

IDG Tech Focus

빅데이터와 DW의 조합, 하이브리드 DW 성공 가이드

최근 비정형 데이터뿐만 아니라 기존에 관리해 왔던 데이터 역시 폭발적으로 증가하고 있다. 이에 DW의 용량 및 성능 이슈가 발생하고 있으며, 비용 측면에서도 기업의 기간계와 정보계 데이터를 처리하기에 곤란한 상황에 이르렀다. 하이브리드 DW라 불리는 기존 DW와 오픈소스 기반 빅데이터 기술의 조합은 이런 문제들을 해소하기에 충분한 솔루션을 제공한다. 하이브리드 DW를 알아보고, 이를 성공적으로 구축하기 위한 방안들을 살펴보자.

❖ Tech Trends

DW의 한계를 극복하려는 노력의 산물, '하이브리드 DW'
빅데이터 시작 전, 꼭 필요한 질문 4가지

❖ Solutions

하이브리드 DW, 데이터 르네상스 시대를 대비하는 기업의 필수 전략

Sponsored by

kt NexR

무단 전재 재배포 금지

본 PDF 문서는 IDG Korea의 프리미엄 회원에게 제공하는 문서로, 저작권법의 보호를 받습니다.
IDG Korea의 허락 없이 PDF 문서를 온라인 사이트 등에 무단 게재, 전재하거나 유포할 수 없습니다.

Tech Trends

DW의 한계를 극복하려는 노력의 산물, ‘하이브리드 DW’

이대영 기자 | ITWorld

기업의 데이터 증가는 이미 예견된 일이었다. 하지만 최근 스마트폰 등 모바일 환경으로의 변화와 SNS의 진화, 사물인터넷 환경의 성숙 등으로 데이터는 예상보다 급격히 증가하고 있으며, 기존의 데이터 관리 기법으로는 도저히 감당하지 못할 상황에 이르렀다.

이렇듯 기존 방식으로는 수용 및 관리하지 못하는 데이터를 통칭해 빅데이터라 부르는데, 이 속에서 가치 있는 정보를 찾아 활용하려는 움직임이 전세계적으로 진행되고 있다. 빅데이터 속에 기업이 원하는 가치가 담겨 있다는 것이다.

비즈니스의 빠른 변화와 데이터 폭증에서 비롯된 DW의 한계

데이터 속에서 가치를 찾으려는 노력은 십 수년 전부터 진행되어 왔다. 그 당시에는 DW(Data Warehouse)가 해법의 터전 역할을 맡았다.

DW는 DB를 활용해 의사결정 프로세스를 지원하도록 데이터를 주제 중심으로 통합해 모아놓은 것을 의미한다. 한 마디로 새로운 형태의 통합 데이터 저장소(Repository)였다. 이러한 특징으로 1990년대 중반부터 DW는 기업 내 여러 곳에서 분산 운영되는 트랜잭션 시스템에서 필요한 정보를 추출해 중앙 집중화된 저장소에 모아놓고, 여러 계층의 사용자들이 좀 더 손쉽게 접근해 활용하기 위한 목적으로 구축됐다.

DW는 기업의 정보계와 분석계 시스템 등을 담당하면서 기업의 의사결정과 판단을 돕기 위한 핵심 데이터를 저장하는 필수 인프라로 자리잡았다. 그러나 다

양한 업무 시스템에서 발생한 데이터를 통합 및 저장하고, 이를 활용해 비즈니스를 지원하는 업무가 적시에 수행되도록 하는 것은 쉽지 않은 과제였다.

시간이 지남에 따라 나날이 증가하는 데이터로 인해 DW는 기대와는 달리 만족할 만한 성능과 확장성을 제공하지 못했다. 특히 급변하는 비즈니스의 속도에 맞추어 필요한 데이터를 적시에 제공하기 어려웠다. 이는 DW가 각 트랜잭션 시스템에서 필요한 데이터만을 추출하고 저장했기 때문이다.

또한 데이터 양이 급증한 것과 달리 실제 기업들에서의 데이터 활용은 매우 저조하다는 것이다. 실제로

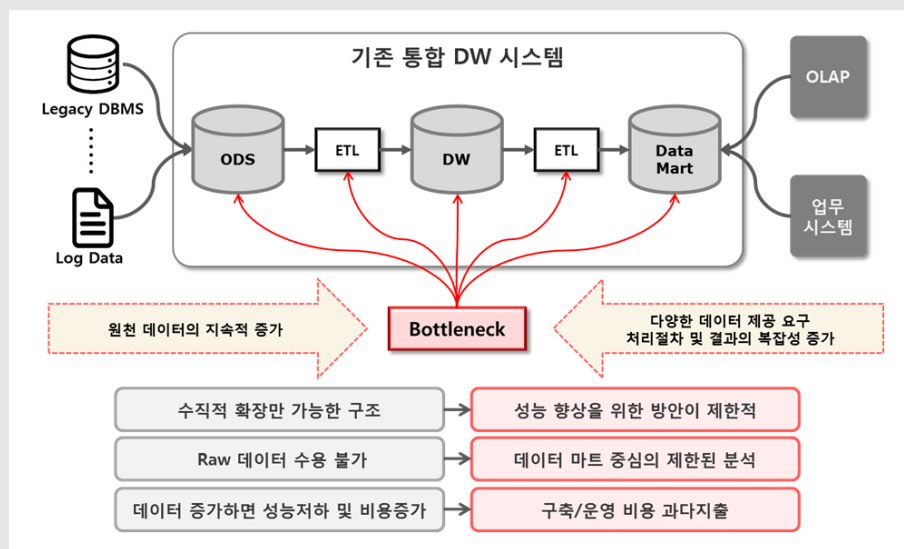


오래 전부터 많은 기업들이 BI/DW 환경을 구축하고 운영했음에도 분석을 위한 데이터 활용도가 전체 데이터의 0.5%밖에 안 된다는 것이 현실이다.

