

Development of CEP System based on Big Data Analysis Techniques and Its Application

Mi-Jin Kim¹, Yun-Sik Yu¹

¹ Convergence of IT Devices Institute Busan(CIDI), Dong-eui University,
176 Eomgwang-ro, Busanjin-gu, Busan, Korea
{agicap, ysyu}@deu.ac.kr

Abstract. In the rapidly business and managerial environment and marketplace, there are increasing needs for massive amount Big Data on a various corporate entity levels. In reality, however, access to Big Data use has been limited and not easy due to the high cost of analytics system. In addition, as real time analysis being emerged as an alternative of conventional data analytics, it requires to develop CEP-based system that analyzes data from the perspectives of real time processing event, rather than Hadoop-based one from the perspectives of batch processing and batch analytics.

In this paper, CEP-based real time analytics system was developed to establish low-cost Big Data system for small-medium size hospitals. Another purpose is to analyze patient-oriented information and equipment data at hospitals in combination of hospital ERP system, thereby implementing the system that can provide systemic and efficient patient management and administrative management at hospitals and suggesting such verified effect.

Keywords: Big data, CEP(Complex Event Processing), Real-time Analysis

1 Introduction

With the development of IT technologies and the recent increase in data amount generated by individuals, governments and corporate entities; more interest in Big Data is growing. Big Data is not just any new concept, it refers to huge and massive amounts of data accumulated in variety of formats, in excess of the scope of data saving, management and analysis allowed by the previous database such as a Relational Database [1].

Big Data is currently generated by and used in many fields, and highlighted especially in health and medical fields. The increase in chronic and degenerative disease cases caused by the population aging, there have been much effort by health and medical fields in many studies, in order to use Big Data to reduce medical expenses, prevent contagious diseases, and raise the bar in quality of medical services. In the results, Big Data is proposed as effective alternatives to explore more efficient diagnosis and treatment, as well as to predict prognosis. It is estimated that there will be cost-saving effects of \$300 billion annually, if Big Data is well applied in the US medical fields[2]. A McKinsey report also said about the added-value in medical

fields which is connected to allowing to save nation-wide medical expenses and to do innovative clinical trials.

The purpose of this study is to verify the effectiveness of Big Data as a useful method for information-oriented health and medical fields, by developing Big Data-based real time monitoring system in manners of real time analysis rather than previous batch-type analytics method, and then by developing the system where massive amounts of data generated by health and medical fields are used and analyzed on a real time basis and systemic service can be provided by medical institutions in more efficient manners in patient and hospital management.

2 Related Research

2.1 Hadoop

Hadoop is a solution with its key focused on distributed processing technology, which is currently most favorable for Big Data processing. It is Java-based framework with Apache open source process to process massive data, using a relatively simple program model[3]. Hadoop is used as a core technology by Yahoo and Facebook, while being applied to many other companies' own solution. Hadoop is composed of the distributed file system called HDFS (Hadoop Distributed File System) and the distributed processing system called MapReduce. The operation method is HDFS, but MapReduce is composed of the masters called Namenode and Datanode as well as multiple slaves in its structure.

2.2 CEP(Complex Event Processing)

CEP is complex event processing technology to extract meaningful data in real time basis from events from various event sources, thereby performing the corresponding actions[4]. Event data herein refers to stream data, which are data of continuous massive inputs, with important time sequences and endless data. It is impossible to process and analyze such stream data in real time basis into a conventional Relational Database. CEP is an event data processing solution that can provide real time analysis of such stream data. That is, it is possible to do real time processing of hundreds/millions of various high speed event stream based on In-Memory without saving it to database, file or Hadoop.

3 System Compositions and Implementation

CEP technology is currently emerging across business fields, as the simplest and strongest method to implement the real time business intelligence based on timely

analysis, providing new values including real time monitoring, with an early alarm and production field management by processing and analyzing various events. In this study, we aimed at providing a systematic and organized business environment for efficient patient management and administrative management at hospitals, by using CEP-based advantages in consideration of the low cost of Big Data and then establishing the Big Data-applicable real time analytics system in combination with hospital ERP systems with yet a insufficient number of cases. In addition, the main Adaptor and data publisher/customizing functions were implemented, to allow to identify and develop UI screens - depending on the needs of each hospital.

