

데이터 관리를 현대화하는 8가지 방법

인공 지능을 활용하고, 증가하는 데이터 다양성에
대처하고, 배포를 더 효과적으로 통합하는 방법





목차

소개

과거와 다른 데이터 3페이지

1

AI로 최적화와 앱 개발 주도
4페이지

2

빠른 보고와 분석 실현 6페이지

3

총소유비용 절감
8페이지

4

필요에 따라 온 프레미스 또는
클라우드를 선택하는 유연성
10페이지

5

확장성과 지속적인 가용성 증대
12페이지

6

새로운 데이터 유형 및 소스 지원
14페이지

7

오픈 소스 기술 통합
16페이지

8

새로운 스트리밍 및 패스트 데이터 기능 도입
19페이지

결론

데이터 관리 현대화를 시작하는 방법
21페이지

소개

과거와 다른 데이터



데이터 관리도 달라져야 할 때입니다.

일부 특정 소스에서 가져오는 고도로 구조화된 정보라는 협의의 데이터 시대는 오래 전에 끝났습니다. 여러 소스에서 가져오는 더 다양한 데이터 유형과 구조를 활용하고자 하는 조직은 이 개념을 바꿔야 합니다. 이러한 모든 요소는 IT 부서부터 현업(LOB) 직원, C레벨 경영진에 이르기까지 모두에게 즉각적이고 전환적인 영향을 미칠 수 있는 인사이트를 제시합니다.

효과적인 정보 사용과 관리는 새로운 비즈니스 기회를 추구하고, 고객을 유치 및 유지하고, 운영을 효율화합니다. 그러나 이러한 요건은 여러 워크로드 과제를 유발하며 기반 IT 인프라와 데이터베이스 시스템에 대한 수요를 증가시키며, 많은 경우 이를 감당할 준비가 되어 있지 않습니다.

이러한 과제를 어떻게 해결하는지가 관건입니다. 기존 시스템에 필요한 패치와 추가 기능, 지속적인 튜닝을 위해 더 많은 인력을 투입할 수도 있고, 풍부하고 새로운 데이터에 존재하는 잠재적인 인사이트를 그냥 무시할 수도 있을 것입니다. 그러나 많은 기업은 인공지능(AI), 그리고 온 프레미스와 퍼블릭 및 프라이빗 클라우드의 다양한 기능과 배포 옵션을 활용하여 데이터 인프라와 비즈니스를 혁신하는 새로운 솔루션을 찾는 방법으로 이 과제에 대처하고 있습니다.

이 ebook은 엔터프라이즈 설계자와 CIO가 데이터 관리를 현대화하여 정보 아키텍처 과제를 극복하고 새로운 비즈니스 기회를 발굴할 수 있는 8가지 방법에 대해, 산업별 전문 애널리스트의 의견과 실제 고객 경험을 함께 제공합니다.

1

AI가 이끄는 최적화와 앱 개발

경쟁력을 확보하려면 조직은 데이터에서 가장 심층적이고 가장 정확하며 실행 가능한 인사이트를 신속하게 추출하여 가장 가치 있는 시점에 인사이트를 구현해야 합니다. 데이터 기반 인사이트는 전사적인 혁신에 박차를 가하여 새로운 제품 또는 시장을 위한 기회를 발굴하고, 영업 직원에게 더 유의미한 논의를 할 수 있는 역량을 부여하고, 개선의 여지가 있는 내부 프로세스를 식별할 수 있습니다. 이러한 이유로 많은 조직이 AI에 맞게 만들어지고 AI를 기반으로 작동하는 데이터 관리 솔루션을 선택하고 있습니다.

AI 기반

AI 및 머신 러닝 기반 솔루션은 데이터 관리의 성능을 직접적으로 최적화하여, 비즈니스 전반에 파급적인 효과를 미칩니다.

예를 들어 머신 러닝을 사용하면 쿼리가 데이터에 이르는 경로를 지속적으로 개선하여 쿼리 시간을 크게 단축할 수 있습니다. 또한 머신 러닝을 활용하여 확실한 쿼리 결과에 더 높은 점수를 매겨 훨씬 더 쉽게, 효율적으로 최선의 인사이트에 도달할 수 있습니다. 자연어 기반 쿼리를 도입하는 것도 도움이 됩니다. 그러면 현업 담당자가 인터넷 검색과 비슷한 인터페이스를 사용하여 데이터로부터 직접 인사이트를 이끌어 낼 수 있습니다. 이를 통해 조직 전반으로 인사이트를 확산하면서 내부 병목 현상을 없애고 데이터 과학자와 비즈니스 애널리스트의 귀중한 시간을 절약할 수 있습니다.



AI 전용 설계

AI 전용으로 설계된 솔루션은 데이터 과학자와 애플리케이션 개발자가 프로토타입을 프로덕션으로 전환할 때 코드를 다시 작성할 필요 없이, 프로젝트에서 AI 구현에 필요한 프레임워크, 언어 및 도구를 신속하게 지원합니다.

AI 전용 솔루션에는 Python, GO, JSON 같은 인기 언어와 Jupyter Notebook에 대한 지원이 포함됩니다. 머신 러닝 라이브러리와 부가적인 도구를 통해 사용자는 AI 애플리케이션을 데이터 관리 솔루션과 통합할 수 있습니다. 이렇게 하면 데이터 과학자와 앱 개발자는 새로운 코드를 배우고 만드는 데 시간을 소비할 필요가 없습니다. 대신 프로젝트 작업을 즉시 시작할 수 있고, 이전 작업 또는 라이브러리의 기존 코드를 차용할 수 있기에, 처음부터 새로 시작하는 경우에 비해 더 빠르게 프로젝트를 완료할 수 있습니다.

데이터 관리 체크인

- 귀사의 데이터 환경은 가장 보편적으로 사용되는 데이터 사이언스 언어와 도구를 지원합니까?
- 쿼리를 머신 러닝에 맞게 최적화하고 자연어를 사용해서 작성할 수 있습니까?
- 쿼리가 반환되는 속도는 빠릅니까? 또한 쿼리의 응답이 정확하게 반환됩니까?

애널리스트들은 AI가 융합된 데이터 관리의 급부상을 예고합니다.

애널리스트들은 회사를 변화시킬 수 있는 AI 데이터 관리의 역량에 주목하고 있습니다. 451 Research의 최근 보고서에 따르면, 가장 데이터 지향적이라고 평가받은 기업의 응답자 중 88%가 AI와 머신 러닝이 데이터 플랫폼 및 분석 이니셔티브의 가장 중요한 구성요소라고 답했습니다.

