

IoT 기반 Big Data 기술 동향

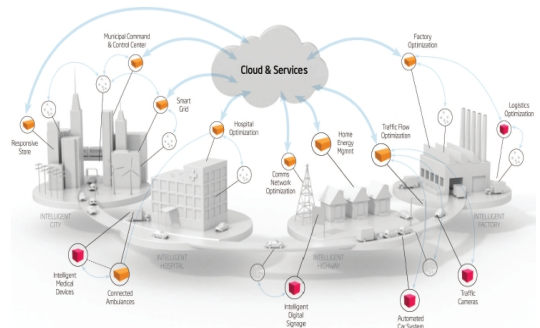
박 현 · 김 세 한
IoT융합연구부,
사물감성융합연구실

I. 서 론

세계는 컴퓨터(computer)의 시대에서 컴퓨팅(computing)의 시대로 극적인 변화의 중심에 와 있다^[1]. 전 세계 인터넷 사용자 20억 명, 국내 SNS(Social Networking Service) 사용자 2천만 명, 스마트기기를 비롯한 수십 억 장치(devices)들의 1조(trillion)개 연결 등, 이러한 환경 하에서 거대한 데이터가 생성되고 있다.

최근 세계 도처의 많은 기업들은 이러한 인터넷 연결과 그로 인해 도출되는 데이터를 통해 기업 운용에 커다란 가치 창출을 얻기 위하여 IoT(Internet of Thing), 빅 데이터(big data) 기술에 관심이 집중되고 있으며, IoT, 빅 데이터 원리들을 구현하려고 한다^[2]. IoT는 모든 사물 혹은 기기가 인터넷과 연결이 가능한 새로운 미래 유비쿼터스(ubiquitous) 서비스 인프라 또는 사람과 사물, 사물과 사물 간에 상호 소통하고 지식을 결합하여 상황 기반의 지능서비스를 제공하는 사물까지 확장된 인터넷이라고 정의하고 있다.

예로써, 집 내부에 스마트 계량기(smart meters), 신호등, 스마트폰의 GPS, 안전 시스템(security system), 원거리 감시 카메라로부터 영상(images) 등, 망 연결된 IoT의 수많은 장치들은 잠정적으로 이러한 빅 데이터의 크기(volume)나 데이터의 다양성(variety)을 만드는 원천이 된다^[3]. 또한, 세계 도처의 산업장비, 자동차, 전기미터, 선박상자 등, 그들은 위치, 움직임, 진동, 기온, 습도, 심지어 공기 중 화학적



[그림 1] IoT를 구성하는 장치들^[1]

변화까지 측정하고, 통신할 수 있다^[4].

이와 더불어 데이터 저장과 통신비용이 대폭 낮아지면서 빅 데이터 이용 여건이 성숙하게 되었으며, IoT의 궁극적인 잠재력은 이러한 시스템의 각 단계(step)에서 획득된 데이터로부터 가치를 도출할 능력을 갖고 실현될 것이다.

빅 데이터 산업 또한 방대한 양의 데이터가 축적되고, 이를 분석함으로써 이전에는 얻을 수 없었던 새로운 가치(insight)를 창출하고자 하는 시도로, IoT가 빅 데이터에 중요한 기여를 할 것이며^[4], IoT 기반 빅 데이터 기술은 센서 네트워크(sensor networks)를 포함, Wi-Fi, 휴대 전화망(cellular networks), 무선망(wired networks) 상에 수많은 물리적 사물들이 연결되어 서로 통신하며, 지능서비스를 위한 데이터를 생성하여, 방대한 빅 데이터를 통해 새로운 가치를 창출할 것이다.

본 연구는 산업통산자원부/미래창조과학부 및 한국산업기술평가관리원의 산업융합원천기술개발사업(산업융합기술-IT융합)의 일환으로 수행하였음[10044580, 농식품의 안전한 유통을 위한 위해인자 신속관리 시스템 기술 개발].

본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음(IITA-2005-C1090-0503-0003).



[그림 2] IoT를 통한 빅 데이터 수집과 APIs를 통한 상호 연결^[2]

II. 빅 데이터 기술

빅 데이터는 일반적인 데이터베이스 소프트웨어가 저장, 관리, 분석할 수 있는 범위를 초과하는 규모의 데이터(McKinsey, 2011)라는 규모 관점의 단순 정의와 다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고, 초고속 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처(IDC, 2011)라는 플랫폼 관점의 정의, 그리고 빅 데이터는 21세기의 원유(가트너, 2011)라고 하는 데이터 활용 관점으로 정의되고 있다.

