



SEAGATE

백서

엔터프라이즈 데이터 오케스트레이션

데이터 전략 및 인프라의 핵심 요소

목차

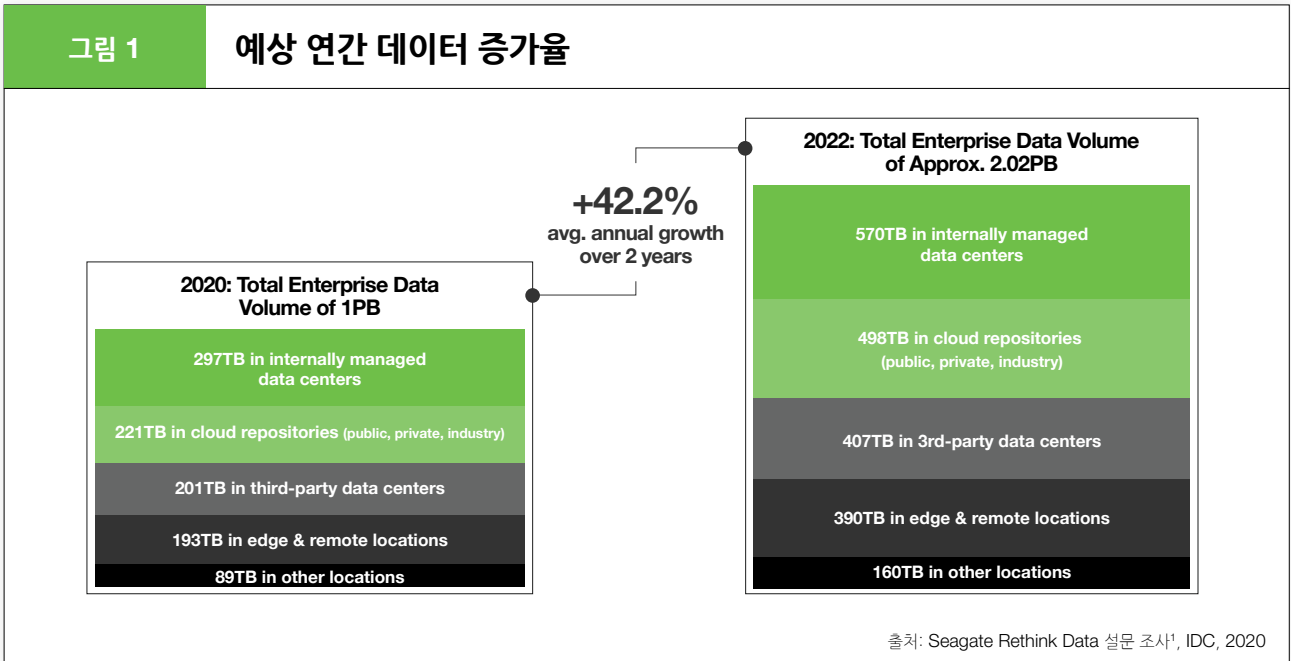
- 3 서론
- 5 데이터가 사용되는 방식:
데이터 수명 주기
- 6 엔터프라이즈 전체에서의
데이터 이동
- 8 대규모 데이터 오케스트레이션
요구 사항
- 12 결론



서론

엔터프라이즈 데이터의 관리가 그 어느 때보다 복잡해졌으며 이러한 복잡성은 계속 커지는 추세입니다. 이제 데이터는 물리적 자본이나 지적 재산과 유사한 필수 자산이 되었습니다. 데이터의 양과 중요성이 증가함에 따라 데이터 관리의 복잡성도 증가합니다. 특히 데이터가 엔드 포인트에서 엣지, 코어 및 클라우드 데이터 센터에 이르기까지 모든 곳에 분산되어 있는 경우 더욱 그렇습니다. 데이터가 어디에나 분산되면서, 현대의 분산된 엔터프라이즈에는 데이터 이동 및 오케스트레이션을 위한 새로운 방법이 필요합니다.

IDC(International Data Corporation)의 조사 및 분석이 담긴 Seagate 보고서인 Rethink Data에 따르면 엔터프라이즈 데이터는 현재 무수한 출처로부터 전례 없는 42.2%의 연간 성장률로 급증하고 있습니다. 이처럼 증가하는 데이터는 여러 수준의 공용, 개인 및 하이브리드 클라우드(멀티 클라우드 아키텍처)를 비롯하여 다양한 IT 구성을 통해 확산되고 있습니다. 점점 더 다양해지는 생태계에서 점점 더 복잡해지는 데이터 이동은 비즈니스 소유자의 데이터 관리 문제를 어렵게 만듭니다.



