

빅 데이터의 효율성 시험 평가 방법

양형식*, 김선배**

호서대학교 벤처전문대학원 융합공학과*, 호서대학교 벤처전문대학원 정보경영학과**

Evaluation Method of Big Data Efficiency

Hyeong-Sik Yang*, Sun-Bae Kim**

Dept. of Fusion Engineering, Graduate School of Venture, Hoseo University*

Dept. of Information Management, Graduate School of Venture, Hoseo University**

요 약 최근 소셜미디어, 산업 간 융합 등이 확대되고, 기존의 PC뿐만 아니라 스마트폰, 태블릿 PC 등 다양한 스마트 기기를 통한 인터넷 이용이 증가하면서 수많은 비정형 데이터를 발생됨에 따라 빅 데이터 시스템에 대한 관심이 증가하고 있다. 시장조사 전문기관에서는 향후 5년에는 현재보다 데이터량이 9배가 증가할 것으로 전망하고 있으며, 빅 데이터 시장도 더 커질 것으로 예상하고 있다. 본 논문에서는 빅 데이터가 갖추어야할 효율성 품질평가 항목을 요구사항 분석을 통해서 도출하고 평가항목을 세분화하여 빅 데이터에 대한 효율성 시험 평가를 하였다.

주제어 : 빅 데이터, 효율성, 평가모델, 품질평가

Abstract Recently, integration between social media and the industry has been expanded, and as the usage of Internet through various smart devices of not only the existing PC but also smart phone, tablet PC and so on, a lot of unstructured data has occurred, leading to increased interest on big data system. According to the institutes which specialize in market research, the data amount is predicted to increase by 9 folds in the next 5 years when compared to the present, and the big data market is also expected to grow bigger. This dissertation evaluates the efficiency test of big data through analysis on the requirements by identifying and fragmenting the items of efficiency quality evaluation that big data should be equipped with.

Key Words : Big Data, Efficiency, Evaluation Model, Quality Evaluation

1. 서론

네트워크의 광대역화(무선망 LTE 도입)와 스마트 디바이스의 보급으로 디지털 정보량이 기하급수적으로 증가함에 따라 빅 데이터 관심이 부쩍 증가하고 있다.

그리고 그 많은 데이터를 통하여 새로운 가치 창출이 기업을 포함한 국가의 경쟁력으로 평가되고 있다. 기존과 차별된 빅 데이터 분석과 전망을 통해 새로운 서비스

를 개발할 가능성은 매우 높다고 볼 수 있다.

전 세계 데이터의 60%가 지난 3년간 만들어진 데이터라고 한다. 지난 3년간 소셜 네트워크와 클라우드 컴퓨팅의 폭발적인 성장으로 데이터는 엄청난 속도로 증가하고 있다. 매일 2.5 쿼틸리언(천의 6승) 바이트의 데이터가 생성된다.

다르게 표현하면 매일 2,273,736.75테라바이트 크기의 데이터가 생성되는 것이다. 이전 세대 데이터 스토리지

Received 22 July 2013, Revised 20 August 2013

Accepted 20 August 2013

Corresponding Author: Sun-Bae Kim(Graduate School of Venture, Hoseo University)

Email: hsyang@hoseo.edu

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

기술(일부는 30 년도 넘은)로는 이 같은 수요를 감당할 수 없어 수십억 달러를 들여 새로운 시스템 전반에 직접 투자할 수밖에 없는 처지에 놓이게 되었다. 시장 조사 기관인 IDC 의 예측에 따르면 빅 데이터 시장은 2010 년에 32 억 불, 2015 년에는 169 억 불의 시장으로 성장한다고 한다. 이는 연평균복합성장률(CAGR)로는 40%로 확산되며 전체 ICT 시장 성장률보다 7배나 높다.

이렇게 빅 데이터 시장은 점점 커져가지만 그에 비해 빅 데이터의 품질평가에 대한 평가체계는 잡혀있지 않다.

따라서 본 논문에서는 먼저 빅 데이터의 기술 개요, 특징, 빅 데이터 기술 동향과 시장현황을 분석하고, 분석된 내용을 기반으로 빅 데이터의 효율성 평가를 위해 국제 표준 ISO/IEC 9126[1]을 참조하여 효율성 평가항목을 구축하였다.

