

빅데이터의 활용을 시작하는 가장 좋은 방법 5가지



빅데이터: 큰 기회

빅데이터가 등장하기 전의 삶이 어땠는지 기억나십니까? 빅데이터는 이미 비즈니스 용어로 널리 쓰이게 되었으며, 우리는 때로 빅데이터가 비교적 최근에 나타난 현상이라는 것을 잊곤 합니다. 빅데이터를 하나의 유행으로 보는 시각도 있지만 사람, 프로세스 및 장치에서 생성되는 데이터의 양은 지속적으로 증가하고 있습니다. 이제 빅데이터는 삶의 일부가 된 것입니다.

그러나 이러한 압도적인 양의 데이터는 잘 활용했을 때에만 중요한 자산이 된다는 점을 명심해야 합니다. 정보화 시대에 이러한 데이터는 가장 유용한 자원 중 하나일 수 있지만, 데이터를 올바르게 활용하는 방법을 모르는 경우에는 큰 실패의 원인이 될 수도 있습니다. 그렇다면 데이터를 어떻게 잘 활용할 수 있을까요?

몇 가지 실제 시나리오를 살펴보도록 하겠습니다.

- 헬스케어 업계는 건강보험과 관련된 부정행위로 인해 연간 약 미화 2,500억 달러를 지출하고 있습니다. 2016년경에는 이러한 지출이 연간 미화 4,000억 달러 이상으로 증가할 수도 있습니다.¹ 미국의 헬스케어 업계에서는 빅데이터를 창의적이고 효과적으로 활용해 효율과 품질을 향상시켜 매년 미화 3,000억 달러 이상의 가치를 창출할 수 있었습니다.²
- 한 선도적인 글로벌 금융 서비스 업체에서 근무하던 악덕 중개인이 회사에 미화 20억 달러의 손실을 입혔습니다. 이로 인해 이 업체는 파산의 위기에 몰렸습니다. 금융 기관들은 이제 대량의 데이터를 활용하여 외부의 부정행위(고객, 예금주 또는 보험 가입자와 관련된 부정행위) 및 직원과 관련된 내부 사고를 모두 손쉽게 방지할 수 있게 되었습니다.
- 유럽의 국가 정부에서는 빅데이터의 활용을 통한 운영 효율 향상으로 1,000억 유로(미화 1,490억 달러) 이상을 절약할 수 있었습니다.³ 이 금액은 빅데이터를 활용해 부정행위와 오류를 줄이고 세금 수입의 징수율을 높여서 얻은 이익은 포함되지 않은 금액입니다.

- 소매 업계는 고객의 요구사항을 충족할 수 있는 적절한 제품의 재고를 확보하지 못해 매년 미화 930억 달러 수준의 판매 기회를 놓치고 있습니다. 한 소매업체는 빅데이터를 활용하여 영업 이익률을 60% 이상 높였습니다.⁴
- 전 세계에 60억 명의 가입자를 보유하고 있으며 매년 두 자리 수의 성장률을 기록하고 있는⁵ 통신 업계에서는 각 개인의 생활방식에 맞게 차별화되고 개인 맞춤형 서비스와 위치 기반 오픈링이 요구되고 있습니다.

높은 비용과 큰 기회를 동시에 가진 빅데이터 솔루션에 대해 시장은 준비가 되어 있습니다. IBM 기업가치연구소가 University of Oxford의 Saïd Business School과 협력하여 수행한 최근의 연구에서는 응답자들에게 현재 자신의 조직의 빅데이터 관련 활동의 수준을 기술하도록 요청했습니다. 이 연구의 결과는 빅데이터의 도입에 대한 네 개의 주요 단계를 제시하고 있으며, 이 단계는 교육, 탐색, 개시 및 실행 단계로 이루어진 연속선상에서 진행됩니다(그림 1 참조).⁶

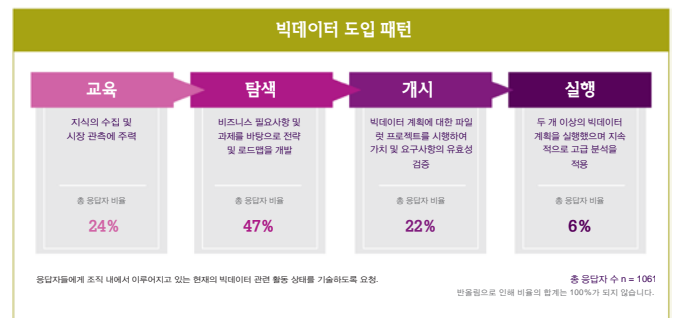


그림 1. 빅데이터 도입의 네 단계

6%의 조직만이 빅데이터와 관련된 계획을 이미 실행하고 있으며, 약 25%의 조직은 파일럿 계획을 실행 중이고, 50%의 조직은 전략을 개발하는 단계에서 빅데이터 솔루션 구매할 것을 고려하고 있으며, 25% 이상의 조직은 정보 수집 단계인 것으로 나타났습니다.

