

[CABOT PARTNERS: 비즈니스 가치 최적화]

IBM Power Systems 의 빅데이터 분석 기능을 통한 고가치 통찰력의 획득



후원:



Srini Chari, 경영학 박사
chari@cabotpartners.com

2014 년 10 월

요약

오늘날 여러 업계에서 스마트폰과 소셜 미디어의 사용이 늘어나고 있습니다. 따라서 비즈니스를 수행하는 방식과 장소가 변화하고 있을 뿐만 아니라, 빅데이터 분석(BDA)에 의해 비즈니스 의사 결정 프로세스의 속도와 범위도 변화하고 있습니다. 이로 인해 Hadoop 인프라 사용도 갈수록 증가하고 있습니다.

기업은 백오피스(레코드 시스템)에 많은 내부 데이터를 보유하고 있으며 이해 당사자의 상호 작용(인게이지먼트 시스템)에서 그보다 많은 데이터를 수집하고 있습니다. 이 모든 방대한 데이터는 정적 데이터일 수도 있고 동적 데이터일 수도 있으며 그 형태도 다양합니다. 예를 들면 고객 관계 관리(CRM), 재고 관리, 대금 청구와 같은 기존 엔터프라이즈 IT 시스템의 정형화된 데이터와, 오디오, 비디오, 소셜 미디어, 이메일, 채팅 등의 비정형화된 데이터가 있습니다. 기업에서 빅데이터 분석을 사용하면 전략적 비즈니스 관련 질문을 구상하고 레코드 시스템과 인게이지먼트 시스템을 결합하여 새로운 고가치 통찰력 시스템을 생성할 수 있습니다.

여러 업계에서 고객 경험 및 충성도 개선, 새로운 수익 기회 파악, 부정 및 위반 행위 감지, 환자 치료 결과 개선, 재무 위험 완화를 위해 빅데이터 분석에 투자하고 있습니다. 마찬가지로 정부도 빅데이터 분석을 통해 긴급 상황에 보다 신속하게 대응하고, 테러 위협을 더 효과적으로 분석하고, 날씨를 더욱 정확하게 예측하는 등 국가 보안, 공공 안전 및 환경에 중요한 다양한 이점을 얻을 수 있습니다. 이처럼 빅데이터 분석의 경제적 가치는 엄청납니다.

그러나 기업에서 이러한 가치를 얻으려면 고가치 통찰력을 지속적으로 생성할 수 있도록 전략, 환경, 비즈니스 프로세스 및 정보 기술(IT)을 조정해야 합니다. 하지만 기업의 주요 인프라 투자 예산은 제한되어 있으므로 고유한 데이터 자산을 활용하여 최대한 바람직한 결과를 도출할 수 있는 비용 효율적이고 안정적이며 신속한 고성능 IT 인프라가 필요합니다. POWER8 프로세서를 기반으로 하는 솔루션은 바로 이러한 기능을 제공합니다.

규모에 관계없이 모든 조직은 Hadoop 워크로드에 최적화된 IBM 빅데이터 분석 솔루션 포트폴리오를 적극적으로 고려해야 합니다. 이러한 솔루션은 비용 효율적이고 확장이 가능한 획기적인 빅데이터 분석용 개방형 플랫폼을 제공합니다.

IBM Power Systems for Big Data Analytics에 투자하는 고객은 x86 대안 시스템에 비해 수는 더 적고 안정성은 뛰어난 서버를 통해 총소유비용(TCO)을 낮출 수 있습니다(고객 사례 참조).



이보다 더 큰 이점은 이러한 고객이 IBM 파트너의 확장 가능한 개방형 에코시스템(오픈파워 파운데이션) 및 CAPI(Coherence Attach Processor Interface)와 같은 혁신적 기술을 통해 제공되는 높은 가치를 누릴 수 있다는 점입니다.

통찰력 시스템: 빅데이터 분석이 제공하는 커다란 기회

데이터의 속도, 다양성, 규모가 급증하는 환경에서 '빅데이터'가 무엇인지 명확하게 정의할 필요가 있습니다. McKinsey¹의 정의에 따르면 '빅데이터'는 일반적인 데이터베이스 소프트웨어 도구로 캡처, 저장, 관리 및 분석할 수 있는 규모를 능가하는 데이터 세트입니다. 기술이 발전하면 빅데이터라고 정의할 수 있는 데이터 세트의 크기도 증가하므로 이 정의는 본질적으로 주관적입니다. 오늘날 많은 분야에서 빅데이터는 규모가 수십 테라바이트(TB)에서 여러 페타바이트(PB)에 달하며, 규모가 클 뿐 아니라 형태도 다양합니다.

비정형화된 데이터를 사용하여 이해 당사자와 상호작용하는 기업이 증가하고 있습니다.

