



특집 05

빅데이터와 인 메모리 컴퓨팅 기술



박호진 ((주)알티베이스)

목 차 »

1. 서 론
2. 과연 빅데이터란 무엇일까?
3. 빅데이터를 보는 관점
4. 빅데이터의 기술
5. 인-메모리 컴퓨팅
6. 결 론

1. 서 론

2012년이 시작되면서 지금까지 IT 업계에서 주기적으로 나오는 단어가 있다. 이 단어는 IT 라는 직종의 업무를 하고 있는 사람이라면 인터넷에서부터 세미나 초청장까지 다양한 곳에서부터 들었을 단어이자 키워드이다. 빅데이터(Big Data)라는 단어가 그 키워드이다. 현재도 이 빅데이터라는 주제는 이미 다양한 곳에서 세미나와 학술회의를 진행하고 있으며 기업들은 새로운 사업의 기회를 창출하고자 많은 노력을 기울이고 있다.

보통 사람들이라면 이 키워드를 한 번 들었을 때는 그냥 어느 정도 큰 볼륨의 사이즈 데이터이겠거니 하고 흘려듣는 사람도 있을 것이다. 하지만 이 말은 단지의 큰 볼륨의 사이즈가 아닌 어마어마한 사이즈에 대한 문제로 부터 시작된다. 이 ‘어마어마’라는 표현은 일반 사람들에게는 마음

에 와 닿는 말들은 아니겠지만 IT 업계에 종사하는 사람이라면 어느 정도 상상이 되는 사이드라고 생각한다.

이러한 ‘어마어마’한 데이터의 환경은 오래전부터 IT 환경의 데이터 발생량에 대해 전 세계적으로 폭발적으로 증가하는 추세에 이미 와 있다고 경고해오고 있었다. 이미 4년 전인 2008년도에 가용는 스토리지의 발전 속도에 대비하여 디지털 정보량의 성장률이 추월하였다는 발표의 내용도 있었고 2008년 기준으로 10여 년간 데이터 연 평균 증가율이 50.6%에 달한다고 하였다. 이는 5.5년 마다 관리되어야 하는 데이터의 양이 10 배로 증가한다는 것을 의미한다. [1]

이 내용을 풀이하면 현재 기업의 데이터 수용이 10TB라면 5년 후에는 100TB 그 후 5년 후에는 1000TB에 해당되는 데이터를 수용하고 관리하여야 한다고 이야기 하고 있다. [2]

이 부분에서 우리가 알아야 하는 시사점은

2008년도 발표 내용들을 현재의 2012년 기준에서 돌아본다면 이는 충분한 근거가 있는 내용이라고 볼 수 있고 현 시점에도 이미 그 이상의 훨씬 더 많은 데이터가 발생된다는 것 또한 분명히 모든 사람들이 공감할 수 있을 것이다. 이러한 막대한 양의 데이터의 발생으로부터 현재의 빅데이터라는 키워드를 만들어 낸 것이 아닐까 생각해 본다.

2. 과연 빅데이터란 무엇일까?

빅데이터에 대하여서는 다양한 내용들이 그 정의를 내리고 있다.

일반적인 의미의 빅데이터는 우리가 현재 운영하고 관리하고 있는 데이터 외에 다양한 데이터에 대해서 정보화하는 일련의 과정(수집, 저장, 관리, 분석)의 한계를 넘어선 수치의 데이터 발생에 따른 부분을 지칭하고 있다.

과거의 데이터 발생은 단지 기업이 서비스 제공을 목표로 콘텐츠를 만들거나 사용자 및 업무 프로세스를 위한 데이터들이 주를 이루었다. 기업 내부의 직원들이 생산해내는 데이터 량은 그 당시에 돈을 지불 하여 구입할 수 있는 저장매체로 충분히 저장하여 운영 및 관리 할 수 있는 정도의 데이터양이었고 DW나 DM을 이용하여 충분히 분석하고 통계를 가짐으로써 데이터를 생산하여 관리만 잘하면 더 이상 문제가 발생하지 않아 단지 꾸준히 데이터들을 콘텐츠화하여 고객이나 담당 업무자에게 공유하면 아무 문제가 되지 않는 구조로 관리되었다. 이것은 마치 과거 기업들은 대부분 내부의 데이터 발생에만 신경을 쓰는 시대였다고 볼 수 있을 것이다.