현재 자신의 조직이 빅데이터 전략과 관련된 작업을 추진하고 있지 않다고 하더라도, 경쟁 업체에서는 이미 이러한 작업을 추진하고 있을 수 있습니다. 빅데이터 전략에 있어 어려운 점은 어떻게 그리고 어디서부터 시작해야 할지를 결정하는 것입니다.

빅데이터 활용 방안

과거에 시장에서 이루어졌던 빅데이터 관련 활동은 대부분 빅데이터 기술의 학습에 초점을 두고 있었습니다. 벤더들은 빅데이터가 어떠한 문제를 처리할 수 있는지에 대한 이해를 돕기 위한 노력을 기울이지 않았지만, IBM은 달랐습니다.

IBM은 설문조사를 실시하고, 분석가가 분석한 결과를 연구하고, 300명 이상의 고객 및 유망 고객과 대화하고, 수백 개의 빅데이터 솔루션을 구현한 후, 높은 가치를 제공하는 가장 우수한 활용 방안 5가지를 다음과 같이 식별했습니다. 이러한 활용 방안은 빅데이터를 향한 첫 걸음이 될 수 있습니다.

1. **빅데이터 탐색:** 빅데이터를 검색하고, 시각화하고, 이해하여 의사결정을 향상
2. **고객에 대한 전방위적인 시야 확보:** 내부 및 외부의 추가적인 정보 소스를 통합하여 고객에 대한 시야를 확장
3. **보안/인텔리전스의 확장:** 실시간으로 위험을 낮추고, 부정행위를 발견하고, 사이버 보안을 모니터링
4. **운영 분석:** 다양한 머신 데이터를 분석하여 비즈니스 성과 및 운영 효율을 향상
5. **데이터 웨어하우스의 보강:** 빅데이터와 기존의 데이터 웨어하우스 기능을 통합하여 새로운 비즈니스 통찰력을 얻고 기존의 웨어하우스 인프라를 최적화

활용 방안 1: 빅데이터 탐색

빅데이터 활용의 첫 번째 단계는 어떠한 데이터를 보유하고 있는지 확인하고 이러한 데이터를 액세스할 수 있는 기능을 구축하여 의사결정 및 일상적인 업무를 지원하는 데 활용하는 것입니다. 이는 빅데이터 탐색을 의미합니다.

빅데이터에 대한 논의의 대부분은 데이터의 양, 속도 및 다양성이라는 세 가지 주제로부터 시작됩니다. 이러한 논의를 통해 대규모 조직은 정보 자원으로부터 가치를 얻고, 더 나은 의사결정을 하고, 운영을 향상시키고, 위험을 줄이기 위해 일상적으로 처리하는 과제의 규모를 식별할 수 있습니다. 중요한 의사결정, 고객과의 의사소통 또는 분석에는 여러 데이터 소스의 정보가 필요합니다. IBM의 빅데이터 플랫폼에 포함된 IBM® InfoSphere® Data Explorer는 엔터프라이즈 시스템 내의 정보 및 조직 외부의 데이터를 쉽게 탐색할 수 있는 기능을 제공합니다.

센서, 머신 로그, 클릭스트림, 웹사이트 등에서 얻을 수 있는 이른바 “원시(raw)” 데이터는 또 다른 과제를 제시합니다. 조직들은 어떻게 이러한 데이터에 컨텍스트를 추가하여 분석 및 의사결정을 향상시키는 데 활용할 수 있을까요? 이러한 반정형(semi-structured) 소스의 정보를 엔터프라이즈 데이터와 융합하기 위한 InfoSphere Data Explorer의 기능과 IBM의 빅데이터 플랫폼의 다른 기능을 이용하면 가치 있는 컨텍스트를 추가하여 데이터에서 더 높은 가치를 얻을 수 있습니다.

또한, IBM이 제공하는 빅데이터 탐색 기능을 이용하면 위험의 발생을 방지할 수 있습니다. 정보 인프라의 광범위한 영역을 탐색할 수 있는 역량이 부족한 조직들은 개인 식별이 가능한 정보(PII)와 같은 기밀 정보를 유출하고, 중요한 영업 비밀 및 전략적 정보를 경쟁업체에 노출하고, 소송 및 기타 기업 관리를 위해 필요한 정보를 제때에 검색 및 확인하지 못 할 수도 있습니다.

빅데이터 탐색이 귀사에 적합한 활용 방안인지 확인하는 방법

다음과 같은 질문에 대해 보십시오.

- 쓸모 없는 정보와 유용한 콘텐츠를 어떻게 구별하고 있습니까?
- 대규모의 복잡한 데이터에 대해 어떠한 방법으로 데이터 탐색을 실행하고 있습니까?
- 어떠한 방법을 이용해 새로운 유형의 데이터 또는 비정형 데이터(예: 소셜 미디어 및 이메일)로부터 통찰력을 얻고 있습니까?
- 귀사의 사용자는 정보를 이용해 불완전한 비즈니스 의사결정을 내리고 있거나, 정보 검색 역량의 부족으로 인해 업무를 효율적으로 실행하는 데 방해받고 있습니까?
- 직원들이 엔터프라이즈 콘텐츠 및 외부의 콘텐츠를 탐색할 수 있도록 하기 위해 어떠한 방법을 이용하고 있습니까? 이러한 콘텐츠를 단일 사용자 인터페이스에 표시할 수 있습니까?
- 데이터와 관련된 위험이 발생할 수 있는 영역을 미리 식별하고 데이터와 관련된 문제의 발생을 방지하기 위해 어떠한 방법을 이용하고 있습니까?
- 빅데이터 계획의 출발점은 어디입니까?