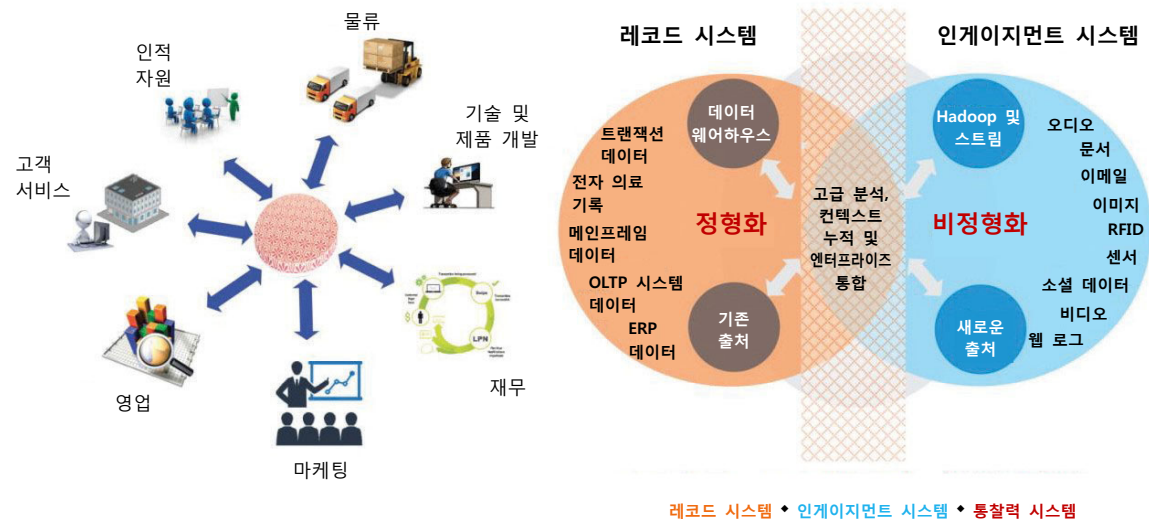


데이터는 정형화(structured), 반정형화(semi-structured), 비정형화(unstructured), 또는 다중 정형화(multi-structured) 데이터로 구분됩니다.² 정형화된 데이터는 관계형 데이터베이스 및 스프레드시트와 같이 형식과 서식이 명확하고 모델링 및 체계화되어 사용 및 관리가 쉬운 데이터로, 일반적으로 *레코드 시스템*에 존재합니다. 비정형화된 데이터는 세상의 거의 모든 데이터에 해당하지만, 정형화된 데이터를 위한 기존 데이터베이스에는 맞지 않습니다. 비정형화된 데이터는 언어 기반 데이터(예: 이메일, Twitter 메시지, 서적)와 비언어 기반 데이터(예: 이미지, 슬라이드, 오디오, 비디오)로 구성됩니다. 반정형화된 데이터는 정형화된 데이터의 특정한 의미 구조를 따르지만 정형화된 데이터에 일반적으로 적용되는 관리 및 자동화 유형의 기준을 충족할 만큼 표준화되지는 않은 데이터입니다. 예를 들면 일부 필드가 누락되었거나, 데이터베이스 시스템에서 쉽게 표현할 수 없는 정보가 포함되어 있을 수 있습니다. 반정형화된 데이터의 예로는 도로의 교통 표지, 웹 페이지 등이 있습니다. 다중 정형화된 데이터는 둘 이상의 데이터 구조 유형이 결합된 형태입니다. 일정한 형태의 비정형화된 데이터를 사용하는 *인공지능 시스템*을 통해 이해 당사자와 상호작용하는 조직이 계속 늘어나고 있습니다.

데이터의 85% 가량이 비정형화된 데이터로 추정되며 이러한 데이터는 오디오, 문서, 이메일, 이미지, RFID, 소셜 미디어, 비디오, 웹 로그 등의 출처에서 생성됩니다. 레코드 시스템 및 인공지능 시스템에서 통찰력 시스템을 구축하는 것은 많은 조직이 당면한 과제입니다. 하지만 이러한 과제는 우수한 고객 경험을 제공하고, 마케팅 효율성 및 운영 효율성을 높이고, 재무 위험을 줄이고, 제품 품질과 안정성을 높이기 위한 획기적인 비즈니스 기회를 의미하기도 합니다.



그림 1: 기업에서 정형화된 데이터 및 비정형화된 데이터를 통합하여 얻는 고가치 통찰력



다양한 업종에서 반복적인 비판적 사고 프로세스를 통한 빅데이터 분석의 고가치 사용 사례가 많이 나타나고 있습니다. 이를 통해 조직은 전략적 비즈니스 질문을 작성하고 레코드 시스템과 인게이지먼트 시스템을 결합하여 새로운 고가치 통찰력 시스템을 생성할 수 있습니다.

McKinsey는 미국 의료 시스템에서 빅데이터 분석을 창조적, 효율적으로 사용하여 품질과 효율성을 개선할 경우 의료 분야의 데이터를 통해 얻을 수 있는 잠재적 가치가 매년 3천억 달러에 달하고, 이 중 3분의 2가 국가 의료 경비를 약 8% 절감하는 효과로 나타난다고 예측합니다. 민간 부문의 경우 빅데이터를 완벽하게 활용하는 소매업체는 운영 수익을 60% 이상 높일 수 있는 가능성이 있습니다. 유럽 선진국의 정부 관리 부문에서는 빅데이터 분석을 사용하는 경우 운영 효율성 개선으로만 1천억 유로(1,490억 달러)를 절감할 수 있습니다. 이러한 추정액에는 부정 행위, 오류, 세금 격차 등을 더욱 줄일 수 있는 빅데이터의 효과는 포함되어 있지 않습니다.

그러나 이와 같이 기업에서 빅데이터 분석을 통해 고가치 통찰력을 확보하려면 전략, 환경 및 기술을 보다 효율적으로 조정해야 합니다.³ 실제로 최근 IBM이 실시한 CEO 연구에 따르면 CEO는 시장 요인보다 기술 요인을 가장 중요한 전략적 문제로 간주합니다.⁴



엔터프라이즈급 정보 기술 인프라의 중요성

개별적인 데이터 마트를 이용하는 기존의 오프라인 분석 또는 비즈니스 인텔리전스 방식에는 한계가 있습니다. 오늘날의 기업은 차세대 네트워크 구현, 스마트폰의 급증, 모바일/소셜 미디어의 대두, 다른 비정형화된 데이터의 폭증으로 인해 빠르게 증가하는 다양한 데이터를 대량으로 처리해야 하는데 기존의 방식으로는 이를 원활하게 처리할 수 없습니다. 또한 느린 속도의 데이터 로드 및 쿼리, 장시간의 네트워크 대기, 낮은 시스템 안정성 및 이용률, 분산된 인프라의 관리와 관련한 비용 및 복잡성과 같은 기술적 어려움이 있습니다.