엔터프라이즈에서 데이터를 이동해야 하는 빈도 또한 증가하면서 문제가 가중됩니다. IDC는 엔터프라이즈가 엣지에서 코어로 주기적으로 전송하는 데이터양이 향후 2년 이내에 36%에서 57%로 증가할 것이라고 예측합니다. 한편 수집되는 즉시 엣지에서 코어로 즉시 전송되는 데이터는 8%에서 16%로 2배 증가할 것입니다. 이는 엔터프라이즈가 이전보다 더 많은 데이터를 관리해야 함을 의미합니다.



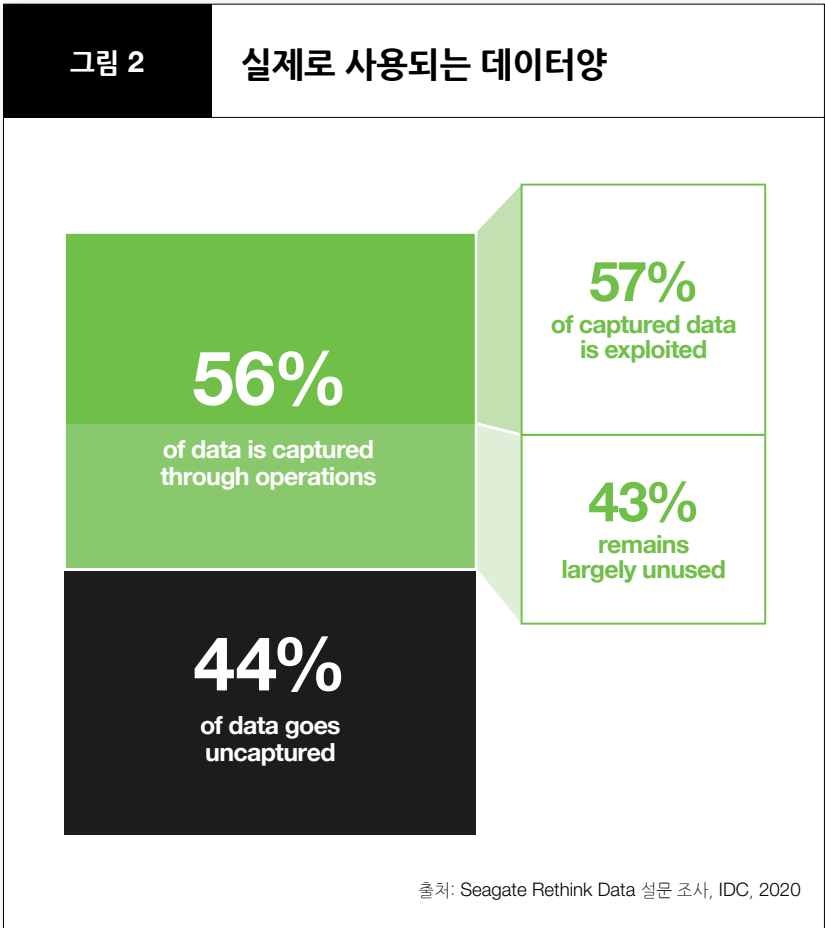
마찬가지로 중요한 것은 모든 종류의 데이터가 이제는 과거에 비해 훨씬 더 큰 잠재적 가치를 지니고 있다는 사실입니다. 이는 고급 분석, 인공지능(AI), 머신 러닝(ML)이 혁신적인 방식으로 데이터의 가치를 발견하고 얻으며 개발할 수 있기 때문입니다. 따라서 각 엔터프라이즈는 모든 데이터를 유지, 추적 및 사용하는 것이 중요합니다. 그러나 Rethink Data 보고서의 IDC의 연구에 따르면 현재 엔터프라이즈는 사용 가능한 데이터의 32%만 활용하고 나머지 68%는 활용되지 못한 상태입니다.

더욱 정교해져야 하는 데이터 오케스트레이션

대부분의 엔터프라이즈 IT 의사결정권자는 데이터 생성자와 데이터 소비자를 연결하는 제어 능력인 DataOps를 구현해야 한다는 것을 잘 알고 있습니다. 보고서에 따르면 현재 조직의 약 10%만이 엔터프라이즈 전체에 DataOps를 완전히 구현했습니다. 그러나 IDC는 다른 데이터 관리 솔루션과 함께 DataOps를 배포하면 고객 충성도, 매출, 수익 등 훨씬 더 나은 비즈니스 성과를 얻을 수 있다는 사실을 밝혔습니다.

경영진과 다른 의사결정자는 데이터 수명 주기의 모든 차원과 데이터를 효과적으로 활용하는 조직의 능력에 미치는 영향을 고려해야 합니다. 특히 조직은 데이터 사용 방법, 데이터 이동 방법, 데이터를 대규모로 관리하는 방법을 이해하여 데이터를 사용할 수 있게 만들고 보호하며 분석 및 활용할 수 있도록 만들어야 합니다.