활용 방안 2: 고객에 대한 전방위적인 시야 확보

고객의 쇼핑 선호도, 거래처를 옮기는 이유, 다음에 구매할 제품, 다른 사람들에게 어떤 기업을 추천하는 이유 등을 포함해 고객을 완전히 이해하는 것은 모든 기업에게 전략적으로 중요한 문제입니다. 이를 위해 기업은 내부 및 외부의 정보 소스를 활용하여 고객 감성을 평가하고, 어떠한 유의미한 행동이 고객과의 관계를 발전시키는 데 도움을 줄 수 있을지 파악해야 합니다.

빅데이터의 실제 활용 방안에 대한 IBM 기업가치연구소의 최근 보고서⁷에서는 “고객의 필요사항을 진정으로 이해하고 향후의 행동을 예측할 수 있도록” 고객 분석에 초점을 두고 빅데이터를 활용할 것을 제안하고 있습니다. “고객”이라는 용어는 광범위한 의미로 이용될 수 있으며, 치료를 받고 있는 환자, 정부 내에서 의심을 받고 있는 인물, 또는 제조업계의 공급업체도 될 수 있습니다.

고객의 행동에 대한 전략적 통찰력을 제공하는 분석은 물론, 전방위적인 시야도 영업 최전선의 직원들에게까지 중요한 요소입니다. 앞선 사고의 조직들은 고객 대면 직원이 고객을 응대하여 신뢰 관계를 형성하고, 긍정적인 결과(예: 문제의 해결, 제품의 상향 판매 및 교차 판매)를 달성할 수 있도록 적합한 정보를 제공해야 할 필요성을 인지하고 있습니다. 이를 위해서는 대량의 정보를 신속하게 탐색하여 특정 고객에게 무엇이 필요한지 파악해야 합니다.

IBM InfoSphere Data Explorer는 IBM InfoSphere Master Data Management(MDM)와 함께 작동하여 고객 정보가 포함된 모든 애플리케이션 및 저장소(CRM, ECM, 공급망, 주문 추적 데이터베이스, 이메일 등)의 컨텍스트의 정보를 결합하므로, 사용자가 로그인하거나 여러 시스템을 검색하지 않아도 고객에 대한 완전한 시야를 확보할 수 있습니다.

이와 같은 단일 정보를 통해 고객을 응대하는 직원들은 고객이 구매한 제품, 최근에 수행된 고객 지원, 고객 회사에 대한 소식, 최근 대화 내용 등 고객에 대한 모든 정보를 확인할 수 있습니다. 화면 중앙의 활동 피드는 고객, 제품 또는 확인 중인 기타 개체에 대한 최신 정보를 표시합니다. 또한, InfoSphere BigInsights™, InfoSphere Streams, IBM Cognos® Business Intelligence 및 IBM SPSS® 제품이 제공하는 분석을 표시할 수 있으며, 이러한 분석은 애플리케이션에 의해 정의되는 분석의 컨텍스트와 함께 표시됩니다. 이를 통해 직원의 업무 부담을 줄일 수 있으므로, 직원은 더욱 개인 맞춤형 방식으로 고객과 의사소통할 수 있습니다. 이에 따라 적절한 답을 신속하게 제공할 수 있을 뿐만 아니라, 상향 판매를 성사시킬 수도 있습니다. 이러한 가시성은 고객의 충성도 및 만족도를 높이고, 궁극적으로는 더 많은 매출로 이어집니다.

그림 2에 나타난 것과 같이, 마스터 데이터 관리를 활용하면 조직 내의 다양한 시스템 전체에 걸쳐 데이터의 정확성과 신뢰성을 보장할 수 있습니다. 이러한 일관성은 InfoSphere Data Explorer가 생성하는 정보에 한 개체에 대한 일관적이고 정확한 데이터가 포함되도록 보장합니다.



그림 2. 신뢰할 수 있는 고객 데이터 정보를 제공하기 위해 InfoSphere Data Explorer Application Builder를 통해 개발되고 InfoSphere Master Data Management를 이용하는 애플리케이션에 표시되는 고객 정보.

정형 데이터 소스와 비정형 데이터 소스에 있는 관련 콘텐츠의 통합을 통해, InfoSphere Data Explorer는 신뢰할 수 있는 마스터 데이터에 대한 비즈니스 사용자 인터페이스를 제공합니다.

고객에 대한 전방위적인 시야가 귀사에 적합한 활용 방안인지 확인하는 방법

다음과 같은 질문에 대해 보십시오.

