



본 보고서에서는 빅데이터와 엔터프라이즈 애널리틱스 부문을 자세히 살펴보고, 애널리틱스 사용 모델을 비교하여 차이점을 알아보면서, 몇 가지 정의를 제시하고자 합니다. 또한, 비용 및 성능 면에서 x86보다 뛰어나다는 IBM의 새로운 플랫폼도 자세히 알아보겠습니다. IBM은 POWER 기반 시스템으로 경쟁 제품인 x86을 능가할 수 있을까요?

일반적으로 "빅데이터"로 알려진 엔터프라이즈 애널리틱스가 경제 혁명을 일으키고 있습니다. 엔터프라이즈 애널리틱스는 빠르게 확산되어 비즈니스에서 경쟁사를 압도할 수 있는 가장 중요한 무기가 되고 있습니다. 빅데이터를 통해 조직은 어떤 것을 잘하는지, 더 잘하려면 무엇이 필요한지, 어떻게 고객을 좀 더 효과적으로 지원할 수 있는지, 미래를 준비하려면 어떻게 해야 하는지를 이해할 수 있습니다.

'빅데이터'를 활용할 수 없는 조직은 결국 차세대 경제 성장 및 경쟁에서 뒤처지게 될 것입니다. 한마디로 빅데이터는 너무나도 중요합니다. 그런데 빅데이터가 정확히 무엇일까요?

간단히 말하자면 '빅데이터' 구현이란 대량의 데이터를 수집하고 분석하여 직관적이지 않은 관계를 찾아내고 새로운 기회와 위협을 발견하는 능력을 설명하는 일반 용어가 되었습니다.

월스트리트에서는 거래 알고리즘 및 모델을 개발하는 데 이를 사용하며, 체인 소매업체에서는 적절한 시점의 적절한 장소에 적절한 재고를 확보해두는 데 사용합니다.

고객과 경쟁사에 대한 정보를 수집하는 것은 전혀 새로운 일이 아닙니다. 지금까지 늘 정보를 모으고, 이를 분석하여, 유용한 통찰력을 도출해왔습니다. 다만 충분하지 않았을 뿐입니다. 정보가 충분하지 않았고, 분석 성능이 충분하지 않았고, 통찰력이 충분히 정확하지 않았었습니다.

빅데이터가 "새로운" 점은 폭발적인 양의 정보에 액세스할 수 있다는 것입니다. 센서, 신용 카드, 전화, 컴퓨터, 소셜 미디어, 네트워크에 연결된 디바이스 등 모든 곳에서 정보가 전달됩니다. 그 양은 놀라울 정도로 많습니다. 어떤 조사에 따르면 전 세계 데이터의 90%가 지난 2년간 생성된 데이터라고 합니다.

빅데이터의 중요한 고려 사항

이러한 데이터 홍수 속에 어떤 데이터는 의미가 있고 유용하며 어떤 데이터는 그저 "잡음(noise)"에 불과합니다. 자주 인용되는 "빅데이터의 4개 V"에 하나를 추가해보겠습니다.



볼륨(Volume) – 얼마나 많이
속도(Velocity) – 얼마나 빠르게
신뢰성(Veracity) – 얼마나 안정적으로
다양성(Variety) – 어떤 형태로

추가해야 할 하나는 가치(Value)입니다. 모든 데이터가 가치있는 것은 아닙니다. 하지만 잡음(noise) 속에서도 꼭 필요한 결정적인 데이터가 있습니다. 찾을 수만 있다면 통찰력, 전략 및 목표를 바꿀 수 있는 데이터 말입니다.

어디를 찾고 있습니까? 많은 데이터 과학자가 아주 작은 수의 변수(아마도 5% 정도)를 완벽하게 만들면 비즈니스 매출에 90% 이상의 효과를 얻을 수 있다고 말합니다. 다시 말해 제품/서비스의 일부 특성이 '최적점(sweet spot)'에 적중하면, 100%를 적중하는 것과 거의 같은 효과를 낼 수 있다는 것입니다.

예를 들어, 자동차 제조사가 고객 요구에 맞추도록 특정 모델의 항목 100개를 조정하거나 변경할 수 있다고 가정해보겠습니다. 고객과 시장에 대한 광범위하고 철저한 조사를 통해 제조사는 5~10개 항목만을 변경하여 100개 항목 모두를 변경하는 것과 거의 동일한 효과를 거둘 수 있는 "마법의 데이터"를 발견할 수 있을지도 모릅니다.

이러한 "마법의 데이터"를 파악하고 수량화하는 것이 새로운 분석 목표입니다. 고객에게 가장 중요한 핵심 특성은 산업 및 제품마다 다릅니다. 마법의 데이터 관계 또한 시장이 성숙해 지면서 달라집니다.

마법의 데이터를 발견한 기업은 많지 않으며, 이미 발견한 기업은 이를 공개하지 않습니다. 대부분은 고객이 경쟁 제품 대신 자사의 제품을 선택한 이유를 정확히 알지 못합니다. 분석은 짐작이나 추측을 대신하여 조직의 가장 중요한 자산인 고객과의 관계에서 진정한 통찰력을 제공할 수 있습니다.