하지만 IT 산업의 환경이 하루가 다르게 바뀌고 하루가 다르게 신기술이 쏟아져 나오듯이 이 IT 환경은 다시 한 번 변화라는 구름위에 띄워지

게 된다. 이 변화라는 내용은 기본적인 요건은 더 이상 기업 내부에서 생겨나는 데이터 발생량만을 고민해서는 안 되는 시대가 되었다는 것이다.

예를 들어 몇 년 전에 국내에 UGC(UGC) 라는 키워드가 생겨났고 이를 뒷받침하듯이 다양한 콘텐츠들이 사용자들에 의해서 생산되어 공유되기 시작했다. 이러한 시기 초기에서는 단지 사용자 간에 생성된 콘텐츠를 단지 공유하여 주기 위한 서비스를 고려하였고 운영함으로써 큰 문제점이 발생하지는 않았었다. 그러나 현재 UGC(UGC)를 넘어선 SNS같은 소셜네트워크(Social network) 서비스의 탄생과 스마트폰의 보급이 급속도로 확산되면서 데이터 발생을 가속화시킴으로 서비스 보다는 데이터를 수용에 따른 고민을 먼저 해야 하는 상태로 변화하게 된다. 이러한 소셜 네트워크의 데이터 발생량은 트위터의 경우 매일 7~8Tera Byte에 해당 하는 데이터를 생산해내고 페이스북은 매일 9억 개에 가까운 콘텐츠를 생산해내고 있고 있으며 학계에서는 이러한 데이터를 주목하여 이 데이터들이 분석과 통계를 바탕으로 정보화 할 경우 가치 있는 정보로 활용이 가능하다는 주장을 내 놓기 시작한 것이다. 이 말은 즉, 다양한 사용자들이 생산해내는 데이터의 대부분은 기존에는 기업이 크게 비중을 두지 않았던 비정형 데이터로 분류되지만 현재의 기업들은 이러한 복잡하고 다양한 형태의 비정형 데이터로부터 분석과 통계를 통하여 의미 있는 정보로 가공해냄으로써 기업의 의사결정이나 기업에서 활용 할 수 있는 핵심 가치를 가진 정보로 만들 수 있다고 보기 시작한 것이다.

기존 인터넷 기반의 기업들은 대부분 3C 라는 기본적인 틀을 가지고 있었다. 3C는 Contents, Community, Commerce를 일컫는 말이며 콘텐츠를 통하여 모임이 생성되고 이것이 기업의 서비스를 창출해낼 수 있게 하고 이는 곧 수익에 직결

된다는 e-Biz시대의 공식이었다. 하지만 현재의 빅데이터의 키워드의 3V는 Volume, Variety, Velocity 이다. 대용량의 복잡하고 다양한 데이터들을 실시간에 가깝도록 처리하고 분석함으로써 기업이 다양한 의사결정을 할 수 있는 정보로 활용함으로써 고객이 기업의 서비스에 대한 만족도를 분석하거나 기업의 새로운 수익을 창출 시킬 수 있다는 공식이다. 즉 3V의 목적은 비정형의 사용자의 다양한 정보들을 Value로 보아야 한다는 것이고 빅데이터의 개념은 이러한 데이터에 주목해야 할 필요가 있다는 개념인 것이다.

3. 빅데이터를 보는 관점

빅데이터는 어떠한 관점으로 보느냐에 따라서 다양한 형태로 의미를 도출해 낼 수 있다. 보통의 분류로 빅데이터는 비정형 데이터와 정형 데이터 두 가지 형태로 구분하여 볼 수 있다.

