

# Optimizing Data-Centric IT Environments

Accelerate time to insights for HPC and Analytics apps

IBM Power CPU와 NVIDIA GPU가 그리는 차세대 컴퓨팅 솔루션

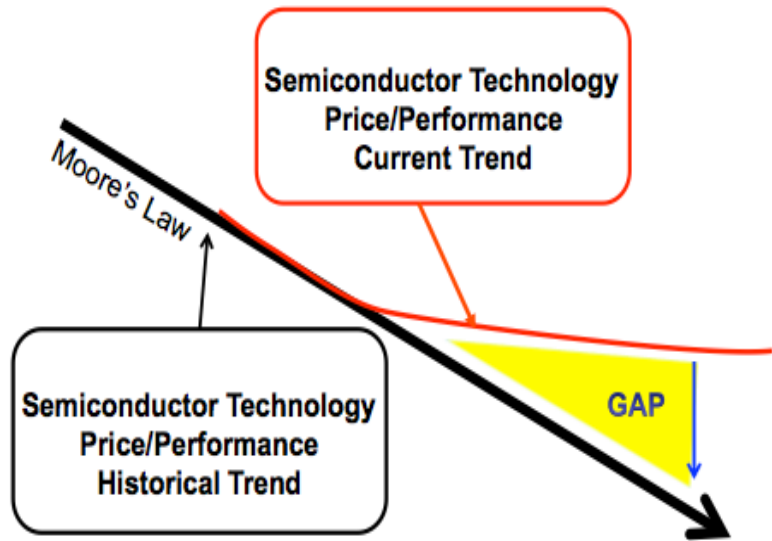
IBM 허욱 실장

Systems-Hardware 부서

(010-4995-5809, [whuh@kr.ibm.com](mailto:whuh@kr.ibm.com))



## 시장 흐름과 IBM의 HPC 및 HPA 시스템 혁신 전략




OpenPOWER™

- 지속적인 가격/성능비 향상 필요
- 데이터센터 효율화를 위해 시스템부터 소프트웨어까지 전체 스택의 혁신 유도
- 개방형 협업
- Little-endian 지원으로 어플리케이션 생태계 조성
- POWER + 가속기 조합으로 어플리케이션 성능 향상

## 차세대 컴퓨팅 솔루션을 위한 IBM의 전략적 방향

- 대규모 연산 처리 향상
  - ✓ 가속기 활용: GPU, FPGA
  - ✓ CPU와 가속기 통합: Shared coherent memory
- 가속기 활용 편이성 증대
  - ✓ CPU-GPU 프로그래밍 모델: UVM, OpenMP 4.0/OpenACC
  - ✓ CAPI-FPGA
- 대규모 데이터 처리 성능 향상
  - ✓ Move Computation to Data
  - ✓ Burst Buffer, NVRAM
  - ✓ Spectrum Scale



**OpenPOWER**  
파운데이션을 통한  
개방형 협업 및  
혁신 추구

## 차세대 컴퓨팅 솔루션을 위한 IBM의 기술 스택



- 워크로드 및 워크플로우 중심 설계
- 최고 성능을 달성하기 위한 최적화
- 밀접하게 통합된 소프트웨어 스택
- 모듈러 빌딩 블록 기반 솔루션
- OpenPOWER 기반의 혁신

# 컴퓨팅 – POWER8 프로세서

x86 대비 2배 이상의 코어 당 성능

어플리케이션 워크로드  
워크플로우 관리

소프트웨어 스택  
& 인프라스트럭처 서비스

POWER8  
서버

가속기

- Compute
- Memory, I/O

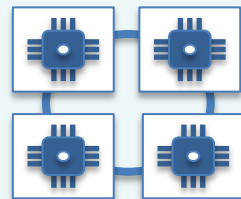
확장성 있는 네트워크

포괄적인 데이터 스토리지  
& 라이프 사이클 관리

Flash  
Disk  
Tape

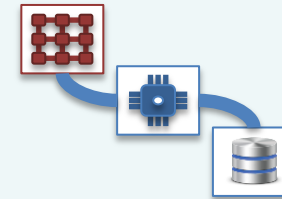
Elastic Storage  
& SDS

## 강력한 연산 능력



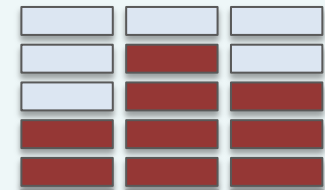
12 코어 / 소켓  
8 스레드 / 코어  
96MB 온-칩 L3 캐시 / 소켓