- 어떠한 방법으로 고객, 제품 또는 경쟁업체에 대한 모든 데이터를 식별하고 이를 필요로 하는 사람에게 제공하고 있습니까?
- 어떠한 방법으로 정형 데이터 및 비정형 데이터를 결합하여 분석을 실행하고 통찰력을 검색하고 있습니까?
- 고객, 고객사 및 파트너에게 제공할 정보 자산 전체에 대해 어떠한 방법으로 일관성을 높이고 있습니까?
- 어떠한 방법으로 고객에 대한 완전한 시야를 제공하여 비즈니스 라인의 사용자가 더 나은 비즈니스 결과를 얻을 수 있도록 하고 있습니까?
- 어떠한 방법으로 통찰력을 적용하고 조치를 취하고 있습니까?

활용 방안 3: 보안/인텔리전스의 확장

새롭게 등장한 정교한 보안 위협에 대처하려면 조직들은 이상 항목을 발견하고 공격에 대한 미세한 지표를 발견할 수 있는 접근법을 도입해야 합니다. 많은 조직들이 이미 빅데이터 기술을 이용해 기존의 보안 솔루션을 보강 및 강화하였으며, 인텔리전스, 보안 및 법의 집행과 관련된 통찰력을 크게 향상시키고 있습니다.

보안/인텔리전스의 확장을 위한 접근법을 이용하면 조직은 다음을 실행할 수 있습니다.

- 내부와 외부에 존재하는 대량의 데이터를 면밀히 조사하여 숨겨진 관계를 밝히고, 패턴을 발견하고, 보안 위협의 발생을 방지
- 실시간 계정 활동 및 과거 계정 활동의 상관관계를 파악해 비정상적인 사용자 행위 및 의심스러운 트랜잭션을 확인하여 부정행위를 발견
- 인터넷, 모바일 장치, 트랜잭션, 이메일 및 소셜 미디어와 같은 새로운 소스와 다양한 데이터를 검사하여 범죄 활동의 증거를 발견

보안/인텔리전스의 확장을 위한 활용 방안에는 다음과 같은 세 가지 주요 분야가 있습니다.

1. **인텔리전스 및 감시 관련 통찰력의 강화:** 동적 데이터 및 정적 데이터를 분석하여 연관성을 찾거나 패턴을 발견할 수 있습니다. 이러한 유형의 실시간 또는 거의 실시간(near real-time) 통찰력은 매우 높은 가치를 제공할 수 있으며, 이를 통해 생명을 구할 수도 있습니다.

2. **사이버 공격의 실시간 예측 및 관리:** 사이버 기반 테러, 스파이 행위, 컴퓨터에 대한 침입 및 주요 사이버 부정행위 등과 같이 첨단 기술을 이용하는 범죄의 수가 점점 증가하고 있으며, 이러한 범죄는 큰 위협이 되고 있습니다. 네트워크 트래픽을 분석하면 새로운 위협을 조기에 검색하여 실시간으로 대응할 수 있습니다.
3. **범죄의 예측 및 예방:** 사법 기관은 통신 데이터(예: 통화 상세 기록) 및 소셜 미디어 데이터를 분석하는 기능을 이용하여 정보로부터 범죄의 위협을 확인하고 증거를 수집할 수 있습니다. 사법 기관은 범죄가 발생하는 것을 기다리는 것이 아니라 범죄가 발생하는 것을 예방하고 사전에 범죄자를 파악할 수 있습니다.

조직들은 시나리오에 따라 범죄 정보 추적 시스템, 감시 및 모니터링 시스템 또는 보안 정보 및 이벤트 관리(SIEM) 등의 보안/인텔리전스 플랫폼 중 하나를 이용할 수 있습니다.

현재 이러한 플랫폼은 다양한 정형 데이터 소스의 데이터(트랜잭션, 데이터베이스, 네트워크, 방화벽 등)를 액세스하고 있습니다. 플랫폼 데이터는 플랫폼의 자체적인 데이터베이스 또는 웨어하우스에 저장되어 관리됩니다. 그러나 이러한 시스템은 새롭게 등장하고 있는 빅데이터 트렌드를 처리할 수 없으며, 이를 처리하기 위해서는 실시간 스트리밍 데이터나 비정형 데이터 유형의 분석이 필요합니다(그림 3 참조).

스트림 컴퓨팅(InfoSphere Streams), 엔터프라이즈급 Apache Hadoop 분석(InfoSphere BigInsights)과 같은 빅데이터 기술은 통신 기록, 스마트 장치, Twitter 스트림, Facebook 게시물, 이메일, POS 모니터링, 위치 기반 센서, 동영상, 오디오, 열 감지 데이터 및 장치에서 생성하는 기타 데이터 등의 비정형 및/또는 스트리밍 방식의 빅데이터 소스에서 제공하는 데이터를 네이티브 방식으로 액세스하여 기존의 보안 및 인텔리전스 분석 플랫폼을 강화합니다.

