

Hewlett Packard
Enterprise

BayNex

HPE-베이넥스 서버 비즈니스 파트너를 위한 솔루션 소개

2020년 10월 13일(화) 15:00 - 16:30

베이넥스 AMD 기반의 HPC 구성사례

베이넥스 | 임영환 전무

차 례

1. HPC 및 시장 소개
2. HPC 적용 분야
3. HPC 아키텍처
4. HPC 시스템 설계
5. 파트너사 협력

1. HPC 및 시장 소개

1. HPC 및 시장 소개

HPC(High Performance Computing)의 필요성

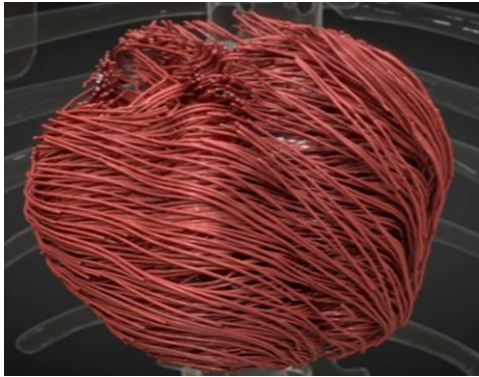
- Analysis 능력 보다 더 많은 Simulation 수행
- Simulation 증대에 따른 Data Volume의 폭발
- Grand Challenge Problem 도전
- 협업의 필요성 증대
- Power & Cooling 비용 증가
- 랙 상면 증가
- 보다 간편한 시스템 설치 & 운영 요구



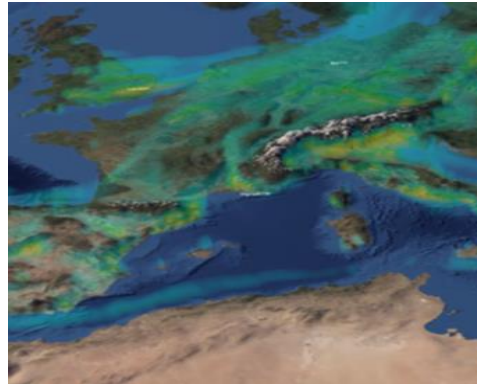
계산 성능에 대한 Demand는
계속 증가함

1. HPC 및 시장 소개

HPC는 다양한 연구 분야에 필요한 핵심 자원



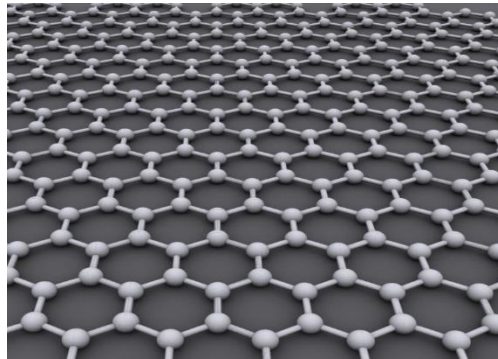
생명과학 & 제약



지구 과학



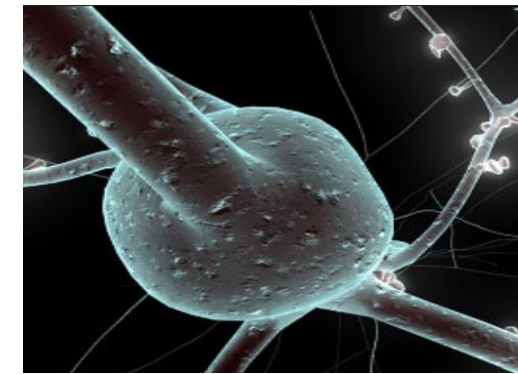
우주, 에너지, 플라즈마 물리학



신소재, 화학, 나노과학



공학



신경 과학

1. HPC 및 시장 소개

HPC 성능의 일반적 지표

(1) 이론 성능 (Rpeak)