빅데이터 유형: 정형 대 비정형

수집하고, 처리하고, 분석해야 하는 빅데이터에는 크게 두 가지 유형이 있습니다. 첫 번째는 "정형 데이터"로서 우리에게 친숙한 유형입니다. 고객이 웹 사이트에서 입력하는 이름, 청구 주소, 배송 주소 등은 정형 데이터입니다. 도로 주소와 우편 번호 같은 정보는 모두 일정한 형식으로 정리되어 있으므로 운영 또는 분석 목적으로 사용하는 기존 관계형 데이터베이스의 필드에 깔끔하게 들어갑니다.



그러면 “비정형 데이터”는 무엇일까요? 기본적으로 그 외의 모든 것이라고 보면 됩니다. 관계형 데이터베이스 테이블에 적합하지 않은 모든 디지털 데이터입니다. 비정형 데이터는 전혀 다른 유형으로 분석을 상당히 어렵게 만드는 장본인입니다.

이메일이 비정형 데이터의 좋은 예가 될 수 있습니다. 표준 형식 없이 문장으로 구성되어 있지만, 매우 중요한 정보를 담고 있습니다. 웹 트래픽, Facebook 게시물, Twitter 피드 및 블로그 또한 비정형 데이터이면서 중요한 데이터 소스입니다.

비정형 데이터는 분석하기 어렵습니다. 이유는 정형화되지 않았기 때문입니다. 유용한 정보를 추출하는 것은 일반적으로 텍스트의 패턴을 분석하여 맥락을 이해하고 그 뒤에 숨겨진 의미를 찾아내는 작업입니다. 이를 지원하기 위해 많은 도구가 개발되었고, 본 문서의 뒷부분에서도 이러한 도구를 살펴보겠습니다.

정형 데이터와 비정형 데이터 모두 폭발적으로 증가하고 있습니다. 관련 통계는 너무나도 많지만 매년 생성되는 디지털 데이터의 양이 40~50% 증가하고 있다고 보면 됩니다. 즉, 전 세계의 데이터가 2년마다 약 두 배씩 증가하고 있습니다. IDC에서는 2020년이면 1년에 생성, 복제 및 사용되는 전체 데이터가 40제타바이트(40조 기가바이트에 해당)에 이를 것으로 예상합니다.

빅데이터 분석

그렇다면 지금까지 이야기한 중요하고 혁신적인 데이터는 모두 어디에 있을까요? 어디에나 있습니다. 여러분의 회사 내부에서도 상당한 데이터를 생성합니다. 아마도 조직의 데이터 웨어하우스 사일로에 저장되어 있을 것입니다. 다른 유용한 데이터는 센서, 웹 크롤러, 타사 데이터 브로커, 소셜 미디어 네트워크 등을 통해 조직 외부에서 유입됩니다.

하지만 한가하게 이 모든 데이터를 저장하고 분석할 수는 없습니다. 먼저 스토리지 및 컴퓨팅 능력이 충분하지 않습니다. 또한, 엔터프라이즈에서는 적시에(대부분 거의 실시간) 의사 결정을 내리도록 빅데이터의 통찰력을 사용할 수 있어야 합니다. 의사 결정을 내리기 전에 모든 데이터를 철저히 분석할 시간 여유가 있는 회사는 거의 없습니다.

데이터 분석 모델은 크게 ‘데이터 호수(lake)’와 ‘데이터 강(river)’ 두 가지로 나눌 수 있습니다. 데이터 호수는 일반적인 엔터프라이즈 데이터 웨어하우스에 보관되는 정형 데이터입니다. 이 데이터는 보통 “저장 데이터(data at rest)”라고 부르며, 장기간(몇 개월 또는 몇 년) 저장되고, 다양한 분석 프로세스에 사용됩니다.



하지만 많은 엔터프라이즈에서 실시간에 가까운 운영이나 전략적 의사 결정을 내리려면 데이터 스트림 즉, 빠르게 흐르는 데이터 강(river)도 분석해야 합니다. 이러한 분석 유형의 가장 좋은 예는 신용 카드 처리일 것입니다. 고객이 매장에서 신용 카드나 체크 카드를 긁으면, 트랜잭션을 처리하는 금융 기관에서는 많은 분석 루틴을 수행해야 합니다. 이러한 프로세스를 통해 해당 카드가 유효하고, 고객이 한도를 넘지 않았으며, 부정하게 사용된 것이 아님을 확인합니다. 1~2초 만에 완료되어야 하는 매우 복잡한 프로세스입니다.

