International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 18 (2015) pp 39281-39284

Big Data Analysis System Model using Field Data

Jae-Joong Kim[1], Baek-Min Seong[2], Min-Gyu Lee[3], Hyo-jung Sohn[4] and and Jong-Bae Kim[5*]

[1]Dept. of IT Policy and Mgmt., Graduate School of Soongsil Univ., Seoul 156-743, Korea
[2, 3, 4, 5]Graduate School of Software, Soongsil University, Seoul 156-743, Korea*
E-mail: [1]andy@forcewin.com, [2]feeling127@naver.com, [3]marse101@naver.com, [4]hyojung.sohn@gmail.com, [5]kjb123@ssu.ac.kr*
[5]Jong-Bae Kim (kjb123@ssu.ac.kr) is the corresponding author of this paper.*

Abstract
Today, thanks to the emergence of many methods of handling multi-dimensional data, enterprises have been able to analyze data above the TB size by utilizing BI (Business Intelligence) systems, and can apply the analyzed data to the decision-making process. However, small/medium-sized enterprises that lack IT experts or cannot make a sufficient investment in IT have difficulties catching up with this trend. In addition, the majority of enterprises have not introduced a Manufacturing Execution System (MES), and most of the enterprises that have managed onsite data have stored the data in the form of memos or Excel files, and thus cannot apply it to the decision making process. Thus, in this research, in order to strengthen the competitiveness of small/medium manufacturing enterprises, a Big Data analysis system model was developed to automatically collect, refine process and analyze the data used on the sites of manufacturing enterprises. Through ODE (Office Data Excavation) modules, this analysis system model automatically recognizes the pattern of documents for information (ex. Excel) recorded non-typically on sites, and then extracts, refines, and collects them in typical information. By providing the stored data through various charts by using D3.js, the open-source data visualization library, this model prepares the correlation between pieces of information and the multi-dimensional analysis base, and then effectively supports the decision making system. In addition, an economical Big Data analysis can be made by using the open-source Spago BI instead of the high-priced Big Data solution.

Keywords: BI(Business Intelligence), ODE(Office Data Excavation), Text Analyzing, Data Analysis, Field Data

1 Introduction

In today's rapidly changing business environment, companies are constantly trying to optimize their external processes through collaboration, as well as to develop their internal processes effectively in order to respond quickly to market demand and secure their competiveness [1]. In accordance with this trend, even the technologies called the business process have been proposed[2].Today, while many methods of handling multi-dimensional data have been proposed so that many BI systems can handle data larger than TB, small/medium manufacturing enterprises that do not have IT experts or who cannot make a sufficient investment in IT have difficulties catching up with this trend [3][4]. In addition, compared with large conglomerates only a very small percentage of small to medium-sized manufacturing enterprises have adopted a Manufacturing Execution System (MES), and in many such cases field data has not been saved [5]. Even today, as most field data is saved in handwritten form or to an Excel file, data analysis and decision-making involve manual processes. As a result, it is difficult to analyse the rate of operation of manufacturing processes, production yield, and product quality, as well as to identify the causes of poor products and abnormal phenomena. As can be seen, small and medium-sized manufacturing enterprisesare inevitablyvulnerablewhen competing with conglomerates.For this reason, this research aims to develop a customized analysis solution for small and medium-sized manufacturing enterprises, which are vulnerable compared to conglomerates. The final aim of this research technology development is to contribute to enhancing the competitiveness of small and medium-sized manufacturing enterprises by developing a combined/related analysis and visualization system for all the data needed tostrengthena company's competiveness that is cost efficient and allows for easy adoption and management.