광범위한 데이터 세트로부터 통찰력을 확보하여 위협을 식별하고 위협으로부터 보호

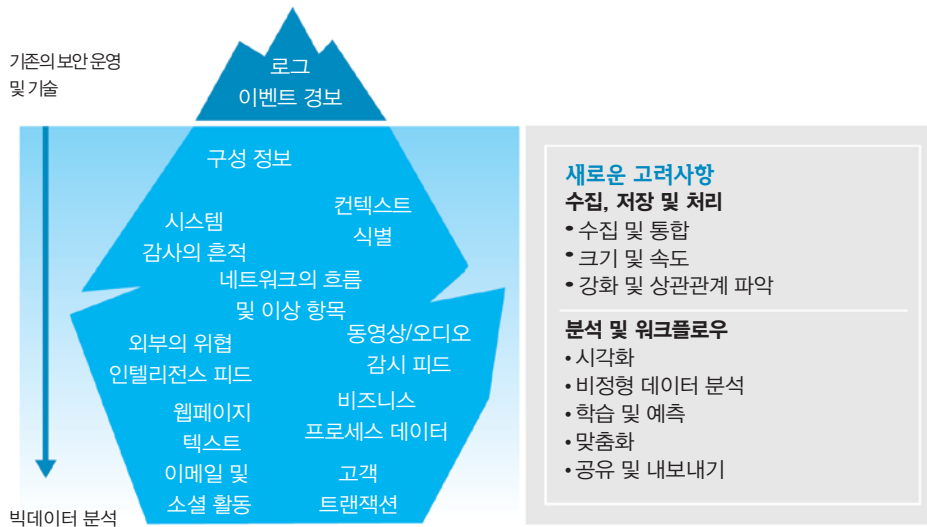


그림 3. 광범위한 데이터 세트로부터 더욱 깊이 있는 보안 통찰력을 확보.

실시간 데이터는 InfoSphere Streams를 이용해 처리 및 분석할 수 있으며, 그 결과물은 데이터 웨어하우스 또는 InfoSphere BigInsights에 저장할 수 있습니다. IBM i2® Analyst's Notebook®을 이용하는 클라이언트에서는 InfoSphere Streams와의 통합을 통해 실시간 데이터를 직접 확인 및 분석할 수 있습니다.

보안/인텔리전스의 확대가 귀사에 적합한 활용 방안인지 확인하는 방법

다음과 같은 질문에 대해 보십시오.

- 이용되지 않고 있는 데이터 소스의 실시간 데이터(동영상, 오디오, 스마트 장치, 네트워크, 통화 데이터 기록 또는 소셜 미디어)를 이용해 보안 또는 인텔리전스 시스템을 강화할 필요가 있습니까?
- 1초 미만의 짧은 시간 동안 물리적 위협 또는 온라인 위협을 발견, 식별 및 해소할 필요가 있습니까?
- 범죄자, 테러리스트 또는 감시 목록에 포함된 사람의 활동을 감시할 필요가 있습니까?
- 대량의 기술적 데이터 및 소스 또는 인지(human intelligence) 데이터 및 소스의 상관관계를 파악하여 연관성 또는 패턴을 발견할 필요가 있습니까(빅데이터 포렌식)?
- 사이버 위협을 더 효율적으로 발견 및 해소하기 위하여 비정형 데이터(이메일, 소셜 미디어)를 이용해 SIEM 솔루션을 강화할 필요가 있습니까?

활용 방안 4: 운영 분석

컴퓨터 및 네트워크 장치, 그리고 센서, 계량기 및 GPS 장치 등에서 생성되는 머신 데이터는 지속적으로 증가하고 있으며, 이는 빅데이터 솔루션의 도입의 주요 요인 중 하나입니다. 이러한 데이터는 큰 용량으로 생성되고 동적 데이터 또는 스트리밍 데이터와 같은 다양한 형식을 가지며, 다양한 유형의 데이터 세트 전반에 걸친 복잡한 분석 및 상관관계의 파악을 필요로 합니다. 또한, 데이터 유형 및 산업 분야나 응용 분야에 따른 고유한 시각화 기능도 필요로 합니다.

하지만 이러한 광범위하고 풍부한 정보의 원천을 활용하지 못하고 데이터의 일부만을 사용하여 비즈니스 의사결정을 내리는 기업들이 아직 있습니다. 운영 분석을 통해 머신 데이터를 기존의 엔터프라이즈 데이터와 결합하면 조직은 다음을 실행할 수 있습니다.

- 운영, 고객 경험, 트랜잭션 및 행동에 대한 실시간 가시성을 확보
- 사전 계획을 통한 운영 효율의 향상
- 이상 항목의 식별 및 조사
- 인프라 전체를 모니터링하여 서비스 품질 저하 또는 서비스 제공 중단을 사전에 회피