## 고성능 시스템 대역폭



230GB/s 메모리 대역폭 / 소켓  
96GB/s I/O 대역폭 / 소켓  
Wide SMP bus  
CAPI

## 대용량 메모리



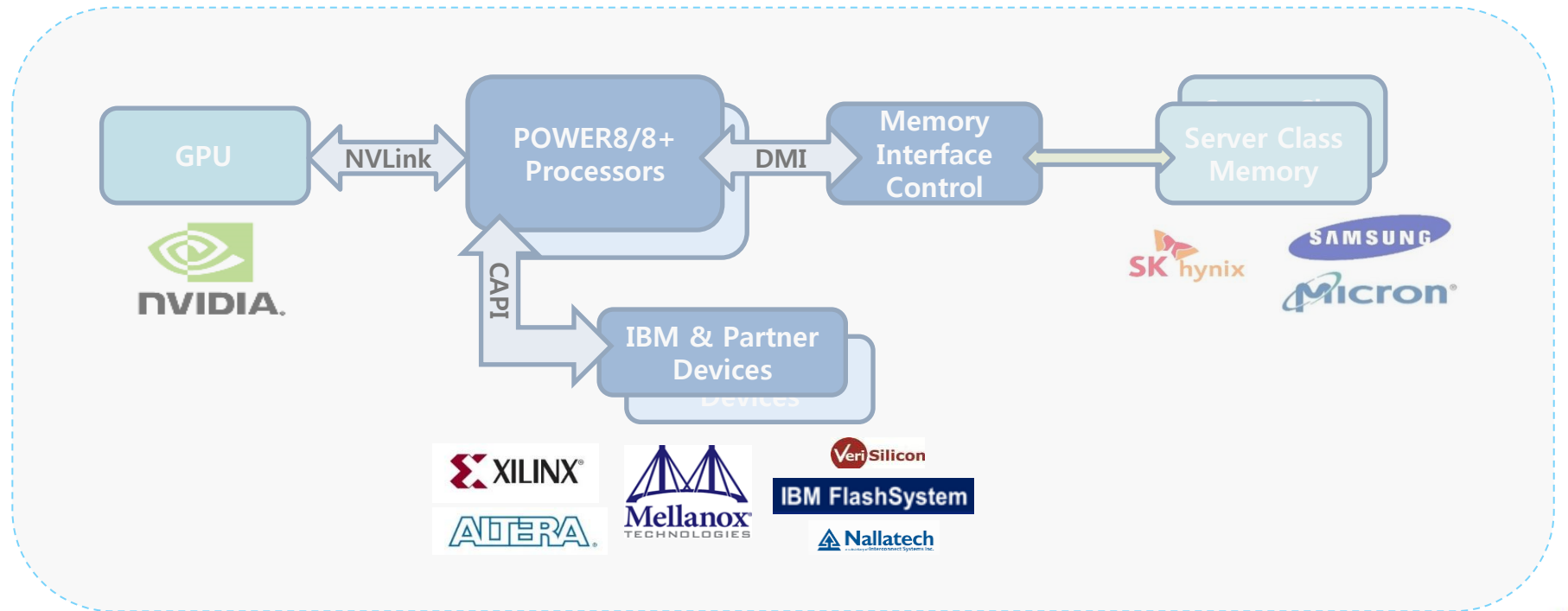
8개의 메모리 채널 / 소켓  
1TB(2TB) 메모리 / 소켓  
128MB의 버퍼 캐시 / 소켓

# 컴퓨팅 – POWER8 프로세서 & 개방형 협업

OpenPOWER를 통한 개방형 혁신



어플리케이션 워크로드 워크플로우 관리	
소프트웨어 스택 & 인프라스트럭처 서비스	
POWER8 서버	가속기 • Compute • Memory, I/O
확장성 있는 네트워크	
포괄적인 데이터 스토리지 & 라이프 사이클 관리	
Flash Disk Tape	Elastic Storage & SDS





