백서

빅데이터 보호를 위한 차세대 백업 및 복구 방안

모던 워크로드에 최적화된 데이터 보호 솔루션

산발적으로 이루어지는 실험 차원의 프로젝트였던 빅데이터가 불과 몇 년 사이에 디지털 기업의 미션 크리티컬 요소로 발전하면서 그 중요성도 점차 커지고 있습니다. IDC에 따르면, 2020년에는 관련 데이터를 분석하고 실행 가능한 정보를 제공하는 능력을 갖춘 기업이 상대적으로 분석을 중시하지 않는 기업에 비해 4,300억 달러 상당의 수익을 더 거둘 것으로 예상됩니다.[1] 간헐적으로 진행되던 빅데이터 분석이 현재는 Amazon, Walmart, UPS 등 많은 기업에서 매일같이 수행되고 있습니다.

빅데이터 백업 과제

빅데이터가 최고의 정보 자산으로 자리매김하면서 그에 걸맞는 관리가 필요합니다. 한때 빅데이터 볼륨에 따라 고려 대상에 불과하던 백업이 지금은 신속하고 확실하게 완수해야 하는 필수 과제가 되었습니다. IT 리더들은 이에 촉각을 곤두세우고 있습니다. 실제로 최근 ESG 설문 조사 결과, 2018년 데이터 센터 현대화를 위해 우선적으로 투자할 5대 분야를 묻는 질문에서 응답자의 31%가 데이터 백업 및 복구 개선을 꼽았습니다. 점차 더 많은 기업이 빅데이터 전담 백업 및 복구 리소스를 배치하고 있습니다. IDC는 2017년 기준 1,508억 달러인 빅데이터 시장의 규모가 2020년에는 2,100억 달러에 이를 것으로 전망했습니다.[2]

데이터를 효과적으로 백업하려면 많은 기업이 운영 면에서 대대적인 변화를 거쳐야 합니다. 기존 워크로드에 적합하던 레거시 아키텍처는 Hadoop과 같은 빅데이터 워크로드 실행 환경을 보호하기 위한 모던 아키텍처로 대체해야 합니다. 예를 들면 다음과 같습니다.

- **현재 많은 기업이 구현한 클라이언트-서버 백업 아키텍처**는 기존 워크로드에 효과적이지만 방대한 양의 빅데이터 처리나 비정형 데이터의 지속적인 분석에는 적합하지 않습니다. 이러한 데이터(및 해당 시점)는 기존 CRM 또는 ERP 시스템에 저장된 대량의 트랜잭션 데이터 못지않게 중요합니다.

- **동일한 데이터 세트에 대해 여러 카피본을 백업하면** 스토리지 리소스가 낭비됩니다. 빅데이터 환경에 효과적인 백업 아키텍처라면 불필요한 데이터 세트 복제를 없애 스토리지 리소스를 절약하고 데이터 처리 속도를 높이며 비용을 줄여야 합니다.

- **데이터 무결성은 매우 중요합니다.** 미션 크리티컬 데이터 분석을 바탕으로 비즈니스 의사 결정이 수행되기 때문입니다. 일반적인 Hadoop 빅데이터 환경에서는 사람의 실수로 데이터가 손상되면 부정확한 정보가 복제됩니다.

- **일일 분석 프로세스가 필요한 경우** SLA(Service-Level Agreement)에 백업 작업을 신속하게 수행하도록 규정되어 있습니다. 더구나 컴플라이언스 요건에 따라 신속한 데이터 검색이 이루어져야 합니다.

- **오래된 백업 아키텍처**는 빅데이터 트래픽을 지연시킵니다. 예를 들어 복제된 데이터 블록이 여러 데이터 노드에서 Hadoop NameNode로 전송된다면 병목 현상이 발생하여 빅데이터 스트림이 느려질 것입니다.

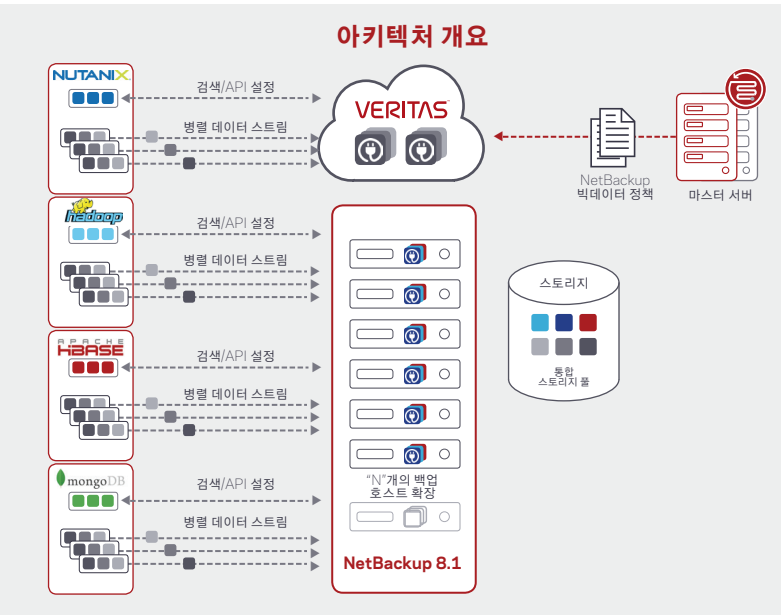
1 DC FutureScape: Worldwide Big Data, Business Analytics, and Cognitive Software Predictions, International Data Corporation, 2016년

2 Worldwide Semiannual Big Data and Analytics Spending Guide, IDC, 2017년

VERITAS™
The truth in information.

