Cross-Border Challenges in Informatics with a Focus on Disease Surveillance and Utilising Big-Data 65
L. Stoicu-Tivadar et al. (Eds.)

doi:10.3233/978-1-61499-389-6-65

Supporting Diagnosis and Treatment in Medical Care Based on Big Data Processing

Oana-Sorina LUPŞE [a,1], Mihaela CRIŞAN-VIDA [a], Lăcrămioara STOICU-TIVADAR [a], and Elena BERNARD[b]
[a] *Department of Automation and Applied Informatics/University Politehnica Timişoara, Romania*
[b] *Department of Obstetrics and Gynecology/University of Medicine and Pharmacy "Victor Babeş", Timişoara, Romania*

Abstract With information and data in all domains growing every day, it is difficult to manage and extract useful knowledge for specific situations. This paper presents an integrated system architecture to support the activity in the Ob-Gin departments with further developments in using new technology to manage Big Data processing - using Google BigQuery - in the medical domain. The data collected and processed with Google BigQuery results from different sources: two Obstetrics & Gynaecology Departments, the TreatSuggest application - an application for suggesting treatments, and a home foetal surveillance system. Data is uploaded in Google BigQuery from Bega Hospital Timişoara, Romania. The analysed data is useful for the medical staff, researchers and statisticians from public health domain. The current work describes the technological architecture and its processing possibilities that in the future will be proved based on quality criteria to lead to a better decision process in diagnosis and public health.

Keywords. Healthcare Big Data, Cloud Computing, Google BigQuery, Treatments, HL7 CDA.

Introduction

We live in an era when people collect a lot of data and try to deal with it and use in different situations with maximum benefits. This led to a new term used when a lot of information is involved, Big Data. Big Data is a collection of large and complex data and information which is difficult to process with common database management tools. Big Data involve tools, processes and procedures that create, manipulate and manage very large data, and information in storage facilities. Four "V" dimensions describe the Big Data concept: volume, velocity, variability and veracity [1].

The healthcare domain is more and more based on the electronic medical record to store the patient's data and information about treatments and physicians. Using the electronic data makes it easier to access medical data, but the information grows bigger and bigger making difficult to draw useful conclusions from it without proper tools.

[1]Corresponding Author.

Day by day, the database of medical information is growing and the medical staff and big hospitals have difficulties to manage it. For this reason, the use of Big Data is very important in this domain. New technology will support the medical staff to manage information easier. Based on the suggested solution, which is a product of current work, information is inferred from complex heterogeneous patient sources. Correlations to the electronic health records of the patients are possible, large volumes of medical imaging data and extraction of useful information is easier, possibilities to analyse data and compare it with standard clinical data are provided, and also integration with data obtained from sensors located in the house of the patient [1].

The paper describes a case study of Big Data in healthcare, emphasizing the advantages of this technology in medical applications for collecting and processing data and extracting useful information for better treatments and surveillance of pregnant women. The Chapter 1 describes the software architecture of the Big Data system in healthcare with a practical approach in obstetrics & gynecology. The Chapter 2 presents Google BigQuery, a software tool from Google, which is useful for Big Data management, followed by the Conclusions reflecting the possibilities of technology use for the medical domain.

1. Big Data system architecture for healthcare

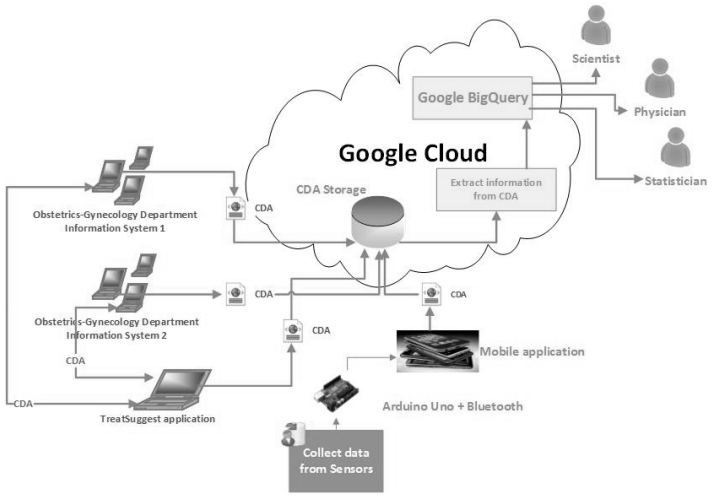

Figure 1. System Architecture

Figure 1 presents the system architecture with four components connected in a cloud, particularly Google Cloud, managing information about the patient. The components of the integrated system are:

- Two Obstetrics & Gynaecology Department Information Systems,
- The "TreatSuggest" application,
- A home fetal surveillance system that collects data from sensors and sends them to the cloud using Arduino Uno and Bluetooth,
- A mobile application. Communication between the components and in the cloud is using HL7 Clinical Document Architecture (HL7 CDA).