일반적으로 빅 데이터는 데이터 규모가 크고, 다양한 데이터를 실시간 처리를 통해 훌륭한 가치 창조 및 중요한 예측 분석을 제공하는 의미에서 4V(Volume, Variety, Velocity, Value)를 특징으로 한다. 이러한 빅 데이터 처리를 위해서는 다음과 같은 기술들이 요구된다.

2-1 수집기술

사용자에 의해서 생성되거나, 각종 장치 및 시스템에서 자동으로 생성되는 데이터를 수동 또는 자동으로 모으는 프로세스로 다양한 데이터 소스에서 이종 데이터를 저장소로 효과적으로 수집하는 기술이다. 수집되는 데이터들은 고정된 필드에 저장된

데이터로, 데이터베이스와 스프레드시트 등과 같은 정형(Structured) 데이터, 고정된 필드에 저장되어 있지는 않으나, 메타데이터나 스키마 등을 포함하는 데이터로 XML이나 HTML 텍스트 등의 반 정형(Semi Structured) 데이터, 고정된 필드에 저장되어 있지 않은 데이터로, 텍스트, 이미지, 동영상, 음성 데이터 등의 비정형(Unstructured) 데이터 등의 형태로 나눌 수 있다. 대표적인 수집 툴인 Chukwa, Flume, Scribe 등을 이용하여 분산 환경에서 생성되는 데이터를 하둡(Hadoop)과 같은 분산 환경 파일 시스템으로 수집하고 저장하는 것이 가능하다^[5].

정형 데이터는 오래 전부터 DW(DataWare house) 및 BI(Business Intelligence) 솔루션 등을 활용하여 다양한 분석이 이루어져 왔으나, 반 정형 데이터와 비정형 데이터는 최근해야 주목을 받기 시작한 분야이다.

2-2 저장, 처리기술

대규모 자료를 여러 대의 소규모 컴퓨터에 나누어서 저장, 처리하는 기술로 방대한 분석 대상 자료를 저렴하게 처리해야 할 필요성에 의해 개발됐다. 하둡과 같은 분산 환경 파일 시스템은 대량의 자료를 저렴한 서버로 가상으로 묶어 대형 스토리지를 구성하고, 저장된 거대한 데이터 세트를 간편하게 분산 처리하는 역할을 한다.

하둡의 HDFS(Hadoop Distributed File System)는 거대 자료를 여러 대의 소규모 컴퓨터에 나누어서 저장하는 기술로 아주 큰 파일을 조그마한 파일 여러 개로 나눈 후 여러 대의 소규모 컴퓨터에 나누어서 저장하는 기술이며, 맵리듀스(MapReduce)는 이렇게 분산 저장된 자료에서 정보를 신속하게 추출하는 기술로 분산 처리된 자료로부터 빠르게 정보를 추출할 수 있는 프로그램 언어라고 생각하면 된다. 이 기술의 특징은 분산된 자료에서 정보를 추출할 때, 필요한 자료를 한 곳에 모아 분석을 하는 게 아니라, 분산 저장된 개별 컴퓨터에서 개별 자료를 분석한 후 분석 결과만 한 곳에 모으는 방식이다.

2.3 분석기술

빅 데이터 분석은 수집된 데이터를 분석에 적합한 형태로 가공하고, 데이터 간의 관계분석을 통해 현상을 설명하거나, 이후에 벌어질 일들을 예측하는 기술로 분석 툴, 분석 방법, 데이터 마이닝, 데이터 시각화를 위한 알고리즘 등의 기술들이 포함된다.

분석 툴로써, 최근 오픈소스 데이터 분석 엔진인 R이 주목받고 있다. R 언어와 개발환경을 통해 기본적인 통계 기법부터 모델링, 최신 데이터 마이닝, 시물레이션, 수치해석 기법까지 구현할 수 있어 구글, 페이스북, 아마존 등이 대용량 데이터 통계분석과 데이터 마이닝 등에 사용하면서 관심을 받고 있다. 데이터 마이닝을 활용한 고객 데이터 분석도 이러한 효과적인 데이터 관리시스템의 지원이 용이하게 되었다.

빅 데이터 마이닝은 데이터의 형태와 범위가 다양해지고 그 규모가 방대하여, 그 중요성이 더욱 부각되고 있다. 특히 웹에서 엄청나게 빠른 속도로 생성되는 웹 페이지(web page) 콘텐츠와 웹 로그(web log), 소셜 네트워크 서비스의 텍스트 정보와 영상과 같은 비정형 데이터를 분석하기 위한 다양한 방법론이 등장해 데이터 마이닝의 포괄 범위는 확장되고 있다.

