



특집 01

빅데이터 정의, 활용 및 동향



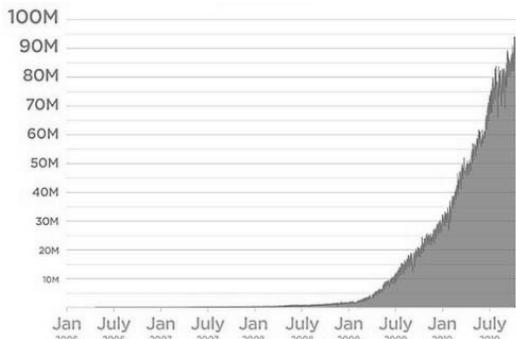
최 성·우성구 (남서울대학교)

목 차 »

1. 서 론
2. 빅데이터의 출현배경
3. 빅데이터의 정의
4. 빅데이터 정보처리기술의 필요성
5. 빅데이터 정보처리 기술현황
6. 빅데이터 처리기술 도입시 고려사항
7. 향후 전망

1. 서 론

현재 정보화 사회는 데이터가 넘쳐나는 세상에서 살고 있으며, 수많은 디지털 정보 데이터의 지



자료: Twitter (2010). Tweets per day. (Rao, L.(2010). Twitter Seeing 90 Million Tweets Per Day, 25 Percent Contain Links. TechCrunch.에서 재인용)

(그림 1) 일간 발신되는 트위터 양 (2006년 1월 ~ 2010년 9월)

배를 받고 있다. 지금 이 순간에도 디지털 정보 데이터는 기하급수적으로 쌓이고 있다.

미국 MIT 슬론스쿨에서 조사에 의하면 모바일 기기, 온라인 상거래, 소셜네트워크서비스(SNS) 등에서 하루에 250경 바이트 분량의 비정형 정보가 쏟아지고 있다고 한다. 또한, 매달 10억여 개의 트윗과 300억여 개의 페이스북 등 다양한 소셜미디어 메시지가 형성되고 있다,

이러한 데이터 속에서 우리 현실은 더 이상 부인할 수 없는 빅데이터(Big Data)시대를 살고 있는 것이다. 이는 기업에게는 위기이자 새로운 기회가 되고 있다. 본고에서는 빅데이터와 스마트폰 및 클라우드와의 관계와 빅데이터의 산업현장에서의 활용 및 동향에 대해서 고찰하고자 한다.

2. 빅데이터의 출현배경

빅데이터의 출현배경은 TV, PC에 이어 스마트

폰의 확산을 들 수 있으며, 스마트 폰 출현으로 가장 가치 있는 개인정보 및 위치정보가 양산되고 있다. 규모로 보면 크지는 않지만 경제적 가치는 최고인 개인정보와 결합되는 빅데이터에서 최고의 부가가치를 창출하고 있다.

과거 산업혁명에서는 철과 석탄이 세계 경제를 주도를 했으며, IT혁명에는 인터넷이 세계 경제 변화를 촉진하였다. 그러나 빅데이터는 스마트폰 혁명의 핵심자원으로서 세계 경제 변화를 이끌어 4의 경영자원으로 부상하였다.

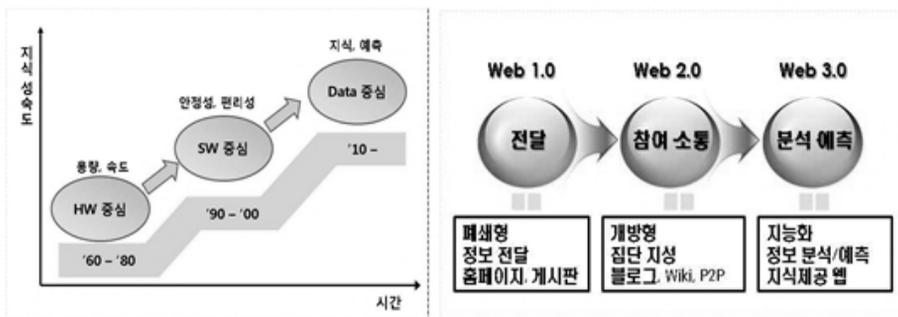
빅데이터의 활용은 경영혁신을 촉진하고 기업 간 경쟁력을 강화하며, 생산성 향상에 기여하고 있다. 스마트폰 혁명과 소셜 네트워크 효과로 디지털 공간의 데이터 빅뱅이 발생하여 정부에서는 데이터 지식 확보 및 활용 방안을 요구하고, 이에 데이터 활용 시 예산 절감 및 대내외 변화에 대한 신속한 대처, 삶의 질과 정부 신뢰도 향상이 가능해 졌다고 한다. 그에 대한 사례로는 오바마 정부의 'Pillbox 프로젝트'로 연간 5,000만 달러 비용 절감 효과를 나타내었으며, 독일의 경우에는 연방 노동기관에서 빅데이터 활용 맞춤형 고용으로 3년간 백억 유로 비용 절감하였다.