비정형 데이터의 경우 설계를 통해 구성되는 고정된 종속관계에 해당되는 데이터가 아닌 이미지나 동영상, 일반적인 텍스트들로 이루어진 데이터를 이야기하며 이는 대체적으로 일반적인 사용자들이 생산해내는 SNS나 키워드, 태그 등이 비정형 데이터라고 본다. 이런 비정형 데이터들의 구조화 수준은 대체적으로 낮다고 평가하지만 분석 대상이 되는 것임에는 확실하다. 이유는 실제로 기업이 발생하고 운영하는 정형 데이터와 달리 사용자가 생산해내고 공유하고 있는 비정형의 빅데이터가 기업 내부에서는 분석을 통하여 의사결정에 대한 훨씬 더 중요 정보로 활용될 수 있기 때문이다. 또한 빅데이터내의 약 70%는 비정형 데이터로 정형 데이터 발생량보다 더 큰 비중을 차지하고 있다.

기업들은 SNS와 같은 비정형 데이터들을 실시

간 속도의 요구와 폭발적인 데이터 증가로 부터 위협이라는 견해로 보기 보다는 다양한 기회로 보고 있다. 현재의 이러한 현상들이 빠른 데이터 처리 속도를 바탕으로 다양한 분석기법을 통해 소비자를 더욱 자세하게 분석할 수 있게끔 도와 주고 빠르고 정확한 예측이 가능하도록 하게 해 줌으로 수익성에 대한 모델을 다시 한 번 검토하게 하고 다양한 대비책 및 방안을 세울 수 있는 방법을 알게 해주기 때문에 빅데이터부터 분석한 소비자의 패턴 정보를 이용하여 다양한 기반 광고와 고객 맞춤형 서비스, 고객 만족도에 대한 정보를 수집하여 활용 할 수 있는 점이 기업으로는 빅데이터를 위기보다는 기회로 볼 수 있다라고 판단하고 있는 것이다.

정형 데이터는 실제로 기업이 운영하는 핵심 업무 내에서 발생하는 데이터이고 데이터 아키텍처 같은 설계 기법을 통하여 설계된 데이터가 주를 이룬다. 핵심 업무를 진행하는데 있어서 꼭 필요한 데이터들로 지칭한다. 빅데이터 시대에서의 정형 데이터의 범주는 M2M 같은 사물 통신 데이터로부터 센서, 유비쿼터스 기술에 적용되는 다양한 컴퓨팅 데이터를 지칭하고 실제로 유비쿼터스 환경에서 발생하는 데이터들은 개인 맞춤형 데이터의 성격이 강하여 이를 바탕으로 분석과 통계를 거쳐 가치 있는 정보로 생성해내고 사용자 및 기업의 서비스 개선 및 설계에 상당히 큰 긍정적 영향을 줄 수 있기 때문에 이런 정형 데이터에도 많은 관심을 가지고 있는 것이다.

4. 빅데이터의 기술

빅데이터 기술은 대체적으로 분석과 기반 인프라로 비교 해 볼 수 있다. 빅데이터를 기업들이 주목 하는 이유는 의미 없다고 믿었던 데이터로

부터 새로운 서비스 및 신규 사업에 대한 창출의 기회로 보기 때문이라고 설명하였다. 그렇기 위해서는 다양한 형태로 분석과 통계에 대한 제반 작업이 이루어져야만 가능하다. 그러면 이런 제반 작업을 위한 기술은 어떤 것이 있을 지 확인해 본다.

4.1 빅데이터 분석 기법

빅데이터의 분석기법은 네 가지로 구분하여 볼 수 있다.

4.1.1 Text Mining

대용량의 데이터에서 사용자가 원하는 정보를 찾아내는 프로세스를 의미하여 비정형/반정형 텍스트 데이터에서 정보를 찾아내는 기술을 말한다. 이 기법은 자연어처리에 기반으로 하고 있다. [2] 데이터 마이닝과 다른 점은 데이터 마이닝의 경우 구조화된 데이터베이스로 부터 패턴을 찾는 것이지만 텍스트 마이닝은 다양한 텍스트로부터 숨겨진 패턴을 찾고 다른 정보와의 연계성을 파악하여 카테고리 들을 찾아내는 등 이를 토대로 잠재적인 유용한 정보를 제공하기 위함으로, 컴퓨터가 인간이 사용하는 자연어를 분석하고 그 안의 정보를 추출해낸다. 이러한 텍스트 분석을 하기 위해서는 대용량 언어자원과 통계적, 규칙적 알고리즘이 사용된다. [3]

4.1.2 Opinion Mining

기존 Mining을 기술을 바탕으로 웹상의 정형/비정형의 텍스트 상에서 사용자 의견이나 선호도에 대한 정보를 추출하는 기술을 말하며 평판 분석(Sentiment Analysis)이라고 불린다. 대체적으로 긍정(Positive), 중립(Neutral), 부정(Negative)의 선호도를 판별하며 상품의 시장규모 예측이나

소문 분석 등에 활용된다.

