

IDG Deep Dive

## Big Data Special Report

## DW 선택 가이드 1

## ❖ 데이터량 증가 추세 : 과거, 현재 그리고 미래 전망

- 데이터량, 향후 10년 50배 증가
- 데이터량 증가, IT 역할 변화 필요

## ❖ 빅데이터 트렌드

- CIO가 알아야 할 9가지 DW 동향
- 인터뷰 | IBM 스티브 밀스, 왓슨의 미래를 논하다
- IBM의 네티자, 대규모 분석 어플라이언스로 발돋움

## ❖ DW어플라이언스 시장 현황

- 빅데이터 선점을 위한 M&A 활발
- 인터뷰 | 네티자 CEO 짐 바움

## ❖ 빅데이터를 위한 제언

- 기고 | 단순할수록 좋다

Sponsored by

IBM®

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.

# 데이터량, 향후 10년 50배 증가

Lucas Mearian | Computerworld

**2011**년 생성되는 데이터량은 총 1.8ZB(zettabytes), 즉 1.8조 GB이며, 향후 10년 동안 이 숫자는 50배 증가하고 전 세계 데이터를 관리하는 총 서버 수는 10배 늘어날 전망이다.

이는 2011년 6월 발간한 IDC 디지털 유니버시티에 포함된 내용의 일부다. 이 보고서에는 또 다른 흥미로운 점도 언급됐다. 사람들이 작성한 이메일 메시지, 촬영한 사진들, 다운로드한 음악 및 영화보다 이들에 대해 간접적으로 만들어진 데이터가 더 많을 것이라는 점이다.

IDC는 2020년 생성되는 전세계의 데이터 중 상당 비율이 임베디드 시스템에 의해 생성될 것이라고 예측했다. 의료 및 의료기기, 건축물 등에 내장된 센서와 같은 임베디드 시스템에서 막대한 데이터를 만들어낼 것이라는 예상이다.

보고서에는 또 파일과 이메일, 동영상 등과 같은 비정형 정보들이 향후 10년 동안 생성되는 전체 데이터의 90%에 달할 것이라는 전망도 담겼다.

IT 관리자에게 좋지 않은 전망도 있다. 이들 데이터를 관리하는 IT 전문가 숫자가 현재에 비해 불과 1.5배 증가한다는 것이다. 이 밖에 데이터 보안이 앞으로도 핵심 이슈일 것으로 전망된다는 내용도 포함돼 있다.

IDC는 이와 관련, 오늘날 생성된 데이터의 1/3만이 최소한의 보안장치를 가지고 있으며, 보호되어야 할 데이터 중에서 실제로 보호되는 비율은 절반 정도에 그친다고 지적했다.

IT 관계자들에게 반가운 소식도 있다. 정보를 생산, 포착, 관리, 저장하는 새로운 하드웨어 및 소프트웨어 기술들의 가격이 2005년에 비해 1/6 정도로 내려갈 것으로 예측됐다.

예를 들어 데이터 중복제거 및 압축 기술이 전송 및 저장 데이터량을 감소시켰으며 가상화 및 신포비저닝 기술이 스토리지 시스템의 활용률을 높여줬다는 설명이다.

IDC의 스토리지 및 반도체 연구 부문 부사장 데이빗 레인슬은 “데이터 저장 비용이 획기적으로 내려갔다. 이에 따라 저장하려는 데이터가 더욱 늘어나고 있다”라고 말했다.


그에 따르면 2005년 이후 하드웨어와 소프트웨어, 클라우드 서비스, 관련 임직원에 대한 기업들의 연간 투자는 50% 상승해 4조 달러에 달하는 상태다.

아울러 데이터 캡처, 검색, 발굴, 분석 도구 등이 ‘데이터에 대한 데이터’를 쏟아낼 것으로 전망됐다. 이를테면 페이스북 사진에 태그를 붙이도록 도와주는 안면 인식 루틴이 한 예다.

레인슬은 ‘데이터에 대한 데이터’가 디지털화 되는 세상보다 두 배나 빠르게 성장하고 있다고 진단했다. 1GB 데이터가 1PB의 정보를 사용할 수도 있다는 설명이다.

그는 “즉 오늘날의 1.8ZB 데이터에는 직접 저장한 데이터와 데이터로부터 생성된 데이터가 혼재돼 있다. 유튜브를 통해 HD 영상을 감상하는 경우를 생각해 보라. 영상을 보는 동안 시청자의 시청 행태에 대한 메타데이터가 생성된다”라고 말했다.

IDC는 또 2015년경에는 모든 정보의 20% 정도가 어떤 식으로든 클라우드 서비스로 제공될 것이며 10% 정도는 클라우드 인프라에 저장될 것이라고 예상했다.

레인슬은 데이터와 관련한 차세대 단계로 ‘대용량 데이터 분석’을 언급했다. 기업들이 방대한 데이터를 통해 유용한 데이터를 추출해낼 수 있어야 한다는 지적이다. 그는 “대용량 데이터 활용에 대한 성공 이야기들이 나오기 시작하면 데이터 속에 금광이 있었다는 점이 알려질 것이다. 많은 기업들이 온라인에 더욱 많은 데이터를 밀어 넣는 방안을 강구하게 될 것”이라고 말했다. 

# 데이터량 증가, IT 역할 변화 필요

Lucas Mearian | Computerworld

**기**업 데이터 저장용량이 매년 50% 이상씩 증가함에 따라 방대한 정보를 찾고 데이터를 저장해야 하는 업무의 특성 때문에 IT 종사자들을 훈련시키는 방법이 바뀔 필요성이 대두되고 있다.

IT 호스팅 업체 선가드(Sungard)의 저장장치 설계자 크리스 맥넬리는 확대된 IT환경에서 다양한 시스템을 맞추는 방법을 IT직원들에게 어떻게 교육할 지에 대해 고객사를 지원하고 있다.

맥넬리에 따르면, 최근 AIX와 백업 관리자들에 선가드의 스토리지 에어리어 네트워크(Storage Area Network, 이하 SAN)와 클라우드 스토리지(Cloud Storage)교육에 자원해서 참가했다.

“내가 전문가들에게 이것이 어떻게 작동하고 우리가 해야 할 일은 무엇인지에 대해 가르치는 대신에 IT직원들 스스로 스토리지에 관한 지적인 논의를 하는 분위기다. 이러한 변화는 앞으로 더 나은 시스템을 만들어내게 될 것이다”라고 맥넬리는 강조했다.



유전자 염기 서열 분석기기(Genome Sequencer)기업인 트랜스레이셔널 지노믹스 인스티튜트(Translational Genomics Institute, 이하 TGen)의 네트워크 및 컴퓨터 시스템 책임자 제임스 로웨이는 자사 IT부서 직원들이 네트워크와 운영체제, 스토리지가 어떻게 상호작용하는지를 이해해야 할 필요가 있다고 주장했다.

로웨이에 따르면, 인간 게놈지도를 만들면서 매주 2TB분량의 새로운 데이터가 생성되며 올해 말부터는 주당 10TB수준으로 증가할 것으로 예상된

다. 게놈 데이터는 인간 유전자 정보를 분석하는 데 활용되며 이는 특정 질병을 치료하는 맞춤형 의약품 개발하는 데 사용된다.

“문제는 뭐가 중요한 지, 필요할 때 정보를 찾아내는 것이다”라고 로웨이는 말했다.

TGen의 경우 생성된 데이터를 모두 보관하는 문제와 씨름하고 있다. 맞춤형 의약품 개발의 과학적 깊어치를 생각해 볼 때 저장된 데이터가 정보의 유용한 출처가 될 수 있기 때문이다. 게다가 게놈 데이터는 매우 개인적인 정보일 수 있다. 따라서 모든 데이터를 안전하게 보관해야 한다는 의미도 된다.

2010년 EMC에 인수된 아이실론도 정제되지 않은

데이터의 중요성에 대해 강조했다. 이 회사의 CTO 폴 루터포드는 “모든 것을 끊임없이 보관하는 이유는 우리가 갖고 있는 것이 무엇인지 확실히 알지 못하기 때문이다. 생명과



학 분야에는 아직 배워야 할 것도, 알려지지 않은 것들도 너무나 많다”라고 설명했다.

