A Big Data-driven Model for the Optimization of Healthcare Processes

Vassiliki KOUFI\textsuperscript{1,2}, FloraMALAMATENIOU\textsuperscript{2} and
George VASSILACOPOULOS\textsuperscript{2,3}

\textsuperscript{2}Department of Digital Systems, University of Piraeus, Greece
\textsuperscript{3}New York University, New York, NY

Abstract. Healthcare organizations increasingly navigate a highly volatile, complex environment in which technological advancements and new healthcare delivery business models are the only constants. In their effort to out-perform in this environment, healthcare organizations need to be agile enough in order to become responsive to these increasingly changing conditions. To act with agility, healthcare organizations need to discover new ways to optimize their operations.

To this end, they focus on healthcare processes that guide healthcare delivery and on the technologies that support them. Business process management (BPM) and Service-Oriented Architecture (SOA) can provide a flexible, dynamic, cloud-ready infrastructure where business process analytics can be utilized to extract useful insights from mountains of raw data, and make them work in ways beyond the abilities of human brains, or IT systems from just a year ago. This paper presents a framework which provides healthcare professionals gain better insight within and across your business processes. In particular, it performs real-time analysis on process-related data in order reveal areas of potential process improvement.

Keywords. Big Data, Business Process Analytics, Real-time Analytics, Healthcare Process Optimization

Introduction

The last few years, healthcare systems around the world are experiencing fundamental transformation as they move from a volume-based to a value-based healthcare delivery model in an effort to meet growing demand for care and enhance healthcare quality while restraining costs [1]. Moving to a value-driven model demands agility from people, processes and technology. Thus, healthcare providers are pushed to focus anew to business processes that guide healthcare delivery and on the technologies that support them. Recently, business process analytics have emerged as a powerful tool to assist healthcare organizations in gaining useful insights regarding the efficiency and effectiveness of their processes. Thus, areas of potential process improvement can be identified, ultimately leading to better financial and budgetary performance, deeper citizen/patient-centric relationships and significant improvement in the way health care is conceived and delivered for better outcomes across the entire spectrum of health services.

\textsuperscript{1} Dr V. Koufi, Dept. of Digital Systems, University of Piraeus, 185 34, Greece; Email: vassok@unipi.gr.
There are three types of business process analysis (BPA), namely validation, verification, and performance, all of which require processing large volumes of process-related data (process and event data) [4]. These data, lumped under the term ‘big data’, can be both structured and unstructured and are obtained from several sources. Currently, little insight can be gained from these data regarding the efficiency and effectiveness of organizational processes due to the volume and the complexity of the data or to data access restrictions (e.g., real-time access is not feasible, certain parts of the data are accessible, etc). The last few years, a number of approaches have been proposed for the optimization of business process performance [2][3]. Most of them use traditional approaches, which, nowadays, may prove to be inadequate since healthcare organizations are already maintaining very large process model repositories representing the various healthcare processes (e.g., clinical, administrative, etc) involved in everyday clinical practice. This, along with the proliferation of clinical information systems, electronic health records and connected medical devices, have created an unprecedented volume of performance-related information, which cannot be managed efficiently by using traditional tools and techniques. To address these issues, an innovative framework is required which will be able to sort through this torrent of complexity and data, and help healthcare organizations deliver on the aforementioned demands. A couple of approaches have already been proposed for performing business process analytics using a big data approach [4][5]. However, these approaches have certain shortcomings such as significant delay in processing data related to business process performance.

This paper presents a framework which facilitates healthcare process optimization by evaluating in real time the business processes enabled by Service Oriented Architecture (SOA) in order to ensure that they meet the stated operational and performance objectives. In particular, it monitors the deployed healthcare processes during operation and analyzes the relevant process-related data by means of a cloud-based business process analytics service. The latter integrates process-related data from all possible resources into a structured view and analyses it in order to provide better insights for decision making to the key stakeholders. Thus, areas of potential process improvement can be identified and better performance of business processes can be achieved without having to recruit and train personnel or continually invest in new technology.

1. Methods

The basic motivation for this research stems from our involvement in a recent project concerned with developing an efficient and robust prototype ePrescribing system for the needs of the Greek National Health Service with the objective to improve quality of care while containing cost. Figure 1 illustrates a high-level architectural view of the proposed framework which comprises three main components: (a) healthcare provider infrastructure, (b) Healthcare Process Analytics Service (HPAS), and (c) client portal.

Healthcare provider infrastructure comprises five main components. These are: the existing IT infrastructure, namely the health information systems (HIS) hosted by healthcare providers, which are heterogeneous and reside at different settings; the Enterprise Service Bus (ESB) which includes adapters to expose existing systems and provide transport connectivity; the Business Process Management System (BPMS) part of which is a Business Process Execution Language (BPEL) engine that is capable of
interpreting and executing business processes described in BPEL by orchestrating existing services; the web services developed or created from the existing systems and deployed on the cloud infrastructure and the authorization server that manages who, in terms of role, can perform the various BPEL activities, such as invoking a service or assigning a new value in an XML document.

