

완벽 가이드

# 빅데이터 분석

# 목차

3	이 책의 내용
4	<b>1장.</b> 현대의 기업에서 빅데이터의 의미
7	<b>2장.</b> Nutanix 솔루션으로 빅데이터 문제 해결
12	<b>3장.</b> Nutanix 기반 Splunk
18	<b>4장.</b> Nutanix 기반 Elasticsearch
20	<b>5장.</b> Nutanix 기반 Hadoop
23	<b>6장.</b> Nutanix 기반 NoSQL
27	<b>7장.</b> 빅데이터에 Nutanix HCI를 사용해야 하는 이유
31	Nutanix 시작하기

# 이 책의 내용

전용 사일로와 베어메탈을 사용하는 기존의 빅데이터 인프라 접근 방식은 가치 창출 시간이 길고, 복잡한 라이프사이클 관리와 성능 저하로 인해 비즈니스의 지속적인 성공 실현에 장애가 되고 있습니다. 빅데이터 분석을 위한 Nutanix HCI(하이퍼컨버지드 인프라) 솔루션은 이러한 문제를 해결합니다. Nutanix HCI는 빅데이터 배포 속도를 가속화하고, 규모에 맞게 효율적으로 운영하며, 더 높은 수준의 보안 및 복원력을 제공합니다. 그 결과 Nutanix HCI는 위험 요소를 최소화하고 비용을 최대 50% 절감합니다.

이 eBook에서는 Splunk, Elasticsearch, Hadoop 및 NoSQL을 비롯하여 많이 사용되는 빅데이터 워크로드를 처리하는 데 Nutanix HCI가 어떤 이점을 제공하는지 살펴봅니다.



## 1장

# 현대의 기업에서 빅데이터의 의미

지난 10년 동안 기업들은 빅데이터를 활용하여 데이터 기반의 기업으로 거듭나기 위해 상당한 노력을 기울여 왔습니다. 유수의 기업들은 현재 디지털 혁신을 신속하게 진행하기 위해 빅데이터 분석 결과로 도출된 유용한 정보를 비즈니스의 모든 측면에 통합하여 다음을 실현합니다.

- 비즈니스에 대한 경쟁 우위 확보
- 더 나은 비즈니스 성과를 실현하고 고객에게 개선된 서비스 제공
- 고객 경험, 고객 참여 및 고객과의 상호 작용 개선
- 실행 가능한 분석 정보를 통해 위험 최소화
- 자동화 및 오케스트레이션을 통해 데이터 기반의 운영 지원

많은 IT 팀이 빅데이터를 일상적인 IT 및 비즈니스 운영의 필수적인 요소로 만들기 위해 기존의 빅데이터 접근 방식을 재고하고 있습니다. 전용 사일로와 베퍼메탈을 사용하는 기존의 빅데이터 인프라 접근 방식에는 다음과 같은 문제가 있습니다.

- **가치 창출에 긴 시간이 소요.** 새로운 빅데이터 프로젝트를 위한 초기 파일럿을 진행하는 데 몇 달이 걸릴 수 있으며 이 파일럿을 프로덕션 수준으로 확장하는 데는 훨씬 더 오랜 시간이 걸릴 수 있으므로 자본비용(CapEx)과 운영비용(OpEx)이 증가합니다.
- **성능 및 사용자 경험 저하.** 빅데이터 애플리케이션은 비효율적인 스토리지 접근 방식으로 인해 용량을 낭비하고, 인덱싱 및 쿼리 프로세스 처리 속도를 늦추며, 처리량이 급증하는 수집 요구 사항을 처리하지 못하는 경우가 많습니다.
- **복잡한 라이프사이클 관리.** 업데이트 주기는 거의 영구적이고, 확장 작업은 예측이 불가능하고 비용이 많이 들며, IT 전문가가 취약한 인프라를 관리하는 데 몇 시간씩 할애합니다.



빅데이터 분석의 경우 데이터에서 가치를 도출하는 기능이 무엇보다 중요합니다. 일부 데이터는 그 가치가 빠르게 손실되므로 신속하게 수집하고 처리해야 합니다. 또한 데이터 유형이 다양하기 때문에 유연하게 활용할 수 있는 스토리지 옵션이 필요합니다.

빅데이터 분석용 Nutanix 솔루션은 Splunk 및 Elasticsearch와 같은 소프트웨어 플랫폼을 강화하여 다음 기능을 포함한 다양한 이점을 제공합니다.

- 고급 가상화를 사용하여 성능은 그대로 유지하면서 규모에 맞게 효율적으로 운영합니다.
- 더 높은 수준의 보안, 복원력, 가용성 및 복구 성능을 제공합니다.
- 파일럿에서 프로덕션 단계로 진행 시 빅데이터 분석 배포 속도가 최대 8배 빨라집니다.
- 기존 IT 접근 방식에 비해 비용을 최대 50% 절감하고 위험을 최소화합니다.

이 eBook에서는 Nutanix의 빅데이터 분석 솔루션을 사용하여 적은 노력으로 데이터에서 더 큰 가치를 창출하는 방법에 대해 설명합니다.



# 정형, 비정형, 반정형 데이터 비교

정형 데이터는 잘 구성된 데이터 유형으로, 고객 이름, 계정 번호, 전화 번호, 구입한 제품, 구입 날짜 등과 같이 일반적이고 명확하게 정의된 유형 필드로 구성된 관계형 데이터베이스에서 찾을 수 있습니다. 정형 데이터는 관계형 데이터베이스를 사용하여 관리되고 기존의 비즈니스 인텔리전스 툴을 사용하여 분석됩니다.

**비정형 데이터**의 경우 정형 데이터와 같은 구성이 결여되어 있습니다. Facebook 게시물, 트위터, 블로그 게시물과 같이 흥미로움을 유발하는 소셜 콘텐츠가 비정형 데이터의 예입니다. 이 유형의 데이터는 관계형 데이터베이스에서 쉽게 관리되지 않으므로 데이터에서 가치를 도출하려면 다양한 툴 세트가 필요합니다. 예를 들어 빅데이터 툴은 소셜 미디어 데이터에서 고객 감정을 추출하는 데 사용되는 경우가 많습니다.

**반정형 데이터**는 어느 정도의 구조를 갖췄지만 정형 데이터베이스와 같이 잘 정의된 스키마는 없습니다. 이메일 메시지는 반정형 데이터의 한 예로, 발신인, 수신인 등과 같은 헤더에는 잘 정의된 데이터 필드가 포함되지만 메시지의 실제 본문은 구조화되어 있지 않습니다. 이메일을 보내고 받는 사람과 시간(헤더에 포함된 정보)을 확인하려면 관계형 데이터베이스를 사용하는 것이 좋습니다. 그러나 메시지 내용에 더 관심이 있다면 자연어 처리와 같은 빅데이터 툴이 더욱 적합합니다.

IT 장비, 애플리케이션, 제조 장비 등에서 생성된 출력 로그인 **머신 데이터**는 비정형 또는 반정형 데이터로 구성되며 환경이 계속 확장되고 애플리케이션이 더욱 복잡해지고 서로 연결됨에 따라 IT 운영에 매우 중요한 요소로 자리잡았습니다.

기업 데이터의 약 90%가 비정형 데이터이므로 정형 데이터와 비정형 데이터의 양 간의 격차는 더욱 커질 수 있습니다. 이것이 비정형 데이터를 이해하는 데 도움이 될 수 있는 빅데이터 툴이 중요한 이유입니다.