본 논문에서는 적용된 결과를 분석하여 빅 데이터 시스템의 효율성을 평가하기 위한 품질평가모델을 구축하고자 한다.

본 논문의 2장에서는 빅 데이터의 관련 동향에 대해 살펴보고 3장에서는 빅 데이터 기반 기술과 효율성에 관한 품질 요구사항을 분석하여 4장에서는 빅 데이터의 효율성에 관한 품질특성에 따른 평가 항목을 구축하였다. 5 장 구축한 항목을 직접 시험 평가하여 시험사례에 대해 정리하였고, 6장에서는 결론과 향후 연구 과제를 제시하였다.

2. 빅 데이터의 관련동향

2.1 빅 데이터

지금은 데이터 빅뱅 시대로 접어들고 있다. 기업의 IT 활용단계는 제 1의 벽의 전자화, 정보화 단계를 넘어 제2의 벽인 데이터 기업의 활용단계로 넘어가고 있다. 데이터 활용 단계는 새로운 이익의 창출에 초점을 맞추고 있어 빅 데이터에 관심이 모아지고 있다.

빅 데이터는 데이터 규모가 방대하고, 데이터 종류가 다양하며, 데이터 분석 및 처리 해결한다.

넓은 의미의 빅 데이터는 3V (Volume, Variety, Velocity)라 하며, 복잡하고 다양한 데이터 모두 저장, 처리 및 분석[3]한다.

2.1.1 크기 (Volume)

여기에서 크기는 데이터의 물리적 크기를 말하며, 기업 데이터, 웹 데이터, 센서 데이터 등 페타데이터 규모로 확장된 데이터이다. 단순 저장되는 물리적 데이터양의 증가뿐만이 아닌 이를 분석 및 처리하는데 어려움이 따르는 네트워크 데이터의 급속한 증가는 빅 데이터의 가장 기본적인 특징이다.

2.1.2 속도 (Velocity)

빅 데이터의 속도적인 특징은 실시간 처리와 장기적인 접근으로 나눌 수가 있다. 인터넷 검색을 할 때도 모든 검색어가 저장이 되고, 쇼핑몰이나 포털사이트 같은 곳을 이용할 때도 우리가 클릭한 이력이 모두 저장된다. 스마트폰에서 SNS나 지도 같은 APP을 이용할 때도 우리의 위치 정보를 남기게 된다. 이와 같이 오늘날 데이터는 매우 빠른 속도로 생성되기 때문에 데이터의 생산, 저장, 유통, 수집, 분석이 실시간으로 처리되어야 한다.

빅 데이터는 데이터의 실시간 처리 및 장기적 접근을 요구한다. 데이터 생산 및 유통, 수집 및 분석 속도의 증가와 이에 대한 실시간 처리 및 장기간에 걸쳐 데이터를 수집, 분석하는 장기적 접근이 데이터의 속도적 특징이다.

2.1.3 다양성 (Variety)

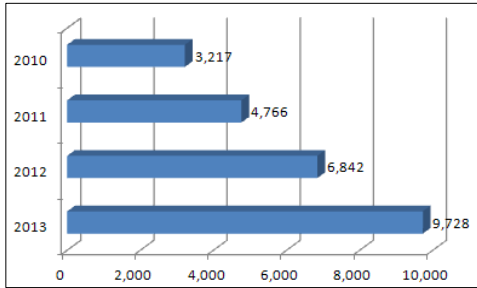
다양한 종류의 데이터들이 빅 데이터를 구성하고 있다. 데이터의 정형화의 종류에 따라서 정형 데이터, 반정형 데이터로 나눌 수 있다. 정형 데이터는 문제 그대로 정형화된 데이터로, 고정된 필드에 저장되는 데이터를 의미한다[4]. 예를 들어 우리가 온라인 쇼핑몰에서 제품을 주문할 때 이름, 주소, 연락처, 배송주소, 결제정보 등을 입력한 후 주문을 하면 데이터베이스에 미리 생성되어 있는 테이블에 저장된다. 이때 테이블은 고정된 필드들로 구성이 되는데, 이렇게 일정한 형식을 갖추고 저장되는 데이터를 정형 데이터라고 한다. 다양성은 데이터가 만들어 내는 정보의 가치를 높게 만드는 것이다.