# 컴퓨팅 – POWER 프로세서 로드맵

## 지속적인 프로세서 기술 혁신 선도

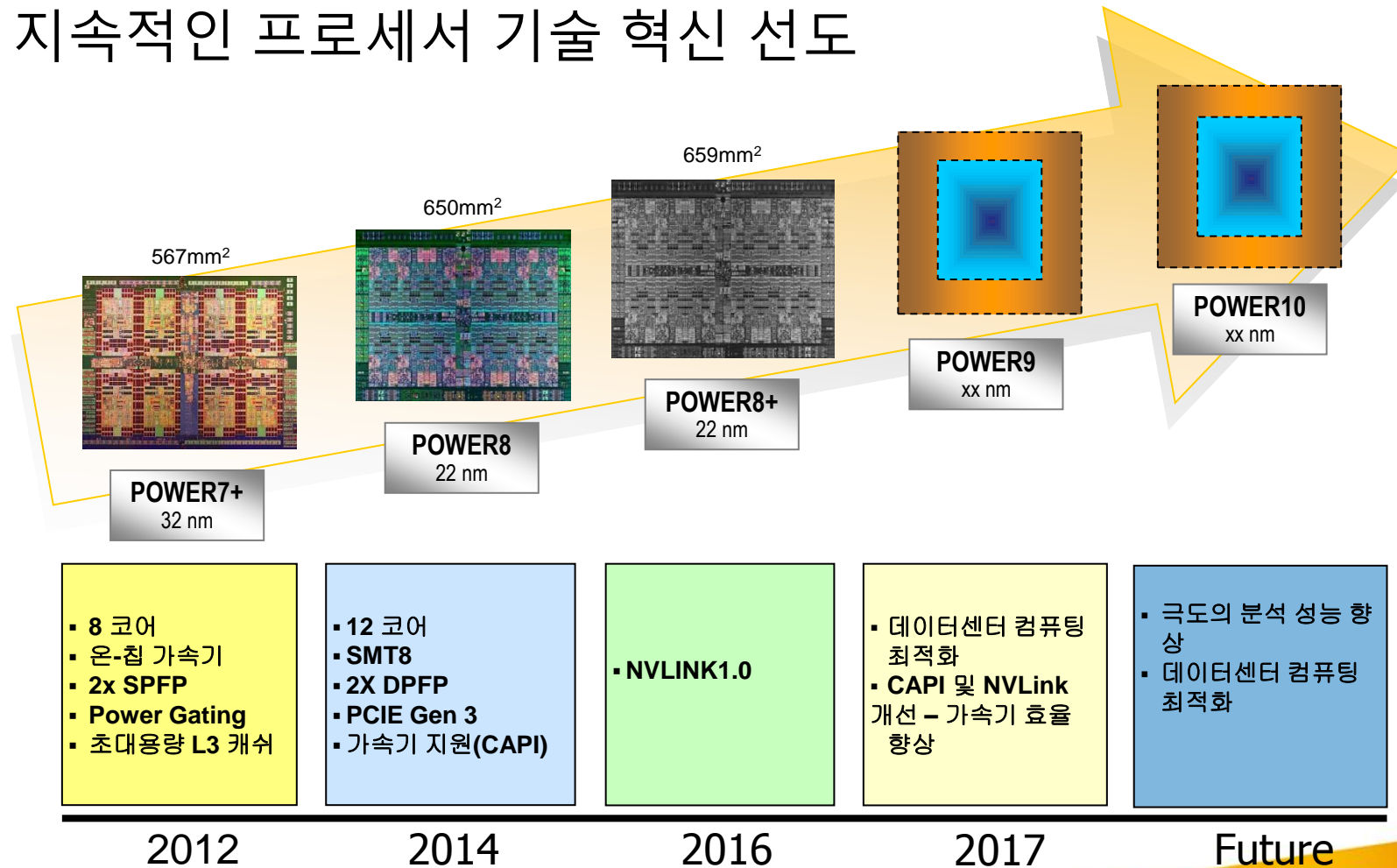
어플리케이션 워크로드 워크플로우 관리

소프트웨어 스택 & 인프라스트럭처 서비스

**POWER8 서버**  
가속기  
• Compute  
• Memory, I/O

확장성 있는 네트워크

포괄적인 데이터 스토리지 & 라이프 사이클 관리  
Flash Disk Elastic Storage & SDS  
Tape & SDS



# 컴퓨팅 – 가속기 기술 통합

## 가속기 기술을 통한 컴퓨팅 효율 극대화

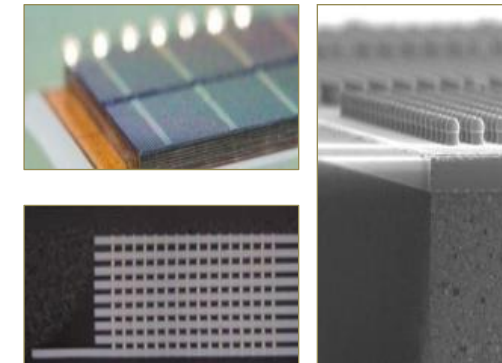
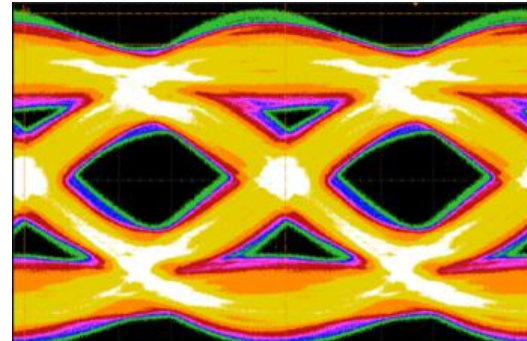


- GPU
  - ✓ 연산 가속
  
- FPGA
  - ✓ 연산 가속: Compression, Encryption, Monte Carlo, ...
  - ✓ 스토리지 I/O 가속: CAPI attached Flash
  - ✓ 네트워크 I/O 가속



# 컴퓨팅 – NVIDIA GPU

## IBM과 NVIDIA 간의 기술 협업



### NVLink

- GPU high speed interconnect
- 80-200 GB/s; 5-12X PCI-E Gen3
- POWER CPU support

### Stacked Memory

- 4x Higher Bandwidth (~1 TB/s)
- 3x Larger Capacity
- 4x More Energy Efficient per bit