또한 기업에서는 고객 데이터 추적행위 및 수집행위 증가 추세에 있으며, 멀티미디어 콘텐츠 사용 관련 정보 증가도 한 몫을 해 왔다. 멀티미

디어 콘텐츠 사용 관련 정보를 사용한 고화질 동영상의 증가는 인터넷 트래픽의 막대한 영향을 주고 있다. 그리고 SNS의 급격한 확산과 비정형 데이터의 폭증과 정보 수집 가능한 센서 저변의 확대로 매달 수백 억개의 콘텐츠가 페이스북에서 공유되고 있다.

빅데이터란 Very large DB, Extremely large DB, Extreme data, Total data 등 다양한 용어로 지칭되고 있으며, Wikipedia에서는 기존 방식으로 저장·관리·분석하기 어려운 정도로 대규모의 자료를 의미한다. 그리고 Mckinsey(2011)에 따르면, 일반적인 데이터베이스 SW가 저장, 관리, 분석할 수 있는 범위를 초과하는 대규모 데이터라고 한다.

빅데이터는 3가지 관점에서 고려해야 한다. 첫째, 비즈니스 가치를 생각해야 한다. 빅데이터를 얘기할 때 '불륨'에 주안점을 둔 기술적 논의가 흥미롭기는 하지만, 초점은 '비즈니스 가치'에 두어야 한다. 즉, 비즈니스 과제와 목표를 식별하는 것이 중요하다. 웹로그 및 소셜 미디어를 사용하여 고객 가입 변동을 분석해야 하며, 클릭 스트림 및 다른 형태의 콘텐츠를 분석하여야 한다. 기업은 분석기술 및 솔루션 접근법에 적합한 비즈니스 가치에 초점을 두어야 한다. 둘째, 단지 거대함만은 아니다. 하둡(Hadoop)은 놀랄 정도로 관심



(그림 2) 빅데이터의 출현 과정

을 받고 있다. 그러나 하둡은 솔루션 측면에서 중요한 역할을 할 수 있는 훌륭한 기술일 뿐 빅데이터 전략을 하둡에 의해 정의해서는 안된다. 하둡은 단지 빅데이터를 다룰 도구 중 하나이지 전사 관점의 비즈니스 과제와 목표를 수행하기 위한 프레임워크의 솔루션이 될 수는 없다.

셋째, 분석의 활용을 고려해야 한다. 대부분의 경우 분석을 위한 데이터를 준비하기 위해 데이터 통합, 데이터 품질과 같은 정보관리 기술을 활용하려고 한다. 이러한 생각은 중요하지만, 전체 빅데이터 프로세스에 분석을 어떻게 적용할 수 있는지가 가장 중요한 변화를 가져온다. 즉, 빅데이터로 무엇을 할지 결정하기 위해 먼저 분석을 활용하고, 어떤 데이터가 관련이 있는지, 데이터를 저장할지 여부와 방법을 결정하는 것 등을 고려해야 한다.

3. 빅데이터의 정의

빅데이터의 전통적 개념은 구글과 같은 대기업이나 NASA의 연구과학 프로젝트에서 분석하는 대용량의 데이터를 일컫는 것이었다. 그러나 대부분의 기업에게 빅데이터란 상대적인 개념이다. 여기서 “빅(Big)”이란 기업 규모에 따라 달라지기 때문이다. 중요한 것은 형식적인 데이터소스 내 외부에서 새로운 가치를 창출하는 것이다.

테크타겟(Tech Target) 조사기업에 의하며, 설문조사 응답기업의 2/3가 데이터를 온라인으로 1년 넘게 유지하고 있으며, 43%가 3년이상 유지하고 있다고 한다. 빅데이터란 상대적인 개념이다. 즉 “빅(Big)”이란 기업 규모에 따라 달라지며, 중요한 것은 형식적인 데이터소스 내 외부에서 새로운 가치를 창출할 수 있다.