텍스트 마이닝(text mining)은 대규모의 문서(text)에서 의미 있는 정보를 추출하는 기술이다. 분석 대상이 비구조적인 문서정보라는 점에서 정형화된 데이터 마이닝과 대별된다. 텍스트 마이닝은 텍스트 분석(text analytics), 텍스트 데이터베이스로부터 지식 발견, 문서 마이닝(document mining) 등으로 불리기도 한다.

웹 마이닝(web mining)은 인터넷 사용시 생성되는 웹 로그(web log) 정보나 검색어로부터 유용한 정보를 추출하는 웹을 대상으로 한 데이터 마이닝을 말한다. 웹 마이닝은 전통적인 데이터 마이닝의 분석

방법론을 사용하기도 하지만, 웹 데이터의 속성이 반 정형 혹은 비정형이고, 링크 구조를 형성하고 있기 때문에 별도의 분석기법이 필요하다.

오피니언 마이닝(opinion mining)은 이슈가 되는 인물, 사건, 이벤트에 대한 사람들의 의견이나 평가, 태도, 감정 등을 분석하는 기술로 특정 주제에 대해 사람들의 주관적인 의견을 모아 문장을 분석한다. 문장 분석에서는 사실과 의견을 구분해 의견을 뽑아 내어 긍정과 부정으로 나누고 그 강도를 측정한다.

버즈 모니터링(buzz monitoring)은 온라인에서 특정 주제에 대한 여론을 분석하는 기술이다. 예로써, 기업의 경우 트위터와 페이스북 등 인터넷에 올라온 기업 관련 댓글을 실시간으로 분석해 자사 이미지를 파악하고 대응전략을 세운다. 버즈는 별이나 기계가 웁웁 대는 소리를 지칭하며, 소비자가 특정 상품이나 서비스에 대해 긍정적인 입소문을 퍼트려 좋은 이미지를 만드는 것을 의미하는 버즈 마케팅(buzz marketing)에서 비롯됐다.

데이터 시각화(data visualization)는 데이터 분석 결과를 쉽게 이해할 수 있도록 시각적으로 표현하고 전달하는 과정으로 데이터 시각화의 목적은 도표라는 수단을 통해 정보를 명확하고 효과적으로 전달한다. R에는 다양한 시각화의 라이브러리가 내재되어 있어, 결과를 쉽게 표현하는데 유용한 툴로 각광을 받고 있다.

III. 국내외 기술 동향

국내 빅 데이터는 최근 관심이 집중되는 분야로 기술의 도입 단계라 할 수 있으며, 학계뿐만 아니라 IT업계에서도 연구와 관심이 활발하며, 과거에 버려졌던 대량의 데이터를 효과적으로 처리, 분석하여 새로운 데이터를 만들어 이를 통해 새로운 서비스를 제공하여 많은 부가가치를 만들어낼 수 있다. 이를 위해 빅 데이터에 대한 인프라 구축 및 지자체와 기

업 CEO들의 관심이 필요하다라는 지적이 나오고 있다.

빅 데이터는 창조경제 구현을 위한 신규 성장 유망분야로 주목 받고 있으나, 국내의 경우, 글로벌 기업에 비해 기술 격차가 존재한다. 공공, 민간에서는 아이디어 및 데이터는 있으나, 빅 데이터 분석 환경 및 활용 경험이 부재해, 서비스 도입에 어려움을 겪고 있다.

국내에서 삼성전자에 이어 현대자동차가 빅 데이터 대열에 합류하면서 제조 산업에 빅 데이터 도입 열풍이 불 것으로 예상된다. 이는 빅 데이터 기술이 일시적 마케팅에 그치지 않고 제조업의 제품 품질 향상과 매출 상승으로 이어질 수 있다는 업계의 판단에 따른 것으로 풀이되며, 그동안 관망하던 제조업체들의 도입도 속도를 낼 전망이다. 포스코와 LG 전자, SK하이닉스 등이 후속으로 빅 데이터 프로젝트를 추진할 것이라는 분석도 나오고 있다.