클라우드 스토리지 제공업체인 라이브오피스(LiveOffice)의 CEO 닉 메타 역시 루터포드의 의견에 동의했다. 메타는 “데이터가 제대로 저장되었는지 여부에 상관없이 데이터 보관은 지속된다”라고 말했다.

“우리의 문제는 어떻게 모든 것을 비용효율적으로 저장하느냐다. 우리는 모든 것을 보관하고 보관된 것들을 가치 있게 할 수 있는 방법을 원한다. 모든 데이터를 갖고 있으면 더 능률적으로 일할 수 있기 때문이다”라고 메타는 전했다.

라이브오피스의 고객이 4PB의 데이터를 디스크에 저장할 경우, 실제 그 기업은 매일 5TB를 추가로 사용하게 된다. 라이브오피스는 고객안전을 위해 모든 데이터를 암호화한다.

데이터량 증가에 따라 이를 관리하거나 통제할 수단이 필요하다는 의견도 대두됐다.

하버드 의대에서 생산관리를 담당하는 스티븐 마르티노는 “기업 사용자들의 마이닝 서비스에 대한 수요가 발생할 때가 가장 좋은 시점”이라고 말했다. 마르티노는 “IT 관리자들에게 필요한 것은 누가 무엇을 사용하고 있는지 추적하는 수단이며 아직까지 그 어떤 툴 업체도 그런 수단을 제공하지 못하고 있다”라고 말했다.

“연구원에게 저장할 수 있는 한계라는 것은 없다. 심지어 하루에 1TB 또는 2TB도 가능하다. 가장 큰 문제는 IT관리자가 연구실의 누군가가 차지 백(Chargeback)을 갹아먹는 저장공간을 사용하고 있는지 설명할 수 있는 데이터를 수집할 필요가 있다고 것이다”라고 마르티노는 강조했다.

재생 에너지 기업에 독자적인 날씨데이터를 제공하는 쓰리티어(3TIER)의 IT 책임자 폴 잉글리쉬는 “데이터가 어디로 이동하고 누가 그에 대한 관리책임을 맡고 있는지 알아내기 위해 IT 직원이 하루에 몇 시간씩 회의에 참석한 적이 있었다”라고 말했다. “우리는 항상 대용량 데이터를 취급해 왔다. 10년 내지는 20년치의 기후학적 데이터를 보관하기를 원한다. 조만간 수 PB의 데이터가 추가로 필요하게 될 것으로 관측된다”라고 마르티노는 덧붙였다.

데이터 폭주에 대비하는 차원에서 확장할 수 있는 풀을 생성하기 위해 쓰리티어는 자사의 임직원 모두가 접속할 수 있는 14개의 아이실론 NAS 어레이를 설치했다. “연구실에서 스토리지는 기본 옵션이 됐다”라고 마르티노가 밝혔다.

데이터량 폭주와 씨름하는 IT 관리자들은 데이터 이동 즉, 데이터를 정확한 스토리지 tier로 이동시키고 이 데이터를 사용하는 사람들에게 가능한 한 가깝게 보관하는 것을 또 하나의 과제로 꼽았다.

로웨이는 “TGen의 피닉스(Phoenix)에 위치한 유전자 배열을 밝히는 컴퓨터와 결과를 처리하기



위해 사용되는 템피(Tempe)의 슈퍼컴퓨터 사이에 끊임없는 데이터 교환이 이뤄진다”라고 말했다.

“우리는 1GB의 전용선을 사용했다. 하지만 오래지 않아 한계를 느꼈다. 현재는 10기가빗의 [이더넷] 전용선을 사용하고 있으며, 사실 인피니밴드(InfiniBand) 사용을 고려 중이다”라고 로웨이는 말했다.

모든 IT 관리자들이 동의하는 오늘날의 가장 큰 고민거리는 ‘도대체 얼마나 더 데이터가 늘어나야 정보저장에 관한 인식이 바뀔 것인가’이다. 대부분은 차후에 데이터가 어떤 식으로 회사에 이득이 될지 모르기 때문에 모든 것을 보관하고 싶어하기 때문이다. **CIO**

# CIO가 알아야 할 9가지 DW 동향

Georgina Swan | CIO



**가**트너가 CIO들을 대상으로 '2011년 및 2012년 주요 데이터 웨어하우징 동향'에 대해 메시지를 발송했다. 여기에는 이 기간 동안의 데이터 웨어하우징 동향이 비즈니스 애널리틱스 확산 관련 기술에 어떻게 영향을 미칠지에 대해 기술돼 있다.

가트너에 따르면, DW는 IT 인프라의 핵심 요소로 남을 전망이다. 비즈니스 인텔리전스(BI)와 비즈니스 분석에 대한 수요가 커지고, 최적화와 유연한 설계, 대안 전략이 점차 중요해질 것이 확실시 되기 때문이다.

가트너 리서치 담당 부사장 마크 베이어는 “DW는 대기업에서 가장 큰 정보 저장소 중 하나로 남아 있게 될 것”이라며, “따라서 CIO는 시장 동향을 파악하고, 새로운 기술 솔루션이 어떤 방식으로 기존 방법들과 혼합될지 이해해야만 예산상의 낭비를 피할 수 있다”라고 설명했다.

가트너는 2011년~2012년 데이터 웨어하우징 시장의 9가지 주요 동향을 다음과 같이 전망하고 있다.

- CPU/메모리 밸런싱, 디스크 스토리지, I/O 등 하드웨어 관리를 위한 최신 기능은 이제 DW 구현 플랫폼에 있어 당연한 요소로 자리잡았다. 일부 신규 진출 업체들은 차별화를 위해 최적화에 초점을 맞추고 있으며, 대다수의 DW 업체들은 압축과 사용 기반의 데이터 배치 전략을 통해 스토리지 최적화를 강조하고 있다. 그러나 이러한 것들이 실 사용에 있어 반드시 중요한 것만은 아니다.

- 기업들이 어플라이언스를 구매하는 이유는 많지만, 가장 주된 이유는 간소함을 위해서다. 업체에서 하드웨어, 소프트웨어, 서비스를 하나로 묶어서 성능 검증까지 마쳤기 때문이다. 또한 어플라이언스는

완성된 상태기 때문에 설치도 빠르다. 따라서 여타 문제가 발생하면 어플라이언스 업체에게 전화를 거는 것으로 1차 조치를 취할 수 있다. 이외에도 하드웨어 조정에 드는 시간 소모를 피하도록 해준다는 점에서 2차 효과도 창출될 수 있다.

- 기업 대부분은 DW 데이터베이스 관리 시스템을 선정하는 단계에서 주요 업체들을 대상으로 POC (proof of concept: 개념 또는 기능 검증)를 수행해야만 한다는 점을 잘 알고 있다.

가트너는 가능하면 운영 시스템에서 데이터를 추출한 실제 소스 시스템을 이용해 POC를 하라고 권장한다. 또 많은 사용자를 대상으로 해야 하고, 실제로 사용되는 환경에 부합하는 방식으로 DW 작업 부하를 생성해야 한다고 권고하고 있다.

- DW가 일으키는 작업부하에는 벌크/배치(Bulk/batch) 부하, 기본 리포팅, 기본 OLAP(온라인 분석 처리), 실시간/지속 부하, 데이터 마이닝 및 운영 BI의 6가지가 있다.

이들 6가지 작업부하를 일으키는 웨어하우스에 대해 작업부하 성능의 예측성을 위해 평가해야만 한다. 작업부하 분담에 대한 계획이 잘못되면 관리 비용이 올라가고, 볼륨과 추가 작업부하가 더해지면 지속가능성(sustainability) 문제를 초래할 수 있기 때문이다.

- 데이터 마트는 특정 애플리케이션 전용의 분석 저장소로 사용되지만 규모에 관계없이 일반적으로 DW보다 특정 소규모 사용자 그룹을 갖는다. 데이터 마트는 DW를 최적화하는데 사용될 수 있다. 작업부하의 일부를 데이터 마트로 분담시킴으로써 웨어하우징 환경의 성능을 높이는 데도 활용된다.