대부분 비즈니스는 저장 데이터(데이터 호수) 및 전송 데이터(스트리밍 데이터 또는 데이터 강)를 처리해야 합니다. 이러한 두 가지 유형의 데이터는 숨겨져 있거나 직관적이지 않은 트렌드에 대한 통찰력을 확보하기 위해 결합되어 있을 때가 많습니다. 지난 몇 년간 이러한 작업으로 인해 발생하는 거대한 양의 데이터를 처리하기 위해 많은 도구가 설계되었습니다. 현재 빅데이터 처리의 경우 하둡이 아마도 가장 익숙한 이름일 것입니다(하둡은 뒷부분에서 자세히 다루겠습니다).

현재 데이터 분석 시장

이러한 트렌드를 주도하는 공급업체의 시장 규모는 매우 커졌습니다. 모든 조직이 이미 이와 관련해 무언가를 하고 있습니다. 데이터 분석 시장은 2017년까지 연평균 58% 성장하여 500억 달러 규모에 이를 것으로 예측됩니다.

빅데이터 시장에서 가장 두각을 나타내는 업체는 IBM입니다. 지난 10년간 IBM은 전체 빅데이터 솔루션 개발에 150억 달러 이상을 투자했습니다. 투자에는 Watson Lab, IBM Global Services business advisory 및 Systems & Technology Group을 통한 유기적인 기능 및 기술 개발도 포함됩니다.

또한, IBM은 빅데이터 포트폴리오를 구성하기 위해 다수의 주요 업체를 인수하기도 했습니다. 서비스 부문으로는 컨설팅 회사인 Price Waterhouse를, 하드웨어/소프트웨어 회사로는 SPSS, Cognos, Netezza, Platform Computing, SoftLayer 등 다수 업체를 인수했습니다. IBM은 이러한 회사들의 "비밀 소스"를 IBM의 빅데이터 오퍼링에 통합하여 다양하고 심층적인 빅데이터 솔루션 포트폴리오를 구성했습니다.

IBM의 빅데이터 전략: 자세히 살펴보기

물론 업계의 모든 공급업체가 어느 정도 빅데이터를 지원한다고 주장합니다. IBM은 빅데이터 솔루션을 구축하기 위한 서비스, 소프트웨어 및 시스템을 갖추고 있을 뿐 아니라 대부분 공급업체에서 따라올 수 없는 경험을 보유하고 있다는 점에서 다른 업체와 차별화됩니다.

IBM은 10년 가까이 HPC(슈퍼컴퓨팅) 부문에서 독보적인 위치를 유지해왔으며, 전 세계에서 가장 크고 빠른 컴퓨터 상위 500개 중 40% 이상이 IBM 컴퓨터입니다.

대량의 복잡한 컴퓨팅 요구 사항의 경우, HPC 워크로드보다 큰 경우는 없습니다. 이러한 고객은 일기 예보, 지진 처리, 원자 규모의 상호 작용 모델링과 같은 작업을 수행합니다. 이러한 유형은 처리가 매우 까다롭고 짧은 시간에 결과를 제공하도록 시스템을 확장할 수 있어야 합니다(예: 허리케인 모델링). 그 뿐만 아니라 분자 프로세스를 모델링하도록 몇 일 또는 몇 달 동안 안정적으로 실행할 수 있어야 합니다.

정교한 소프트웨어, 뛰어난 성능

앞에서 트렌드에 대한 통찰력을 얻고 기회를 활용하려면 “저장 데이터(data at rest)”와 “전송 데이터(data in motion)” 모두 분석해야 한다는 점을 설명했습니다. 빅데이터의 특성 중 하나는 대부분 비정형 데이터이며, 일반적인 관계형 데이터베이스에 저장하여 처리하기가 불가능하거나 비현실적인 대량 볼륨으로 전달된다는 것입니다.

하둡은 현재 빅데이터 워크로드 처리에 가장 많이 사용되는 도구입니다. 하둡과 기존 데이터 웨어하우스의 가장 큰 차이점은 아키텍처입니다. 기존 데이터 웨어하우스는 사용자 및 다른 애플리케이션을 위한 중앙 데이터 저장소로 사용되는 관계형 데이터베이스의 단일 인스턴스상에 배포됩니다. 쿼리는 호스트 서버에서만 데이터베이스를 대상으로 실행됩니다.

하둡은 원래 대량의 비정형 데이터를 수집하고 신속하게 분석하기 위해 설계되었습니다. 단일 시스템에서는 대량의 데이터를 감당하지 못할 것이므로 하둡은 여러 시스템(경우에 따라 수천 개)에서 실행되도록 구축되었습니다. HDFS(Hadoop Distributed File System)는 클러스터의 여러 시스템(또는 노드)에 데이터를 분산하여 스토리지 작업을 처리합니다.

하둡이 작업을 처리할 때, 관련 데이터를 가지고 있는 노드로 작업을 분산합니다. 노드에서 작업이 완료되면, 네임노드(NameNode)로 결과를 보고하고, 네임노드에서 결과를 사용자에게 전달합니다.

하둡이 유용한 도구이지만 문제점도 있습니다. 가용성을 확보하기 위해서는 3개의 데이터 복사본이 유지되어야 합니다. 또한, 하둡은 좀 더 정교한 스케줄러와 효과적인 데이터 배치를 활용할 수 있습니다.