한계를 극복해야만 새로운 가치와 기회가 찾아오는 것이므로, “빅(Big)”이란 현재 자신의 출발

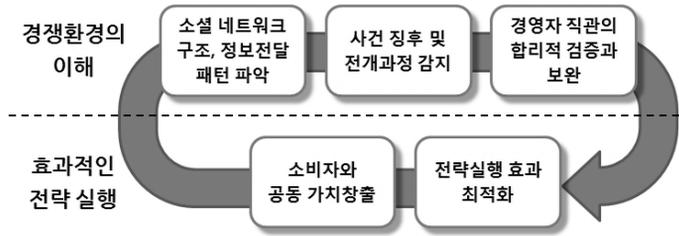
선에 따라 달라진다. 빅데이터는 사용자 집단을 위해 허용 경과시간 내에 데이터를 캡처하고, 관리하고, 처리하는 범용 하드웨어 환경 및 소프트웨어 도구의 영역을 넘어서야 한다.

사회영역에 활용되는 빅데이터란? 국가정보화 전략위원회(2011) 따르면, 대용량 데이터를 활용, 분석하여 가치 있는 정보를 추출하고, 생성된 지식을 바탕으로 능동적으로 대응하거나 변화를 예측하기 위한 정보화 기술이라고 정의한다. 이에 따라 빅데이터 정보처리기술을 정의하면, 대용량 데이터의 저장, 수집, 관리, 유통, 분석을 처리하는 일련의 기술을 의미한다. 실제로 글로벌 기업들은 빅데이터 분석을 통해 새로운 비즈니스 모델을 만들어 성공적으로 활용하고 있으며 빅데이터 처리기술 역량을 강화하고 있다.

새로운 가치를 창출하고, 지속적인 마케팅 프로그램을 테스트하고, 최적화하는 등 다수의 작업을 진행하면서 분석가들은 다양한 해결과제를 제기한다. 그렇다고 해도 유사한 분석은 자동화하여, 최적의 고객 혜택을 위해 설계된 콜 센터 스크린을 통해 알려준다. 위와 같은 빅데이터는 데이터 웨어하우스를 기반으로 해왔다. 그러면 앞으로도 가능 할 것인가? 그 해답은 데이터의 사용 빈도에 따라 달라진다. 데이터의 사용 빈도가 낮고 변동도 거의 없는 경우, MapReduce와 같은 도구가 등장하면서 효율적인 데이터 처리가 가능하여 저렴한 하드웨어의 분산파일시스템이

정형 data		Biz 혁명
반정형 data	Information	Smart 정부
Social data		Life Smart
Sensor data		

(그림 3) 빅 데이터 정의



(그림 4) 빅 데이터 분석의 활용효과

효율적인 옵션이 될 수 있다.

빅데이터는 데이터 문제의 진화 속도만큼 혁신적인 접근 방식을 이끌어내고 있다. 이제는 새로운 아이디어와 정보기술에 문을 열고 꾸준한 실험정신에 귀를 기울여야 한다.

모바일 혁명시대, 스마트 사회에서는 빅데이터가 핵심자원이며, 이를 수집하여 분석하는 새로운 빅데이터 기업과 신산업이 출현하고 있다. 공공 분야에서 빅데이터 활용 시 경제적인 효과는 국내의 경우 최소 10.7조원 이상이 추산되며, EU에서는 비용절감, 부정 및 오류에 따른 손실 감소, 세수 증대에 따른 비용 효과가 무려 1,500~3,000억 유로의 규모가 된다.

분석 및 예측 시대의 살아가는 현재에 빅데이터를 활용하는 선진국의 효율적인 정부 운영을 도모하는 제반 현황으로 볼 때, 빅데이터를 활용하지 못하는 정부는 국제 사회에서 국가 경쟁력 하락의 위험과 대면하게 된다. 또한 데이터 활용에 따른 우려를 불식시키기 위해서는 ‘정보의 익명성’과 ‘수집 데이터 비저장’ 등을 포함하는 공공데이터 활용 가이드라인이 필요하다.

4. 빅데이터 정보처리기술의 필요성

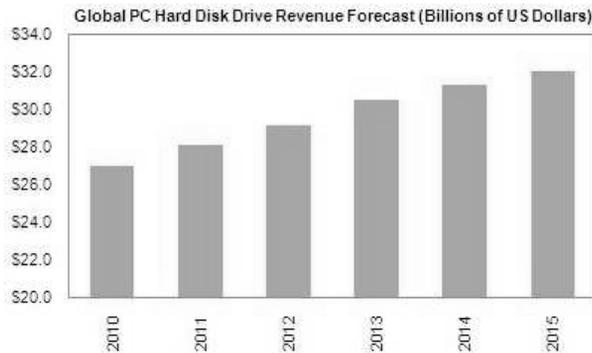
미국 시장조사기관 IDC에 따르면 지난 10년간 생성된 데이터보다 최근 2년간 생성된 데이터양이 더 많다고 한다. 실제로 불과 10년 전 까지만

해도 데이터베이스가 다루는 데이터양이 수천 만 건에 달하면 기업들은 대용량 데이터 처리 솔루션을 도입하고 전문 인력을 확충하였다. 최근 2년 사이 데이터베이스가 감당해야 할 데이터의 양은 급속도로 증가했으며 이제는 수백억 건에 달하는 빅데이터를 처리해야하는 현실에 부딪치고 있다.