IBM 솔루션은
엔터프라이즈급
빅데이터 분석을
위해 POWER8을
사용합니다.

이러한 문제를 해결하고 비용을 통제하는 동시에 수익을 창출하는 새로운 제품/서비스를 제공하기 위해, 많은 기업들이 실시간 및 대규모 분석 기능을 모두 지원하는 확장 가능한 고성능의 애자일(Agile) 정보 기반을 구현하기 시작했습니다. 여기에는 (특히 분산 컴퓨팅 환경에서) 갈수록 증가하는 데이터의 처리 시간을 줄이기 위해 Hadoop 및 다양한 NoSQL 제품과 같이 새롭게 대두하는 오픈 소스 기술의 사용이 포함됩니다. 이와 함께 뛰어난 안정성-가용성-서비스 가능성(RAS) 및 보안/관리 프로세스가 필요한데, 이는 IBM **POWER8** 프로세서를 기반으로 하는 IBM의 빅데이터 분석 솔루션 포트폴리오와 같은 엔터프라이즈급 IT 솔루션에서 일반적으로 제공합니다. 이 제품군의 핵심 솔루션으로는 빅데이터 워크로드에 최적화된 **IBM Solution for Hadoop - Power Systems Edition** 및 다양한 빅데이터 분석 워크로드를 효율적으로 지원/통합하는 **IBM Data Engine for Analytics - Power Systems Edition**이 있습니다.

빅데이터 분석에 POWER8이 필요한 이유

데이터 중심으로 설계된 IBM Power Systems는 빅데이터 워크로드를 위해 긴밀하게 통합되고 성능 향상을 위해 최적화된 인프라를 제공합니다. 또한 IBM은 오픈파워 이니셔티브를 통해 파트너 에코시스템을 구축하고 있습니다. Hadoop 및 기타 NoSQL 오픈 소스 빅데이터 분석 인프라를 지원하려면 이러한 에코시스템이 필요합니다.

대량의 스레드 및 대역폭: POWER8은 대규모 멀티스레드 프로세서로, 각 코어가 동시에 하드웨어 스레드를 8개씩 처리할 수 있으므로 12개 코어 칩에서 총 96개 스레드를 동시에 실행할 수 있습니다. 이 프로세서는 많은 양의 온칩/오프칩 eDRAM 캐시를 사용하며, 온칩 메모리 컨트롤러를 통해 매우 높은 대역폭의 메모리 및 시스템 I/O를 지원할 수 있습니다.



그림 2: POWER8 프로세서 세부사항



고성능 프로세서: POWER8은 약 4.5GHz의 클럭 속도를 제공하며 열 설계 전력(TDP)은 약 250와트입니다. 이 속도의 POWER8 CPU는 단일 스레드에 적용한 POWER7에 비해 속도가 약 60% 더 빠르며 멀티스레드 작업의 경우에는 2배 이상 더 빠릅니다. IBM에 따르면, 경우에 따라 POWER8은 RAM과 코어 수가 같은 기존의 행 기반 데이터베이스를 이용하는 상용 x86 시스템에 비해 빅데이터 워크로드를 82배나 더 빠르게 분석할 수 있습니다. 실제로 IBM 고객들은 Power Systems에서 BLU 가속화 기능을 사용한 결과 분석 쿼리 실행 속도가 50배에서 1,000배 이상 빨라졌다고 말했습니다.⁵

뛰어난 RAS 및 가상화: 광범위한 기업을 대상으로 진행된 최근 연구^{6, 7} 결과에 따르면 IBM Power Systems는 RAS, 성능, TCP, 보안 및 전반적인 만족도 측면에서 x86 시스템에 비해 우수한 것으로 나타났습니다. 이러한 결과를 통해 POWER8 아키텍처의 기본적인 이점을 보다 명확하게 확인할 수 있습니다. 여기에는 높은 수준의 동시 오류 발견, 결함 격리, 복구, 가용성을 지원하는 기능이 포함됩니다.

PowerVM을 통한 고급 가상화 기능을 사용하면 프로덕션 데이터베이스, 애플리케이션, 고가용성 서버, 백업/복구 서버 및 기타 서버를 포함한 여러 워크로드/작업을 더 적은 수의 단일 서버 세트에 통합할 수 있습니다. 따라서 다양한 빅데이터 분석 애플리케이션에서 확인할 수 있는 병렬 작업을 대부분의 서버 x86 솔루션에 필요한 것보다 적은 수의 서버에서 훨씬 쉽게 최적화할 수 있습니다.