2.2 시장 동향

IDC(Internet Data Center)에서 빅 데이터를 앞으로 IT의 주요 기술로 예상했으며, 빅 데이터 시장 규모가 2012년 68억 달러에서 2013년은 작년 대비 42% 증가한

97억 달러에 이를 것을 전망하고 있다.

가트너는 빅 데이터 관련 IT 지출 규모를 빅 데이터 관련 기업용 소프트웨어, 소셜 미디어, IT 서비스로 구분해 추정하였고, 2012년 278억 달러 2013년에는 339억 달러로 작년대비 22% 증가할 것으로 전망하고 있다.

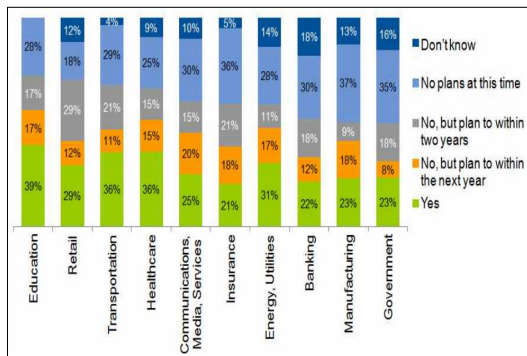


Source : IDC(2012)

[Fig. 1] Big Data in 2013 World Market Prospects

산업별 빅 데이터 관련 투자 현황 및 계획을 살펴보면, 현재는 교육, 운송, 에너지/유틸리티 등의 산업에 빅 데이터 관련 투자가 많은 것으로 조사[5]되었다.

그리고 향후 1~2년 내에 빅 데이터 관련 투자 계획이 높은 산업은 소매, 보험, 통신 및 미디어등으로 나타났다.



Source : Gartner(2012)

[Fig. 2] Industrial by Big Data Relation Investment Present Condition and Plan

2.3 국가별 빅 데이터 동향

주요 선진 국가는 빅 데이터의 중요성을 인식하고, 국가 차원에서 빅 데이터 활성화를 위한 다양한 노력을 기울이고 있다.

2.3.1 미국

미국은 2012년에는 연간 2억 달러 이상을 투입하는 빅 데이터 R&D 이니셔티브를 출범하였다.

<Table 1> Status of Big Data Initiative of United States

Central Department and Provincial Government	Details
National Crime Council (NCC)	<ul style="list-style-type: none"> - Receives data of more than 10,000 cases daily related to terror on SNS, newspaper, magazines and so on and analyzes data - Analyzes terror tendencies and predicts signs of terror in advance
National Science Foundation (NSF)	<ul style="list-style-type: none"> - Core research for data management, analysis, visualization, knowledge extraction and so on, Support for technology development and fostering of human resources - Establish long-term strategy for knowledge extraction methodology using big data - Size of annual investment : UC.Berkeley(10 million Dollars), Workforce fostering institute(2 million Dollars), Research group(1.4 million Dollars) etc
Combined DNA Index System (CODIS)	<ul style="list-style-type: none"> - Analyze information of over 11 million cases on criminals - Utilizes it for tracking suspects and utilized it in arresting 180,000 cases
National Institute of Health (NIH)	<ul style="list-style-type: none"> - Execute '1,000 dielectric project' through partnerships with 75 corporate and institutes - Analyze genetic diversity through genetic data. Cure incurable diseases. development of new medicine - Utilize the citizen's medicine search statistics of Pillbox to predict the distribution of main diseases which are subject to management such as AIDS and the increasing trend per year
Defense Advanced Research Projects Agency (DARPA)	<ul style="list-style-type: none"> - Develop high-performance computing software technology for large-capacity data analysis - Develop HCI technology which can easily be applied to various missions - XDATA project (2,500 annually)
Department of Defense (DoD)	<ul style="list-style-type: none"> - Establish decision making system using data - 250 million Dollars budget allocated with regards to separate big data
Department of Energy (DoE)	<ul style="list-style-type: none"> - Scientific discovery through Advanced Computing - Lawrence Berkely National Research Center (25 million Dollars)
U.S Geological Survey (USGS)	<ul style="list-style-type: none"> - Big data technology for Earth System Science
State of Michigan	<ul style="list-style-type: none"> - Establish government institute's (Integrated Data Warehouse, IDW), Detect illegal public medical insurance actions, Improve service quality such as improving individual health management

이는 고성능 컴퓨터와 인터넷 등의 주도로 확보한 강력한 IT 기술과 빅 데이터 기술을 접목하여 지속적인 경쟁우위를 확보하는데 목적이다.