## STRUCTURED VS UNSTRUCTURED VS SEMI-STRUCTURED DATA

### STRUCTURED

- OLTP
- Transactional DB



### SEMI-STRUCTURED

- IoT
- Sensors Email
- Call Center Logs



### UNSTRUCTURED

- Social Networking Posts
- Documents



# Nutanix 솔루션으로 빅데이터 문제 해결

기존의 많은 빅데이터 애플리케이션은 설계, 배포 및 관리가 어려운 전용 베어메탈 서버 클러스터에 배포되었습니다. 일부 기업은 별도의 서버, SAN(Storage Area Network) 및 스토리지 어레이가 있는 기존 3-티어 인프라 사용을 선호하지만, 이러한 접근 방식에는 자체 문제가 있으며 다음과 같은 이유로 복합적인 결과가 발생하는 경우가 많습니다.

- **데이터 로컬리티 부족.** Hadoop과 같은 빅데이터 프레임워크는 운영 중인 데이터가 해당 작업을 수행하는 서버에 국한될 것이라는 개념을 기반으로 설계되었습니다. 네트워크 스토리지를 도입하면 지연과 병목 현상이 발생합니다.
- **I/O 블렌더 효과.** 까다로운 빅데이터 작업에서 발생하는 I/O 요청의 스트림으로 인해 I/O가 무작위로 혼합되면 스토리지 미리 읽기 알고리즘을 무효화할 수 있으므로 중앙 스토리지 시스템이 효율적으로 작동하기 어렵습니다.
- **복합적인 I/O 요구 사항.** 일부 빅데이터 작업에는 높은 스트리밍 성능이 필요한 반면 다른 작업은 임의 I/O를 생성합니다. 대부분의 스토리지 어레이는 두 유형을 모두 제대로 처리하지 못합니다.

위에서 언급한 접근 방식 중 하나를 사용하면 새로운 배포, 인프라 업그레이드, 전면 교체 시 사전 계획과 상당한 노력이 필요하며 관리 작업은 매우 어렵고 일관성이 떨어지게 됩니다. 따라서 업그레이드나 다른 유지 보수 작업이 항상 필요한 것처럼 보일 수 있습니다.

Nutanix는 인프라 배포 및 관리를 간소화하여 빅데이터 애플리케이션을 지원함으로써 적은 노력으로 더 나은 성능을 제공하고 빅데이터 애플리케이션이 나머지 작업에서도 잘 작동하도록 지원합니다. 주요 이점은 다음과 같습니다.

- **가치 실현 시간 단축.** 파일럿에서 프로덕션 단계로 진행 시 배포 속도를 최대 8배 단축하여 위험을 최소화합니다.
- **대규모 처리 성능.** 소규모로 시작하여 쉽게 확장 가능하며 선형 확장 기능은 요구 사항이 증가함에 따라 예측 가능한 성능을 제공합니다.
- **라이프사이클 관리 간소화.** 업데이트에 소요되는 시간을 단축하고 다운타임 없이 몇 분 만에 새로운 스토리지 서비스를 추가할 수 있습니다.
- **TCO 절감.** 총 소유 비용을 최대 50% 절감할 수 있습니다.

Nutanix HCI는 강도 높은 인프라 및 애플리케이션 자동화를 통해 증가하는 요구 사항에 발맞춥니다. 고급 데이터 서비스, 기본 제공되는 보안, 지능형 운영 및 기본 데이터 보호 기능이 플랫폼에 내장되어 있습니다. Nutanix는 훨씬 적은 노력으로도 빅데이터 애플리케이션에 필요한 보안, 가용성, 복원력, 중복성 및 복구 기능을 제공할 수 있습니다.

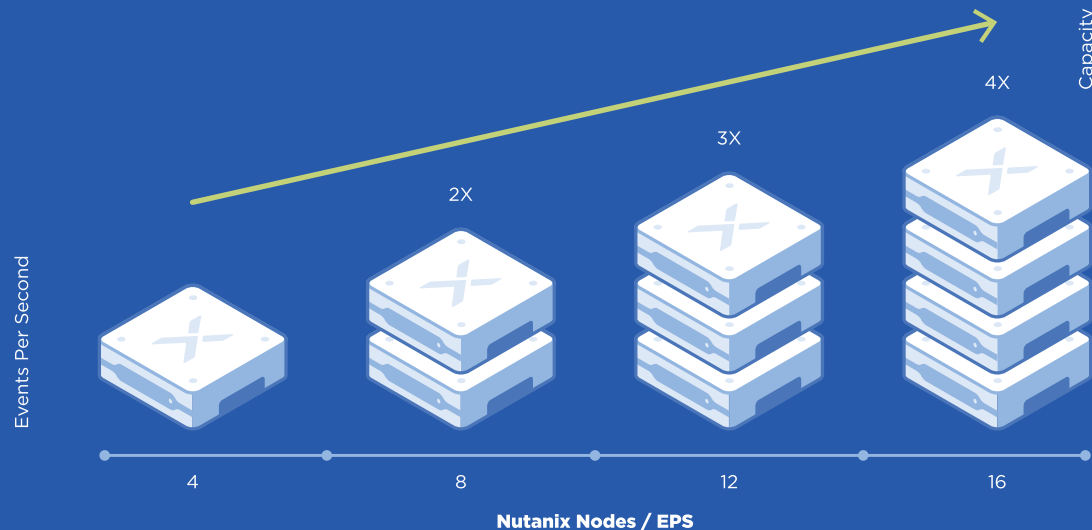


그림 1. Nutanix HCI는 빅데이터 워크로드에 대해 예측 가능한 선형 확장 기능을 제공합니다.



## 지능형 운영 및 자동화

Nutanix HCI는 전체 라이프사이클에 걸쳐 인프라 관리를 간소화하고 운영을 자동화하며 셀프 서비스를 지원하여 가장 끈질긴 빅데이터 문제를 해결합니다.

지능형의 원클릭 작업만으로 소프트웨어 설치, 업그레이드 및 워크로드 배치와 같은 일상적인 작업의 번거로움과 수고를 덜어줍니다. 고급 분석 기능은 빅데이터 환경에 유용한 분석 정보를 제공하므로 활용률, 성장률 및 기타 중요한 계획 정보를 즉각적으로 파악할 수 있습니다.

Nutanix Calm을 사용하면 표준 또는 맞춤형 애플리케이션 블루프린트를 배포하고 몇 분 내에 실제 사용이 가능한 환경을 구축할 수 있으므로 새로운 배포 과정을 획기적으로 간소화하고 가치 실현 시간을 단축할 수 있습니다.

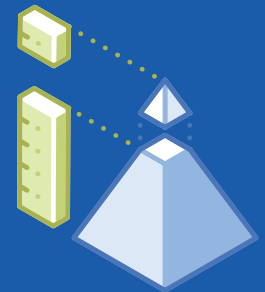
## 종합적인 데이터 플랫폼 서비스

스토리지 용량과 성능 관리는 빅데이터 애플리케이션에 있어 중요한 문제입니다. Nutanix는 컴퓨팅을 통해 동일한 플랫폼에서 모든 데이터 서비스를 통합합니다. Nutanix 데이터 서비스는 프로비저닝 및 관리를 간소화하는 동시에 활용률을 높입니다. 데이터 로컬리티와 지능형 계층은 지속적인 조정 없이도 성능을 최적화하고 별도의 스토리지 시스템을 설계하고 관리할 필요를 없애줍니다. 사용 가능한 데이터 서비스에는 현재는 물론 미래의 요구 사항을 해결하기 위해 블록, 파일 및 개체 스토리지가 포함됩니다.

## 성능 및 확장성

Nutanix HCI 기술을 사용하면 간단한 빌딩블록으로 강력하고 유연한 스케일 아웃 아키텍처를 구축할 수 있습니다. 불확실한 짐작에 의존하거나 비용이 많이 드는 초과 프로비저닝 없이 소규모로 시작하여 필요에 따라 확장할 수 있습니다. 빅데이터 애플리케이션은 다른 애플리케이션과 인프라를 공유할 수 있습니다. 또는 사용자가 클러스터를 애플리케이션 전용으로 지정하여 격리를 강화할 수 있습니다. 적당한 규모의 Nutanix 클러스터는 하루에 테라바이트급의 데이터를 수집하고, 초당 수백만 개의 이벤트를 처리하며, 필요에 따라 성능을 점진적으로 확장할 수 있습니다.