최고의 성능을 제공하는 CAPI(Coherence Attach Processor Interface): POWER8은 뛰어난 SPECint 및 SPECfp 성능을 제공할 뿐 아니라, CPU에 대한 직접 링크인 CAPI가 도입되어 주변 장치와 보조 프로세서가 CPU와 직접 통신할 수 있기 때문에 운영 체제 및 드라이버 오버헤드를 대폭 줄일 수 있습니다. Intel의 버스 프로토콜 및 인터페이스는 독점적인 기능이지만, IBM은 써드파티 공급업체도 CAPI를 자유롭게 사용할 수 있도록 개발했으며 설계 지원 킷도 제공합니다. CAPI를 통해 연결되는 플래시 메모리의 경우에는 오버헤드가 24:1 비율로 감소합니다. 더욱 중요한 점은 CAPI를 사용하면 GPU, FPGA 등의 보조 프로세서를 POWER8 CPU에 직접 연결하여 워크플로우 관련 성능을 크게 높일 수 있다는 것입니다. 이와 같이 CAPI에 보조 프로세서를 연결하면 POWER8 시스템의 속도가 동급 x86 시스템에 비해 1,000배나 빨라질 수 있습니다.

CAPI는 액셀러레이터를 사용하는 혁신적인 이종 컴퓨팅 방식입니다. 앞으로 등장할 수평 확장 데이터 센터는 워크로드와 애플리케이션에 따라 달라질 가능성이 높으므로 이러한 방식이 중요합니다. 이러한 방식에서 서버는 특정 작업용으로 최적화된 특수 하드웨어를 사용하여 각기 다른 워크로드에 전용으로 사용됩니다. 즉, 서버 환경이 더 이상 동종 환경이 아닙니다. IBM은 CAPI 및 오픈파워를 통해(아래 그림 참조) NVIDIA, Altera, VeriSilicon 및 기타 파트너의 특수 액셀러레이터를 사용하여 워크로드를 현재 가능한 수준보다 대폭 가속화할 수 있는 환경을 구현했습니다. 이전 아키텍처보다 쉽게 프로그래밍할 수 있는 아키텍처에서 이와 같이 워크로드 속도를 높일 수 있습니다.

그림 3: CAPI의 이점

일반적인 I/O 모델 흐름



코히어런트(Coherent) 모델의 흐름



기존 I/O 연결과 비교한 CAPI의 이점은 다음과 같습니다.

가상 주소 지정 및 데이터 캐싱

- 대규모 공유 메모리. 데이터 집중적 애플리케이션에 중요
- 자주 참조하는 데이터의 응답시간 단축

보다 쉽고 자연스러운 프로그래밍 모델

- 기존의 스레드 수준 프로그래밍은 더 정교한 분산 메모리 방식을 사용할 수 없음
- I/O 응답시간이 길어 일반적으로 애플리케이션을 광범위하게 재구조화해야 함

기존의 I/O로는 불가능한 애플리케이션을 지원

- 예: 포인터 추적

혁신을 위한 개방형 에코시스템을 조성하는 오픈파워 파운데이션: IBM은 POWER8 출시와 함께 오픈파워 파운데이션(<http://openpowerfoundation.org/>)에 대한 참여를 지속한다고 발표했습니다. IBM은 3년의 개발 기간 동안 POWER8 기술에 이미 24억 달러를 투자했으며, 오픈파워 파운데이션에서 이 기술을 이용하여 POWER8 플랫폼을 더욱 혁신할 수 있기를 기대합니다.

오픈파워 파운데이션은 IBM에서 출시한 Power 아키텍처 기반 제품을 중심으로 하는 협업 커뮤니티로서 2013년 8월에 출범했습니다. IBM은 오픈파워 파운데이션을 통해 프로세서 사양, 펌웨어, 소프트웨어 등 Power 아키텍처 제품과 관련된 기술을 공개합니다. IBM은 이를 자유 라이선스 방식으로 제공하며, 파트너와 협업 개발 모델을 이용할 계획입니다. 목표는 서버 공급업체가 향후 데이터 센터 및 클라우드 컴퓨팅을 위해 사용자 정의된 서버, 네트워킹 및 스토리지 하드웨어를 직접 구축할 수 있는 서버 공급업체 에코시스템을 조성하는 것입니다.

오픈파워 파운데이션 - 전 세계에서 55개 이상의 기술 공급업체가 참여하는(참여 업체 수는 증가 중) 오픈파워 파운데이션은 IBM, Google, NVIDIA, Mellanox 및 Tyan에 의해 설립되었습니다. 오픈파워 파운데이션은 회원사의 기여 계획이 상세히 기술된 혁신 로드맵을 발표했으며, 오픈파워 기술을 활용하는 최초의 서버로 IBM Power Systems를 지정했습니다. 오픈파워 파운데이션의 기본 개념은 마이크로프로세서만으로는 더 이상 비용과 성능을 향상시킬 수 없다는 것입니다. 비용과 성능을 개선하려면 시스템 스택 혁신이 필요합니다.

이미 오픈파워 파운데이션을 통해 혁신과 성능 향상을 실현한 사례가 많습니다. 여기에는 Monte Carlo 재무 위험 모델링, 빅데이터 및 Java 가속화, NoSQL 가속화, 키 값 저장소(KVS) 가속화 등 다양한 비즈니스 범위가 포함됩니다.

개방형 표준을 이용하는 IBM Power Systems는 상용 아키텍처에 적용되는 제약을 제거함으로써, 개발자에게 플랫폼에 맞게 조정되어 생산성과 성능을 높이는 도구를 제공합니다. Power



Systems는 Linux, OpenStack, KVM 등의 개방형 표준을 지원하는 아키텍처를 사용하고, Power Systems Linux 센터, Watson & Power 개발 클라우드 및 오픈파워 파운데이션에서 지원하는 커뮤니티 혁신을 이용하여 애플리케이션 및 인프라의 새로운 기준을 제시하는 한편 컴퓨팅과 데이터 집중 작업의 속도를 크게 높여 줍니다.

