Design of Infrastructure to Analyze Big Data

Sungbum Park* · Sangwon Lee** · Hyunsup Ahn*** · In-Hwan Jung****

*Department of Management Planning, National Information Society Agency
**D. of Information & Electronic Commerce (I. of Convergence & Creativity), Wonkwang University
***Department of Wirtschaftsinformatik, Technische Universität Braunschweig
****Department of Computer Engineering, Hansung University

E-mail: parksb@nia.or.kr, sangwonlee@wku.ac.kr, hs.ahn@tu-bs.de, ihjung@hansung.ac.kr

ABSTRACT

Nowadays, owing to the development of hardware, most of enterprises should operate and manage bigger amount of data than they used to do in the past. For this reason, the enterprises ceaselessly and urgently need systemized tools, platforms, and analysis methodologies in order to collect, store, and handle explosively increasing data. In this paper, we firstly understand key elements of Big Data, and secondly define critical factors for Big Data application with these elements. Thirdly, we research on various analysis techniques for Big Data analysis. Lastly, we propose an infrastructure to analyze big data.

I. Introduction

The recent development of hardware let most of enterprises operate and manage bigger amount of data than they used to do in the past. So, the enterprises ceaselessly and urgently need systemized tools, platforms, and analysis methodologies in order to collect, store, and handle explosively increasing data. Our research focuses on designing an infrastructure so as to analyze Big Data. Above all, we define key elements of Big Data and critical factors for Big Data application. And then, after researching on various analysis techniques for Big Data, we propose an infrastructure to analyze bid data.

II. Techniques of Big Data

There are some key elements in discussing about...
Big Data. The key elements could be classified into several parts from two major points of view (Figure 1). The key technique of Big Data is composed of the volume of data, the input/output velocity of data, and the variety of data. And the factors in practically using Big Data are resource, technique, and human resource. In addition, there are four major techniques for analysis on Big Data: text mining, opinion mining, social network analytics, and cluster analysis. Many enterprises have been using their own infrastructures for Big Data analysis such as Hadoop, R, NoSQL, Hbase. Data processing module is supported by MapRecude framework. Data access module is made up of Pig, Hive, and Avro. Management module is performed by ZooKeeper and Chkwa.

III. Practical Use of Big Data

Big Data have been developed according data processing paradigm (Figure 2). The paradigm has been going along nicely with functional stages such as storage via databases, exploration via search engines, management via Knowledge Management Systems (KMS), sharing via Web 2.0, analysis via Big Data, and inference for situational recognition. The practical use of Big Data is carried out in the exploration stage and its diffusion or sharing is accelerated by KMS or Web 2.0. Nowadays, the stages of analysis and inference are creating various values.

IV. Conclusions

It is said that 90 percent of all data that exist all over the world has been made for the past three years. In these days, Big Data are widely used in the fields of Information Telecommunication, Education, Medical Industry, Finance, and so forth. Although the characteristics of the future society are uncertainty, risk, smartness, and convergence, Big Data would certainly play important roles of insight, response ability, competitive power, and creativity for these characteristics each.

References