반복적인 작업의 자동화부터 복잡한 작업의 자동화에 이르기까지 모든 작업에서 사람의 일을 보강하는 AI의 광범위한 적용 사례도 보고서에 나와 있습니다.



AI와 머신 러닝은 데이터 플랫폼 및 분석 이니셔티브에서 가장 중요한 구성요소입니다.

451 Research: 데이터 관리로 AI를 더 빠르게, AI로 데이터 관리를 더 빠르게 [보고서 읽기](#) →

88%

2

빠른 보고와 분석 실현

컴퓨팅과 처리부터 압축, 데이터 스킵에 이르기까지 보고와 분석의 속도를 높이는 여러 고급 기술을 사용할 수 있습니다. 이와 같은 여러 기술을 결합하면 인사이트를 제공하는 속도를 대폭 개선할 수 있습니다. 이러한 기술은 쿼리가 더 신속하게 반환되도록 하고, 기업에서 적시에 의사결정을 내릴 수 있게 하기 때문에, 데이터 과학자와 비즈니스 애널리스트에게 특히 중요합니다.

인메모리 컴퓨팅과 컬럼 기반 새도우 테이블

인메모리 컴퓨팅은 성능을 높일 수 있다는 점으로 인해 인기가 높아졌습니다. 그러나 인메모리 데이터베이스라고 해서 모두 같은 방식으로 데이터를 다루지는 않습니다. 데이터베이스는 읽기 및 쓰기 기능이 필요하므로 일반적으로 행으로 구성되며, 원하는 성능 개선을 달성하기 위해서는 모든 데이터가 메모리 내에 저장되어야 합니다. 그러나 데이터가 꾸준히 증가함에 따라 모든 데이터를 메모리 안에 집어넣기가 불가능한 경우도 있습니다. 행 기반 테이블과 나란히 분석할 수 있는 컬럼 기반 새도우 테이블이 필요한 이유가 여기에 있습니다. 컬럼 읽기는 더 효율적이며, 분석을 위해 정확히 필요한 컬럼을 메모리에 넣을 수 있습니다. 또한 IBM® Db2® BLU Acceleration®과 같은 인메모리 컬럼 기술 기반 스토리지의 경우 시스템 메모리, CPU 메모리로의 최적화된 데이터 이동 기능은 쿼리 완료 속도를 높여 줍니다. 따라서 OLTP 및 OLAP 워크로드를 동일한 데이터베이스에서 병렬로 실행할 수 있으며, 이는 데이터 웨어하우스로의 마이그레이션 없이 쿼리를 실행하고 더 신속하게 답을 반환할 수 있음을 의미합니다.



능동적 압축

능동적 압축은 데이터의 압축을 해제하지 않고도 데이터를 스캔하거나 비교할 수 있게 합니다. 따라서 압축 해제를 필수 데이터로 제한하고 꼭 필요할 때까지 연기할 수 있습니다. 데이터베이스에서 이를 달성하는 방법은 자체 압축 사전으로 각 컬럼을 압축하고 다른 값보다 더 자주 나타나는 데이터 값을 압축하는 것입니다. 이렇게 하면 I/O 처리량을 개선해서 인사이트를 얻는 시간을 줄이는 동시에, 처리 역량을 아껴 전체 스토리지 비용을 직접적으로 낮출 수 있습니다.

데이터 스킵

데이터 스킵은 쿼리와 무관한 데이터를 식별하여 무시함으로써 쿼리 수행 시간을 절약합니다. 이를 위해 데이터베이스는 시놉시스 테이블이라는 메타데이터 개체를 사용하여 컬럼의 최솟값과 최댓값을 추적합니다. 데이터 스킵 기술은 쿼리를 실행할 때 시놉시스 테이블에서 쿼리에 적용되는 값이 포함된 다양한 데이터를 검토해서 일치하지 않는 데이터를 건너뛵니다. 필요한 데이터만 읽고 사용하므로 스캔의 효율성이 높아지고, 불필요한 I/O가 줄어들고, 쿼리 실행 속도가 빨라집니다.

데이터 관리 체크인

- 분석 중 데이터 압축이 해제되기를 기다리는 시간이 얼마나 됩니까?
- 인사이트를 최대한 활용하기에 충분할 만큼 빠른 속도로 쿼리가 완료됩니까?
- 운영에 지장을 초래하지 않으면서 효율적으로 트랜잭션 데이터를 분석할 수 있습니까?

글로벌 제조업체는 주요 기술을 조금 사용하여 보고 시간을 98% 단축했습니다

브레이크 시스템 제조업체인 Knorr-Bremse는 인수합병과 조인트 벤처를 통해 전 세계적으로 사업 영역을 넓혔고, 이로 인해 데이터 볼륨이 크게 증가했습니다. 이 회사는 매일 최대 4TB의 데이터를 처리해야 했습니다. Knorr-Bremse는 능동적 압축, 데이터 스킵, 인메모리 컴퓨팅과 같은 기술을 선택하여 상당한 혜택을 얻었습니다.



Db2 BLU Acceleration을 통해 보고 시간을 3분에서 3초로 단축했습니다.

Thomas Brauchle
Knorr-Bremse의 IT 아키텍처 및 전략 임원

98%

3

총소유비용 절감

IT 예산이 계속 정체되면서 조직에서는 재무 자원과 인력의 효율적인 사용이 과거 어느 때보다 중요해졌습니다. 데이터베이스는 경제적이고 업계를 선도하는 데이터 관리 기능을 제공하는 동시에, 광범위한 완전 관리형 서비스와 엔터프라이즈급 지원을 통해 서비스 수준 협약(SLA)도 충족해야 합니다.

자동화 기능

관리 작업을 자동화하면 더 전략적인 이니셔티브에 인적 자원을 집중하고 워크플로의 속도를 높일 수 있으므로 위의 두 가지 목표 달성에 모두 도움이 됩니다. 설정 및 배포, 워크로드 관리, 리소스 활용 및 스토리지 관리와 유지보수, 업그레이드, 용량 확장과 같은 다양한 관리 작업을 자동화할 수 있는 데이터베이스 솔루션은 대단히 유용합니다.

각 배포에 걸쳐(경쟁사 클라우드까지 포함) 공통 코드 베이스를 공유하는 클라우드 및 온프레미스 옵션을 갖춘 데이터 솔루션을 선택하면 관리 시간과 노동을 줄이고 절약한 시간을 코드 재작성에 투입할 수 있습니다. 좋은 예가 Db2 데이터 관리 제품군입니다.