데이터의 성격도 바뀌었다. 과거에는 형식이 정해져 있는 텍스트 위주의 데이터가 많았던 반면 이제는 그림, 동영상, 음성 위주의 비정형 데이터가 급속히 증가하고 있다. 성능적인 측면이나, 가용 데이터양을 측정해 보았을 때 수천만 건의 텍스트 중심의 정형 데이터를 처리했던 관계형 데이터베이스로는 수백억 건의 비정형 데이터의 처리는 현실적으로 어렵다. IT기업들은 이러한 빅데이터를 처리하기 위해 처음에는 하드웨어적인 시각으로 접근하였다.

DAS(Direct Attached Storage), NAS(Network Attached Storage), SAN(Storage Area Network) 등의 스토리지 장비를 추가하는 방식으로 늘어나는 데이터를 감당하였으며, 이를 데이터베이스가 처리했다. 그러나 이는 곧 한계에 부딪혔다. 스토리지만 늘린다고 해서 빅데이터를 처리할 수 있는 것이 아니기 때문이다.

시장조사업체인 아이서플라이의 통계를 보면 세계적인 데이터 증가 추세에도 불구하고 스토리지 서버에 탑재되는 HDD(Hard Disk Drive)시장



(그림 5) 세계 HDD 시장 예측

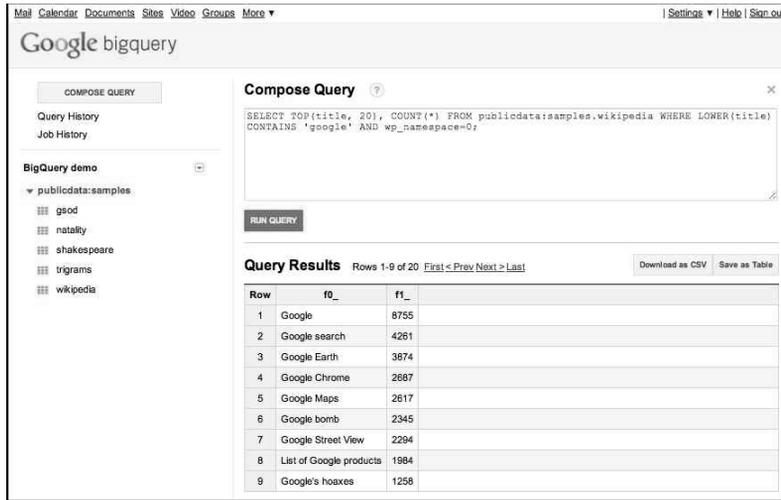
이 둔화되고 있다고 한다. HDD는 스마트폰, 태블릿 PC 등에 사용되는 플래시 메모리 기반의 저장장치로 빠르게 대체되고 있지만 HDD에 비하여 비용이 많이 발생하고 가용용량 또한 빅데이터를 처리하기 위해선 부족 한 것이 사실이다. (그림 5)는 세계 HDD 시장 예측이다. 이로서 IT 기업들은 스토리지 증설과 함께 대용량 병렬 처리(MPP, Massively Parallel Processing)기술을 도입하였다. MPP기술은 프로그램을 여러 부분으로 나누어서 다수의 프로세스가 각 부분을 동시에 수행시키고 하나의 프로그램을 수행하는데 다수의 프로세스를 이용하여 처리할 수 있었다. 그러나 이 역시 정형 데이터를 처리할 때는 효율적이었지만 폭증하는 비정형 데이터를 처리하기에는 비용이 효율적이지 못했다. 위에서 언급한 기업들의 빅데이터 처리를 위한 접근 사례에서 보면, 기존의 대용량 데이터 처리방식으로는 빅데이터의 처리가 실질적으로 어렵다. 그러므로 대용량의 비정형 데이터를 처리하기 위한 새로운 기술이 필요하다는 점을 알 수 있다. 실제로 글로벌 IT기업 및 오픈소스 진영에서 새로운 정보처리기술들이 개발되어 지속적으로 발전하고 있으며, 이미 해외 글로벌 기업들은 성공적으로 도입활용하고 있다.