2.3.2 영국

영국은 가장 투명한 정부를 목표로 오픈 데이터 전략을 추진하고 있다. 기업혁신기술부(BIS)는 공공정보 공개 데이터 가치창출을 위한 자문기구인 데이터 전략위원회 설립하고, 내각 사무처와 함께 데이터 접근성을 강화 및 데이터 공개지침등의 정책을 발표[6]하였다.

영국도 데이터 공개를 위해 'data.gov.uk'의 데이터 공유 플랫폼을 재정비하여 운영하고 있다.

2.3.3 EU(European Union)

EU는 'Faemework Programme 7' 일환으로 금융 위기 극복과 사회의 복잡성을 이해하기 위한 FuturICT 프로젝트 추진하여 개별적이고 단편적인 데이터 분석을 지양하고, 거시적이며, 통계적 의존성이 높은 복잡계 시스템을 기반으로 체계적 연구를 진행하였다. UN은 Global Pulse 프로젝트에서 데이터 연구, 데이터 분석 도구 개발, 연구개발 네트워크 구축 등 할 수 있는 환경을 구축하여 국제적인 개발 정책을 수립하고, 취약계층을 보호하여 글로벌 쇼크에 대한 회복력을 강화하고 있다.

<Table 2> National by data opening to the public present condition

Country	Site Name	data set
USA	data.gov	392,498
Canada	data.gc.ca	13,069
England	data.gov.uk	8,702
Singapore	data.gov.sg	5,000
New Zealand	data.govt.nz	2,261
Germany	poralu.da	1,630
Australia	data.gov.au	1,120
Denmark	data.digitaliser.dk	749
Spain	datos.gob.es	655
Republic of Korea	data.gov.kr	441
Netherlands	data.overheid.nl	351
Finland	suomi.fi	266
Saudi Arabia	saudi.gov.sa	247

2.4 빅 데이터 기술 동향

빅 데이터는 Hadoop와 NoSQL의 성공으로 시장으로 나왔다. 인터넷 업체나 미디어 업체들은 관계형 DBMS 없이 아주 저렴한 비용으로 대용량 데이터를 처리할 수 있는 인프라를 구축할 수 있게 된 건 Hadoop 등장으로 가능하게 되었다.

<Table 3> Big data Technical Element

Item	Details	Open Source
Data Collection	<ul style="list-style-type: none"> Execute features that stores data from data source to stable storage area 	Flume, Scribe, Chukwa
Save Original Data	<ul style="list-style-type: none"> Storage area which stores collected data stably As a non-structural data storage area, it mainly stores large-capacity files 	Hadoop File Sytem (HDFS), mOGILEfs
Save Transaction Data	<ul style="list-style-type: none"> Storage area to save original data real-time and to conduct search Executes simple aggregation calculation such as count, sum and so on rather than complex analysis 	NoSQL (Cloudata,Hbase, Cassandra), Katta
Real-time Analysis Platform	<ul style="list-style-type: none"> Executes data collection and analysis simultaneously 	S4, Storm
Allocation Analysis Platform	<ul style="list-style-type: none"> Complex and various analysis on the entire or partial data Necessary for dispersion, parallel handling for large-capacity handling Supports various analysis model from simple text analysis to graph analysis 	Hadoop MapReduce(Hive, Pig), Giraph, GoldnOrb
Data Mining Statistics Tool	<ul style="list-style-type: none"> Basic algorism library and tool for data mining such as Cluster, Classification and so on 	Mahout, R
Cluster Management and Monitoring	<ul style="list-style-type: none"> As the majority of them are comprised of dispersion system, management and monitoring of the entire cluster is complex too 	ZooKeeper, HUE, Cloumon
Data Serialization	<ul style="list-style-type: none"> As it uses heterogeneous platform and various types of solutions, there is a need for standard framework on sending and handling 	Thrift, Avro, ProtoBuf

Hadoop은 원래 구글이 학계 논문으로 발표한 자사의 인프라 운영에 적용되었던 구글 파일 시스템(GFS)과 분산 데이터베이스인 빅 테이블(Big Table) 기술을 바탕으로 시작되었다. 구글 논문을 바탕으로 시작된 오픈소스 프로젝트로 더그 커팅(Dug Cutting)이 오픈소스 기반의 분산 컴퓨팅 플랫폼을 개발했는데 이게 바로 Hadoop이다. 2007년 야후가 더그 커팅을 영업해 Hadoop 개발에 대한 지원을 하면서 상용화 단계까지 개발이 가능하게 되었다.