Spiceworks “2019 State of IT” 연구에서는 직원 수가 5,000명을 넘는 기업이 스토리지에 소비하는 비용을 조사했습니다. 이는 하드웨어, 호스팅 및 클라우드 서비스, 관리형 서비스 영역의 전체 IT 예산 중 8% 이상이었습니다. 따라서 스토리지 요구사항을 낮추면 큰 효과를 얻을 수 있습니다. 이는 스토리지 요구사항을 직접적으로 낮추는 데이터 압축과 다중 온도 데이터 관리 기술을 통해, 또는 콜드 데이터를 더 비용 효율적인 스토리지 옵션으로 덜어 내는 방법으로 달성할 수 있습니다.

현재 애플리케이션을 수정하지 않으면서 비용을 절감하고 프로세스 속도를 높이고자 데이터베이스 전환을 모색하는 기업은 클러스터 토폴로지 투명성을 갖춘 솔루션도 고려할 수 있습니다. 이러한 솔루션을 사용할 경우 애플리케이션에서 기반 클러스터와 데이터베이스 배포를 인식할 필요가 없고, 따라서 코딩과 테스트의 속도를 높이고 애플리케이션 개발자의 생산성을 향상할 수 있습니다.

데이터 관리
체크인

- 기존 시스템에서 사용 가능한 자동화는 어느 수준입니까?
- 조직에서 재작성에 소요되는 시간은 얼마입니까?
- 공급업체별로 데이터 관리 가격을 비교한 적이 있습니까?

Owens-Illinois는 현대화 후 7자릿수나 비용을 절감했습니다

글로벌 제조업체인 Owens-Illinois에서 7자릿수 비용 절감은 상당히 큰 성과입니다. 이러한 비용 절감과 함께 더 빨라진 트랜잭션, 50%에 가까운 데이터베이스 및 스토리지 공간 축소 효과까지 많은 혜택을 얻었습니다. 아래 동영상에서 전체 이야기를 알아보십시오.

총소유비용이 7자릿수 절감되고 스토리지는 50% 줄었으며 트랜잭션 응답 시간은 최소 20~30% 빨라졌습니다.”

Rod Masney
Owens-Illinois의 CIO



50%

4

필요에 따라 온 프레미스 또는 클라우드를 선택하는 유연성

고유한 필요에 따라 온 프레미스, 프라이빗 클라우드, 호스팅형 클라우드, 퍼블릭 클라우드 또는 다양한 플랫폼을 활용하는 하이브리드 모델 중에서 요구사항에 가장 잘 맞는 배포가 무엇인지를 고려할 수 있습니다. 데이터나 데이터의 용도는 제각기 다르므로 폭넓은 배포 옵션을 활용할 수 있어야 합니다.

조직 전략의 모든 측면을 지원할 수 있는 광범위한 배포 옵션을 제공하는 공급업체를 선택하면 단일 배포 형태만 제공하는 공급업체 선택에 비해 더 높은 유연성을 얻을 수 있습니다. 또한 배포의 모든 요소가 동일한 데이터 관리 제품군에 속한다면 포인트 솔루션을 끼워 맞추는 아키텍처에 비해 통합 성능이 더 우수할 가능성이 높습니다.

온 프레미스

온 프레미스 배포는 최적화, 데이터 액세스 및 가시성에 대한 높은 수준의 통제력을 제공하지만 하드웨어와 IT 인력에 많은 투자가 필요합니다. 온 프레미스는 보안이 무엇보다 중요할 경우, 또는 특정한 규정에 따라 반드시 필요한 경우에 보편적으로 사용됩니다.

프라이빗 클라우드

프라이빗 클라우드 배포는 자체 방화벽의 보안을 활용하는 동시에 일반적으로 퍼블릭 클라우드의 장점인 유연성도 제공할 수 있습니다. 컨테이너 기술을 사용할 수 있다는 점은 이 유형의 배포가 가진 중요한 이점입니다. 컨테이너는 가상 머신에 비해 신속하게 배포할 수 있고 작동을 위해 필요한 서버 공간도 더 작은 경우가 많습니다.

호스팅형 클라우드

호스팅형 클라우드 환경에서는 하드웨어 구매와 유지보수 책임은 없지만, 어느 정도는 시스템을 직접 관리할 수 있습니다. 호스팅형 클라우드 배포는 추가 인프라를 구매할 필요가 없고 기존 데이터베이스 라이선스를 사용할 수 있으므로 재해복구 프로젝트에 적합합니다. 호스팅형 클라우드는 온 프레미스 시스템과 원활하게 통합됩니다.

퍼블릭 클라우드

퍼블릭 클라우드 배포는 대역폭 및 모니터링과 같은 측면에서 직접 제어할 수 있는 수준이 가장 낮은 대신 유연성이 높습니다. 클라우드 버스팅이라 알려진 방법을 통해 퍼블릭 클라우드 용량을 신속하게 추가하여 프라이빗 클라우드를 보완하고, 컴퓨팅과 스토리지 수요를 분리하고, 사용한 만큼만 비용을 지불한다는 점은 온 관리형 클라우드 배포가 백업 및 개발 용도 또는 테스트 환경 용도로 적합하다는 것을 의미합니다.

데이터 관리 체크인

- 동일한 코드 베이스로 온 프레미스와 클라우드 배포를 지원할 수 있습니까?
- 현재 얼마나 쉽게 다양한 데이터 배포를 추가하고 통합할 수 있습니까?
- 멀티 클라우드 옵션을 사용할 수 있습니까? 멀티 클라우드 옵션이 얼마나 원활하게 통합됩니까?

전체적이고 연결된 하이브리드 환경이 중요한 이유는 무엇일까요?



기업은 더 이상 온 프레미스와 클라우드 중 양자택일이라는 잘못된 이분법에 끌리지 않고, 잘 통합된 온프레미스와 클라우드 옵션을 함께 제공할 수 있는 공급업체를 찾습니다. 다양한 데이터 관리 배포 사이의 공백을 메우기 위해 필요한 기능에 대해 자세히 알아보십시오.

클라우드로 가는 연결 통로 [백서 읽기](#) →



5

확장성과 지속적인 가용성 증대

비즈니스 운영을 지속하기 위해서는 엔터프라이즈에 의해 생성되고 갈수록 증가하는 데이터를 처리하기에 충분한 성능을 갖춘 상시 가동되는 데이터 시스템이 필요합니다. 고객 및 파트너뿐만 아니라 회사 내부에서도 지속적인 업무와 즉각적인 응답을 기대합니다. 실패는 인사이트를 기반으로 한 부가적인 비즈니스 기회의 손실, 파트너 관계 약화 및 호감도 손실로 이어집니다. 여기에서는 확장성과 가용성을 위한 가장 유용한 세 가지 옵션에 대해 설명합니다.