성능 및 용량은 예측 가능하게 선형으로 확장되므로 필요에 따라 리소스를 점진적으로 추가하여 예상치 못한 대규모 자본 지출을 방지할 수 있습니다.



# 보안, 가용성 및 데이터 보호

Nutanix HCI 플랫폼은 보안, 가용성 및 데이터 보호를 최우선으로 고려하여 설계되었습니다.

- **보안.** Nutanix는 보안을 최우선으로 하는 접근 방식을 취합니다. 기본 제공되는 심층적인 방어 기능을 통해 소중한 데이터를 언제나 보호합니다. 암호화 및 마이크로세그멘테이션은 플랫폼의 기능을 확장합니다.
- **복원력.** Nutanix HCI는 단일 장애 지점이 없고 병목 현상이 발생하지 않는 내결함성을 갖추고 있습니다. 또한 장애를 감지, 격리 및 복구하도록 구축되어 단순한 하드웨어 고장으로 인해라도 종종 발생하는 비상 상황을 피할 수 있습니다.
- **가용성.** Nutanix는 운영자 개입 없이 완전하게 복원되는 자가 회복 아키텍처를 통해 기존 인프라보다 뛰어난 기능을 제공합니다. 중단 없는 확장 및 업그레이드 기능과 결합된 자가 회복 기능은 계획된 다운타임과 계획되지 않은 다운타임을 감수해야 하는 불편을 없애줍니다.
- **데이터 보호 및 DR.** 기본 제공되는 데이터 보호 및 DR 기능은 인프라를 간소화하고 병목 현상을 제거하며 적은 노력으로도 뛰어난 보호 기능을 제공하는 동시에 데이터 복제본의 수를 최소화하도록 지원합니다. 자세한 내용은 **비즈니스 연속성 완벽 가이드**를 참조하십시오.

빅데이터 환경에서 사용하는 Nutanix HCI에 대한 자세한 내용은 7장을 참조하십시오.

## NUTANIX HCI 자세히 알아보기

- **Nutanix가 하는 일**
- **보안 솔루션 브리프**
- **HCI 완벽 가이드**
- **Nutanix Bible**
- **[nutanix.com/bigdata](https://nutanix.com/bigdata)**
- **Nutanix HCI 테스트 드라이브 체험**

# 빅데이터의 5대 요소

빅데이터 운영은 다음과 같은 5가지 주요 요소로 특성이 지정됩니다.

- 데이터의 **볼륨**
- 데이터의 **처리 속도**
- 데이터 유형의 **다양성**
- 데이터의 **정확성**, 신뢰성 및 유용성
- 데이터의 **가치** 또는 데이터에서 가치를 도출하는 방법

위의 5대 요소 측면에서 데이터를 파악하면 요구 사항에 가장 부합하는 툴을 선택하는 데 도움이 됩니다.



# Nutanix 기반 Splunk

Splunk Enterprise는 IT 인프라와 모든 유형의 장비에서 수집한 머신 데이터의 잠재력을 극대화할 수 있는 선도적인 소프트웨어 플랫폼으로, 운영, 규정 준수, 보안 및 감사를 간소화하는 등 IT 부문에서 다양하고 폭넓은 용도로 사용됩니다. Splunk는 모든 종류의 머신 데이터에서 작동하기 때문에 IoT(Internet of Things) 지원을 비롯하여 금융 서비스, 헬스케어, 석유 및 가스, 제조 등의 업종에 대한 다양한 사용 사례를 보유하고 있습니다.

Splunk는 세 가지 주요 구성 요소를 사용하여 머신 데이터를 캡처, 인덱싱 및 상호 관계를 파악합니다.

- **포워더(Forwarder).** 데이터를 수집한 후 인덱싱을 위해 해당 데이터를 Splunk로 전송합니다.
- **인덱서(Indexer).** 수신한 데이터를 수집, 인덱싱, 저장합니다.
- **서치헤드(Search head).** 검색 요청을 인덱서에 분산합니다.



그림 2. Splunk 배포를 처리하는 세 가지 주요 구성 요소의 유형입니다.

이러한 각 기능의 CPU, 메모리 및 스토리지 요구 사항은 서로 다릅니다. 이로 인해 이상적인 인프라 환경을 설계하기가 매우 어려워질 수 있습니다. Splunk 환경의 최우선 과제인 복잡성을 해소하고 데이터 보안을 강화하며 병목 현상을 제거하는 것은 기존 인프라 접근 방식을 통해 해결하기에는 부적합합니다.

# Splunk 배포 간소화 및 가속화

Nutanix는 Splunk에 대한 인프라 관리의 복잡성을 해소합니다. 이를 통해 Splunk는 기존 인프라에서 베어메탈 또는 가상화 솔루션의 제약 없이 서버 가상화를 최대한 활용할 수 있으므로 인덱서와 서치헤드의 수와 구성을 신속하게 미세 조정할 수 있습니다. Nutanix는 모든 인덱서에서 데이터에 로컬로 액세스할 수 있게 함으로써 기존 3-티어 인프라에서 스토리지 시스템의 걸림돌이었던 "I/O 블렌더 효과"를 없애줍니다.

Splunk 배포는 새로운 데이터 소스가 추가될 때마다 빠르게 확장됩니다. Nutanix를 사용하면 기존 아키텍처에서 발생하는 병목 현상에 대한 걱정 없이 소규모로 시작하여 확장할 수 있습니다.

- **하루에 테라바이트급 데이터 수집.** 소형 4노드, 2U 클러스터는 초당 GB급의 대량 데이터를 순차적으로 처리합니다.
- **초당 수백 개의 이벤트 처리.** 4노드 클러스터는 초당 수십만 개의 이벤트를 처리할 수 있습니다.
- **점진적 확장.** 소규모로 시작하여 노드를 추가하면서 선형으로 확장할 수 있습니다.

Nutanix HCI는 선형 확장 기능을 제공하므로 Splunk 배포도 걱정 없이 확장할 수 있습니다. 각 추가 노드는 예측 가능한 성능을 제공하여 Splunk 서치헤드 및 인덱서를 지원합니다. 분산 아키텍처 덕분에 Nutanix 엔터프라이즈 클라우드는 워크로드 간의 균형을 맞추므로 필요한 경우 빅데이터 작업은 동일한 인프라를 다른 워크로드와 공유할 수 있습니다. 기본 제공되는 용량 계획은 확장을 위한 추가 용량이 필요할 때 사전 대응식으로 관리자에게 알림을 전송합니다.







# Verint가 Nutanix 기반 Splunk로 전환한 이유

Verint Systems는 고객 참여 및 관리, 보안, 영상 관제, 비즈니스 인텔리전스를 위한 소프트웨어 및 하드웨어 솔루션을 제공합니다. Verint는 증가하는 Splunk 설치를 지원하기 위해 수많은 물리적 서버를 활용하고 있었기 때문에 상당한 관리 문제가 발생했습니다.

Nutanix HCI는 모든 관리 문제를 해소하고 회사에 필요한 성능을 제공하기 때문에 Nutanix 기반 Splunk로 전환했습니다. Splunk로의 전환에 아주 만족한 Verint는 미션 크리티컬 VDI 환경과 엔터프라이즈 애플리케이션도 Nutanix로 이전했습니다.

전체적인 이점은 다음과 같습니다.