5. 빅데이터 정보처리 기술 현황

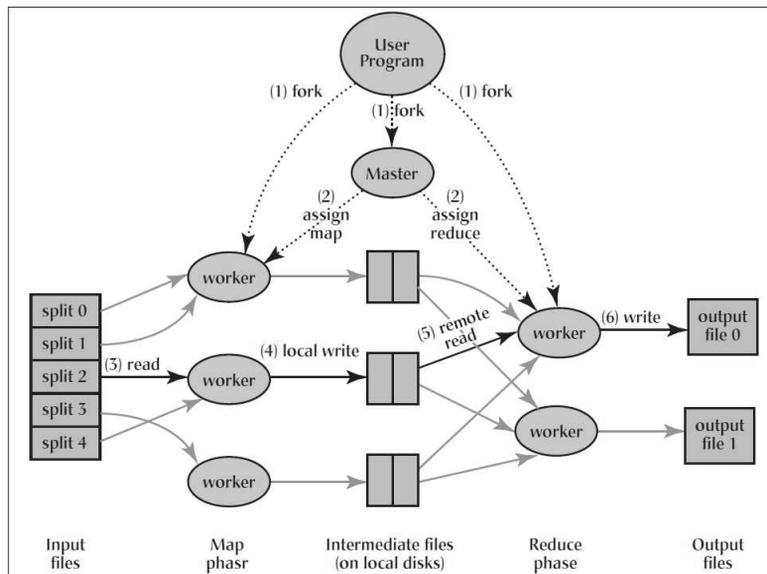
5.1. 구글의 빅데이터 처리기술

글로벌 인터넷 검색 서비스기업 구글은 2003년 GFS(Google File System)를 시작으로 2004년 MapReduce, 2005년 Sawzall, 2006년 Bigtable과 같은 대용량 데이터 처리 기술을 논문을 통해 발표하였다. 구글의 이러한 기술 공개는 새로운 빅데이터 처리기술 개발의 초석의 역할을 하고 있으며, 이로 인하여 IT 전문가들로부터 이 분야의 선두기업으로 인정받고 있다.

최근 구글은 Bigquery서비스(그림 6)를 구글 컨퍼런스에서 공개하여, 이 서비스를 이용하는 개인과 기업에게 데이터 센터에 자원을 투자하지 않고도 최대 70테라바이트의 비압축된 데이터 분석 기능을 시범 서비스로 제공하고 있다. (그림 7)은 구글이 발표한 대표적인 빅데이터 병렬처리 모델 MapReduce의 실행구조이다. MapReduce는 Map과 Reduce의 함수 기반으로 구성되어 있으며, (Key, Value)기반의 데이터를 병렬처리 기술이다. Map함수에서 데이터를 처리하고 Reduce함수는 원하는 결과 값의 계산을 담당한다.



(그림 6) 구글의 bigquery 서비스

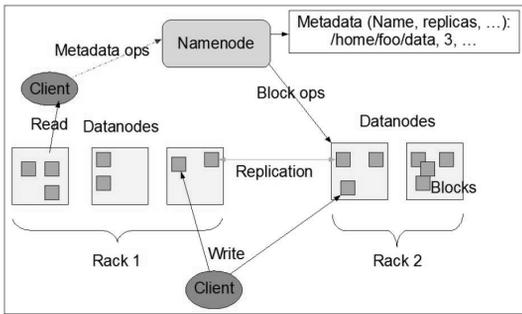


(그림 7) Google의 MapReduce 처리구조

5.2 하둡(Hadoop)

오픈소스인 하둡(Hadoop)은 구글의 GFS를 대체할 수 있는 분산 파일시스템(HDFS, Hadoop Distributed File System)과 MapReduce를 구현한 빅데이터 처리 기술의 대표적인 프레임워크이다.

하둡의 분산 파일시스템(그림 8 참조)은 파일을 블록단위로 나누어 각 노드 클러스터에 저장하며, 데이터 유실을 막고 부하처리를 위해서 각 블록의 복사본(Replication)을 만든다. 하둡은 적은 비용으로 빅데이터의 처리가 가능할 뿐만 아니라 높은 사용 편의성을 제공하며, 이에 기업들은 하

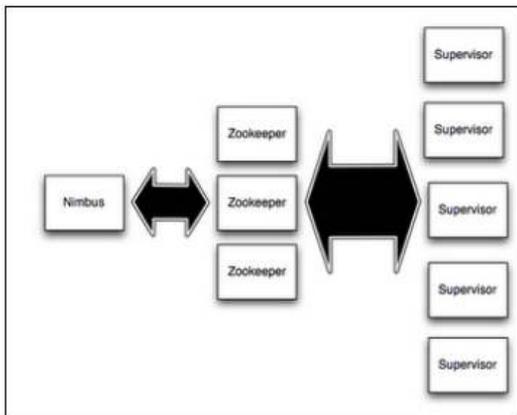


(그림 8) 하둡(hadoop)의 분산 파일시스템(HDFS) 처리구조

둡을 빅데이터의 분산처리에 활용하기 시작하였다. 뿐만 아니라, 최근에는 IBM, 오라클, 테라데이터, EMC 등의 글로벌 데이터분석 전문기업들도 자사의 상용 솔루션에 하둡을 포함시키거나 연계가 가능하도록 지원하고 있다.