이러한 요소를 적절히 해결하지 못하면 가장 중요한 리소스 중 하나를 활용하는 조직의 능력에 부정적인 영향을 미칠 수 있습니다.



데이터가 사용되는 방식: 데이터 수명 주기

데이터는 데이터로 수행되는 일이 시간이 지남에 따라 변한다는 의미에서 수명 주기가 있습니다. 또한 데이터는 원점에서 초기 수집 플랫폼으로 그리고 스토리지와 분석 시스템을 따라 움직이면서 이동합니다. 데이터의 전체 수명 주기와 해당 데이터를 위해 작동하는 다양한 종류의 컴퓨팅, 스토리지 및 네트워크 인프라 전체에서 데이터에 대한 보호가 필요합니다. 이는 4단계에 걸쳐 발생합니다.

출처 단계

수명 주기의 출처 단계는 데이터가 생성되는 시점입니다. 구별되는 특성을 가진 여러 가지 다양한 데이터 출처가 있으며 각 출처를 수용해야 합니다. 예를 들어, 제조 장비 또는 자율 주행 차량에 부착된 센서는 엄청난 양의 데이터를 생성할 수 있습니다. 이런 데이터를 모두 중앙 집중식 처리 시스템으로 전송하는 것은 실용적이지 않을 수 있습니다. 이러한 엔드포인트에서 수집된 데이터는 로컬 또는 엣지에서 더 적절히 처리될 수 있으며, 초기 처리 결과는 네트워크를 통해 또는 데이터양이 방대하고 속도가 중요한 경우 물리적 스토리지 장치를 통해 중앙 집중식 시스템인 코어로 전송될 수 있습니다.

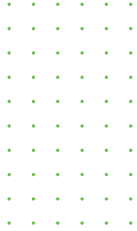
또는 모든 데이터를 수집해야 하는 사용 사례가 있을 수 있습니다. 이러한 상황에서는 데이터를 확장 가능하고 대기 시간이 짧은 데이터 저장소로 전송할 수 있습니다. 모든 데이터를 수집해야 하지만 데이터에 즉시 액세스할 필요가 없는 애플리케이션은 개인 또는 공용 클라우드에 일괄 업로드하여 더 나은 서비스를 제공할 수 있습니다.

수집 단계

수집 단계는 데이터를 확보하여 조직의 통제 하에 가져올 때 발생합니다. 이 단계에서 눈에 띄는 문제는 데이터양, 수집 속도, 데이터 보안, 검증 절차입니다. 경우에 따라 데이터가 로컬로 캐시된 후 일괄 전송될 수 있습니다. 예를 들어, 시계열 데이터 스트림은 최대한 빨리 데이터 스트림의 이상을 감지하기 위해 로컬에서 분석될 수 있습니다. 그런 다음 데이터를 스트림이나 일괄 전송을 통해 영구 스토리지 시스템으로 전송할 수 있습니다. 또는, 예비 검증 및 분석을 수행하여 도출된 결과를 전송하는 ML 모델이 엣지에 배포될 수 있는 경우도 있습니다.

분석 및 통합 단계

분석 및 통합 단계에서는 데이터가 비즈니스 운영의 일부로 사용됩니다. 이는 트랜잭션 실행에서 ML 알고리즘 교육에 이르는 모든 과정을 포함합니다. 과거에는 대용량 데이터 세트가 제한적으로 사용되었지만 비즈니스에 ML을 적용함으로써 바뀌었습니다. ML 모델의 품질은 알고리즘과 데이터의 조합에 따라 달라집니다. 대부분의 경우 더 많은 데이터를 사용하여 모델을 교육시킴으로써 성능을 향상시킬 수 있습니다. 인간이 더 많은 경험을 통해 예외와 특이 사항에 대해 배울 수 있는 것처럼, ML 모델은 모델의 예측에 영향을 미치는 드문 패턴을 식별할 수 있습니다. 데이터 관리에서 불가능하지는 않더라도 가장 어려운 과제 중 하나는 미래에 유용할 데이터를 결정하는 것입니다. 지금은 겉보기에 가치가 없는 데이터가 어느 시점에는 가치가 있을 수 있습니다. 예를 들어, 자율 주행 차량은 방향을 전환할 때 노란색 셔츠를 입은 보행자를 식별하는 데 문제가 있었습니다. ML 엔지니어는 기존 데이터 세트의 메타데이터 태그가 지정된 이미지를 사용하여 이러한 경우의 예를 더 많이 찾아 모델 교육을 심화하는 데 사용했습니다.



미래에 어떤 데이터가 유용할지 알 수 없고 데이터가 ML 모델의 형태로 중요한 지적 재산을 생성하는 데 사용되기 때문에 조직은 사실상 모든 데이터를 저장해야 할 당위성이 있습니다.