- 빠른 배포를 통해 성장에 대응
- 간편하고 운영 중단 없는 업그레이드
- 데이터센터 상면 50% 축소
- 탁월한 단일 공급업체 지원

“ 기존의 IT 인프라는 증가하는 Splunk 설치 요구 사항을 해결하기에 적합하지 않았습니다. Nutanix 엔터프라이즈 클라우드는 Splunk에는 **가상화의 운영 효율성을 제공하고 I/O 집약적인 워크로드에는 최고의 성능**을 보장합니다. Nutanix는 트랜잭션 중심의 Splunk 워크로드를 위해 **확장성이 뛰어난 고성능 검색 및 인덱싱 플랫폼**을 제공했습니다. ”

— Cyrille Valery, Verint Systems Inc. IT 책임자

# Splunk SmartStore와 통합

스토리지 용량과 성능 관리는 Splunk에 있어 중요한 문제입니다. Nutanix는 컴퓨팅을 통해 동일한 플랫폼에서 모든 데이터 서비스를 통합하며 프로비저닝 및 관리를 간소화하는 동시에 활용률을 높입니다. 데이터 로컬리티와 지능형 계층은 지속적인 조정 없이도 최적의 성능을 보장하는 동시에 Splunk 데이터를 위한 별도의 스토리지 티어를 설계하고 관리할 필요를 없애줍니다. Splunk SmartStore는 자동으로 콜드 Splunk 데이터를 온프레미스 또는 클라우드 오브젝트 저장소로 이전합니다. Nutanix Objects는 Splunk SmartStore를 실행할 수 있습니다. Objects는 기존 또는 전용 Nutanix HCI 클러스터에서 실행되며, 관리의 복잡성을 가중시키지 않으면서 필요에 따라 다양한 성능 특성을 제공하도록 최적화할 수 있는 유연한 온프레미스 오브젝트 저장소를 제공합니다.

Nutanix Objects는 SmartStore를 지원하기 위해 테라바이트에서 페타바이트에 이르는 다양한 규모의 비정형 데이터를 수용하며, 불변성과 관리 연속성을 위해 WORM 기능을 제공합니다.

# Splunk 배포 자동화

Nutanix HCI는 전체 라이프사이클에 걸쳐 인프라 관리를 간소화하고 운영을 자동화하며 셀프 서비스를 지원하여 Splunk 운영을 효율화합니다. Nutanix Calm을 사용하면 표준 또는 맞춤형 Splunk 블루프린트를 배포하고 몇 분 내에 실제 사용이 가능한 환경을 구축할 수 있으므로 새로운 배포 과정을 획기적으로 간소화하고 가치 실현 시간을 단축할 수 있습니다. Calm을 사용하면 프라이빗 클라우드에서도 Splunk 서비스를 쉽게 사용할 수 있습니다.

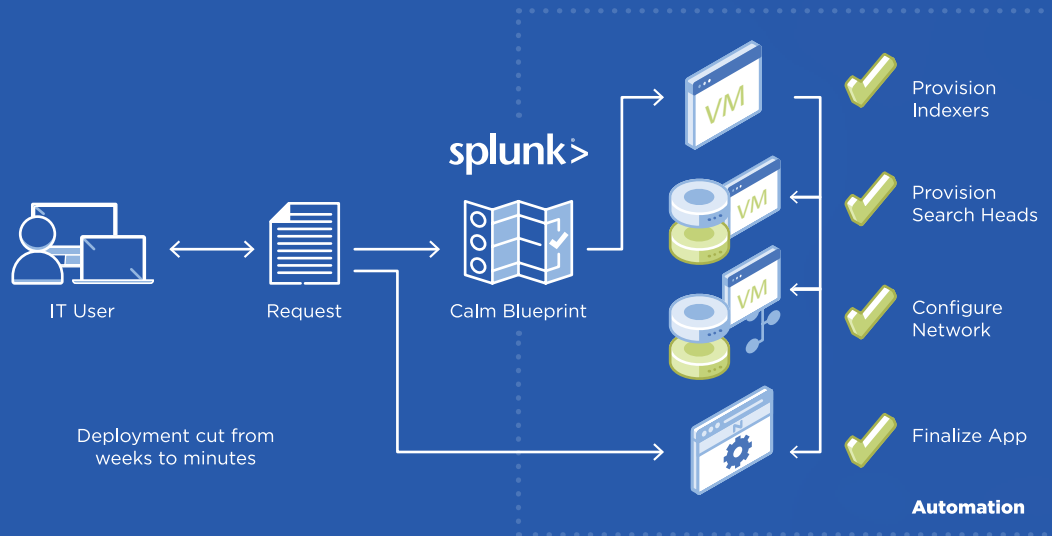


그림 3. Nutanix Calm은 새로운 Splunk 배포 과정을 간소화하고 가속화합니다.

## 보안 정보 및 이벤트 관리

기업은 내부 및 외부 공격을 식별하고 차단하기 위해 보안 정보 및 이벤트 관리 또는 SIEM에 점점 더 많이 의존하고 있습니다. Splunk Enterprise Security는 보안 모니터링, 고급 위협 탐지, 사고 대응은 물론 광범위한 추가 보안 사용 사례에 널리 사용됩니다.

Nutanix HCI에 기본 제공되는 보안 기능과 사용 용이성 때문에 Nutanix 기반 Splunk는 성능 저하 없이 확장 및 관리를 간소화하여 SIEM을 지원하는 이상적인 플랫폼입니다. 많은 고객들이 중요한 보안 프로세스의 격리를 강화하기 위해 Nutanix 클러스터를 SIEM 전용으로 사용하고 있습니다. Nutanix는 인프라 관리의 오버헤드를 제거하여 데이터 과학자와 보안 전문가가 데이터에서 유용한 분석 정보를 추출하는 데 더 많은 시간을 할애할 수 있게 해줍니다.

### Nutanix 기반 Splunk 자세히 알아보기

- [솔루션 브리프](#)
- [Nutanix AHV 기반 Splunk 가상화](#)
- 블로그: [Nutanix Objects 2.0](#)

# Nutanix에 SIEM용 Splunk 배포

Nutanix 사이버 보안 팀은 회사의 보안 요구 사항을 해결하기 위해 2017년부터 Nutanix 플랫폼에서 Splunk를 실행해 왔습니다. 또한 회사의 성장에 맞춰 Splunk 환경을 신속하게 확장해오고 있습니다. 수집된 데이터는 2017년부터 2018년까지 10배 이상 증가(하루 1TB 이상)했지만 노드 수는 3배만 증가했습니다.

Nutanix 팀은 인프라의 제한을 넘어 Splunk 운영을 위협하는 예상치 못한 이벤트에 신속하게 대응해야 했습니다.

- 30명의 분석가 추가로 인해 Splunk의 메모리 및 CPU 제한에 도달했습니다.
- 데이터센터 마이그레이션으로 인해 데이터 수집이 50% 급증했습니다.
- 코로나19로 인한 재택 근무로의 전환으로 수집량이 급증했습니다.

Prism 관리를 통해 사이버 보안 팀은 인프라 팀의 개입 없이도 이러한 이벤트를 신속하게 처리할 수 있었습니다. 몇 번의 클릭만으로 메모리, CPU 및 스토리지 용량을 추가할 수 있으므로 문제를 즉시 해결할 수 있습니다.

자세한 내용은 **Nutanix의 사이버 보안 관련 프레젠테이션**을 참조하십시오.



# Nutanix 기반 Elasticsearch



Elasticsearch는 속도와 확장성을 고려하여 설계된 업계 최고 수준의 오픈 소스 검색 및 분석 솔루션으로, Logstash, Kibana 및 Beats와 함께 사용되는 경우가 많고 **ELK Stack** 또는 Elastic Stack이라고도 합니다.

Elasticsearch 클러스터는 여러 Elasticsearch 노드 유형에 의존합니다. 이러한 기능 분리는 솔루션을 확장 가능하게 하는 요소의 일부입니다. 특히 Elasticsearch 데이터 노드는 데이터를 저장하고 검색을 포함한 모든 데이터 관련 작업을 수행합니다.