- 컬럼스토어(Column-store) 데이터베이스 관리 시스템은 일반적으로 기존의 로우 기반(row-based) 시스템에 비해 쿼리 응답이 빠르다. 따라서 뛰어난 데이터 마트 플랫폼으로, 더 나아가 주요 DW 플랫폼 역할을 했다.

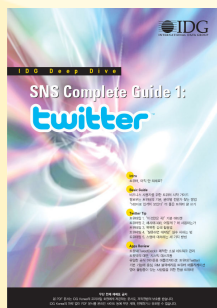
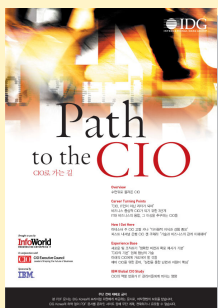
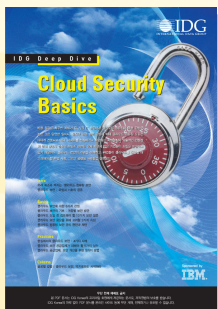
- 인-메모리 데이터베이스 관리 기술은 쿼리 응답과 데이터 커밋(data commit) 시간이 아주 빠르다. 또 분석 시스템과 트랜잭션 시스템이 동일한 데이터베이스를 공유할 수 있는 가능성이 한층 높아진다.

미들 티어(middle tier)의 데이터 서비스, 마스터 데이터 접근법, 분석 데이터 모델이 지배적인 방식으로 부상하기 시작하면서, 기존의 로우 기반 업체들은 컬럼 방식과 인-메모리 방식을 권장하는 추세다. BI 솔루션은 곧 부상할 것으로 보인다. 이로 인해 고성능 BI 제품에 인-메모리 데이터베이스의 기능이 강화될 것으로 예상된다. 아울러 대형 업체들

의 인수 대상이 될 것으로 보인다.

- 2011년, 서비스로 제공되는 DW는 SaaS(Software as a service)와 아웃소싱 DW 두 가지 형태다. 클라우드의 DW는 근본적으로 인프라 설계를 어떻게 할 것인가의 문제다. 여전히 데이터 모델을 개발해야 하고, 통합 전략을 적용해야 하며, BI 사용자가 접근할 수 있도록 관리해야 하기 때문이다. 프라이빗 클라우드는 일부 기업에 있어 주요 인프라 설계 방안의 하나로 부상하고 있다.

- 오픈소스 데이터베이스 관리 시스템은 여전히 실험적으로, 그리고 구색 맞추기로 사용되고 있다. 현재 오픈소스 웨어하우스는 드물며, 전통적인 웨어하우스보다 소규모이고, 더 많은 수동 지원을 필요로 하는 게 일반적이다. 하지만 일부 솔루션은 데이터 웨어하우징 전용으로 최적화되어 있다. **CIO**



## IT 트렌드 종합 정보센터 IDG Tech Library

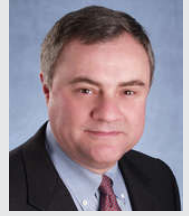
IDG Tech Library는 IDG 글로벌 네트워크를 통해 축적된 전문 정보를 재구성하여 최신 기술의 기본 개념부터 현황, 전략 및 도입 가이드까지 다양한 프리미엄 IT 정보를 제공합니다. Computer World, Info World, CIO, Network World 등의 세계적 IT 유명 매체의 심도 깊은 정보를 무료로 만나보세요

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



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

[www.idg.co.kr](http://www.idg.co.kr) [twitter.com/idgkr](https://twitter.com/idgkr) [www.facebook.com/IDGKorea](https://www.facebook.com/IDGKorea)



# IBM 스티브 밀스, 왓슨의 미래를 논하다

## 오라클에 대한 신랄한 비판도 함께 거론

Chris Kanaracus | IDG News Service

### interview

**IBM**의 스티브 밀스는 소프트웨어 분야에서 뛰어난 리더십으로 유명하다. 그러나 2010년 하드웨어와 시스템 분야까지 맡게 되면서, 밀스는 이제 10만 명이 넘는 직원과 400억 달러가 넘는 돈을 관리하는 자리에 앉게 됐다.

2011년 6월 보스턴에서 열린 IBM의 네티자 엔지 유니버스 컨퍼런스(Netezza Enzee Universe conference)에서 IDG 뉴스 서비스가 밀스를 만나 인터뷰했다. 그는 라이벌인 오라클에 대한 자신의 생각을 전하고, 게임 쇼 '제퍼디(Jeopardy)'에서 인간 경쟁자를 물리쳐 화제가 된 왓슨 슈퍼컴퓨터의 미래에 대한 견해를 밝혔으며, SaaS의 미래를 전망했다.

● IDG뉴스: 래리 엘리슨이 CEO로 있는 오라클은 소프트웨어와 하드웨어가 합쳐진 시스템이라는 비전을 제시했다. 많은 이들이 이를 놓고 IBM이 오래 전에 해놓은 일을 고작 따라 하는 것뿐이라고 얘기한다. 오라클의 행보에 대해 어떻게 생각하는가?

● 스티브 밀스: 오라클의 관심사는 매우 단순하다. 래리 회장을 부자로 만들자, 그게 그 회사가 설립된 이유이자 회사의 핵심 목표다. 회사의 이익은 가장 많은 주식을 가진 래리에게 가게 돼 있다.

오라클은 고객 중심의 회사가 아니다. 그들은 고객들의 돈을 목적으로 하고 있고, 그 돈을 얻어내기 위해 매우 독창적인 방법을 사용하기도 한다. 일하는 환경 자체가 무한 경쟁을 부추기는 분위기라 파트너끼리도 경쟁하게 한다. 사실 아주 실용적인 전략이라 할 수 있다. 원하는 돈을 벌 수 있다면 어떤 방법이든 사용하겠다는 것이다. 협력을 해야 돈을 벌 수 있겠다 싶으면 그렇게 하고, 그게 아니면 다

른 방식을 찾는다.

IT를 움직이는 주요 동력이 돈인 것은 맞다. 모든 기업 역시 돈을 벌고자 한다. 하지만 일부 회사들은 다른 기업들보다 그 목적을 성취하는 데 있어 지나치게 적극적이다.

오라클의 경우 고객들에게 빨대를 꽂다시피 해 이 목적을 성취하는 편인데, 나는 바로 그런 점에서 오라클을 라이벌로 높이 평가한다. 그들은 학교에서 애들을 괴롭히는 불량 청소년 같은 전략을 사용한다. 계속해서 판지를 걸고, 기술에 대한 평가는 애써 피하는 편이다. 개인적으로 볼 때, 오라클은 타고난 장사꾼이라고 생각한다. 고객들은 오라클과 협상을 되도록 피하려고 한다. 협상하기 까다롭고, 고객들을 화나게 만드는 데 일가견이 있기 때문이다.

● IDG뉴스: HP는 자사 서버에 인텔의 아이태니엄 칩을 사용한다. 그런데 오라클이 인텔 아이태니엄 칩을 지원하는 소프트웨어 개발을 중단했다. 이에 대해 어떻게 생각하는가?

● 밀스: 근본적으로 [그 결정은] HP에게 싸움을 거는 것이라 생각한다. 하지만 고객의 입장에서 보면 '내가 HP의 아이태니엄 박스를 잔뜩 산데다 HP-UX를 가지고 있고 오라클의 제품까지 가지고 있는데 더이상 그걸 지원하지 않겠다니, 내 뒤통수를 친 것 아닌가' 싶을 수도 있다. 결과만 놓고 보면, 오라클은 고객을 배신한 것이다. 고객들에게 묻지도 않고 아이태니엄에 대한 지원을 중단했으니 말이다. 하지만 오라클은 단순히 HP에 시비를 걸고 싶어서 그렇게 했을 뿐이다.

● **IDG뉴스: 장기적인 관점에서, IBM은 아이태니엄 지원에서 어떤 입장인가?**

● **밀스:** 우리는 아이태니엄을 지원한다. HP-UX도 마찬가지다. 덧붙이자면, 아이태니엄 시스템을 대체하려는 노력도 하고 있다. HP와 경쟁해 우리 제품을 판매하고 있긴 하지만, 어떤 면에선 HP를 지지하기도 하다. IBM은 오래도록 이런 복합적인 입장을 취해 왔다.