3. 빅 데이터의 효율성 요구사항

3.1 시간효율성 품질평가

시간 효율성에 관한 요구사항으로는 다음과 같다.

- 제품 사용시 사용자의 처리 요구에 대한 반응이 나타나기까지의 평균반응 시간은 명시된 규정을 준수하거나 최소화되어야 한다.
- 제품에 주어진 시간 내에 성공적으로 작업을 수행할 수 있는 평균 처리량은 규정된 수준을 유지하여야 한다.
- 제품 사용 중 특정한 업무를 성공적으로 수행하는 평균 처리 시간은 규정된 수준을 유지하여야 한다.

3.2 자원효율성 품질평가

자원 효율성에 관한 요구사항으로는 다음과 같은 항목들이 있다.

- 제품의 I/O자원 사용은 시스템의 전체적인 성능에 영향을 미칠 수 있으므로 가능한 최소화되어야 한다.
- 제품의 메모리 사용 정도는 최소화되어야 한다.
- 제품의 데이터 전송 속도는 규정된 수준을 유지하여야 한다.
- 제품의 CPU 사용은 시스템의 성능에 영향을 미칠 수 있으므로 가능한 최소화되어야 한다.

3.3 준수성 품질평가

준수성에 관한 요구사항으로는 다음과 같은 항목들이 있다.

- 제품이 효율성과 관련된 표준 및 관례에 따라 제품이 구현되어 있어야 한다.

4. 빅 데이터의 효율성 평가항목

빅 데이터 효율성 요구사항과 국제표준 ISO/IEC 9126의 효율성을 참고하여 부특성과 평가항목을 구축하였다. 효율성이란 명시된 조건에서 사용되는 자원의 양에 따라 요구되는 성능을 제공하는 소프트웨어 능력을 의미하며, 효율성에서는 시간효율성, 자원효율성, 성능, 준수성 등의 품질 부특성으로 세분화된다.

4.1 시간효율성 평가항목

시간 효율성이란 명시된 조건에서 그 기능을 수행할 때 적절한 반응 및 처리 시간과 처리율을 제공하는 소프트웨어의 능력을 의미한다.[8] 시간 효율성에는 평균 반응 시간의 적절성, 평균 처리율, 평균 처리시간의 적절성 등의 평가항목을 가진다.

<Table 4> Time utilization Evaluation items

Subcharacteristics	Evaluation items	Purpose of the evaluation items
Time utilization	Appropriateness of average response time	Measures the average response time on the user input when using the product
Time utilization	Average activity ratio	Measures the average processing amount of the product conducting works successfully within the given time
Time utilization	Appropriateness of average process time	Average process time of which a certain work is successfully executed when using the product

4.2 자원 효율성 평가항목

자원 효율성이란 명시된 조건에서 소프트웨어가 그 기능을 수행할 때 적절한 양과 종류의 자원을 사용하는 소프트웨어의 능력을 의미한다.[9] 자원 효율성에는 입출력 자원 사용률, 메모리 사용률, 데이터 전송률, CPU 사용률 등의 평가항목을 가진다.

<Table 5> Resource utilization Evaluation items

Subcharacteristics	Evaluation items	Purpose of the evaluation items
Resource utilization	Activity ratio of input and output resource	Degree of usage of I/O resource
Resource utilization	Memory usage rate	Degree of memory usage
Resource utilization	Data transmission rate	Data transmission speed
Resource utilization	CPU usage rate	Degree of CPU usage

4.3 보안성능 평가항목

보안성능이란 효율성에 관한 특성 평가항목 중에서 시간적인 측면과 자원 사용에 관한 측면을 제외한 보안성 측면에서의 성능 수준을 의미한다.

보안성능은 Throughput, CPS(Connection Per Second, TPS(Transaction Per Second), 전송지연(Latency), Open Connection 등의 평가항목을 가진다.