많은 회사가 내부 드라이브가 있는 베어메탈 서버에서 대규모 Elasticsearch 배포를 실행합니다. 이로 인해 서버 구성을 선택 및 최적화하고 클러스터를 배포할 때 상당한 문제가 발생합니다. Elasticsearch를 Nutanix HCI로 이전하면 인프라가 크게 간소화되며, 이외에도 다음과 같은 이점을 얻을 수 있습니다.

- **신속한 프로비저닝.** Nutanix 솔루션을 이미 사용 중인 곳에서는 예비 컴퓨팅 및 용량을 사용하여 배포할 수 있습니다.
- **가용성 증대.** 베어메탈 서버 장애로 복구 시간이 길어지면 운영에 영향을 미칠 수 있습니다. Nutanix는 드라이브, VM 및 Nutanix 노드 등의 장애에 대해 우수한 복원력을 갖췄습니다. 올바르게 설계된 대형 Nutanix 클러스터는 전체 블록(1, 2 또는 4대의 서버를 포함하는 쉼) 또는 전체 랙의 장애에도 견딜 수 있습니다.
- **효율성 향상.** Nutanix HCI는 기본 제공되는 압축 및 이레이저 코딩 기술로 데이터 효율성을 높입니다.
- **유연성 증대.** Nutanix HCI 환경에서는 훨씬 더 유연한 방식으로 각 Elasticsearch 노드를 구성하고 Nutanix 클러스터 전반에 걸쳐 해당 노드의 배포를 최적화할 수 있습니다. 실험 능력은 더 나은 결과, 효율성 증대 및 운영 비용의 감소로 이어집니다.

Nutanix는 Nutanix AHV 하이퍼바이저를 통해 Nutanix 기반 Elastic Stack을 배포할 경우의 이점에 대해 설명하는 **솔루션 노트**를 발표했습니다.

자세히 알아보기

- **AHV 기반 Elastic Stack**
- **Nutanix에서 Elasticsearch 실행: 기술 심층 학습**(블로그)
- **Nutanix 기반 Elasticsearch를 선택해야 하는 이유? Elasticsearch 데이터 복원력**(블로그)



# 까다로운 Elasticsearch 애플리케이션을 엔터프라이즈 클라우드에 배포한 IHG

InterContinental Hotels Group(IHG)은 Elastic Stack을 사용하여 맞춤형 애플리케이션에 대한 실시간 로그 분석과 성능 지표를 제공합니다. 운영 및 개발 팀은 이 데이터를 활용하여 애플리케이션을 파악하고 인프라를 관리합니다. 사용자는 쿼리를 제출하고 모든 데이터를 신속하게 검색할 수 있습니다.

IHG는 널리 사용되는 컨버지드 인프라 솔루션(별도의 서버와 스토리지로 사전 패키징한 솔루션)에서 Elastic Stack을 실행하고 있었습니다. 환경은 빠르게 성장하고 비즈니스와 관련해 중요도가 점점 더 높아지고 있지만, 이 솔루션은 확장하기가 쉽지 않고 소프트웨어 및 펌웨어 업그레이드, 유지 보수 및 기타 프로세스로 인해 운영이 중단될 수 있으며 시간과 비용이 많이 드는 것으로 확인되었습니다.

IHG 팀은 Nutanix를 선택하기 전에 성능, 사용 편의성, 관리 기능, 총 비용을 기준으로 여러 HCI 솔루션을 평가했습니다. Nutanix 네이티브 AHV 하이퍼바이저를 통해 IHG 팀은 이전 솔루션에서 막대한 비용이 발생했던 VMware 가상화 라이선스 비용을 최소화할 수 있었습니다.

Nutanix로 업그레이드한 후 운영 팀은 다음과 같은 즉각적인 이점을 누릴 수 있었습니다.

- **실시간 성능 확인.** IHG 팀은 이제 더 많은 데이터를 수집하여 실시간으로 분석할 수 있습니다.
- **신속하고 쉬운 업그레이드.** 실행 중인 애플리케이션에 영향을 주지 않으면서 중단 없이 원클릭 업그레이드를 수행할 수 있습니다.
- **인프라 가시성 향상.** Prism을 통해 IHG는 다수의 클러스터를 관리하면서 매주 상당한 시간을 절약하고 있습니다.
- **확장 간소화.** 확장하려면 노드만 추가하면 됩니다. Prism이 최적의 구성을 추천해 줍니다.

IHG 배포에 대한 자세한 내용은 [기술 사례 연구](#)를 참조하십시오.

“Elasticsearch 분석을 위해 Nutanix 엔터프라이즈 클라우드 소프트웨어를 사용하면서 이제 **더 많은 양의 데이터를 실시간으로 수집하고 분석할 수 있게 되었습니다.** 향상된 분석을 신속하게 수행하여 보다 능동적으로 대처하면서 IT 환경을 안정적으로 유지할 수 있습니다.”

— Brian Malia, IHG 시스템 및 스토리지 관리자

**IHG®**

# Nutanix 기반 Hadoop



Apache Hadoop은 대규모 데이터 세트를 동시에 처리하기 위한 분산 플랫폼입니다. Hadoop은 머신 클러스터에서 실행되고 최대 수천 개의 노드까지 확장할 수 있으며, 다음과 같은 주요 요소로 구성됩니다.

- **Hadoop Common.** 다른 구성 요소에서 공유 및 사용하는 유틸리티가 포함됩니다.
- **HDFS(Hadoop Distributed File System).** 여러 서버에 분산되어 있는 데이터를 관리하는 데 사용되는 분산 파일 시스템입니다. HDFS는 일반적으로 중복성, 가용성 및 성능을 위해 3개의 데이터 복제본을 관리합니다.
- **YARN(Yet Another Resource Negotiator).** 작업 일정을 예약하고 클러스터 리소스를 관리하기 위한 프레임워크입니다.
- **MapReduce.** YARN 기반으로 구축된 MapReduce 알고리즘을 구현한 Hadoop 결과물입니다.

Hadoop은 원래 상용 베어메탈 서버의 클러스터용으로 설계되었으며, 각 서버마다 독립적인 내부 스토리지를 갖추고 있습니다. 복원을 위해 여러 개의 데이터 복제본이 서버에 분산됩니다. Hadoop은 기존의 3-티어 아키텍처 또는 HCI에서도 실행될 수 있습니다.

Nutanix HCI는 다른 워크로드와 마찬가지로 Hadoop을 가상화함으로써 리소스를 신속하게 배포 및 재배포하는 기능 등 가상화의 민첩성을 최대한 활용하도록 지원할 수 있습니다. Nutanix HCI의 유연성 덕분에 주말이나 저녁 시간과 같이 근무 외 시간에 빅데이터 작업을 위한 기존 인프라를 재배포할 수 있으므로 빅데이터 분석 비용을 더욱 절감하고 더 큰 투자 수익을 얻을 수 있습니다.

다른 오픈 소스 소프트웨어와 마찬가지로, 많은 공급업체가 Hadoop 분산 기능을 제공하여 소프트웨어를 더 쉽게 사용하고 지원을 제공하며 추가 기능과 서비스를 제공할 수 있게 해줍니다. Nutanix는 Hadoop 지원을 위해 Cloudera 및 Hortonworks와 파트너십을 체결했습니다. Cloudera는 2019년 HortonWorks를 인수했으며 계속해서 제품을 제공하고 있습니다.

## Nutanix 기반 Cloudera

Nutanix는 Cloudera와 협력하면서 회사의 Hadoop 운영을 간소화하고 가속화할 수 있게 되었습니다. Nutanix에서 실행하는 Cloudera의 중요한 이점은 Hadoop이 다른 애플리케이션과 동일한 인프라에서 실행되므로 사일로가 제거되고 인프라 관리를 간소화하고 비용을 절감할 수 있다는 것입니다.

Nutanix와 Cloudera의 조합으로 얻을 수 있는 이점을 다음과 같습니다.

