Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 113 (2017) 585–590

Procedia Computer Science

www.elsevier.com/locate/procedia

International Workshop on Healthcare Interoperability and Pervasive Intelligent Systems (HiPIS 2017)

Towards of a Real-time Big Data Architecture to Intensive Care

André Gonçalves[a], Filipe Portela*[a], Manuel Filipe Santos[a] and Fernando Rua[b]

[a]*Algoritmi Research Center, University of Minho, Guimarães, Portugal*
[b]*Intensive Care Unit, Centro Hospitalar do Porto, Portugal*

Abstract

These days the exponential increase in the volume and variety of data stored by companies and organizations of various sectors of activity, has required to organizations the search for new solutions to improve their services and/or products, taking advantage of technological evolution. As a response to the inability of organizations to process large quantities and varieties of data, in the technological market, arise the Big Data. This emerging concept defined mainly by the volume, velocity and variety has evolved greatly in part by its ability to generate value for organizations in decision making. Currently, the health care sector is one of the five sectors of activity where the potential of Big Data growth most stands out. However, the way to go is still long and in fact there are few organizations, related to health care, that are taking advantage of the true potential of Big Data. The main target of this research is to produce a real-time Big Data architecture to the INTCare system, of the Centro Hospitalar do Porto, using the main open source big data solution, the Apache Hadoop. As a result of the first phase of this research we obtained a generic architecture who can be adopted by other Intensive Care Units.

Peer-review under responsibility of the Conference Program Chairs.

Keywords: Intensive Medicine; Intensive Care Units; Real-time; Big Data; Architecture; Hadoop.

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: cfp@dsi.uminho.pt

1877-0509
Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2017.08.294

1. Introduction

Over the past 20 years, the amount of data has increased in large scale in several areas [1]. According to an International Data Corporation (IDC) report [2], in 2010, the amount of data created and replicated exceeded the Zettabytes barrier, reaching in 2011 the 1.8 Zettabytes. As at the date of the report, in 2011, the prospect was that this growth would increase nine-fold over the next five years [2], meaning that in 2016 was expected that the volume of data was close to 17 Zettabytes. According to an EMC report with research and analysis by IDC, health care accounts for a significant percentage of the data in the digital universe. In 2013, digital data about health care was 153 Exabytes, which, with an accelerated growth rate, in the order of 48% per year, is expected that the volume reach 2314 Exabytes in 2020 [3]. Yet, in the same report, they indicate that the global digital universe presents a growth of 40% a year, allowing to conclude that the growth of digital data, in health care, will be faster than the rest of the digital universe.

The constant, predictable and significant data increase has led to the introduction of a new paradigm, the Big Data. The term Big Data has come up with the explosive increase in data globally and is mainly used to describe huge data sets. In comparison to traditional data sets, Big Data is often associated with real-time analysis of large amounts of unstructured data [1]. At the level of health care, the information technology, the business world and the clinical research are creating an emerging movement under the Big Data banner [4]. Proponents of this movement argue that it is a new approach that will transform and accelerate health care, correcting decades of misguided research, and reshape clinical science as well as how to care for patients. Intensive Care Units (ICUs) are an environment in which the demand for change is increasing due to the constant recording and processing of large amounts of data, related with patients' health conditions [5]. According to Portela et al. [5], the complex condition of critically ill patients and the enormous amount of data, may hinder the decision-making, by intensivists, at the time of providing the best health conditions. The authors also add, that intensivists do not have the time to analyse the conditions of the patient in an assertive and consolidated way, and it is becoming increasingly necessary to develop systems capable of assisting intensivists in the shortest period, i.e., in real-time. In the first phase of this research, based on the INTCare architecture, other analysed big data architectures, and a superficial analysis of the Apache Hadoop ecosystem, we produced a Big Data architecture, which we believe can be used in any Intensive Care Unit. As a way of validating and release a final version of the architecture, to the INTCare, in the second phase of this research, we'll perform a detailed analysis of the Hadoop ecosystem and produce a prototype that will simulate the processing and storage of streaming data from bedside monitors, collected in the ICU of the Centro Hospitalar do Porto (CHP).

Finally, in addition to the introduction, this article is composed of four other sections. Second section, Background, provides background knowledge on topics related to Intensive Medicine and Intensive Care Units, the INTCare system, the main theme, Big Data and related work. Next, the third section identifies the research methodology used and how it applies in this research project. The fourth section describes the proposed real-time Big Data architecture, based on open source technologies. In the last section, the conclusions are presented, as well as the future work.

2. Background

2.1. Intensive Medicine, Intensive Care Units and INTCare Project

Intensive Medicine (IM) is a multidisciplinary area of medical sciences focused on the prevention, diagnosis and treatment of serious diseases, that are considered as threats to the lives of patients and which are characterized by causing failure of one or more vital organs [6]. Intensive Care Units (ICUs) are special hospital units prepared to provide health care to patients whose survival depends on intensive care. In these units, patients' vital signs are continuously monitored by various life support devices, which together with drug delivery enable patient recovery [6].

INTCare is a project developed at the Intensive Care Unit (ICU) of the Centro Hospitalar do Porto (CHP), which emerged from the need to build an intelligent system to automate the data collection and analysis process and, consequently, predict organ failure and their effects on patients [7]. The original system has suffered many changes, resulting from the expansion of ICU needs and the potentialities generated by the growing amount of electronic data. As a result, it is currently a Pervasive Intelligent Decision Support System (PIDSS) that acts automatically and in real-time, to provide more information to the intensivists of the UCI [7]. The INTCare system presents an architecture [8]