Hewlett Packard Enterprise

올인원 빅데이터 플랫폼 어플라이언스

HPE Ezmeral Data Fabric Appliance

INDEX

Data Challenges 03

혁신을 위한 빅데이터 플랫폼, HPE Ezmeral Data Fabric 04

HPE Ezmeral Data Fabric의 기술적 장점 05

여러 분산파일시스템 공인성능수치 비교 09

성능과 안정성을 한번에 잡은 빅데이터를 위한 최적의 플랫폼, HPE Synergy 10

빅데이터를 위한 최고의 선택 – HPE Ezmeral Data Fabric Appliance 11

Data Challenges

현대의 IT 환경에서는 개별 사용자, 스마트 앱, 센서, IoT 장치 등 수많은 곳에서 대량의 데이터가 생성되고 있습니다. 기업들은 이런 다양한 데이터를 통해 비지니스 인사이트를 수집, 저장 및 생성하기 위해 많은 노력을 하고 있지만, 비즈니스에 이 데이터를 제대로 활용하는 것은 매우 어려운 일입니다. 로컬, 원격 및 클라우드 서비스에 걸쳐 대규모 데이터 세트와 복잡한 데이터 구조를 연결해야 하기 때문입니다.

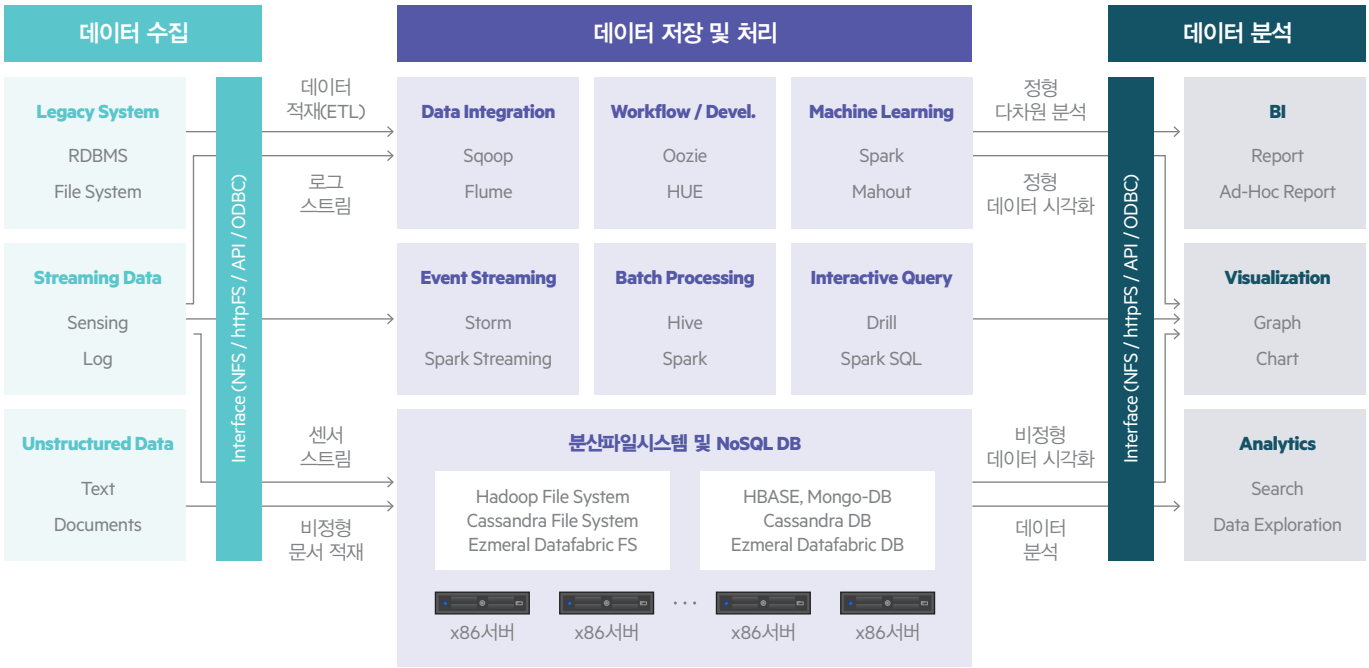
과거의 IT조직은 이를 위해 다양한 장비와 스토리지 솔루션을 구축했지만, 각 시스템들은 사일로로 구축되어 IT 환경을 복잡하게 만들고 비즈니스가 모든 데이터를 활용하는 데 제약이 있었습니다. 현재의 퍼블릭 클라우드 스토리지는 효과적일 수 있지만, 지연 시간이 많고 전송 비용이 매우 높습니다. 그래서 많은 조직이 다음과 같은 도전을 해결하기 위해 많은 노력을 하고 있습니다.

