

하이브리드 환경을 위한 데이터 통합 및 라이프사이클 관리 전략 개발

IBM

목차

개요: 혼돈 속 질서의 정립	3
데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의	7
데이터 통합 및 라이프사이클 관리를 위한 IBM 솔루션	17
다음 단계: 클라우드 거버넌스 논의의 지속	19

개요: 혼돈 속 질서의 정립

- 네 가지 핵심

데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

- 기업이 필요로 하는 데이터는 어디에 있는가?
- 데이터를 조합하여 기업이 필요로 하는 정보를 생성하는 방법은 무엇인가?
- 이 모든 데이터를 반드시 보관해야 하는가?

데이터 통합 및 라이프사이클 관리를 위한 IBM 솔루션

다음 단계: 클라우드 거버넌스 논의의 지속

개요: 혼돈 속 질서의 정립

클라우드 기반의 데이터는 업계에서 경쟁 우위를 확립하고 유지하고자 하는 기업에 풍부한 잠재적 정보를 제공합니다. 하지만 ["The truth about information governance and the cloud"](#)(정보 거버넌스 및 클라우드에 대한 진실)에서 설명한 대로, 대부분의 기업은 기존의 사내 데이터를 타사의 새로운 클라우드 기반 데이터에 맞게 조정해야 하는 어려운 과제에 직면합니다. 이러한 "하이브리드" 환경에서 기업은 중요한 의사결정을 내릴 수 있는 인스턴스를 찾으려고 합니다.

일반적으로 하이브리드 환경이 충분한 사전 계획 없이 확장되기 때문에 계속 증가하는 데이터 저장소를 관리하는 것은 매우 어렵습니다. 그리고, 하둡처럼 확장 가능한 데이터 플랫폼은 비교할 수 없이 우수한 비용 이점과 분석 가능성을 제공합니다.

하지만 하둡과 하둡 기반 솔루션은 빅 데이터 볼륨을 처리하는 경우에 이점이 있을 뿐, 하둡 자체는 데이터 통합을 위해 설계되지 않았습니다. 데이터 통합에는 고유한 요구사항(예: 거버넌스 지원, 메타데이터 관리, 데이터 품질 및 유연한 데이터 제공 스타일)이 수반됩니다. 이러한 혼란스런 상황을 명확히 이해할 수 있는 방법이 있습니다. 항상 그렇듯이 첫 번째 단계는 문제의 본질을 이해하는 것입니다. 데이터 자체에 초점을 두고, 데이터의 소스나 데이터 관리에 사용되는 시스템에는 덜 집중해야 합니다. 데이터 및 데이터에서 파생된 정보의 소유권에 가장 높은 우선 순위를 두면 나머지 우선 순위는 쉽게 결정됩니다.

개요: 혼돈 속 질서의 정립

- 네 가지 핵심

데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

- 기업이 필요로 하는 데이터는 어디에 있는가?
- 데이터를 조합하여 기업이 필요로 하는 정보를 생성하는 방법은 무엇인가?
- 이 모든 데이터를 반드시 보관해야 하는가?

데이터 통합 및 라이프사이클 관리를 위한 IBM 솔루션

다음 단계: 클라우드 거버넌스 논의의 지속

네 가지 핵심

기업에서 클라우드의 재무 이점을 실현하는 동시에, 클라우드 소스에서 선별한 정보가 안전하고 신뢰할 수 있는 정보인지 확인하는 방법은 무엇일까요? 해답은 바로 '거버넌스'입니다.

이상적인 하이브리드 정보 거버넌스는 IT 및 비즈니스의 네 가지 주요 우선순위를 기반으로 합니다.

1. 정보의 개념에 대한 광범위한 합의

여기에는 비즈니스에 필요한 정보와 그러한 정보의 처리 방식에 대한 공통의 정책 및 일반 언어 규칙의 메타데이터가 포함됩니다.

2. 소유한 정보 자산을 유지관리 및 모니터링하는 방식에 대한 명확한 합의

예를 들면 사내 시스템의 마스터 데이터 관리에 대한 운영 데이터 품질 규칙이 있습니다.

3. 전략적 정보 자산의 보안 및 보호를 위한 전사적/부서별 표준 방침

예를 들면 정보에 대한 역할 기반 액세스의 명시, 정보 공유 방법을 규정하는 규칙의 작성, 중요한 정보를 제3자로부터 보호 등이 있습니다.

4. 엔터프라이즈 데이터 통합 전략

여기에는 데이터를 이동하고 전략적 정보로 취합하는 방법을 계획하고, 이러한 정보를 시간의 경과에 따라 유지 관리하는 방법을 파악하는 라이프사이클 관리가 포함됩니다.