우리는 고객들이 아이태니엄으로 무엇을 하느냐에 따라 거기에 충실할 뿐이다. 만약 고객이 IBM 제품을 사용하고 있다면, 거기에 맞춰 장기간 지원해 준다. 심지어 우리가 1960년대에 만들었던 것을 사용하고 있는 고객도 있다.

● **IDG뉴스: 그렇다면 자연어를 이해할 수 있고 질문에 대답도 할 수 있는 왓슨의 슈퍼컴퓨터와 같은 신기술은 어떤가? IBM은 그런 기술을 어떻게 상용화할 생각인지 궁금하다.**

● **밀스:** 제퍼디 쇼에는 슈퍼컴퓨터가 하나의 개체인 것처럼 나왔지만, 한 겹만 들춰보면 그렇지 않다는 걸 알 수 있다. 하둡(Hadoop)과 IBM DB2도 사용하고 있다. 물론 그 뒤에는 많은 양의 정보가 뒷받침돼 있고 말이다.

왓슨은 정보를 검색하거나 되찾아 올 수도 있고, 특정 정보를 반복해 검색할 수도 있다. 따라서 단순히 답을 찾는 것이 아니라, 입력된 다른 정보들과 자신의 해답을 비교하는 능력까지 있다는 것이다.

하나의 주제를 시스템에 입력하고 계속해서 데이터를 넣으면, 정보 베이스가 생긴다. 풍부한 정보 베이스는 질의응답 측면에서 보자면, 시스템의 정확도를 높이는데 가장 중요하다고 할 수 있다.

컴퓨터는 훌륭한 도구이고 많은 도움을 주지만, 그렇다고 사람이 하는 일을 대체할 수 있다는 의미는 아니다.

만약 은행에서 수표를 관리하는 업무라면, 그다지 복잡한 일이 아니므로 컴퓨터가 대신하게 되도 괜찮다. 하지만 컴퓨터가 의사의 치료를 대신해야 한다면(지금 왓슨에서는 실제로 그러한 것을 개발하려 노력 중이다) 의사의 특정한 상황에 맞춰 정보를 넣는다. 그러면 왓슨이 가장 큰 방해물이 무엇인

지 몇 가지 가능성을 보여준다. 또 증세에 대한 정보를 얻기 위해 환자에게 어떤 질문을 던져야 할지에 대해서도 알려준다.

불행하게도 의학 분야에서는 부정확한 정보가 발생할 수도 있다. 모든걸 다 알 수 있는 사람은 없기 때문에 다른 사람의 의견을 묻기도 하고 전문가를 찾는 것 아닌가. 그럼 컴퓨터를 사용해 모든 사람의 지식을 합칠 수 있을까? 대답은 '그렇다'이다. 그리고 왓슨 역시 인간의 지식에 보탬이 되려는 목적에서 만들어 졌다.

우리는 이제 왓슨에 대해 상용화를 검토하고 있다. 올해와 내년에 왓슨을 적용한 다양한 사례들을 발표할 예정이며, 제품 발표도 있을 것이다. 일부는 티셔츠의 사이즈처럼 여러 가지의 패키지 형식으로 할 예정이다.

● **IDG뉴스: SaaS에 쓰는 지출이 전체 소프트웨어 예산에 비하면 적은 편인데 반해, 전문가들은 앞으로 몇 해 동안 SaaS의 사용이 급격히 증가할 것이라고 말하고 있다. IBM의 입장은 어떠한가?**

● **밀스:** IBM 예산에서 SaaS는 약 20억 달러를 차지한다. SaaS의 규모가 커지기를 바라고 있긴 하지만, 클라우드나 클라우드 관련된 것들은 정확히 측정하기 어려운 것이 사실이다. SaaS를 협의로 보면 소프트웨어로 국한되지만 광의로 보면, 비즈니스 프로세스 아웃소싱(BPO)도 포함된다.

ADP가 좋은 사례다. ADP는 클라우드 업체다. 물론 자기들은 아니라고 하지만, 그들이 하는 일은 클라우드링이다. 세일즈포스닷컴(Salesforce.com)과 다를 것 없는 SaaS 모델이다. 이같은 맥락에서 볼 때, SaaS는 거대한 시장이다. 희망적인 것은 IBM이 여전히 BPO 기업으로 남아 있다는 점이다. **CIO**



# IBM의 네티자, 대규모 분석 어플라이언스로 발돋움

Chris Kanaracus | IDG News Service

**IBM**의 네티자 솔루션이 10PB의 데이터를 몇 분 안에 처리할 수 있는 새로운 분석 어플라이언스로 발돋움하고 있다. 2011년 6월 보스턴에서 열린 엔지 유니버스(Enzee Universe) 컨퍼런스에서 이같이 발표했다.

고성능 DW어플라이언스는 주로 은행과 통신 기업들을 대상으로 한다. 또한 폭증하는 데이터에서 통찰력 있는 분석을 얻고자 하는 기업, 경쟁우위를 획득하거나 좀더 효율적으로 운영하고자 하는 기업, 신규 사업에서 수익을 얻고자 하는 기업들에게 유용하다. 고성능 DW어플라이언스는 대규모 DW 구성에도 활용 가능하다.

2010년 9월 IBM에 인수된 이후 처음으로 네티자 어플라이언스 신제품이 소개됐다. 네티자는 오라클의 엑사데이터(Exadata)에 필적하면서 좀더 비용 효율적이고 설치하기 쉬운 대안 솔루션으로 입지를 굳혔다.

네티자 솔루션을 적극적으로 활용하는 기업으로는 켈리 블루 북(Kelley Blue Book)이 있다. 신차 및 중고차 가격 정보를 제공하는 KKB의 애널리틱스 인사이트 담당 부사장인 댄 앵글에 따르면, 네티자 트리핀 활용을 지속적으로 확대할 계획이다.

KKB의 통계 전문가들은 현재 SAS의 분석 솔루션과 함께 네티자의 DW어플라이언스인 트윈핀 6를 사용해 방대한 가격 정보를 분석하고 있다. “자동차 관련 데이터들은 출처마다 표기도 제각기라 표준화되지 않았다. 이들을 분석하려면, 출처에 따라 좀더 복잡한 작업이 필요하다”라고 앵글은 말했다.

“KBB는 경매 사이트 등 여러 출처에서 자동차 가격에 대한 데이터를 얻는다. 특수 제작된 자동차 가격을 지난해보다 더 정확하면서도 간단하게 분석

하고 전망하고자 했다. 하지만 경매 거래 정보는 자동차 옵션과 다른 다양한 특이사항을 빠뜨릴 수 있다. 그리고 차에 장착된 다양한 특수 기능들은 시장 가격에 영향을 끼칠 수 있다”라고 그는 설명했다.

트럭의 경우, 상상을 초월할 정도로 옵션이 다양한데 예를 들어 트럭에 장착된 침대 길이와 같이 운전자들이 중요하게 여기는 정보는 VIN(Vehicle Identification Number)에 포함되지 않는다고 그는 전했다.

또한 2010년 오토트레이더닷컴(AutoTrader.com)에 인수된 KBB는 자사 웹사이트 방문자들의 엄청나게 많은 양의 클릭스트림을 분석하는 데 주력하고 있다. 앵글에 따르면, 현재 온라인 자동차 구매자 3명 중 2명 이상이 오토트레이더닷컴에 방문할 정도로 대중적인 사이트다.

앵글에 따르면, 현재 많은 클릭스트림 데이터가 어도비 옴니채어(Omniture) 시스템에 저장돼 있다.

KBB는 클릭스트림 데이터를 바탕으로 새로운 매출원을 찾고자 하며 새로운 네티자 어플라이언스를 통해 이것이 가능할 것으로 기대하고 있다.

“우리는 DW어플라이언스를 가장 잘 사용하는 기업이 되고자 한다. 아마도 내년쯤이면 트윈핀이 한계에 다다를지도 모른다”라고 앵글은 밝혔다.

KBB는 네티자로 갈아타기 전 마이크로소프트의 BI 스택과 오라클의 엑사데이터를 사용했다. 엑사데이터와 네티자는 성능 면에서 비슷하지만, 네티자의 비용이 더 저렴하고 설치와 유지보수도 더 쉽다는 것이 그의 설명이다.