5.3 글로벌 인터넷 서비스 업체들의 빅데이터 처리기술

아마존, 야후, 페이스북, 트위터 등의 대형 글로벌 인터넷 서비스 업체들도 자사의 기술과 플랫폼을 공개하여 빅데이터 처리 기술의 발전에 공조하고 있다. (그림 9)는 트위터가 공개한 데이



(그림 9) 트위터 Storm의 설계구조

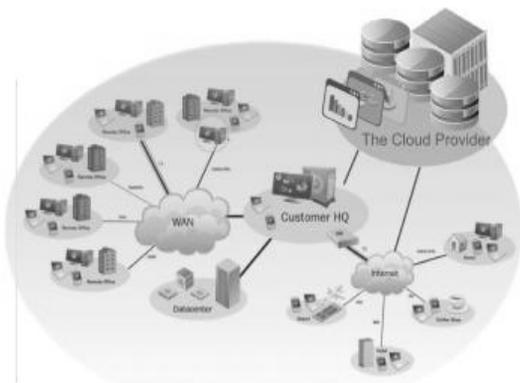
터 분석 “스톰”의 설계구조이다. 주 컴퓨터에서 실행되는 님버스(nimbus)와 그 명령을 따르는 하위 컴퓨터에서 실행되는 슈퍼바이저(supervisor), 그리고 그 사이의 최적화를 맡는 클러스터 기술 주키퍼(zookeeper)를 두어 성능을 최적화하였다.

트위터는 스톰 기술을 도입하기 전까지는 역시 하둡 기술을 사용하였다. 페이스북, 야후, 아마존은 하둡 기반의 검색 서비스를 채택하고 있으며, 국내의 경우 NHN이 일부 서비스의 로그 검색 및 데이터 분석에 하둡을 사용 중이다.

5.4 클라우드와 빅데이터의 관계

클라우드는 표준화되어 있는 IT기반 기능으로 컴퓨팅, 저장장치, 네트워크 SW등을 포괄하는 전반적인 정보기술 자원으로 제공되는 서비스 이외에는 없기 때문에 제공자에 의해 제공되는 표준을 따른다. IP망을 통하여 접근하기 때문에 IP망과 HTTP, REST, SOAP등 웹기반 컴퓨팅 프로토콜을 활용하여 사용자 인터페이스를 위해 운영체제에 중립적인 표준 웹 브라우저와 웹 표준을 지원한다.

빅데이터는 인터넷을 기반으로 컴퓨터 네트워



(그림 10) 클라우드 컴퓨팅 네트워크

크를 통하여 사용자가 필요하면 언제든지 인터넷 접속을 통하여 데이터를 공유를 할 수 있다. 클라우드가 사용량과 데이터양이 증가를 하게 되면서 클라우드 대용량화가 진행되어 방대한 데이터를 활용을 하지 못하는 사례가 늘어가게 되었는데 이러한 부분들을 빅데이터라고 볼 수 있다. 그래서 빅데이터를 처리를 하는 가장 큰 이유가 흔히 저장공간을 늘리기 위해서라고 생각을 할지 모르지만 기하급수적으로 늘어난 데이터를 효율적으로 관리를 하는 것뿐만이 아니라 이를 분석해서 정보자산으로 활용을 하고자 하기 위하여 스토리지 및 네트워크 기술 도입을 하여 기존에 존재하는 데이터를 활용할 수 있도록 하는 것이 빅데이터 처리에 대한 부분이라고 볼 수 있다. 클라우드란 다양한 클라이언트 디바이스에서 필요한 시점에 인터넷을 이용해 공유폴에 있는 서버, 스토리지, 애플리케이션, 서비스 등과 같은 정보기술 리소스에 쉽게 접근할 수 있는 모델이다.