공유 스토리지 클러스터링

옵션 중 하나는 모두 동일한 데이터 소스에 의존하는 다양한 노드 간에 작업을 분산할 수 있게 해 주는 공유 스토리지 클러스터링입니다. 공유 스토리지 클러스터링의 두 가지 주요 기능은 아키텍처의 노드 간 자동 워크로드 분산, 그리고 공유 스토리지 클러스터의 각 노드를 서로에 대한 장애 조치 노드로 사용하는 내장된 자동 장애 조치입니다. 이러한 여러 기능이 조합되어 예비성을 늘리고, 특정 노드의 과부하를 방지하여 장애 발생 가능성을 낮추고, 재해에 직면한 상황에서도 더 효과적으로 가용성을 제공할 수 있습니다.



퍼블릭 클라우드로 온
프레미스와 프라이빗
클라우드 보완

온 프레미스와 프라이빗 클라우드 배포에 퍼블릭 클라우드를 추가하면 확장성과 가용성을 모두 개선할 수 있습니다. 이를 위해서는 온 프레미스와 프라이빗 클라우드 서비스를 기본 용량으로 사용하면서 추가 스토리지 또는 컴퓨팅 용량이 필요할 때 퍼블릭 클라우드 서비스를 사용하면 됩니다. 효율적으로 이를 달성하려면 데이터가 배포에 따라 격리되지 않도록 클라우드와 온 프레미스 시스템 간의 강력한 통합이 필요합니다. 또한 기존 온 프레미스 또는 프라이빗 클라우드 서비스를 보완하는 퍼블릭 클라우드 환경을 가동하면 엔터프라이즈 내의 예정된 또는 예정되지 않은 가동중단 발생 시 더 나은 가용성을 제공할 수 있습니다.

클라우드 스토리지와
컴퓨팅을 독립적으로
확장

기업은 클라우드 서비스의 스토리지와 컴퓨팅을 독립적으로 늘리거나 줄여 클라우드별 워크로드의 변동에 대처할 수 있어야 합니다. 독립적인 확장 기능을 사용하면 클라우드에서 최대 성능 수요를 충족하기 위해 컴퓨팅 용량을 늘렸다가, 이러한 부가적인 성능이 더 이상 필요 없을 때 다시 축소하여 비용을 절감할 수 있습니다. 가용성이 높으면 고가용성 재해 복구(HADR), 일일 백업, 유지보수, 엔터프라이즈 지원과 같은 완전 관리형 클라우드 서비스를 통해 필요한 작업도 줄어듭니다.

데이터 관리 체크인

- 현재 데이터 전송 비용을 얼마나 지출하고 있습니까?
- 사용 중인 클라우드 및 온 프레미스 서비스가 공통 코드 베이스 공유를 통해 더 효과적인 통합을 지원합니까?
- 클라우드 데이터베이스 스토리지와 컴퓨팅을 독립적으로 확장할 수 있습니까?

네덜란드의 의료 비즈니스 인텔리전스 센터에서 클라우드 확장을 적용한 이유는 무엇일까요?

네덜란드의 의료 비즈니스 인텔리전스 센터인 Vektis는 데이터 자체(변경 및 조작이 가능)를 제공하는 방식에서 수집하는 데이터의 단일 버전에 대한 액세스를 제공하는 방식으로 전환함으로써 단일 버전의 진실(one version of the truth)을 생성하고자 했습니다. 또한 이 회사는 다수의 사용자가 동시에 어려운 쿼리를 입력하는 상황을 수용하기 위해 확장이 필요하다는 사실도 인식하고 있었습니다.



확장과 축소가 가능한 유연성을 자랑하는 IBM Db2 Warehouse on Cloud로 마이그레이션하고자 했습니다. 덕분에 민첩성도 더 높아졌고 비용도 통제할 수 있게 되었습니다.

Herman Bennema
Vektis의 CEO

6

새로운 데이터 유형 및 소스 지원

이제는 복수의 데이터 유형과 소스가 보편적입니다. 소셜 미디어, 사물 인터넷(IoT)에 연결된 기계의 센서 및 멀티미디어 콘텐츠 저장소와 같은 소스에서 생성되는 반구조화된 데이터와 비구조화된 데이터는 비약적으로 증가하고 있습니다. 또한 이러한 소스의 상당수는 매우 높은 속도와 규모로 스트리밍 데이터를 생산합니다.

조직 외부의 데이터를 기업이 이미 수집한 데이터와 결합하면 더 나은 의사결정을 이끌어 낼 수 있는 더 심층적이고 정확한 인사이트를 얻을 수 있습니다. 예를 들어 내부 데이터만으로 신제품의 초기 판매량이 견실한 것을 알 수 있다고 가정해 보겠습니다. 그러나 이 데이터를 소셜 미디어 및 기타 외부 비구조화된 데이터와 결합하면 다른 양상이 나타날 수 있습니다. 기업은 감성 분석을 사용하여 해당 제품이 많이 팔렸지만 부정적인 경험을 제공했는지 여부를 알 수 있고, 이를 통해 판매가 줄어들기 전에 문제를 포착하여 수정할 수 있습니다.

그러나 내부 데이터와 외부 데이터의 결합은 다양한 소스에서 다양한 속도로 생성되는 광범위한 정보 유형을 제공, 캡처, 저장, 처리하기 위한 유연성을 제공해야 하는 데이터 인프라 관점에서 새로운 배포 및 개발 과제를 의미합니다. 기업에는 필요할 때, 필요한 방식으로 데이터를 제공하고 언제나 현재 일어나는 상황을 정확히 포착할 수 있는 데이터 환경이 필요합니다.

일반적이지 않은 데이터의 대표적인 사례:



시청각 콘텐츠 - 조직에서 다른 방법으로는 찾거나 검토하기 어려운 지식을 얻을 수 있는 기회를 제공하며 유용한 맥락 정보를 추가합니다.



이벤트 데이터 - 대체로 시간에 매우 민감한 정보를 제공합니다. 관련성이 손실되기 전에 캡처 및 분석된 이벤트 데이터는 실시간 조치를 위한 방향타 역할을 합니다.



소셜 미디어 데이터 - 고객에 대한 360° 시야를 제공하여 마케팅 대상을 파악하고, 판매를 늘리고, 고객 또는 전체적인 대중의 인식과 관련된 문제를 신속하게 식별하거나 완화합니다.