# 컴퓨팅 – NVIDIA GPU

## NVLink 인터커넥트의 차별점

어플리케이션 워크로드 워크플로우 관리 S1 S2

---

소프트웨어 스택 & 인프라스트럭처 서비스

---

POWER8 서버 가속기  
• Compute  
• Memory, I/O

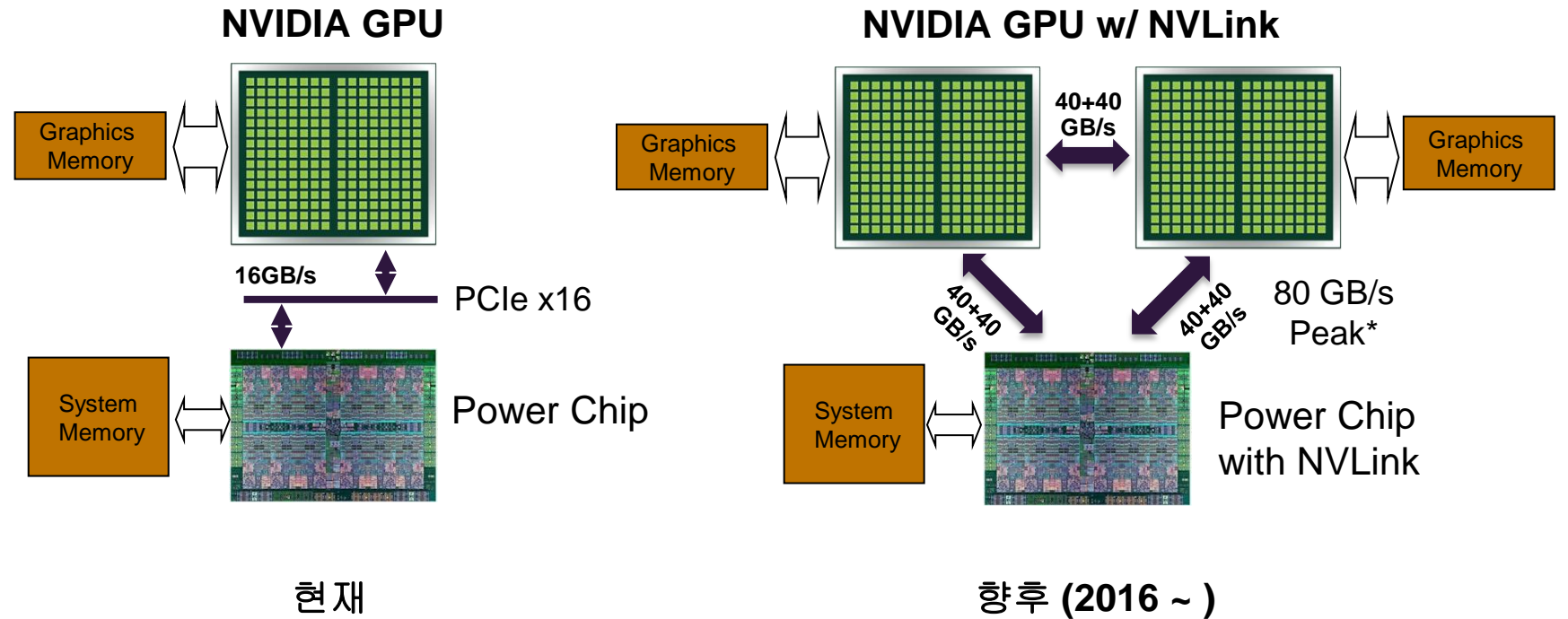
---

확장성 있는 네트워크

---

포괄적인 데이터 스토리지 & 라이프 사이클 관리

Flash Disk Elastic Storage & SDS  
Tape



# 컴퓨팅 – NVIDIA GPU

## POWER CPU와 Nvidia GPU 기반 시스템 로드맵

어플리케이션 워크로드 워크플로우 관리 S1 S2

---

소프트웨어 스택 & 인프라스트럭처 서비스

---

POWER8 서버

가속기

- Compute
- Memory, I/O

---

확장성 있는 네트워크

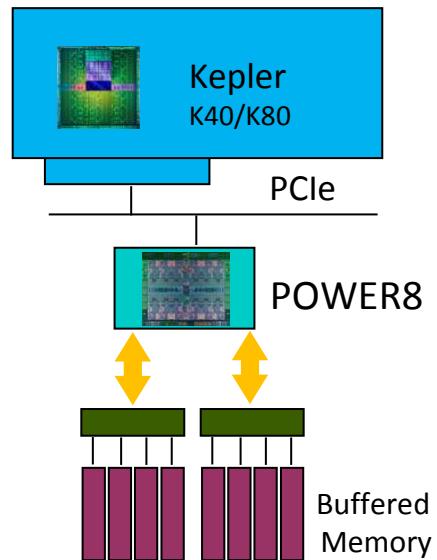
---

포괄적인 데이터 스토리지 & 라이프 사이클 관리

Flash Disk Tape

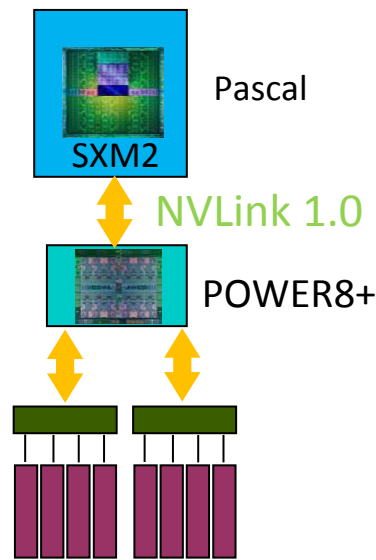
Elastic Storage & SDS

**Kepler**  
CUDA 5.5 – 7.0  
Unified Memory



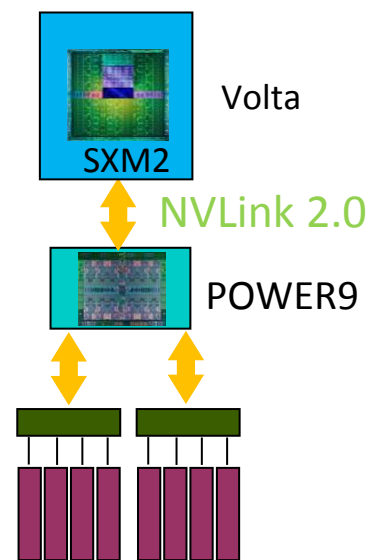
2014-2015

**Pascal**  
CUDA 8  
Full GPU Paging



2016

**Volta**  
CUDA 9  
Cache Coherent



2017

# 네트워킹 - Mellanox

## Mellanox와의 협업

어플리케이션 워크로드  
워크플로우 관리

소프트웨어 스택  
& 인프라스트럭처 서비스

POWER8  
서버

가속기

- Compute
- Memory, I/O

확장성 있는 네트워크

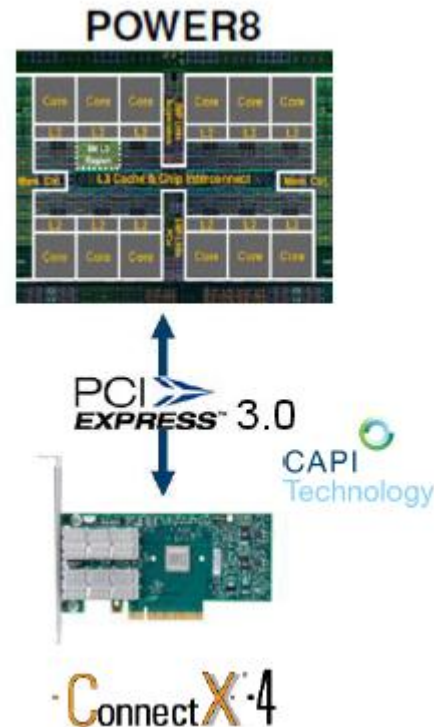
포괄적인 데이터 스토리지  
& 라이프 사이클 관리

Flash

Disk

Tape

Elastic Storage  
& SDS



- Networking evolution for system to system connection
  - ✓ High bandwidth, low latency networking Ethernet(RoCE)
  - ✓ Infiniband RDMA
- Connect-X, SwitchIB, LinkX
- CAPI