4.1.3 Cluster Analysis

군집분석이라 불리며 비슷한 특성을 가진 개체들을 분석하여 유사 특성의 그룹을 발굴하는데 사용된다. 관심사나 취미 등에 따른 사용자를 군집 분석하여 내부에서 가치 있는 정보를 추출하는데 목적을 가진다.

4.1.4 Social Network Analytics

소셜 네트워크 연결구조 및 연결 강도 등을 바탕으로 사용자의 명성, 영향력을 측정, 소문의 중심이나 허브 역할을 하는 사용자(Influencer)를 찾는데 주로 활용된다. 이러한 사용자를 바탕으로 모니터링이나 관리를 통하여 마케팅에 활용 할 수 있는 중요 정보를 수집하는 데에 목적을 가진다.

이 외에도 기계학습, 통계 모델링, 자연어 처리, 시멘틱 분석 등... 다양한 형태의 분석 기술들이 존재하고 있으며 대시보드, 웹기술, 시계열등의 표현 기술들이 있다.

4.2 빅데이터 분석 인프라 기술

인프라 기술들은 현재 발생하고 있는 빅데이터에 대하여 어떤 방식으로 분석, 처리 할 것인지를 포함하고 있는 솔루션들이다. 대체적으로 오픈소스 기반의 솔루션들이 주를 이루지만 이 외에도 수많은 분석 인프라들이 존재한다. 대표적인 몇 가지를 가지고 예를 들어본다.

4.2.1 Hadoop

분산 시스템 상에서 대용량 데이터 처리 분석을 지원하는 오픈소스 소프트웨어 프레임워크로 정형, 비정형 빅데이터 분석에 많은 활용도를 가지고 있다. 실제로 야후나 Facebook에 사용되고

있으며 하둡 분산 파일 시스템인 HDFS(Hadoop Distributed File System), 분산 컬럼기반 데이터베이스인 (HBase), 분산컴퓨팅 지원 프레임워크인 Map Reduce가 있다. [4]

4.2.2 R

오픈소스 R 은 S언어를 기반으로 통계계산 및 시각화를 위한 언어 및 개발 환경을 제공하며, R 언어는 객체 지향 프로그래밍 언어로 일련의 데이터 처리 및 분석 작업을 대화형으로 처리 할 수 있는 장점을 가지고 있고 개발 환경을 통해 기본적인 통계 기법부터 모델링, 최신 데이터 마이닝 기법까지 구현과 개선이 가능하다. 구현한 결과는 그래프 등으로 시각화할 수 있으며, Java 나 C, Python 등의 다른 프로그래밍 언어와 연결도 용이하다. 하둡 환경 분산 처리를 지원하는 라이브러리 덕분에 구글, 페이스북, 아마존 등의 빅데이터 분석이 필요한 기업에서 대용량 데이터 통계 분석 및 데이터 마이닝을 위해 널리 사용되고 있다.

4.2.3 NoSQL

NoSQL 은 Not-Only SQL 혹은 No SQL 을 의미하며, 전통적인 관계형 데이터베이스(RDBMS)와 다르게 설계되어 Join이나 고정된 스키마를 갖지 않는 비관계형 데이터베이스를 의미한다. 대표적인 NoSQL 솔루션으로는 Cassandra, Hbase, Big Table, Dynamo, Riak, CouchDB, MongoDB 등이 존재한다.

NoSQL은 수평적 확장(Horizontal Scalability)이 용이하다는 특징을 가진다. 관계형 데이터베이스의 경우, 일관성(Consistency)과 유효성(availability)에 중점을 두고 있는 반면, NoSQL 기술은 분산 가능성(Partition tolerance)에 중점을 두고 일관성과 유효성은 보장하지 않는다. 이것은 일관성,

유효성, 분산가능성 중 2가지만 보장이 가능하다는 분산 데이터베이스 시스템 분야의 CAP 이론에 따른 것이다. [5]

따라서 NoSQL은 대규모의 유연한 데이터 처리를 위해서는 적합하지만, 안정성이 중요한 시스템에서는 관계형 데이터베이스를 채택할 필요가 있다.