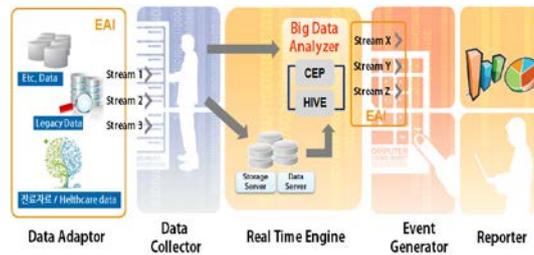


Fig. 1. CEP-based real-time analysis system architecture.

To be brief on the processing of the entire system, Legacy saves log details generated in its own process as in Fig. 2, while performing real time analysis processing as far as major tasks among such processes. Batch layer collects data through data aggregation and extracts data through analysis works. Depending on such extracted results, there may be either storage to middle repository or request for real time analysis. Real time processing layer processes the requests from Legacy and from batch layers, requesting for data as necessary for processing and saving the results to the middle repository. For such processed results, there will be monitoring system API calls. Dashboard visualizes the data depending on the user's request, and then saved the Dashboard-processed data to the middle repository.

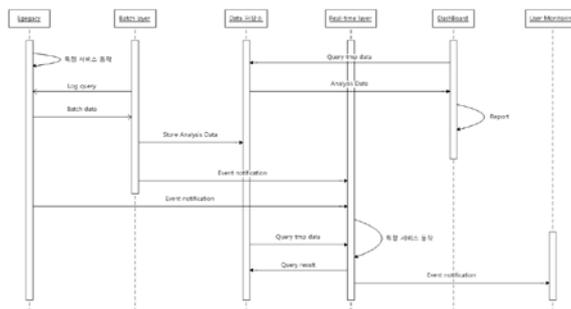


Fig. 2. The entire system process definition.

Fig. 3 is the screenshot of monitoring analyzed of all data on Output Adaptor settings on Dashboard. Through filtering functions, Server/Adaptor/Event/Hour can also be set up in order to analyze necessary parts alone. Fig. 4 is the screenshot of analysis, when filtering Server selected as 210.109.9.177, Output Adaptor selected as Event, Event selected as event_10 and Hour selected as 07:00 ~ 24:00



Fig. 3. Result screen of full data.



Fig. 4. Result screen of the selected filter.

4 Conclusion

It can be said that the fundamentals for Big Data analytics, using the rapidly developing information systems in medical institutions. Nevertheless, the potentials of such systems are not sufficiently utilized. If the system developed by this study can be used not only by government bodies but also by SMEs in such environment established to enable clinical trials, patient management and managerial management, it will lead to amazing developments in these fields in the future. Although such developed systems can be customized as necessary, continuous research activities are needed so that the system can be also applicable to many other fields including shipping and logistics.

Acknowledgments. This article was written and the study was conducted as a part of 2014~2015 Nurimaru R&BD Project (SW convergence R & BD project) supported by the National IT Industry Promotion Agency and supervised by the NIPA Busan.

References

1. Lee Sung-hoon and Lee Dong-woo. "Current Status of Big Data Utilization" The Journal of Digital Policy, The Society of Digital Policy & Management, 11(2), pp.229~233, 2013.
2. Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers A. H. "Big data: The next frontier for innovation, competition, and productivity," McKinsey Global Institute, 2011.
3. K.-W. Park, K.-J. Ban, S.-H. Song, and E.-K. Kim, "Cloud-based Intelligent Management System for Photovoltaic Power Plants," J. of the Korea Institute of Electronic Communication Sciences, vol. 7, no. 3, June 2012, pp. 591-596.

4. Luckham, David C., and Brian Frasca. Complex event processing in distributed systems. Computer Systems Laboratory Technical Report CSL-TR-98-754. Stanford University, Stanford 28, 1998.