

중소기업을 위한 하둡 클러스터의 프로토타입과 분석 소프트웨어의 통합된 검증

Integrated Verification of Hadoop Cluster Prototypes and Analysis Software for SMB

차병래¹ · 김남호² · 이성호³ · 지유강⁴ · 김종원^{1*}

¹광주과학기술원 정보통신공학부

²호남대학교 인터넷소프트웨어학과

³(주)UC 연구소

⁴휴인테크

Byung-Rae Cha¹ · Nam-Ho Kim² · Seong-Ho Lee³ · Jong-Won Kim^{1*}

¹School of Information and Communications, GIST, GwangJu, 500-712, Korea

²Dept. of Internet Software, Honam Univ., GwangJu, 506-714, Korea

³Research Center of UC Ltd., GwangJu, 500-712, Korea

⁴HUIN-Tech Co., GwangJu, 500-712, Korea

[요 약]

최근 IT 분야의 화두인 클라우드 컴퓨팅과 빅데이터 패러다임을 중소기업(Small and Medium Business: SMB) 차원에서 용이하게 활용하도록 지원하는 시도가 증가하고 있다. 이러한 노력의 일환으로, 본 논문에서는 프라이빗 클라우드 인프라 환경을 대상으로 하둡(Hadoop) 클러스터를 시험적으로 구축하는 프로토타입을 설계하고 구현한다. 프로토타입 구현은 싱글보드, PC, 그리고 서버를 이용하여 각각 수행하고, 그 성능을 테스트한다. 또한, ASA (American Standard Association) Dataset을 이용한 빅데이터 분석을 통해서 구축된 하둡 프로토타입을 활용하는 분석 소프트웨어 시스템의 성능을 통합적으로 검증한 결과를 제시한다. 이를 위해, R, 파이썬, D3, 자바와 같은 오픈소스를 이용하여 분석 소프트웨어 시스템을 구현하고, 테스트를 수행한다.

[Abstract]

Recently, researches to facilitate utilization by small and medium business (SMB) of cloud computing and big data paradigm, which is the booming adoption of IT area, has been on the increase. As one of these efforts, in this paper, we design and implement the prototype to tentatively build up Hadoop cluster under private cloud infrastructure environments. Prototype implementation are made on each hardware type such as single board, PC, and server and performance is measured. Also, we present the integrated verification results for the data analysis performance of the analysis software system running on top of realized prototypes by employing ASA (American Standard Association) Dataset. For this, we implement the analysis software system using several open sources such as R, Python, D3, and java and perform a test.

Key word : Big Data and Cloud, Private and personal cloud, Hadoop cluster, Prototyping and verification.

<http://dx.doi.org/10.12673/jant.2014.18.2.191>



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.

Received 27 February 2014; Revised 18 April 2014

Accepted (Publication) 21 March 2014(30 April 2014)

*Corresponding Author; Jong-Won Kim

Tel: +82-70-8269-0591

E-mail: jongwon@gist.ac.kr

I. 서론

클라우드 컴퓨팅과 빅데이터는 2010년부터 가트너 그룹[1]이 선정한 10대 전략기술이며, 퍼스널 클라우드를 2013년에 새롭게 추가하면서 (그림 1 참조) 점차적으로 세부화되고 있다. 이러한 상황을 감안하여 중소기업(SMB)의 빅데이터 처리에 적합한 수준으로 클라우드 인프라에 대응하는 Hadoop 클러스터와 이를 운영하여 빅데이터를 분석하는 해석 소프트웨어 시스템이 필요하다. 특히 중소기업 대상으로 진행된 ENISA[2]의 2009년도 설문조사 질의 3번에서는 중소기업이 클라우드 컴퓨팅을 채용하는 이유를 크게 세 가지를 언급하고 있다. 첫 번째는 인프라/플랫폼/서비스를 아웃소싱할 때 겪는 정보 보안, IT 지원, S/W, H/W의 자본 지출을 피할 수 있고, 두 번째는 IT 자원의 융통성과 확장성, 그리고 세 번째는 비즈니스의 연속성과 재난 극복 능력 때문에 클라우드 컴퓨팅을 채용한다는 것이다 [3].

본 논문에서 이러한 점들을 감안하여 스케일 아웃방식으로 하드웨어를 추가할 수 있는 프라이빗 클라우드 방식으로 하둡 클러스터 하드웨어를 구성한다. 이를 통하여 경제적으로 하둡 기반의 다양한 빅데이터 업무에 적용이 가능한 고가용성을 지원하면서, 다양한 오픈소스 Hadoop, R, Python, D3 소프트웨어를 이용한 빅데이터의 분석을 위한 프로토타입 시스템을 제안한다. 즉 빅데이터를 Hadoop으로 처리한 후에 R에 의한 통계분석을 처리하고, Python에 의한 다양한 데이터 가공/변환을 지원하면서 D3에 의한 가공된 데이터의 시각화를 연결한다. 논문의 구성은 2장에서 하둡, 빅데이터, 프라이빗과 퍼블릭 클라우드, 그리고 기존에 개발된 프로토타입에 관한 관련 내용을 서술하였다. 이어서 3장에서는 제안한 하둡 클러스터 프로토타입을 설계하고 부분적으로 구현하여 테스트한 결과를 설명한다. 4장에서는 구축된 프로토타입 하드웨어와 분석 소프트웨어를 통합하는 빅데이터 분석 성능을 검증하고 5장에서 결론을 맺는다.

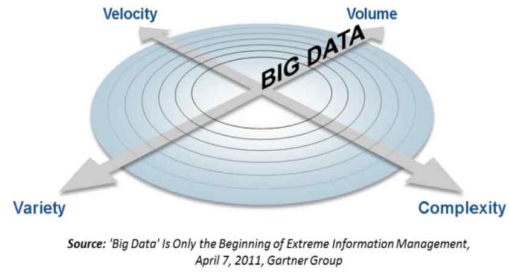


그림 2. 빅데이터의 3V+1C 정의 [1]

Fig. 2. Definition 3V+1C of big data [1].