<Table 6> Security Performance Evaluation items

Subcharacteristics	Evaluation items	Purpose of the evaluation items
Security Performance	Throughput	Measures maximum packet throughput. Measures the maximum traffic of which the equipment can process the packet without packet loss (Express in bps, which refers to the data amount being processed per second, or pps unit which refers to the number of packet being processed per second)
Security Performance	CPS(Connection Per Second)	Measures whether the connection number per second is of appropriate standard
Security Performance	TPS(Transaction Per Second)	Measures whether the transaction per second is of appropriate standard
Security Performance	Latency	Verifies transmission delay in the regulated % standard of handling capacity which the product can process
Security Performance	Open connection	Measures the maximum accumulated HTTP connection number's handling amount which the product can process

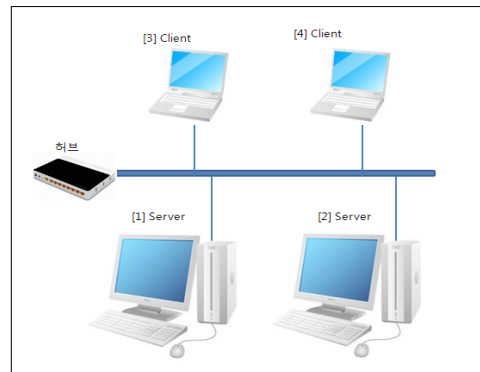
5. 빅 데이터의 효율성 시험 평가

모의 평가 시험은 베타버전인 빅 데이터 시스템 제품군을 시험 평가하였으며, 효율성에 대한 부분에 대해 평가하였다.

5.1 시험환경

평가하기위한 시험 환경을 구축하였다.

[1]서버와 [2]서버에 설치한 프로그램은 시험대상 제품과 서버모듈을 설치하였으며 [3]번 클라이언트와 [4]번 클라이언트에 설치한 프로그램은 시험대상 제품인 클라이언트 모듈과 일반 응용프로그램을 설치하였다.



[Fig. 3] Test Environment

성능측정도구는 [3]클라이언트에 Ldaperf(부하생성) 설치하였으며, [2]번 서버에는 TeamQuest Manager를 설치하였고 [3]번 클라이언트에는 TeamQuest Viewer를 설치하였다.

<Table 7> Detailed list

No	Role	OS	CPU	Mem	HDD
1	Server	Microsoft Windows Server 2008	Xeon Processor 5160 3.0GHz	16GB	500GB
2	Server	Microsoft Windows Server 2008	Xeon Processor 5160 3.0GHz	16GB	500GB
3	Client	Microsoft Windows 8	Intel Core2Duo 2.4GHz	4GB	300GB
4	Client	Microsoft Windows 8	Intel Core2Duo 2.4GHz	4GB	300GB

5.2 시험결과

시험대상을 평가하여 제품을 크게 기능성, 사용성, 신뢰성, 유지보수성, 효율성, 이식성의 주특성 6가지 품질 특성을 중심으로 시험한 후 시험 결과서를 작성하였다.

시험을 실시한 결과는 <Table 8>과 같이 각 특성별 결함수가 집계되었다.

<Table 8> Number of defects

Quality Characteristics	Defects
Functionality	4
Reliability	1
Usability	0
Maintainability	6
Efficiency	0
Portability	1
Total	12

품질 특성별 작성된 결과표를 바탕으로 각 결함에 대해서 결함의 정도를 파악하였다. 결함의 정도는 각각의 기준에 따라서 설정하며 강결함(High), 중결함(Medium), 약결함(Low)로 구별하여 결함을 평가하였다. 시험결과 아래 표와 같이 결함정도별 결함수가 산출되었다.

<Table 9> Degree of fault by the number of defects

Degree of fault	Defects
High	2
Medium	6
Low	4
total	12

결함정도는 기능이 정상적으로 동작하지 않거나, 시스템(HW) 혹은 프로그램이 비정상적으로 종료되는 강결함(H), 프로그램 운영에는 문제가 없으나, 기능이 정확하게 동작하지 않거나 사용자의 혼란을 야기하는 정도의 결함이 발생하는 중결함(M), 프로그램 운영에 문제가 없고 기능도 정확하게 동작하거나 권고 사한 수준의 약결함(L)으로 3단계 결함정도로 나눠 평가하였다.

5.3 성능시험결과

성능시험은 자원효율성과 시간효율성을 다음의 시나리오에 따라 실시하였다.