App & Data Silos | **New Technologies & Use Cases** | **Real-Time Data Access**

- 각 데이터 사일로에 대한 유지 보수 및 업데이트는 많은 비용과 시간이 소요됩니다. 어떠한 데이터가 공유되기 전에 IT는 시스템 간에 이 데이터들을 매뉴얼로 이동해야 합니다.
- 기존 레거시 솔루션은 디지털 혁신을 위해 새롭게 탄생하는 새로운 기술을 채택하는데 매우 보수적이며 느립니다. 이러한 새로운 도전이 그들의 솔루션의 경쟁력을 위험에 빠뜨릴 수 있다고 생각하기 때문입니다.
- 오늘날의 연중무휴 24x7 비즈니스 환경에서는 사용자와 최신 애플리케이션 모두 데이터에 실시간으로 액세스해야 합니다.

빅데이터 플랫폼기반 데이터 처리 프로세스

빅데이터 플랫폼의 데이터 처리 프로세스는 대용량 데이터를 수집하고 저장하는 것을 시작으로 데이터 분석, 워크플로우 관리, 시각화 등이 진행됩니다. 빅데이터를 활용하기 위한, 비즈니스 인사이트 발굴을 위한 데이터 흐름은 하기와 같이 진행됩니다.



기존 오픈소스기반 분산파일시스템은 성능과 편의성에 한계가 있습니다.

조금 더 빨리

느린 성능

- 하둡 파일 시스템에서 데이터 조회시 장시간 소요 및 추가작업 부하발생
- 이로 인한 운영업무시스템 서비스 응답속도 지연현상 발생
- 데이터 저장 및 제공을 위한 플랫폼 구축에 따른 데이터 과학자의 서비스 요청에 온라인 운영시스템과 관계없이 안정적인 서비스 응답시간 제공

적시에 비즈니스 인사이트(Business Insight) 발굴 대응을 위한 빠른 데이터 조회 및 서비스 제공

보다 손쉬운 관리

관리 어려움

- 오픈소스 컴포넌트의 결합으로 인해 빅데이터 플랫폼의 복잡도가 증가
- 수많은 오픈소스 모듈, 어려운 기술, 그리고 수많은 관리 포인트
- 즉, 하둡 전문가만이 시스템 운영 및 개발 가능

다양한 오픈소스 모듈관리에 대한 보다 손쉬운 관리를 위한 플랫폼 구현

HPE 어플라이언스는 분산파일시스템 및 NoSQL DB 영역에 해당 되는 범위에서 고객은 데이터 처리와 분석에 집중할 수 있도록 심플한 구성과 뛰어난 성능을 제공하는 인프라를 제공합니다.

혁신을 위한 빅데이터 플랫폼, HPE Ezmeral Data Fabric

HPE Ezmeral Data Fabric은 여러분의 기대 수준을 만족할 유일한 솔루션입니다.

- 데이터 이동을 최소화한 단일 제어 플랫폼
- 신뢰성 높은 데이터에 대한 실시간 접근
- 통합된 대쉬보드, 분석툴과 실질적인 인사이트
- 강력한 암호화 기술과 접근 권한 제어 기능