보관 단계

데이터 수명 주기의 네 번째이자 마지막 단계는 보관입니다. 이 단계에서는 데이터 센터에서 클라우드에 이르는 인프라를 포괄하는 오케스트레이션 플랫폼을 통해 관리해야 합니다. 과거에는 최상의 장치 및 스토리지 소프트웨어 모음이 최고의 솔루션이었을 수 있지만 이제 더 이상 그렇지 않습니다. 이러한 앙상블 솔루션의 구성 요소는 수명 주기의 일부를 로컬로 최적화하는 경향이 있지만 이는 최적의 글로벌 솔루션으로 이어집니다.

엔터프라이즈 전체에서의 데이터 이동

조직은 데이터가 기기 간에 어떻게 이동하는지 검토하여 모든 인프라 전체에서 오케스트레이션을 처리하는 솔루션을 배포해야 합니다. 고려해야 할 몇 가지 스토리지 인프라 클래스가 있습니다.

엔드포인트에서는 예비 분석을 비롯하여 조직이 관리해야 하는 대부분의 데이터가 생산됩니다. 엔드포인트에는 모바일 장치, 사물 인터넷(IoT) 센서 및 자율 장치를 비롯한 다양한 구성 요소가 포함됩니다.

즉, 엣지 장치에서 생성된 모든 데이터가 데이터 센터나 클라우드로 직접 전송되는 것은 아닙니다. 일부 필터링, 집계 및 예비 분석은 데이터 생성 지점에 더 가까운 곳에서 발생할 수 있습니다. 이 방식을 엣지 컴퓨팅이라고 합니다.

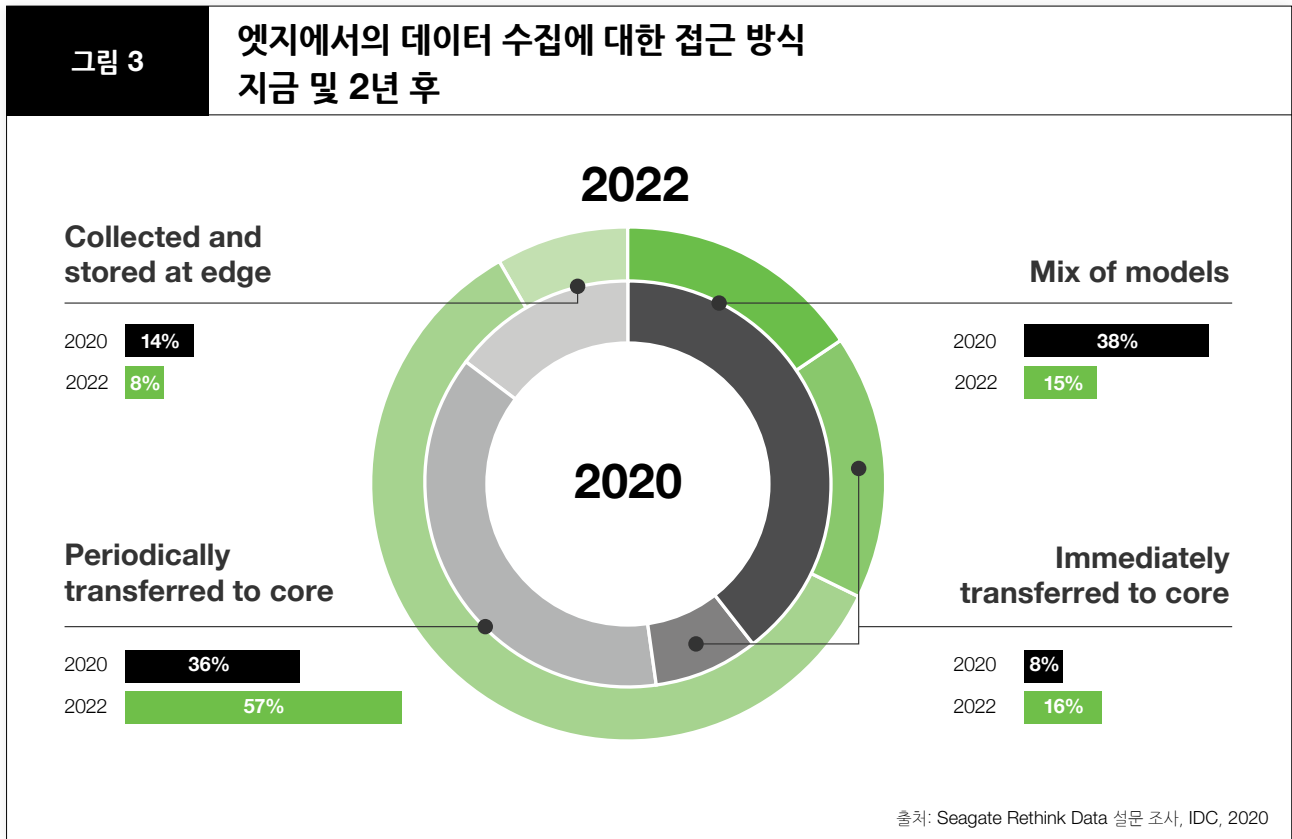
엣지 컴퓨팅은 즉각적인 데이터 분석을 허용하는 이점을 제공하지만 추가적인 관리 책임이 따릅니다. 로컬 처리는 애플리케이션이 실시간 의사 결정을 요구할 때 유용합니다. 그러나 심층 분석, AI 및 ML 방식을 위해서는 여전히 코어에서 대규모로 모델 교육이 진행되어야 합니다. 따라서 데이터가 필요에 따라 활용되고 더 효과적으로 활용될 수 있을 만큼 이해되도록 데이터 관리자는 많은 데이터를 코어로 이동하여 분석 모델을 계속 개선해야 합니다. 따라서 스토리지 관리자는 엣지에서 데이터를 보호하고 이동하는 가장 간단하고 효율적인 방법을 고려해야 합니다.