IBM Solution for Hadoop – Power Systems Edition

IBM Solution for Hadoop - Power Systems Edition은 빅데이터 워크로드를 위해 전문적으로 설계되고, 긴밀하게 통합되고, 성능이 최적화된 참조 아키텍처를 제공합니다. 기존의 x86 인프라와 달리 이 IBM Solution for Hadoop은 CPU와 스토리지의 대략적인 조합을 애플리케이션 요구사항에 맞추는 간편한 빌딩 블록 방식을 사용함으로써 빅데이터 워크로드를 충족하도록 구체적으로 조정할 수 있습니다. 이러한 방식은 x86 솔루션과 비슷한 가격으로 빅데이터 워크로드를 위해 인프라를 최적화할 수 있습니다. 이 아키텍처는 다음과 같은 정의를 제시합니다.⁸

완벽한 클러스터. 손쉽게 구매, 구현 및 작동할 수 있도록 설계되었으며 포괄적이고 긴밀하게 통합된 클러스터입니다. 이 참조 아키텍처는 서버, 네트워크, 스토리지, 운영 체제, 관리 소프트웨어, Hadoop 호환 소프트웨어, 런타임 라이브러리 등 빅데이터 애플리케이션에 필요한 모든 구성요소를 표현합니다. 고객은 스토리지를 상황에 따라 유연하게 기존 Hadoop 중심 분산 토폴로지에 구현할 수도 있고, 블록 수준 복제를 수행할 필요 없이 스토리지를 보다 경제적으로 사용할 수 있는 공유 디스크 토폴로지에 구현할 수도 있습니다.

Hadoop에 최적화되고 클러스터에 맞게 사용자 정의된 아키텍처가 완벽하게 지원되어 최고의 성능 및/또는 비용/성능 제공

애플리케이션 최적화 구성. 클러스터 구성이 애플리케이션 성능을 최적화하고 총소유비용을 낮추도록 설계되었습니다. 이러한 클러스터가 IBM Platform Cluster Manager 및 IBM InfoSphere BigInsights Enterprise Edition과 통합되며, 여기에는 IBM Platform Symphony MapReduce 및 IBM General Parallel File System – File Placement Optimizer(GPFS - FPO)가 임베디드 구성요소로 포함됩니다. 이와 같이 최적화된 구성을 통해 사용자는 생산성을 보다 빠르게 높일 수 있습니다.



성과와 견고성을 위한 고급 기술. 최상의 성능 또는 가격/성능 비율을 구현하도록 클러스터의 기본 하드웨어 및 소프트웨어 구성요소를 사용자 정의할 수 있습니다.

상업용 솔루션 수준 지원. 구성이 모든 하드웨어 및 소프트웨어 구성요소에 대해 검증되고 지원됩니다.

이 아키텍처는 일반 빅데이터 애플리케이션에 대하여 최적화되어 있고 대부분의 소프트웨어 구현이 완전히 자동화되어 있으므로 클러스터를 1~2일 안에 구현할 수 있습니다. 사용자는 최소한의 교육만 받으면 이러한 클러스터가 제공하는 성과와 기능을 활용할 수 있습니다.

Hadoop

워크로드가 많은
경우 컴퓨팅
기능과 스토리지를
동적으로 조정하는
기능이 유용합니다

IBM Data Engine for Analytics - Power Systems Edition

일반적인 Hadoop 솔루션은 대량의 스토리지를 제공하는 서버 기반의 수평 확장 솔루션을 사용합니다. 이러한 방식의 문제점은 스토리지에 다수의 컴퓨팅 서버가 존재하고 별도의 스토리지 비용 계획 기능이 없으므로 사용자가 스토리지에 대한 제어를 상실하는 경향이 있다는 점입니다. 게다가 일반적으로 백업, 아카이브, 재해 복구 기능도 없으며 스토리지 보안 제어 및 감사 기능도 엄격하지 않습니다. 따라서 디스크 오류 시 많은 재작성 작업을 수행해야 하고, 네트워크 리소스를 많이 사용하게 되며, 애플리케이션 성능이 저하될 수 있습니다.

반면 워크로드에서 여러 작업이 같은 데이터에 반복해서 동시에 액세스(대개 읽기 작업)할 필요가 없는 경우에 여러 개의 스토리지 복제본을 사용하면 불필요한 비용이 많이 소요됩니다. 또한 Hadoop 솔루션은 대개 Hadoop 이외의 워크로드와 리소스를 공유할 수 없고 기존 인프라를 재사용할 수 없으므로 문제가 더 복잡해집니다. 이러한 모든 요인으로 인해 구성을 최적화할 수 없게 됩니다.

IBM Data Engine for Analytics는 고성능 상호 연결 기능을 사용하여 IBM General Parallel File System(GPFS)를 기반으로 하는 Elastic Storage Server에 POWER8 기반 컴퓨팅 서버를 연결함으로써 이러한 문제를 해결하도록 설계되었습니다. IBM GPFS는 공유 디스크 또는 비공유 분산 병렬 모드로 구현할 수 있는 고성능 클러스터형 파일 시스템으로, 보통 엔터프라이즈 컴퓨팅 및 슈퍼컴퓨팅 환경에서 모두 사용됩니다. GPFS는 여러 클러스터 노드에서 실행되는 애플리케이션에 대해 동시 고속 파일 액세스 기능을 제공합니다.