IBM의 InfoSphere BigInsights는 하둡의 오픈 소스 버전을 훨씬 더 빠르고 리소스 사용이 더 효율적이도록 개선했습니다. 같은 하드웨어에서 테스트했을 때, BigInsight 버전 하둡(Platform Symphony 스케줄러 사용)이 표준 Apache 하둡보다 300% 이상 뛰어난 성능을 보였습니다.



IBM의 InfoSphere BigInsights를 활용하고 있는 고객 중 하나가 네덜란드 국립천문대인 ASTRON입니다. ASTRON의 SKA(Square Kilometer Array)는 지구에서 가장 크고 가장 강력한 망원경입니다. 또한, ASTRON은 지구에서 가장 큰 데이터 스트리밍 문제를 안고 있습니다. SKA를 구성하는 3,000개의 접시는 매일 1엑사바이트 이상을 출력합니다. 이는 24시간 동안 전 세계 인터넷에서 생성하는 데이터의 2배에 해당합니다.

ASTRON은 IBM BigInsights를 사용하여 빗발치는 스트리밍 데이터를 며칠이 아니라 몇 분 만에 천문학자가 사용할 수 있도록 처리하고 있습니다. 이는 다른 솔루션 대비 99%의 가속도입니다. 또한, IBM 솔루션은 데이터를 꼼꼼히 검토하여, 추가 분석을 위해 다양한 위치로 이를 전달하고 필요 없는 데이터는 폐기합니다. 대량의 스트리밍 데이터 로드를 처리하는 성능은 IBM의 아키텍처 접근 방식과 소프트웨어 포트폴리오가 성공적임을 보여줍니다.



좀 더 실제적인 예로는 온라인 채팅과 영화 박스 오피스 매출의 관계를 이해하기 위해 BigInsights와 IBM InfoSphere Streams를 사용하는 서던캘리포니아 대학교의 Annenberg Lab을 들 수 있습니다.

구체적으로 설명하자면, 이 연구소에서는 예를 들어 TV의 영화 광고로 소셜 미디어 활동이 증가하는 경우, 소셜 미디어 활동이 박스 오피스 매출로 이어지는지를 파악하기 위해 IBM 분석 시스템을 사용하고 있습니다. 또한, 아카데미 어워드 방송 같은 생방송 이벤트 PD에게 거의 실시간 피드백을 제공하는 데도 사용합니다. 쇼가 청중에게 어떻게 받아들여지는지 측정할 수 있습니다.

이를 위해 Annenberg 팀은 수백만 개의 트윗과 다른 소셜 미디어 언급을 추적하고, 분석한 다음, 가능한 한 빠르게 결과를 보고합니다. 사용자가 시스템에게 자연어의 뉘앙스를 평가하는 방법을 가르칠 수 있어야 하므로 소프트웨어 패키지는 유연해야 합니다. 소셜 미디어 게시물의 큰 부분을 차지하는 풍자적 표현을 구분할 수 있도록 시스템을 가르칠 수 있습니다.

Annenberg 팀은 IBM 하드웨어 및 소프트웨어 스택과 함께 BigInsights 도구 상자를 사용하여 소셜 네트워크를 분석함으로써 어떤 이벤트에 대한 피드백도 안정적으로 아주 빠르게 제공할 수 있습니다.

새로운 프로세서는 크게 도약하는 디딤돌이 됩니다.

IBM 빅데이터 역사상 가장 흥미로운 개발 제품은 최근에 출시된 POWER8 프로세서를 탑재한 서버 제품군일 것입니다. 새로운 프로세서는 IBM으로서는 커다란 도약이며 어떤 면에서는 이전 프로세서와 완전히 차별화된다고 할 수 있습니다.

지난 몇 년간 IBM은 Power Linux Scale-out 시스템이 특히 빅데이터와 분석 워크로드 부분에서 x86 기반 시스템을 대체한다고 홍보해왔습니다. 이러한 시스템이 표준 Linux에서 실행되긴 하지만(POWER 프로세서용으로 컴파일된 버전일지라도) 고객과 ISV에게는 여전히 전혀 다른 제품으로 여겨졌습니다.

가장 큰 차이점은 POWER 프로세서는 “빅 엔디언”이고 x86 프로세서는 “리틀 엔디언”이라는 것입니다. 이러한 ‘엔디언’은 바이트가 메모리에 저장되는 순서를 말합니다. 빅 엔디언 시스템은 가장 큰 바이트를 가장 작은 주소 공간에 저장하고 리틀 엔디언 시스템은 가장 작은 바이트를 가장 작은 주소 공간에 저장합니다.

고객은 알 수 없는 작업이며 개발자와 ISV 외에는 의식하지 못하는 부분입니다. 빅 엔디언과 리틀 엔디언 시스템 간의 코드를 변환하는 것은 가능한 일일 뿐 아니라 그다지 어려운 일이 아니지만, 이를 위해서는 ISV와 개발자가 추가 작업을 해야 합니다. 일부에서는 x86의 대안으로서 POWER 기반 시스템을 제안하는 IBM의 노력을 무산시킬 만큼 이 문제가 큰 이슈라고 생각합니다.