개요: 혼돈 속 질서의 정립

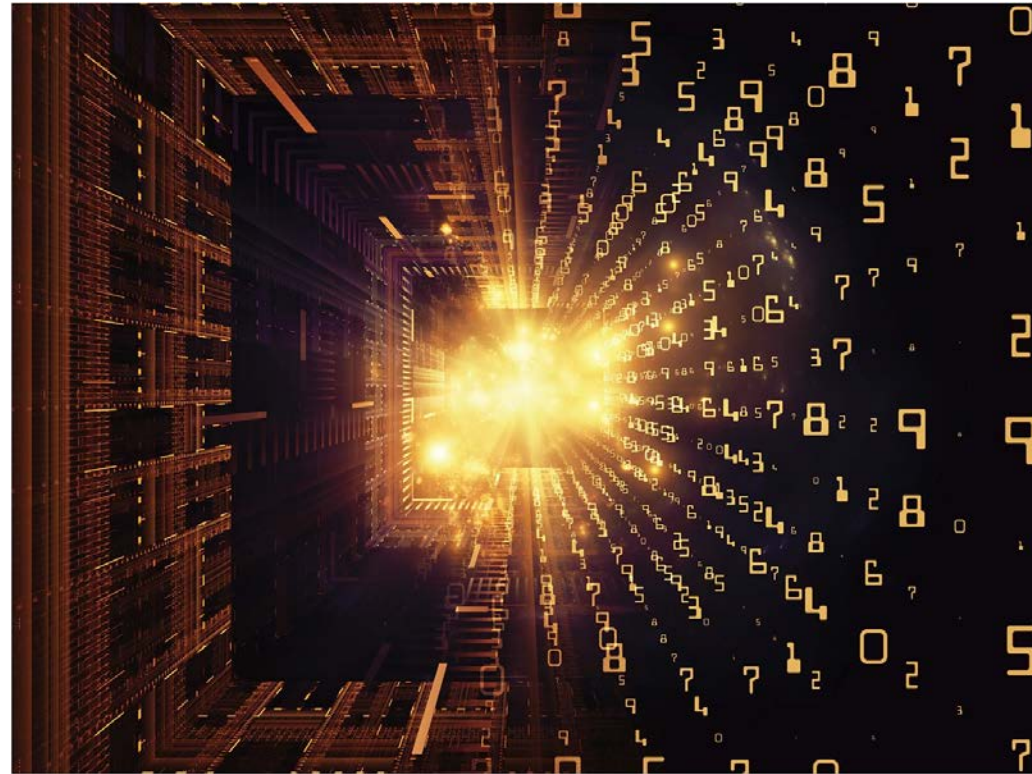
- 네 가지 핵심

데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

- 기업이 필요로 하는 데이터는 어디에 있는가?
- 데이터를 조합하여 기업이 필요로 하는 정보를 생성하는 방법은 무엇인가?
- 이 모든 데이터를 반드시 보관해야 하는가?

데이터 통합 및 라이프사이클 관리를 위한 IBM 솔루션

다음 단계: 클라우드 거버넌스 논의의 지속



개요: 혼돈 속 질서의 정립

- 네 가지 핵심

데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

- 기업이 필요로 하는 데이터는 어디에 있는가?
- 데이터를 조합하여 기업이 필요로 하는 정보를 생성하는 방법은 무엇인가?
- 이 모든 데이터를 반드시 보관해야 하는가?

데이터 통합 및 라이프사이클 관리를 위한 IBM 솔루션

다음 단계: 클라우드 거버넌스 논의의 지속

이러한 요소들이 하이브리드 환경의 정보 거버넌스를 위한 기반을 형성합니다. 개별 사례에서는 프로세스와 조직적/기술적 지원 요소들이 조화를 이루어야 성공을 거둘 수 있습니다. 이러한 핵심 요소가 제 기능을 다할 때 기업은 유연하면서도 확신을 가지고 신속하게 움직일 수 있습니다.

이 전자책은 네 번째 핵심: 데이터 통합 및 라이프사이클 관리 전략의 개발 및 실행에 중점을 둡니다.

전략적 정보의 소유권 획득

하이브리드 환경을 채택하기 위해 IT 전략을 완전히 바꿀 필요는 없습니다. 사실 환경에서 클라우드에 기반한 부분은 비즈니스 우선 순위에 따라 신속하게 변화합니다. 하지만 클라우드 기반 소스에서 유래하는 데이터의 비율이 아주 작더라도, IT에는 데이터 통합 및 보안을 위한 계획이 필요합니다. IT는 기업이 모든 데이터 및 처리 과정에서 생성되는 정보를 위치에 관계없이 소유할 수 있도록 지원해야 합니다. 하이브리드 인프라와 탈중심화된 컴퓨팅은 전략적 정보 자산 생성이라는 궁극적인 목표를 달성하기 위한 수단입니다.

데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

개요: 혼돈 속 질서의 정립

- 네 가지 핵심

데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

- 기업이 필요로 하는 데이터는 어디에 있는가?
- 데이터를 조합하여 기업이 필요로 하는 정보를 생성하는 방법은 무엇인가?
- 이 모든 데이터를 반드시 보관해야 하는가?

데이터 통합 및 라이프사이클 관리를 위한 IBM 솔루션

다음 단계: 클라우드 거버넌스 논의의 지속

지금까지 기업에서 전략적으로 중요한 데이터 소유의 중요성에 대해 알아보았습니다. 이는 다시 분명하게 명시된 데이터 통합 및 라이프사이클 전략의 필요성으로 이어집니다. 이러한 전략은 허용 가능한 데이터 소스, 소유할 데이터 요소, 소유한 데이터를 보관 또는 유지 관리하는 방식, 그리고 이러한 데이터가 조직에서 존속되는 기간을 정의합니다.

이러한 결정을 내리려면 먼저 세 가지 질문의 답을 생각해 보십시오.

1. 기업이 필요로 하는 데이터는 어디에 있는가?
2. 데이터를 조합하여 기업이 필요로 하는 정보를 생성하는 방법은 무엇인가?
3. 이 모든 데이터를 반드시 보관해야 하는가?

각 질문을 하나씩 살펴 보겠습니다.

기업이 필요로 하는 데이터는 어디에 있는가?

비즈니스 리더는 데이터가 어디에 있든 데이터의 힘을 활용하고자 합니다. 하지만 광범위한 데이터 세계에 접근하기 전에 알아야 할 점은 데이터 볼륨이 증가할수록 소스 정보의 신뢰성과 정확성을 확인하기가 무척 어려워진다는 사실입니다. 이러한 신뢰성 문제를 해결하지 않으면 최종 사용자가 데이터에서 생성된 인사이트에 대한 신뢰를 잃고, 그 결과 기회나 위협에 대응하지 못할 수 있습니다.