II. 관련 연구NH

2-1 빅데이터 관련 동향

빅 데이터 (big data)란 통상 일반적인 데이터베이스, 소프트웨어로 관리가 어려운 대용량의 데이터를 의미하며, 기존 데이터베이스 처리 방식의 데이터 수집, 저장, 관리, 분석 역량을 넘어서는 데이터 셋이라고 정의된다[4]. 특징으로는 기존의 데이터 단위를 넘어서는 엄청난 양(volume), 데이터의 생성과 흐름이 매우 빠르게 진행되는 속도(velocity), 사진, 동영상 등 기존의 구조화된 데이터가 아닌 다양한(variety) 형태의 정보 등 3가지 속성이 있다[5, 6]. 이러한 빅데이터의 특징을 가트너 그룹은 그림 2와 같이 3V+1C 로 나타내며, IBM은 complexity 대신에 veracity를 사용하여 4V로 설명한다[7].

그림 3과 같이 빅데이터에서 다루는 정보들은 형식이 정해져 있는 텍스트 위주의 데이터로부터 그림, 동영상, 음성 위주의 형식이 정해져 있지 않은 비정형 데이터들이 주류를 이루고 있

	2010	2011	2012	2013	2014
1	클라우드 컴퓨팅	클라우드 컴퓨팅	미디어 태블릿 그 이후	모바일 대전	다양한 모바일 기기 관리
2	진보적 분석	모바일 앱과 미디어 태블릿	모바일 중심 애플리케이션과 인터페이스	모바일 앱 & HTML 5	모바일 앱과 애플리케이션
3	클라우드 컴퓨팅	소셜 커뮤니케이션 및 협업	상황인식과 소셜이 결합된 사용자 경험	퍼스널 클라우드	만들 인터넷
4	그림을 위한 IT	비디오	사물 인터넷	사물 인터넷	하이브리드 클라우드와 서비스 브로카의 IT
5	데이터센터 재설계	차세대 분석	앱스토어와 마켓 플레이스	하이브리드 IT & 클라우드 컴퓨팅	클라우드/클라우드/클라우드 아키텍처
6	소셜 SW & 소셜 컴퓨팅	소셜 분석	차세대 분석	전략적 빅데이터	퍼스널 클라우드의 시대
7	사용자 액티비티 모니터링(보안)	상황인식 컴퓨팅	빅데이터	실용 분석	소프트웨어 정의
8	플래시 메모리	스토리지급 메모리	인메모리 컴퓨팅	인메모리 컴퓨팅	클-스케일 IT
9	가용성을 위한 가상화	유비쿼터스 컴퓨팅	저전력 서버	통합 생태계	스마트 머신
10	모바일 애플리케이션	패브릭 기반 컴퓨팅 및 인프라스트럭처	클라우드 컴퓨팅	엔터프라이즈 앱스토어	3-D 프린팅

그림 1. 2010년부터 2014년까지 가트너 그룹의 10대 전략기술

Fig. 1. 10 strategic techniques of Gartner Group to 2014 from 2010.

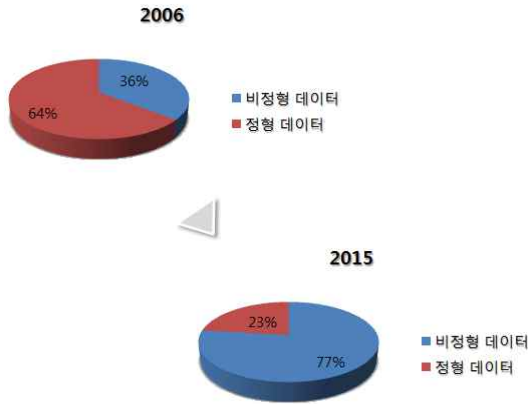


그림 3. 정형/비정형 데이터의 변화 추이
Fig. 3. Change of fixed/unfixed data.

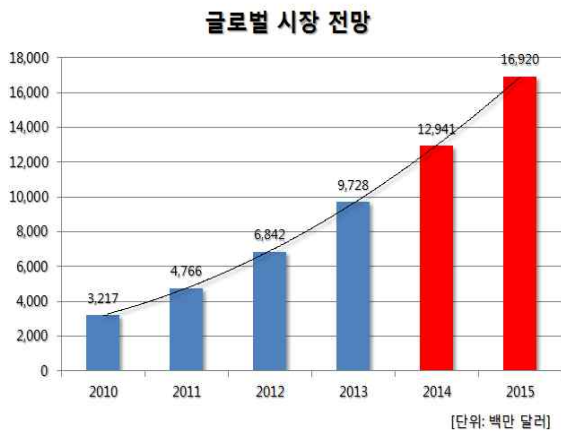


그림 4. 글로벌 빅데이터 시장 전망
Fig. 4. View of global big data market.

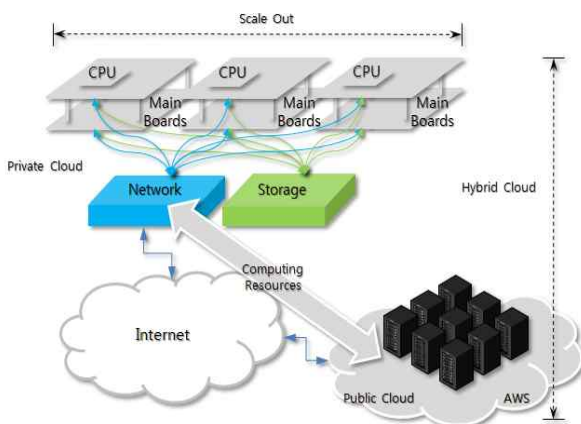


그림 5. 프라이빗 클라우드 클러스터의 스케일 아웃과 하이브리드 클라우드 연동 지원
Fig. 5. Scale out and hybrid cloud extension of private cloud cluster.