# 스토리지 – Spectrum Scale

## Spectrum Scale : Elastic Storage Server

어플리케이션 워크로드 워크플로우 관리 S1 S2

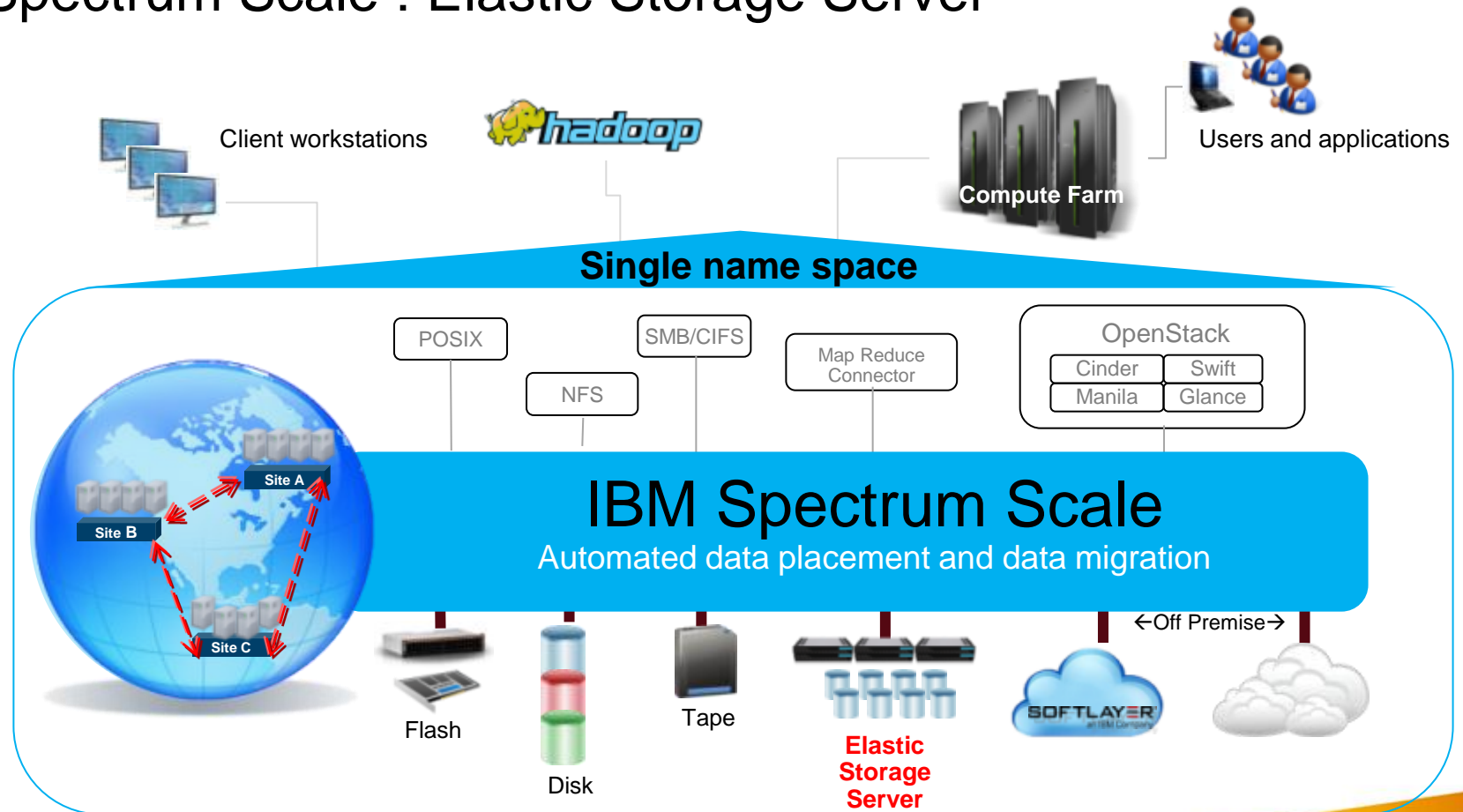
소프트웨어 스택 & 인프라스트럭처 서비스

POWER8 서버      가속기  
• Compute  
• Memory, I/O

확장성 있는 네트워크

포괄적인 데이터 스토리지 & 라이프 사이클 관리

Flash Disk Tape      Elastic Storage & SDS





# 스토리지 – Elastic Storage Server

어플리케이션 워크로드 워크플로우 관리 S1 S2

---

소프트웨어 스택 & 인프라스트럭처 서비스

---

POWER8 서버

가속기

- Compute
- Memory, I/O

---

확장성 있는 네트워크

---

포괄적인 데이터 스토리지 & 라이프 사이클 관리

Flash Disk Tape

Elastic Storage & SDS

Power S822L

JBOD Disk Enclosure  
2, 4 or 6 TB NL SAS  
1.2TB SAS or 400/800GB SSD

대용량

고성능 (IOPS)