IDG
STRATEGIC MARKETING SERVICES

▶ **많은 기업이 Hadoop 환경을 백업하기 위해** 마련한 방편은 시간 소모적이고 여러 단계를 거쳐야 하는데, 여기에는 보조 Hadoop 클러스터 및 별도의 스토리지를 추가하는 것도 포함됩니다. 이러한 조치의 복잡성으로 인해 각 단계에서 데이터가 유출되거나 복원 시간 목표를 달성하지 못할 위험이 있습니다.



베리타스 솔루션: NetBackup 8.1

Veritas NetBackup 8.1은 빅데이터 백업의 요구 사항을 해결하도록 설계되었습니다. 여기에 구현된 Veritas NetBackup Parallel Streaming Framework는 Hadoop 환경의 대규모 스케일아웃 백업 및 멀티노드 클러스터 워크로드를 지원합니다. 비에이전트 NetBackup 8.1은 기본적으로 Hadoop Distributed File System(HDFS) 백업을 지원하므로 복잡한 방법에 의존하지 않아도 됩니다. 비에이전트 아키텍처는 현대화된 환경에서 중요한 개념입니다. 클러스터 노드에서 에이전트에 의한 공간 점유가 없고 각각의 에이전트를 관리해야 하는 부담이 사라져 시간과 비용이 절약됩니다. 그리고 Hadoop 업그레이드에 대해 걱정할 필요가 없습니다.

이 솔루션의 작동 방식을 설명하자면, 다운로드 가능한 NetBackup 8.1용 플러그인이 Hadoop 클러스터를 검사하여 메타데이터를 찾고 워크로드를 분할한 다음 여러 백업 호스트에 분산시키며 NetBackup 미디어 서버 또는 클라이언트의 리소스로 풀을 구성합니다. 이때 각 Hadoop 노드가 백업 호스트에 데이터를 보낼 때 병렬 스트리밍을 위한 백업 및 복구 성능을 극대화하도록 작업 리소스를 관리합니다. 또한 빅데이터 정책이 NetBackup 사용자 인터페이스에 완전히 통합되어 있습니다. 플러그인을 사용하므로 다음 Hadoop 릴리스가 나올 때까지 기다릴 필요가 없습니다.

베리타스 솔루션은 Hadoop 빅데이터 아키텍처와 다릅니다. 즉 각 노드에서 데이터의 카피본 3개를 복제한 다음 NameNode를 거쳐 백업 호스트를 통해 스토리지에 모두 전송하느라 스토리지를 낭비하고 스토리지 비용을 증가시킬 염려가 없습니다. NetBackup 8.1은 고유한 데이터만 수집하고 저장하므로 스토리지 공간 사용량이 2/3가량 줄고 보조 스토리지의 비용도 67% 절감됩니다. 뿐만 아니라 백업 성능이 3배 향상됩니다.

Hadoop은 데이터 복제를 수행하지만 포인트 인 타임 데이터 보호 기능이 없으므로 사람의 실수 및 데이터 손상으로부터 보호할 수 없습니다. 반면 NetBackup 8.1은 포인트 인 타임 백업으로 중요 데이터를 보존하고 SLA 및 컴플라이언스 요건을 충족하면서 그러한 허점을 해결합니다.

이 베리타스 솔루션은 (Hadoop 클러스터에 에이전트가 상주하지 않는) 비에이전트식 NetBackup 플러그인을 사용하므로 클러스터의 분석 성능이 백업 및 복구 프로세스에 좌우되지 않습니다. 또한 에이전트와 관련된 관리상의 어려움, 이를테면 설치, 관리, 업그레이드 문제가 사라집니다.

확장성은 빅데이터 환경에 중요하므로 팽창하는 Hadoop 데이터 볼륨의 요구 사항에 맞게 베리타스 솔루션을 증강할 수 있습니다. Hadoop 클러스터에 노드가 추가될 때 NetBackup 노드도 추가하면 됩니다.

NetBackup 8.1은 Hadoop뿐만 아니라 Nutanix 하이퍼컨버지드 인프라스트럭처, Apache HBase, MongoDB와 같은 여러 환경도 지원합니다. NetBackup 8.1은 통합된 클라우드 커넥터를 통해 40가지의 클라우드 환경을 지원합니다.

결론

빅데이터 분석은 엔터프라이즈 환경의 주류 기술로 자리 잡으면서 여러 선두 기업에서 전략적 가치를 입증해 왔습니다. 비즈니스 성패에 따라 기업은 방대한 데이터를 실시간으로, 신속하게 분석해야 합니다.

빅데이터 백업 및 복구를 지원하는 인프라스트럭처 역시 이러한 미션 크리티컬 역할에 맞게 변화해야 합니다. Veritas NetBackup Parallel Streaming Framework는 빅데이터 환경에 내재된 백업의 장애물을 해소합니다. Veritas NetBackup 8.1은 더 우수한 성능을 제공하는 동시에 리스크를 완화하고 스토리지 리소스 사용량을 줄임으로써 빅데이터 분석으로 전략 우위를 확보하려는 디지털 기업에게 새로운 기회를 제시합니다.

자세한 내용은 여기에서 확인하십시오.