다. 이로 인해 텍스트 중심의 정형 데이터를 처리했던 기존 도구로는 수백억 건의 비정형 데이터에 대한 수집, 저장, 검색, 분석, 시각화 등이 어렵게 되었다. 또한 그림 4와 같이 빅데이터 시장의 확대가 예상되므로 빅데이터를 위한 분석기술의 확보를 통하여 다가오는 빅데이터 시대에 효과적으로 대응하는 것이 필요하다.

2-2 Public Cloud와 Private Cloud

빅데이터 분석을 해결하기 위한 대표적인 인프라 확보 방안으로 클라우드 컴퓨팅 기술이 사용된다[8]. 빅데이터 분석을 위해 IaaS (infrastructure as a service) 형태로 인프라 자원을 확보하여 하둡을 연동하는 PaaS (platform as a service) 수준으로 연결하는 방식이 널리 활용되는 추세이다.

이러한 클라우드 컴퓨팅은 배치 방식에 따라 공용(public)과 프라이빗(private) 방식으로 분류된다. 일반에게 공개하여 서비스하는 공용 클라우드 컴퓨팅은 유틸리티 컴퓨팅 형태로 사용량에 따른 과금을 통해 사용 목적에 따른 서비스 이용의 탄력성을 최대화 할 수 있다. 하지만 반복되는 불확정된 비용 지불 방식에 따른 번거로움이 발생하며, 서비스를 독자적으로 심화시키기 위한 지원 비용이 증가할 수도 있다. 또한 서비스가 어디서, 어떻게 제공되는 지에 대한 정보가 제한되므로 서비스 이용에 대한 통제 권한이 부족한 문제점이 있다. Amazon Web Service, Google Cloud, Microsoft Azure, Salesforce 등이 주도적으로 서비스를 제공하고 있으며 Carolyn Purcell & David Floyer에 따르면 매출 10억달러 이하 기업에서 이용하는 것이 적절하다.

한편 프라이빗 클라우드 컴퓨팅은 기업 내부의 클라우드 컴퓨팅 데이터 센터를 중심으로 클라우드 컴퓨팅 환경을 구성하여 내부 고객에게 서비스를 제공하는 방식으로 구성원 개개인의 시스템에 대한 관리 부담이 적다. 특정 임무 중심의 애플리케이션 구성이 일반적이기 때문에 기업 입장에서 자료를 통합하여 관리하기 용이하며 전체 인프라에 대한 통제권을 가질 수 있다는 장점이 있다. 인프라에 대한 통제권을 가질 수 있기 때문에 보안 및 신뢰성이 제고되며 네트워크 대역폭의 제약이 줄어들며 서비스 수준 관리(SLA: service level agreement)가 가능하다. 그에 반해 사용량에 따른 비용을 정산할 수는 없다는 단점이 있으며 별도의 구축비용이 발생할 수 있다. 장비, 하드웨어, 가상화 기술에 대한 비용이 발생하며 별도의 데이터 센터 구축비용과 높은 인력 비용이 예상되며 탄력성이 비교적 낮게 나타난다. IBM, HP, VMware, EMC 등 대형 벤더사에서 서비스 제공이 가능하며 Carolyn Purcell & David Floyer에 따르면 매출이 10억 달러 이상 되는 기업에서 구축하는 것이 유리하다[9].

2-3 중소기업에 위한 시험적인 프라이빗 클라우드 인프라



그림 6. 프로토타입 0.1의 구현
 Fig. 6. Implementation of prototype 0.1.

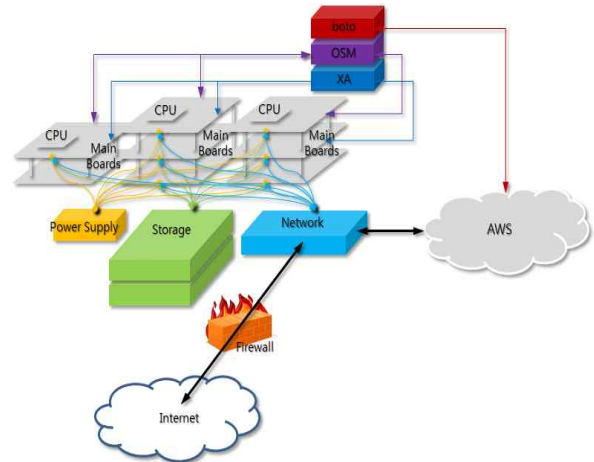


그림 9. 프로토타입 1.0 설계
 Fig. 9. Design of prototype 1.0.



그림 7. 프로토타입 0.2의 구현
 Fig. 7. Implementation of prototype 0.2.

중소기업이 경제적으로 클라우드 기반 빅데이터 분석을 이해하도록 지원하고자 하는 목적으로 구성하는 시험적인 규모의 프라이빗 클라우드는 블레이드 서버 기술, 가상화 기술, 보안, 고가용성, 오픈 소스 관리, 확장성 등의 기능들이 필요하다. 즉 그림 5에 제시한 바와 같이 아래에 기술한 관련 이슈들을 감안 하면서 또한 다양한 업무에 적용이 가능하도록 컴퓨팅 및 스토리지 저장 측면에서 고가용성과 하이브리드 클라우드 확장성을 제공하여야 한다[10].