GL product Line

**Model 22:**  
*Analytics Focus*  
2 Enclosures, 12U  
116 NL-SAS, 2 SSD  
**5 GB/Sec**

**Model 24:**  
*Analytics & Cloud*  
4 Enclosures, 20U  
232 NL-SAS, 2 SSD  
**15+ GB/Sec**

**Model 26:**  
*Petascale*  
6 Enclosures, 28U  
348 NL-SAS, 2 SSD  
**25+ GB/sec**

GS product Line

**Model 21s**  
24 SSD

**Model 22s**  
48 SAS or SSD  
12 GB/Sec

**Model 24s**  
96 SAS or SSD  
19+ GB/Sec

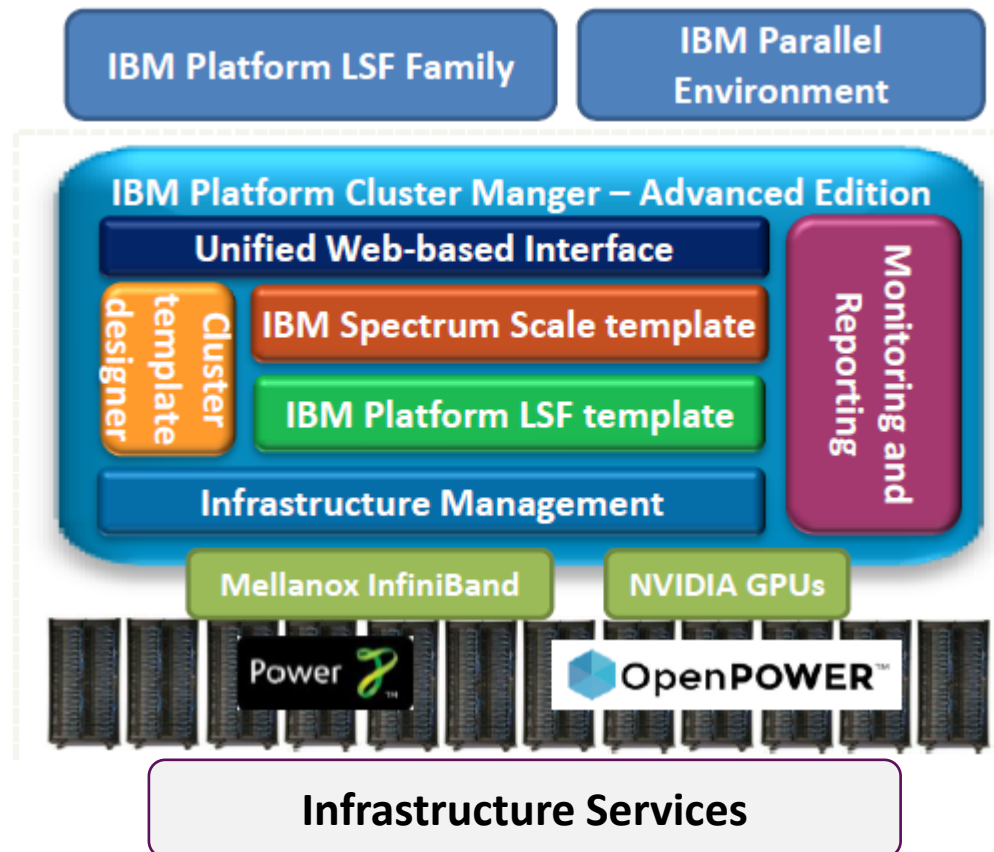
**Model 26s**  
144 SAS

- 최고의 확장성
- 최고의 성능
- 뛰어난 관리 기능
- 소프트웨어 정의



# 솔루션 스택 – IBM Platform Computing

포괄적인 HPC 및 HPA를 위한 시스템 소프트웨어 스택



Workload Management

Application Runtime

System Management

# 차세대 슈퍼컴퓨팅 시스템을 위한 여정의 시작

Oak Ridge와 Lawrence Livermore Lab을  
위한 2개의 슈퍼컴퓨터 구축 사업  
(2017년 구축 완료 예정)



현재 DoE의 최고 수준 슈퍼컴퓨터 시스템

Titan (ORNL)  
2012 - 2017



Sequoia (LLNL)  
2012 - 2017



Mira (ANL)  
2012 - 2017

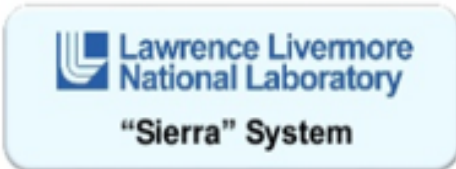


현재 시스템보다 5X – 10X 높은 어플리케이션 성능

>100 PF, >2GB/core main memory, 800 GB/node local NVRAM, ~10MW  
120 PB, 1 TB/s GPFS™ File System  
Mellanox® Dual-Rail InfiniBand, IBM POWER® CPUs, NVIDIA® Volta™ GPUs