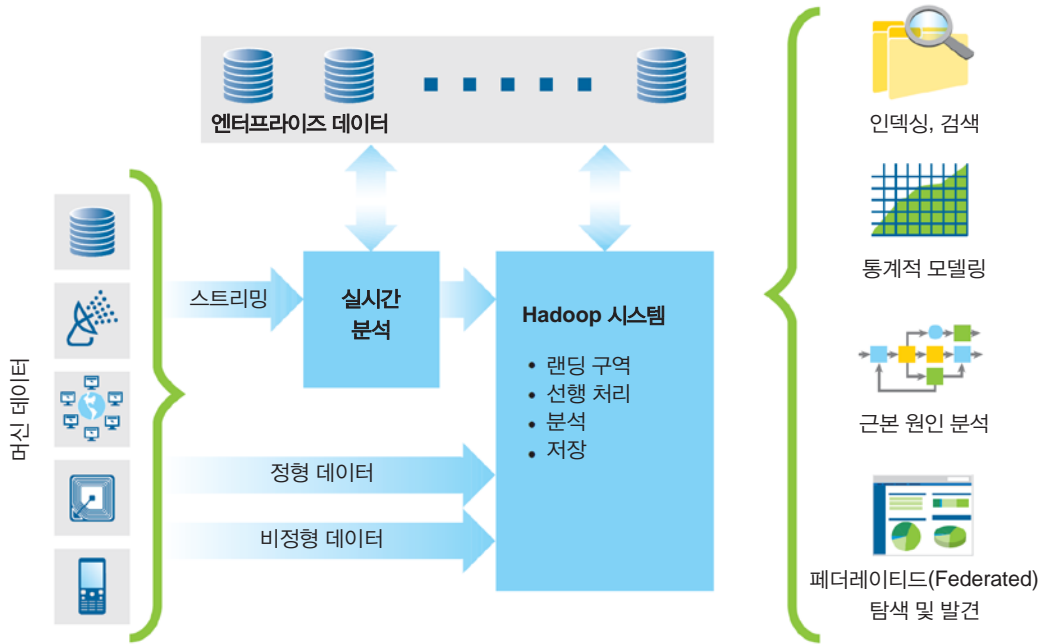


그림 4. 운영 분석은 머신 데이터와 엔터프라이즈 데이터를 결합하여 통찰력을 강화합니다.

그림 4에 나타난 것과 같이, 사용자는 서로 잘 연동되지 않는 다양한 형식으로 되어 있는 대량의 머신 데이터를 보유하고 있을 수 있습니다. 이러한 데이터는 Hadoop Distributed File System(HDFS)에 입력될 수 있으며, 스트리밍 데이터를 보유하고 있을 수도 있습니다. InfoSphere BigInsights에는 대량의 머신 데이터를 수집 및 처리하여 깊이 있는 비즈니스 통찰력을 제공하기 위한 머신 데이터 가속기가 내장되어 있습니다. 이 기능을 통해 고객 정보 또는 제품 정보 등의 다른 엔터프라이즈 데이터와 머신 데이터의 상관관계를 파악할 수 있습니다.

머신 데이터와 비즈니스 데이터를 결합하면 이러한 데이터를 운영 관련 의사결정자에게 제공하여 운영 인텔리전스 및 효율을 높일 수 있습니다. 의사결정자는 여러 시스템 전체의 데이터를 시각화하여 충분한 정보를 바탕으로 시야를 확보한 후 비즈니스 변화와 이벤트에 신속하게 대응할 수 있습니다.

운영 분석이 귀사에 적합한 활용 방안인지 확인하는 방법

다음과 같은 질문에 대해 보십시오.

- 로그, 센서, 스마트 계량기, 메시지 큐, 공공 설비 시스템, 시설 시스템, 클릭 스트림 데이터, 구성 파일, 데이터베이스 감사 로그 및 표 등을 통해 생성되는 원시 데이터와 같은 머신 데이터를 대량으로 처리하고 있습니까?
- 다양한 데이터 세트 및 이벤트 전체에 걸쳐 정보와 핵심 성과 지표의 상관관계를 실시간으로 파악하는 데 필요한 복잡한 분석을 실행할 수 있습니까?
- 모든 머신 데이터를 검색 및 액세스할 수 있습니까?
- 스트리밍 데이터를 시각화하고 이에 실시간으로 대응할 수 있는 역량을 보유하고 있습니까?
- 위와 같은 데이터를 이용해 근본 원인 분석을 실행할 수 있습니까?

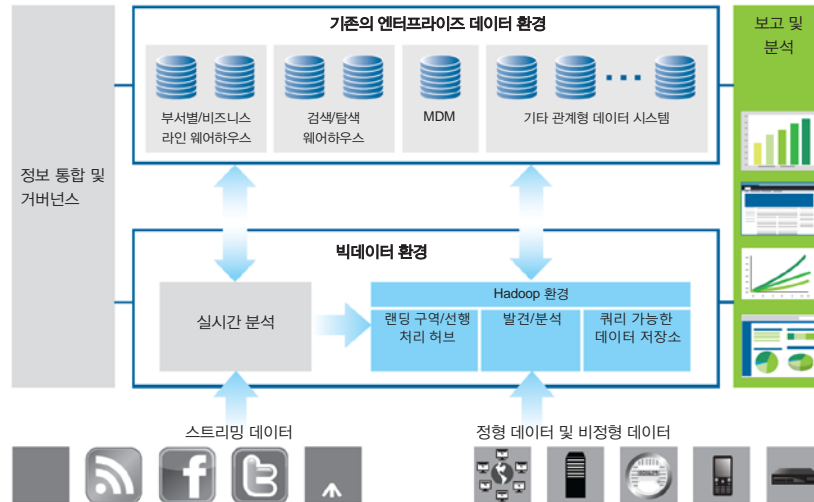


그림 5. 데이터 웨어하우스를 보강하면 데이터의 가치를 최대화할 수 있습니다.