하지만 POWER8 프로세서의 등장으로 이 문제는 말끔히 해결되었습니다. 새로운 POWER8 프로세서는 리틀 엔디언을 지원합니다. 즉, x86 표준과 호환되어 POWER8 기반 시스템에서 구동되는 소프트웨어에 새로운 세상을 열어줍니다.

Java, Pearl, PHP, Python 등의 스크립팅 언어 또는 해석(interpretative) 언어로 작성된 Linux 애플리케이션은 POWER8 시스템에서 추가 작업 없이 실행할 수 있습니다. C 또는 C++로

© 2002-2014 Gabriel Consulting Group. Inc. All Rights Reserved



작성된 애플리케이션은 소스 코드의 변경 없이 간단히 다시 컴파일만 하면 됩니다. 이는 Power 시스템에 있어 커다란 진전이며 이 기능 하나만으로도 x86의 확실한 대안이 될 수 있습니다.

하지만 이 기능이 전부가 아닙니다. 그 전에 잠시 과거를 되짚어 보겠습니다.

21세기로 접어든 이후로 고성능 서버 프로세서 분야에서 IBM과 Intel은 후발주자였습니다. AMD는 64비트 및 멀티 코어 프로세서를 출시했을 때 잠시 각광을 받았지만 그리 오래가지 못했습니다.

약 2년 마다 IBM은 Intel의 서버 프로세서보다 훨씬 뛰어나고 새로운 기능이 추가된 POWER 프로세서를 출시했습니다. 이후 몇 년 동안 Intel은 IBM의 선두와 접근 방식을 조금씩 따라오거나 동등한 위치를 지켜왔습니다. 그러나 IBM이 또 다른 Power 버전을 출시하면서 상황은 새롭게 바뀌었습니다.

POWER8의 출시로 IBM은 다시 한 번 Intel을 뛰어넘었습니다. IBM POWER8과 Intel Xeon(Ivy Bridge)의 최상위 제품군을 비교해보겠습니다.

| | Intel E7-8890 v2 | Intel E7-8893 v2 | IBM POWER8 |
|-------------------|--------------------------|--------------------------|---------------------------------|
| 출시일 | 2014년 1분기 | 2014년 1분기 | 2014년 2분기 |
| 최대 코어/스레드 | 15개 코어 30개 스레드 | 6개 코어 12개 스레드 | 12개 코어 96개 스레드 |
| 주파수 - 모든 코어 (GHz) | 2.8 | 3.4 | 3.02 - 4.0 |
| 최대 주파수 | 3.4, 터보 (일부 코어 유휴 상태) | 3.7, 터보 (일부 코어 유휴 상태) | 일부 코어에서 최대 약 5.0 |
| 온칩 캐시 | 37.5MB | 37.5MB | 100 MB (128MB 오프칩 추가) |
| I/O | PCIe Gen 3.0 | PCIe Gen 3.0 | PCIe Gen 3.0 |
| 최대 메모리 | 1,536GB | 1,536GB | 1,000GB |
| 메모리 대역폭 | 초당 85GB | 초당 85GB | 초당 230GB |

위의 표에서 굵은 글씨는 사양별로 가장 뛰어난 프로세서를 나타냅니다. IBM의 새로운 POWER8 프로세서가 최대 메모리를 제외하고 거의 모든 항목에서 Intel보다 뛰어나다는 것을 볼 수 있습니다.

하지만 머지않은 시점에 프로세서당 1TB라는 POWER8의 RAM 크기 제한이 2TB로 상향 조정될 것이라는 소문에 따르면 최대 메모리조차도 Intel을 앞서게 될 것입니다.

© 2002-2014 Gabriel Consulting Group. Inc. All Rights Reserved



CAPI 소개

IBM은 POWER8에 빅데이터 및 분석 워크로드에 적합한 새로운 인터페이스를 적용했습니다. CAPI(Coherent Accelerator Processor Interface)라고 불리는 인터페이스로서 CPU 및 PCIe 디바이스가 메모리를 공유하도록 해줍니다. PCI 기반 액셀러레이터와 CPU는 공유된 메모리를 사용하여 더 적은 시간에 훨씬 더 많은 데이터를 처리할 수 있습니다.

IBM의 초기 테스트는 플래시 메모리와 CAPI를 사용하여 수행되었으며, 매우 인상적인 결과를 보였습니다. IBM에 따르면, 24개의 분산 x86 시스템을 사용하는 NoSQL 데이터베이스를 40TB CAPI 지원 플래시 메모리 어레이가 장착된 단일 POWER8 노드에 담을 수 있습니다. POWER8 솔루션은 80% 저렴한 비용으로 같은 성능을 제공할 수 있습니다.