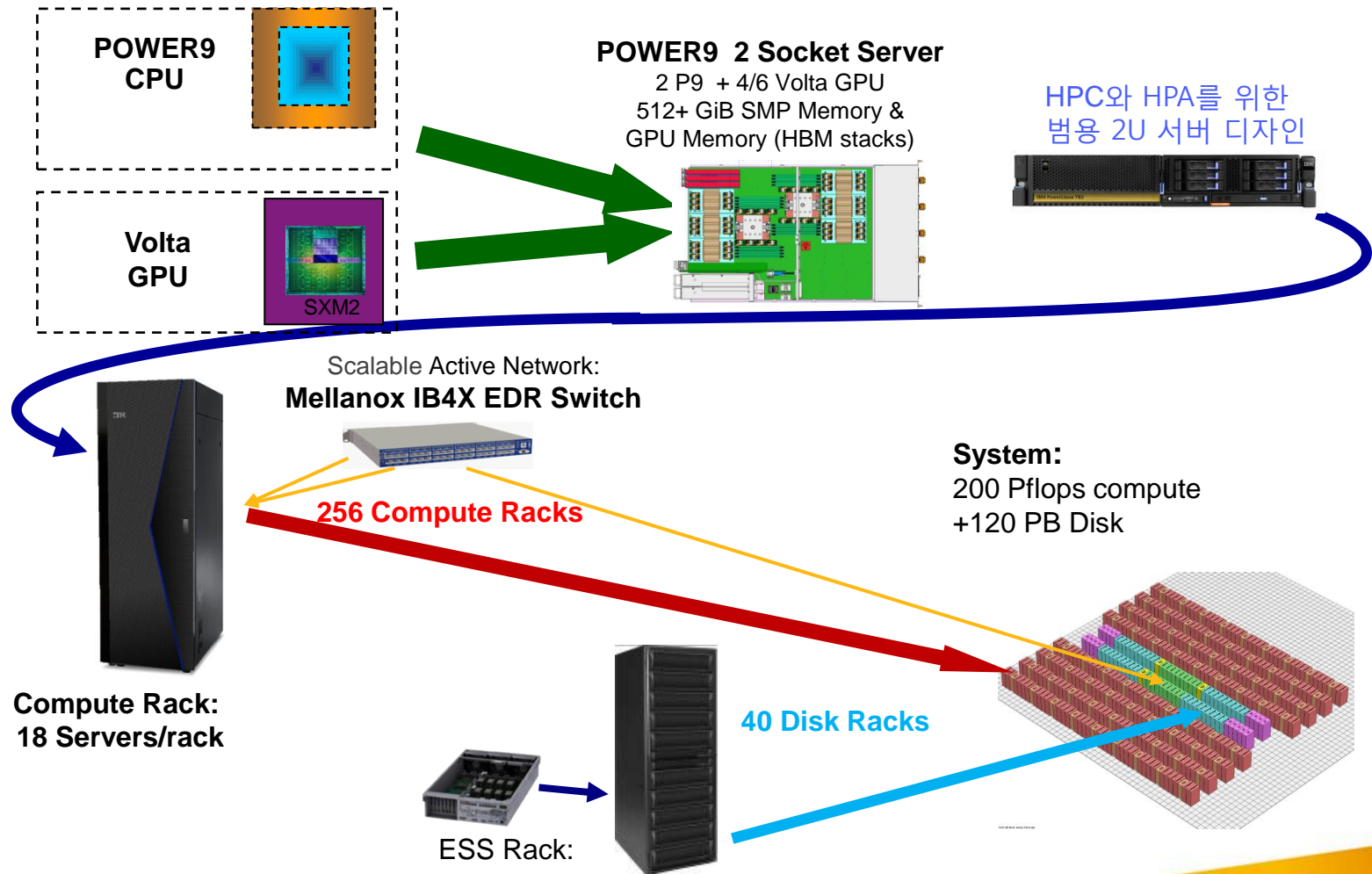


## 차세대 슈퍼컴퓨팅 시스템을 위한 여정의 시작



Feature	Summit	Titan
Application Performance	5-10x Titan	Baseline
Number of Nodes	~3,400	18,688
Node performance	> 40 TF	1.4 TF
Memory per Node	>512 GB (HBM + DDR4)	38GB (GDDR5+DDR3)
NVRAM per Node	800 GB	0
Node Interconnect	NVLink (5-12x PCIe 3)	PCIe 2
System Interconnect (node injection bandwidth)	Dual Rail EDR-IB (23 GB/s)	Gemini (6.4 GB/s)
Interconnect Topology	Non-blocking Fat Tree	3D Torus
Processors	IBM POWER9 NVIDIA Volta™	AMD Opteron™ NVIDIA Kepler™
File System	120 PB, 1 TB/s, GPFS™	32 PB, 1 TB/s, Lustre <sup>®</sup>
Peak power consumption	10 MW	9 MW