활용 방안 5: 데이터 웨어하우스의 보강

마지막 활용 방안인 데이터 웨어하우스의 보강에서는 빅데이터 기술을 이용해 빅데이터의 가치를 보강하여 기존 데이터 웨어하우스 인프라를 강화합니다. 이 활용 방안은 현재 보유하고 있는 데이터 웨어하우스 환경을 대체하는 것이 아니라 데이터 웨어하우스의 가치를 최대화하도록 설계되었습니다.

데이터 웨어하우스의 보강은 두 가지 기초적인 필요사항으로부터 시작됩니다. 조직들은 다중 구조를 가진 데이터를 분석하기를 원하지만, 웨어하우스는 이러한 목적으로 개발된 것이 아닙니다. 따라서, 데이터 웨어하우스에만 의존하는 경우에는 가치 있는 데이터를 방치하는 결과를 가져올 수도 있습니다. 조직들은 레이턴시를 낮추기를 요구하면서 몇 주나 몇 달이 아닌 몇 시간 또는 몇 분 만에 정보를 이용할 수 있기를 원하며, 데이터에 대한 쿼리 액세스를 필요로 합니다.

두 번째 기초적인 필요사항은 웨어하우스 인프라의 최적화입니다. 오늘날의 웨어하우스 데이터의 양은 빅데이터의 수준에 육박하고 있으며, 이는 데이터 웨어하우스에 부담을 가하고 있습니다. 웨어하우스 자체는 비용이 높지 않을 수 있지만, 이러한 환경에 모든 것을 저장하고 이를 분석하려면 성능이 악화되고 비용이 상승할 것입니다.

데이터 웨어하우스의 보강에는 다음과 같은 세 가지 유형이 존재합니다(그림 5 참조).

1. **선행 처리 허브:** 어떠한 데이터를 데이터 웨어하우스로 이동시켜야 할지 결정하기 전에, 데이터에 대한 스테이징 영역 또는 “랜딩 구역(landing zone)”으로서의 엔터프라이즈급 Hadoop 기능 (InfoSphere BigInsights)이 필요할 때 이용됩니다. InfoSphere Data Explorer는 어떠한 데이터를 이동시킬지 결정하여 더욱 심층적인 분석을 실행하거나 스토리지 비용을 낮추는 초기 탐색에 이용할 수 있습니다. 이러한 작업은 필수적인 단계는 아니며, 일부 데이터를 정적 데이터로 남겨두기 위한 영역에 이용될 수 있습니다.

또한, 데이터를 먼저 저장할 필요 없이, 그리고 어떠한 데이터를 HDFS나 데이터 웨어하우스에 저장해야 할지 결정할 필요 없이, 스트리밍 데이터를 처리 및 분석하여 스트림 컴퓨팅(InfoSphere Streams)을 실시간 구성요소로 이용할 수 있습니다. 일부 경우에는 데이터를 저장할 필요가 없으며, 정보가 발생했을 때 즉각 처리하고 수정하여 웨어하우스 내의 스토리지를 줄일 수 있습니다. 이러한 랜딩 구역 접근법을 이용하면 데이터 웨어하우스에 로드하기 전에 데이터를 정리 및 변환할 수 있습니다.

2. **검색/분석:** 이 접근법은 동적 데이터에 대한 스트림 컴퓨팅 분석을 이용합니다. 이를 통해 이전에는 데이터 웨어하우스에서 실행되었던 분석을 스트림 컴퓨팅에서 실행할 수 있는 기능을 제공하여 웨어하우스를 최적화하고 새로운 유형의 분석을 실행할 수 있게 합니다. 다양한 데이터 유형을 웨어하우스 데이터와 결합할 수 있으며, 이를 통해 심층 분석을 실행하여 이전에는 얻을 수 없었던 통찰력을 얻을 수 있습니다. 또한, 스트림 컴퓨팅은 큰 가치의 데이터를 찾기 위한 분석 필터의 역할을 할 수 있으며, 이러한 데이터는 이후 InfoSphere BigInsights 또는 데이터 웨어하우스에 저장할 수 있습니다.
3. **쿼리 가능한 데이터 저장소:** 이 접근법에서는 자주 액세스되지 않거나 오래된 데이터를 정보 통합 소프트웨어 및 툴을 이용해 웨어하우스 및 애플리케이션 데이터베이스로부터 오프로드할 수 있습니다. 이러한 기능을 이용하면 자주 이용되지 않는 콜드(cold) 데이터를 저비용 스토리지에 보관할 수 있으며, 이러한

데이터는 쿼리 또는 BI 툴을 이용해 InfoSphere BigInsights 내에서 액세스할 수 있습니다. InfoSphere Data Explorer를 이용하면 InfoSphere BigInsights에 저장된 모든 데이터를 확인 및 탐색할 수 있습니다.

데이터 웨어하우스의 보강이 귀사에 적합한 활용 방안인지 확인하는 방법

다음과 같은 질문에 대해 보십시오.