해외에서는 빅 데이터를 활용한 사례가 지속적으로 나오고 있다. 미국 FBI는 2013년 4월 보스턴 마라톤 행사장 근처 이동통신기지국의 로그 기록과 주변 사무실, 주유소, 아울렛 등의 감시카메라(CCTV), 청중의 휴대폰 카메라 등에서 수집한 10 테라바이트 데이터를 분석해 용의자를 찾아냈다. 또한, 이미 BMW, 볼보, 포드 등 글로벌 자동차 제조업체들은 최근 2~3년간 빅 데이터 시스템을 발 빠르게 도입해 차량 품질 개선 등에 활용하고 있다.

중국 정부는 자국 기업이 빅 데이터를 저장, 분석하는 인프라 개발에 나설 수 있도록 지원하고 있다. 중국 정부는 2010년 확정된 12차 5개년 개발 계획(2011~2015년)에서 에너지 절약, 환경보호, 바이오 산업, 첨단 장비 제조, 신 재료, 신에너지, 친환경 자동차와 함께 빅 데이터 관리까지 포함한 차세대 정보기술을 '7대 신흥 전략 산업'으로 꼽았다. 그러나 중국 기업 대다수는 아직 충분한 데이터를 갖고 있지 않다.

미국은 2012년 3월부터 2억 달러(약 2,160억 원)

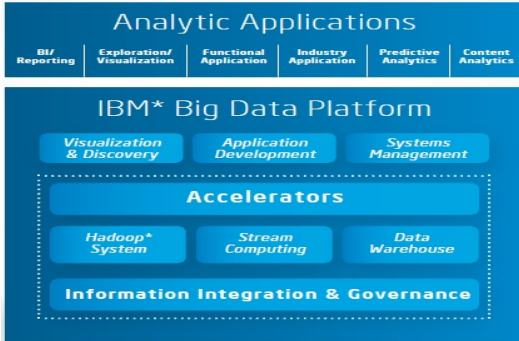
규모의 '빅 데이터 연구 개발 이니셔티브' 전략을 추진하고 있다. 이를 위해 미 정부는 과학기술정책실 산하에 '빅 데이터 고위운영그룹(BDSSG)'을 별도로 마련했다. 이 그룹에는 국립과학재단(NSF), 국립보건원(NIH), 국방부(DoD) 등 6개 연방 부처와 기관이 참여하고 있으며, 미국항공우주국(NASA), 미국해양대기관리처(NOAA) 등이 추후 합류할 예정이다. 이들 부처와 기관들은 연계 프로젝트로 얻을 수 있는 편익을 분석하고, 실현 가능한 협업 프로젝트를 개발, 제안하고 있으며, 빅 데이터 관련 R&D에 연간 2억 달러 이상을 투입하고 있다.

미국 McKinsey의 전망에 따르면 2018년까지 빅 데이터를 해석해 효과적인 결정을 내릴 수 있는 기술이나 방법을 가지고 있는 관리자나 해석자들이 미국 내에서만도 약 150만 명이 부족하고, 데이터과학에서 중요 해석기술을 가지고 있는 전문가 20만 명이 추가로 필요하다고 하였다.

영국은 기업 혁신부를 비롯해 16개 부처를 중심으로 2012년 6월 '오픈 데이터 전략'을 마련했다. 각 부처는 현재 보유하고 있는 데이터를 빅 데이터, 개인정보와 관련된 '마이 데이터', 서비스에 대한 이용자 경험과 만족도와 관련된 데이터 등으로 구별하는 작업을 추진 중이다. 그리고 의료, 교육, 세금, 고용, 기상 데이터 등에 대해 2015년까지 순차적으로 확대해 개방할 예정이다.

일본은 민간 분야를 고려하면서 빅 데이터 활용과 활성화 정책을 추진 중이다. 이를 위해 일본 정부는 데이터 개방, 기반기술 연구 개발, 활용 인재 확보, 표준화 방안 등을 마련하고 있다.

SAP은 최근 메모리 가격 하락과 실시간 데이터처리에 대한 요구사항 증가로 인 메모리(In-memory) 데이터베이스 'SAP HANA 솔루션'을 출시했다. 자체 하둡 버전은 없으며, 하둡 상용 버전 공급사와 제휴가 예상되며, SAP Integrator를 통해 하둡 연계를 지원하고 있다.