시나리오 1

- 제품구동 전과 구동 후

시나리오 2

- Monitoring 기능 수행

시나리오 3

- Report 기능 수행

5.3.1 자원 효율성

자원 효율성은 CPU 사용률과 메모리 사용률을 측정하였다.

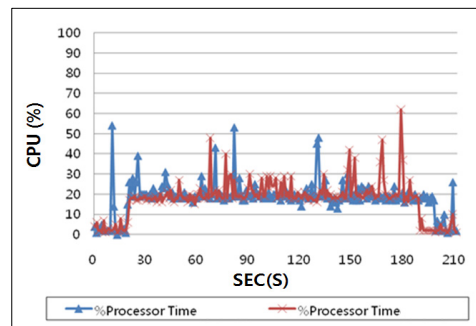
CPU 사용률은 제품 구동 전과 구동 후 [1]서버에서는 각각 54%, 62%까지 증가하였으나, 작업 완료 후 이전 상태로 복귀되었다.

Monitoring 기능 수행 시 [1]서버와 [3]클라이언트의 CPU 사용률은 각각 최대 27%, 21%까지 증가되었으나, 작업 완료 후 이전 상태로 복귀되었다.

Report 기능 수행 시 [1]서버와 [3]클라이언트의 CPU 사용률은 각각 최대 36%, 63%까지 증가되었으나, 작업 완료 후 이전 상태로 복귀되었다.

메모리 사용량은 제품 구동 전과 구동 후 [1]서버의 메모리 사용량은 각각 평균 3150.20MB, 3155.46MB가 사용되었고, Monitoring 기능 수행 시 [1]서버와 [3]클라이언트의 메모리 사용량은 각각 평균 3203.69MB, 724.47MB가 사용되었다.

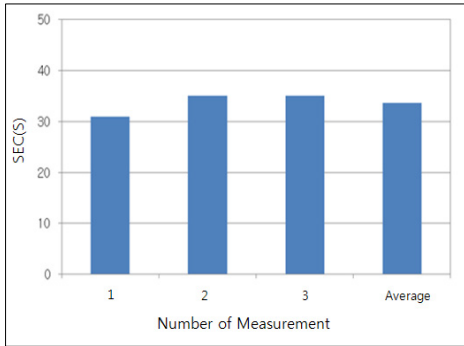
Report 기능 수행 시 [1]서버와 [3]클라이언트의 메모리 사용량은 각각 평균 3204.97MB, 679.96MB가 사용되었다.



[Fig. 4] Server – CPU Activity Ratio

5.3.2 시간 효율성

응답시간은 Report 기능 수행 시 [3]번 클라이언트의 응답 시간은 평균 33.7초가 소요되었다.



[Fig. 5] [3]Client Response Time

6. 결론

현재 휴대인터넷 및 태블릿의 기기의 다양화로 하루에 수많은 데이터가 생산되고 있다. 다시 말하면 스마트 디바이스 시대가 열리면서 일상생활과 함께 인터넷 사용 증가로 거대한 양의 정보가 쏟아지고 있으며, 우리는 빅 데이터 시대에 살고 있다.

이렇게 매일 수많은 데이터를 처리 할 수 있는 빅 데이터의 등장으로 관련된 시장이 점차 커져가고 있으며 향후 3년내에 지금의 빅 데이터 시장 규모의 30% 더 성장 할 것으로 보고 있다.

그러나 빅 데이터 시장은 날로 증가하는데 그에 대한 빅 데이터 품질에 대한 연구가 부족하다. 지금까지 품질 시험에 대한 연구는 패키지 소프트웨어, 산업용 소프트웨어, 임베디드 소프트웨어, 보안용 소프트웨어, 시스템 관리 소프트웨어, 의료용 소프트웨어, 생체인식 소프트웨어, 등 다양한 분야에서 연구되어 왔으며 시험 인증 현장에서 활용하고 있다.

지금 빅 데이터 분야의 품질평가 모델에 대한 연구는 아직까지 미흡한 부분이 있는 실정이다.

빅 데이터 분야에 대한 품질측정 체계가 이루지기 위해서는 먼저 품질을 측정하기 위한 방법과 기준에 대한 연구가 먼저 이루어져야하며, 다양한 소프트웨어 분야를 전반적으로 품질을 특정할 수 있는 수준에 이르기

위해서는 지속적인 연구 개발이 이루어져야 할 것이다.