“하드웨어 사용이 편리하다는 네티자의 마케팅 전략이 상당히 마음에 든다. 운영요원 1명이 근무 시간의 절반을 다른 업무에 활용해도 될 정도다”라고 앵글은 말했다.



재무 소프트웨어 업체 인튜잇(Intuit) 역시 마케팅 및 상품 개발용 클릭스트림 데이터를 분석하기 위해 네티자를 사용하고 있다.

“인튜잇의 퀵북스 온라인(Quickbooks Online)과 같은 소프트웨어에는 다양한 정보들이 담겨 있다. 상품 기능들이 화면에서 나타나면, 버튼 하나만 눌러서 고객이 사용한 제품 기능에 대한 정보를 모으거나 여타 깊이 있는 정보를 조사할 수 있다”라고 이 회사 데이터 시스템 개발 담당 디렉터 애럴레이는 말했다. 그에 따르면, 네티자의 DW어플라이언스는 인튜잇의 개발 부서가 어떤 기능을 그대로 두고 어떤 기능을 개선해야 하는지와 같은 깊이 있는 정보를 파악하도록 해준다.

네티자는 오라클, EMC, 테라데이타, HP와 같은 업체들과 시장에서 경쟁을 벌여 왔다. HP는 2011년 2월 DW업체 버티카(Vertica)를 사들여 이 시장에 뛰어들었다.

네티자 CEO 짐 바움은 인터뷰에서 공식적으로 HP의 행보에 대해 다음과 같이 말했다. “경쟁이 치열하고 시장이 매우 역동적인 것은 사실이다. 버티카를 인수했지만 DW와 BI 시장에서 HP가 기회를

얻을 지는 미지수라고 본다.”

“버티카 인수는 꽤 인상적이었다. 그렇지만 HP에게는 분석 소프트웨어 제공업체라는 정체성이 부족하다”라고 바움은 말했다. “데이터베이스는 있지만 그것만 있다고 해서 DW어플라이언스를 제공할 수 있는 것은 아니다. 우리는 하드웨어와 함께 데이터베이스를 번들로 제공할 것이다. 그리고 어플라이언스도 있다. 그건 네티자만의 차별점이다.”

네티자가 오라클과 매년 경쟁하는 동안 엑사데이터는 일부 분야에서는 잘 작동하고 있었다. 바움은 오라클에 대해 지향점이 다르다고 언급했다. “오라클 엑사데이터가 일부 업무에서 성공적으로 돌아가고 있는 것은 사실이다. 그러나 우리가 목표로 하는 분석 업무에서는 사정이 다르다”라고 그는 덧붙였다.

“현재 DW어플라이언스를 표방한 업체들은 그들의 잠재적인 경쟁력이 있다. 거기에 대해서는 추후도 의심하지 않는다. 그리고 우리는 지금 IBM으로 편입돼 앞으로 더 크게 성장할 발판을 기대하고 있다”라고 바움은 전했다. **CIO**

**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.

**IDG**  
INTERNATIONAL DATA GROUP

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/idgkr www.facebook.com/IDGKorea

# 빅데이터 선점을 위한 M&A 활발

Marc Ferranti | IDG News Service

**IT** 경기가 2008년 뒷서리를 맞은 이후 서서히 회복 국면으로 접어드는 가운데 업계 전문가들은 M&A가 계속될 것으로 전망했다. 컨설팅 기업 PwC의 보고서에 따르면, IT기업들의 현금 여력과 공격적인 인수 시도로 2009년 말 이후 IT업계의 M&A 규모와 가치가 지속적으로 상승하고 있다.

PwC는 “경기 낙관론, 현금 유동성 우세, 기술 변화, 통합 및 혁신 도입 등이 IT업계의 M&A에서 성장을 계속해서 주도할 것”이라고 전망했다.

최근 가장 큰 M&A는 2011년 인텔이 맥아피를 76억 8,000만 달러에 인수한 것이다. HP는 보안 업체인 아크사이트(ArcSight)를 23억 5,000만 달러에 인수했고 IBM은 네티자를 17억 달러에 사들였다.

인수를 이끄는 동인은 경제가 지속적으로 안정화됨에 따라 변화하고 있다.

PwC의 미국 트랜잭션 서비스 리더 마틴 커러프는 “지난 18개월에서 24개월 동안 기업들은 비용 구조를 개선하는 전략을 수립하는데 중점을 뒀다”라고 말했다. 커러프는 “2011년에도 M&A는 계속되겠지만 과거와 달리 전략적인 성장으로 관심이 변화하고 있다”라고 덧붙였다.

톱슨 로이터와 PwC의 자료에 따르면, 2009년에 비해 2010년의 M&A는 크게 증가했으며 미국 전체의 M&A 규모는 2009년 7,220억 달러에서 2010년 7,860억 달러로 9%나 늘어난 것으로 집계됐다. 이 가운데 IT기업들의 인수합병은 12%를 차지했다.

주요 기업의 보도자료에 따르면, IBM은 2010년 9월 DW어플라이언스 전문 업체 네티자를 인수했고, EMC는 이보다 2달 먼저 그린플럼을 사들였다.

빅데이터 관련 대형 M&A 현황

	피인수 기업	시기	규모
IBM	네티자	2010년 9월	17억 달러
HP	버티카	2011년 2월	비공개
EMC	그린플럼	2010년 7월	비공개
테라데이타	아스터데이타	2011년 3월	2억6,300만 달러에 지분 89% 인수

출처 : 각사 보도자료

HP는 2011년 2월 버티카를 인수했다. 빅데이터 분석 솔루션 업체 아스터데이타의 지분 11%를 이미 확보한 테라데이타의 경우, 나머지 89%를 추가로 매입하겠다고 밝혔다.

M&A 동향은 시장을 선도하는 기업들이 고객사에 윈스톱 쇼핑을 제공하는 방향으로 계속 진행될 전망이다. “장기적인 안목으로 기술 시장을 바라보면, 선도 IT업체가 클라우드 제공을 위해 부족한 기술을 채우거나 차별화하기 위해 기술 업체를 인수하면서 하드웨어, 네트워킹, 소프트웨어, 서비스 기업들이 하나로 통합되는 엔드-투-엔드 서비스 기업으로 커나갈 것이다”라고 PwC는 말했다.

PwC 보고서는 이밖에 디바이스, 서비스, 모바일 소비자 및 기업 사용자용 미디어 등에 대한 M&A가 내년에도 계속되며 인터넷 기업들은 지역 서비스 제공을 확대하기 위한 인수 대상을 찾아 나설 것으로 전망했다. **CIO**



# ‘One-Size-Fits-All 모델’ 은 DW 시장에는 맞지 않아



Jaikumar Vijayan | Computerworld(US)

## interview

**IBM**은 지난해 전세계 DW 시장에서 가장 파괴적인 기술을 보유한 회사 중 하나로 알려진 네티자(Netezza)를 17억 달러에 인수했다. 컴퓨터월드(Computerworld)는 네티자의 짐 바움 CEO와의 이번 인터뷰에서 인수가 갖는 의미와 시장을 견인하고 향후 이 회사 제품의 토대가 되는 트렌드에 대해 들어봤다.

**Q. IBM의 네티자 인수는 고객들에게 어떤 의미가 있나? 이를 중시해야만 하는 이유가 뭔가?**

**A.** IBM의 네티자 인수는 비즈니스 애널리틱스(Business Analytics) [애플리케이션]에 필요한 인프라의 지원과 창출 측면에서 살펴봐야 한다. 네티자는 여전히 네티자다. 팀과 엔지니어링, 현장 지원 메커니즘이 지속되고 있다. 따라서 고객 관점에서는 지금까지 거래해왔던 네티자와 거래를 하고 있는 것이다. 우리는 상당한 규모로 비즈니스를 키워가고 있다. 과거에는 없었던 지역과 영역에 더 많은 자원을 갖게 됐다. 전반적으로 고객들은 더 많은 확장성과 자원, 기회를 얻고 있다. 우리가 IBM으로부터 확보한 레버리지 효과다.