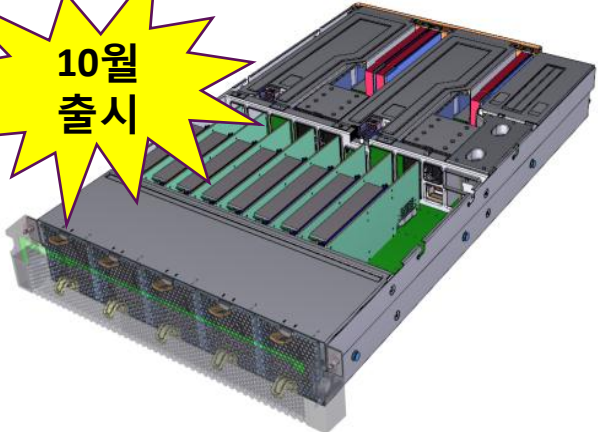
# 차세대 슈퍼컴퓨팅 시스템을 위한 여정의 시작



- Scalable system software and data architecture
- LLVM Open Source compiler
- Water cooling
- Integrated Local Active Storage



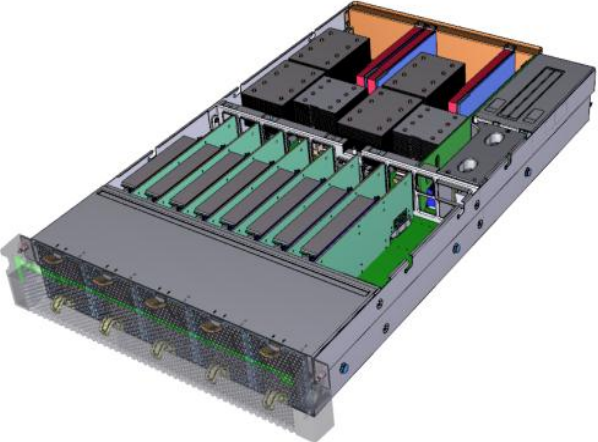
# HPC와 HPA를 위한 IBM의 서버 솔루션



**10월 출시**

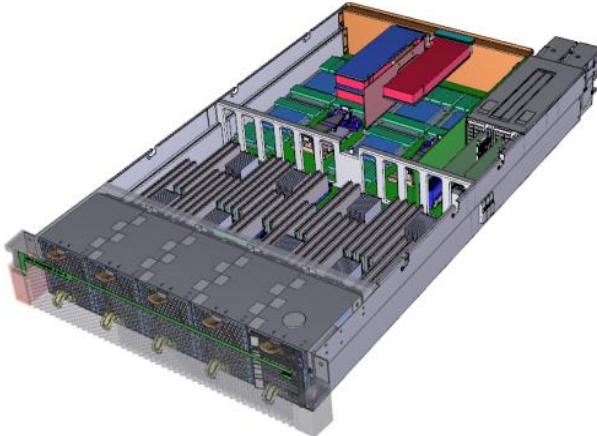
**Firestone (2015)  
Power S822LC**

- 2 x POWER8 processor
- 2 x Nvidia K80 via PCIe slot
- 2U form factor



**후속 모델 (2016)**

- 2 x POWER8+processor
- 2 ~4 x Nvidia Pascal via SXM2
- NVLink 1.0
- 2U form factor



**후속 모델 (2017)**

- 2 x POWER9 processor
- 4 ~6 x Nvidia Volta via SXM2
- NVLink 2.0
- 2U form factor

Thank You

감사합니다.