본 연구에서는 빅 데이터의 효율성에 관한 품질평가 항목을 구축하기 위해 빅 데이터 기술 및 동향을 분석하여 빅 데이터 효율성 요구사항을 도출하였다. 도출된 요구사항과 국제표준 ISO/IEC 9126을 바탕으로 빅 데이터의 효율성 평가항목을 구축하였다.

앞으로 빅 데이터 시장은 계속 증가하고 있는 점을 착안하여 다른 주특성에 대한 평가항목과 측정 방법에 대한 연구가 필요하며, 측정하는 방법에 대한 모색도 필요 하겠다.

향후 지속적인 빅 데이터의 평가사례 구축을 바탕으로 객관성과 타당성을 갖춘 평가체제로 발전시키기 위한 지속적인 연구를 수행할 필요가 있다.

REFERENCES

- [1] ISO/IEC 9126, "Information Technology - Software Quality Characteristics and metrics.
- [2] ISO/IEC 12119, "Information Technology - Software Package - Quality requirement and testing".
- [3] Andrew McAfee, Erik Brynjolfsson, "Big Data: The Management Revolution", Harvard Business School Publishing, Vol. 90 No. 10, pp. 60-68, 2012.
- [4] Steve LaValle, Eric Lesser, Rebecca Shockley, Michael S. Hopkins and Nina Kruschwitz, "Big Data, Analytics and the Path From Insights to Value", SLOAN MANAGEMENT REVIEW, Vol. 52 No. 2, pp. 21-32, 2011.
- [5] Paul C. Zikopoulos, Chris Eaton, et al, "Understanding Big Data - Analytics for Enterprise Class Hadoop and Streaming Data, McGraw Hill, 2012.
- [6] Sherri Rose, "Big data and the future", Wiley Blackwell, Vol. 9 No. 4, pp. 47-48, 2012.
- [7] In-Oh Jeon, Sang-Won Kang, Hae-Sool Yang, "Development of Security Quality Evaluate Basis and Measurement of Intrusion Prevention System", Journal of the Korea academia-industrial cooperation society, Vol 11, No.4, 2010.

- [8] Kyung-Muk Kim, Hae-Sool Yang, "VPN (Virtual Private Network) SW's examination example analysis", Journal of the Korea academia-industrial cooperation society, Vol 11, No.8, 2010.
- [9] Sang-Won Kang, Hae-Sool Yang, "Quality Evaluation of Criterion Construction for Open Source Software", The journal of digital policy & management, Vol 11, No.2, pp. 323-330, 2013.
- [10] Philip Russom, "Big Data Analytic s", TDWI Best Practices report, TDWI Reserarch, 4th Quarter 2011.
- [11] F. Chang, J. Dean, S. Ghemawat, W. C. Hsieh, D. A. Wallach, M. Burrows, T. Chandra, A. Fikes, R. E. Gruber, "Bigtable: A Distributed Storage System for Structured Data", Google, Inc.

양 형 식(Hyeong-Sik Yang)



- 2008년 2월 : 호서대학교 컴퓨터공학과 졸업(공학사)
- 2011년 2월 : 건국대학교 컴퓨터공학과 졸업(공학석사)
- 2012년 2월 ~ 현재 : 호서대학교 벤처전문대학원 융합공학과 박사과정 재학 중

- 2011년 3월 ~ 현재 : ALTIBASE(주) 기술서비스본부 재직중
- 관심분야 : 데이터베이스 시스템, DB모델링, SQL튜닝, DB 최적화, 분산데이터베이스, GIS, 운영체제

김 선 배(Kim Sun Bae)



- 1973년 3월: 연세대학교 경영학과 (경영학사)
- 1991년 5월 : 美國 뉴욕대 경영대학원 (MBA)
- 2006년 3월 : 건국대 컴퓨터정보통신공학 (공학박사)
- 1993년 9월 ~ 2004년 3월 : 현대정보기술 대표이사사장

- 2005년 6월 ~ 2006년 12월 : 한국정보통신수출진흥센터 원장
- 2007년 1월 ~ 2009년 1월 : 정보통신국제협력진흥원 원장
- 2009년 3월 ~ 현재 : 호서대학교 교수
- 관심분야 : 정보통신, 인터넷비즈니스, 소셜미디어
- E-Mail : sunbkim@gmail.com