[그림 3] IBM 분석 응용^[7]

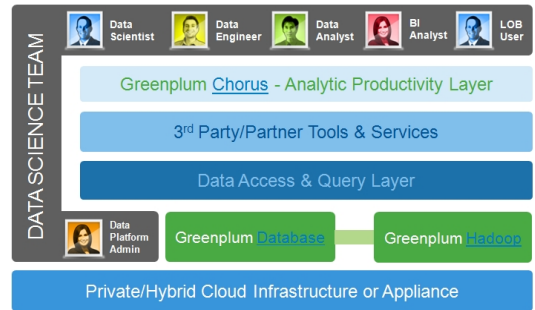
테라데이터는 ‘Teradata Aster Big Analytics Appliance’를 출시, 애스터(맵리듀스와 SQL을 결합한 분산 플랫폼)와 호튼웍스 협력을 통해 애스터 SQL-H를 개발, 어플라이언스에 통합했다.

호튼웍스는 ‘호튼웍스 데이터 플랫폼(HDP)’을 출시, 하둡, 하이브, 마훗, 우지, 피그, 주키퍼, 휴와 다른 오픈 소스 도구를 포함하고, 업체 고유 제품은 포함하지 않으며, 모든 소프트웨어는 무료 제공되며, 교육과 지원 프로그램을 통해 수익을 창출하고 있다.

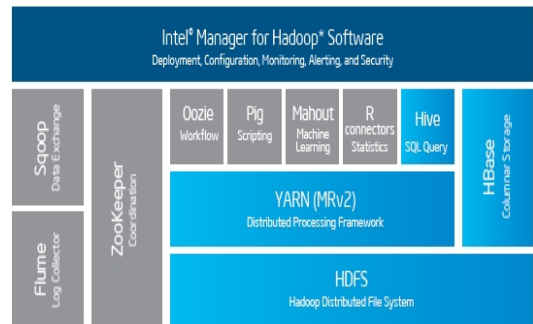
IBM은 ‘SPSS 애널리틱 서버(IBM SPSS Analytic Server)’ 출시했는데, 기존의 데이터 웨어하우스에 저장되어 있는 데이터를 분석하거나, 다른 저장장치로 데이터를 이동하지 않고도 하둡 플랫폼 안에서 직접 데이터 접근 및 실시간 고급 분석이 가능한 것이 특징이다.

EMC는 오픈소스 하둡 기술을 적용한 DW 솔루션인 ‘EMC 그린플럼’을 2011년 시장에 선보였다. 그린플럼은 단일 어플라이언스 내에서 정형, 비정형 데이터를 처리하는 장비로, 정형 데이터 분석을 위한 DB 모듈과 비정형 데이터 분석을 위한 하둡 모듈을 단일 어플라이언스로 지원한다.

인텔은 빅 데이터를 사용할 수 있도록 지원하는 ‘인텔 아파치 하둡 소프트웨어’를 2013년 출시했다.



[그림 4] EMC 그린플럼 통합분석 플랫폼^[8]



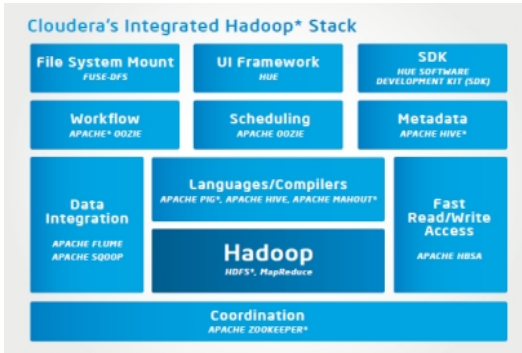
[그림 5] Intel 하둡 소프트웨어 요소^[9]

이 소프트웨어는 인텔 아파치 하둡 매니저를 포함, 업계의 선도적인 성능과 향상된 보안성을 제공하도록 인텔 반도체 제품군의 관련 기술을 활용해 설계됐다.

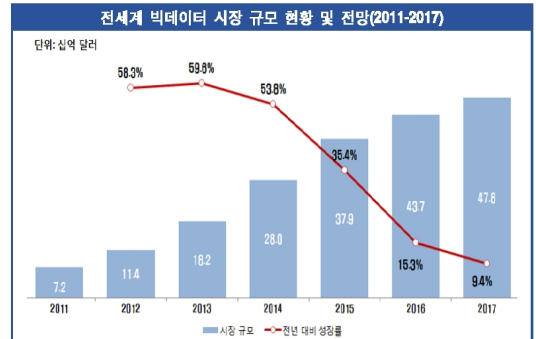
오라클은 정형화된 데이터와 비정형화된 데이터를 모두 감당할 수 있는 ‘빅 데이터 어플라이언스’를 출시했다. 역사 데이터와 연동되며, 기존 데이터베이스 시장에서 검증된 오라클 시스템의 신뢰성, 유연성과 성능에 기반해 빅데이터 시장을 목표로 하고 있다.