종이 문서 내의 데이터 - 액세스는 더 어려울 수 있지만 수동으로 기록된 정보를 통해 인사이트를 얻을 수 있습니다. 비슷한 온라인 양식에 비해 더 풍부하고 세부적인 맥락에 관한 정보가 포함될 수 있습니다.



IoT 센서 데이터 - 기계 또는 생산 라인에 오작동 위험이 있을 경우 알람을 전달하여 사후 대응이 아닌 예방적 조치를 이끕니다.



이메일 콘텐츠 - 시청각 구성요소를 많이 포함하고 있고, 설설문 데이터와 같은 보고 데이터나 트랜잭션과 같은 관찰 데이터에 비해 더 견고한 정보를 제공합니다.

데이터 관리 체크인

- 현재 액세스 및 분석이 가능한 구조화된 데이터의 양은 어느 정도입니까?
- 인사이트를 끌어내기 위해 내부 데이터와 외부 데이터를 얼마나 자주 결합합니까?
- 귀사 또는 고객이 일상적으로 사용하는 IoT 장치는 몇 개입니까?

Capitalogix는 구조화된 데이터와 비구조화된 데이터를 혼합하여 헤지 펀드를 위한 의미 있는 맥락을 만들었습니다.

주식의 평균 보유시간이 20초에 불과한 상황에서 금융기술기업 Capitalogix는 더 견고한 인사이트를 더 빠르게 얻고 시장 트렌드보다 앞서 나가기 위한 방법을 지속적으로 고민하고 있습니다. 이를 위해 다양한 속도의 수많은 구조화 또는 비구조화된 데이터를 사용합니다.

완전히 비구조화될 수 있는 대안 데이터인 Twitter 피드의 감성과 위성 영상에 액세스할 수 있습니다.

Howard Getson
Capitalogix의 CEO



7

오픈 소스 기술 통합

오픈 소스 기술은 현대 데이터 관리 환경의 기반입니다. 오픈 소스 옵션을 통해 전체 개발 커뮤니티의 노력을 활용하고 벤더 종속성을 피하고 더 많은 데이터 소스를 활용할 수 있는 역량의 중요함은 아무리 강조해도 지나치지 않습니다. 오픈 소스는 적절히 구현되면 더 낮은 비용과 더 우수한 인사이트로 이어질 수 있습니다.

커뮤니티 혁신을 활용하고
벤더 종속성 방지

오픈 소스 솔루션은 개인, 그리고 때로는 기업이 참여하는 커뮤니티에 의해 지속적으로 개선됩니다. 어떠한 문제 또는 새로운 개념에도 대처할 수 있는 광범위한 기술 지식과 수많은 숙련된 참여자는 단일 기업이 따라갈 수 없는 영역이며, 이는 더 혁신적인 기술로 이어집니다. 이러한 첨단 기술 외에도 오픈 소스 솔루션은 여러 방법으로 벤더 종속성 위험을 완화합니다. 여러 벤더를 통해 동일한 오픈 소스 기술에 액세스할 수 있으므로 다시 코딩하는 번거로움 없이 더 쉽게 다른 벤더를 선택할 수 있습니다. 또한 오픈 소스 코딩 기술은 사유 시스템에 필요한 사유 지식과 달리 다른 사람에게 전달할 수 있습니다. 벤더 종속성을 없애면 경쟁 기회의 증가로 인해 비용이 절감됩니다.



더 많은 데이터에
액세스

또한 오픈 소스 솔루션은 기존 데이터베이스 또는 데이터 웨어하우스에 저장할 수 있는 것보다 더 많은 데이터에 액세스할 수 있게 해 줍니다. 예를 들어 Hadoop은 데이터를 저장할 때 스키마 할당을 피할 수 있는 기능 덕분에 비구조화된 정보의 주 저장소로 사용되는 경우가 많습니다. 또한 데이터 웨어하우스의 콜드 데이터를 Hadoop으로 덜어내는 것도 가능합니다.

그러나 오픈 소스 기술을 통한 더 많은 데이터 액세스는 Hadoop에 국한되지 않습니다. MongoDB 및 PostgreSQL과 같은 데이터베이스는 사유 데이터베이스를 보강합니다. 예를 들어 MongoDB는 NoSQL 기반 문서 데이터베이스입니다. 데이터를 JSON 문서에 저장함으로써 더 구조적인 관계형 데이터베이스에서는 일반적으로 찾아볼 수 없는 빠른 스키마 진화가 가능합니다. 따라서 모바일 애플리케이션, 실시간 분석, 콘텐츠 관리 등 스키마 유연성의 이점을 활용할 수 있는 적용 사례에 적합합니다. 마찬가지로, PostgreSQL은 관계형 데이터베이스지만 조직의 기존 관계형 데이터베이스와 충돌하기보다는 오히려 이를 보강합니다. PostgreSQL은 객체 관계형 데이터베이스이기 때문입니다. 즉 사용자 정의 객체와 함께 더 복잡한 데이터 구조를 생성, 저장, 검색하기 위해 이러한 객체의 동작 방식을 지원합니다.

PostgreSQL은 UUID, 통화, 열거형, 지오메트릭, 바이너리, 네트워크 주소, 비트 문자열, 텍스트 검색, XML, JSON, 배열, 합성 및 범위 유형과 같은 데이터 유형과 객체 식별 및 로그 위치를 위한 몇 가지 내부 유형을 지원합니다. 전체 범위의 유연성은 항상 필요한 것은 아니지만 표준 관계형 데이터베이스에 대한 유용한 보완 역할을 합니다. 궁극적으로 이러한 보완적 오픈 소스 기술은 더 많은 소스의 더 많은 데이터 유형에 액세스할 수 있게 해 주며, 더 큰 재무적 이익을 위한 비즈니스 혁신과 최적화를 이끌 수 있는 더 깊은 인사이트에 필요한 추가적인 정보를 제공합니다.

애널리스트들이 IBM과 Cloudera 전략적 파트너십의 총소유가치에 주목하기 시작했습니다.

Cabot Partners의 애널리스트들은 인공 지능(AI)의 성장과 긍정적인 영향은 적절한 정보 아키텍처(IA)를 통해서만 가능하다는 사실을 이해하고 있습니다. 여기에는 IBM과 Cloudera가 제공하는 솔루션과 같은 엔터프라이즈급 오픈 소스 솔루션이 포함됩니다. 애널리스트들은 총소유가치(TVO) 연구를 수행한 후 이렇게 결론을 내렸습니다. “IBM과 Cloudera는 함께 고객이 분석/AI를 산업화하는 여정에서 전례 없는 유연성과 선택권, 가치를 제공합니다.” 전체 보고서를 읽고 이들 두 오픈 소스 리더의 조합이 총소유비용 절감, 생산성 향상, 수익 증대 및 위험 완화에 어떻게 도움이 될 수 있는지 알아보십시오.



