

---

# 빅데이터 분석을 위한 인프라 설계

박승범\* · 이상원\*\* · 안현섭\*\*\* · 정인환\*\*\*\*

\*한국정보화진흥원 경영기획부

\*\*원광대학교 정보전자상거래학부(융복합창의연구소)

\*\*\*브라운슈바이크공과대학 경영정보학과

\*\*\*\*한성대학교 컴퓨터공학과

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

## 요 약

요즘에는 하드웨어의 발달 덕분에, 많은 기업들이 과거에 사용했던 데이터보다 훨씬 많은 양의 데이터를 조작하고 관리해야만 한다. 이런 이유에서, 기업들은 폭발적으로 증가하는 데이터를 수집하고 저장하고 다루기 위해서, 체계화된 도구, 플랫폼, 분석 방법론을 끊임없이 긴급하게 필요로 하고 있다. 본 논문에서는 우선 빅 데이터의 주요 요소를 이해하고, 둘째로 이러한 요소들을 활용한 빅 데이터 애플리케이션을 위한 주요 요소를 정의한다. 셋째로, 빅 데이터 분석을 위한 다양한 분석 기법에 대해 연구하고, 마지막으로 빅 데이터 분석을 위한 인프라를 제안한다.

## 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 bid data.

## 키워드

Big Data, Infrastructure, Data Analysis, Practical Use

### I. Introduction

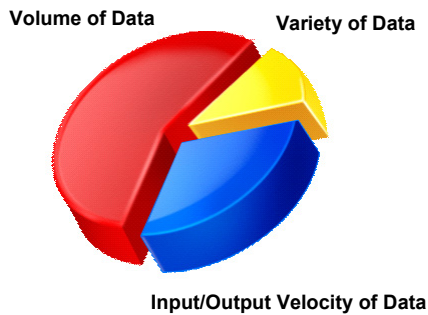
The recent development of hardware let most of enterprises operate and manage bigger amount of data than they used to do in the past. So, the enterprises ceaselessly and urgently need systemized tools, platforms, and analysis methodologies in order to collect, store, and handle explosively increasing 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.

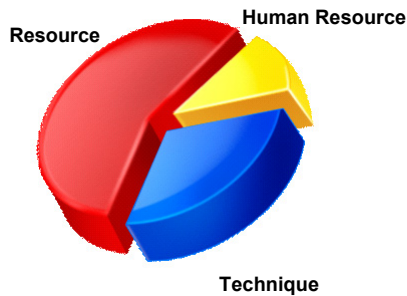
### II. Techniques of Big Data

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. Hadoop has three major modules; data storage, data processing, data access, management, and data connections. Data storage module comprises Hadoop distributed file systems and Hbase. Data processing module is supported by MapReduce framework. Data access module is made up of Pig, Hive, and Avro. Management module is performed by ZooKeeper and Chkwa.



(a) Key Techniques of Big Data



(b) Key Factors of Using Big Data

Figure 1. Factors of Big Data

### 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.

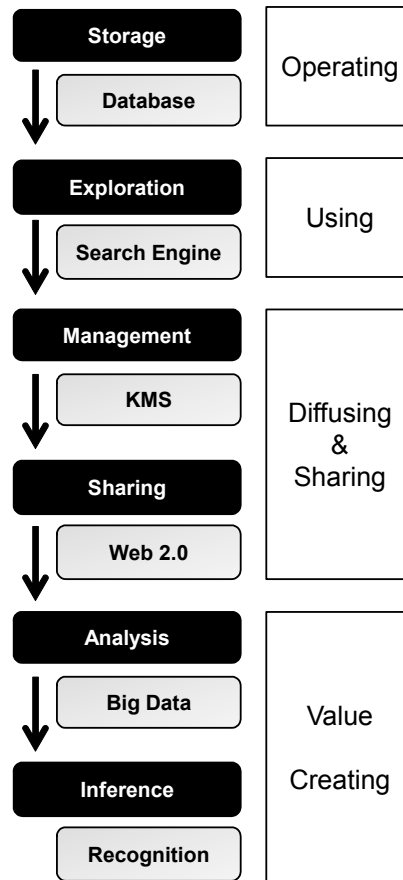


Figure 2. New possibilities of Big Data

### 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

[1] Gartner, Report on Big Data Analytics, Gartner Group, 2011.  
 [2] Jungsook Kim, "Research on Big Data Utilization and Related Works," Journal of Korea Contents Society, Vol. 10, No. 1, pp. 9-116, 2012.  
 [3] Manjae Lee, "Practical Use of Big and Public Data," Internet and Information Security, Vol. 2, No. 2, pp. 47-64, 2011.  
 [4] McKinsey, "Big Data: The Next Frontier for

Innovation, Competition, and Productivity,” McKinsey & Company, Vol. 5, No. 1, 2011.

[5] Vertica, “Managing Big Data with Hadoop & Vertica,” Vertica Systems, Vol. 10, No. 1, 2009.

[6] Tome White, Perfect Guide of Hadoop, Hanvit Media Publishing Co., 2010.