**Q. 네티자는 독자적인 어플라이언스 접근법으로 DW 시장에 상당한 영향을 미쳤다. 성공의 열쇠는?**

**A.** 초기 데이터 웨어하우스가 복잡하지 않았다면 더욱 성공적이었을 것이다. 고객들은 사용하고자 하는 서버 유형에 따라 스토리지와 컴퓨터 역량을 찾아 구매해야 했다. 또 소프트웨어를 찾아 구매하고, 서비스 공급자를 찾아야 했다. 그리고 이런 모든 것들을 고객 각자의 환경에 통합해 DW를 만들어야 했다. 이렇게 해야 DW가 다양한 분석과 기업이 필요로 하는 리포팅을 제공할 수 있었다. 아울러

최종 사용자가 이 웨어하우스에 있는 정보에 접근하기 위해서는 또 다른 많은 요소를 충족해야 했다. 따라서 과거에는 다수의 설치 요소들을 확장하는데 많은 비용이 들었다. 우리의 어플라이언스 모델은 아주 쉬운 설치와 유지보수, 빠른 가치 창출을 기반으로 이들 환경의 성능을 개선할 수 있는 기회를 제공했으며 그것이 성공의 열쇠였다고 분석하고 있다.

**Q. 시스템이 실제 어느 정도 뛰어난가?**

**A.** DW 산업의 많은 사람들은 TB당 비용에 대해 이야기한다. 그러나 TB당 비용보다 중요한 것이 있다. 실시간으로 의사 결정을 내릴 때 사용하고 있는 애플리케이션들이 문제가 된다. 따라서 정말 중요한 요소는 성능이다. ‘얼마나 많은 데이터에 얼마나 빠르게 액세스 할 수 있을까? 얼마나 많은 사용자가 데이터에 접근할 수 있을까?’ 이다.

뉴욕에 있는 우리 고객 중에 미디어매쓰(Media-Math)라는 회사가 있다. 실시간으로 부동산 가격을 책정하고 홍보하는 사업을 하고 있다. 마이크로초(100만 분의 1초) 단위로 부동산에 가격을 책정하고 경매하는 사업이다. 이런 사업은 대용량 데이터를 포괄적으로 분석하는 역량 없이는 불가능하다.





따라서 미디어매쓰에게 TB당 비용은 문제 거리가 아니다. 가격 책정 성능이 얼마나 뛰어나냐가 중요하다. 고객의 성능을 2배로 향상 시켜줄 수 있다면, 그리고 이를 바탕으로 2~3배의 주문을 실현했다면, 이들의 비즈니스는 송두리째 바뀔 수 있다.

**Q. 네티자 제품의 수요를 견인하는 요소는 무엇인가?**

**A.** 아주 뚜렷한 트렌드가 있다고 생각한다. 기업들이 데이터로부터 유용한 비즈니스 의사 결정 정보를 추출하는 역량과 그 필요성을 인식하고 있다는 것이다. 현대에는 어느 기업이든 정말 많은 데이터를 생산해낸다. 공급업체 시스템과 고객 시스템, 고객 관리 경영 시스템, ERP 시스템에서 이런 데이터를 생산하고 있다.

오늘 날의 기업들은 전통적인 BI 이상을 추구하고 있다. 개인적으로는 비즈니스 애널리틱스, 비즈니스 최적화라고 부르는 것들이다. 고객들은 데이터를 살펴보고, 이 데이터를 토대로 실제 예측하고자 한다.

두 번째 트렌드는 데이터의 규모다. 불과 얼마 전만 해도 1TB의 데이터가 크다고 생각했었다. 하지만 이제는 수백 TB, PB를 이야기하고 있다. 또 엑사바이트를 이야기할 날도 멀지 않았다. 우리가 생각하는 바는 이렇다. 데이터를 더욱 관리 가능하도록 만들 수 있다면 더 광범위한 최종 사용자 그룹에 접근할 수 있다는 것이다.

**Q. 이런 트렌드가 앞으로 몇 년 동안 네티자의 제품 전략에 어떻게 영향을 미칠 것으로 판단하나?**

**A.** 이제 곧 대용량의 데이터를 지원할 수 있는 제

품들을 보게 될 것이다. 바로 워크로드 최적화 환경이라는 아이디어다. 네티자와 IBM의 궁합이 잘 맞았리라고 판단한 이유 중 하나이기도 하다. IBM이 이야기하고 있는 워크로드 최적화 컴퓨터 중 하나가 왓슨(Watson) 컴퓨터다.

왓슨은 특별한 목적을 위해 구축된 시스템이다. 제퍼디(Jeopardy)쇼를 보라. 사실 제퍼디의 질문에 단 몇 초 만에 답을 내놓을 수 있다는 건 자연어 처리와 문자 분석에서 획기적인 기술을 가졌다는 의미다. 따라서 왓슨은 워크로드 최적화 시스템의 아주 좋은 사례다.

**Q. 새 BI 애플리케이션이 DW를 계속해서 필요로 할 것으로 보는가?**

**A.** 그렇다. 이 문제와 관련해 시장에서 많은 이야기를 들었으리라 본다. 그 중 가장 눈길을 끄는 이야기는 오라클과 관련해 나오고 있다. 오라클은 엑사데이터를 모든 사람들을 위한 모든 것들로 표방하고 있다. 오라클은 OLTP(Online Transaction Processing), ODS, 데이터 웨어하우징을 제공하려 할 것이다. 아주 흥미로운 포지셔닝이다. 그러나 네티자와 IBM이 지난 수십 년 동안 발견한 사실은 하나로 모든 것을 충족하는 'One-Size-Fit'은 없다는 것이다.

이와 관련해 워크로드 최적화 시스템에 대한 아이디어는 사실 아주 중요하다. 엑사데이터 같은 제품으로 뭘 할 수 있을까? 모든 사람의 요구사항을 충족하도록 설계된 소프트웨어와 하드웨어를 통합하기 아주 복잡한 시스템으로 구축하는 것뿐이다. 따라서 그 누구의 요구사항도 완벽하게 충족 못할 가능성도 있다. 우리는 특정 목적을 위한 워크로드 최적화 솔루션이 분명히 필요할 것이라고 믿고 있다. 웨어하우징은 네티자, 제퍼디 게임에는 왓슨, 그리고 OLTP, 트랜잭션이 다른 워크로드 처리에는 IBM의 다른 솔루션 같은 것들이다. **CIO**

# 기고 | 단순할수록 좋다

Netezza | 한국IBM

**세**계는 지금 폭발적인 정보의 증가를 경험하고 있다. 매일 15PT의 새로운 정보가 생성되고 있으며 그 중 80%는 이메일, 이미지, 비디오 등에서 대량으로 생성되는 구조화되지 않은 자료다.

기업 전체에서 요구사항과 복잡성이 증가하고 있지만 아직도 대부분의 기업이 정보에 대한 고민에 빠져있으며 40% 가량의 비즈니스 리더들은 신뢰할 수 없는 정보를 토대로, 혹은 아예 정보 없이 결정을 내리곤 한다.

83%의 CIO가 경쟁력 향상을 위한 전략 계획의 필수 요소로 'BI 및 분석 기술'을 꼽을 만큼 데이터를 보다 효율적으로 분석해 비즈니스 통찰력을 가지는 것은 현대 기업의 생존 문제로 부각되기 시작했다.

이제 현대 기업의 성공 여부는 올바른 의사결정을 내릴 수 있도록 적시에 최적의 정보를 보유하고 있는지에 달려 있다고 해도 과언이 아니다. 시간이나 자원을 덜 사용하는 정보라 할지라도 조직에 위기를 초래할 수 있다. 따라서 분석 업무를 끊임없이 개선하는 기업은 비즈니스의 모든 영역에 긍정적인 결과를 가져오는 인텔리전스라는 뜻밖의 혜택을 누릴 수 있다.

그러나 결과를 예측하고 최신 경향을 파악하거나 그에 따른 가장 적합한 조치를 판단하는 데 중요한 정보를 찾는다는 것은 수십억 개의 데이터로 연결된 PT 규모의 데이터를 분석하는 과정을 의미한다.