The HPAS is essentially a cloud-based service that performs real-time analysis on the data generated during the execution of the healthcare processes. Its main component is an agent, which is a Java Virtual Machine (JVM) process that hosts the sub-components through which events flow from the external source (ESB) to a central repository. After being archived in the central repository, events are retrieved in order to be analyzed.

Real-time analytics on process-related data is performed in three steps:

(a) **Collection of process-related data:** In this step, process-related big data are being collected in real time from the various sources that generate them, namely BPMS events, SOA events etc. These data are events that occur within process steps, such as failure to invoke a service. Essentially, they constitute a continuous and changing sequence of data that continuously arrive at central event-handling backbone of the HPAS, namely the event channel. Usually, they make up a massive volume (e.g., terabytes) and are temporally ordered, fast changing, and potentially infinite.

(b) **Processing of process-related data:** As soon as the process-related data are collected, their processing is initiated (streaming data processing). Each one of the events collected has a certain proprietary format which depends on the system which generated it. In order for these events to be integrated and further analyzed, they first need to be converted according to a standardized event format, like XML Business Process Analytics Format (BPAF) which has been proposed by Workflow Management Coalition [6]. According to this standard, each event contains a minimum set of information, namely a unique identifier, the identifiers of the process definition and the process instance it originated from, the date and time the event occurred and the state of the process at the time the event occurred. However, since many of the healthcare processes may by cross-organizational, additional information is required. Thus, apart from the aforementioned set of data, a number of additional data items are included in each event recorded. These are, among others:
i. **Location:** it refers to the node where the event was generated and is particularly important in a highly distributed environment, such as the healthcare environment where healthcare providers participating in a healthcare process execution may be geographically distributed.

ii. **Business process participant ID:** it refers to the specific person who initiated the process activity instance.

iii. **Data relevant to the process instance:** These may be the prior state of the process activity and may be particularly useful when a process activity state can be reached from multiple other states.

Actions performed during data processing may include, among others, merging data, doing calculations, connecting data with outside data sources that contain structured data (e.g. a HIS) etc.

(c) **Data analysis and visualization of process metrics:** After data have been processed and converted in BPAF format they are delivered to a central repository for analysis and visualization of results. The analytic figures which are used to measure process performance (process metrics) are absolute measurements, which can be obtained by analyzing the time stamps of the process-related events that belong to the same process or activity instance. Calculations performed on these time stamps can provide useful insights into the behavior of each process instance. For example, activity instance metrics can be:

i. **Completion time** – it is the time required for the completion of a process instance, that is the time that elapses from the moment an instance becomes ready for execution to the moment the same instance becomes completed (state transition from Ready to Completed, regardless of the intermediary state transitions that might have occurred, e.g. from Ready to Suspended).

ii. **Assignment time** – it is the time it takes for a process instance to be assigned to a certain process participant (e.g. doctor, radiologist etc.) (state transition from Ready to Assigned).

iii. **Instance initiation time** – it is the time that elapses from the moment a process participant picks a process instance for execution to the time he actually initiates the execution of the instance (state transition from Assigned to Running).

iv. **Suspension time** – it is the time a process instance may be put to suspension by the process participant executing it.

v. **Processing time** – it is the actual time it took for a process instance to be completed and can be calculated by subtracting (ii), (iii) and (iv) from completion time.

The output generated by the analysis of the process-related data can then be used in order to generate alerts and relevant reports which provide useful insights regarding process performance. These data can also be used for the generation of dashboards and ad-hoc queries.

The client is essentially a web portal, whereby medical professionals and decision makers can access reports on the performance of certain healthcare processes.

### Results

To illustrate the functionality of the proposed framework, a prototype cloud-based system was implemented. The prototype system implementation was based on the Oracle 11g SOA Suite which was set up in the Amazon Elastic Compute Cloud
(Amazon EC2) [7][8]. In particular, the following components of this Suite have been used: (i) Integrated Service Environment (ISE) for developing the web services; (ii) a services registry for discovering and managing the lifecycle of services; (iii) a BPEL-based orchestration engine for tying the services into business processes which are also exposed as web services; and, (iv) an enterprise portal for healthcare professionals, patients and collaborating healthcare organizations to access content, access relevance performance metrics, collaborate and take actions via interaction with healthcare processes. The implementation of HPAS was based on Apache Flume, a distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of log data [9].

3. Discussion

Nowadays, healthcare environment is substantially more volatile and complex than in years past. Healthcare processes have become increasingly important in healthcare enterprise and in conjunction with health information systems organizations are producing increasingly vast volumes of high velocity data in a myriad of formats. Currently, data are captured, stored and analyzed in order to extract insights after the fact. However, many benefits would be realized if data could be continuously analyzed in real-time thus enabling the uptake of immediate action. This paper presents a framework that sets the ground to realize these goals by performing real-time healthcare process analytics. In particular, the proposed framework performs continuous, real-time analysis of streaming and, when appropriate, of stored data, with the ability to take immediate action on the discovered insights. System evaluation is a task to be undertaken in the near future aiming at determining the system usability. Thus, its potential weaknesses may be revealed suggesting alterations in the system design and directions for future work.

References