IBM과 Cloudera는 함께 고객이 분석/AI를 산업화하는 여정에서 전례 없는 유연성과 선택권, 가치를 선사합니다.

고급 분석과 AI를 위한 더 넓은 선택권과 더 큰 가치 [보고서 읽기](#) →

적절한 오픈 소스 선택

엔터프라이즈는 오픈 소스 솔루션이 안전하고 잘 통합되며 적절히 관리되는지를 확인해야 합니다. 자체적으로 할 수도 있지만, 엔터프라이즈급 공급업체 지원과 부가가치 솔루션을 사용하는 편이 더 효율적인 경우가 많습니다. 보안을 예로 들어 보겠습니다. 보안을 기본 제공하는 사전 패키징된 엔터프라이즈 솔루션을 사용하면 개발자는 회사 방화벽 뒤에서 시스템을 통합하고 액세스 제어를 설정하는 데 시간을 소비하는 대신 더 가치 있는 활동에 집중할 수 있습니다.

통합의 경우, 공통 SQL 엔진을 사용한 Hadoop과 데이터 웨어하우스 전반의 데이터 연함은 엔터프라이즈 오픈 소스 솔루션이 큰 가치를 제공하는 또 다른 영역입니다. 이 통합이 없으면 예를 들어 첫 주 판매 데이터와 소셜 미디어 감성 데이터를 연계하여 미래 수요에 대한 예측하기 위해 훨씬 더 많은 작업이 필요할 것입니다. 엔터프라이즈 수준의 거버넌스도 데이터 결합에 필요한 작업을 줄여 줍니다. 데이터 레이크를 구성하는 엔터프라이즈급 오픈 소스 솔루션의 메타데이터, 카탈로그 및 일반 구성이 없으면 꼭 필요한 데이터를 정확하게 가져오기가 어렵습니다. 데이터 레이크가 일명 데이터 늪이 되고 맵니다.

데이터 관리 체크인

- 현재 오픈 소스로 대체할 수 있는 사유 솔루션에 갇혀 있습니까?
- 오픈 소스 데이터베이스 기능에 더 잘 맞는 특정 적용사례가 있습니까?
- 현재 오픈 소스 솔루션에 필요한 보안, 통합 및 거버넌스 기능이 있습니까?



8

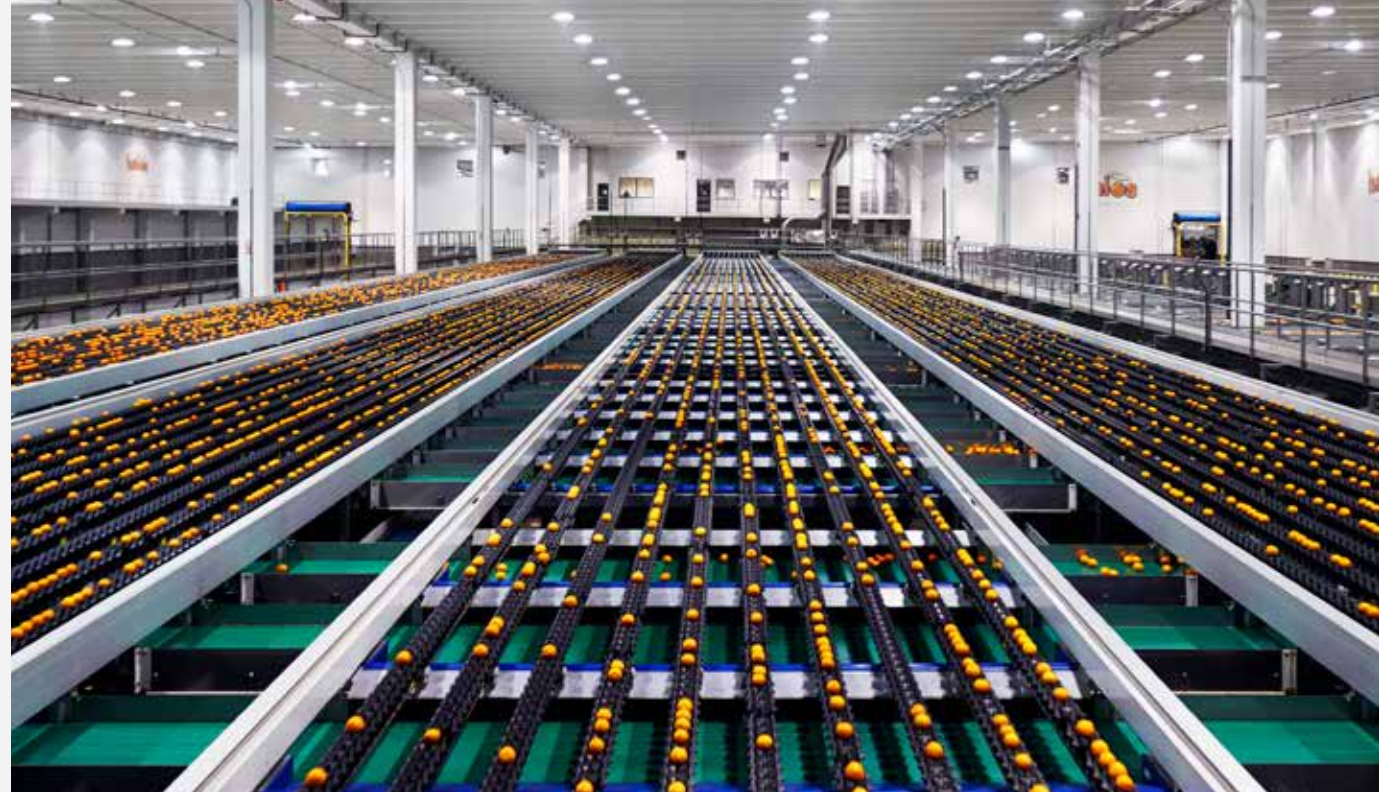
새로운 스트리밍 및 패스트 데이터 기능 도입

데이터의 폭발적인 증가, AI 프로젝트의 확산, 오픈 소스의 이점에 대한 이해 증대와 같은 요인들로 인해 기업이 스트리밍 데이터를 캡처하고 사용하는 방식이 바뀌고 있습니다. 실제로 이 전환은 워낙 두드러져서, 일부 애널리스트들은 이 데이터 관리 범주를 비교적 좁게 정의된 스트리밍 분석 대신 더 넓은 범위의 패스트 데이터 (fast data)라는 용어로 지칭하기 시작했습니다. 패스트 데이터 솔루션으로 업그레이드하면 조직은 빠르게 데이터를 흡수 및 사용하고, 오픈 소스 기술을 더 원활하게 도입하고, 더 강력한 분석과 AI를 수행할 수 있습니다.