유감스럽지만, 데이터 및 데이터 소스의 볼륨과 복잡성 때문에 정보를 검색, 관리 및 수정하기 위한 기존의 수동적인 방법이 더 이상 효과적이지 않습니다. 이제는 정보의 소스가 사내든 클라우드든 관계없이, 다양한 데이터 애플리케이션, 데이터 웨어하우스 및 데이터 웨어하우스 보강 이니셔티브를 지원하는 정보 통합 및 거버넌스(Information Integration and Governance, IIG) 솔루션을 구현해야 합니다.

개요: 혼돈 속 질서의 정립

- 네 가지 핵심

데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

- 기업이 필요로 하는 데이터는 어디에 있는가?
- 데이터를 조합하여 기업이 필요로 하는 정보를 생성하는 방법은 무엇인가?
- 이 모든 데이터를 반드시 보관해야 하는가?

데이터 통합 및 라이프사이클 관리를 위한 IBM 솔루션
다음 단계: 클라우드 거버넌스 논의의 지속

중요한 또는 잠재적으로 중요한 데이터 소스가 발견될 경우, 이상적인 데이터 통합 전략은 데이터 생성 지점의 정보를 통합하여 무수한 데이터 토폴리지를 제시합니다. **견고한 IIG 프로그램은 다양한 데이터 세트에서 컨텍스트를 제공하고 직원들이 정보에 기반한 의사 결정을 내리도록 지원하기 위해 자동화된 검색, 프로파일링 및 메타데이터 관리 기능을 포함해야 합니다.**

이러한 맥락에서 분명한 사실은, 사내 데이터를 획득하여 분석을 위한 상태로 준비하는 것만으로는 충분하지 않다는 점입니다. 대신, 하이브리드 환경에서는 사내 및 클라우드 기반의 통합 방식과 기술(예: Platform-as-a-service, PaaS)을 결합해야 합니다. 사내 데이터를 사내 기술로 통합하여 관리하고, 클라우드 기반 데이터를 클라우드 기반 기술로 통합하여 관리해야 합니다.

또한 데이터 통합 기술은 사내 데이터든 클라우드 데이터든 다양한 형태의 데이터를 지원하고, 데이터 마트에서 Apache 하둡 시스템에 이르기까지 다양한 데이터 스토리지 기술을 통합할 수 있을 만큼 민첩해야 합니다.

데이터를 조합하여 기업이 필요로 하는 정보를 생성하는 방법은 무엇인가?

데이터 소스에 데이터 통합 기술을 적용하는 경우, 기업이 필요로 하는 정보 자산을 생성하려면 데이터를 어떻게 조합해야 할까요? 데이터는 요청 시 다양한 소스에서 빠른 속도로 스트리밍되므로, 무엇보다 성능이 중요합니다.

비즈니스 리더가 변화하는 시장 상황에 최대한 신속하게 대응할 수 있도록 하려면 빠르게 변화하는 데이터를 시스템의 다양한 애플리케이션에 신속하게 입력해야 합니다. 이러한 데이터를 성공적으로 처리하려면 다음과 같은 특성을 갖춘 엔터프라이즈급 데이터 통합 솔루션이 필요합니다.

- **동적:** 현재 및 미래의 성능 요구사항을 충족할 수 있습니다.
- **확장 가능:** 신속하고 간편한 확장성을 위해 파티셔닝됩니다.
- **통합:** 하둡 또는 기타 빅 데이터 스토리지 기술과 통합됩니다.

개요: 혼돈 속 질서의 정립

- 네 가지 핵심

데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

- 기업이 필요로 하는 데이터는 어디에 있는가?
- 데이터를 조합하여 기업이 필요로 하는 정보를 생성하는 방법은 무엇인가?
- 이 모든 데이터를 반드시 보관해야 하는가?

데이터 통합 및 라이프사이클 관리를 위한 IBM 솔루션

다음 단계: 클라우드 거버넌스 논의의 지속

대용량 확장 가능 기반

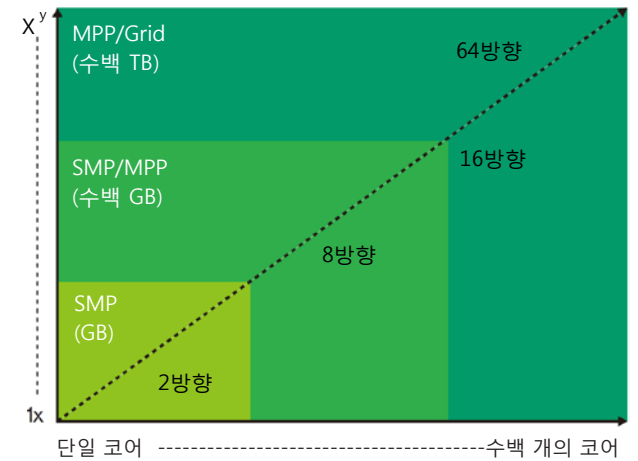
데이터 통합을 위해 대규모 하이브리드 데이터 볼륨을 처리하는 데 있어서 가장 중요한 요구사항은 대량 데이터 확장성(Massive Data Scalability, MDS)입니다. MDS는 대량의 데이터를 동시에 처리하는 기능을 제공하여 다양한 워크로드를 처리하는 데 소요되는 시간을 크게 줄여줍니다. 다른 프로세싱 모델과 달리, MDS 시스템은 사용 가능한 하드웨어 리소스를 최적화하여 노드당 처리되는 데이터의 양을 극대화합니다(그림 1 참조).