또한 증가하는 데이터 용량에 따른 성능 이슈를 해결하기 위해 추가적으로 고가의 솔루션을 증설함에 따라 과도한 비용이 추가적으로 발생했을 뿐만 아니라 이에 따른 기대효과도 매우 낮아 DW에 대한 회의적인 시각도 존재했다. 가트너는 오는 2016년까지 기존 DW의 75%가 무용지물이 될 것이라고 전망하면서 기존 DW의 한계를 지적하기도 했다.

IDC의 최근 보고서에 따르면, 기업의 40% 가량은 데이터 볼륨이 매년 50%씩 급증하고 있는 반면, DW는 규모 면에서 매년 18%만 증가한다고 밝혔다. 이는 취급해야 할 데이터 용량이 급증하고 이에 따른 DW 구축 요구가 커지고 있지만 실제로 기업들은 이에 대응하지 못하고 있음을 의미한다. 기존에 활용되지 않던 비정형 데이터에 대한 분석 요구뿐만 아니라 전통적으로 활용했던 정형 데이터 관리에 있어서도 감당할 수 없는 근본적인 한계에 다다른 것이다.

그림 1 | 상용 DW 기반 분석 시스템의 구조와 한계



DW 어플라이언스, “저장한 만큼 늘어나는 비용”

기존 DW의 한계를 극복하고 폭증하는 데이터를 해결하기 위한 방안으로 등장한 것이 바로 DW 어플라이언스다. 오라클, EMC, IBM 등 글로벌 업체들이 앞서 언급한 전통적인 DW의 한계를 극복하기 위해 HW와 SW를 통합하여 성능을 극대화한 DW 어플라이언스를 내놓기 시작했다. 근래 들어서는 새로운 빅데이터까지 수용하기 위해 자사의 DW 어플라이언스에 하둡(Hadoop) 기술을 적용한 제품까지 선보이기까지 했다. 그리고 저마다 빅데이터의 시대에 시장에 가장 적합하고 고성능의 기술을 탑재한 DW 솔루션임을 강조하고 있다. DW 어플라이언스를 쉽게 설명하면 DBMS와 서버, 스토리지 등을 구조적으로 통합한 것으로, 데이터 처리 성능을 최대화할 수 있도록 HW와 SW를 최적화한 것이다.



DW 어플라이언스는 폭증하는 데이터를 저장하고 이를 좀 더 빠르게 처리하기 위해 MPP(Massive Parallel Processing) 방식이 적용됐다. MPP는 컴퓨팅 성능이 조금 떨어지더라도 저렴한 서버들을 고속 네트워크로 엮어 서버 간 병렬처리를 최대한 유도함으로써 처리 성능을 최대한 향상시킨 것이다. 이 방식은 DW의 데이터 처리 패턴을 아주 효율적으로 유지함으로써 서버 비용을 낮추고도 만족할 만한 처리 성능을 기대할 수 있다. 특히 시스템 확장성이 높아 지속적으로 늘어나는

대용량 데이터의 수용과 분석처리 시스템의 유지관리에도 좋은 성능을 보인다.

그러나 DW 어플라이언스는 저장 용량이 증가할수록 증설 비용이 비싸지는 단점(소프트웨어 라이선스 및 유지보수 비용 증대)을 갖고 있다. 이로 인해 기업 환경에서의 데이터가 지속적으로 증가하는 것과 함께 DW 투자비용도 급격하게 증가하고 있다.

실제 자체적으로 분석한 결과를 보면 1TB 당 가격으로 5년간 TCO를 비교했을 경우 거의 2배까지의 비용이 더 들어가는 것을 확인할 수 있었다. 따라서 DW 어플라이언스가 좋은 성능을 제공하는 것은 분명하지만, 대부분의 기업들은 차세대 DW를 구축하는 데 있어 투자할 수 있는 예산이 한정적일 수 밖에 없기 때문에 부담스러울 수 밖에 없다. 결국 비용 효율적인 측면에서 DW 혁신을 가능케 하는 최적화된 솔루션이 무엇인지 파악할 필요성이 생길 수 밖에 없는 시기가 도래한 것이다.

20년만의 패러다임 변화, 하이브리드 DW

이런 상황에서 구원 투수로 등장한 것이 바로 하이브리드(Hybrid) DW다. 하이브리드 DW는 기존 DW의 문제점과 한계를 오픈소스 기반의 빅데이터 기술을 통해 극복한다는 방식이다. 대용량 및 비정형 데이터를 위한 저장, 처리 기술인 빅데이터 기술을 활용해 그 동안 처리, 저장, 분석하지 못했던 데이터뿐만 아니라 기존 정형 데이터의 고질적인 문제점인 확장성과 고비용 문제를 개선할 수 있다는 것이다.

예를 들어 HP는 한 보고서를 통해 기존 관계형 데이터베이스에서 1TB의 데이터를 1년간 저장하는 비용이 3만 7,000달러, DW 어플라이언스의 경우 5,000달러가 소요됐지만, 하둡의 경우 2,000달러만으로 가능한 것으로 추정했다. 또한 kt NexR 측은 1PB 당 5년 TCO 기준으로 1TB 당 가격을 따져봤을 때 DW 어플라이언스는 4,900만 원, 전통적인 스토리지는 2,800만 원, 그리고 하이브리드 DW는 1,300만 원이라고 계산했다.

하이브리드 DW의 비용절감에 대해서는 이미 KT BIT 경영인프라혁신담당

BI/DW 구축팀에서 진행한 가입자 분석 시스템 구축 프로젝트에서 실제로 검증된 바 있다.

통신업체에서 가장 중요한 데이터 가운데 하나가 고객들이 유선전화나 휴대폰으로 통화할 때마다 통화위치와 대상, 시간 등을 로그 데이터로 기록하는 CDR(Call Detail Record)이다. 그런데 스마트폰의 등장 이후 무선 데이터 로그가 급증하고, 데이터 프로세스 전 구간에서 병목현상이 발생했으며, 시스템 성능 저하와 분석 업무에도 차질을 빚게 됐다.