오늘날의 이른바 고급 분석론은 그 필요조건이 갈수록 까다로워지고 있어 데이터의 양과 복잡성이 기하급수적으로 늘어난 방대한 정보를 안정적으로 받아들여 관리하고 데이터 웨어

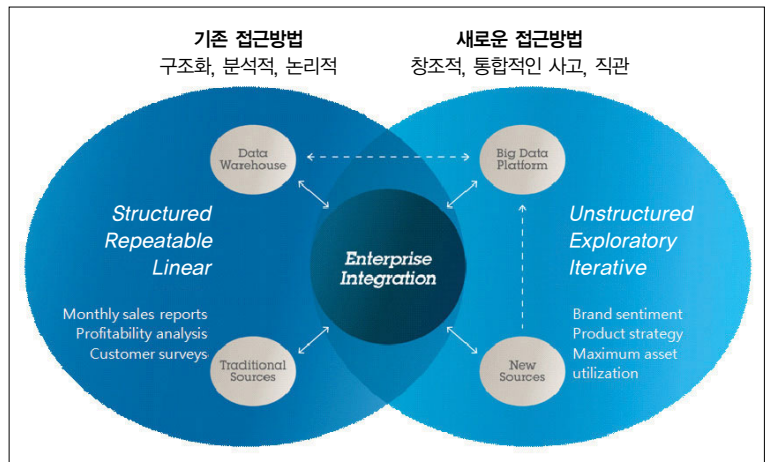
하우징을 통해 언제든 즉각 분석할 수 있도록 만드는 능력이 매우 중요해졌다.

중요한 정보를 긴급하게 찾아야 하는 경우, 이 정보를 제공하는 플랫폼이 관건이 된다. 속도와 정확성에 방해되는 엄청난 볼륨의 워크로드를 간단하게 처리할 수 있도록 조명 스위치처럼 단순하고 신뢰할만하며 즉각적으로 작동해야 한다. 또한 점차 복잡해지는 워크로드를 수행하는 사용자가 늘어나고 데이터 볼륨이 급속하게 증가하는 경우에도, 성능을 그대로 유지하는 아키텍처를 가지고 있어 오랜 기간 사용할 수 있고 최소한의 총 소유 비용(TCO)만 소비해야 한다.

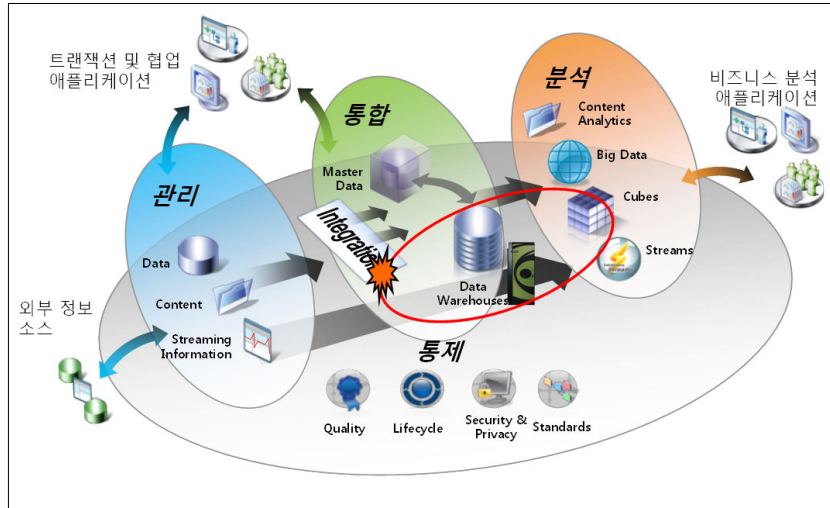
### 전통적인 DW는 너무 복잡하다

가트너의 조사에 따르면, DW의 70% 가량이 성능 문제를 야기시키는 것으로 나타났다. 전통적인 DW들(1세대 데이터웨어하우스)이 트랜잭션 처리 작업에 최적화된 데이터베이스를 기반으로 하기 때문이다. 기본적으로 이 시스템들은 범용 데이터베이스, 서버 및 스토리지 플랫폼으로 짜집기돼 있다. 이 때문에 너무 복잡하고, 또 급속하게 늘어나고 변화하는 TB급 대량의 데이터를 토대로 한 고

빅데이터와 같은 구조화되지 않은 정보들은 새로운 방식의 접근 필요



정보의 Life cycle과 그 안에서 DW의 역할



급 분석 작업의 요구사항을 충족하기에는 전문성이 떨어진다.

또 컴퓨팅 인프라가 갈수록 관리하기 복잡해짐에 따라 기술팀은 대부분의 업무 시간을 문제 해결에 허비하게 된다. 그리고 기술 전문가들은 컴퓨팅 인프라에서 가치를 창출하기 위해 일할 시간을 데이터베이스, 서버, 그리고 스토리지 관리에 허비한다. 그런 여건 때문에 웨어하우스에 소중한 정보가 저장돼 있을지라도 기술팀은 자사의 업무 역할에서 데이터를 근거로 한 의사 결정에 참여하고 이를 주도할 시간이 턱없이 부족하다.

이러한 복잡성은 데이터 본연의 잠재력을 약화시킨다. 결국 비즈니스 기여도가 줄어들고 성능의 이슈가 발생하게 되는 것이다. 이러한 이유로 DW 시장에 어플라이언스라는 개념이 등장하기 시작했다. 어플라이언스는 전용 디바이스와 용도에 맞는 최적화 솔루션으로 이를 단순하게 바꿔 사용자 경험을 변화시키고 비즈니스 사용자들은 원하는 장소와 시점에 통찰력을 신속하게 제공받아 기업의 비즈니스를 가속화하고 경쟁력을 높이기 시작했다.

### 적시에 제공되는 정보

분석은 어렵다. 좋은 분석은 특히 어렵다. 하지만 결과는 가급적 단순해야 한다. 간단명료함을 이루는 것은 어려운 일이다. 경영진에게 직관적인 의견을 전달하고자 할 때, 단순함은 대단히 유용하다. 분석에 영향을 주는 요소들은 제품, 솔루션 제공업

체, 유지 보수 체계, ETL(Extract Transform & Loading), MDM(Master Data Management), DQ(Data Quality), DI(Data Integration), DBMS, OS의 패치 및 업데이트, 서버, 스토리지, 가상화 및 도구 등 다양하다.

따라서 어플라이언스의 관건은 통합 지원이다. 특수 목적 설계, 사전 최적화, 그리고 사전 통합이 이뤄진 솔루션이라면 데이터를 추가하기만 하면

된다. 화가가 물감을 관리하고 붓을 빼는 시간을 없애고 작품 구상과 채색에 대한 고민에 쓸 시간을 더 가질 수 있다면 좋은 작품을 더 많이 볼 수 있는 것은 당연하다. 최소화와 자동화가 결국 단순함을 가져오고 결과에 집중할 시간을 늘려준다.

좋은 솔루션은 때로 가장 규모가 크고 값비싼 제품이 아니라 가장 독특한 설계를 기초로 한다. IBM 네티자의 스트리밍 처리 방식이 기존의 컴퓨팅 아키텍처보다 더 고유한 장점을 지니고 있다. 그리고 대용량 데이터 처리와 분석에 가장 최적화된 시스템을 위해 '데이터베이스+서버+스토리지'를 하나로 통합해 모든 구성을 최적화시킨 어플라이언스 개념을 세계 최초로 DW 시스템에 적용했다.

그 결과 대형 시스템보다 작지만, 수천 명의 사용자가 동시에 이용하는 대용량 데이터와 다양한 워크로드로 구성된 복잡한 알고리즘을 매우 빠른 속도로 실행하는 성능을 갖춘 컴팩트한 솔루션을 선보이고 어플라이언스 가격/성능에 대한 새로운 기준 정립했다. IBM 네티자의 데이터 분석 특허 기술은 업계에서 입증된 기술이다.

데이터가 물리적 속도로 처리되지만 데이터 이동은 최소화된다. 데이터 분석 시 대규모 데이터가 병렬로 처리되는데, 그 내부에는 사용하기 쉽고, 매우 빠르고, 비용도 저렴한 분석 어플라이언스가 근간을 이루고 있다. 이전에는 불가능했던 것으로 생각되었던 분석도 수행할 수 있다. IBM 네티자 고객들은 IBM 네티자의 분석 어플라이언스를 구현해 현

재의 시스템이 갖고 있는 한계를 극복하고 데이터에 대한 처리 방식을 크게 간소화했다.