이 외에도 분산 스토리지, 분산 데이터베이스, 분산 병렬처리, 분산 데이터 수집 등 다양한 형태의 인프라 기술들이 존재하고 인프라 기술들은 빠르게 변하는 소비 경향과 시장 흐름에 민첩하게 대응할 수 있는 새로운 정보 관리체계를 제시한다. 데이터 규모가 본질적으로 크기 때문에 분산처리, 확장성이 필요하여 인프라 기술들을 이용하여 기반 처리를 진행하고 데이터 마이닝을 통한 패턴분석을 통해 숨겨진 정보나 지식을 탐색해 낸다.

5. 인-메모리 컴퓨팅

5.1 인-메모리 컴퓨팅 기술

2012년 가트너에서 발표한 10대 핵심 기술에는 7위인 빅데이터(Big data)와 함께 8위에 인-메모리 컴퓨팅(in-Memory Computing)이라는 키워드를 올려놓았다.

인-메모리 컴퓨팅은 가장 핫 이슈가 되고 있는 빅데이터와 함께 가장 뜨거운 이슈로 함께 급부상하고 있는 키워드이다. 이 인-메모리 기술은 빅데이터로부터 분석되어진 정형화 데이터의 처리와 실시간 의사결정이라는 빅데이터의 또 하나의 요건을 만족시키기 위한 기반 기술 중 하나이다.

이미 기업들은 실시간 의사 결정을 위한 실시간 데이터 처리 기술에 대해 활용 가치를 높게 보고 있기 때문에 기존 디스크 기반의 데이터 저장

How Do Technology Trends Impact the Human, Business and IT Experiences?

Top 10 Strategic Technology Trends for 2012	
Human Experience	1. Media tablets and beyond
	2. Mobile-centric applications and interfaces
	3. Contextual and social user experience
Business Experience	4. Internet of Things
	5. App stores and marketplaces
	6. Next-generation analytics
IT Dept. Experience	7. Big data
	8. In-memory computing
	9. Extreme low-energy servers
	10. Cloud computing

Gartner

(그림 1) 2012 주목해야 할 10대 전략 기술 from 가트너

영역보다 훨씬 빠른 속도의 데이터 저장 및 전달 체계를 갖추는 것이 가능한 인-메모리의 영역을 관심있게 지켜보고 적용해오고 있다. 이 인-메모리 컴퓨팅은 이미 우리 환경에서 많이 사용되고 있는 기술 중 하나이다.

2012년 가트너는 소비자 장치, 엔터테인먼트 장치 및 기타 다른 내장형 IT 시스템에서 플래시 메모리 사용이 크게 증가할 것으로 내다봤다.

또 공간, 열, 성능 및 내구성에서 우위를 가진 서버에 새로운 서열의 메모리 층을 추가 하여 새로운 저장 층을 구현하는 것 외에 대용량 메모리 가용성은 새로운 애플리케이션 모델을 견인하고 있다고 강조했다. 인-메모리 애플리케이션 플랫폼에는 인-메모리 분석, 이벤트 프로세싱 플랫폼, 인-메모리 애플리케이션 서버, 인-메모리 데이터 관리 및 인-메모리 메시징 등이 포함된다.^[6]

그럼 어떻게 2012년도에 인-메모리 컴퓨팅 기술이 등장하게 된 것인지 과거로 거슬러 올라가 확인해 볼 필요가 있다.