KT는 초기에 고가의 유닉스 서버 대신 x86 서버를 사용하는 하둡을 선택했지만 호환성과 연동성 측면에서 어려움이 많았고, 온라인 분석처리(OLAP)처럼 빈번하고 빠른 응답성을 요구하는 쿼리를 효과적으로 지원하지 못한다는 단점이 발견돼 kt NexR의 NDAP 기반 하이브리드 DW 모델을 채택했다.

KT의 가입자 분석 시스템은 용량이 큰 원시 데이터와 계산량이 많은 배치 작업은 하둡에서, 데이터 마트처럼 작고 중요한 데이터와 BI 도구 연동 작업은 기존 DW에서 처리하는 하이브리드 DW 모델의 대표적인 사례가 됐다.

이는 스마트폰 등장 이후 전통적인 DW로는 급증하는 데이터 로그를 처리하기 어렵다는 점, 하둡의 경우 뛰어난 경제성에도 불구하고 지원하는 인터페이스가 기존 데이터베이스 언어인 SQL과 다르며, 주요 BI 도구와의 연동도 쉽지 않다는 단점을 극복한 것이다.

이를 통해 KT는 가입자 분석 시스템의 성능개선과 함께 RDBMS 기반의 아키텍처와 비교했을 때 5년 TCO 기준으로 약 500억 원의 비용절감 효과를 얻었다. 또한 성능개선 부문에서는 선형적 배치를 통해 1.5배의 개선을 이뤘으며, ETL

하이브리드 자동차와 DW의 공통점

하이브리드라는 단어는 원래 이질적인 요소가 서로 섞인 것으로 혼합과 혼성이라는 의미인데, 넓게는 기존 기술과 서비스에 이질적인 기술의 조합으로 새로운 형태의 기술 및 서비스를 창조한다는 의미를 포함하고 있다.

하이브리드의 대표적인 사례는 자동차다. 기존 가솔린 엔진의 자동차와 전기 엔진이 함께 들어있는 차량인 하이브리드 자동차는 가솔린 엔진의 사용자 경험을 기반으로 전기 엔진의 장점을 받아들인 중간단계의 자동차다.

하이브리드 자동차의 공통된 특징

은 고속 주행시 가솔린 엔진을 가동시켜 힘을 받아 전기를 충전하고, 저속 주행시 전기에너지를 활용해 주행한다는 원리다. 하이브리드 자동차의 가장 큰 장점은 바로 연비와 저공해라는 점이다. 가솔린 엔진 기반의 자동차로는 따라올 수 없는 연비와 함께 배출가스 또한 줄일 수 있다.

기존 DW 기술에 빅데이터 기술을 접목한 하이브리드 DW는 하이브리드 자동차와 유사한 장단점을 갖고 있다. 하이브리드 DW의 비용절감 효과는 하이브리드 자동차의 장점과 확실히

맞닿아 있다. 기존 DW 장비를 주로 사용하면서 보조적인 차원에서 빅데이터 기술인 하둡을 사용한다는 점 또한 유사하다.

하이브리드 DW는 고가의 DW 어플라이언스로는 저장하지 못했던 정형 데이터와 비정형 데이터를 모두 처리 및 저장하고 이를 분석함으로써 얻을 수 있는 비즈니스 통찰력이야말로 십수년 전, DW를 처음 도입했을 때부터 기업이 염원하던 본연의 목적 자체다.


병목현상과 BI 성능을 개선했다.

가까운 곳에서 찾자

현업의 빅데이터를 통한 통찰력 찾기는 무엇보다 중요하다. 현재의 빅데이터 이해 수준에서 데이터를 가장 잘 이해하고 기술적으로 잘 활용할 수 있는 곳은 바로 IT부서다. 또한 기업에서 빅데이터가 가장 많은 곳이 전산실이다.

빅데이터의 가치 찾기는 현실적이고 가까운 곳에 있다. 예를 들어 전산실에는 제대로 처리하지 못하고 그냥 지우기만 하는 어마어마한 로그들이 있다. 이 로그 데이터들은 지금껏 관리되지 못하고 버려져 왔었다. 이 로그 데이터는 저장하기에는 비용이 너무 많이 들었기 때문이다.

현재 일부 기업에서는 이렇게 버려지는 로그 데이터를 저장하고 분석해 새로운 가치를 창출하고 있다. 대표적인 예가 로그 분석을 통한 보안 관제와 서버 관리다. 빅데이터 분석을 통한 보안 관제와 서버 관리는 이미 그 효과가 증명된 빅데이터 활용의 주요 분야가 됐다. 이는 빅데이터 통찰력을 찾는 이들에게 많은 시사점을 안겨준다.

마찬가지로 DW는 이미 비즈니스 통찰력을 찾기 위해 십 수년 전부터 기업이 보유한 데이터 인프라다. 예전부터 의사 결정 프로세스에 있어 중추적인 인프라로 이미 가치 있는 데이터라고 인증 받은 셈이다. 하이브리드 DW는 이 인프라에 기존에 관리하지 못했던 정형 데이터와 비정형 데이터를 모두 함께 관리하는 것이다. 하이브리드 DW가 중요한 이유는 지금 당장 빅데이터를 포함한 모든 데이터에서 가치를 찾아낼 수 있다는 것이다. 나아가 다양한 데이터를 요구하는 현업 사용자들이 비즈니스 통찰력을 얻을 수 있도록 하는 것이야말로 본연의 목표를 달성하는 것이다. 이 지점이 바로 하이브리드 DW가 필요한 곳이다. 

Global IT Standard IDG

PC World, Computer World, CIO 등으로 잘 알려진 IDG는 90여 개국에서 180여 미디어를 발행하는 글로벌 테크놀로지 미디어로, 전 세계에 1억 4,000만 명의 독자를 대상으로 미디어, 리서치, 컨퍼런스, 이벤트 등 다양한 테크놀로지 관련 서비스를 제공하고 있습니다.