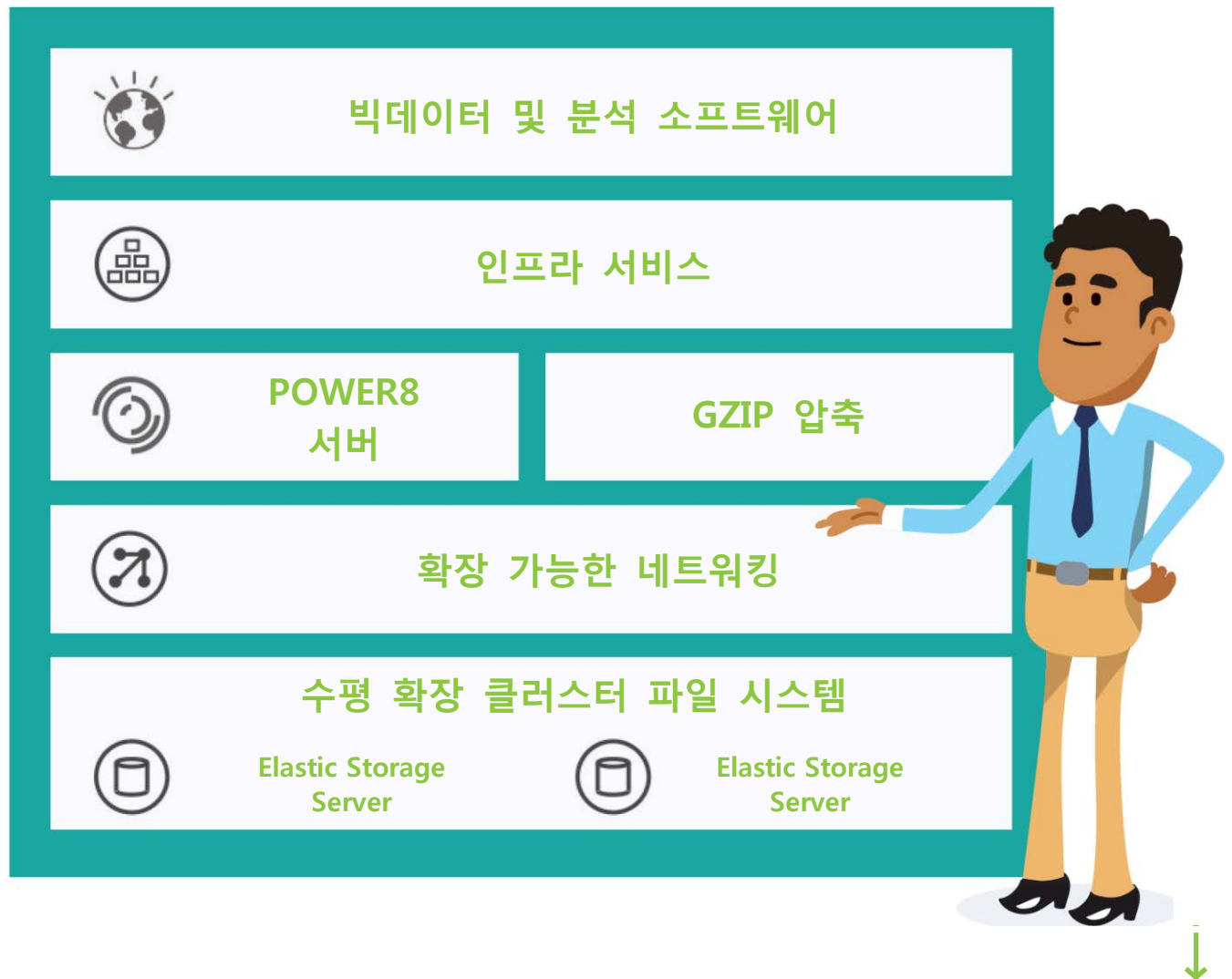


IBM Data Engine for Analytics에 사용되는 IBM GPFS 기반 Elastic Storage Server는 SAN보다 저렴한 비용으로 공유 스토리지를 제공하며, RAS 및 관리 기능을 포함합니다.

Elastic Storage Server를 사용하여 얻는 다양한 이점:

- 복제본이 필요 없으므로 복제 오버헤드 및 네트워크 부담 감소
- 디스크 오류 시 네트워크 성능 저하 및 애플리케이션 성능에 대한 영향이 더 적으므로 보다 신속하게 재작성
- 중앙 집중식 스토리지 관리, 향상된 백업, 아카이브, 재해 복구, 보안 제어 및 감사

그림 4: IBM Data Engine for Analytics – Power Systems Edition의 구성요소 아키텍처



GPFS Elastic Storage Server, 플랫폼 컴퓨팅 및 네트워크 구성요소의 혁신적 기능을 Linux on Power 수평 확장 시스템과 완벽하게 통합합니다.

- 서로 다른 워크로드에서 공통 인프라를 사용할 수 있으므로 Hadoop 및 Hadoop 이외의 워크로드를 모두 포함하는 워크플로우를 보다 쉽게 구축할 수 있으며 솔루션 수준의 구성, 정렬, 이행, 구현, 지원이 가능

IBM에 따르면 IBM Data Engine for Analytics는 일반적인 x86 Hadoop 솔루션 구성에 비해 다양한 이점을 제공합니다.