과거 메모리 기반 기술 관련된 내용으로써는 2010년부터 2011년까지 메모리 관련 키워드가 연속으로 나타나고 있다. 2010년 플래시 메모리 (Flash Memory)라는 키워드를 8위로 내어 놓은

기록이 있다. 이는 플래시 메모리의 단가 인하로 인한 사용량 증가와 새로운 사용처가 발생됨을 예측한 결과로 풀이 되어 왔다. 이런 내용은 현재 많은 개발과 많은 학회의 관심이 되고 있는 SSD (Solid State Disk)가 범용화 되는 시점이기도 하다. 그 이후 2011년 다시 한 번 스토리지 클래스 메모리(Storage Class Memory) 라는 키워드를 9 번째 기술 분야로 올려놓았다. 이 스토리지 클래스 메모리는 플래시 메모리(Flash memory) 와 같이 비휘발성에 해당되는 속성을 제공하면서 동시에 전형적인 DRAM이나 SRAM(Static Random Access Memory)처럼 고속 바이트 단위 랜덤 접근을 지원하는 메모리 기술이다. 즉 하드디스크와 CPU 아래에 있는 DRAM의 사이의 속도적인 차이를 완화시키기 위한 구성이라고도 한다. 스토리지 클래스 메모리는 더 이상 메모리를 연산 수행의 영역으로 보지 않고 저장 매체로써의 활용이 가능한 수준임을 암시하는 내용이라고 할 수 있다. 이러한 메모리 저장 매체의 방향성은 기존 시스템의 수행 성능과 안전성 향상에 기여 할 수 있으며 시스템 입출력에 대한 성능이 개선될 수 있는 장점을 가지고 있고 기존 휘발성 메인 메모리의 한계 극복 할 수 있을 것이라는 전망을 가지고 있었다.^[7]

올해 가트너에서는 다시 한 번 메모리에 관련된 인-메모리 컴퓨팅(In-memory computing)이라는 키워드를 다시 올려 놓는다. 이 인-메모리 컴퓨팅에서는 기존의 플래시 메모리와 스토리지 클래스 메모리를 포괄적으로 포함하면서도 다양한 어플리케이션 및 분석용 데이터를 메모리를 바탕으로 구현 할 수 있음으로 실시간 의사 결정에 조금 더 가까이 갈 수 있음을 나타내며 빅데이터 환경에서 핵심 정형 데이터들을 비정형 데이터들과 함께 분석된 정보를 바탕으로 새롭게 가공하고 활용 할 수 있도록 함으로 실시간 데이터 처리 영

역과 실시간 데이터 분석, 실시간 업무 환경 구성이 가능해지도록 한 것이다..

5.2 빅데이터와 인-메모리 컴퓨팅

빅데이터에 대한 이슈들은 향후 몇 년간 키워드의 이름만 변경될 뿐 계속하여서 데이터 발생량과 데이터 분석 기법을 통하여 계속하여 이슈화 될 것으로 보고 있다. 현재도 빅데이터는 핫이슈로 계속 다양한 개념과 기술들을 생산해내고 있으며 BI, 데이터 수집을 위한 데이터 모델링, 데이터 처리 속도 향상을 위한 인-메모리 기술 등 다양한 형태로 키워드를 확장해나가고 있다. 오라클이나 SAP에서도 인-메모리 분석 기술이 적용된 어플리케이션들을 상용화 솔루션으로 출시하고 있으며 그 외로 알티베이스, IBM, 오라클, 카이로스 등 다양한 메모리 DBMS들이 각광 받고 있으며 필요성과 중요성을 인식하기 시작했다.

마이크로소프트 eScience group에 근무했던 짐 그레이는 2006년에 ‘테이프는 더 이상 사용되지 않고, 플래시가 디스크의 역할을 대신하고 RAM 영역에서 모든 일이 처리된다(Tape is Dead. Disk is Tape. Flash is Disk, RAM Locality is King)’라는 내용을 이야기 했고 ‘메인메모리 컴퓨팅’이 곧 보편화 될 것이라고 전했다.

데이터는 수집과 관리라는 용도가 있지만 가장 핵심은 처리 후 정보의 활용이라는 용도이다. 데이터는 처리라는 가공 과정을 거치면서 의미 있는 정보로 탈바꿈 하게 된다. 이러한 정보는 기업의 중요한 의사 결정의 기반지식으로 활용되며 기업의 대 고객 서비스의 생사를 논할 수 있는 정도의 가치로 변화된다. 우리가 알고, 쓰고 있던 정형화 되어 있는 데이터들은 그 나름대로의 핵심 정보를 생산해내는데 큰 비중이 있었지만 비정형 데이터를 함께 바라보아야 하는 지금 시기

에서는 우리가 핵심 정보를 추출해 낼 수 있는 범위가 극대화 되어 다가왔기 때문에 다양한 데이터 저장 영역에 대하여 심각하게 고민해야 할 필요를 가지고 있다.