한국IDG(주) 서울시 중구 봉래동 1가 108번지 창화빌딩 4층 100-161

Tel : 02-558-6950 Fax : 02-558-6955 www.idg.co.kr twitter.com/ITWorldKR www.facebook.com/IDGKorea

Tech Trends

빅데이터 시작 전, 꼭 필요한 질문 4가지

Reda Chouffani | CIO



빅데이터는 비즈니스 프로세스를 개선하는 방법을 찾고 있는 기업들에게 많은 가능성을 제시해줄 것이다. 그 전에 어떤 데이터를 필요로 하고, 어떻게 그 데이터를 확보해 사용할 것이며, 거기서 어떤 시사점을 찾을 지를 알아야 한다. 이 글에서는 이런 고민들을 어떻게 풀어야 할 지 알아보자.

여러 설문조사에서 많은 경영진들이 최근 주요 프로젝트로 빅데이터 도입을 지목한 것으로 나타났다. 빅데이터는 업무의 능률을 올리고 간소화 시키며, 소비자 구매성향에 대한 통찰력을 준다. 의료 산업에서는 생명을 구

하는 데도 도움을 줄 수 있다는 것을 고려하면 이러한 경영진들의 계획은 그다지 놀라운 것은 아니다. 그러나 빅데이터에 대해 몇 가지 짚고 넘어가야 할 것들이 있다.

- 어떤 데이터를 고려해야 하나?
- 데이터를 어떻게 수집하나?
- 빅데이터 도입이 회사에 어떤 실질적 혜택을 주나?
- 빅데이터의 투자대비 성과는 어떻게 측정하나?

경영진들 머리 속에는 이외에도 수많은 질문들이 떠오를 것이다. 하지만 이 네 가지 질문이 가장 빈번히 거론되고 있기에 이에 대한 해답을 소개하고자 한다.

어떤 데이터를 고려해야 하나?

데이터에는 정형, 반정형, 비정형, 이 세 가지 형태가 있다. 정형 데이터는 컴퓨터와 인간 모두 읽을 수 있는 데이터다. 가장 대표적인 예로 관계형 데이터베이스(relational database, RDB)가 있다. 반정형 데이터에는 XML, 이메일, EDI 등이 포함되는데, 정형화되어 있지 않지만 시맨틱(semantic) 요소들을 분리하는 태그들을 가지고 있다. 마지막으로 비정형 데이터는 이미지, 오디오, 비

디오 등의 데이터베이스에 들어가지 않는 데이터다.

데이터의 접근제한을 풀고 접속권한을 제공해 데이터를 저장하고 사용할 수 있게 하는 것이 가장 중요한 관건이다. 그러면 정보를 미가공(raw) 형태로 남길 수 있게 되어, 분석 시스템으로 실시간으로 정보를 분석하고 보고할 수 있다.

정형 데이터에서는 이 과정이 상당히 직관적으로 일어난다. 이와 반대로, 비정형 데이터는 고급 알고리즘과 강력한 엔진을 반드시 거친 후 들어오는 데이터만을 처리할 수 있다.

데이터를 어떻게 수집하나?

세상에는 셀 수 없이 많은 데이터 소스에서, 역시 셀 수 없이 많은 유형의 데이터를 제공한다. 궁극적으로는 데이터 조합을 수집해야 한다는 결론에 이르게 된다.

요즘 기업들이 고객과 브랜드 선호에 대한 정보를 얻기 위해 사용하는 데이터 소스가 바로 소셜 미디어다. 페이스북, 트위터 등의 주요 소셜 미디어 사이트들은 모두 API를 통해 일종의 데이터 접속 기능을 제공한다.

다음으로 중요한 데이터 소스는 위치와 이동 패턴(movement patterns)에 존재한다. RFID, 적외선, 무선 기술 등이 점점 소형화되고 단가도 낮아지면서 기업들의 적절한 비즈니스 애플리케이션에 이들의 위치를 보고하는 자원, 직원, 고객들이 점점 증가할 것이다.

기업들은 이런 소스에서 수집한 데이터를 클라우드와 웹에 저장돼 있는 정형 및 비정형 데이터들과 결합시킬 수 있다. 이 때 기업들은 이 데이터들을 반드시 가치 있는 방식으로 사용할 수 있도록 유념해야 한다.

빅데이터 도입이 어떤 실질적 혜택을 주나?

오늘날 대부분의 업종에서 빅데이터 도입이 필요한 대다수의 정보는 지금도 활용할 수 있다. 하지만 몇몇의 경우 대용량 데이터가 아니거나 데이터 표준화가 이뤄지지 않았다. 많은 기업들은 각각의 업무용 애플리케이션에서 데이터를 수집 및 추출하고 분석에 활용할 수 있는 적절한 플랫폼을 신속히 도입해야 하는 과제를 안고 있다.

현재 마케팅 기업들이나 내부 마케팅 부서는 우선 더 많은 리드(lead)를 발생시키고, 그 후에 정확히 마케팅 캠페인의 효과를 측정하는데 있어 데이터를 활용해야 하는 어려움에 직면해 있다.

많은 고객들이 다양한 소셜 미디어 사이트에서 브랜드를 평가하기 때문에, 웹에서 트위터의 트윗, 사용 후기, 페이스북의 좋아요(Likes) 등을 추적하는 일은 아주 어려울 수 있다. 그러나 많은 소셜 사이트들이 데이터를 볼 수 있는 기능을 제공하기 때문에, 조직들은 빅데이터를 이용해 고객들의 생각과 실시간 분석을 얻어낼 수 있다.

특히 유통 기업이 유용하게 사용할 수 있는 수많은 온라인 데이터 소스도 있

다. 이들 가운데 상당수는 고객의 둘러보는 행동과 전체적인 브랜드 이미지에 관련 있다. 앞서 말한 바와 같이, 데이터는 소셜 미디어 서비스로부터 API를 사용하거나, 구글과 웹 서버 로그를 사용하여 추출할 수 있다. 이에 더해, 많은 소매상들은 매장 안에서 데이터를 수집하기도 한다. 실제 쇼핑카트를 추적하거나 고객 보상카드를 읽어 고객들의 쇼핑 패턴을 모니터링할 수 있다.