- 블레이드 서버와 가상화에 의한 컴퓨팅 자원의 확장성 (scalability)
- 보안 (security) 및 고가용성 (high availability)
- 오픈소스 소프트웨어 지원

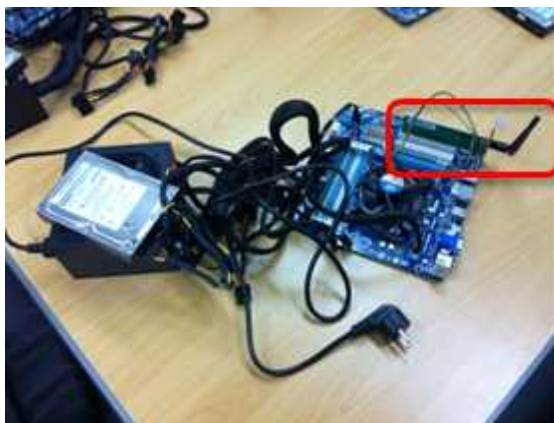


그림 8. 프로토타입 0.5의 HA 기능 구현
 Fig. 8. HA function implementation of prototype.

중소기업을 위한 시험적인 프라이빗 클라우드를 구성하기 위해서 ENISA[2]의 보고서에서 제안된 첫 번째는 인프라/플랫폼/서비스를 아웃소싱할 때 겪는 정보 보안, IT 지원, S/W, H/W의 자본 지출 문제의 최소화하고, 두 번째는 IT 자원의 융통성과 확장성을 지원할 수 있는 하둡 클러스터 프로토타입을 활용한다.

이러한 사전 연구로 하둡 클러스터의 프로토타입들을 설계 하였으며, 구축된 하둡 클러스터의 프로토타입의 성능을 테스트하였다. 그림 6과 같이 프로토타입 0.1은 PC 4대에 name node 1대와 data node 3대를 구성 및 Hadoop 설치를 하여 테스트를 완료하였으며, 그림 7의 프로토타입 0.2은 인텔 i3 CPU, RAM 4 GB, 하드디스크 320 GB로 구성된 소형 메인보드를 4개를 공유기로 연결하여 NameNode와 DataNode를 구성하였다.

프로토타입 0.2 형태의 구축된 Hadoop 시스템에 보안을 위한 방화벽과 네트워크의 고가용성(XA), 그리고 오픈 소프트웨어 관리를 위한 OSM(open source management)을 추가하여 프로토타입 0.5을 설계하였으며, 그림 8과 같이 나타내었다.

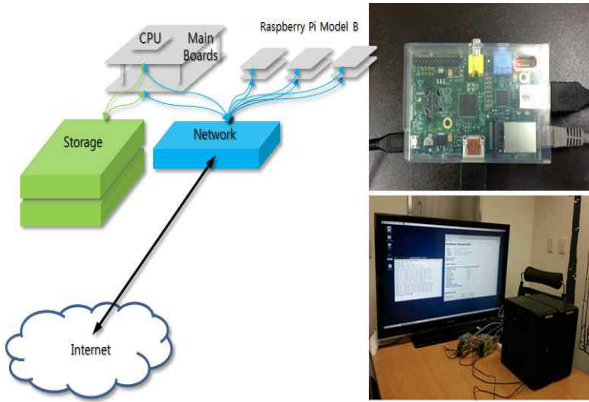


그림 10. Type SB-1의 프로토타입의 설계 및 구현
 Fig. 10. Design and implementation of prototype Type B-1.

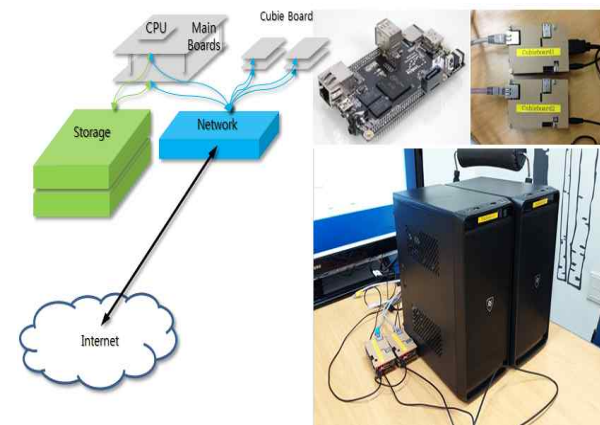


그림 11. Type SB-2의 프로토타입의 설계 및 구현
 Fig. 11. Design and Implementation of prototype Type SB-2.

표 1. Type SB-1의 스펙과 테스트 결과
 Table 1. Spec.and test results of Type SB-1.

항목	스펙	갯수	성능
CPU	Intel i3, 3.3 GHz, Dual Core	2	7분 27초
	ARM 1176JZ-F, 700 MHz	3~6	
Memory	4G	2	
	512 MB SDRAM	3	
Disk	320 GB	2	
	4 GB	3	
Network	NetGear SafePro, 20 포트 스위칭 허브	1	

표 2. Type SB-2 스펙과 테스트 결과
 Table 2. Spec.and test results of Type SB-2.

항목	스펙	갯수	성능
CPU	Intel i3, 3.3 GHz, Dual Core	2	16분 52초
	1 GHz ARM	4	
Memory	4 G	2	
	1 GB	2~4	
Disk	320 GB	2	
	4 GB	4	
Network	NetGear SafePro, 20 포트 스위칭 허브	1	

표 3. Type S의 스펙과 테스트 결과
 Table 3. Spec.and test results of Type S.

항목	스펙	갯수	성능
CPU	Intel Xeon E5-2640, 2.0 GHz	1	3분 44초
Memory	16 G	1	
Disk	200 G	1	
Network	NetGear SafePro, 20 포트 스위칭 허브	1	

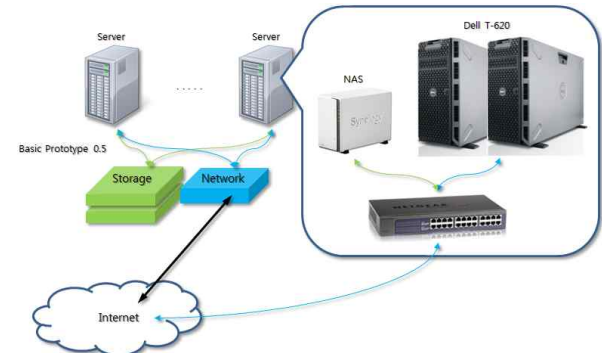


그림 12. Type S의 프로토타입의 설계 및 구현
 Fig. 12. Design and implementation of prototype Type S.