플래시 메모리로의 초고속 연결성을 제공하는 CAPI는 많은 엔터프라이즈와 HPC 애플리케이션에 매력적인 사용 사례가 됩니다. 또한, CAPI는 GPU 액셀러레이터로의 빠른 공유 메모리 액세스를 제공하며, 향후 FPGA도 지원 예정입니다.

POWER8과 x86 비교: 시스템 수준 성능

칩 사양 비교도 흥미롭지만, IBM의 최신 프로세서와 최첨단 x86 시스템 간의 시스템 수준의 비교가 매우 궁금합니다. 아직 1대1 비교 벤치마크가 많지 않지만, 살펴본 결과는 POWER8 시스템의 장점을 확인하기에 충분했습니다.

처음 소개된 POWER8 벤치마크 중 하나는 자주 사용되는 SPECint_rate와 SPECfp_rate입니다. IBM 시스템은 4개의 6코어 CPU와 512GB RAM이 탑재된 2개의 듀얼 소켓 노드로 구성되었습니다.

| | 프로세서 | 프로세서 수 | 코어 | 스레드 | 메모리 | 벤치마크 결과 | 날짜 | 결과와 Power8 비교 |
|------------------|-----------------------|--------|----|-----|-------|---------|--------|---------------|
| SPECint2006_rate | Power8 | 4 | 24 | 192 | 512GB | 1,750 | 5월 14일 | 100% |
| | Intel Xeon E7-8892 v2 | 4 | 24 | 48 | 1TB | 1,200 | 5월 14일 | -46% |
| | Intel Xeon E5-4607 v2 | 4 | 24 | 48 | 512 | 905 | 4월 14일 | -93% |
| SPECfp2006_rate | POWER8 | 4 | 24 | 192 | 512GB | 1,370 | 5월 14일 | 100% |
| | Intel E7-8893 v2 | 4 | 24 | 48 | 1TB | 1,060 | 5월 14일 | -29% |
| | Intel E5-4607 v2 | 4 | 24 | 48 | 1TB | 766 | 4월 14일 | -79% |

© 2002-2014 Gabriel Consulting Group. Inc. All Rights Reserved



POWER8과 x86 프로세서의 최고 성능을 비교하기 위해 표에서 유사하게 구성된 Intel E7 시스템의 결과를 선택했습니다. POWER8 기반 시스템이 Xeon E7 기반 시스템보다 정수(integer) 워크로드에서 46% 그리고 부동 소수점 테스트에서 30% 앞선 것을 볼 수 있습니다.

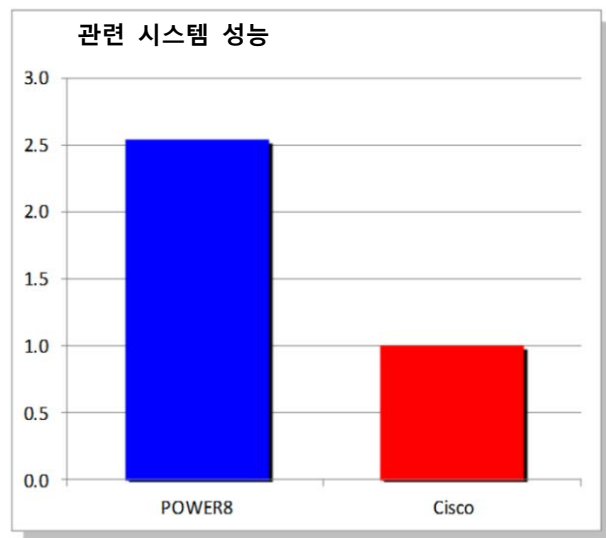
POWER8과 Intel 서버 칩의 대부분을 차지하는 Intel E5 프로세서를 비교하면 그 차이는 더욱 극명해집니다. E5 프로세서는 E7 칩처럼 고가는 아니지만 좋은 성능을 제공합니다.

Power 시스템과 같이 4개의 프로세서(24코어)로 구성된 E5 프로세서를 지원하는 시스템에서 가장 높은 결과를 선택했습니다. 이렇게 벤치마크를 선정하여도 POWER8 시스템의 점수가 가장 높은 E5 정수(integer) 점수의 두 배 그리고 부동 소수점 결과에서는 79% 높은 성능을 나타냈습니다.

이러한 벤치마크가 흥미롭고 의미가 있지만, POWER8이 빅데이터 및 분석 워크로드에서 얼마나 뛰어난 성능을 보이는지를 나타내는 것은 아닙니다. 안타깝게도 아직은 시스템의 일반적인 분석 기능에 대한 산업 표준 벤치마크가 수립되지 않았습니다. 또한, 현재 빅데이터/분석 벤치마크의 경우 일반적으로 사용할 수 있는 결과가 많지 않습니다.

IBM이 새로운 POWER8 시스템을 최적의 빅데이터/분석 솔루션으로 포지셔닝하고 있으므로, POWER8 성능을 뒷받침하기 위해 IBM이 제공한 첫 번째 벤치마크 중 일부가 빅데이터/분석과 관련된 기준임은 당연한 일일 것입니다.