상대적으로 적은 양의 데이터는 5G 네트워크를 사용하여 효율적으로 전송할 수 있지만 대용량 데이터의 경우 비용이 많이 듭니다. 유선 네트워크는 비용은 저렴하고 용량은 크지만 이를 사용할 수 없는 엔드포인트나 엣지 장치도 있습니다.



데이터 센터는 컴퓨팅 및 스토리지 서비스의 허브이며 조직의 스토리지 인프라 대부분이 위치한 장소입니다. 이러한 데이터 센터는 공용 및 개인 클라우드 스토리지로 보완되면서 증가하고 있습니다. 온프레미스 및 클라우드 리소스의 조합은 하이브리드 클라우드라고 하며 특히 스토리지 분야에서 점차 널리 활용되고 있습니다.

데이터는 데이터 센터 또는 공용 클라우드에 있는 수집 파이프라인 및 분석 작업 흐름을 통해 엣지로부터 엔터프라이즈 인프라 전반에 걸쳐 다양한 방식으로 이동합니다. 이 흐름을 관리하려면 엣지에서 클라우드에 이르는 인프라를 포괄하는 오케스트레이션 소프트웨어가 필요합니다.



대규모 데이터 오케스트레이션 요구 사항

데이터 오케스트레이션의 첫 번째 요구 사항 중 하나는 데이터 오케스트레이션이 작동해야 하는 규모를 인식하는 것입니다.

엔터프라이즈에서는 스토리지 시스템 간에 페타바이트 규모의 데이터를 효율적으로 이동할 방법이 필요합니다. 네트워크를 사용하여 이 규모의 데이터를 이동하려고 한다면 장기간 모든 처리량을 소모해 버리고 말 것입니다. 100Mbps 업로드 속도로 20TB의 데이터만 이동해도 20일 이상이 소요됩니다. 기가비트 속도의 광섬유 서비스를 사용하더라도 평균 지속 업로드 속도가 800Mbps라고 가정하면 20TB를 업로드하는 데 2.5일이 걸리고 50TB를 업로드하는 데는 거의 일주일도 걸립니다. 네트워크는 테라바이트 또는 페타바이트 규모의 데이터 세트를 정기적으로 이동할 만큼 빠르거나 비용 효율적이지 않습니다.

데이터 이동을 가능하게 하려면 네트워크를 보완할 추가 인프라가 있어야 합니다. 이 보완 인프라에는 데이터 센터로 물리적으로 전송되기 전에 데이터를 신속하게 로드할 수 있는 물리적인 데이터 셔플과 같은 저장 장치가 포함되어야 합니다. 가변 경로와 프로세스를 사용하여 데이터를 이동하려면 결과적으로 복잡한 데이터 흐름을 구성할 수 있는 오케스트레이션 계층이 필요합니다. 현재 위치에서 필요한 위치로 데이터를 가져오려면 정책 기반 관리가 필요합니다. 개별 데이터 세트 또는 작업 부하마다 지시하여 데이터 흐름을 관리하는 것은 사실상 불가능합니다. 대신, 해당 데이터의 속성에 따라 데이터를 이동하고 저장하는 방법을 지정하는 정책을 정의해야 합니다.

정책 기반 데이터 관리 정의

정책은 데이터를 이동하고 저장하는 방법에 대한 몇 가지 사항을 지정합니다.

