



Solution Brief

EMC Isilon Scale-out NAS and RainStor with Hadoop for Efficient Big Data Management



기업은 최소비용으로 신뢰성, 보안성 그리고 컴플라이언스가 보장되는 하둡 인프라 구축가능

기업들은 빠르게 증가하고 있는 다중구조의 대규모 데이터 볼륨에 대한 관리 및 분석을 위하여 최저 비용으로 구축이 가능한 하둡 빅데이터 플랫폼 도입을 우선적으로 검토하고 있습니다. 그러나 하둡 인프라 도입에 따른 운용의 용이성, 효율적인 인프라 확장 지원 그리고 신뢰성 확보가 매우 중요합니다. 그리고 최적의 하둡 운영환경을 계속 유지하기 위해서는 데이터와 스토리지 증가에 대응하는 CPU와 스토리지 간의 상호 밸런싱을 위한 아키텍처가 필요합니다.

RainStor 데이터베이스를 실행하는 EMC Isilon(아이실론)은 복수의 데이터 소스의 대용량 데이터 볼륨 증가에 대응하기 위하여 만들어진 검증된 솔루션입니다. 최고의 데이터 압축을 제공하는 RainStor와 하둡 기능 그리고 EMC의 효율적인 스케일-아웃 NAS(네트워크 연결 스토리지) 플랫폼인 Isilon의 결합으로 데이터 관리의 용이성, 완벽한 확장성으로 가장 효과적인 빅데이터 인프라 구축이 가능하여 기존의 단순한 DAS 기반의 하둡 인프라보다 신뢰성, 보안성 그리고 적용의 용이성 등 다양한 이점들을 얻을 수 있게 되었습니다.

EMC Isilon 기반의 RainStor와 하둡 - NAS 신뢰성을 충족시키는 최저비용 확장성 제공

아파치 하둡 오픈소스 플랫폼은 DAS 상에서 동작되는 하둡 분산파일 시스템(HDFS)을 사용하여 저비용으로 다중구조의 데이터 볼륨을 저장하고 관리하는데 이상적인 솔루션으로 자리를 잡았습니다. 맵리듀스, 하이브 그리고 피그는 기업의 필요에 따라 비즈니스 분석을 가능케 하는 오픈소스 툴입니다. 하둡은 데이터 볼륨이 증가함에 따라 간단하게 스케일-아웃 형태의 확장 및 실행이 가능한 저가형 클러스터 서버를 사용하여 구축이 가능하도록 설계되었습니다. 그러나 기존 DAS 기반의 스토리지 사용하여 구축한 하둡 고객들은 NAS 솔루션에서 제공받은 정도의 신뢰성, 보안성 그리고 성능을 기대할 수는 없습니다.

따라서 기업들은 다양한 빅데이터 스토리지 요구에 대응이 가능한 EMC Isilon 스케일-아웃 NAS 기반의 HDFS와 네이티브 연동이 가능한 RainStor 데이터베이스에 의존하게 되었습니다.

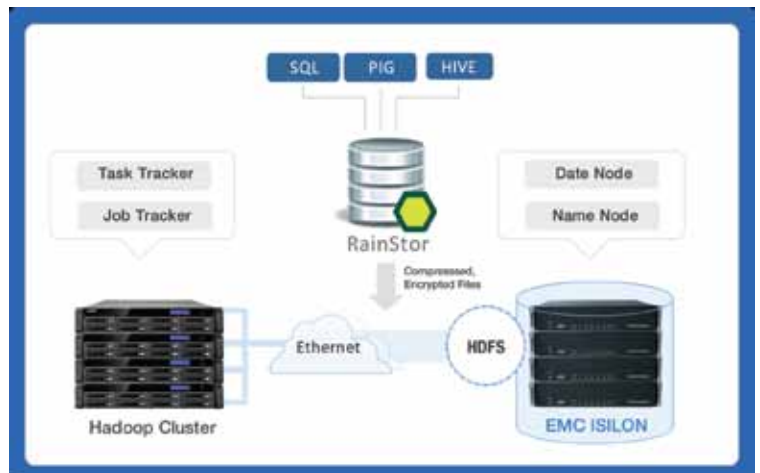
RainStor는 단순히 기계적인 측면에서 뿐만 아니라 기존 전통적인 스토리지 시스템에서 제공하는 무제한의 스케일-아웃 확장을 지원하는 Isilon OneFS 운영체제와 결합으로 20배에서 40배에 해당하는 데이터 압축을 지원합니다.