24코어 POWER8 시스템은 Intel Ivy Bridge 기반 시스템과 비교하여 67% 빠른 성능을 제공합니다. 두 시스템 모두 24코어, 384GB 메모리 및 같은 Linux/DB2/Cognos 10.2 소프트웨어 스택으로 구성되었습니다.



또한, IBM은 하둡의 경우 상당히 많은 투자를 했습니다. IBM의 POWER8 출시 이벤트에서 다음과 같은 결과와 왼쪽의 도표를 공개했습니다.

*** POWER8 S822L은 최고 사양의 x86 시스템보다 2.5 배 뛰어난 성능을 제공할 것이며 계속해서 훨씬 탁월한 RAS를 제공할 것입니다.**

이는 POWER8 TeraSort 벤치마크가 오퍼링에 있다는 의미입니다. TeraSort는 시스템이 분석을 위해 대용량(10TB) 데이터를 생성, 정렬 및 확인하는 데 걸리는 시간을 측정하는 하둡 기반 벤치마크입니다.

몇몇 주요 공급업체에서 자사 시스템의 TeraSort 점수를 공개했습니다. 가장 최신 보고서 중 하나는 Cisco의 2013년 8월 보고서로 HP의 2012년 TeraSort 결과를 앞질렀습니다. IBM이 POWER8과 TeraSort로 Cisco와 HP를 모두 누를 것으로 보입니다.

경제성: POWER8과 x86 비교

IBM의 새로운 Linux 구동 POWER9 기반 시스템은 구성이 비슷한 x86 시스템보다 비용 우위를 유지할 것입니다. 예, "비용 우위"라는 말이 맞습니다.

2012년 초 Power Linux 시스템을 처음 출시하면서 IBM은 같은 급의 x86 솔루션보다 구매 비용이 낮고, 운영 비용도 낮지만, 성능은 뛰어난 시스템을 제공하겠다고 약속했습니다.

실제로 IBM POWER8 하드웨어 가격은 HP 또는 Dell이 비슷한 구성의 x86 시스템에 책정한 가격과 큰 차이가 없습니다. 그러나 고객은 Linux 지원에 비용을 더 적게 지불하고 가상화 같은 시스템 도구(VMware가 아니라 KVM 사용)에는 더 적은 비용을 지불하게 될 것입니다.

이러한 부분과 POWER8의 x86 대비 뛰어난 성능이 합쳐지면 고객은 x86 시스템에 지불한 비용과 같은 비용으로 최소 20% 이상의 성능을 확보할 수 있습니다. 일부 애플리케이션에서는 POWER8의 x86 대비 가격/성능 이점이 20%보다 훨씬 클 것입니다.

POWER8 개방성 강화

IBM은 POWER8을 소개할 때 칩과 박스에 대해 언급하지 않았습니다. 대부분 대화와 뉴스가 POWER8 플랫폼상에서 개발하는 업계 참가자의 그룹인 새로운 오픈파워 파운데이션에 집중되었습니다(<http://OpenPowerFoundation.org/>).

이러한 컨소시엄은 누구나 아는 기술 기업(Google, 삼성, NVIDIA, Micron, Mellanox 등)과 그 보다는 잘 알려지지 않은 기업(Inspur, ZTE, Xilinx 등)으로 다양하게 구성되었습니다.

IBM의 POWER8 프로세서가 지원하는 진정한 새로운 개방형 플랫폼을 위한 에코시스템을 구축할 기회가 다양한 파트너를 함께 모이게 했습니다. 흥미로운 점은 IBM이 아니라 Google이 의장직을 맡고 있고 IBM이 사장직을 맡았다는 것입니다. 나머지 이사진은 NVIDIA, Mellanox, Altera, Tyan 및 다른 파트너의 대표로 구성되어 있습니다.

가장 큰 서버 고객인 Google은 시스템 성능, 가격/성능, 효율성 등의 부문을 더 발전시키는 데 관심이 매우 높습니다. Mellanox와 NVIDIA는 다른 파트너와 함께 각각 I/O와 컴퓨팅 효율성을 극대화하는 작업을 진행하고 있습니다. 지난 3월 NVIDIA가 고속 NVLink(현재 기술보다 5~12배 빠르게 상호 연결하는 시스템)를 출시했을 때, IBM의 Power Systems에서 이를 조기에 도입했다고 발표했습니다.

컨소시엄에는 Inspur, ZTE, Servergy 같은 시스템 공급업체도 있습니다. 이러한 업체는 x86 시스템 및 심지어 IBM 브랜드인 Power Systems의 대안으로서 자체적인 POWER8 기반 시스템을 개발, 생산 및 판매할 수 있을 것입니다. 경쟁이 치열한 시장은 최종 사용자에게 더 많은 선택권을 주고 공급업체는 고객의 선택을 받기 위해 지속적으로 제품 및 가격을 개선하게 합니다.