클라우드래는 ‘CDH(하둡, 하이브, 마훗, 우지, 피그, 주키퍼, 휴와 다른오픈 소스 도구 포함)’와 ‘클라우드래 매니저(CDH 환경 관리 도구)’를 출시하였다.

HP는 2011년 인수한 데이터웨어 하우스 솔루션인 ‘버티카’와 지능형 검색 엔진 솔루션인 ‘오토노미’ 기



[그림 6] 클라우데라 통합 하둡 스택^[10]



[그림 7] 전세계 빅 데이터 시장 규모 현황 및 전망 (2011~2017)

술을 통합했다. 대량의 데이터를 실시간으로 분석할 수 있는 환경을 제공하며, 영상, 텍스트, 음성 등 다양한 종류의 데이터를 의미 기반으로 분석해 주는 오토노미의 핵심 검색 기술인 ‘IDOL(Intelligent Data Operating Layer)’을 통해 감성적 분석까지 가능하다.

스플링크는 사용자 컨퍼런스 행사장에서 비공개 베타 버전인 ‘스플링크 하둡 오퍼스 앱(Splunk App for HadoopOps)’에 대해 소개했다. 이 제품은 싱글 엔드 투엔드 인터페이스가 전체 하둡 환경을 모니터하고, 사용자들은 하둡 노드, HDFS, 맵리듀스를 모니터링, 경고, 문제해결, 치료, 검색과 분석을 할 수 있다.

IV. 국내외 시장 동향

시장 조사 업체 Wikibon의 발표에 따르면, 2012년 전 세계 빅 데이터 관련 시장 규모는 전년 대비 약 59% 성장한 114억 달러로 추산되며, 글로벌 빅 데이터 시장은 향후 5년간 31%에 달하는 연평균 성장률을 기록하며 2017년에는 478억 달러 규모에 육박할 것으로 전망했다^[6].

빅 데이터 분야의 산업 구조적 측면에서 여타 정보통신 기술 산업과 마찬가지로 크게 하드웨어와 소프트웨어, 서비스 부문으로 구분한다. 각 세부 산업별 비중을 살펴보면 빅 데이터 서비스 부문이 41.5~

44%, 하드웨어 부문이 28.9~31%, 소프트웨어 부문이 25~29.7%의 비중을 차지하고 있다. 이와 같이 빅 데이터 서비스 부문이 가장 높은 비중을 보였다. 이 결과는 빅 데이터의 주요 기술이 구현되어 적용되는 서비스 영역이 전체 빅 데이터 시장에서 매우 중요한 비중을 차지하고 있음을 알려준다.

한국과학기술정보연구원(KISTI)이 2013년 4월에 발표한 ‘KISTI 마켓 리포트’에 따르면 국내 빅 데이터 시장 규모는 단기간적으로 2015년 약 2억 6,300만 달러(약 2,950억 원)에 달해 세계 빅 데이터 시장의 약 1.6%의 비중을 점유할 것으로 예측했다. 국내 정보통신기술관련 산업에서 빅 데이터 분야가 차지하는 비중은 2013년 0.6%에서 지속적으로 증가해 2020년에는 약 2.6%로 높아질 것으로 예상했다.

주요 해외기업들이 빅 데이터 시장을 주도하고 있는 환경에서 국내 기업들은 가격을 핵심경쟁력으로 설정하고, 2012년 하반기부터 2013년 상반기에 걸쳐 본격적으로 제품을 출시해 다양한 솔루션을 시장에 진출시킬 예정이다.

빅 데이터 산업의 경쟁상황을 보면, 현재 서비스 영역에서는 구글, 아마존 등 글로벌 인터넷 기업들이 시장을 선도하고 있다. 구글은 2011년 말 ‘Google

