

지능형 데이터 티어링 솔루션 ‘루마다 데이터 옵티마이저’, 하둡에 날개를 달다



2025년에는 전 세계의 데이터가 지금보다 5배 이상 증가할 전망이다. 데이터의 가치 극대화는 이미 기업 경쟁력을 좌우하는 핵심 요소가 됐다. 물론 데이터의 가치를 끌어낼 수 있는 하둡(Hadoop)이라는 툴이 있다. 그러나 하둡도 지금처럼 급증하는 데이터를 처리하기에는 역부족이다.

지능형 데이터 티어링 솔루션

HDFS(Hadoop Distributed File System)는 수많은 데이터 노드를 컴퓨팅과 스토리지가 위치한 단일 풀(pool)로 통합하는 클러스터 스토리지를 제공한다. 오래된 데이터가 누적돼 클러스터가 꽉 차면 스토리지 용량을 추가해야 하는데, 하둡(Hadoop)¹⁾ 환경에서는 스토리지와 컴퓨팅 용량을 함께 추가해야 한다. 이렇게 되면 리소스 간 불균형이 발생하거나 효율성도 떨어질 수 있다. 결과적으로 비용만 증가하는 상황이 된다.

그러나 앞으로는 걱정하지 않아도 된다. 지능형 데이터 티어링 솔루션인 하둡용 루마다 데이터 옵티마이저(Lumada Data Optimizer)가 이런 문제를 간단히 해결할 수 있기 때문이다.

하둡용 루마다 데이터 옵티마이저는 스토리지와 컴퓨팅 리소스를 독립적으로 확장할 수 있는 솔루션이다. 이로써 유연성이 향상될 뿐 아니라 리소스 활용 범위도 커진다. 볼륨 구성도 간단하다. 자동화된 데이터 티어링을 통해 자주 사용하지 않는 데이터는 HCP에 저장하므로 스토리지 타입, 스토리지 정책, 무버(Mover) 서비스 등 하둡의 기본 기능을 최대로 활용할 수 있다. 하둡 노드에는 자주 사용하는 데이터를 저장해 최적화된 리소스 관리가 가능하다.

HDFS에서 파일을 제거하는 S3A의 데이터 오프로딩과 달리, 루마다 데이터 옵티마이저는 HDFS와 통합된다. 따라서 용량을 충분히 확보할 수 있고, HDFS와 HCP 간 다이나믹한 데이터 티어링으로 언제나 HDFS 데이터에 안전하게 액세스할 수 있다.

1) 하둡(Hadoop): 여러 개의 저렴한 컴퓨터를 마치 하나인 것처럼 묶어 대용량 데이터를 처리하는 기술. 수천 대의 분산된 x86 장비에 대용량 파일을 저장할 수 있는 기능을 제공하는 분산 파일 시스템과 저장된 파일 데이터를 분산된 서버의 CPU와 메모리 자원을 이용해 쉽고 빠르게 분석할 수 있는 컴퓨팅 플랫폼인 맵리듀스(MapReduce)로 구성

한눈에 알아보는 루마다 데이터 옵티마이저

1 자동화된 정책 기반 티어링

자동화된 정책 기반 티어링으로 하둡 데이터가 가장 경제적인 스토리지 리소스로 이동한다. 비용 대비 사용 빈도를 최적화해 데이터를 저장하므로 데이터 활용도와 효율성이 높아진다.

자주 사용하는 데이터는 하둡에,
자주 사용하지 않는 데이터는 HCP에 저장

3 컴퓨팅과 스토리지 리소스의 독립적 확장

하둡을 확장하면 활용도를 높일 수 있다. 필요에 따라 하둡 노드를 추가하고, HCP로 스토리지를 확장할 수 있어 데이터를 장기간 보관할 수 있다. 페타바이트 규모의 오래된 데이터를 저장하기 위해 하둡 컴퓨팅 노드를 확장할 필요가 없다.

HCP로 데이터 확장성 향상,
더 효율적인 하둡 리소스 활용

2 원활한 하둡 데이터 액세스

루마다 데이터 옵티마이저는 HDFS와 HCP 간 데이터를 자동으로 티어링 하므로 데이터 경로와 애플리케이션 구성을 변경하지 않고도 언제나 데이터에 액세스할 수 있다.