**안정된 아키텍처 기반 성능**

IBM 네티자 아키텍처는 PT급 데이터를 단시간에 분석할 수 있도록 특별히 설계된 어플라이언스를 구축하기 위해 상용 디스크 스토리지와 개방형 블레이드 기반 서버를 조합하는 IBM 네티자의 고유한 MPP(초고속 병렬 처리) 아키텍처와 FPGA(Field Programmable Gate Arrays)를 사용한 IBM 네티자의 데이터 필터링 특허 기술을 결합했다.

이러한 기술을 조합함으로써 수만 명의 사용자를 지원하는 매우 복잡한 혼합 워크로드에서 쿼리 성능이 빨라지고, 정교한 분석 결과를 생각의 속도로 제공하며, PT 단위까지 모듈식 확장이 가능해진다.

개방형 블레이드 기반의 구성요소를 활용해 성능이나 스토리지 중심의 요구사항에 맞게 구성에서 디스크 프로세서와 메모리 속도를 쉽게 변경할 수 있다. 또한 이 아키텍처는 업무 위주의 애플리케이션에 빠른 실시간 분석을 제공하는 메모리 기반 시스템도 지원할 수 있다. 기본적인 성능 외에도 안정된 아키텍처는 선형 확장성을 제공하므로 천여 개의 처리 스트림을 동시에 실행할 수 있으며 TCO 부담이 적다.

**간소화**


IBM 네티자 시스템은 인덱싱이나 튜닝하지 않고도 바로 고성능 기술을 활용할 수 있는 사용이 편리한 어플라이언스다. 네티자는 정규 업무를 자동화 및 간소화하므로 사용자가 플랫폼의 복잡성을 고려할 필요가 없다. 이 시스템은 데이터를 즉시 로드해 쿼리를 실행할 수 있도록 준비된 상태로 제공된다. 또 표준 ODBC(open database connectivity), JDBC(Java Database Connectivity) 및 OLE(object linking and embedding) DB 인터페이스를 통해 모든 중요 ETL, BI 및 분석 애플리케이션과 통합된다.

IBM 네티자는 모든 하드웨어, 소프트웨어 및 스토리지가 통합된 어플라이언스로 구축 기간이 줄어들고 기존 데이터센터에 쉽게 배치할 수 있으며, 간단한 설치만으로 수 시간 내에 데이터를 로드하고 쿼리를 실행할 수 있다. 또 업그레이드, 고가용성 및 비즈니스 연속성 확보 등 일반적으로 시간이 많이 걸리는 업무도 매우 단순화되므로 시간과 자원을 절약할 수 있다

**고급 분석용 플랫폼**

소스와 밀접한 데이터 처리 및 MPP 원칙은 대용량 데이터 세트의 고급 분석에도 동일하게 적용할 수 있다. IBM 네티자는 조잡한 병렬식 또는 그리드 프로그래밍 과정을 거치지 않고도 복잡한 비 SQL 알고리즘을 MPP 스트림의 처리 요소에 쉽게

**IBM 네티자 - 간단하고 빠른 분석 전용 어플라이언스**



- 저렴한 도입비용, 1TB에서 수 PB까지 확장 가능성
- 고급 분석을 위한 플랫폼 - 뛰어난 성능 (빠른 로딩 속도: 시간당 최대 2TB, 빠른 백업 속도 : 최고 시간당 4TB)
- 수천 명의 사용자는 물론 복잡하고 혼재된 워크로드 지원
- 인덱싱 또는 튜닝 없이 최소 관리 작업만해도 곧바로 사용 가능
- 산업 표준 인터페이스 지원 (SQL, ODBC, JDBC, OLE DB) 및 멀티코어 인텔 기반의 블레이드 아키텍처
- BI, ETL 및 분석 애플리케이션과 완벽한 호환성 확보
- 엔터프라이즈급 안정성 및 가용성 - 99.99% 이상 가동 시간
- 친환경 제품 - 탄소 배출을 저감시키며 저전력 사용 및 저냉각 필요
- 고객 의견
  - "어떤 업무에 24시간이 걸린다면 다른 일은 업무도 못 내지만 10초가 걸린다면 비즈니스 자체를 완전히 다시 생각할 수 있게 된다."
  - "20분만에 미국 전체 인구의 신용정보를 파악할 수 있게 되었다. 게다가 사용방법도 간단하다."
  - "이틀이 걸리던 작업이 한 두 시간으로 줄었다. 별도의 튜닝없이도 획기적인 성능 향상이 이루어져서 새로운 비즈니스를 구상하기 시작했다"





“분석 시스템 정의에서 물리적 파티션, 인덱스, 사전 정의형 집합 함수 및 총계 함수는 몇 개나 됩니까? 디스크 메모리는 충분히 배정했습니까?”와 같은 질문에 IBM 네티자 어플라이언스는 “생각할 필요 없다”라고 답한다. 중요 BI 이니셔티브를 평가하는 데 걸리는 시간을 단축시키면 기업 수익에 긍정적인 영향을 주기 때문이다.

변화하는 시장 환경에 따라 민첩하게 대응해야 하고, 분석 요건이 더욱 고도화되고 복잡해지는 현실에서 설치가 쉽고 신속 실행 및 PT급 데이터 분석이 가능한 시스템을 갖추는 것이 가장 이상적이다.

소비자의 성향은 점점 다양해지고, 비즈니스 환경은 급변하고 있다. 빠른 속도로 끊임 없이 변화하는 고객과 시장의 니즈에 능동적으로 대처해야 기업의 지속가능성(sustainability)을 보장받을 수 있다. 더 많은 데이터, 더 많은 요구 사항, 더 많은 비용으로 이어지는 악순환의 고리를 끊기 위한 해답이 필요한 때다. **CIO**

적용할 수 있게 해준다.

대용량의 데이터를 복잡도에 관계없이 ‘원활하게’ 분석할 수 있는 기능은 데이터를 각 하드웨어로 이동시키는 시간과 비용을 절감시켜 준다. 또한, 네티자는 데이터 크기에 따라 성능을 가속화할 수 있어 데이터 웨어하우징과 고급 분석을 통합하는 이상적인 플랫폼이다.





# Netezza. Up and running in 24 hours, not 24 days.

Get set up in hours instead of days, and start counting returns in minutes instead of hours. All with IBM's Netezza data warehouse appliance for high-performance analytics. It gives you analytics reports at supersonic speeds. At a fraction of the cost of Oracle Exadata. Get real, actionable business results fast.

[ibm.com/facts](http://ibm.com/facts)

# IDG Korea Contacts



## Sales Contact

**김성일, Eddy Kim**

Tel: + 82-2-558-6939 Mobile: + 82-10-2702-0360

Email: sungil\_kim@idg.co.kr

**이포원, Tony (Pown) Lee**

Tel: + 82-2-558-6924 Mobile : + 82-10-2053-2413

Email : tony\_lee@idg.co.kr

## CIO Editorial Contact

**천신웅, Brian Cheon**

Tel: + 82-2-558-9882 Mobile: + 82-10-3221-7824

Email: psy\_cheon@idg.co.kr

**박해정, Jenny Park**

Tel: + 82-70-7725-9955 Mobile: + 82-10-2810-4518

Email: jenny\_park@idg.co.kr

## IT World Editorial Contact

**박재곤, Jay Park**

Tel: + 82-2-558-9887 Mobile: + 82-10-2782-6722

Email: jay\_park@idg.co.kr

**김현아, Hyuna Kim**

Tel: + 82-2-558-6956 Mobile: + 82-10-8300-8511

Email: hyuna\_kim@idg.co.kr

## PR & Marketing Contact

**한정규, Andrew Han**

Tel: + 82-2-558-6909 Mobile: + 82-10-4110-1172

Email: andrew\_han@idg.co.kr

**김종민, Matt Kim**

Tel: + 82-2-558-6079 Mobile: + 82-10-2408-4121

Email: matt\_kim@idg.co.kr



IDG KOREA

4F ChangHwa Bldg., Bongnae-dong 1Ga, Jung-Gu, Seoul, Korea 100-161