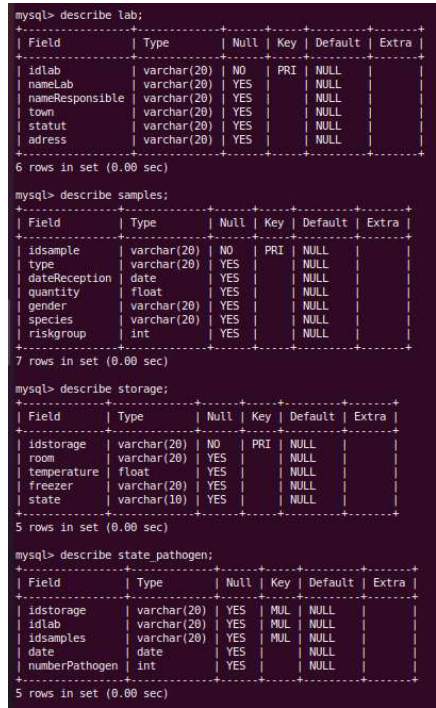


Figure 3. Data Warehouse Logical Diagram under MySQL

Conclusion

This paper introduced a model for building a data warehouse for high pathogen management in a country. The activity measured is the number of dangerous pathogens stored as well as their variation in order to make decisions in the public health area. The data warehouse is an informational database whose data are extracted from an already existing Excel files. The purpose of the proposed design is to help decision makers in performing data mining and data analysis over the data stored in the warehouse which eventually helps them in discovering critical patterns and trends.

References

[1] El-Hage W, Hingray C, Lemogne C, Yrondi A, Brunault P, Bienvenu T, Etain B, Paquet C, Gohier B, Bennabi D, Birmes P, Sauvaget A, Fakra E, Prieto N, Bulteau S, Vidailhet P, Camus V, Leboyer M, Krebs MO, Aouizerate B. Les professionnels de santé face à la pandémie de la maladie à coronavirus (COVID-19): quels risques pour leur santé mentale ? [Health professionals facing the coronavirus disease 2019 (COVID-19) pandemic: What are the mental health risks?]. Encephale. 2020 Jun;46(3S):S73-S80.

[2] Lefever N, Pouillet Y. Entrepôts de données et vie privée [Data warehouses and privacy]. Revue du droit des technologies et de l'information. 2008;30:7-20.

[3] Tinto B, Salinas S, Dicko A, Kagone TS, Traore I, de Rekeneire N, Bicaba BW, Hien H, Meda N, van de Perre P, Kania D, Simonin Y. Spreading of SARS-CoV-2 in West Africa and assessment of risk factors. Epidemiol Infect. 2020 Sep 14;148:e213.

- [4] World Health Organization. International Health Regulations. Areas of work for implementation (2005). June, 2007. p. 28. Available at: http://apps.who.int/iris/bitstream/10665/69770/1/WHO_CDS_EPR_IHR_2007.1_eng.pdf?ua=1.
- [5] Vaisman A, Zimányi E, Vaisman A, Zimányi E. Conceptual Data Warehouse Design, Data Warehouse Systems, 2014, Springer p. 89-119.
- [6] Mercado KE, Perez CB, Lopez-Arredondo LP, Caro K, Castro LA, and Rodriguez FL., Agile Dimensional Model for a Data Warehouse Implementation in a Software Developer Company, Research in Computing Science;147(3):27–34.
- [7] Prater SL, Davis CL. A perinatal intervention program for urban american Indians, part 2: the story of a program and its implications for practice. J Perinat Educ. 2002 Spring;11(2):23-32.
- [8] Bréant C, Thurler G, Borst F, Geissbuhler A. Design of a Multi Dimensional Database for the Archimed DataWarehouse. Stud Health Technol Inform. 2005;116:169-74.