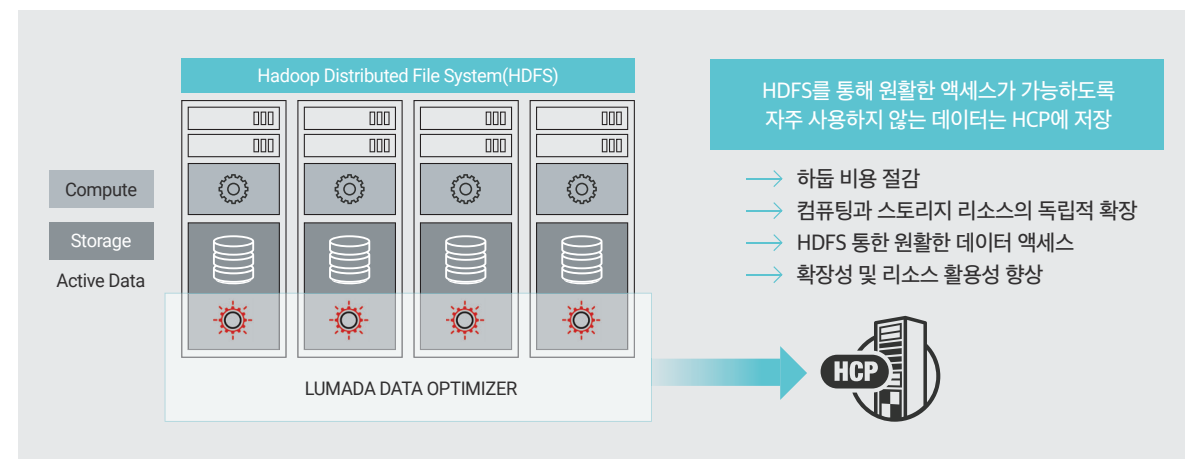
HCP로 하둡 데이터 티어링,
장기적인 데이터 보관 가능

4 지능형 하둡 데이터 관리

루마다 데이터 옵티마이저는 HDFS의 정책 기반 메타 데이터를 취합해 데이터 관리를 간소화한다. 따라서 하둡 데이터의 처리와 관리 방식도 최적화될 수 있다.

데이터 효율성과 통찰력 향상으로
최고의 비즈니스 의사결정

↓ 루마다 데이터 옵티마이저



5 하둡과 통합돼 HDFS 볼륨으로 운영

루마다 데이터 옵티마이저는 하둡과 통합돼 HDFS 볼륨으로 운영되므로 HDFS 데이터를 HCP로 이동할 수 있다. HDFS에 파일이 계속 남아있기 때문에 용량에는 아무런 영향을 미치지 않으면서 HDFS로 원활하게 데이터에 액세스할 수 있다.

HCP에 티어링된 하둡 데이터로
안전하고 원활한 HDFS 액세스

8 안전성, 유연성, 비용 효율성을 모두 갖춘 솔루션

HCP는 대규모로 확장 가능한 장기적인 클라우드 오브젝트 스토리지 플랫폼을 제공한다. 데이터 모빌리티, AES-256 암호화, 퍼블릭 클라우드에 대한 정책 기반 티어링, 강력한 데이터 보호, 최고 수준의 가용성 등 컴플라이언스 준수와 데이터 복구가 용이하다.

확장성, 경제성, 고가용성을 보장하는
HCP 스토리지에 하둡 데이터 보관

6 하둡 리소스 사용 및 활용도 최적화

하둡은 이중화와 가용성을 위해 3개의 데이터 복제본을 저장하므로 스토리지와 컴퓨팅 리소스를 늘 확보해야 한다. 사용하지 않는 데이터를 HDFS에서 HCP로 티어링 하면 스토리지와 컴퓨팅 용량을 줄일 수 있고, 하둡에는 자주 사용하는 데이터만 보관하면 된다.

오래된 데이터에 대한 하둡의 3중 데이터 복제 제거
→ 활용도는 높이고 비용은 절감

7 데이터 복원, 자체 보호 기능 탁월

HCP는 복원력이 뛰어나 데이터 복구가 간편하고, 자체 보호 기능도 탁월한 스토리지 플랫폼이다. 강력한 데이터 내구성, 복제, 이중화 구성, 자동 수리 및 버전 관리와 같은 기능을 제공한다.

HCP로 이동한 데이터 보호 정책 불필요,
가용량도 최대 40%까지 절감

빅데이터를 구축할 때는
컴퓨팅과 스토리지를
분리하는 방법이 낫다.
이렇게 하면 리소스 활용도와
유연성은 향상시키면서
비용은 절감할 수 있다.

리투 조티 / IDC AI 전략 부문 프로그램 부사장

출처: Lumada Data Optimizer for Hadoop, www.hitachivantara.com, 2019년 9월