그리고 빅데이터란 기존 데이터베이스 관리도구의 데이터 수집, 저장, 관리, 분석의 역량을 넘어서는 대량의 데이터 셋으로 IT업계의 핵심 양대 트렌드인 클라우드와 빅데이터가 결합하여, 클라우드 환경은 다양한 데이터가 폭발적으로 증가하는 빅데이터 시대에서 핵심적인 플랫폼 역할을 할 것이다. 즉, 빅데이터의 활용을 가능하게 만든 것이 클라우드 컴퓨팅 기술이다. Hadoop과 MapReduce는 거대 데이터를 개별 클러스터에 분산, 저장하고 보관노드에서 병렬 처리함으로써 고속 처리 및 저비용화를 모색했다. 앞으로 대부분의 데이터는 클라우드에서 생산 및 유통되며, 2015년 생산되는 데이터 중 클라우드 내에 있는 비율이 5/35Zettabyte로서 대략 14.3%를 차지하며, 2020년 생산되는 데이터 중 클라우드로 유통되는 비율은 12/35Zettabyte로서 대략 34.3% 차지하게 된다.

5.5 빅데이터 분석의 활용효과

기업은 빅데이터의 분석을 통해 경쟁 환경을 이해하고 효과적으로 전략을 실행할 수 있는 기반을 강화하고 있다. 그동안 빅데이터를 사용한 것 중에 가장 많은 분야가 카드 분야라고 볼 수 있다. 카드 내역서를 분석하려면, 개개인의 카드 사용 내역서를 기본으로 쇼핑행사, 쿠폰행사 등 여러 가지 행사에 대한 쿠폰 및 적립금 행사를 하는데 사용해 왔다. 예를 들면, 커피를 좋아하는 사람에게는 카드 결제시 추가 디스카운트 된다는 내용의 문자를 받으면 이왕이면 그 카드로 결제하는 것처럼 빅데이터를 사용한 마케팅 방법은 점점 변화되어 가고 있다. 위치기반 소셜커머스를 사용하고 있으면, 자동차 보험 갱신 시 여러 업체에서 연락오는 것도 다 빅데이터를 사용하는 것이라고 볼 수 있다. 쇼핑몰에 있는 검색 및 장바구니 일단 저장한 후 며칠 후에 이메일로 그 검색 내용 및 장바구니에 있는 물건에 대한 내용의 메일을 받는다면 그 역시 빅데이터를 활용한 마케팅 기법이라고 할 수 있다. 위치기반의 GPS 정보를 활용한 마케팅에서도 빅데이터 부분을 사용한다. 일정한 거리의 행사 및 이벤트 내용을 문자로 전송하려 많은 사람들이 참석할 수 있도록 하는 마케팅에서도 빅 데이터라는 것을 사용한다. 따라서 빅데이터를 정보를 수집하여 저장 및 분석하려 새로운 정보를 만들어 내는 기술력이라고 할 수 있다.

일본 노무라 연구소는 빅데이터의 활용 방안에 관한 특집 보고서에 의하면 향후 빅데이터는 클라우드 컴퓨팅과 소셜미디어의 연결을 통해 2012년 이후 적극적인 활용 움직임이 가속화되고, 비정형 데이터를 활용하여 개개인에 최적화된 맞춤형 서비스를 제공할 수 있을 뿐 아니라, 정확도가 높은 예측분석이 가능하다고 한다.

6. 빅데이터 처리기술 도입 시 고려사항

6.1 오픈소스 도입 이슈

2011년 12월 현재까지의 빅데이터 처리기술의 동향을 보면, 하둡을 포함한 오픈소스 진영을 중심으로 다양한 기술이 빠르게 진화하고 있으며, 아직 상용 솔루션 보다 오픈소스의 비중이 높다고 한다. 기업은 오픈소스를 활용하여 초기 투자 비용 절감, 서비스 성능향상, 가용성 향상을 기대할 수 있다. 하지만 오픈소스의 공통 이슈인 유지보수를 위한 기술지원, 안정성에 대한 불안감 등이 문제점으로 공존하고 있다. 기업은 오픈소스 도입 시, 유지보수 등 기술지원 체계 확보 및 안정성 분석 등을 실시하여 사전에 해당 문제점을 최소화할 수 있는 방안을 검토해야 한다.

6.2 전문 기업과 전문 인력 부족

국내 기업은 아직까지 빅데이터 처리에 대한 기술적 이해와 경험이 부족하다. 빅데이터 처리의 필요성에 대한 시각 차이가 문제 일 수 있지만 향후 빅데이터의 중요성을 고려한 인식 변화가 필요하다. 현재 소수의 기업만이 빅데이터 처리 및 활용을 위한 고민을 하고 있으며 도입을 한다 해도 기술력 있는 전문 기업과 인력이 턱 없이 부족한 실정이다. 매킨지 보고서에 따르면 미국의 경우 2018년경 데이터 분석 전문가가 14~19만명 정도 부족할 것으로 전망하고 있으며, IT전문가들은 국내의 경우 이보다 더 심각한 인력난을 예상하고 있다. 기업뿐만 아니라 정부 차원의 전문 기업과 전문 인력의 양성이 필요한 시점이다.