더 높은 속도,
더 많은 데이터,
더 나은 인사이트
확보

IoT, 클릭스트림, 모바일 앱을 비롯해 개별 이벤트로부터 빠르게 흘러가는 기타 데이터 유형은 본질적으로 수명이 짧습니다. 따라서 이러한 유용한 데이터 포인트를 최대한 활용하려면 빠른 흡수가 가능한 시스템이 필요합니다. 이제 IBM Db2 Event Store와 같은 솔루션에서 노드 단 3개로 하루에 최대 2,500억 개의 이벤트를 저장하고 분석할 수 있습니다. 이전 기술로 같은 결과를 달성하려면 100개에 가까운 노드가 필요했습니다.

이를 통해 조직은 두 가지의 이점을 얻게 됩니다. 첫 번째는 비용 절감입니다. 노드 요구 사항이 낮아지면 기업은 더 높은 효율성을 통해 인프라 비용을 없애고, 그 비용을 더 혁신적이고 유의미한 프로젝트에 투입할 수 있게 됩니다. 두 번째는 각 노드가 더 효과적으로 운영되므로 비슷한 수의 노드를 유지하더라도 더 많은 데이터를 캡처할 수 있어 더 확실하고 완전한 인사이트를 얻을 수 있습니다. 클릭스트림 데이터의 경우, 캡처되는 각각의 부가적인 이벤트는 고객이 웹사이트에 머무는 동안 고객을 더 잘 이해하고 대응하기 위한 기회이자, 고객의 요구에 맞게 조정하고 수익을 늘릴 수 있는 역량을 제공하는 것이기도 합니다.



오픈 소스로 패스트
데이터 강화

이전 섹션에서 다룬 오픈 소스의 이점 중 상당수는(예를 들어 벤더 종속성 감소, 더 많은 데이터에 액세스) 오픈 소스를 패스트 데이터 솔루션에 도입할 때도 동일하게 적용됩니다. 그러나 이 경우 이벤트와 스트리밍 데이터의 속도로 인한 과제가 발생합니다. Apache Parquet와 같은 오픈 포맷 오브젝트 스토리지에 작성할 수 있는 솔루션 역시 대량의 스트리밍 데이터를 저장하는 용도로 최근 인기를 끌고 있습니다. Hadoop과 같은 다른 환경에 비해 더 빠르고 비용도 더 저렴할 수 있기 때문입니다. 놀라운 속도로 이벤트 데이터를 캡처하는 기능은 흡수 및 분석 엔진 못지않게 저장소에 따라서도 좌우됩니다.

새로운 패스트 데이터 솔루션은 광범위한 분석 및 AI 적용사례를 통해 과거 데이터와 스트리밍 데이터를 더 쉽게 결합한 인사이트를 제공합니다. 패스트 데이터 솔루션이 캡처하는 많은 데이터 유형에 대한 조치에는 실시간에 가까운 분석이 중요하므로 데이터 흡수가 부실하게 수행되지 않도록 하는 것이 중요합니다. 메모리 최적화를 통해 더 높은 체계성을 달성하면 데이터를 더 손쉽게 사용할 수 있습니다. 좋은 예가 알림 모니터링입니다. 핵심 시스템을 모니터링하는 IoT 디바이스는 대량의 고속 스트리밍 데이터를 알림용으로 전송합니다. 이를 수용할 때 적절하게 체계화된 데이터 분석 기능이 있으면, 누군가 데이터 구조를 할당할 때까지 기다리는 대신 더 쉽게, 거의 즉각적으로 대응할 수 있습니다. 중요한 알림이 수신하여 즉각적인 조치를 취하느냐 아니냐는 큰 차이가 있습니다.

또한 그 순간의 스트리밍 데이터를 과거 데이터 또는 최근에 스트리밍된 데이터와 결합하면 앞뒤 맥락을 포함하여 더 정확한 분석 결과를 확보할 수 있습니다. 최근 스트리밍된 데이터는 데이터 포인트가 트렌트인지 단발성 현상인지 여부를 판단하는 데 도움이 됩니다. 과거 데이터를 사용하면 알림을 보내야 할 정도의 장애에 앞서 일반적으로 발생하는 데이터 포인트를 확인할 수 있습니다. 또한 이는 알림 대상 여부를 판단을 지속적으로 개선하기 위해 머신 러닝을 도입하기에도 좋은 기회입니다. 이렇게 하면 알림 모니터링 시스템의 정확성이 더 높아지고, 중대한 실패를 방지하면서도 일반적으로 잘못된 알림과 관련된 번거로움을 없앨 수 있습니다.

데이터 관리
체크인

- 현재 실시간에 가까운 의사결정을 위해 스트리밍 데이터와 과거 데이터를 결합할 수 있습니까?
- 현재 사용하는 노드는 몇 개이며 이를 통해 몇 개의 이벤트를 처리할 수 있습니까?
- 객체 스토리지에 개방형 데이터 유형을 사용할 수 있다면 필요한 유연성을 얻을 수 있습니까?

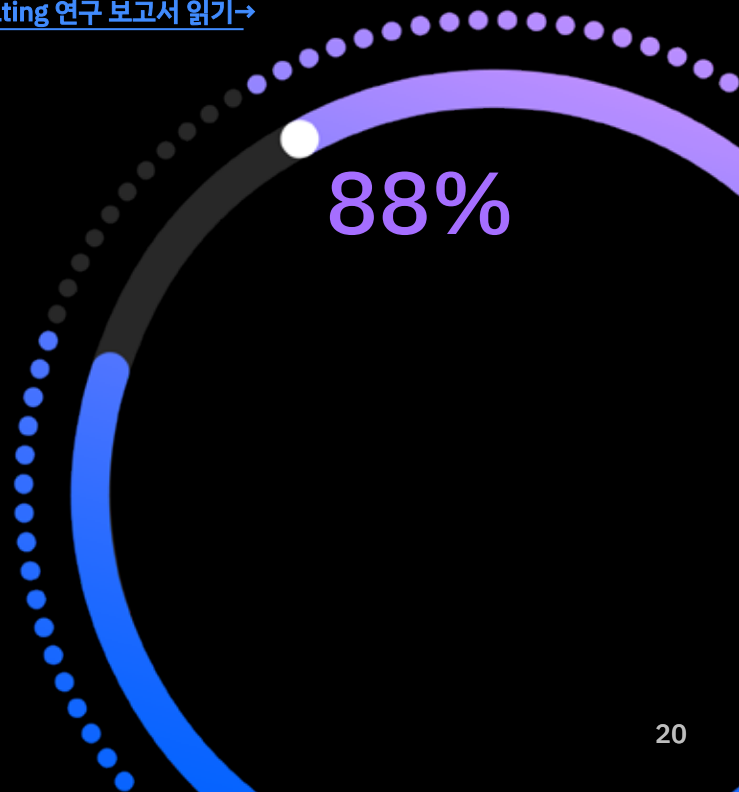
Forrester Consulting의 최근 연구는 빠르게 변화하는 패스트 데이터 분야에 대해 더 자세한 내용을 보여줍니다.