POWER8의 이점

- 코어당 하드웨어 스레드 수가 4배
- 하이퍼스레딩(HT)보다 우수한 동시 멀티스레딩(SMT) 성능
- 메모리 용량 30% 증가
- 메모리 대역폭 170% 증가 및 내장형 하드웨어 암호화/암호 해독 엔진
- x86 코어에 비해 2배의 성능

Platform Symphony의 차별화된 기능

- 보다 빠른 MapReduce 엔진
- 리소스를 더욱 효율적으로 활용하여 인프라 중복 최소화 및 비용 감소
- Hadoop 이외의 워크로드와 인프라 공유 가능
- 여러 테넌트의 인프라 공유
- 테넌트의 SLA 보장
- 저널링을 통해 세션과 작업을 복구할 수 있으므로, 오류 발생 시 장시간 실행되는 작업을 지연 없이 재개할 수 있어 시간과 비용 절감
- 테넌트 간의 보안 격리
- 사용 가능한 인프라를 사용하도록 테넌트를 확장 가능
- 사용자 공간 소켓을 통해 응답시간 단축 및 처리 속도 증가



NCSU는 확장할 수 없는 x86 시스템과 달리 POWER8에서 더 많은 동시 쿼리를 더 빠르게 실행할 수 있습니다

GPFS의 가치

- POSIX 파일 시스템을 통해 Hadoop 및 기타 워크로드를 모두 포함하는 워크플로우를 보다 쉽게 구축할 수 있으며 데이터를 더욱 쉽고 빠르게 공유 및 수집
- 파일 시스템 성능이 향상되어 RDMA 가능 네트워크를 기본적으로 활용함으로써 워크로드 속도 개선
- GPFS ACL(액세스 제어)을 통해 공유 인프라 환경 내의 여러 테넌트 간에 데이터 보안 제어 효율성 개선
- GPFS 암호화 및 안전한 삭제 기능을 통해 보안 수준 향상
- 분산 메타데이터 서버를 통해 단일 실패점을 방지하고 단일 이름 노드보다 우수한 성능 제공
- Active File Management(AFM), Information Lifecycle Management(ILM), 다중 클러스터 등의 기타 GPFS 기능과 결합 가능
- Elastic Storage Server는 보다 속도가 빠르고 복원력이 뛰어난 스토리지를 제공하므로 애플리케이션 성능에 대한 영향을 최소화하면서 오류가 발생한 디스크에서 더욱 빠르게 복구 가능

POWER8 빅데이터 분석 사례

Power Systems for Big Data Analytics가 제공하는 고가치 이점을 보여 주는 몇 가지 사례를 소개합니다.

노스 캐롤라이나 주립대: 노스 캐롤라이나 주립대(NCSU)에 속한 Poole College of Management의 Center for Innovation Management Studies(CIMS)에 따르면, 신규 창업 기업에서 Fortune 500대 기업에 이르기까지 다양한 규모의 회사⁹가 비정형화된 데이터에서 전략적 조회를 생성하는 데 어려움을 겪고 있습니다. 연구자들이 다양한 사례를 살펴본 결과, 데이터에 기반하여 결정을 내리려면 소프트웨어를 지원하고 데이터를 수집 및 저장할 수 있도록 IT 부서 전체에서 긴밀한 팀 작업을 수행해야 하며 빅데이터 프로세스를 실행할 수 있도록 통계 분석가들이 협업해야 한다는 사실을 확인했습니다.



또한 NCSU 연구진은 이 프로세스를 위해서는 소프트웨어 및 서버 플랫폼의 적절한 조합이 중요하다는 사실을 확인했습니다. 이 연구에서는 x86 시스템과 Power 기반 시스템을 모두 사용하여 프로젝트를 시작했지만, x86 기반 시스템은 불안정하며 빅데이터 환경에 맞게 확장할 수 없으므로 곧 사용을 중지했습니다. 연구진이 확인한 바로 빅데이터에 가장 적합한 시스템은 POWER8 프로세서 기반 기술을 사용하는 Power Systems 서버였는데, 이 서버에서는 여러 코어에서(코어당 스레드 수가 더 많음) 더 많은 병렬 쿼리를 보다 빠르게 병렬로 실행할 수 있기 때문입니다. 또한 이 서버는 메모리 대역폭도 더 높고 데이터를 수집, 이동 및 액세스하는 데이터 I/O 속도도 더 빠릅니다.

Bloor Research: 영국의 유명 IT 연구/컨설팅 기업인 Bloor Research는 최근 Power Systems에서 실행되는 BLU 가속화 기능을 포함하는 IBM DB2의 상대적 이점을 x86 기반 플랫폼에서 실행되는 SAP HANA, Oracle Exadata 및 Microsoft SQL Server와 비교 측정했습니다.¹⁰ 이 프로젝트는 주요 비즈니스 분석 플랫폼의 성능 기능을 비교하기 위한 연구의 일환으로 진행되었습니다.

Bloor Research는 Power Systems가 DB2 with BLU Acceleration에 상당한 이점을 제공한다는 사실을 확인했습니다. POWER8에 소켓당 최대 12개의 코어와 코어당 8개의 스레드가 포함되어 있기 때문입니다. 따라서 POWER8의 경우 소켓당 병렬 처리 수가 96개로 크게 증가합니다.

또한 Bloor Research는 Power Systems에는 마이크로코드를 통해 가상화 기능이 내장되어 있으므로 Power 서버를 가상화할 경우 보통 성능에 큰 영향이 없다는 점도 지적했습니다(워크로드가 균형을 이루고 있을 때). 반면 x86 시스템은 가상화 기능이 내장되어 있지 않으므로 VMware 같은 소프트웨어를 설치해야 하는데, 이러한 소프트웨어는 시스템 리소스를 20%까지 사용할 수 있으므로 성능이 저하됩니다.

뿐만 아니라 Power Systems는 더 큰 메모리 크기를 지원합니다. IBM은 대규모 시스템에 필요 시 용량을 추가하는 CoD(Capacity on Demand) 기능을 제공하며, Power Systems는 복원성 및 보안 효율성도 보다 뛰어납니다. x86 프로세서를 사용하는 솔루션은 중복 기능만 사용하므로 서버 수와 획득 및 유지보수 비용이 증가하고 관리 작업의 수준이 더 복잡해집니다.