MDS가 중요한 이유는 대규모의 하이브리드 데이터 볼륨을 한 번에 처리하여 많은 고가치 비즈니스 문제를 단번에 해결할 수 있고 하드웨어 플랫폼이 예상되는 이점을 제공하도록 보장할 수 있기 때문입니다.

MDS 시스템에는 네 가지 공통점이 있습니다.

1. 비공유(Shared-nothing, SN) 아키텍처
2. 소프트웨어 데이터 흐름을 사용하여 구현
3. 선형적인 데이터 확장성을 위해 데이터 파티셔닝을 이용
4. 설계 분리(design isolation) 환경 사용

하드웨어 아키텍처의 데이터 확장성



- 동일한 기능
- 모든 아키텍처
- 선형적인 속도 향상
- "N방향"으로 확장하기 위한 결정(런타임 시)

그림 1. 무제한의 데이터 볼륨을 처리하기 위한 MDS 요구사항을 지원하는데 필요한 중요 특성

목차 > 데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

개요: 혼돈 속 질서의 정립

- 네 가지 핵심

데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

- 기업이 필요로 하는 데이터는 어디에 있는가?
- 데이터를 조합하여 기업이 필요로 하는 정보를 생성하는 방법은 무엇인가?
- 이 모든 데이터를 반드시 보관해야 하는가?

데이터 통합 및 라이프사이클 관리를 위한 IBM 솔루션

다음 단계: 클라우드 거버넌스 논의의 지속

비공유 아키텍처의 소프트웨어는 컴퓨팅 노드에서 데이터 세트를 파티셔닝하고 각 데이터 파티션에 대하여 실행되는 것과 동일한 애플리케이션 논리로 단일 애플리케이션을 실행하여, 비공유 대량 병렬 처리(Massively Parallel Processing, MPP) 아키텍처를 이용하도록 철저하게 설계되었습니다(그림 2 참조). 이는 시스템 어디에도 단일 경합 지점이나 단일 처리 병목 지점이 존재하지 않음을 의미합니다. 따라서 데이터 볼륨, 처리량 또는 프로세스 및 노드 개수에 상한이 없습니다.

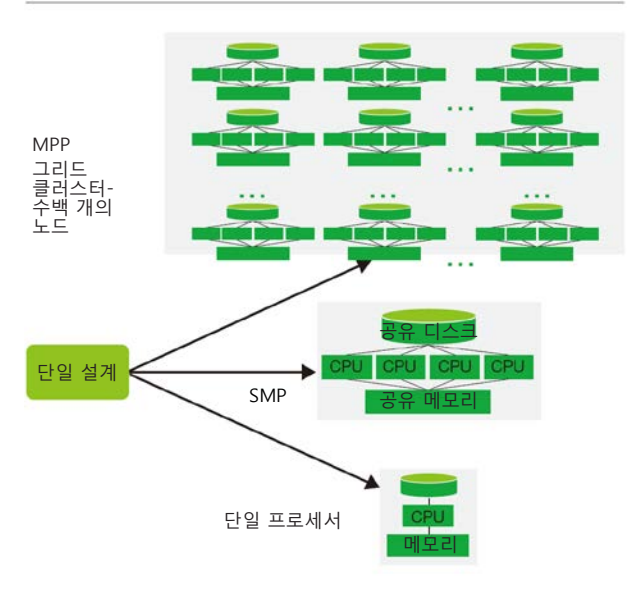


그림 2. 비공유 아키텍처의 예

개요: 혼돈 속 질서의 정립

- 네 가지 핵심

데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

- 기업이 필요로 하는 데이터는 어디에 있는가?
- 데이터를 조합하여 기업이 필요로 하는 정보를 생성하는 방법은 무엇인가?
- 이 모든 데이터를 반드시 보관해야 하는가?

데이터 통합 및 라이프사이클 관리를 위한 IBM 솔루션

다음 단계: 클라우드 거버넌스 논의의 지속

소프트웨어 데이터 흐름은 단일 노드 및 서로 다른 노드 간에 데이터 파이프라이닝과 데이터 파티셔닝을 손쉽게 구현하고 실행할 수 있도록 하여 SN 아키텍처의 완벽한 활용을 지원합니다. 또한 소프트웨어 데이터 흐름은 병렬 처리 애플리케이션을 빌드하고 조정하는 작업의 복잡성이 사용자에게 드러나지 않게 숨깁니다.

소프트웨어 데이터 흐름은 대칭 멀티프로세싱(SMP) 서버 내에서 멀티코어 프로세서를 사용(애플리케이션 수직 확장)하고 여러 시스템으로 확장(애플리케이션 수평 확장)하기 위한 최고의

아키텍처이며, 이는 하이브리드 사내/클라우드 데이터 환경에서 앞으로 더욱 널리 사용될 시스템 아키텍처입니다. 이 아키텍처의 특성은 다음과 같습니다.

- SMP 노드 내에서 그리고 SMP 노드 간에 파이프라이닝되고 파티셔닝된 병렬 처리를 지원
- 모든 하드웨어 아키텍처에서 병렬 처리를 위한 단일 메커니즘을 제공하여 복잡성을 제거하도록 지원
- 병렬 처리 애플리케이션의 구축, 조정 및 실행과 관련한 어려움을 줄여줌
- 데이터 볼륨, 처리량 및 처리 노드의 개수에 상한이 없음

목차 > 데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

개요: 혼돈 속 질서의 정립

- 네 가지 핵심

데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

- 기업이 필요로 하는 데이터는 어디에 있는가?
- 데이터를 조합하여 기업이 필요로 하는 정보를 생성하는 방법은 무엇인가?
- 이 모든 데이터를 반드시 보관해야 하는가?