프로토타입 0.2에서 고가용성 HA의 결함 감내 기능을 위하여 무선랜을 추가한 그림이다. 유선과 무선에 의한 네트워크 이중화에 의한 결함 감내 및 네트워크 속도 향상의 효과를 갖게 된다.

프로토타입 0.5에 scalability를 제공하므로써 프로토타입 1.0으로 프로토타입을 업그레이드하였다. scalability를 제공하기 위한 방법으로 boto[11] 등의 오픈 소스에 의한 AWS 등의 public cloud에 인프라 기반 확장성을 부여함으로써 public cloud의 자원을 공유가 가능하도록 private cloud인 프로토타입 1.0을 그림 9와 같이 설계하였다.

III. 하둡 인프라의 프로토타입 설계 및 구현

빅데이터의 다양한 요구 사항을 수용하기 위하여 설계되었으며, single 보드와 PC, 그리고 서버를 이용한 각각의 하드웨어 프로토타입들을 구현 및 테스트를 수행하였다.

3-1 Single Board 기반의 인프라

그림 10은 Type SB(single board)-1의 설계와 라즈베리 파이 모델 B에 의한 구현된 결과를 보여주며, 오픈 소스 하둡을 설치하였고, ASA에서 공개한 미국 항공편 운항 통계 데이터 11 GB를 이용하여 테스트를 수행하였다. 표 1은 Type SB-1의 사양과 테스트 결과 시간을 나타낸 것이다. 또한 빅데이터를 분석하고 그 결과를 시각화하여 보여주기 위하여 오픈 소스 R을 라즈베리 파이 모델 B에 직접 설치 및 테스트를 완료하였다.

그림 11은 Type SB-2 설계와 Cubie 보드에 의한 구현된 결과를 보여주며, 오픈 소스 하둡을 설치 및 ASA의 11GB 데이터를 이용하여 테스트를 수행하였다. 표 2는 Type SB-2의 사양과 테스트 결과를 나타낸 것이다.

3-2 서버 기반의 인프라와 테스트 결과의 성능 분석

그림 12는 서버들과 NAS를 이용한 Type S 설계 및 구현 환경을 나타내며, 오픈 소스 하둡을 설치하고, 데이터 11 GB를 이용하여 테스트를 수행하였다. 표 3는 Type S의 사양과 테스트 결과를 나타낸 것이다.

프로토타입 Type SB와 프로토타입 Type S의 하드웨어 프로토타입을 이용하여 하둡을 테스트를 수행하였으며, 그림 13 ~ 그림 16은 테스트 결과를 차트로 나타낸 것이다. 그림 15는 프로토타입 Type SB-2의 DataNode를 단계적으로 추가하여 성능을 비교하였다. 그림 16은 서버 기반 프로토타입 Type S는 가상화 기술을 이용하여 DataNode를 단계적으로 추가하여 성능을 측정된 결과를 나타내고 있다.

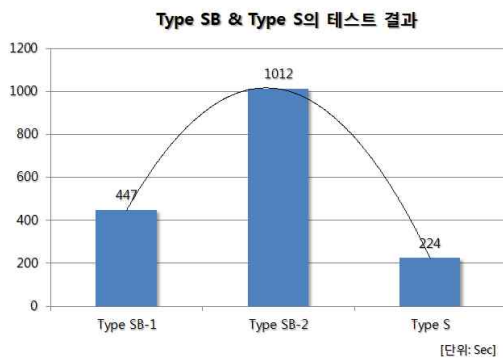


그림 13. Type SB와 Type S의 테스트 결과
Fig. 13. Test results of Type SB and Type S.

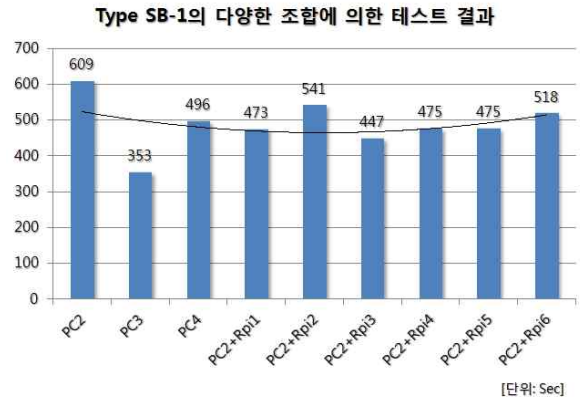


그림 14. Type SB-1의 다양한 조합에 의한 테스트 결과
Fig. 14. Test results of various combination of Type SB-1.

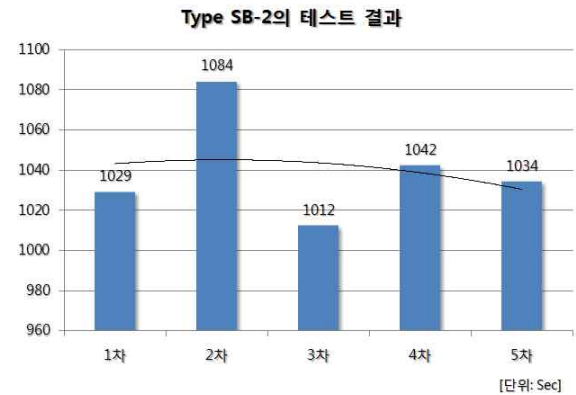


그림 15. Type SB-2의 테스트 결과
Fig. 15. Test results of Type SB-2.