의료 산업에서는 전자의료기록 시스템에 저장된 데이터를 활용할 수 있다. 거기에다 미국의 개인정보정보보호 관련 법안(HITECH Act)과 같은 최근 미 정부가 추진하고 있는 경기 부양책들과 의료개혁은 의사, 경영진, 의료 관계자 모두에게 정형 혹은 반정형 데이터들의 활용을 돕는 의료정보 기술 도입에 재정적인 인센티브를 부여한다.

마지막으로, 웨어러블 의료기기와 모바일 의료 애플리케이션들의 인기가 높아지고 있다. 기기와 앱 모두 의료기관들이 환자 서비스 향상을 목적으로 수집과 이용을 준비 중인 데이터의 지속적인 흐름을 유발한다.

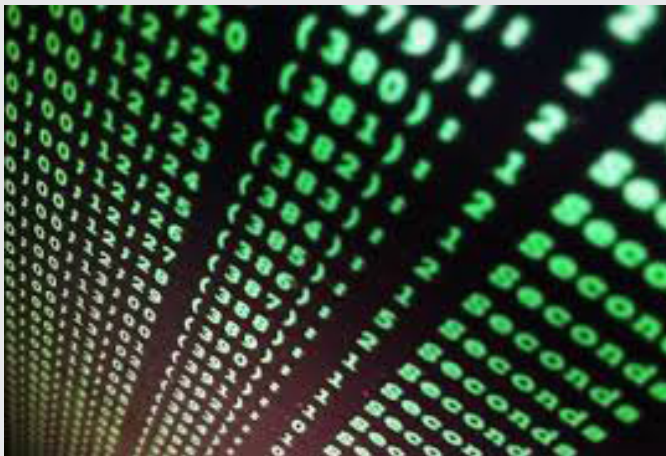
의료 연구의 발전으로 제약사들은 환자를 위한 복제약 임상실험 목표를 높일 수 있었다. 이 변화는 과학자들이 이용할 수 있는 압도적인 데이터의 규모와 인간 유전자 데이터에 주석 달기(Annotated Human Genome Data) 등과 같은 국제적인 데이터 공유 운동으로 걸음을 같이 했다.

공급망관리(SCM)와 물품통제 그리고 GPS 기술, 공급자와 운송 업체로부터 날아온 EDI 메시지, 물품 운반대(pallets)와 케이스, 고객 정보를 담은 모바일 기기, 내부 ERP 시스템, 소셜 미디어 등의 수많은 소스에서 수집한 빅데이터를 다뤄야 하는 물류 업계에서도 빅데이터는 중요한 통찰력을 제공하고, 조직의 재충전을 지원한다.

시티그룹 같은 금융 업체의 경우에는 이미 빅데이터 도입을 시작했다. 여기서 주안점은 사기를 감소시키고, 보통 데이터 세트 안에 숨겨져 있는 패턴들을 발견하며, 돈세탁 같은 범죄 행위를 감지해 미국과 세계의 은행법 준수를 돕는데 맞춰져 있다.

미국 오바마 대통령의 최근 지시사항은 연방정부 내의 빅데이터 지원을 목

표로 하고 있다. 모바일 기기를 사용해 데이터 접속을 가능하게 하는 등 미국 국민들이 정부 서비스와 정보에 대한 접근성을 향상시키는 것이 핵심 목표다. 연방 정부 CIO 스티븐 반로예클은 백악관의 공공기관과 민간기업에 무료 데이터 접속권한을 제공할 계획인 디지털 정부 로드맵(Roadmap for a Digital Government)을 감독할 예정이다. 이를 통해 많은 기업들은 자사에 빅데이터를 도입하고 발전시킬 수 있을 것이다. 물론, 여기에서 다루지 않은 다른 많은 산업들도 변화와 혁신



을 주도하기 위해 빅데이터의 혜택을 누릴 수 있다.

빅데이터의 투자 대비 성과는 어떻게 측정하나?

일부 IT업체들이 빅데이터의 즉각적인 투자수익률을 내세우고 있다. 하지만 그들이 접속권을 가지고 있는 데이터의 여러 측면들과 ROI가 어떤 가치를 제공할 수 있는지 파악해 볼 필요가 있다.


대부분의 ROI 시나리오에서 비용효율 분석과 IT업체가 제안하는 솔루션을 연결시켜야 한다. 빅데이터 도입 사례에서 보면, 측정할 수 없는 몇 가지가 있다. 고객은 비즈니스 프로세스에 잠재적 변화를 줄 만한 통찰력을 발견하려는 의도로 대규모 데이터를 간주하기 때문에 이를 통해 얻을 수 있는 가치를 예측하기가 어려워지는 것이다.

빅데이터 투자의 함정에 빠지지 않는 방법

빅데이터 투자의 함정에 빠지지 않으려면 두 가지의 경험 법칙을 명심해야 한다.

빅데이터 기술은 확장성이 매우 좋기 때문에, 빅데이터 도입에 비용을 지출하더라도 이에 따른 데이터 규모의 증가는 잘 일어나지 않는다. 그리고 비록 데이터가 여러 가지 형태(정형, 반정형, 비정형)를 가지고 지속적인 성장률을 보이며 들어오더라도, 하둡을 지원하는 플랫폼들의 통합과 관리는 전통적인 데이터베이스 관리 시스템에 비해 훨씬 비용효율이 높다. 새로운 솔루션은 오픈소스 SW를 구동하는 HW에서도 구동할 수 있기 때문이다.

모든 새로운 트렌드가 그러하듯이 빅데이터를 둘러싼 과대 광고들이 많다. 과거부터 많은 기업들은 데이터 기반의 의사 결정을 위해 데이터 분석용 전사 DW를 운영해왔다.