비즈니스를 주도적으로 이끌어가는 데이터 관리 솔루션

HPE 에즈메랄 데이터 패브릭: Software Defined Data Store

고성능 데이터 지속성

- 글로벌 네임 스페이스
- 멀티 프로토콜
- POSIX-NFS-HDFS-REST
- 엣지 | 코어 | 클라우드

스케일 최적화

- 스케일 아웃
- 콜드 | 웜 | 핫
- 하이브리드 환경 전반에 걸친 확장

탄력성와 이동성

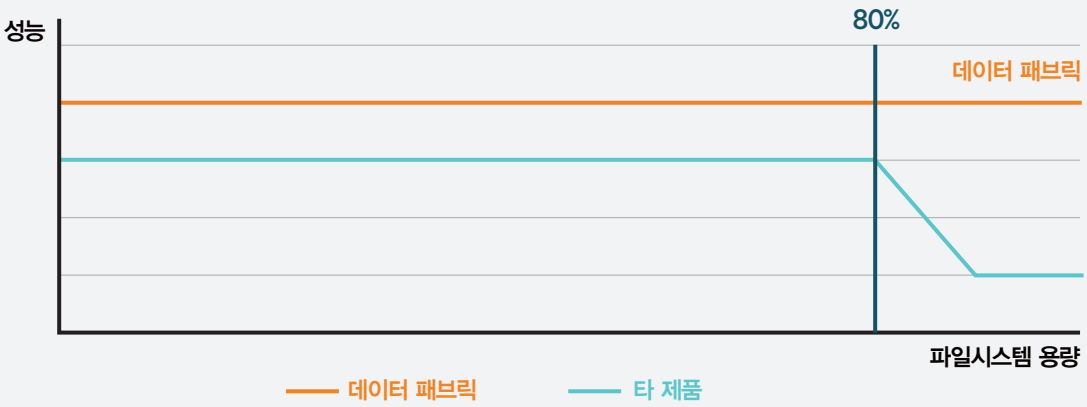
- 데이터와 앱 모빌리티
- 회복력
- 장애조치 | 복제
- 엣지 | 코어 | 클라우드

HPE Ezmeral Data Fabric의 기술적 장점

01 HPE Ezmeral Data Fabric 빅데이터 플랫폼은 타 빅데이터 플랫폼에서 제공하지 못하는 **성능과 용량 확장성, 그리고 확장시 IT 예산투자에 대한 가치를 제공**합니다.

성능 측면

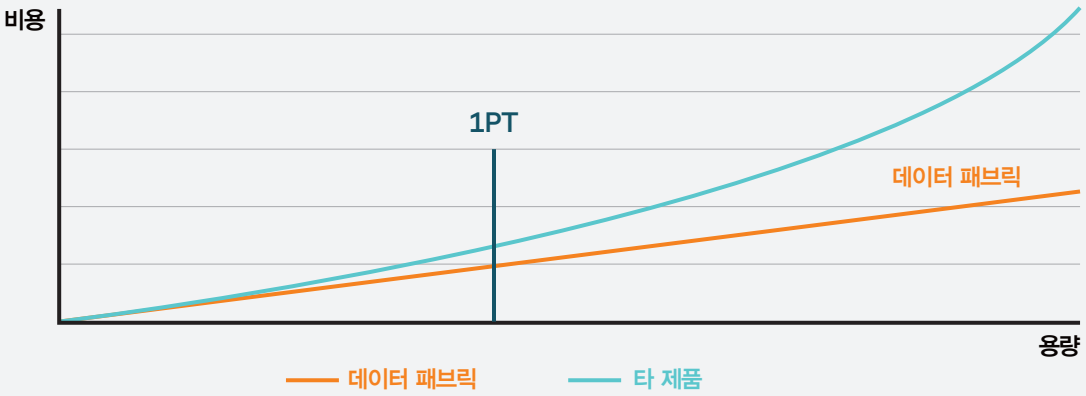
- 빅데이터 플랫폼 및 파일시스템의 경우 일정 시점의 용량 사용시(약 80% 수준) 성능저하가 발생하게 됩니다. 따라서 이를 해결하기 위해서는 데이터 노드를 추가해야 합니다.
- 하지만, 제안사의 데이터 패브릭 플랫폼은 80% 볼륨 사용시에도 동일한 성능을 제공하며, 용량 확장 시에도 선형적인 성능을 보장합니다.



〈 데이터 증가에 따른 성능 비교 〉

비용 측면

- 시스템 확장시, 데이터 패브릭 플랫폼의 경우 처음과 동일한 비용으로 서비스를 제공하기 때문에 예산 투자에 대한 비용예측이 가능합니다.



〈 시스템 확장 시 추가 비용 비교 〉

02 빅데이터 플랫폼으로 Data Fabric을 선택하는 이유는 **성능과 손쉬운 관리 그리고 기업환경에 맞는 고가용성**입니다.