현재 인-메모리 기술 중 다양한 방면에서 활용 중인 기술은 인-메모리 데이터베이스(In-Memory DBMS)이다. 인-메모리 데이터베이스는 외산 상용 DBMS에서부터 국산 상용 DBMS까지 다양한 시스템 소프트웨어들이 상용화하여 다양한 형태로 기업 내부에 적용해오고 있다. 이러한 인-메모리 데이터베이스는 기존의 하드디스크(HDD)기반의 DBMS에 비하여 빠른 성능을 보장 받을 수 있으며 트랜잭션이 빈번한 최번시간(Peak time) 시에도 안정적인 응답속도를 보장 받을 수 있기 때문에 그 활용 가치가 높다는 평가이다.

빅데이터와 같은 비정형 데이터를 인-메모리 DBMS 내부의 Temporary Table 같은 No logging Table에 데이터를 기록하고 실시간으로 분석하여 활용 가치에 따라서 다시 주기적으로 Memory Table에 데이터를 기록하면 비정형 데이터를 실시간으로 분석하고자 하는 요구와 핵심 정보를 재 저장할 수 있는 요구를 충족시킬 수 있다. Temporary Table의 경우 일반적인 Memory Table 보다 logging 자체가 없기 때문에 빠른 데이터 적재와 빠른 데이터 처리가 가능한 두 가지 장점을 함께 가져 갈 수 있기 때문이다.

빅데이터 환경에서의 또 하나의 맹점은 잦은 데이터 복제를 지양해야 한다는 점이다. 이러한 데이터 복제는 정형화 데이터에서 자주 일어나는 일 중 하나이다. 핵심 정보를 추출하기 위하여 임의의 위치에 각각의 정보를 복제하여 저장하고 그 데이터를 기반으로 핵심 정보를 가공해내는 형태는 빅데이터 시대에서는 반드시 지양해야 하는 행태이다. 그에 반면 인-메모리 데이터베이스는 물리적인 메모리 영역에서 해당 정보를 가지

고 있기 때문에 빠른 응답시간을 보장받음으로 임의의 위치에 데이터를 복제하여 분석하는 것과는 다르게 요구되어지는 시간 단위로 적재된 메모리의 데이터를 바탕으로 분석에 대한 결과 값을 도출해 낼 수 있음으로 심각한 데이터 복제에 대한 현상을 지양해낼 수 있다.

6. 결론

빅데이터의 환경과 인-메모리 컴퓨팅 기술에 대해 이야기를 해보았다. 빅데이터 환경은 이미 우리 앞에 놓여 있는 문이다. 이 문을 열고, 열지 않고는 극히 개인적이거나 기업의 역할이지만 중요한 것은 그 문을 열게 됨과 동시에 많은 고민과 투자가 이루어져야만 궁극적으로 원하는 정보를 손에 얻음으로 그 앞에 놓인 다음 문 앞에까지 나아갈 수 있게 된다. 늘 그래왔듯이 기술은 개념을 포함한다는 논쟁이나 개념이 기술을 포함한다는 논쟁이 줄을 잇고 이야기 되어 오고 있다. 어느 MC의 말처럼 모든 사람들이 옆 사람과 손을 맞잡는 다면 좌도 없고 우도 없게 되며 오른손을 들 때에 왼손이 들리고 왼손을 들 때에 다른 사람의 오른손을 들어주는 것처럼 개념과 기술은 포괄적으로 함께 고민 되어야 한다. 빅데이터 환경은 날이 갈수록 더욱 심각하게 우리에게 다가 올 것이다. 이유는 간단하다. 지금의 현상이 아주 잘 이야기 해주고 있듯이 빅데이터 환경에서 실시간 의사 결정을 위한 정보를 추출해 내는 것은 힘든 일이지만 그만큼 가치 있고 분명한 역할을 하는 것임에는 틀림없기 때문에 기업은 끊임없이 새로운 가치의 정보를 찾아내고자 끝없이 노력할 것이다. 예를 들어 스마트폰 내에 사람의 생체 리듬을 감지하는 기능이 있다고 생각해보자. 그리고 사용자의 생체 정보를 실시간으로 중앙 서버에 전달해주고 문제가 있을 것이란 사용자의 정