Forrester Consulting은 최근 패스트 데이터(스트리밍 및 이벤트 데이터 포함)에 대한 연구를 실시했는데, 이 연구에서는 패스트 데이터의 지속적인 중요함이 재차 확인되었습니다. Forrester는 각 개인을 대상으로 설문을 수행하여 현재 사용과 미래 계획, 두 가지 측면 모두에서 패스트 데이터의 확산을 확인했습니다. 또한 Forrester는 패스트 데이터 솔루션의 효과를 저하시킬 수 있는 장애물을 살펴보고 몇 가지 중요한 권장 사항도 제안했습니다.



88%는 저장된 스트리밍 데이터에 대해 실시간에 가까운 분석을 수행해야 한다고 말합니다.

패스트 데이터를 그냥 기다려서는 안 됩니다.
[Forrester Consulting 연구 보고서 읽기](#)→



결론

데이터 관리 현대화를 시작하는 방법

데이터에 대한 조직의 의존도는 매일 더 높아지고 있습니다. 이 리소스는 빠른 속도로 증가하며 유형과 소스도 과거 어느 때보다 많으므로 기회와 과제가 함께 발생합니다. 그래서 이와 같은 요구사항에 대처하고 요구사항과 함께 성장하는 하이브리드 데이터 관리 인프라가 중요한 것입니다.

IBM Hybrid Data Management에 대한 자세한 정보는 IBM 담당자 또는 IBM 비즈니스 파트너에게 문의하십시오. 또한 IBM Cloud Pak for Data를 무료로 체험할 수도 있습니다.

이 eBook이나 하이브리드 데이터 관리, 또는 IBM Cloud Pak for Data에 대해 질문이 있을 경우 IBM 전문가가 30분 동안 무료 상담을 통해 답변 드립니다.

전문가에게 질문하기 →

IBM HDM 및 Cloud Pak for Data에서 비롯되는 차이

아키텍처가 미래에 대비하고, 경쟁이 치열한 이 환경에서 탁월함을 지속적으로 유지하기 위해 필요한 AI, 머신 러닝 및 데이터 과학 툴을 수용하는 동시에 온 프레미스, 프라이빗 클라우드 또는 다양한 퍼블릭 클라우드 공급업체의 배포에 대한 선택권을 제공하는 것 역시 중요합니다. 데이터베이스, 데이터 웨어하우스, 데이터 레이크 및 패스트 데이터 기능을 포함한 광범위한 데이터 관리 상품도 필수적입니다.

IBM Hybrid Data Management는 귀사에 필요한 AI 기능, 배포 옵션 및 다양한 솔루션을 제공합니다. IBM Db2는 AI를 통해 작동하며 AI에 맞게 설계되어 성능을 개선하면서 AI 애플리케이션을 구축하는 데 필요한 도구를 제공합니다. IBM Db2 제품의 기반인 공통 SQL 엔진은 개선된 통합 기능, 그리고 한 번 작성 후 어디서나 실행 가능한 기능을 제공합니다.

IBM의 Hybrid Data Management 제품은 IBM Cloud Pak for Data의 일부로 제공됩니다. Cloud Pak for Data를 통해 Kubernetes 컨테이너에서 솔루션을 실행하면 이러한 솔루션은 하드웨어 종류나 클라우드 제공업체에 상관없이 각각 Linux 컨테이너 또는 RedHat을 지원하는 환경에서는 모두 실행 가능합니다. 부가적인 혜택으로, 데이터 가상화가 Cloud Pak for Data에 포함되어 데이터 이동 없이 더 효율적으로 워크로드를 실행할 수 있습니다.

Cloud Pak for Data를 사용하면 용량도 간단히 늘릴 수 있습니다. 추가 VPC를 더하기만 하면 추가 용량을 이용할 수 있습니다. 이는 하이브리드 데이터 관리 확장 및 기본 제공 상품을 넘어 IBM Watson Studio Open Scale과 같은 솔루션까지 확장되므로 AI Ladder의 모든 부분이 연결된 상태로 유지되고, 이를 통해 AI 프로젝트에서 자동으로, 매끄럽게 데이터를 관리하고 사용할 수 있습니다.



© Copyright IBM Corporation 2020

IBM Corporation
Route 100
Somers, NY 10589

Produced in the United States of America
2020년 3월

IBM, IBM 로고, **ibm.com**, Db2 및 BLU Acceleration은 전 세계에 등록되어 있는 International Business Machines Corp.의 상표입니다. 기타 제품 및 서비스 이름은 IBM 또는 타사의 상표일 수 있습니다. 최신 IBM 상표 목록은 웹사이트 “저작권 및 상표 정보 (www.ibm.com/legal/copytrade.shtml)”에 있습니다.

이 문서는 처음 발행될 당시의 날짜를 기준으로 업데이트되었으며 IBM은 언제든지 문서 내용을 변경할 수 있습니다. IBM이 사업을 운영하는 국가라도 일부 제품은 공급되지 않을 수 있습니다.

여기서 소개된 성능 데이터와 고객 사례는 오로지 예시를 목적으로 작성된 것입니다. 실제 성능 결과는 특정 구성 및 운영 환경에 따라 달라질 수 있습니다.

이 문서의 정보는 상품성에 대한 보증, 특정 목적의 적합성 여부 및 저작권을 침해하지 않는다는 보증 또는 조건을 포함해 명시적 또는 암묵적 보증 없이 “있는 그대로” 제공됩니다. IBM 제품은 제공된 약정에 명시된 조항 및 조건에 따라 보증됩니다.