- 대량의 데이터 세트(테라바이트 단위 또는 페타바이트 단위)로 인해 큰 부담을 느끼고 있습니까?
 - 현재의 웨어하우스 환경을 모든 데이터에 대한 저장소로 이용하고 있습니까?
 - 콜드 데이터 또는 자주 이용하지 않는 데이터를 대량으로 보유하고 있습니까?
 - 저장할 수 없거나 처리할 수 없는 데이터를 폐기할 필요가 있습니까?
 - 어떠한 데이터를 웨어하우스에 저장해야 할지 실시간으로 결정하기 위해 동적 데이터를 분석하기를 원합니까?
 - 복잡한 대량의 데이터에 대한 데이터 탐색을 실행하기를 원합니까?
 - 운영 데이터 이외의 데이터를 분석하기를 원합니까?
 - 현재 보유하고 있는 데이터를 이용해 일반적인 분석 및 새로운 유형의 분석을 실행하는 데 관심이 있습니까?
-

IBM의 빅데이터 플랫폼

본 백서에서 소개한 빅데이터 활용 방안 5가지는 빅데이터의 여정을 시작하고자 하는 기업에게 이상적인 출발점을 제공합니다. IBM의 빅데이터 플랫폼은 빅데이터로의 전환에 있어 중요한 역할을 할 수 있습니다.

빅데이터 활용 방안에는 다양한 유형을 가진 대량의 고속 데이터를 처리하는 업무를 위해 특별히 설계된 통합된 기술이 필요합니다. 이는 단일 제품 솔루션과 관련된 단순한 문제가 아닙니다. IBM의 플랫폼을 이용하면 사전에 통합된 구성요소를 활용하여 빅데이터 프로젝트에 필요한 시간과 비용을 줄이고 신속하게 ROI를 달성할 수 있습니다. 바로 이용할 수 있는 표준 기반의 서비스는 배치를 용이하게 합니다. 초기의 활용 방안에서 소규모의 빅데이터 프로젝트를 시작한 후, 빅데이터의 여정을 진행하면서 다른 활용 방안을 위한 프로젝트로 발전시킬 수 있습니다.

추가 정보

빅데이터 활용 방안 및 IBM의 빅데이터 플랫폼에 대해 자세히 알아보려면 IBM 영업 담당자 또는 IBM 비즈니스 파트너에게 문의하거나 다음 사이트를 참조하시기 바랍니다.

ibm.com/software/data/bigdata/use-cases.html

또한 IBM Global Financing은 고객이 소프트웨어 기능을 최대한 비용 효과적이고 전략적으로 얻을 수 있도록 지원합니다. 또한, 신용이 보장된 고객들과 협력하여 고객의 비즈니스와 개발 목표에 부합하는 파이낸싱 솔루션을 제공함으로써 효과적인 현금 관리와 총 소유비용 개선을 돕습니다. IBM Global Financing을 통해 중요한 IT 투자 자금을 확보하고 비즈니스 성장을 가속화하십시오. 자세한 정보는 다음 사이트를 참조하시기 바랍니다. ibm.com/financing



© Copyright IBM Corporation 2013

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
June 2013

IBM, IBM 로고, ibm.com, Analyst's Notebook, BigInsights, Cognos, i2, InfoSphere 및 SPSS는 전세계 여러 국가에 등록된 International Business Machines Corp.의 상표입니다. 기타 제품 및 서비스 이름은 IBM 또는 해당 회사의 상표입니다. 현재 IBM 상표 목록은 웹 "저작권 및 상표 정보" (www.ibm.com/legal/copytrade.shtml)에 있습니다.

본 문서는 발행일 기준으로 최신이고 IBM은 이를 통지없이 변경할 수 있습니다.

본 문서의 모든 정보는 타인의 권리 비침해, 상품성 및 특정 목적에의 적합성에 대한 묵시적 보증을 포함하여 묵시적이든 명시적이든 어떠한 종류의 보증 없이 "현상태대로" 제공됩니다. IBM 제품은 제공된 제품에 적용된 계약의 이용 약관에 따라 보증됩니다.

법적 요구사항을 준수하는지 확인해야 할 책임은 IBM 고객에게 있습니다. IBM은 법률 자문을 제공하지 않으며 IBM의 서비스나 제품을 통해 관련 법률이나 규정에 대한 고객의 준수 여부가 확인된다고 진술하거나 보증하지 않습니다..

¹ Financial Crimes Report to the Public: Fiscal Years 2010-2011. www.fbi.gov/stats-services/publications/financial-crimes-report-2010-2011

² McKinsey Global Institute. "Big data: The next frontier for innovation, competition, and productivity." May 2011. www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

³ McKinsey Global Institute. "Big data: The next frontier for innovation, competition, and productivity." May 2011. www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

⁴ McKinsey Global Institute. "Big data: The next frontier for innovation, competition, and productivity." May 2011. www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

⁵ International Telecommunication Union. "Measuring the Information Society." September 2012.

⁶ IBM Institute for Business Value in collaboration with the Saïd Business School at the University of Oxford. "Analytics: The real-world use of big data." November 2012. <http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-big-data-at-work.html>

⁷ IBM Institute for Business Value in collaboration with the Saïd Business School at the University of Oxford. "Analytics: The real-world use of big data." November 2012. <http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-big-data-at-work.html>



Please Recycle