오늘날 우리 모두 알다시피, 빅데이터는 이전의 사일로(Silo)화된 DW 사용자들이 이용할 수 없었던 소셜 미디어 행동과 같은 추가적인 정보와 결합해 과거보다 훨씬 적은 비용으로 데이터를 클라우드 내에 저장하고 관리할 수 있게 되었다. 그 데이터를 가장 잘 활용하려면 빅데이터가 자사에 의미하는 바와 기업이 향후(혹은 첫) 빅데이터 도입에서 얻고자 하는 것을 반드시 알고 있는 것이 무엇보다도 중요하다.  ITWORLD

Solutions

하이브리드 DW, 데이터 르네상스 시대를 대비하는 기업의 필수 전략

최성호 부장 | kt NexR 기술컨설팅팀

오늘 폰소스 기반의 빅데이터 기술과 범용적인 x86 서버의 확산으로 인해 이제 모든 데이터를 활용할 수 있는 데이터 르네상스 시대가 도래했으며, 비즈니스 환경도 점차 데이터 중심으로 이동하고 있다. 이에 많은 기업들은 기존 DW를 포함한 IT인프라를 보호하면서도 기존에 활용되지 않았던 데이터도 수용하는 차세대 DW에 대한 고민이 높아지고 있다. 데이터 중심의 새로운 IT인프라를 제공하는 가장 경제적인 방법 중 하나가 바로 kt NexR이 제시하는 ‘하이브리드 DW’이다.

DW의 패러다임 전환이 ‘진행 중’

지난 2000년 이후 많은 국내외 기업들이 자사 정보시스템의 핵심 인프라로 관계형 데이터베이스를 기반으로 한 DW 시스템을 구축 및 운영해 왔다. 하지만 실제 기대와는 다르게 빠르게 진화하는 비즈니스 환경에 적절하게 대응하지 못하거나 지속적으로 증가하는 데이터 양을 효과적으로 분석하기 위해 적합한 인프라를 제공하지 못했던 것이 현실이었다.

이에 대한 대안으로 지난 몇 년간에는 어플라이언스 기반의 DW 시스템이 주목받았으며, 이를 통한 DW 고도화 사업을 진행하는 경우도 있었다. 물론 이런 방식을 통해 전통적인 DW 시스템의 데이터 관리와 분석의 문제점들이 다소 해결되기도 했지만 어플라이언스 제품을 도입하는 데는 많은 비용과 시간이 소요됐다. 더불어 특정 글로벌 업체의 종속적인 DB 솔루션만을 선택할 수 밖에 없

는 한계도 있었다.

최근 ‘빅데이터’라는 새로운 IT의 흐름이 기사화되면서 DW 분야에서도 패러다임 전환이 진행 중이다. 이제 기업들은 기존 정형 데이터뿐만 아니라 이전에는 활용되지 않았던 다양하고 엄청난 양의 비정형 데이터도 이용해 새로운 비즈니스 기회를 창출하고, 기업의 경쟁력을 강화하려는 추세이기 때문이다.

이와 같은 빅데이터를 활용한 비즈니스의 혁신을 위해 차세대 DW에 대한 관심이 고조되고 있다. 기업들은 활용 가능한 모든 형태의 데이터에





대한 실시간 및 다차원 분석을 할 수 있는 DW 시스템을 필요로 하고 있다. 더불어 차세대 DW를 구축할 때 갈수록 불확실해지는 경제 상황에서 비용효율 측면도 고려하지 않을 수 없는 중요한 요소로 부각되고 있다.

즉, 빅데이터까지 수용하는 차세대 DW의 요건을 충족하면서도 비용효율적으로 이를 구축 및 운영할 수 있는 대안이 필요하게 된 시점이 도래한 것이다. 그리고 이를 가능케 하는 것이 바로 경제적인 오픈소스 SW 기술을 활용한 ‘하이브리드 DW’라 할 수 있다.

데이터 중심의 IT 인프라 구축 시 경제적인 기술로써 하이브리드 DW

몇 년 전부터 빅데이터가 기사화 되기 시작하고 글로벌 성공사례가 나타나면서 국내 또한 지난 2013년을 기점으로 기업들이 적극적으로 검토를 시작했다. 당시 검토의 목적은 빅데이터 기술의 실체와 효과 그리고 적용 가능한 범위를 확인하는 정도였다.

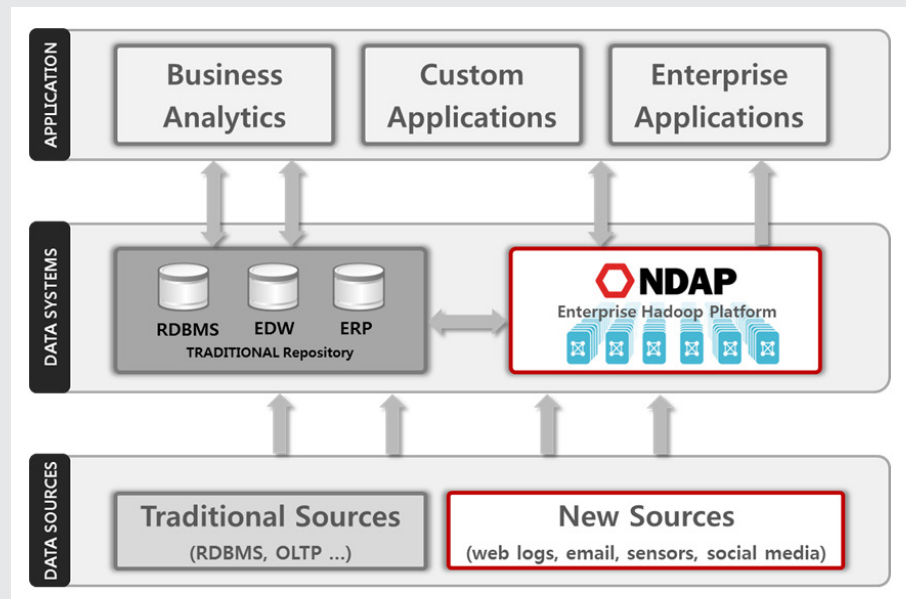
실제 쉽게 고객들이 이해했던 빅데이터의 접근법은 비정형 로그파일의 통합 저장장을 통해 OI(Operational Intelligence) 환경을 구축하는 것이었다. 하지만 빅데이터 기술은 이미 비정형 데이터뿐만 아니라 정형 데이터를 적재하여 처리할 수 있는 기술을 이미 제공하고 있었고, 이것이 실제 가능하다는 사례가 국내외 알려지면서 많은 관심을 갖게 되었다.