- 빅데이터 프로젝트의 프로비저닝 간소화 및 가속화
- 베어메탈의 성능으로 가상화의 유연성과 관리 용이성 제공
- 빅데이터를 위한 인프라 구입, 설치 및 관리 비용 절감
- 온프레미스 인프라에서 빅데이터 프로젝트의 사용 편의성 향상
- 가용성을 높이고 계획된 다운타임의 필요성 제거

Nutanix는 Cloudera와의 협력을 통해 Nutanix 엔터프라이즈 클라우드에서 실행되는 Cloudera Hadoop을 더 쉽게 배포할 수 있도록 지원하는 완벽한 레퍼런스 아키텍처를 구축했습니다.

### 자세히 알아보기

- [솔루션 브리프](#)
- [Cloudera 레퍼런스 아키텍처](#)
- [Hortonworks 레퍼런스 아키텍처](#)

The Cloudera logo is displayed in white lowercase letters against a dark blue background. The background of the entire slide features a faint, abstract pattern of white lines and dots, resembling a network or circuit board.

# Nutanix 기반 Hadoop 배포

Hadoop을 지원하기 위해 인프라를 설계하고 설정하는 일은 시간이 많이 걸리는 복잡한 작업이 될 수 있습니다. Nutanix는 Hadoop 배포의 골칫거리를 해결하고 가치 실현 시간을 단축하는 동시에 탁월한 성능을 제공합니다. Hadoop을 Nutanix 엔터프라이즈 클라우드의 일부로 가상화하면 다음과 같은 이점을 얻을 수 있습니다.

- **Hadoop을 다른 애플처럼 관리.** 원클릭 관리로 인프라 관리와 업그레이드를 간소화합니다.
- **하이퍼바이저 비용 절감.** Nutanix에는 AHV가 추가 비용 없이 포함됩니다.
- **가용성 및 보안 개선.** VM-HA 및 자동화된 STIG(Security Technical Implementation Guide)는 데이터의 가용성과 보안을 보장합니다.
- **데이터 로컬리티 활용.** Nutanix는 데이터 로컬리티를 제공하는 유일한 HCI 공급업체입니다.
- **하드웨어 활용률 개선.** Nutanix는 높은 CPU 활용률과 탁월한 유연성을 제공합니다.
- **Hadoop의 경제성 변화.** 다운타임과 활용도가 낮은 하드웨어로 인해 빅데이터 프로젝트가 위험해질 수 있습니다. Hadoop의 가상화는 경제성을 변화시킵니다.
- **데이터 효율성 증대.** VM 또는 파일 수준에서 압축을 활성화합니다.
- **자동 레벨 설정 및 자동 아카이브의 이점.** 데이터는 클러스터 전체에 고르게 분산됩니다. 콜드 데이터는 컴퓨팅 노드에서 콜드 스토리지 노드로 자동으로 이동합니다.

# Nutanix 기반 NoSQL

새로운 클라우드 및 빅데이터 요구 사항을 해결하기 위해 10년이 넘는 시간 동안 다양한 NoSQL 데이터베이스가 개발되어 왔습니다. 초기에 "NoSQL"이라는 이름은 "SQL이 아님"을 의미했지만 현재는 "SQL뿐만이 아님"이라는 의미로 널리 사용되고 있습니다. NoSQL 데이터베이스는 포함된 데이터의 양에 관계없이 예측 가능한 응답 시간이 요구되는 워크로드에 적합한 경우가 많습니다. 많은 회사에서 사용자 정의 코딩된 프런트 엔드가 있는 모바일 애플리케이션의 백엔드로 NoSQL 데이터베이스를 사용합니다.

NoSQL 데이터베이스의 세부 사항은 크게 다르지만 특히 기존의 관계형 데이터베이스와 비교하면 빅데이터 애플리케이션에 적합한 여러 가지 특성을 공유합니다. 이러한 데이터베이스는 여러 서버에 걸쳐 확장되어 대량의 데이터를 처리하고 높은 가용성을 제공하도록 설계되었습니다. 대부분의 경우 "최종 일관성"(관계형 데이터베이스의 엄격한 일관성보다 가용성을 우선 순위에 둠) 및 최대 속도를 고려하여 설계되었습니다. 관계형 데이터베이스가 정형 데이터에 맞춰 설계된 경우 NoSQL 데이터베이스는 빅데이터 환경에서 보편적인 비정형 및 반정형 데이터 유형에 훨씬 적합합니다.

NoSQL 데이터베이스에는 다음과 같은 여러 범주가 있습니다.

- **키-값 쌍:** 인덱스 토큰 또는 "키"를 특정 데이터 요소 또는 "값"과 페어링합니다. 예로는 DynamoDB, Riak, Redis, Voldemort가 있습니다.
- **문서 저장소:** 일반적으로 XML 또는 JSON과 같은 반정형 형식으로 문서를 저장하고 검색합니다. 예로는 MongoDB 및 CouchDB가 있습니다.
- **열 형식:** 하나 이상의 키를 여러 값과 연결하여 키-값 패러다임을 확장합니다. 여러 열과 행으로 구성되며 각 행이 동일한 열로 구성될 필요는 없습니다. 예로는 Cassandra 및 HBase가 있습니다.
- **기타:** 계층 구조형, XML, 그래프 및 개체 데이터베이스가 포함됩니다. 예로는 LDAP, Active Directory, MarkLogic, Versant가 있습니다.



**NoSQL 데이터베이스에는 네 가지 범주가 있습니다.**

**기타:**

계층 구조형, XML, 그래프 및 개체 데이터베이스가 포함됩니다. 예: LDAP, Active Directory, MarkLogic, Versant

**열 형식:**

하나 이상의 키를 여러 값과 연결하여 키-값 패러다임을 확장합니다. 여러 열과 행으로 구성되며 각 행이 동일한 열로 구성될 필요는 없습니다. 예: Cassandra 및 HBase

**문서 저장소:**

일반적으로 XML 또는 JSON과 같은 반정형 형식으로 문서를 저장하고 검색합니다. 예: MongoDB 및 CouchDB

**키-값 쌍:**

인덱스 토큰 또는 "키"를 특정 데이터 요소 또는 "값"과 페어링합니다. 예: DynamoDB, Redis, Voldemort

**Figure 4.** Major categories of NoSQL databases.

# MongoDB

Nutanix HCI 아키텍처가 NoSQL 데이터베이스 운영을 어떻게 간소화하는지 알아보려면 탁월한 성능과 가용성을 제공하고 손쉽게 확장 가능하며 널리 사용되고 있는 문서 데이터베이스인 MongoDB 사용을 고려해 보십시오. MongoDB는 물리적 및 가상 메모리를 집중적으로 사용합니다. 각 Nutanix 노드의 로컬 플래시 SSD는 페이지 오류의 영향을 최소화하고 급격히 증가하는 쓰기 활동과 대규모 작업 세트 크기를 수용합니다.

수직 확장(또는 스케일 업)은 RAM 증가, SSD 기반 스토리지 레이어 증가, CPU 추가 등과 같은 개별 VM의 하드웨어 기능을 업그레이드하여 성능을 향상시킵니다.

Nutanix를 사용하면 MongoDB 배포를 소규모로 시작한 후 수직 및 수평으로 확장할 수 있습니다.

- **수직 확장**(스케일 업)은 MongoDB VM에서 사용할 수 있는 CPU, RAM 또는 스토리지 등의 리소스 규모를 확대합니다.
- **수평 확장**(스케일 아웃)은 MongoDB 워크로드를 지원하기 위해 리소스를 추가할 때마다 VM을 추가합니다. MongoDB는 분할을 사용하여 스케일 아웃합니다. 분산을 사용하는 이 접근 방식은 클러스터 전체에 부하 및 리소스 공유를 분산시킵니다.