데이터 통합 및 라이프사이클 관리를 위한 IBM 솔루션

다음 단계: 클라우드 거버넌스 논의의 지속

데이터 파티셔닝은 데이터 세트가 독립적인 여러 노드에서 파티셔닝되고, 파티셔닝된 모든 데이터를 대상으로 단일 작업이 동일한 애플리케이션 논리를 실행하는 것을 의미합니다(그림 3 참조). 작업 파티셔닝과 같은 다른 방식에서는 정렬, 병합 및 집계할 수 있는 데이터의 양이 단일 노드에서 처리할 수 있는 수준으로 제한되므로 데이터 볼륨 증가에 따라 선형적인 데이터 확장성을 제공할 수 없습니다. 이는 또한 필요한 경우나 원하는 경우 다양한 소스의 데이터를 고유한 노드에 지정할 수 있음을 의미합니다.

데이터 파티셔닝을 사용하는 시스템의 특성은 다음과 같습니다.

- 데이터 파티션을 여러 노드에 분할
- 하나의 작업을 여러 노드에서 병렬 실행
- 디스크에 랜딩하지 않고 파이프라이닝 및 재파티셔닝을 여러 스테이지 및 여러 노드 간에 구현
- 빅 데이터에 저비용의 그리드 하드웨어를 이용

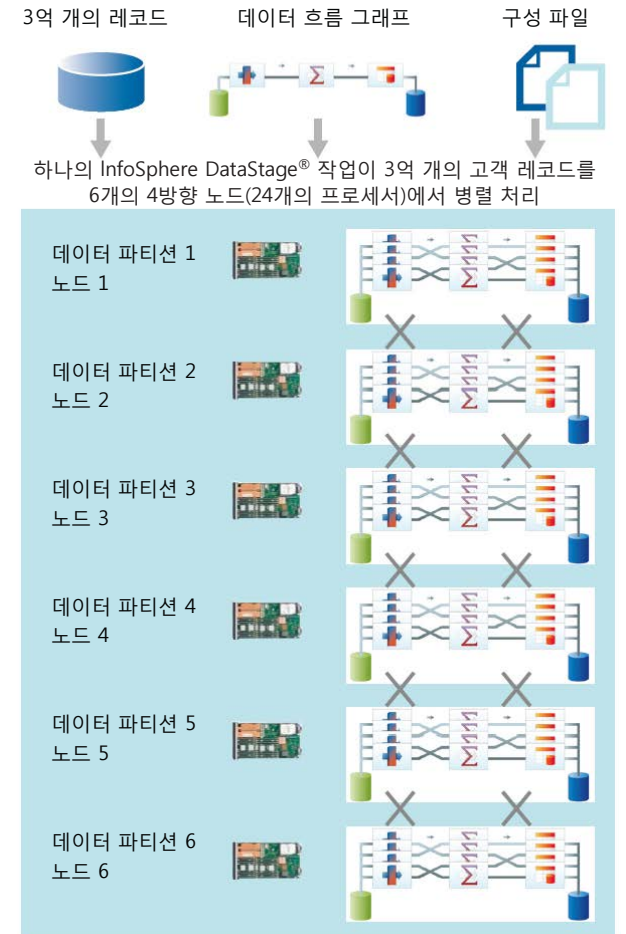


그림 3. 데이터 파티셔닝 아키텍처

개요: 혼돈 속 질서의 정립

- 네 가지 핵심

데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

- 기업이 필요로 하는 데이터는 어디에 있는가?
- 데이터를 조합하여 기업이 필요로 하는 정보를 생성하는 방법은 무엇인가?
- 이 모든 데이터를 반드시 보관해야 하는가?

데이터 통합 및 라이프사이클 관리를 위한 IBM 솔루션

다음 단계: 클라우드 거버넌스 논의의 지속

마지막으로 설계 분리는 개발자가 데이터 처리 작업을 한번에 설계한 다음, 작업을 재설계하거나 재조정할 필요 없이 모든 하드웨어 구성에서 사용할 수 있는 방식입니다. 특징과 이점은 다음과 같습니다.

- 한 번 빌드하여 어디서나 수정 없이 실행
- 병렬 처리를 위해 통합된 단일 메커니즘으로 복잡성을 줄여줌
- 작업의 개발과 런타임 시 병렬 처리의 표현을 확실하게 분리
- 하드웨어 아키텍처를 변경할 때마다 성능 조정이 필요하지 않음
- 데이터 확장성의 상한 없이 하드웨어 추가가 가능

확장성이 가장 우수한 플랫폼(IBM® Netezza®, IBM PureData™ System, IBM DB2® Database Partitioning Feature, Teradata, Hadoop 및 IBM InfoSphere® Information Server)들은 이러한 네 가지 특성을 완벽하게 지원하도록 설계되었으며 MPP 및 CoG(Commodity Grid) 아키텍처를 원활하게 활용할 수 있습니다.

신속한 인사이트를 위해 데이터 복제 이용

하이브리드 데이터 소스에서 파생되는 인사이트의 양을 극대화하려면 사용 사례에 따라 다른 데이터 제공 방식을 이용해야 합니다. 어떤 사용 사례에서는 신뢰할 수 있는 결정을 내리기 위해 1분 이내(또는 1초 이내)의 데이터 뷰가 필요합니다. 사기 감지, 여러 채널의 재고 분석, 실시간 운영 분석과 같은 시나리오에서 한 달, 한 주 또는 하루가 지난 데이터를 기반으로 한 의사 결정은 비생산적입니다.