2 Related Works

Major global SI and IT venders, including EMC, IBM, ORACLE, SAP, GOOGLE, and MICROSOFT are focusing on solution and core technology development to achieve early domination of theBig Data market.M&A and technical cooperation of related businesses for technology development werecarried out until 2011, and related services and solutions werelaunched in earnest starting in 2012. IT service companies run Big Data business or develop the platform forBig Data analysis, and social network analysis companies work on various Big Data businesses for ordinary people and companies [6][7].It is explained that most Big Data sales are created by huge IT companies like IBM and HP, and Big Datasolution companies such as Vertica and Cloudera. It is predicted that the size of the Big Data market may increase at a 58% compound annual growth rate(CAGR) for the next 5 years in the future, from 5 billion dollars in 2012 to 53 billion dollars in 2017 [8].IDC, an IT market analysis and consulting

institute, predicted that the size of the Big Data market would increase from 3.2 billion dollars in 2010 to 16.9 billion dollars in 2015. This is a CAGR of 40%, which is seven times larger than the overall ICT market growth rate. In particular, it is expected that the technology and service market of Big Datain the Asia-Pacific area (other than Japan) may show a high growth rate, at an annual average 46.8% for the next 5 years [9].In the rapidly changing era of smart digital, the direction of the viewpoint is changing from how to predict and prepare for the era of Big Data, to how to process and use it. The series of techniques of saving, collecting, managing, distributing, and analysingBig Data are called Big Data processing technology. In reality, global companies have built new business models through Big Data analysis and are using them successfully, and have also strengthened the capacity of Big Data processing technology [10].Global competitorsare already achieving high levels of effectiveness. It is critical to develop new business models that can create value through Big Data analysis in diverse areas, by benchmarking the success of Hadoop, an open source based Big Data platform [11].

3 Big Data analyzing system on site

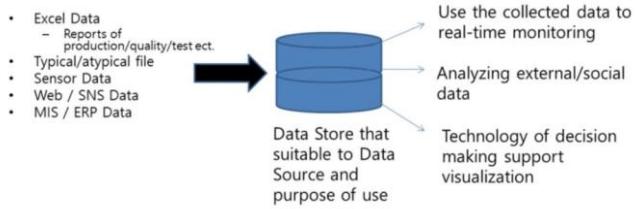

Fig1.Concept of the System

In this research, a platform structure is developed that can be adopted to analyze data flexibly on a company's site by automatically collecting, refining, and processinga manufacturing enterprise's on-site data(typical, atypical). This system provides effective decision making support by offering a correlation between information and a multi-dimensional analysis base, and maximizing system utilization through an intuitive and user-friendly user interface. It also applies diverse visualization technology that is customized tothe executive group and site staff of small and medium-sized manufacturing enterprises.The system for on-site analysis of Big Data that is developed in this research involvesan ETL(Extract Transform Loading) process that extracts, refines and collects data using the DB function of each system that exists in ERP, MIS, and so on. Data that is recorded atypically on site (ex. Excel) is extracted, refined, and collected as typical data by an ODE(Office Data Excavation) module that automatically recognizes the patterns of the data. The digital information is saved in a NoSQL storage after the index process for various statistical analysis, and the stored digital information is enabled to be analyzed in real-time series, visualized in Straight Table, Pivot Table or diverse Chart formations by each index, and calculated using various Aggregation functions.The information that has the same mutual Relation as the information stored in the existing R-DB is saved in Graph-based NoSQL storage after the relations are modeled and analyzed when it is collected. Network After all, the information that stored in Graph based modeling, provides intuitive and user-friendly analyzing tool with visualized in Network format.The first of the final aims of the research is to modularize the Big Data analysis system to enable the development of a system for data collecting/saving/analysis with technology support for early set-up only amounting to about 1~2M/M. The second aim is to providea visually intuitive interface so that workers on the manufacturing site who are not familiarwith the computer system can use it easily. Lastly, the goal is to develop a S/W package to minimize adoption and maintenance costs by actively using an open operating system, web server, open source, and so on.