이렇게 빅데이터라는 새로운 데이터 르네상스 시대를 맞이해 차세대 DW 환경 구축이 본격화되고 있다. 하지만 경제성에 민감할 수 밖에는 없는 기업들은 낮은 도입비용으로 실질적인 차세대 DW를 구축해야 하는 두 마리 토끼를 잡는 방법에 대해 고민할 수 밖에 없게 됐다.

그리고 이에 걸맞은 최적의 데이터 관리를 위한 해결책이 바로 ‘하이브리드 DW’다. 하이브리드 DW는 기존에 담지 못하던 비정형 데이터를 수용하고, 추가로 장기간 DW 데이터까지 포괄해 다목적 데이터 저장소로서 고도화 할 수 있는 인프라 기술을 말한다. 여기서 말하는 장기간 DW 데이터란 기존 DW가 최근 1~3년치 데이터만 유지하는데 반해, 그 이상의(3년 이상의) 대용량 과거 데이터를 의미한다. 이런 장기간의 데이터를 하둡 기반의 빅데이터 저장소에 담음으로써 기존 메인 시스템을 변경하지 않고 신기술을 적용하는 안정적인 IT 전략을 가져갈 수 있다.

또한 기존 DW 시스템을 가볍게 유지함으로써 고가의 라이선스, 유지보수 등에 대한 확장비용을 크게 절감할 수 있다. 확장 데이터와 새로운 데이터를 저장하는데 비용효율적인 빅데이터 플랫폼을 활용함으로써 비용 최적화된 IT 인프라 계획을 고려할 수 있게 된 것이다.

그림 1 | 하이브리드 DW - 차세대 데이터 아키텍처



물론 데이터 저장뿐 아니라 맵리듀스(MapReduce)라는 병렬 분산처리 방식으로 대량의 데이터를 목적에 따라 처리해 결과를 추출한다. 또한 하드웨어 환경의 특징은 기존 하이엔드 스토리지와 대형 서버를 결합한 것이 아닌 경제적인 x86 서버를 멀티 클러스터로 구성하게 되면 다른 요소가 전혀 필요하지 않다.

그림 1과 같이 하이브리드 DW는 기존 DW 시스템(최근 데이터)을 빅데이터 플랫폼(아카이빙 데이터)과 유연하게 결합하는 방식이다.

그림 2와 같이 하이브리드 DW는 기존 환경을 수용하면서 비용효율적인 신기술을 접목해 시너지를 비용 대비 성능 관점으로 내는 것이라고 할 수 있다. 예를 들어 최근 주목 받고 있는 하이브리드 자동차는 주 연료인 가솔린이 비싸져 차량

빅데이터에서 장기간의 데이터를 강조하는 이유

보통 우리가 주식을 할 때 일반적으로 특정 종목을 일봉, 주봉, 월봉 단위로 장기 추이를 통해 다양한 패턴을 살펴보고 투자를 결정한다. 이것은 한 달치 정도의 추이로는 투자하기에 리스크가 너무 크고 불확실성이 높기 때문이다.

또한 집에서 쓰는 가계부도 한달치 입력 값으로는 절약해야 하는 패턴이나 우리 집의 소비 패턴을 판단하기에

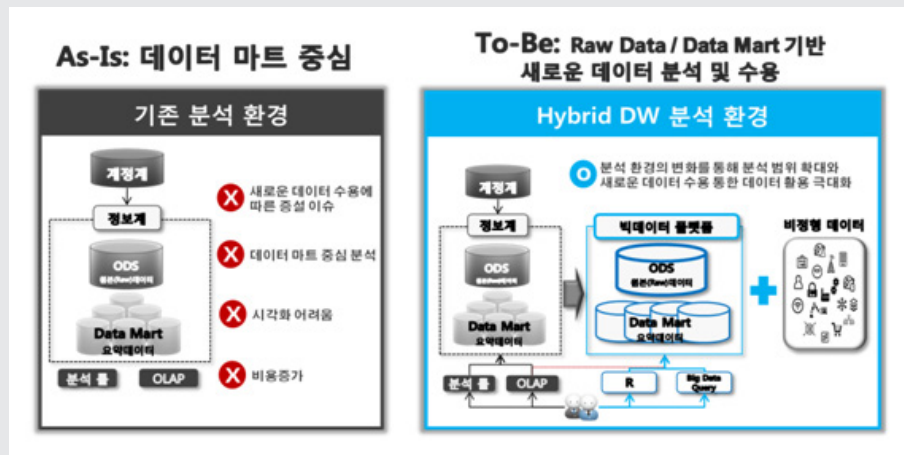
는 턱없이 부족하다. 그래서 1년치 이상의 데이터를 입력해 소비생활의 패턴을 계절별, 날씨별, 이벤트별로 진단하고 새로운 소비 설정을 계획하게 된다.

이런 일상적인 예처럼 장기간의 데이터가 주는 가치는 생각보다 크다. 기존 IT 인프라 환경에서 장기 데이터를 수용하기 위한 고민을 하기에는 기존 솔루션들이 너무 고비용이었고, 다른

형태로 처리하는 사례 또한 찾아볼 수 없었다.

빅데이터 시대가 도래하면서 장기 데이터를 잘 처리하고, 기존에 담지 못하던 데이터까지 수용할 수 있는 하둡이라는 기술이 탄생함에 따라 우리는 새로운 접근을 할 수 있는 도구가 생겼고, 이에 맞는 데이터 기반의 전략을 수립할 수 있게 됐다.