RainStor와 Isilon은 HDFS에 대한 모든 통신을 TCP/IP 상위계층에서 이루어지고 OneFS에 대한 HDFS 실시간 프로토콜이 가능한 HDFS 계층과 통합하였습니다. 이러한 통합으로 고객들은 자사의 하둡 아키텍처를 네이티브 형태의 스케일 아웃 NAS 플랫폼으로 확장이 가능하게 되었습니다.

RainStor는 네이티브 HDFS에서 운용이 가능하여 데이터를 기존 리포지터리에서 HDFS로 이동시킬 필요가 없습니다. 사실 RainStor는 HDFS 데이터에 대한 빌트-인 시큐리티를 제공하면서 하이브 쿼리, 피그 스크립트를 사용하는 것 이외에 표준 SQL을 사용하여 쿼리가 가능한 최초의 엔터프라이즈 레벨의 데이터베이스입니다. 이것은 단순히 사용자에게 빠르고 유연한 쿼리방식을 제공하는 것 뿐만 아니라 처리결과와 신속한 도출로 스토리지, 하드웨어 그리고 직원들의 자원을 효과적으로 활용하여 최대한의 효율을 얻을 수 있게 합니다.

RainStor의 특허기반의 고성능 압축 알고리즘은 스토리지 사용을 획기적으로 줄이는 효과 뿐만 아니라 쿼리성능 또한 형태에 따라 10배에서 100배까지 향상 시킬 수 있습니다. 이것이 가능한 이유는 압축으로 인한 처리 파일수의 감소와 RainStor 특유의 필터 기반의 메타 데이터 관리방식으로 가능하게 되는 것입니다. 이는 기업으로 하여금 최소 비용으로 더 많은 비즈니스 분석이 가능하게 된다는 의미입니다. 따라서 기업은 스토리지 사용량 절감 및 데이터 센터 스페이스, 전력 그리고 인건비를 줄이면서 마케팅을 위한 최적의 비즈니스 사이클을 구축이 가능하게 되었습니다.

Isilon 스케일-아웃 NAS와 하둡 연결



Isilon NAS 기반의 RainStor – 업계 최고 효율의 스케일-아웃 솔루션

NAS는 기업이 성능이나 유연성을 희생하지 않고 테라 바이트 당 낮은 비용으로 미래 지향적인 스토리지를 구성하여 대규모 다중구조 데이터에 대한 효율적인 관리가 가능한 통합 스토리지 플랫폼입니다. 고객은 NAS 스토리지 인프라 구성시에 연계될 안정성, 보안 및 성능을 중요한 요소로 평가해 왔습니다. Isilon과 RainStor 결합은 기업의 특정 비즈니스의 요구사항을 위한 용량과 성능을 모두 만족 시킬 수 있습니다. RainStor와 Isilon 솔루션으로 기업은 전례 없는 데이터 압축 및 안정성 그리고 업계의 최고 수준의 가용성으로 가장 중요한 기업의 데이터 자산을 위탁할 수 있습니다.