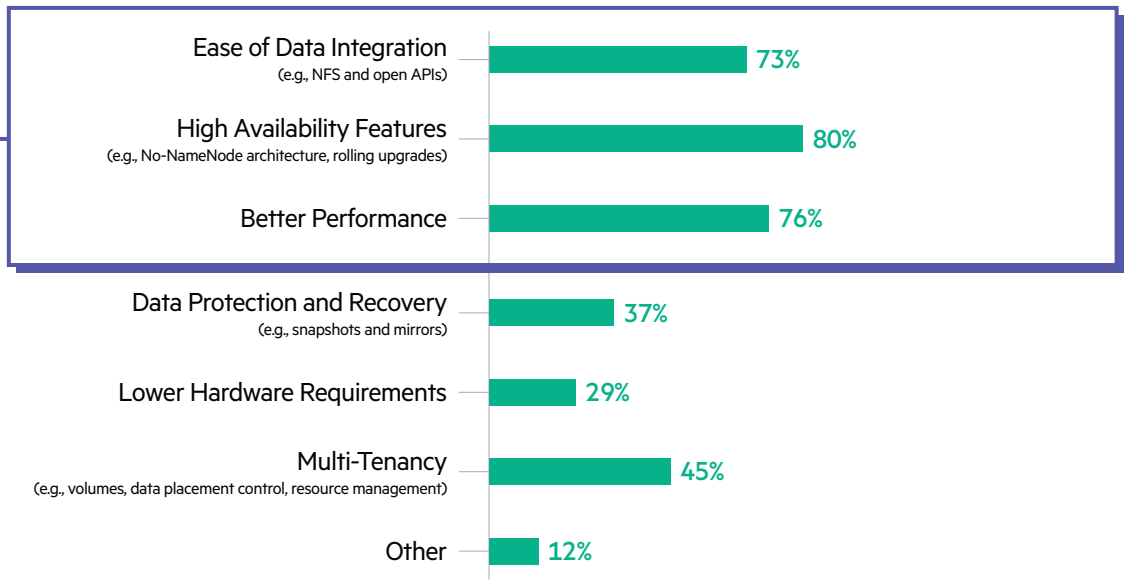
Ezmeral Data Fabric

Research by TechValidate

데이터 패브릭 선택 이유는?

Why Select "Data Fabric" over other distros?

Customers with Experience with Another Hadoop Distribution Chose the Data Fabric Distribution For these Reasons:



Note: This is multiple-choice question – response percentages may not add up to 100.

선택이유 Top 3

고가용성 기능

- 최소 네임노드 2대, 데이터 노드 8대의 일반적인 빅데이터 플랫폼 아키텍쳐가 아닌 No 노드로 CLDB를 이용한 분산배치(무 중단)
- 전 구성요소의 HA 및 스냅샷 기능 제공
- **고가용성, 실시간, 멀티 테넌시, 글로벌 네임스페이스**

뛰어난 성능

- **단순한 데이터 계층구조 및 8K IO 단위로 인한 대용량 및 수천억개의 파일에 대해서도 매우 빠른 성능 보장**
- 수조 단위까지 확장되는 제한 없는 파일 시스템(Scale-Out)

손쉬운 데이터 통합관리

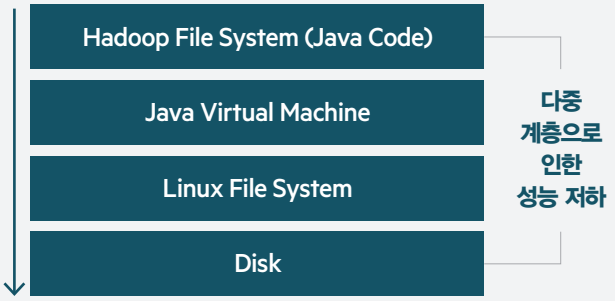
- 기본 POSIX와 100% 호환으로 스크립트 수정 불필요
- **다양한 데이터 타입 및 API 연결 제공을 통한 데이터 통합관리 수행**

※ POSIX : 서로 다른 UNIX OS의 공통 API를 정리하여 이식성이 높은 유닉스 응용 프로그램을 개발하기 위한 만든 API

03 파일 시스템 계층 간소화를 통한 **심플한 관리와 8KB IO 단위로 인한 작은 파일에서 대규모 데이터까지 고속 처리**가 가능합니다.