<표 1> 국내외 빅 데이터 서비스 사례

응용 분야	제목	내용
안전 빅 데이터	보스턴 마라톤 폭파 사건(2013. 4)	이동통신기지의 로그기록, 주변 사무실, 주유소, 아웃렛 등의 감시카메라(CCTV), 청중의 휴대폰 카메라 등에서 수집한 10테라바이트 데이터를 분석해 용의자 추출
통신 빅 데이터	서울시, 빅데이터로 심야버스 노선 구축(2013. 7)	시민들이 자정부터 새벽 5시까지 심야시간대에 사용한 휴대폰 콜데이터 30억여 건과 시민들이 이용한 심야택시 승, 하차 데이터 500만 건의 빅 데이터를 활용한 심야버스 노선이 처음으로 구축
소셜 빅 데이터	Google 독감 트렌드 분석 (2009. 2)	독감 관련 검색 키워드의 트렌드와 지역별 분포를 실시간 분석, 미국 질병통계예방센터(CDC)보다 2주 정도 빠른 통보 성과
	Google 자동차 판매량 예측분석 (2009)	오바마 정부가 경기부양책의 일환으로 “노후 차량 보상 프로그램으로 10억 달러 배정”, 미국 정부는 경기불황으로 인해 예산이 빨리 소진되지 않을 것으로 전망하였으나, 조기 소진되어 20억 달러의 추가 예산 편성, 구글은 웹 검색빈도수로 예산의 조기 소진을 예측함
	Walmart, 소비자 패턴 데이터 분석(2011)	데이터 과학자가 주축이 된 Walmart Labs 운영을 통해 획득한 소비자 패턴을 분석한 자료를 활용하여 재고를 감소시키고, 매장별로 소비자가 원하는 제품의 원활한 공급을 통해 고객만족도를 제고
기술 빅 데이터	IBM Watson 컴퓨터 (2011. 2)	IBM Watson 컴퓨터(인터넷 단절, 데이터 보유, 자연어 이해, 정보 찾고, 3초내 응답)와 사람과의 퀴즈 대결에

기술 빅 데이터		서 Watson 승(빅데이터 활용, 사회자 말 인식, 바로 결정)
	애플, Siri (2011. 4)	IOS용 질문에 답변하고, 동작을 수행할 수 있는 자연 언어 처리를 이용한 단말 소프트웨어, 스마트한 일상을 원하는 수백만 이용자의 요구에 기반, 조건을 사전적으로 제공
	볼보 차량, 제조업 생산 프로세스	과거에는 차를 만들어 처음 문제 파악 어려움, 5만 대정도 만들어야 숨겨진 문제점 파악, 차량 내에 센서 데이터 이용, 천대만 만들어도 미리 생산 프로세스 개선
고객 빅 데이터	도요타, 빅 데이터 교통정보서비스(2013. 7)	차량의 위치나 속도, 주행 상황 등을 포함한 빅 데이터를 기본으로 가공한 교통정보나 통계 데이터 등을 지자체나 기업에 제공해 교통 흐름 개선 또는 방재 대책 등에 활용할 수 있도록 하는 서비스
	아마존, 개인화 추천시스템	고객들의 선호도와 관심 표현을 바탕으로 선호도, 관심에서 비슷한 패턴을 가진 고객들을 식별, 비슷한 취향을 가진 고객들에게 서로 아직 구매하지 않은 상품들은 교차 추천하거나 분류된 고객의 취향이나 생활 형태에 따라 관련 상품을 추천
	넷플릭스, 동영상 추천	10만개의 영화정보, 1천 600만 명 고객의 시청이력을 분석해, 영화 콘텐츠를 추천, 넷플릭스 고객 중 60%는 추천 받은 콘텐츠를 이용
	마이에미 몽블랑 매장, 유통회사 매장 혁신	들어오는 고객의 휴대폰의 신호ID를 검사, 감시카메라의 모션을 인식, 동선 파악, 선반 어느 것을 잡는지 등을 파악, 품목을 재매치, 1주일 설치, 10% 매출 증가

BigQuery' 서비스를 시작했다. 'Google BigQuery'은 이용자가 분석을 원하는 데이터를 웹서비스를 통해 업로드하면 상호작용 방식으로 빅 데이터를 분석해 준다.

V. 국내의 서비스 사례

<표 1>에서는 국내의 대표적인 빅 데이터 서비스 사례를 요약하였다. 빅 데이터 서비스는 다양한 응용분야에 걸쳐 제공되고 있다.

VI. IoT 기반 빅 데이터 기술 이슈

기존 센서 데이터를 포함, 인터넷 연결을 통한 사물들의 데이터는 빅 데이터 분석을 위한 수용이 현재 미약한 상태이며, 가치 정보 창출, 사용자 전달 등의 연계가 필요하며. 다음과 같이 빅 데이터 처리를 위한 각 기술의 고려가 요구된다.

빅 데이터 수집 기술은 데이터를 효과적으로 수집하는 기술로, 정형 데이터는 오래 전부터 DW 및 BI 솔루션 등을 활용하여 다양한 분석이 이루어져 왔으나, 반 정형 데이터와 비정형 데이터는 최근에야 주목을 받기 시작한 분야로 용이하게 수집할 수 있는 방안들이 요구된다.