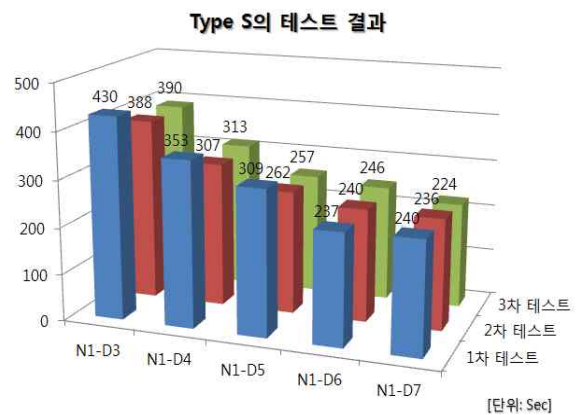


그림 16. Type S의 테스트 결과
Fig. 16. Test results of Type S.

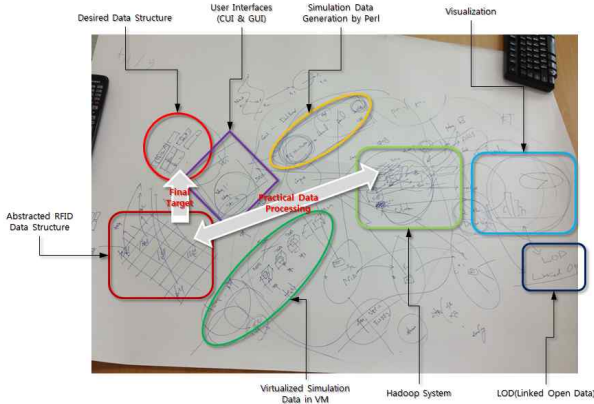


그림 17. 시스템 분석 및 모듈 설계
Fig. 17. System analysis and module design.

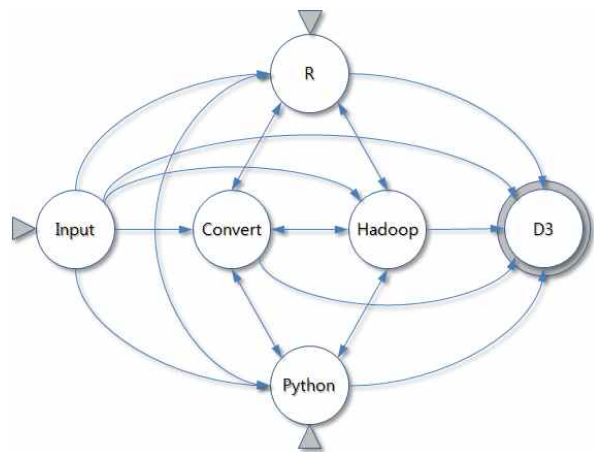


그림 18. 상태 다이어그램
Fig. 18. State Diagram.

IV. 빅데이터 분석 시스템의 설계 및 구현

4-1 시스템 설계

빅데이터 분석을 위하여 하둡 기반의 시스템 전반적인 분석과 모듈을 그림 17과 같이 진행하였으며, 시스템 개발에 사용된 오픈 소스는 하둡, R, Python, D3 등과 JSP, 그리고 Perl, 자바 스크립트와 자바 언어가 시스템 개발에 사용되었다. 그림 18은 설계된 시스템의 운영을 위한 상태 다이어그램을 나타낸 것이다.

4-2 구현 결과

프로토타입 Type S와 프로토타입 0.2의 하드웨어 프로토타입을 기준으로 오픈 소스 하둡과 R, Python, D3 등을 설치 및 통합하여 시스템을 개발하였다. 개발된 시스템은 그림 18의 상태 전이도의 Input 항목, Convert 항목, Hadoop 항목, 그리고 D3 항목이

목이 웹브라우저에 의하여 통합되었으며, R 항목과 Python 항목은 자체적으로 운영되게 구성하였다. 그림 19는 그림 18의 상태 전이도의 Input 항목에 해당하는 UI를 보여주고 있으며, 네트워크를 통한 파일 전송과 로컬 파일 시스템에서 데이터를 가져오게 설계 및 구현되었으며, 그림 19의 오른쪽 부분은 데이터의 상태와 리스트를 확인할 수 있도록 구현되었다.

그림 20은 상태 전이도의 Convert 항목에 해당하는 UI를 나타내고 있으며, 이 Convert 기능을 이용하여 각각의 오픈 소스들 간의 다양한 데이터 타입의 호환성을 제공할 수 있게 된다. 예를 들어, 웹 사이트에 공개된 데이터를 네트워크를 통해 텍스트 파일로 전송 받았다면, Convert 기능을 이용하여 JSON 자료로 변환 후에 곧바로 웹브라우저에서 D3로 시각화가 가능하게 된다. 또한 JSON 자료를 이용하여 R 기반 통계 처리를 위하여 태그를 제거하고 변환하는 기능과 기존 데이터를 Python 도구에 의한 다른 데이터 가공을 위한 전처리 기능을 제공하게 된다.

그림 20은 Input 항목 또는 Convert 항목에서 전달된 데이터를 이용하여 하둡에 의한 MapReduce 처리를 수행하게 된다. 하둡의 성능은 사전에 설계한 프로토타입 Type SB와 Type S의 하드웨어 성능에 따라 달라진다. 그림 21은 하둡에 의한 MapReduce 결과를 웹브라우저 화면에 보여주고 있다. 그림 21의 왼쪽 상단은 하둡을 위한 가공되기 전의 데이터를 보여주며, 그림 21의 하단에서 가공된 데이터를 선택하면 그림 21의 오른쪽 화면에 결과를 보여주게 된다. 그림 22는 빅데이터를 위한 하둡의 맵리듀스 처리 결과를 나타내는 화면이다.