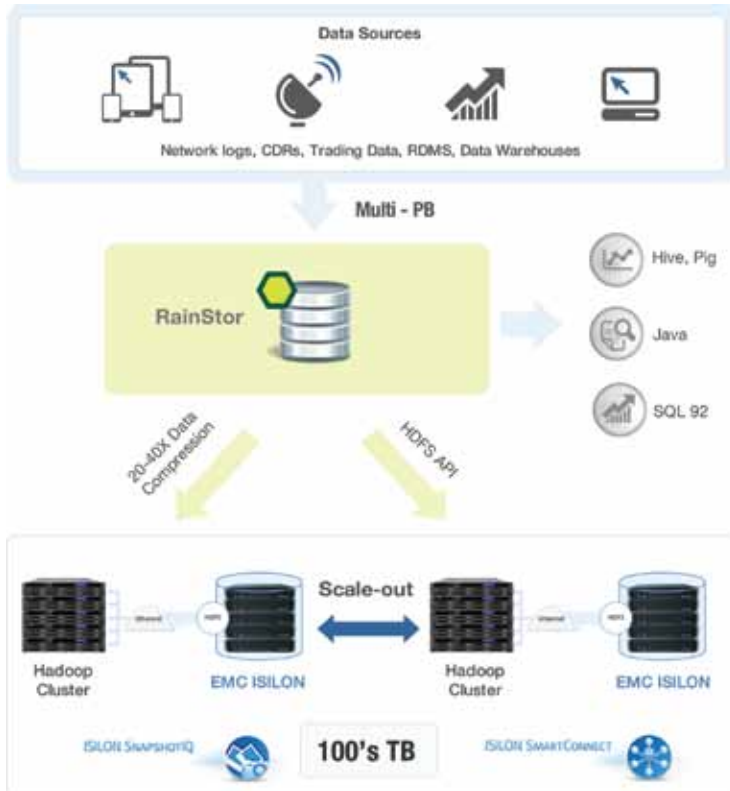
RainStor와 Isilon 결합 솔루션으로 기업은 빠른 속도의 네트워크 데이터, 머신 또는 센서 생성 데이터의 기본 저장소 또는 분석 데이터 허브로 구축이 가능합니다. 또한 이러한 통합 솔루션으로 현재 또는 거래가 발생한 이후의 추가적인 분석을 위한 액티브 아카이브 및 데이터 허브 형태의 이차 데이터 저장소로 활용이 가능합니다.

이것은 중요한 데이터 자산을 활용하려는 기업뿐만 아니라 저장을 위한 간단하면서도 강력한 솔루션입니다. 그리고 기존의 엔터프라이즈 스토리지와 달리 향후 성능에 대한 요구사항 그리고 미래 사업의 변화로 많은 스토리지가 추가되더라도 효율은 그대로 유지가 가능합니다.

빅데이터 도전: Isilon NAS 기반의 RainStor와 하둡과 DAS 기반의 하둡 간의 비교

Isilon NAS 기반의 RainStor와 하둡 결합	DAS 기반의 하둡
스케일-아웃 스토리지 플랫폼 ■ 복수의 어플리케이션	직접연결 스토리지 인프라 ■ 하둡에 직접연결
엔드-투-엔드 데이터 보호 ■ 계정관리, 프라이버시, 통합, 리플리케이션, 데이터 폐기, NDMP 백업	인터프라이즈 보안과 컴플라이언스 기능 결여 ■ 단지 기본적인 커버로스 인증
업계 최고의 스토리지 효율 ■ 제로 데이터 리플리케이션, 20~40배 데이터 압축, >80% NAS 운용성	스토리지 효율 나쁨 ■ 3X 데이터 리플리케이션
독립적인 확장성 ■ 압축으로 I/O 밴드위스 향상, 스토리지와 서버를 별도로 확장 가능	확장이 제한적임 ■ 경직된 스토리지 비용 계산
복수의 프로토콜 지원 ■ NFS, HDFS, FTP, HTTP	수동적인 임포트/익스포트 ■ 프로토콜 지원불가
빠르고 유연한 쿼리 ■ SQL, BI Tools, Hive, Pig, MapReduce, Search	제한적인 쿼리만 가능 ■ Pig, Hive, MapReduce

상위 레벨 시스템 아키텍처



주요기능

- 데이터 압축 및 스토리지 효율
 - > 20-40X 압축
 - > 80% 효율
- 빠르고 유연한 쿼리 지원
 - > SQL, Pig, Hive, MapReduce, BI 툴
- 보안 및 컴플라이언스 제공
 - > 인크립션, 증명 및 인증
 - > SnapshotIQ
 - > FlexProtect
- 성능 & 자원관리
 - > InsightIQ
 - > SMARTPOOLS
- 복수의 프로토콜 지원
 - > HDFS, NFS, HTTP, CIFS
- 데이터 리플리케이션 지원
 - > Active-Active
 - > SyncIQ