기존 하둡 파일 시스템

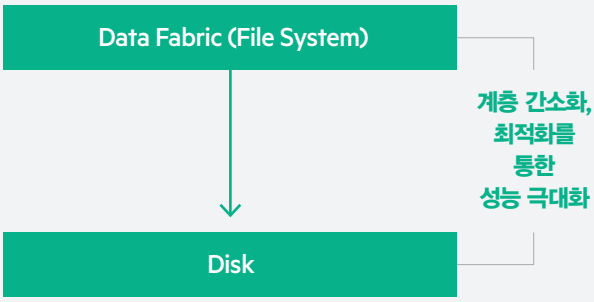
128MB 단위 IO 발생



- JVM및 OS File System의 복잡한 단계로 인해, 다중 계층으로 인한 성능 저하 요소 존재하여, 데이터가 커질 경우 이슈가 발생합니다.
- **전통적인 하둡 파일 시스템의 경우 128MB 단위로 IO가 발생되기 때문에 파일 크기가 작고, 파일 개수가 많은 경우에는 100% 성능 이슈가 발생합니다.**

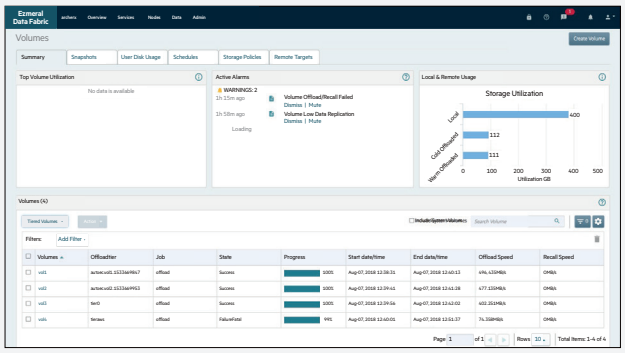
데이터 패브릭 파일 시스템

8KB 단위 IO 발생



- **대규모 분산 처리 환경에서 대용량 데이터 고속 처리**
- 수조 단위까지 확장되는 제한없는 파일 시스템(Scale-Out)
- 기존 POXIS와 100% 호환으로 Script 수정 불필요합니다.
- **다양한 데이터 포맷 지원 - HDFS API, POSIX, NFS, HBase API, JSON API, Kafaka API**

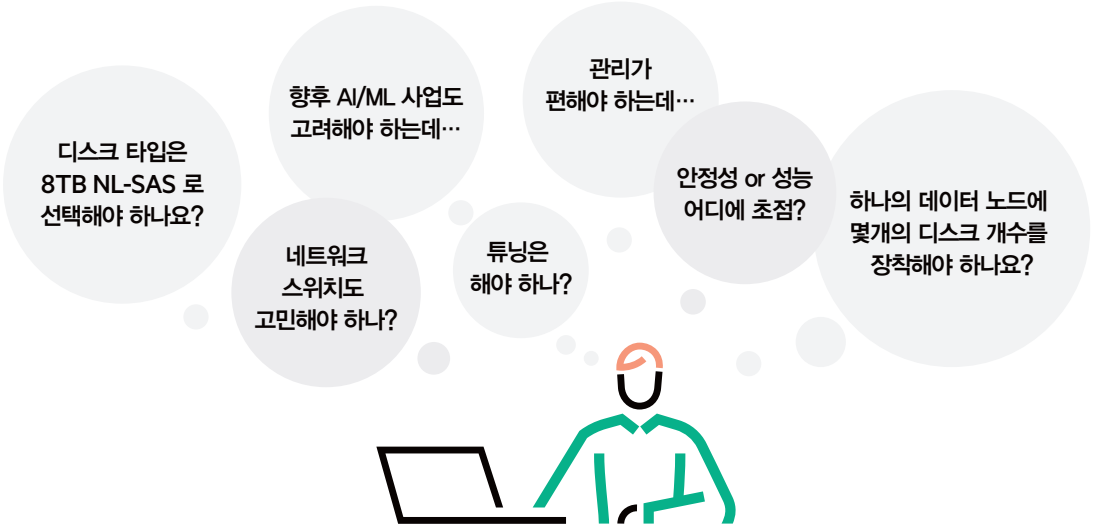
04 Command Line 기반 데이터 플랫폼 운영관리가 아닌 Data Fabric Control System
GUI 통합 운영 관리 툴을 통한 손쉬운 운영 관리 기능을 제공합니다.