그림 23은 통계 처리를 위한 R 도구와 하둡 이외의 다른 데이터 가공을 위한 Python 도구의 구동 및 결과물을 보여주는 UI이다. 그림 24와 그림 25는 시각화 도구인 D3를 이용한 웹브라우저에서의 인포그래픽스를 나타내는 UI이다.

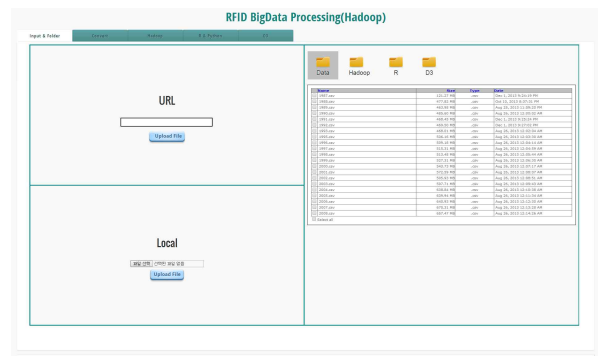


그림 19. 빅데이터 처리를 위한 Input & Folder의 UI
Fig. 19. UI of input & folder for BigData processing.

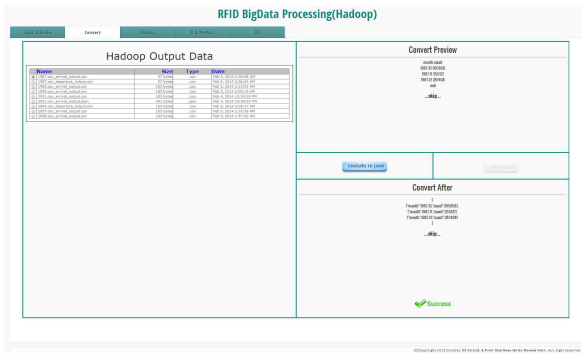


그림 20. 빅데이터 처리를 위한 Convert UI
 Fig. 20. Convert UI for big data processing.

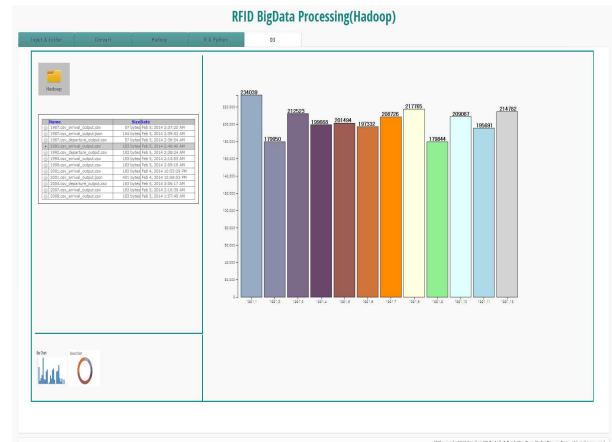


그림 24. 시각화 도구인 D3의 UI #1
 Fig. 24. UI #1 of visualization tool D3.

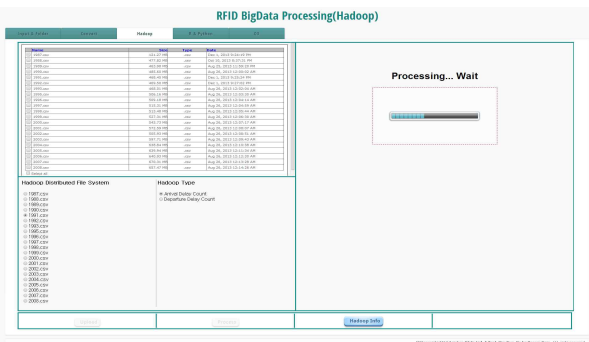


그림 21. 하둠 처리 과정의 UI
 Fig. 21. UI of Hadoop processing.

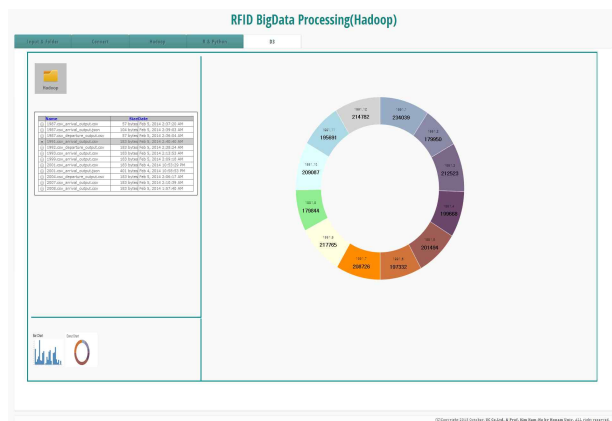


그림 25. 시각화 도구인 D3의 UI #2
 Fig. 25. UI #2 of visualization tool D3.

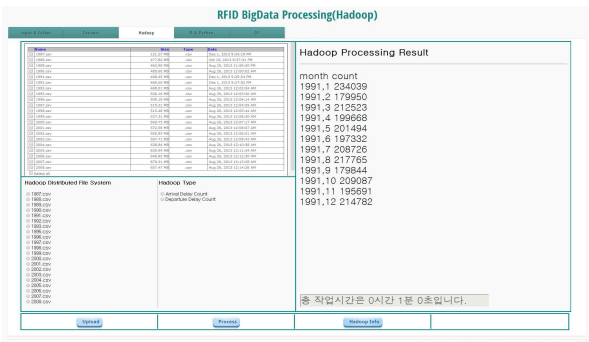


그림 22. 하둠 결과의 UI
 Fig. 22. UI of Hadoop result.