The Obstetrics & Gynecology Department Information System is an application developed in Visual Studio.NET 2010 using ASP.NET and C# language. The database is using Microsoft SQL Server 2008 [2]. The Obstetrics & Gynecology Department systems send a diagnostic and a treatment to the TreatSuggest application that will respond with a message if there is available a better treatment than the suggested one for a certain situation. The communication messages use HL7 CDA.

The component "TreatSuggest" is a medical application that supports physician's activity in selecting treatments for patients. The application is written in Visual Studio 2010, Visual C#. "TreatSuggest" suggests successful treatments and retrieves successful treatments from other physicians [3]. Physicians supply the information related to this component. All the successful treatments are saved in the local database of the physician and sent to a global treatments database. This justifies the necessity of involving the Big Data technology for processing. The "TreatSuggest" application is a local application that sends medical information about successful treatments in a cloud database and retrieves from it successful treatments for specific health conditions. The physicians may improve their work and increase the quality of the medical act dealing more easily with a big amount of data. The patients will receive better treatments based on evidence coming from peers on certain diseases. From the global database of treatments, using Big Data procedures, the treatments for a certain disease are extracted at the physician's request The database with the treatments contains the name of the disease, treatment for the disease previously input by peers of the domain and the characteristics of the patient. Data is not associated with personal information about the patient, so the security measures are the usual ones for this kind of applications. All the treatments are collected and analysed to obtain a better result for next treatments on other patients.

The fetal surveillance system component collects the data from the pregnant woman and sends it using an Arduino Uno component and a Bluetooth port to an Android application. The Android application creates an XML file - HL7 CDA standard and sends it to the cloud.

Communication between the systems and the cloud is based on HL7 CDA. This standard is a document mark-up standard that specifies the structure and semantics of "clinical documents" for the purpose of data exchange. It can include text, images, sounds, and other multimedia data and could be any of the following: discharge summary, referral, clinical summary, history/physical examination, diagnostic report, prescription, or public health report [4]. The information sent from the systems in the Google cloud is anonymous. It is not necessary to know the name and the personal ID of the patient because the Google BigQuery is used for statistical purposes supporting the work of physicians and researchers accessing the system. The security measures are the usual information systems ones, needing the name and the personal ID for identification in the system.

2. Using Google BigQuery

BigQuery is a web service based on RESTful, giving the possibility to analyse massively large datasets working in conjunction with Google Storage. Ii is an Infrastructure as a Service (IaaS) [5]. This web service ensures interactive analysis of up to billions of rows. It is scalable and easy to use, making developers possible to

execute powerful data analytics on demand [6]. Several features of BigQuery are presented [5]:

- Data management (creates and deletes tables from different sources),
- Query (the queries are expressed in SQL language and the results are returned in JSON with a maximum reply length of approximately 64 MB),
- Integration (can use Google App Script, Google Spreadsheets or any language that can work with REST API),
- Access control (shares datasets with arbitrary individuals, groups, etc.).