오픈파워 파운데이션에 합류한 파트너는 모두 자신만의 이유가 있지만, 이러한 개별적인 목표 추구를 통해 플랫폼은 더 빠른 속도로 발전할 것입니다.

요약 및 추천

새로운 POWER8 프로세서와 관련 시스템은 단순히 새로운 칩이 아닙니다. POWER8의 기술 변화(기본적으로 “리틀 엔디언” 메모리 순서로의 전환)는 사용자에게 많은 Linux 애플리케이션을 수정 없이 사용하거나 간단히 다시 컴파일하여 사용할 수 있게 해줍니다.

새로운 POWER8 프로세서를 통해 플랫폼에서 실행될 수 있는 Linux 애플리케이션이 획기적으로 증가했습니다. 이렇게 많은 애플리케이션이 IBM의 POWER8 플랫폼 요금과 결합되었을 때 x86을 상대로 시장에서 경쟁할 수 있을 것입니다. 공급업체가 고객에게 더 큰 가치를 제공하기 위해 경쟁함에 따라 이러한 경쟁의 진정한 수혜자는 고객이 될 것입니다.

오픈파워 파운데이션은 IBM에게 새로운 장을 열어줍니다. IBM이 파트너에게 전체 하드웨어 플랫폼을 모두 개방한 것은 POWER8이 처음입니다. 지금의 치열한 하드웨어 시장을 감안하면 매우 적절한 행보입니다.

오픈파워 파운데이션은 POWER8 플랫폼을 경쟁 우위를 확보할 수 있게 해 줄 도구로 보는 매우 의욕적인 기업을 한데 모았습니다. 컨소시엄 참가 회사는 뒤쳐지는 업체가 아닙니다. 참가한 기업은 각자의 분야에서 선도업체이며 Google의 경우 전 세계에서 가장 큰 IT 소비자입니다.

POWER8 시스템이 성장하고 오픈파워 파운데이션이 물살을 타기 시작하면서 서버 구매자의 마음을 사로잡기 위한 경쟁을 목격하게 될 것입니다. IBM은 새로운 하드웨어, 개방적인 태도 및 컨소시엄 파트너를 통해 상당한 시장 점유율을 확보하기에 좋은 포지션을 형성했습니다.

하지만 쉽지는 않을 것입니다. 데이터 센터 근무자의 대부분이 x86 시스템 이외에는 익숙하지 않으므로 POWER8이 검토해볼 충분한 가치가 있음을 설득해야 할 것입니다. Intel과 AMD 같은 기존의 x86 업체도 손을 놓고만 있지 않고 주도권을 빼앗기지 않기 위해 최선을 다할 것입니다. 이는 IBM에게 그리고 POWER8과 한 배를 탄 모든 이에게 어려운 과제가 될 것입니다.

물론 IBM과 오픈파워 파운데이션에 유리한 요소도 있습니다. 먼저, 많은 고객이 빅데이터/비즈니스 분석 목표가 시설과 예산 제한이라는 벽에 부딪히는 시점이 곧 올 것입니다. x86과 비교하여 뛰어난 POWER8의 성능은 이러한 고객의 관심을 끌게 될 것이며, 특히 기존 데이터 센터 공간의 일부에 얼마나 더 강력한 프로세싱 파워를 설치할 수 있는지 알게 되면 더욱 큰 주목을 받게 될 것입니다.

이제 POWER8 시스템은 사실상 어떤 Linux 패키지든지 실행(수정 없이 또는 다시 컴파일해서)할 수 있다는 점이 이러한 Linux 시스템을 사용하려고 계획하고 있는 고객에게 큰 이점으로 작용할 것입니다. 경쟁 모델인 x86 솔루션보다 20~30% 저렴한 서버 구매 비용은 많은 고객에게 최소한 데이터 센터에서 테스트할 기회를 주게 할 것입니다.

컨설팅 회사로서 당사는 새로운 POWER8 시스템이 x86 서버 시장을 뒤흔들어 놓을 것을 예상하고 있습니다. 새로운 POWER8 시스템의 성능, 애플리케이션 가용성 및 가격은 컴퓨팅 집약적 애플리케이션을 위해 더 큰 성능이 필요한 고객에게 획기적인 전환점이 될 수 있을 것입니다.

본 문서는 발행자의 사전 서면 허가 없이는 어떠한 형태로도 복제 또는 전송할 수 없습니다. 언급된 제품 및 회사의 모든 상표와 등록 상표는 해당 소유자의 재산입니다. 이 책에 포함된 정보는 신뢰할 만한 출처로부터 제공된 것입니다. Gabriel Consulting Group은 본 보고서의 완전성, 정확성 또는 적합성을 보장하지 않으며 이 보고서에 포함된 정보의 오류, 누락, 부적합성 또는 번역본에 대해 책임을 지지 않습니다. 본 문서의 의견은 발행 당시 Gabriel Consulting Group의 판단을 반영하는 것이며 통지 없이 변경될 수 있습니다