- 애플리케이션 및 하드웨어 레벨까지 헬스 체크 모니터링
- 클러스터 관리
- 애플리케이션 리소스 프로비저닝
- 작업 모니터링 및 관리, 헬스 지표 관리
- 경고, 여러 인터페이스 관리

Q. 그렇다면 빅데이터 플랫폼을 위한 하드웨어는 어떻게 해야할까요?

A. 데이터 유형과 서비스 목적, 그리고 IT 예산 규모에 따라 구성하고자 하는 분산파일시스템 종류 및 NoSQL DB 솔루션에 대한 선택이 달라질 수 있습니다.



이제는 기반 인프라가 되는 **x86서버-데이터 노드와 데이터 노드간 통신을 위한 네트워크 설계**를 위한 고민을 하게 됩니다.

분산파일시스템 공인성능수치 비교

여러 분산파일시스템 공인성능수치에 답이 있습니다!

여러 분산 파일 시스템	System Configuration			Results				비고
	Workload Name	Memory (GiB)	Total Capacity	Streams	ORT (ms)	MB/s	RT(ms) per Stream	
A 솔루션	VDA	5,120	1.5 PiB	2,070	12.94	9,538	**0.0063**	8TB NL-SAS, **40Gb 네트워크**
B 솔루션	VDA	4,608	1.1 PiB	1,810	24.95	8,352	**0.0138**	8TB NL-SAS, **40Gb 네트워크**
	VDA	1,536	450.5 TiB	1,700	4.12	7,821	0.0024	**SSD 디스크 및 40Gb 네트워크**
	VDA	1,024	44.1 TiB	1,720	2.92	7,940	**0.0017**	**8TB 플래시 모듈, 100Gb 네트워크**
C 솔루션	VDA	360	3.6 PiB	3,400	50.07	15,703	0.0147	10TB NL-SAS, **100Gb 네트워크**
D 솔루션	VDA	640	36.6 TiB	800	29.98	3,691	0.0375	300GB SAS, 10Gb 네트워크
	VDA	9,144	873.4 TiB	2,520	20.34	11,231	0.0081	**3TB SSD, 40Gb 네트워크**
E 솔루션	VDA	11,712	342.73 TiB	6,800	1.56	31,375	**0.0002**	**3.84TB SSD, 100Gb 네트워크**

SPEC SFS® 2014는 **파일 서버의 성능(처리량 및 응답 시간)을 측정하기 위한 벤치마크** 입니다.
벤치마크 결과는 빅데이터 플랫폼의 성능이 시험 환경을 구성하는 서버 노드의 디스크 타입과 노드 간 네트워크 대역폭 성능에 크게 의존한다는 것을 보여줍니다.

파일시스템의 종류, 디스크 타입 및 네트워크 대역폭에 따라서도 성능차이의 격차가 크게 발생하고 있습니다.

↓

실시간 데이터 분석을 위한 빅데이터 플랫폼을 성공적으로 구현하기 위해서는 성능에 초점을 맞춰야 합니다.

노드내 디스크 타입
80TB 용량 미만으로 SSD 구성

노드간 통신을 위한 네트워크 구성
최소 40Gb~ 100Gb 네트워크 이상

성능과 안정성을 한번에 잡은 빅데이터를 위한 최적의 플랫폼, HPE Synergy

1. 컴퓨트 모듈

· Intel Xeon 전 CPU 지원 (2-4소켓) 모듈당 24개, 최대 6TB 메모리 지원

2. 네트워크 모듈

· 초고속 데이터 망: Frame간 300Gb or 600Gb

· 서버당 25/50Gb 대역폭 제공 (별도 ToR 스위치 불필요)