그림 2 | 하이브리드 DW 기반의 차세대 데이터 분석 환경



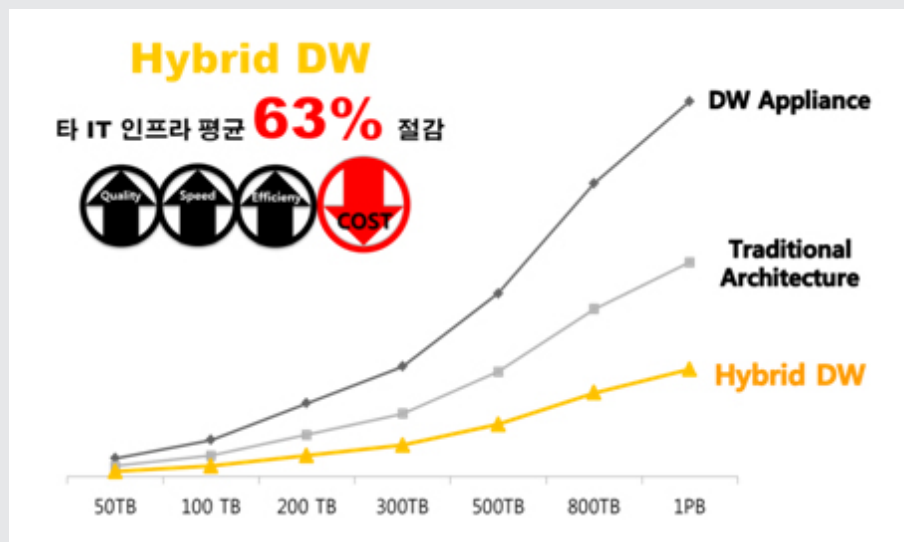
유지비용이 증가하게 되자 이를 타개하기 위한 가장 현실적인 대안 기술로써 시장에 나타난 것이었다. 하이브리드 DW 또한 과거보다 빠른 속도로 증가하는 데이터 수용과 처리를 위해 기존 고가의 소프트웨어와 하드웨어가 결합된 DB 솔루션이라는 틀에서 벗어나 획기적인 비용절감이 가능한 대안 기술로 탄생했다.

타 IT 인프라 대비 평균 63%의 획기적인 비용절감

하이브리드 DW는 비용 측면에서 보면 그림 3과 같이 기존 DW 인프라와 비교해 봤을 때 평균 63% 이상의 비용절감 효과가 있다는 것을 알 수 있으며, 빅데이터 기반 DW 어플라이언스 제품과 비교했을 때는 73% 감소라는 엄청난 비용이 절감되는 것으로 나타났다.

이는 하이브리드 DW가 비용이 경제적인 x86 기반의 하드웨어에서 하둡의 대량 데이터 처리 능력을 제공함으로써 기존 DW의 부담을 줄여주고, 장기간의 데

그림 3 | 하이브리드 DW Cost ROI 비교



이터를 탐색할 수 있는 기반을 만드는 플랫폼 기술이기 때문이다.

구체적으로 하이브리드 DW는 HW 측면에서 비용이 낮고 확장성이 우수한 범용적인 x86 서버와 네트워크 스위치만 있다면 구성이 가능하다. SW 측면에서도 글로벌 업체들이 제공하는 고가의 DBMS를 사용하는 것이 아닌 오픈소스 기반의 빅데이터 플랫폼을 도입함으로써 초기 도입비용은 물론 매년 들어가는 유지보수 비용도 대폭 절감할 수 있다. 또한 스케일아웃(Scale-Out) 방식의 확장 및 증설에 따른 선형적인 성능 보장이 확보됨으로써 향후 데이터 증가에 따른 증설 비용이 기존 방식에 비해 매우 경제적이다.

혁신적인 빅데이터 기술과 경제성의 만남

지금까지 하이브리드 DW가 무엇인지, 빅데이터 기술이 왜 그 역할을 수행하는 도구로 사용되는지, 왜 이것을 해야 하는지를 살펴봤다. 이제는 대량의 데이터라는 주제 아래 비용과 기술의 한계를 극복할 수 있는 오픈소스 기반의 빅데이터 기술을 활용해 기존 DW에서 하기 어려운 것들을 보완할 수 있는 하이브리드 DW 형태의 접근이 가능하다. 혹은 히스토리성 트랜잭션을 아카이빙 하는 전략을 수행할 수 있는 도구로도 쓸 수 있게 됐다.

또한 하이브리드 DW는 획기적인 비용절감이 가능할 뿐만 아니라 대량 데이터를 수용하고 처리할 수 있는 혁신적인 기술이면서 동시에 기존 기술과도 어우러져 시너지를 낼 수 있는 있다는 결론을 도출할 수 있다. 갈수록 치열해지는 시장 경쟁체제에서 데이터를 기반한 의사결정 체제를 확보해 경쟁력을 확보하기 위해서는 하이브리드 DW라는 방향은 반드시 고려해야 할 기술이다. **ITWORLD**






IT 트렌드 종합 정보센터 IDG Tech Library

IDG Tech Library는 IDG 글로벌 네트워크를 통해 축적된 전문 정보를 재구성하여 최신 기술의 기본 개념부터 현황, 전략 및 도입 가이드까지 다양한 프리미엄 IT 정보를 제공합니다. Computer World, Info World, CIO, Network World 등의 세계적 IT 유명 매체의 심도 깊은 정보를 무료로 만나보세요.

IDG Deep Dive, Tech Focus, Summary, World Update 등의 다양한 콘텐츠를 제공 받을 수 있습니다.



한국IDG(주) 서울시 중구 봉래동 1가 108번지 창화빌딩 4층 100-161 Tel : 02-558-6950 Fax : 02-558-6955

www.itworld.co.kr
[www.twitter.com/ITWorldKR](https://twitter.com/ITWorldKR)
www.facebook.com/ITworld.Korea