예를 들어, 이동되는 데이터는 대부분의 경우 암호화해야 합니다. 이를 위해서는 특정 암호화 알고리즘 및 암호화 키를 사용해야 합니다. 암호화 키에는 일반적으로 키 관리 서비스에서 관리하는 자체 수명 주기가 있습니다. 미사용 데이터에도 암호화가 필요한 경우가 많으며 데이터 관리 정책 내에서 추가 사양이 필요합니다.

데이터 세트마다 대기 시간 요구 사항이 다릅니다. 일부 데이터는 대기 시간이 짧으므로 생성 직후 분석해야 합니다. 예를 들어, 이상 징후가 있는지 모니터링하는 IoT 센서 데이터는 최대한 빨리 분석해서 잠재적인 문제를 감지해야 합니다. 신용카드 거래도 사기 위험을 예방하기 위해 즉각적인 분석이 필요한 또 다른 예입니다. 그러나 이런 유형의 데이터에 대한 관리는 여기서 끝나지 않습니다. 이러한 즉각적인 분석 후, 해당 소스의 데이터는, 예를 들어, 트랜잭션 상호 관계 및 장기 추세 분석과 같은 심층적인 이해를 개발하는 데 상당한 가치를 갖습니다. 따라서 정책 기반 오케스트레이션은 모든 데이터에 대한 후속 관리 단계를 고려해야 합니다.

물론 많은 비즈니스 애플리케이션은 센서 판독 값이나 신용카드 거래만큼 시간에 민감하지 않습니다. 생성 후 최대 24시간 안에만 도착하면 여전히 완전한 가치를 제공하는 데이터도 있습니다. 소매점은 한밤중에 재고 데이터를 일괄 업로드하여 향후 며칠 동안 재입고될 물류를 확인할 수 있습니다.

데이터 오케스트레이션과 관련하여 고려해야 할 또 다른 요소는 데이터의 양입니다. 예를 들어, 센서 데이터의 양은 센서 수와 각 센서가 측정값을 전송하는 빈도의 함수입니다. 데이터 증가율은 센서가 배포된 비율에 따라 결정됩니다. 그러나 재고 사례에서 데이터의 양은 재고의 제품 수에 의해 결정되고 증가율은 새로운 제품이 재고에 추가되는 속도에 의해 결정됩니다.

요약하자면, 데이터 오케스트레이션 정책을 개발할 때는 대기 시간, 예상 데이터양 및 보안 제어에 대한 요구 사항을 데이터 이동 방법에 매핑해야 합니다.

또한, 정책은 데이터 세트 메타데이터에 따라 달라집니다. 메타데이터는 데이터 수명 주기의 다양한 단계를 통해 통합, 선택 및 필터링을 지원하려는 태그를 사용하여 구현되는 경우가 많습니다.

초기 저장 및 처리

많은 조직이 기존에 보유하고 있는 방대한 양의 데이터와 앞으로 증가할 것이라 예상되는 데이터의 양을 고려할 때, 데이터 저장 및 이동 비용을 낮추는 것은 분명 이점을 가져옵니다. 이러한 비용은 데이터 손실의 기회비용과도 균형을 이루어야 합니다. 더 많은 엔터프라이즈가 데이터 과학 기술과 ML 모델을 사용함에 따라 이전에는 분명히 사용되지 않았을 데이터를 활용하는 새로운 방법이 발견됩니다.

고도로 분산된 환경에서 일부 데이터는 5G 네트워크를 비롯한 타사 네트워크를 통해 이동합니다. 이러한 경우 비용과 보안 모두를 문제로 고려해야 합니다. 타사 네트워크는 짧은 대기 시간 요구 사항을 충족하기 위한 최선의 선택일 수 있지만 비용은 다른 전송 방법보다 많이 듭니다. 데이터에 대한 즉각적인 액세스의 잠재적 비즈니스 가치는 대기 시간은 낮지만 전송 비용은 더 많이 들게 하는 비용보다 커야 합니다.

물론 데이터가 처리 지점에 도착했다고 해서 이동이 끝났다는 의미는 아닙니다. 데이터는 클라우드로 이동하기 전에 데이터 센터에 도착할 수 있습니다. 한 클라우드 내의 데이터는 다른 클라우드로 이동하거나 데이터 센터로 다시 이동할 수 있습니다. 이 상황에서 엔터프라이즈가 직면한 과제 중 하나는 공용 클라우드 환경에서의 데이터 이동을 위한 표준화된 도구가 없다는 것입니다. 테라바이트 규모의 데이터를 이동하는 클라우드 종립적인 방법은 없습니다. 클라우드 공급업체마다 클라우드 간에 데이터를 옮기는 독점 방법이 있습니다. 이상적으로 엔터프라이즈는 멀티 클라우드 환경 내에서 작동하고 해당 플랫폼 간에 데이터를 여러 방향으로 옮길 수 있는 데이터 오케스트레이션 시스템을 배포해야 합니다.