빅 데이터 저장 기술은 거대 자료를 여러 대의 소규모 컴퓨터에 나누어서 저장하는 기술로, 대량의 자료를 범용의 컴퓨터를 가상으로 묶어 대형 스토리지를 구성하고, 저장된 거대한 데이터를 간편하게 분산 처리하는 역할을 수행하며, 응용에 따라 실시간 데이터 처리가 요구된다.

빅 데이터 분석은 분석하려는 데이터를 수집하는 단계에서 시작하여 분석에 적합한 형태로 가공하고, 데이터 간의 관계분석을 통해 현상을 설명하거나, 이후에 벌어진 일들을 예측하는 과정으로 진행되며,

이를 위한 다양한 분석 방법이 요구된다.

또한, 빅 데이터 분석기술은 데이터의 형태와 범위가 다양해지고 그 규모가 방대해져, 그 중요성은 더욱 부각되고 있다. 웹 페이지 콘텐츠와 웹 로그, 소셜 네트워크 서비스의 텍스트 정보와 영상과 같은 비정형 데이터를 분석하기 위한 다양한 방법론이 등장해 데이터 마이닝의 포괄 범위는 확장되고 있다. 웹 데이터의 속성은 반 정형 혹은 비정형이고, 링크 구조를 형성하고 있기 때문에 별도의 분석기법이 필요하다.

VII. 결 론

IT시장 분석 및 컨설팅 회사인 IDC는 2015년까지 수십억 연결된 장치들 중 3분의 1 이상이 지능 및 지능처리가 가능한 지능 시스템 일거라고 예측했다. 이는 40억 달러 장치들과 2조 달러 이상의 잠재 시장과 더불어 IoT로부터 데이터들이 추출될 중요한 가치를 의미한다. IoT의 궁극적인 잠재력은 이러한 시스템의 각 단계에서 생성된 데이터와 사용자가 만들어낸 데이터를 잘 활용하여 삶을 향상시킬 수 있는 잠재력과 생산성의 놀랄만한 진보의 원동력이 될 것이며, 또한 이 데이터를 기반으로 빅 데이터 기술을 통해, 새롭고 변화되는 사업 서비스를 창출할 것이다.

참 고 문 헌

- [1] Simplifying the Internet of Things - Intel® Intelligent Systems Framework, *Intel White Paper*, 2012.
- [2] Andy Thurai, "Big Data, IoT, API: Newer technologies protected by older security", <http://blog.programmableweb.com/2013/05/17/big-data-iot-api-newer-technologies-protected-by-older-security/>, Intel, May 2013.

[3] Sander Duivestijn "The internet of things makes big data more interesting", <http://vint.sogeti.com/the-internet-of-things-makes-big-data-more-interesting/>, posted on March, 2013.

[4] Josh Leavea, "Big data's relationship with the internet of things", <http://blog.iobridge.com/2012/02/big-datas-relationship-with-the-internet-of-things/>, Feb. 2012.

[5] 손진승, 최규현, "빅 데이터 상용 솔루션 동향과

시사점", 주간기술동향, 2013년 4월.

[6] "2013년 글로벌 빅 데이터 시장, 전년 대비 61 % 성장 기대", 주간인터넷동향, 2013년 3월.

[7] <http://www.ibm.com/us/en/>

[8] <http://korea.emc.com/index.htm>

[9] <http://hadoop.intel.com/pdfs/>

[10] <http://www.cloudera.com/content/cloudera/en/home.html>

≡ 필자소개 ≡

박 현



1985년 2월: 전남대학교 전산통계학과 (이학사)
 1987년 2월: 서울대학교 계산통계학과 (이학석사)
 2005년 8월: 충남대학교 컴퓨터과학과 (이학박사)
 1988년 1월~현재: 한국전자통신연구원

사물감성융합연구실 책임연구원, 과제책임자
 [주 관심분야] 센서네트워크, M2M/IoT/WoT 상황인지, 하둡 클러스터링, 빅 데이터 마이닝

김 세 한



1998년 2월: 한국항공대학교 컴퓨터공학과 (공학사)
 2000년 2월: 한국항공대학교 정보통신공학과 (공학석사)
 2000년 7월~현재: 한국전자통신연구원
 사물감성융합연구실 실장
 [주 관심분야] RFID/USN, M2M/IoT, 농업/환경IT 융합기술