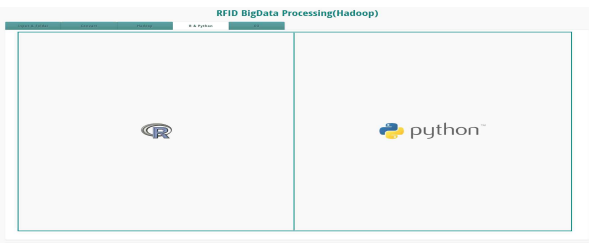


그림 23. R과 Python의 UI
 Fig. 23. UI of R and Python.

V. 결론

최근 IT 분야에서는 빅데이터가 이슈화 되었으며, 2012년부터 가트너 그룹의 10대 항목 중에 빅데이터가 포함되기 시작하였으며, 더불어 클라우드 컴퓨팅 기술이 세분화되고 있음을 보여주고 있다. 본 연구는 중소기업을 위한 프라이빗 클라우드 인프라를 구축하기 위한 싱글보드, PC, 그리고 서버 기반의 하드웨어 인프라 프로토타입을 정의하였으며, 이를 바탕으로 하둠의 성능을 테스트 하였다. 다양한 하드웨어 인프라와 오픈 소스 소프트웨어 Hadoop, R, D3, Python, 등을 이용하여 빅데이터의 분석과 시각화를 위한 시스템을 설계 및 구현하였다. 향후에는 이러한 인프라를 기반으로 Asterisk와 CCNx 등과 같은 다양한 애플리케이션들의 서비스를 테스트 및 지원할 예정이다.

감사의 글

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업(2012R1A1A2041274)과 중소기업청의 창업성장 기술개발과제사업(S2111889, "Development of Dedicated SIP Server based on Virtualization for Voice-OTP Mobile App.")을 지원받아 수행된 것임.

참고문헌

[1] Gartner Group [Internet]. Available: <http://www.gartner.com/>
 [2] ENISA [Internet]. Available: <http://www.enisa.europa.eu>
 [3] ENISA Survey, "An SME perspective on Cloud Computing," 2009 [Online]. Available: <http://www.enisa.europa.eu/activities/risk-management/files/deliverables/cloud-computing-sme-survey>
 [4] J. Manyika and M. Chui, "Big data: the next frontier for innovation, competition, and productivity," *McKinsey Global Institute*, May 2011.

[5] P. Russom, "Big data analytics", *TDWI Research Fourth Quarter*, p.6, 2011.
 [6] Big Data: Expanding on 3 fronts at an increasing rate [Internet]. available: <http://itknowledgeexchange.techtarget.com/writing-for-business/files/2013/02/BigData.001.jpg>
 [7] The Big Data & Analytics Hub [Internet] Available: <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>
 [8] S. Makoto, "The impact of cloud computing," *Jpub Press*, 2009.
 [9] R. Craif, J. Frazier, N. Jacknis, S. Murphy, C. Purcell, P. Spencer and JD Stanley, "Cloud computing in the public sector", 2009, USA.
 [10] B. R. Cha, H. G. Kim, D. G. Kim, J. W. Kim, and Y. I. Kim, "Basic prototype design and verification of Hadoop cluster based on private cloud infrastructure for SMB," *Journal of Advanced Navigation Technology*, Vol. 17, No. 2, pp. 225-233, Apr. 2013.
 [11] boto [Internet], Available: <https://github.com/boto/boto>



차 병 래 (Byung-Rae Cha)

2004년 2월 : 국립 목포대학교 컴퓨터 공학과(공학박사)
 2005년 3월 ~ 2009년 2월 : 호남대학교 컴퓨터공학과 전임강사
 2009년 9월 ~ 현재 : 광주과학기술원(GIST), 정보통신공학부 연구조교수
 2012년 5월 ~ 현재 : (주)UC 대표이사
 ※관심분야 : 정보보안, Intrusion Detection System, 신경망, 클라우드 컴퓨팅, Future Internet 등



김 남 호 (Nam-Ho Kim)

1997년 8월 : 포항공과대학교 정보통신학과(공학석사)
 2000년 8월 : 전남대학교 전산통계학과 (박사수료)
 1991년 4월 ~ 1998년 2월 : 포스데이타(주)
 1998년 3월 ~ 현재 : 호남대학교 인터넷콘텐츠학과 부교수
 ※관심분야 : 데이터마이닝, 유비쿼터스 컴퓨팅, 가상현실 응용, 생체인증 등



이 성 호 (Seong-Ho Lee)

1999년 2월 : 전남대학교 전산학과 이학석사
 2005년 8월 : 전남대학교 전산학과 이학박사
 2011년 4월 ~ 2013년 6월 : 목포대학교 정보산업연구소 전임연구원
 2013년 7월 ~ 현재 : (주)UC 연구소장
 ※관심분야 : 클라우드 컴퓨팅, 빅데이터, NFC



지 유 강 (Yoo-Kang Ji)

2006년 2월 : 동신대학교 정보통신공학과(공학박사)
 2006년 3월 ~ 2009년 8월 : 동신대학교 정보통신공학과 전임강사
 2009년 9월 ~ 2014년 2월 : 동신대학교 초빙교수
 2011년 1월 ~ 현재 : 휴인테크 대표
 ※관심분야 : Computer Network, Embedded System, Networked Media Systems



김 종 원 (Jong-Won Kim)

1997년 8월 ~ 2001년 7월 : University of Southern California 연구 조교수
 1999년 12월 ~ 2000년 7월 : Technology Consultant for VProtect Systems Inc.
 2000년 7월 ~ 2001년 6월 : Technology Consultant for Southern California Division of InterVideo Inc.
 2001년 9월 ~ 2008년 3월 : 광주과학기술원 정보기전공학부 부교수
 2008년 4월 ~ 현재 : 광주과학기술원 정보기전공학부 교수
 ※관심분야 : Networked Media Systems and Protocols focusing "Reliable and Flexible Delivery for Integrated Media over Wired/Wireless Networks"