대규모 데이터 분석

ML 실무자들은 모델의 품질이 알고리즘과 데이터의 조합이라는 사실을 잘 알고 있습니다. 알고리즘은 잘 알려져 있고 잘 파악하고 있기 때문에 ML 시스템을 사용하여 경쟁할 때 차별 요소가 될 수 있는 것은 데이터입니다. 실제로 ML의 또 다른 잘 알려진 사실은 어떤 데이터가 나중에 유용하게 될 것인지 알지 못하는 경우도 있다는 것입니다.



테스트 결과, 때때로 ML 모델이 특정 조건에서 제대로 작동하지 않는 것으로 나타났습니다. 이 문제는 성능이 저조한 영역에 구체적으로 초점을 맞춘 모델에 대한 추가 교육을 제공하여 수정할 수 있습니다. 이 경우에는 설명 태그로 잘 선별된 데이터에 액세스하면 ML 및 데이터 엔지니어는 추가 교육을 위한 관련 데이터를 확인할 수 있습니다.

효과적인 ML 모델을 개발하는 것만큼 효과적인 분석에 중요한 것은 어떤 데이터이며 데이터양이 얼마나 되는지입니다. 상호 관련된 데이터 세트의 많은 부분이 손실되면 머신 러닝 모델에 데이터를 효과적으로 배포할 수 없습니다. 따라서 잠재적으로 가치 있는 데이터가 손실되지 않도록 하는 것은 데이터에서 지적 재산을 개발하기 위한 장기 전략의 핵심적인 부분입니다.

데이터 손실은 다양한 형태로 발생할 수 있습니다. 일정 기간이 지나 단순히 데이터를 삭제할 수 있습니다. 이는 스토리지가 부족하고 비용이 많이 들 때는 의미가 있었지만 더 이상 그렇지 않습니다. 5G 네트워크를 통한 데이터 전송 비용에 대한 우려로 인해 엔터프라이즈는 데이터 센터 또는 클라우드로 데이터를 보내기 전에 엣지 장치에서 데이터를 집계하려고 시도할 수 있습니다. 그러나 집계는 중요한 실시간 가치를 제공하는 세부 데이터를 분석하는 기능의 대기 시간을 높일 수 있습니다. 제조 기계의 성능에 대한 시계열 데이터를 생각해 봅시다. 5초마다 보고되는 데이터에서는 명백했던 이상 현상도 엣지에서 분 단위로 데이터를 집계할 때는 누락될 수 있습니다.

조직이 데이터를 사용하는 새로운 방법을 발견함에 따라 시간이 지나면서 데이터의 가치는 증가할 것입니다. 여러 면에서 데이터는 비즈니스 가치의 통화이며 특히 데이터 보존 비용이 하락할 때 보존되어야 합니다.

소프트웨어가 주도하는 오케스트레이션

스토리지 인프라의 소프트웨어 계층은 효율적이고 효과적인 데이터 조정을 위해 매우 중요합니다. 소프트웨어는 데이터 보호 보장, 운영상의 오버헤드 최소화, 비용 최적화, 정책 기반 관리 활성화와 같은 데이터 관리의 중요한 측면을 지원합니다.

모범 사례는 소프트웨어 정의 스토리지와 함께 오픈 소스 소프트웨어 및 상용 스토리지 시스템을 활용하여 비용을 최적화하는 것입니다. 이 접근 방식은 공급업체에 종속되는 것을 완화할 수 있습니다.

또 다른 모범 사례는 데이터를 폐기하지 않고 저장하는 것입니다. 스토리지 비용을 줄이고 양질의 효과적인 모델을 구축하는 조직의 능력을 제한하는 것보다 분석 및 ML에 사용할 수 있는 대용량 데이터 세트를 보유하는 것이 더 낫습니다.

데이터 오케스트레이션 소프트웨어는 데이터 및 기타 메타데이터의 출처를 추적하여 활성화되는 정책을 기반으로 관리함으로써 비용을 최적화합니다. 또한 잘 선별된 데이터를 통해 조직은 AI를 사용하여 데이터를 수집, 이동, 저장 및 처리하는 방법을 결정할 수 있습니다. 데이터의 가치와 기능을 인식하는 자동화된 AI 서비스는 데이터가 스토리지 인프라를 통해 이동함에 따라 정책을 적용할 수 있습니다. 정책은 데이터 프라이프라인 처리와 통합될 수 있습니다. 예를 들어, 데이터 관리자는 물리적으로 전송된 셔플 장치를 통해 가치가 낮은 데이터를 이동하는 한편 네트워크를 통해 소규모 또는 가치가 높은 데이터 세트를 이동하는 정책을 설정할 수 있습니다. 스토리지 관리자는 네트워크 관리를 위해 VPN이나 물리적 연결이 필요하지 않은 스토리지 관리 애플리케이션을 활용할 수도 있습니다.