6.3 데이터 보안과 성능의 조화

빅데이터를 활용하기 위해서 기업이 저장, 관

리해야 하는 대상 데이터는 꾸준히 증가하고 이 데이터에는 민감한 개인정보와 기밀정보가 포함될 수 있다. 이러한 데이터는 특별히 관리되어야 하나 관리 소홀로 인하여 유출될 수 있다. 이러한 데이터의 유출을 방지하기 위하여 빅데이터 처리시 인증이나 암호화를 적용할 수 있지만 이에 따른 성능저하도 고려해야 한다. 기업은 빅데이터 처리기술 도입시 “보안과 성능”이라는 상충되는 두 마리의 토끼를 잡기 위한 노력이 필요하다.

7. 향후 전망

빅데이터에 관련한 정부의 새로운 정책들도 발표되고 있다. 정부에서는 데이터를 정책에 활용하는 “빅데이터를 활용한 스마트 정부 구현 방안”을 발표하고, ‘빅데이터 활용추진단’ 신설하였다. 이에 따르면, 정부는 공공과 민간에서 축적한 광범위한 지식정보를 누구나 활용할 수 있도록 국가 지식 정보 플랫폼과 민간지식정보 유통을 촉진하는 기능의 사이버 지식정보거래소를 2013년부터 단계적으로 구축하고, 범국가적 지식정보 활용 촉진에 필요한 법, 제도개선과 안전한 지식 정보 공유를 강화할 예정이다. 그리고 빅데이터 활용 및 인프라 기반조성을 선도적으로 추진하기 위하여 빅데이터 경쟁력 함양을 위한 국가적, 사회적 기반을 정비하고, 또한 산, 학, 연 협력을 통한 원천기술 개발로 빅데이터의 핵심 기반을 확보하고 있다.

빅데이터는 개인과 기업에서 가치 있게 활용될 수 있을 뿐만 아니라 국가 신성장 동력으로서의 가치가 높아지고 있다. 향후 국내 빅데이터 처리 기술의 발전과 활성화를 기반으로 빅데이터 시대의 “빅데이터 강국 대한민국”을 기대한다.

참고 문헌

- [1] 김재훈, 금융보안연구원 리포트, 2011-021, “빅데이터 시대 도래에 따른 데이터 처리기술 현황과 전망”.
- [2] 국가정보화전략위원회, “빅데이터를 활용한 스마트 정부 구현(안)”, 2011.10.26
- [3] 채승병 수석연구원, “정보홍수 속에서 金脈 찾기: 빅데이터(Big Data) 분석과 활용”, SERI경영 노트, 2011.2.10. (제91호)
- [4] 노무라연구소: http://www.nri.co.jp/publicity/n_letter/index.html, NRI뉴스레터 Vol.108, 2011.9
- [5] Gartner, <http://www.gartner.com>
- [6] IDC, <http://idc.com>
- [7] iSuppli, Hard Disk Drive Market for Computers to See Lower Revenue Growth in 2011, 2011.5



우 성 구

이메일 : wsk527@naver.com

- 1992년 연세대학교 산업대학원 전자계산학 공학석사
- 2009년 충북대학교 전자계산학과 이학박사
- 1978년~1998년 (주)삼화 전산실, 현대엔진공업(주) 전산실, 해외건설협회 전산실(과장), (주)CHNO건설링(과장), (주)경동보일러 전산실(부장)
- 1999년~2008년 (주)ASE(이사), (주)지아이에스(수석연구원)
- 2001년~2011년 한서대학교 컴퓨터정보공학과 외래교수
- 1998년~현재 평택대학교 컴퓨터학과 외래교수
- 1996년~현재 남서울대학교 컴퓨터학과 외래교수
- 관심분야: Constraint Database, 데이터마이닝, 빅데이터

저 자 약 력



최 성

이메일 : sstar@nsu.ac.kr

- 현재 남서울대학교 컴퓨터학과 교수
- 한민족IT평화봉사단 단장
- 전) 한국생산성본부 대한민국 OA추진국장
- 제주은행 전산실장, 기업은행 전산개발부
- 관심분야: OA 및 SW공학, 이러닝, 기능성게임, VR미디어 융합콘텐츠, 영재교육 등
- 저서: 게임PD가 되는길(국내 최초게임 저서), 리엔지니어링(94년도 베스트셀러), 가상화 스토리지 네트워크, CBD엔지니어링, 사이버대학 가이드 등