· 네트워크 모듈당 6개의 100G Uplink 제공

3. 통합 관리 모듈

· OneView는 REST API 기반으로 설계되어, 모든 서버, 스토리지, 네트워크 인프라를 통합 관리함

· 인프라 프로비저닝과 펌웨어 업그레이드 자동화

· 인프라 연계도(Topology) 형태로 인벤토리 조회

· 서버, 스토리지, 네트워크 모니터링 및 알람 생성

Synergy Frame

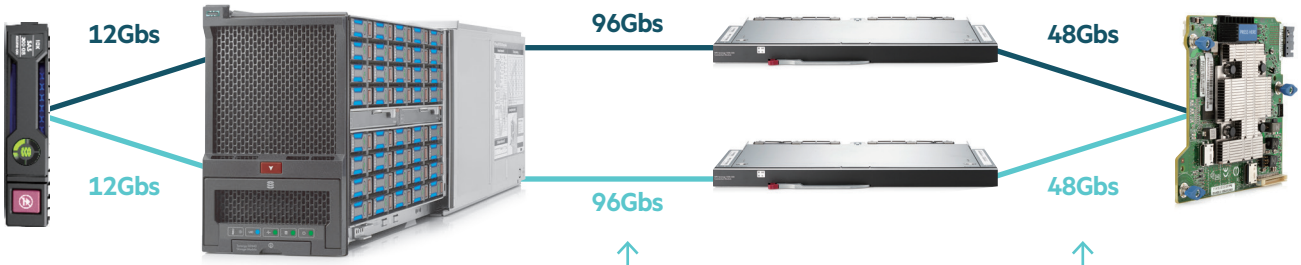
· 최대 5대의 Frame 300Gb(서버당 25Gb)

· 최대 3대의 Frame 600Gb(서버당 50Gb)

4. 스토리지 모듈

모듈 당 SFF SDD 40개, MAX 200만 IOPS+ 지원

(4K random read IO SDD 기준)



Synergy D3940
스토리지 모듈
SDD 40개

스토리지당 SAS Adapter
2개 4포트 포트당
SAS 12G x 4채널

스토리지당 SAS Adapter
1개 2포트 포트당
SAS 12G x 4채널

그래서 하나의 어플라이언스로 만들었습니다!
빅데이터를 위한 최고의 선택 – HPE Ezmeral Data Fabric Appliance

최고의 성능과 안정성을 자랑하는 HPE Ezmeral Data Fabric과 최적의 인프라 환경을 제공하는 HPE Synergy를 하나의 제품으로 통합하였습니다.

HPE Ezmeral Data Fabric 어플라이언스

빅데이터 플랫폼 구현 손쉬운 시작, 그리고 필요할때 하나씩 추가

기본 모델 | 확장 모델 | 무제한

무제한 확장 (단일 클러스터 구성 기준 2000노드 레퍼런스 존재)

구분	중형모델(Medium)		대형모델(Large)	
타입	기본모델	확장모델(증설단위)	기본모델	확장모델(증설단위)
데이터 용량(Usable)	100TB	100TB	200TB	200TB
데이터노드	8대	8대	8대	8대
디스크 수량	3.84TB SSD x 80개	3.84TB SSD x 80개	7.68TB SSD x 80개	7.68TB SSD x 80개
솔루션	Ezmeral Data Fabric 100TB 라이선스	Ezmeral Data Fabric 100TB 라이선스	Ezmeral Data Fabric 200TB 라이선스	Ezmeral Data Fabric 200TB 라이선스
기타	관리 솔루션 탑재		관리 솔루션 탑재	

구분	서비스 네트워크	데이터노드 네트워크	물리용량	가용용량	비고
중형모델(Medium)	10/25GB/s	300GB/s	300TB	100TB	
대형모델(Large)	25/40/100GB/s	300GB/s	600TB	200TB	

© Copyright 2020 Hewlett Packard Enterprise Development LP. 이 문서의 모든 내용은 통보 없이 변경될 수 있습니다. Hewlett Packard Enterprise 제품 및 서비스에 대한 보증은 해당 제품 및 서비스에 포함된 간이 보증서에 기재된 내용에 한합니다. 이 문서에 설명된 내용 중 어느 것도 추가적인 보증을 제공하지 않습니다. Hewlett Packard Enterprise는 이 문서의 기술적 또는 편집상 오류와 누락에 대해 책임지지 않습니다.