Nutanix HCI는 다음을 통해 전체 스토리지 비용도 절감합니다.

- 데이터 액세스 패턴을 지속적으로 모니터링하고 가장 적합한 스토리지 미디어에 데이터를 배치하여 **투명한 데이터 계층**을 구현합니다. 계층은 별도의 관리가 필요하지 않습니다.
- 압축을 사용해 **데이터 효율성**을 높입니다.

MongoDB 배포 시 Nutanix 아키텍처에서 기본 제공하는 데이터 보호 및 비즈니스 연속성 기능의 이점도 누릴 수 있습니다.

## 자세히 알아보기

- **BP-2023: 가상화 모범 사례 Nutanix 기반 MongoDB**

# 동일한 Nutanix 클러스터에서 NoSQL 및 SQL 데이터베이스를 실행하는 백화점

한국에서 가장 큰 백화점인 신세계는 IT 인프라를 통합하고 최신화해야 했습니다. 여러 시스템이 복잡하게 섞여 사용되고 있었으므로 관리와 문제 해결이 어려웠고 IT 환경은 신뢰성 문제를 겪고 있었습니다.

이제 신세계는 Nutanix 엔터프라이즈 클라우드를 통해 관계형 데이터베이스와 NoSQL 데이터베이스가 혼합된 다양한 워크로드를 실행하고 있습니다. MongoDB는 대규모의 Microsoft SQL Server 데이터베이스, 5만 명의 사용자를 위한 Microsoft Exchange 배포 및 기타 여러 일반적인 서버 워크로드와 동일한 클러스터에서 실행됩니다. Nutanix에서 표준화하면 다음과 같은 상당한 이점을 얻을 수 있습니다.

- 크게 향상된 성능
- 자본비용(CapEx) 및 운영비용(OpEx) 감소
- 빠르고 쉬운 배포 및 업그레이드
- 보다 정확한 용량 계획
- 라이선싱 비용이 추가 부과되지 않는 간소화된 가상화 환경

## Hadoop과 NoSQL 비교

빅데이터 프로젝트에 Hadoop과 NoSQL 중 어떤 것을 선택해야 할까요? 두 데이터베이스에는 여러 공통된 속성이 있습니다. 둘 다 스케일 아웃 아키텍처를 사용하고 고가용성을 실현하며 대량의 데이터와 다양한 데이터 형식을 잘 처리합니다. 그리고 이 두 데이터베이스는 함께 사용 가능합니다. 일부 NoSQL 데이터베이스는 Hadoop 프레임워크에 통합될 수 있으며 다양한 빅데이터 사용 사례를 지원하기 위해 두 데이터베이스를 모두 배포하는 조직이 점점 늘어나고 있습니다. Hadoop은 일괄 처리 작업 모드에서 대량의 데이터를 동시에 처리하고 분석하기 위해 가장 많이 사용됩니다. 작업을 제출하고 결과가 계산되는 데 몇 분 또는 몇 시간이 걸릴 수 있습니다.

NoSQL은 실시간 분석, 부정 행위 탐지, 침입 탐지, 기타 스트리밍 애플리케이션을 포함하여 데이터를 실시간 및 대화형으로 활용하는 데 더 적합합니다.

“Nutanix 덕분에 중요한 IT 서비스에 적합한 관리 용이성과 고성능을 실현했습니다. 또한 **Nutanix 플랫폼의 확장성 측면은 타의 추종을 불허합니다.** POC를 진행하는 동안 Nutanix 솔루션은 구현하기가 더 쉬웠고 **우리가 평가한 다른 모든 제품을 능가하는 성능을 보였습니다.**”

— 현윤섭, 신세계 설계팀 관리자



# 빅데이터에 **Nutanix HCI**를 사용해야 하는 이유

Nutanix HCI 환경은 프로세서, 메모리, 로컬 스토리지를 각각 포함하는 여러 고성능 서버(노드)의 스케일 아웃 클러스터로 구성됩니다. 하이브리드 클러스터에서 각 노드에는 성능 및 용량에 적합한 플래시 SSD 및 하드 디스크 드라이브(HDD)의 조합이 포함됩니다. All-Flash 및 NVMe 클러스터는 가장 엄격한 실시간 및 스트리밍 데이터 프로젝트를 지원하는 데 사용할 수 있습니다.

클러스터는 3개의 노드에서 매우 많은 수의 노드로 확장할 수 있습니다. 가상 머신 외에도 각 Nutanix 노드는 해당 노드의 드라이브에 대한 스토리지 컨트롤러 역할을 하는 특수 컨트롤러 VM(CVM)을 실행하여 로컬로 실행되는 VM과 다른 노드에 데이터 서비스를 제공합니다.

Nutanix는 Nutanix 아키텍처와 밀접하게 통합되고 Nutanix 라이선스에 포함된 기본 하이퍼바이저인 Nutanix AHV는 물론 VMware vSphere 및 Microsoft Hyper-V를 지원합니다.

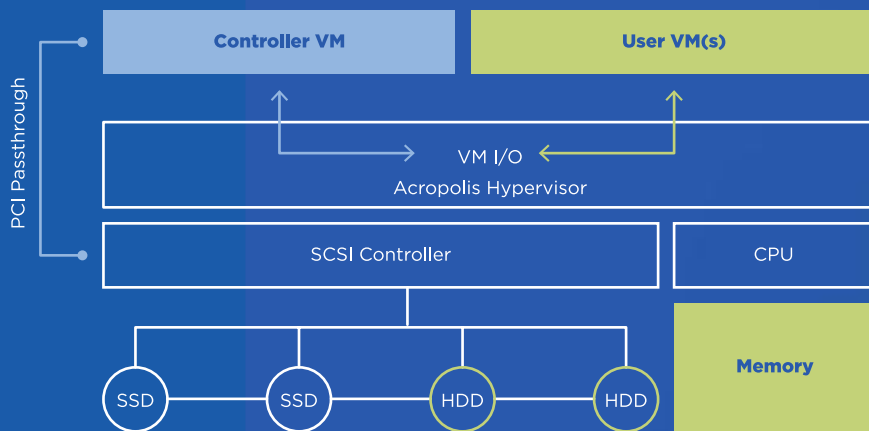


그림 5. Nutanix HCI 노드 아키텍처(하이브리드 HDD/SSD 노드)

Nutanix는 유연한 소프트웨어 라이선싱 모델을 사용하므로 Nutanix NX 어플라이언스 외에 주요 하드웨어 공급업체의 하드웨어를 선택할 수 있습니다.

Nutanix 가상화에는 리소스를 신속하게 배포 및 재배포할 수 있는 기능이 포함되어 있어 빅데이터의 요구 사항에 가장 잘 부합합니다. 특정 인프라 구성이 최적의 상태가 아닌 경우 물리적 변경 없이 손쉽게 조정할 수 있고, 다양한 노드 수와 스토리지 구성을 실험하여 각각의 적합성을 확인할 수 있습니다. 일부 Nutanix 고객은 주간에는 VDI를, 야간에는 동일한 인프라에서 Hadoop을 실행합니다.

## 분산 스토리지 패브릭

Nutanix가 빅데이터 애플리케이션에 제공하는 여러 이점은 Nutanix 분산 스토리지 패브릭(DSF)에서 비롯됩니다. DSF는 클러스터의 모든 노드에 있는 로컬 스토리지를 통합된 풀로 가상화합니다. DSF는 모든 노드의 로컬 SSD 및 HDD(있는 경우)를 사용합니다.