3.1 Internet data collection/analysis system

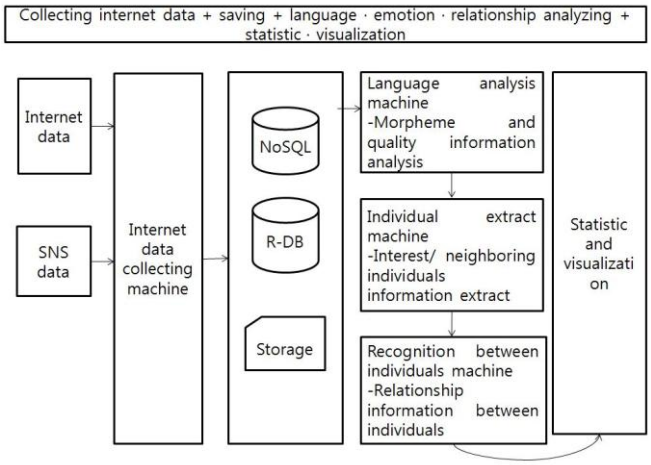

Fig2.Internet data collecting/analyzing system

Many steps are needed from discovering to using Big Data. Firstly, data must be collected from diverse data sources. Next, it must be processed through a preconditioning process, such as by filtering the useless data from collected data or processing it into a proper format, etc. After that, the information is saved and maintained in order and the meaningful information is visualized through the information processing analysis process to obtain useful data, or the hidden information Internet data collection/analysis system follows the steps of collecting-saving-analyzing linguistic and emotional relation – statistics. In the internet data collecting and saving steps, the data related to products, companies, and market information on the internet on news/blogs/SNS and so onis collected and saved. After analyzing morphemes and the quality of atypical information, it is made into simultaneous appearance morphemes.The relationship between individuals is discovered and emotional information is analyzed after interesting words and neighboring wordsare extracted through quality information. Use statistics and visualized analyzed results to analyze present condition/issue and market respond.

3.2 Auto collecting Excel data system on site

Developed a Data Integration connected-collecting tool for collecting auto recognizing atypical documents module(ODE : Office DataExcavation)and typical data(ERP, MIS) for auto collecting module offield data(atypical Excel data).Customer selects the Excel file he/she wants on the web(management machine), and matches it with the same type of template by using a program(management machine) after developing an Excel data pattern consonance algorism. If it is sent to the target after transforming to CSV format and saving inBig Data storage, it is finally sent (automatic) the same type of source in a group.

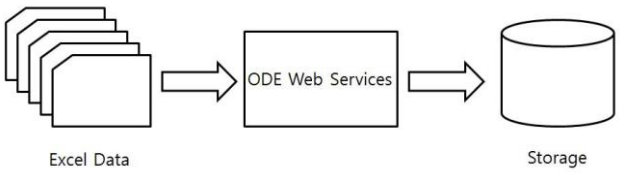

Fig3. Auto Excel data collection System

3.3 Graph storage for data relation analysis

Save and maintain Big Data effectively to use the extracted information through the collection of internal · external unrefined data. The correct way of usingBig Data technology, which is organized with diverse files, is to save all data even it is small, process data in low-cost in real-time, more quickly and easily analyze the processed data, and use it immediatelyin business decision making.In this way, various data storages have been developed, followed by diverse data sources and analysis environments. It is a real-time storage for facility data as well as for multi-analysis of typical company data. It is also graph storage to use a Network Visualization technique that all employees of a company can use easily, including executives and working-level staff. In addition, it is graph storage for analyzing the relationship between data and visualization.

3.4 Visualizing technology for supporting decision making

Visualization is the easiestway to receive and understand the large amount of information that our brains are exposed to. If you visualize the numerous patterns made of data, it helps you to understand clearly and intuitively what kind of work is happening, and how the work will develop in the future. It is also effective for a business working group that needs to easily understand an analysis of Big Data. Most meaningful information has a relationship with other informationand by using it, secures integrity and saves normalized data in R-DB. With existing information analysis tools, it is typical to merely show a 3-dimensional comparison of the information with relationships or just present the Raw Data in a table.Existing information analysis tools have thedisadvantageof not showing the insight with the data due to not being able to express the characteristics of data itself. In this developing technology, the information with the relationship is saved as Graph-based modeling in the data visualization library 'D3.js'. The expression is visualized into a similar structural expression method as a human thinking system, which is called Network Visualization. Monitoring and index analysis is possible through the Network visualization, and multi-dimensional analysis and connection analysis can be performed depending on the data type. Analysis visualization technology is a multi-dimensional analysis tool that enables modeling based on graph analysis, and also has avisualized interface that can be used in diverse environments.The overall composition diagram of theBig Data analyzing system model on site is shown below.