경쟁자보다 신속하게 데이터를 수집하고 활용하여 보다 우수한 연속적인 서비스를 제공함으로써 우수한 고객 경험을 제공할 수 있습니다. 이를 위해 데이터 변환 및 제공 요구사항이 (특히 변경 데이터 캡처와 관련한) 데이터 복제 기능을 기반으로 하는 실시간 데이터 전송을 포함하도록 확장되었습니다.

개요: 혼돈 속 질서의 정립

- 네 가지 핵심

데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

- 기업이 필요로 하는 데이터는 어디에 있는가?
- 데이터를 조합하여 기업이 필요로 하는 정보를 생성하는 방법은 무엇인가?
- 이 모든 데이터를 반드시 보관해야 하는가?

데이터 통합 및 라이프사이클 관리를 위한 IBM 솔루션

다음 단계: 클라우드 거버넌스 논의의 지속

실시간 데이터 통합의 성능과 확장성을 극대화하려면 세 가지 요소를 고려해야 합니다.

1. **소스에서 변경을 캡처하는 방식:** 소스에서 변경을 캡처하는 가장 유연하고 효율적인 옵션은 소스 로그에 변경이 작성될 때 변경을 캡처하는 복제 프로세스입니다. 소스 데이터가 수정되는 즉시 이 메커니즘이 변경 사실을 감지하고, 소스 데이터베이스 및 애플리케이션에 영향을 거의 또는 전혀 미치지 않으면서 변경된 데이터를 전송하며, 따라서 대규모 배치 창고의 필요성을 최소화합니다.
2. **사용되는 메커니즘:** 데이터 복제에는 여러 가지 메커니즘이 사용될 수 있습니다. 적절하게 구현된 로그 기반의 캡처 방식은 소스 데이터베이스에 미치는 영향이 상대적으로 낮은 경우가 많으며, 따라서 전반적인 성능이 더 높습니다.
3. **일시적인 데이터 지속성:** 데이터가 일시적으로 지속되는지 여부도 데이터 복제 성능에 영향을 미칩니다. 이론적으로는, 데이터가 디스크에 기록될 필요가 없고 변환 엔진에 의해 액세스되므로 기업이 성능을 높이기 위해 데이터를 존속시키지 않고 변경 사항을 스트리밍할 수 있습니다.

하지만 데이터 통합 솔루션에서 중요한 요소는 성능만이 아닙니다. 다음과 같은 사항을 추가로 고려해야 합니다.

- **솔루션 유연성:** 유연한 메커니즘은 대용량 데이터를 위한 HDFS(Hadoop Distributed File System)를 포함하여 여러 플랫폼, 소스 및 대상을 지원합니다.
- **기존 IT 인프라 및 프로세스에 미치는 영향:** 이상적인 데이터 복제 솔루션은 스크립팅이나 Java 같은 일반적인 프로그래밍 언어를 통해 간편한 자동화를 허용하여 기존 변경 관리 프로세스에 손쉽게 통합될 수 있어야 합니다.
- **사용 편의성:** 학습 곡선의 효과가 크면 데이터 통합에도 영향을 미칩니다. 손쉬운 구성 및 모니터링을 지원하는 효과적인 그래픽 사용자 인터페이스는 빅 데이터에서 현실적인 인사이트를 얻는 데 소요되는 시간을 최소화합니다.

개요: 혼돈 속 질서의 정립

- 네 가지 핵심

데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

- 기업이 필요로 하는 데이터는 어디에 있는가?
- 데이터를 조합하여 기업이 필요로 하는 정보를 생성하는 방법은 무엇인가?
- 이 모든 데이터를 반드시 보관해야 하는가?

데이터 통합 및 라이프사이클 관리를 위한 IBM 솔루션

다음 단계: 클라우드 거버넌스 논의의 지속

이 모든 데이터를 반드시 보관해야 하는가?

많은 분석 사용자들이 데이터를 전부 보관하는 방식을 요구합니다. 현재 유용하지 않은 데이터에서 나중에 어떤 인사이트를 얻을 수 있을지 모르기 때문입니다. 하지만 이런 방식은 비즈니스에 심각한 위험과 막대한 비용을 초래할 수 있습니다.

"나중에 누군가에게 필요할 경우를 위해 보관한다"는 지침을 따를 경우, 많은 기업에서 너무 많은 이력 데이터를 보관하게 되며 이 중에서 대부분이 가치가 없습니다. 지나치게 큰 저장소와 보관 수단을 마련하면 상황이 악화될 뿐입니다. 또한, 기업은 규모가 증가하는 기밀 정보의 개인정보 보호 및 보안을 보장해야 합니다. 정부 및 산업 규정은 기업이 개인 정보를 위치에 관계없이(테스트 및 개발 환경을 포함하여) 보호할 것을 요구합니다.

게다가 사용자들이 하이브리드 데이터 볼륨에 대한 쿼리를 실행함에 따라 느린 응답 시간과 애플리케이션 성능의 저하가 중요한 문제로 부상하고 있습니다. 지속적인 데이터 증가가 억제되지 않으면 리소스를 역량 이상으로 사용하게 되고 중요한 쿼리 및 보고 프로세스의 응답 시간에 부정적인 영향을 미칩니다. 이러한 문제는 프로덕션 환경에 영향을 미치고 업그레이드, 마이그레이션 및 재해 복구 작업을 방해할 수 있습니다. 이와 같이 비즈니스를 중단시킬 수 있는 문제를 방지하려면 이력 데이터와 휴면 데이터에 대하여 지능적인 데이터 관리를 구현해야 합니다.