보를 바탕으로 의료 서비스를 구현한다고 데이터량은 얼마나 더 증가하게 될까? 현재 급속도로 확산되고 있는 M2M(Machine to Machine)이란 서비스도 앞으로는 모든 사물간의 통신으로 인해 중요 데이터를 발생시키는 주체가 많아질 것이고 이 데이터 또한 수용 한다고 하면 어떤 결과가 발생할까? 앞으로는 데이터는 저장이라는 개념을 벗어나야 한다.

실시간 의사 결정은 데이터를 저장한다라는 주제로부터 시작해서는 안 된다. 이 이상 디스크는 핵심 데이터에 대한 저장 매체가 아니고 30년 전에 보아왔던 드럼이나 테이프를 보아야 한다. 인메모리 기술이 가트너의 2012 전략 기술로 발표된 이유는 그러한 개념을 먼저 깨우쳐야 한다는 이야기라고 생각한다. 메모리 영역은 더 이상 연산을 위한 공간이 아닌 저장을 위한 공간으로 개념이 바뀌어져야 한다. 인-메모리 데이터베이스부터 인-메모리 분석까지 다양하게 인-메모리 기술은 차츰 차츰 현실과 밀접해 질 것이다. 그리고 다른 하나의 견해로 빅데이터와 함께 이슈화 되는 논쟁거리는 개인의 프라이버시에 대한 침해 문제이다.

빅데이터는 간혹 개인으로써는 독이 될 것이고 기업으로써는 늘 기회가 될 것이라고 한다. 개인은 IT 발전에 따라서 늘 개인정보가 언제 어디서든 보여 것이고 기업은 그 기회를 이용하여 인간의 뇌 속의 생각까지 들어가기 위하여 노력 할 것이다. CCTV처럼 내가 언제 어디를 지나가고 있는 지 항상 내가 촬영되고 있는 것처럼 내가 먼저 인식하기 전에 기업이 먼저 인식 할 것이고 언제 어디에 위치하여 있던 시간에 내가 필요한 정보, 내가 원하는 것들이 눈앞에 펼쳐 질 것이다.^[8]

유비쿼터스 사회에서는 인간의 모든 것을 IT 환경이 대신 한다. 늘 환경의 변화가 그러했듯이 장점을 가져 갈 수 있다면 단점을 함께 수용해야

한다는 점은 예나 지금이나 틀리지 않다. 하나의 바람은 모든 환경이 인간을 위하여 구현된다고 하지만 그 환경이 인간을 침해하지는 않았으면 하는 바람을 가져본다.

참 고 문 헌

- [1] worldwide Disk Storage system Forecast 2005~2010.
- [2] IDC korea 2007
- [3] What Is Text Mining?, Marti Hearst, New York Times 2003
- [4] Hadoop <http://wiki.apache.org/hadoop/ProjectDescription>
- [5] Availability and Partition Tolerance <http://pl.atyp.us/wordpress/index.php/2009/11/availability-and-partition-tolerance>
- [6] 가트너 “2012년 이끝 10대 IT 트렌드는?” http://www.zdnet.co.kr/news/news_view.asp?artice_id=20111020112655
- [7] Rich Freitas and Winfried Wilcke, “Storage Class Memory, the next storage system technology,” 52(4/5), 439, IBM Journal of Research and Development, (2008).
- [8] Privacy and Big Data [The Players, Regulators, and Stakeholders] By Terence Craig, Mary E. Ludloff / Publisher: O'Reilly Media / Released: September 2011.

저 자 약 력



박 호 진

이메일 : p.hojin@gmail.com

- 2007 성균관대학교 컴퓨터 공학 석사
- 현 : (주)알티베이스 Technical Consulting Team 책임컨설턴트
- (주)알티베이스 국내사업본부 통신사업팀 차장
- 관심분야 : NVRAM, MMDBMS, DBMS, DA, DW, Data Stream