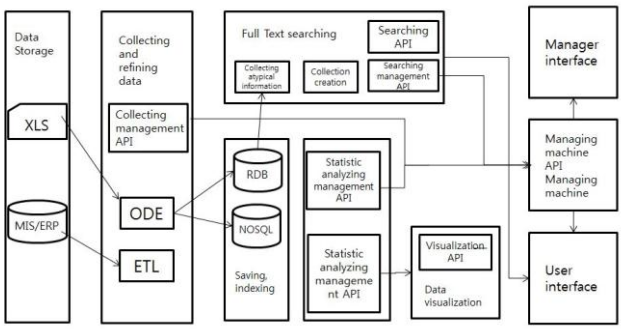

Fig4. System composition diagram

4 Conclusions

There are many examples of small and medium-sized manufacturing enterprises that have not adopted MES. In such cases, it is difficult for companies to accurately check their present conditions due to the fact that most of the main records on site that are written on a manufacturing site are disposed of or are not properly used and maintained. In addition, in comparison with conglomerates small and medium-sized companies obtain and use only fragmentary information in a poor information system environment. There are a large number of inconveniences of common use in a user interface environment that does not consider the field situation.In conclusion, this research has developed a "Big Data analysis system model for on-site use" that collects and saves data from the structural data (typical, atypical, and semi-typical) and internal information system of small and medium-sized manufacturing enterprises. In addition, it provides a base for decision-making support, through monitoring of the main indexes (quality, obstacle, sales and so on), which is refined and analyzed using Big Data technology. In my opinion, the new developed system in research will aid effective decision making, enhance a company's management environment, and accelerate the building of a production system that fits a company's manufacturing environment.

References

[1] M.A Schilling, "Strategic Management of Technological Innovation", 2nd Edition, (2008)

[2] Moon S.J., Jung G.D. andChoi, Y.K., "A Study on Cooperation System design for Business Process based on XMDR in Grid", International Journal of Grid and Distributed Computing, 3, 3, (2010), 1-12.

[3] XianfengY.and Liming L., "A New Data Mining Algorithm based on MapReduce and Hadoop", International Journal of Signal Processing, Image Processing and Pattern Recognition, 7, 2, (2014), 131-142.

[4] J.Dean, S.Ghemawat, "Usenix. MapReduce: Simplified data processing on large clusters", Proc. Operating Systems Design and Implementation, 4, (2005)

[5] Kim J.J., Seong B.M., Lee M.G., Sohn H.J. and Kim, J.B., "On site Big Data analysis system model to promote the competiveness of manufacturing enterprises", Proc. Advanced Science and Technology Letters, 87, (2015), 29-33.

[6] Gartner, "High-Tech Tuesday Webinar: Big Data Opportunities in Vertical Industries", (2012)

[7] Forrester, "The Forrester Wave : Big Data Predictive Analytics Solutions", Q1, (2013)

[8] http://wikibon.org/wiki/v/Big_Data_Market_Size_and_Vendor_Revenues

[9] IDC, "Market Analysis, Worldwide Big Data Technology and Services", (2012)

[10] LeeB.Y., Lim J.T. andYooJ.S., "Utilization of Social Media Analysis using Big Data", KOCON, 13, (2013), 211-219.

[11] Koh J.C., Lee H.U., Jeong J.Y., and Kim K.S. "Correspondence Strategy for Big Data's New Customer Value and Creation of Business", Korea Safety Management & Science, 14, (2012), 229-238.

Received: Month 4, 2015