데이터 오케스트레이션 구성 시 보안 고려 사항

데이터 오케스트레이션 시스템을 구축하는 경우 광범위한 데이터 개인 정보 보호 및 보안 요구 사항을 고려해야 합니다.

데이터 사용, 권한, 공유, 개인 정보 보호 및 보안에 대한 규정 준수 요구 사항은 산업 및 데이터 특성에 따라 다릅니다. 이로 인해 비용 효율적인 방식으로 요구 사항을 충족하기 위한 추가 문제가 발생하며, 데이터 오케스트레이션이 도움이 될 수 있습니다.

엔터프라이즈는 데이터의 지문 또는 메시지 다이제스트를 사용하여 데이터 무결성을 추적할 수 있습니다. 이렇게 하면 데이터가 변조되지 않도록 할 수 있습니다. 지문이 예상과 일치하지 않는 경우 데이터 관리자가 파일을 격리하여 검토한 후 데이터를 수락하거나 거부할 수 있습니다.

또한 데이터 접근 권한을 누구에게 부여할 것인지도 고려해야 합니다. 특정 데이터가 경쟁 우위를 제공하는 경우, 특정 클라우드 서비스 공급자가 경쟁 갈등을 유발할 수 있는지 고려합니다.

규정 준수의 추가 구성 요소에는 블록체인 추적 및 출처 추적이 포함됩니다. 데이터 보안은 데이터의 전체 수명 주기를 포괄하는 책임입니다. 이를 위해서는 데이터 오케스트레이션 시스템 내에 신뢰 루트를 설정해야 합니다. 데이터 출처 정보는 일관되게 수집, 저장 및 관리되어야 합니다. 데이터 무결성과 신뢰성을 보장하는 동시에 민감 정보 또는 기밀 정보의 개인 정보를 보호하기 위한 메커니즘이 마련되어야 합니다.

결론

그 어느 때보다 데이터가 분산되어 있으며, 더 많은 데이터가 경쟁 우위의 원천입니다. 따라서 조직이 이 기회를 활용하는 것이 중요합니다.

데이터 관리의 목표는 데이터에 대한 총체적 보기를 제공하고 이동 중인 데이터이든 미사용 데이터이든 사용자가 데이터에 액세스하여 최적의 가치를 얻을 수 있도록 하는 것입니다.

성공적인 데이터 오케스트레이션 전략을 배포하는 가장 중요한 단계는 엔터프라이즈의 IT 인프라를 개발하는 방법을 조사하고 고려하는 것입니다. 이 인프라는 확장성 외에도 분산 스토리지 아키텍처 및 가장 필요한 곳에 데이터를 제공하는 풍부한 소프트웨어 기반 제어 계층을 사용하여 데이터를 효율적이고 신속하게 수집, 이동, 분석 및 활용할 수 있도록 해야 합니다.

궁극적으로 능동적이고 효율적인 DataOps 프로그램은 AI 모델을 구축 및 교육하고 대규모 분석을 배포하기 위한 기반으로 효과적인 데이터 오케스트레이션에 의존합니다. 결국 이러한 고급 분석 결과는 경쟁 우위로 이어질 수 있습니다.

자세히 알아볼 준비가 되셨습니까?

seagate.com에서 확인하십시오.



Seagate 개인 클라우드 솔루션

www.seagate.com/solutions/cloud/private-cloud/

Lyve Rack – 개방적이고 경제적이며 내구성이 우수한 오브젝트 스토리지 솔루션

www.seagate.com/products/storage/object-storage-solutions/lyve-drive-rack/

CORTX – 오픈 소스 대용량 오브젝트 스토리지

www.seagate.com/products/storage/object-storage-software/

© 2020 Seagate Technology LLC. All rights reserved. Seagate, Seagate Technology 및 Spiral 로고는 미국 및/또는 기타 국가에서 Seagate Technology LLC의 등록 상표입니다. 기타 모든 상표 또는 등록 상표는 해당 소유자의 재산입니다. 드라이브 용량과 관련하여 1기가바이트(GB)는 10억 바이트이며 1테라바이트(TB)는 1조 바이트입니다. 사용 중인 컴퓨터의 운영 체제에 다른 측정 기준이 적용되는 경우 이보다 낮은 용량을 보고할 수도 있습니다. 또한 나열된 용량의 일부는 포맷 및 기타 기능을 위해 사용되는 공간이므로 데이터 저장에 사용될 수 없습니다. Seagate는 별도의 통지 없이 제품의 품목 또는 사양을 변경할 수 있습니다. TP718.1-2011KR, 2020년 11월