데이터 증가는 인프라 및 운영 비용을 증가시킬 수 있으며, 데이터 웨어하우징 또는 데이터 관리 예산의 상당 부분을 소비하는 경우가 많습니다. 데이터 볼륨이 증가하면 더 많은 용량이 필요하고, 확장되는 인프라를 유지, 모니터링 및 관리하기 위해 추가 하드웨어를 구입하고 더 많은 비용을 지출해야 하는 경우가 많습니다. 대규모의 데이터 웨어하우스와 복잡한 데이터 환경은 일반적으로 더 큰 서버, 어플라이언스 및 테스트 환경을 필요로 하며, 따라서 인력, 전력 및 법률적 비용이 증가할 뿐만 아니라 데이터베이스 및 데이터베이스 도구의 소프트웨어 라이선싱 비용도 증가합니다.

개요: 혼돈 속 질서의 정립

- 네 가지 핵심
-

데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

- 기업이 필요로 하는 데이터는 어디에 있는가?
 - 데이터를 조합하여 기업이 필요로 하는 정보를 생성하는 방법은 무엇인가?
 - 이 모든 데이터를 반드시 보관해야 하는가?
-

데이터 통합 및 라이프사이클 관리를 위한 IBM 솔루션

다음 단계: 클라우드 거버넌스 논의의 지속

아카이빙의 역할

데이터 라이프사이클은 데이터가 생성, 사용, 공유, 업데이트, 저장되고 궁극적으로 아카이빙되거나 방어적으로 폐기되는 여러 과정에 걸쳐 확장됩니다. 데이터 라이프사이클 관리는 특히 아카이빙 단계에서 중요한 역할을 합니다.

아카이빙 전략의 일환으로, 중요한 데이터 요소는 참조 및 추후 사용을 위해 보관하고 기업의 비즈니스 또는 법률적 요구사항을 지원하는 데 더 이상 필요하지 않은 데이터는 삭제하는 보존 정책이 설계되었습니다. 효과적인 데이터 라이프사이클 관리는 데이터를 수십 개의 데이터 소스에 걸친 정보를 포함할 수 있는 완전한 컨텍스트로 아카이빙하는 동시에, 데이터를 수명과 같은 특정한 요소 또는 비즈니스 규칙에 따라 아카이빙하는 인텔리전스를 포함합니다. 데이터 라이프사이클 관리는 스토리지 관리자가 계층화되고 자동화된 스토리지 전략을 개발하여 휴면 데이터를 데이터 웨어하우스에 아카이빙하고, 따라서 전반적인 분석 애플리케이션 성능을 향상시킬 수 있게 도와줍니다.

효과적인 데이터 라이프사이클 관리는 다음과 같은 방법으로 IT 및 비즈니스 이해관계자에게 모두 이점을 제공합니다.

- **수익 증가:** 개발 라이프사이클 동안 인프라 및 자본 비용 절감, 생산성 향상 및 애플리케이션 결함 감소
- **위험 감소:** 애플리케이션 정지 시간 감소, 서비스 및 성능 중단 최소화, 데이터 보존 요구사항 준수
- **비즈니스 민첩성 촉진:** 출시 기간 단축, 애플리케이션 성능 강화 및 애플리케이션 품질 향상

데이터 통합 및 라이프사이클 관리를 위한 IBM 솔루션

개요: 혼돈 속 질서의 정립

- 네 가지 핵심

데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

- 기업이 필요로 하는 데이터는 어디에 있는가?
- 데이터를 조합하여 기업이 필요로 하는 정보를 생성하는 방법은 무엇인가?
- 이 모든 데이터를 반드시 보관해야 하는가?

데이터 통합 및 라이프사이클 관리를 위한 IBM 솔루션

다음 단계: 클라우드 거버넌스 논의의 지속

IBM은 수십 년간 여러 소스의 엔터프라이즈급 데이터를 처리하는 솔루션을 제공해 왔으며, 우수한 신뢰성, 유연성 및 확장성으로 인정 받은 데이터 통합, 관리, 보안 및 분석 솔루션을 통해 업계를 선도해 왔습니다.

IBM InfoSphere Information Server는 기업이 데이터를 이해 및 관리하고, 데이터 품질을 생성 및 유지하고, 데이터를 변환 및 제공하도록 도와주는 시장을 선도하는 정보 통합 플랫폼입니다. InfoSphere Data Replication 및 InfoSphere Federation Server와 함께, InfoSphere Information Server는 다음과 같이 강력한 IIG 기능을 제공합니다.

- **데이터 통합:** 데이터를 스타일에 관계없이 변환하고 모든 시스템에 제공하여, 가치 실현 시간을 단축하고 IT의 위험을 줄여줍니다. 또한 이 패키지는 셀프 서비스 데이터 통합을 지원하는 InfoSphere Data Click 기능을 포함합니다.

- **데이터 복제:** 실시간 데이터 복제 요구사항을 지원하고, 복제된 데이터를 통합하여 모바일 애플리케이션, 비즈니스 분석 및 빅 데이터 프로젝트를 보강합니다.
- **정보 거버넌스 카탈로그:** 정보를 이해 및 관리하고, IT 자산을 검색하고 공통의 비즈니스 언어를 정의하기 위한 표준화된 방식을 마련하여 비즈니스와 IT 목표를 더 효과적으로 일치시킬 수 있게 지원합니다.
- **데이터 품질:** 더 효과적인 비즈니스 의사결정을 내리고 비즈니스 프로세스 실행을 개선하도록 지원하여 데이터를 분석, 정리, 모니터링 및 관리하고, 더 큰 가치를 추가할 수 있게 해줍니다.
- **데이터 페더레이션:** 데이터를 유연하게 제공하고 가상 데이터 허브 요구사항을 지원합니다.