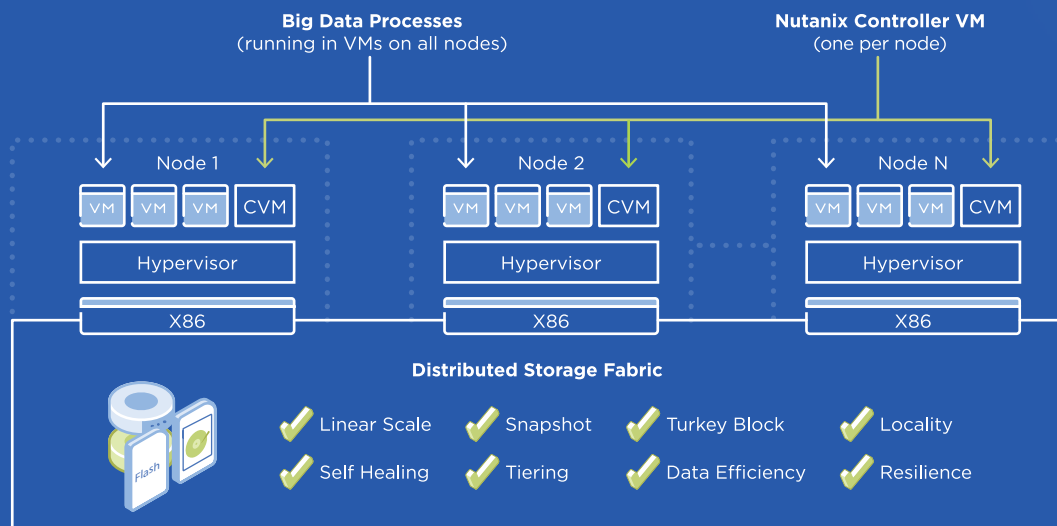


그림 6. Nutanix 분산 스토리지 패브릭의 아키텍처 및 기능



DSF는 빅데이터 환경에 매우 중요한 여러 기능을 제공합니다.

- **자가 회복.** 디스크 또는 노드에 장애가 발생한 경우 추가적인 장애로부터 보호하기 위해 전체 데이터 복구 기능이 빠르게 자동으로 복원됩니다. 필요한 경우 다른 노드에서 VM이 다시 시작됩니다. 대규모 Nutanix 클러스터는 전체 4노드 엔클로저(블록이라고 함) 장애도 극복할 수 있습니다. 빅데이터 애플리케이션을 여러 클러스터에 걸쳐 설계함으로써 전체 랙 장애도 운영을 중단하지 않고 지속할 수 있습니다.
- **데이터 로컬리티.** 각 가상 머신에서 사용되는 데이터는 우선적으로 가상 머신이 실행 중인 노드의 로컬 스토리지에 보관됩니다. 이는 관리 및 가용성 문제없이 내부 스토리지가 있는 베어메탈 서버와 유사한 성능을 제공합니다.
- **자동 계층.** 핫 데이터는 우선적으로 플래시에 저장되고 콜드 데이터(있는 경우)는 HDD 계층에 저장됩니다. 성능을 최적화하기 위해 데이터는 액세스 패턴에 따라 계층 간에 자동으로 이동합니다.
- **자동 조정.** Nutanix는 여러 워크로드 및 서로 다른 워크로드가 동시에 실행되는 경우에도 지속적인 성능 조정 없이 임의 및 순차적 I/O 모두에 대해 탁월한 성능을 실현합니다.

- **자동 레벨 설정.** 고급 알고리즘을 사용하여 여러 노드 간에 데이터의 균형을 맞춥니다. 모든 가상 머신의 데이터는 필요 시 다른 노드의 스토리지를 활용하여 단일 노드의 스토리지 부족을 방지하고 데이터 관리를 간소화할 수 있습니다. 특정 노드에 로컬 SSD보다 더 많은 핫 데이터가 있으면 다른 노드에서 사용 가능한 SSD 용량이 사용됩니다.
- **자동 아카이빙.** Nutanix 클러스터에는 용량 전용 노드가 포함되어 있어 사용 가능한 스토리지를 늘릴 수 있습니다. 이러한 노드가 배포되면 콜드 데이터가 우선적으로 유입됩니다. 이러한 용량 노드는 빅데이터 애플리케이션을 실행하지 않기 때문에 일반적으로 이와 관련된 라이선싱도 없습니다.
- **RF(Replication Factor).** Nutanix HCI는 데이터 보호 및 가용성 유지를 위해 RF(Replication Factor)를 사용합니다. 모든 데이터에 대해 2개 또는 3개의 복제본이 클러스터 내의 서로 다른 노드에서 관리됩니다. Nutanix의 특허받은 이레이저 코딩 알고리즘인 EC-X는 RF로 인한 스토리지 오버헤드를 줄입니다. EC-X는 쓰기 콜드 데이터를 기반으로 작동하므로 WORM(write once read many) 워크로드에 가장 적합합니다. EC를 사용하면 클러스터의 사용 가능한 공간이 70%까지 증가할 수 있습니다.



# Nutanix Prism 관리



베어메탈에 빅데이터 환경을 배포하는 경우 각 서버에서 소프트웨어 업그레이드 및 변경, 드라이브에서 펌웨어 업그레이드 등 다양한 관리 업무의 부담을 안게 됩니다. 따라서 모든 워크로드를 동기화하기가 매우 어려워지고 문제 해결은 길고 골치 아픈 과정이 될 수 있습니다.

Nutanix Prism 관리 플랫폼은 인프라 관리에 사용자의 눈높이에 맞춘 단순성을 구현하며 인프라를 쉽게 가동할 수 있도록 지원합니다. VM 추가 또는 제거와 같은 인프라 변경을 신속하게 수행할 수 있습니다. Prism은 일반적인 IT 워크플로우를 간소화하여 서버, 스토리지, 데이터 보호 및 가상화 관리를 위한 단일 인터페이스를 제공합니다. Prism을 사용하면 Nutanix 솔루션의 구성, 모니터링 및 관리가 크게 간소화됩니다. 원클릭 관리 기능은 관리 부담을 줄이고 운영자가 실수할 가능성을 낮추는 동시에 계획된 다운타임도 줄여줍니다.

- **원클릭 소프트웨어 업그레이드.** 빅데이터 운영 중에 일관되게 겪고 있는 문제 중 하나가 시스템 소프트웨어 및 펌웨어를 최신 상태로 유지하는 것입니다. Prism은 업그레이드로 인한 운영 중단없이 정상 업무 시간 중에 업그레이드를 수행할 수 있도록 합니다. 지능형 소프트웨어가 힘든 일을 모두 수행하므로 상세한 초기 계획을 세울 필요가 없습니다.
- **원클릭 교정.** 경고 또는 장애 발생 시 Prism은 문제의 신속한 해결을 위해 취할 수 있는 교정 조치를 제안합니다. 원클릭 교정을 사용하면 서비스 복구 및 복원에 걸리는 평균 시간이 많이 단축되므로 가용성이 크게 향상됩니다.
- **풍부한 REST API.** Prism을 통해 수행할 수 있는 모든 작업은 REST API 또는 PowerShell cmdlet 라이브러리를 사용하여 수행할 수 있습니다.

# Nutanix 시작하기

Nutanix가 빅데이터 운영을 혁신하고 분석을 비즈니스 프로세스와 통합하는 방식에 대해 자세히 알아보려면 [nutanix.com/kr/solutions/big-data](http://nutanix.com/kr/solutions/big-data)를 참조하십시오. Nutanix([info@nutanix.com](mailto:info@nutanix.com))에 문의하거나 [트위터에서 @nutanix](#)를 팔로우하거나 [www.nutanix.com/kr/demo](http://www.nutanix.com/kr/demo)에 귀사를 위한 맞춤형 브리핑을 요청할 수 있습니다. 하드웨어, 설치 또는 비용의 부담 없이 Nutanix 인프라를 **테스트 드라이브**로 무료 체험해 볼 수도 있습니다. 쉽게 따라할 수 있는 가이드 안내를 통해 퍼블릭 클라우드의 단순성과 민첩성 그리고 온프레미스의 탁월한 성능, 보안 및 제어 기능을 경험해 보십시오.

테스트 드라이브 시작