Bloor는 Power Systems가 x86 기반 플랫폼보다 가격이 비싼 것처럼 보일 수도 있지만, Power Systems와 같은 성능 및 기능을 얻으려는 고객은 더 많은 x86 서버를 구입해야 하므로 x86 솔루션의 비용이 더 높다는 결론을 제시했습니다.



결론

고가치 통찰력을 확보하려는 조직은 POWER8 및 CAPI가 포함된 IBM Power Systems를 사용하는 빅데이터 분석 솔루션의 도입을 적극적으로 고려해야 합니다.

- 모든 기업은 다중 테넌트 클러스터 가상화를 통해 이용률을 개선함으로써 서버 수를 줄여 총소유비용(TCO) 절감 이점을 얻을 수 있습니다. 또한 통합, 중복 사본 수 감소, 효율적인 압축 알고리즘 및 데이터 인식 스케줄링 기능을 통해 스토리지 수와 데이터 병목 현상도 줄일 수 있습니다.
- 이미 많은 기업들이 뛰어난 RAS 및 성능을 제공하는 Power Systems에서 미션 크리티컬 데이터 웨어하우스 및 비즈니스 인텔리전스 애플리케이션을 실행하고 있습니다. 이와 같이 정형화된 데이터와 비정형화된 데이터를 혼합하여 새로운 분석 기능을 추가하는 방식은 구현이 훨씬 쉽습니다. 또한 사용자, 프로세스, 플랫폼 및 애플리케이션 영역의 기존 IT 투자를 보호하는 동시에 Hadoop 및 NoSQL 워크로드 수요를 충족하도록 확장할 수 있는 완벽하고 비용 효율적인 방법을 제공할 수 있습니다.

무엇보다 중요한 점은, CAPI가 포함된 POWER8은 **빅데이터의 고가치 통찰력**을 얻기 위한 반복적인 작업에 필요한 비판적 사고를 지원하는 혁신적인 개방형 플랫폼을 제공한다는 사실입니다.



Copyright© 2014. Cabot Partners Group. Inc. All rights reserved. 다른 회사의 제품 이름, 상표 및 서비스표는 본 문서에서 식별용으로만 사용되며 해당 소유자가 소유합니다. 모든 이미지와 지원 데이터는 IBM 또는 범용 소스로부터 얻은 것입니다. Cabot Partners Group은 공용 정보 및 출처를 근거로 본 정보와 제품 권장사항을 제공하며 이러한 정보와 제품 권장사항에는 Cabot Partners Group과 기타 제3자의 정확하고 신뢰할 만한 견해가 포함될 수 있습니다. 단, 통제가 불가능한 시장 여건의 변동성을 감안하여 해당 정보와 권장사항은 일체의 보증 없이 제공됩니다. Cabot Partners Group, Inc.는 귀하 또는 귀하의 고객이 제공된 정보와 권장사항을 사용하여 발생한 모든 손해(부수적 손해, 결과적 손해 또는 기타 손해 포함) 및 문서상의 부주의에 의한 오류에 대해 책임을 지지 않습니다. 이 문서는 지원금으로 개발되었습니다. 본 문서는 IBM을 포함한 다양한 벤더의 공용 자료를 활용하지만 관련 사안에 있어서 반드시 해당 벤더의 입장을 반영하지는 않습니다.

Cabot Partners는 협업 컨설팅 업체이자 독립적인 IT 분석 기업으로서, 기술 기업과 해당 고객에게 고객 기반을 형성하고 확장하는 방법, 바람직한 수익과 수익성 결과를 달성하는 방법, 그리고 HPC, 클라우드 컴퓨팅, 분석을 비롯하여 새롭게 대두하는 기술을 효율적으로 사용하는 방법에 대한 자문을 전문적으로 제공하고 있습니다. 자세한 내용을 확인하려면 www.cabotpartners.com을 방문하십시오.

- 1 Big Data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute, 2011년 6월
- 2 Big Data, Bigger Opportunities, Jean Yan, President Management Council Inter-agency Rotation Program, Cohort 2, 2013년 4월
- 3 IBM Institute for Business Value 2013 Big Data & Analytics Study, <http://www.ibm.com/services/us/gbs/thoughtleadership/ninelevers/>
- 4 IBM Global CEO Study, 2012, <http://www.ibm.com/services/us/en/c-suite/ceostudy2012/>
- 5 https://www14.software.ibm.com/webapp/iwm/web/signup.do?source=sw-infomgt&S_PKG=ov20419
- 6 Edison Group, "Better Performance, Lower Costs The Advantages of IBM PowerLinux 7R2 with PowerVM versus HP DL380p G8 with vSphere 5.1, <http://public.dhe.ibm.com/common/ssi/ecm/en/pol03161usen/POL03161USEN.PDF>
- 7 Solitaire Interglobal, "Power Boost Your Big Data Analytics Strategy", <http://www.ibm.com/systems/power/solutions/assets/bigdata-analytics.html>
- 8 IBM Solution for Hadoop - Power Systems Edition, IBM Reference Architecture, 2014년 6월
- 9 The Business Case for Using Unstructured Text Analytics on IBM Power Systems for Critical Decision Making, Stephen Markham, Ph.D. and Michael Kowolenko, Ph.D., Poole College of Management, North Carolina State University
- 10 IBM DB2 with BLU Acceleration on Power Systems: how it compares, An Comparison Paper by Bloor Research, Philip Howard, 2014년 4월