개요: 혼돈 속 질서의 정립

- 네 가지 핵심

데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

- 기업이 필요로 하는 데이터는 어디에 있는가?
- 데이터를 조합하여 기업이 필요로 하는 정보를 생성하는 방법은 무엇인가?
- 이 모든 데이터를 반드시 보관해야 하는가?

데이터 통합 및 라이프사이클 관리를 위한 IBM 솔루션

다음 단계: 클라우드 거버넌스 논의의 지속

또한 InfoSphere IIG 기능은 가상화 및 검색, 하둡 기반 분석, 스트림 컴퓨팅, 데이터 웨어하우스 및 텍스트 분석을 위한 도구가 포함된 IBM 빅 데이터 플랫폼을 지원합니다.

IBM InfoSphere Optim™ 솔루션은 엔터프라이즈 요구사항에 맞게 확장 가능한 데이터 라이프사이클 관리 기능으로 IIG 기능을 보강합니다. InfoSphere Optim 솔루션을 단일 애플리케이션, 데이터 웨어하우스 또는 복잡한 데이터 환경 등 무엇을 위해 구현하든, 데이터 라이프사이클 관리를 위한 일관성 있는 전략을 사용할 수 있습니다. InfoSphere Optim의 고유한 관계 엔진은 데이터의 아카이빙, 하위 설정, 검색과 같은 데이터 처리 활동을 관리하는 단일 제어 지점을 제공합니다.

InfoSphere Optim 솔루션은 IIG 및 거버넌스 요구사항을 충족하도록 도와주는 동시에 멀티소스 하이브리드 데이터 환경으로 가중되는 문제를 해결하도록 도와줍니다. 대규모 데이터 웨어하우스 환경에서 오래된 데이터를 아카이빙하여 소중한 스토리지 용량을 확보함으로써 응답 시간을 향상시키고 비용을 절감할 수 있습니다.

개요: 혼돈 속 질서의 정립

- 네 가지 핵심

데이터를 정보로 변환: 통합 및 라이프사이클 전략 정의

- 기업이 필요로 하는 데이터는 어디에 있는가?
- 데이터를 조합하여 기업이 필요로 하는 정보를 생성하는 방법은 무엇인가?
- 이 모든 데이터를 반드시 보관해야 하는가?

데이터 통합 및 라이프사이클 관리를 위한 IBM 솔루션

[다음 단계: 클라우드 거버넌스 논의의 지속](#)

다음 단계: 클라우드 거버넌스 논의의 지속

클라우드 기반의 데이터 및 처리 서비스는 비즈니스 사용자가 간과할 수 없는 다양한 기회를 제공하고, IT는 내부, 사내의 트랜잭션 및 보고 시스템의 무결성을 유지 관리하는 일을 담당합니다. 하이브리드 환경을 위한 거버넌스 전략의 계획은 나중에 고려할 일이 아니라 지금 바로 착수해야 할 일입니다.

이 전자책은 성공적인 하이브리드 데이터 거버넌스의 네 가지 핵심 중 하나인 '하이브리드 환경의 데이터 통합 및 라이프사이클 관리'를 설명합니다. **다른 핵심 요소를 살펴보려면 이 시리즈의 다른 전자책을 다운로드하여 참조하십시오.**

- [The truth about information governance and the cloud](#)(정보 거버넌스 및 클라우드에 대한 진실)
- [Make sense of your data](#)(데이터의 이해)
- [Prepare and maintain your data](#)(데이터의 준비 및 유지 관리)
- [Securing data in the cloud and on the ground](#)(클라우드 및 사내 데이터 보안)

IBM 거버넌스 사고 리더십과 지원 기술에 대한 자세한 정보는 다음 사이트를 방문하십시오.

ibm.com/analytics/us/en/technology/agile/

또한 IBM 글로벌 파이낸싱은 기업이 비즈니스의 성장에 필요한 기술을 획득하도록 지원하기 위해 다양한 상환 옵션을 제공합니다. IBM은 IT 제품 및 서비스의 획득에서 처분에 이르기까지 라이프사이클 전반의 관리를 제공합니다. 자세한 정보는 다음 사이트를 방문하십시오.

ibm.com/financing/kr



© Copyright IBM Corporation 2016

Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
September 2016

IBM, IBM 로고, ibm.com 및 InfoSphere는 전세계 여러 국가에서 등록된 International Business Machines Corp.의 상표입니다. 기타 제품 및 서비스 이름은 IBM 또는 타사의 상표입니다. 현재 IBM 상표 목록은 웹 "저작권 및 상표 정보" (ibm.com/legal/copytrade.shtml)에 있습니다.

Netezza는 IBM Company인 IBM International Group B.V.의 상표 또는 등록 상표입니다.

Java 및 모든 Java 기반 상표와 로고는 Oracle 및/또는 그 계열사의 상표 또는 등록상표입니다.

이 문서는 최초 발행일을 기준으로 하며, 통지 없이 언제든지 변경될 수 있습니다. IBM이 영업하는 모든 국가에서 모든 오퍼링이 제공되는 것은 아닙니다.

이 문서의 정보는 상품성, 특정 목적에의 적합성에 대한 보증 및 타인의 권리 비침해에 대한 보증이나 조건을 포함하여(단, 이에 한하지 않음) 명시적이든 묵시적이든 일체의 보증 없이 "현상태대로" 제공됩니다. IBM 제품에 대한 보증은 제품의 준거 계약 조항에 의거하여 제공됩니다.

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



재활용하십시오.