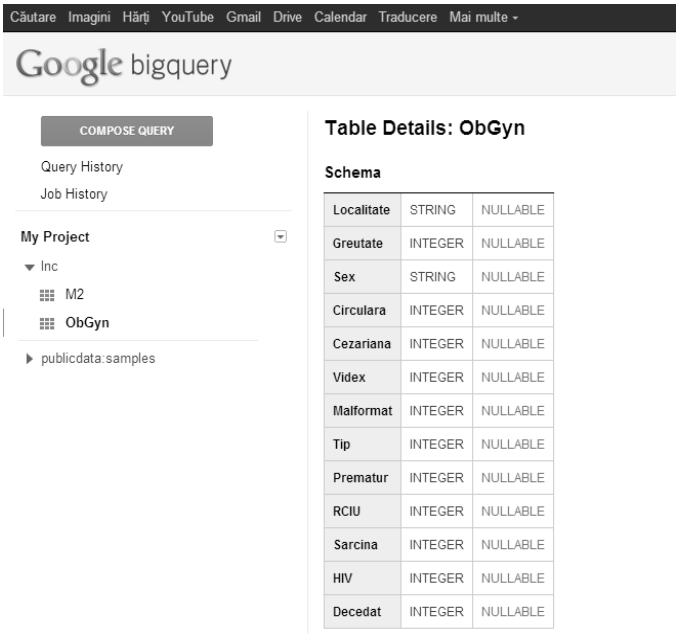

Figure 2. Data from the Obstetrics & Gynaecology Department information systems

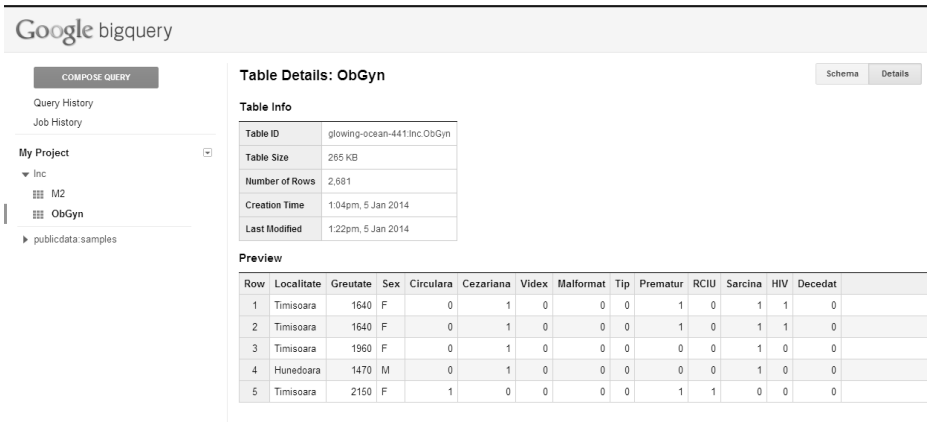

Figure 3. Medical data from ObGyn table

Google BiqQuery gives the possibility to analyse big amount of medical data received from different sources as a good solution for medical staff, researcher or

statisticians. Figure 2 and Figure 3 present medical data uploaded in Google BigQuery that will be accessed with different tools supporting any language working with REST API, or Google App Script or Google Spreadsheets. The applications will send the information to Google cloud where will be stored. The stored information from Google cloud is used in BigQuery. The data is received in CDA format and on Google cloud will be available an application that extracts the needed data for the specific situation.

3. Conclusions

The paper presents a system architecture with integrated components: two Obstetrics-Gynaecology Department information systems, a treatment suggestion application and a home fetal surveillance application that sends medical information to Google cloud. The information is analysed using Google BiqQuery specialized in analysing big amount of data, responding to the medical staff need to analyse big amounts of data in the decision process.

The presented architecture and tools give a flexible frame for processing Big Data in order to obtain better results in patient treatment. Physicians have available an easy to use tool to obtain personalized diagnosis information for their patients. They describe in the system a particular situation and this is matched with the best result in the database in the cloud. The database is improving in real time with the results evaluated by the associated physicians. The processing is conveniently displayed as charts and graphics.

The work presented in this paper will be followed by proposing evaluation criteria for the solutions and parameters to validate the improvement in treatments.

References

[1] J. Sun, C. Reddy, Big Data Analytics for Healthcare, *Tutorial presentation at the SIAM International Conference on Data Mining*, Austin (2013).

[2] O. Lupşe, M. Vida, L. Stoicu-Tivadar, V. Stoicu-Tivadar, Using HL7 CDA and CCD standards to improve communication between healthcare information systems, *9th IEEE International Symposium on Intelligent Systems and Informatics, SISY, Serbia,* (2011), 435-457.

[3] O. Lupse, L. Stoicu-Tivadar, C. Golie, Assisted prescription based on successful treatments, *IEEE International Conference on e-Health and Bioengineering,* (2013).

[4] HL7 version 3 Interoperability Standards Normative Edition 2009, Based on HL7 v3 Data Types, Release 1, Disk 1 – Standards Publication

[5] Google BigQuery, http://en.wikipedia.org/wiki/BigQuery (accessed December 2014)

[6] Google BigQuery, https://developers.google.com/bigquery/ (accessed December 2014)