- FLOPS (Floating Point Operation per Second)가 지표 (Hz당 배정밀도 연산 횟수)
- Intel Xeon (32 Flops/Hz) vs AMD Epyc (16 Flops/Hz)
- 예:
 - Xeon Gold 6248R (3 GHz 24 Core Processor): $3 \text{ GHz} * 32 \text{ Flops/Hz} * 24 \text{ Core} = 2,304 \text{ GFlops}$
 - AMD Epyc 7702 (2 GHz 64 core Processor): $2 \text{ GHz} * 16 \text{ Flops/Hz} * 64 \text{ core} = 2,048 \text{ GFlops}$

(2) HPL 실측 성능 (Rmax)

- 서로 다른 CPU/아키텍처와의 성능 비교를 위해 **LINPACK**이라고 부르는 역행렬 계산 프로그램에 대한 실측 성능 수치를 기준으로 상호 비교
- Rpeak 대비 Rmax가 비례하지 않음
- 실제 성능은 아키텍처 구성, Application, data 등 여러 요소에 의해 결정됨
- Memory Latency, Bandwidth, I/O Bottleneck, Communication Overhead 등에 의해 제한이 됨

1. HPC 및 시장 소개

세계 Top 5 HPC 시스템

(1) Top500 발표

- Rmax 기준 매년 2회(6월, 11월) 슈퍼컴퓨팅 컨퍼런스에서 발표
- 최근 발표(2020년 6월) Top 1 시스템:
 - 32 bits (Single Precision) 연산: 최초 1 ExaFlops (1초에 10^{18} 번(100만 조, 100 경) 연산 성능)
 - 62 bits (Double Precision) 연산: 약 537 PetaFlops (1초에 53만7천조 번 연산 성능)

(2) Top 5

(출처: www.top500.org)

순위	이름	제조사	국가	연도	총 코어	Rmax [TFlop/s]	Rpeak [TFlop/s]	프로세서
1	Fugaku	Fujitsu	일본	2020	7,299,072	415,530	513,855	Fujitsu A64FX 48C 2.2GHz
2	Summit	IBM	미국	2018	2,414,592	148,600	200,795	IBM POWER9 22C 3.07GHz
3	Sierra	IBM / NVIDIA / Mellanox	미국	2018	1,572,480	94,640	125,712	IBM POWER9 22C 3.1GHz
4	Sunway TaihuLight	NRCPC	중국	2016	10,649,600	93,015	125,436	Sunway SW26010 260C 1.45GHz
5	Tianhe-2A	NUDT	China	2018	4,981,760	61,445	100,679	Intel Xeon E5-2692v2 12C 2.2GHz

1. HPC 및 시장 소개

HPC 하드웨어의 발전

산업표준기술을 통한 구축비용의 절감이 아키텍처 변화를 이끌어 옴

ASCI White (IBM)
2000년 11월 #1 (Top500)



Peak performance: 12,288 GF
375MHz Power3, 8,192 core, 6TB
무게: 106 Tons (w/ 160 TB storage)
전력: 3MW(운영) + 3MW(냉각)

2006: 동일한 성능의
시스템이 1랙에 구현



Peak performance: 12,288 GF
무게: 100x 배 이상 감소
전력: 100x 배 이상 전력 감소
비용: 100x 배 이상 비용 감소

2020년: 2대 서버



서버 2 대

1 랙: 1백만 GF (1PF) 가능



2. HPC 적용 분야

2. HPC 적용 분야

빠른 계산 시간이 중요한 영역이 모두 HPC의 영역

Computer Aided Engineering



- Structures
- Fluids
- Impact

Life and Materials Sciences



- R & D
- Clinical Trials
- Data Analysis
- Material Sciences

Government / Classified / Defense



- Crypto
- Custom
- Classified
- Weather

Scientific Research



- Custom
- LS R&D
- Weather



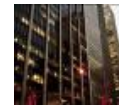
Digital Content Creation

- Rendering
- Gaming



Electronic Design Automation

- Verification
- Layout



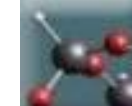
Finance/ Securities

- Financial Modeling/ Analytics



GeoSciences

- Seismic
- Reservoir Modeling



Product Lifecycle Management

- PDM
- PLM

2. HPC 적용 분야

국내 HPC 영역

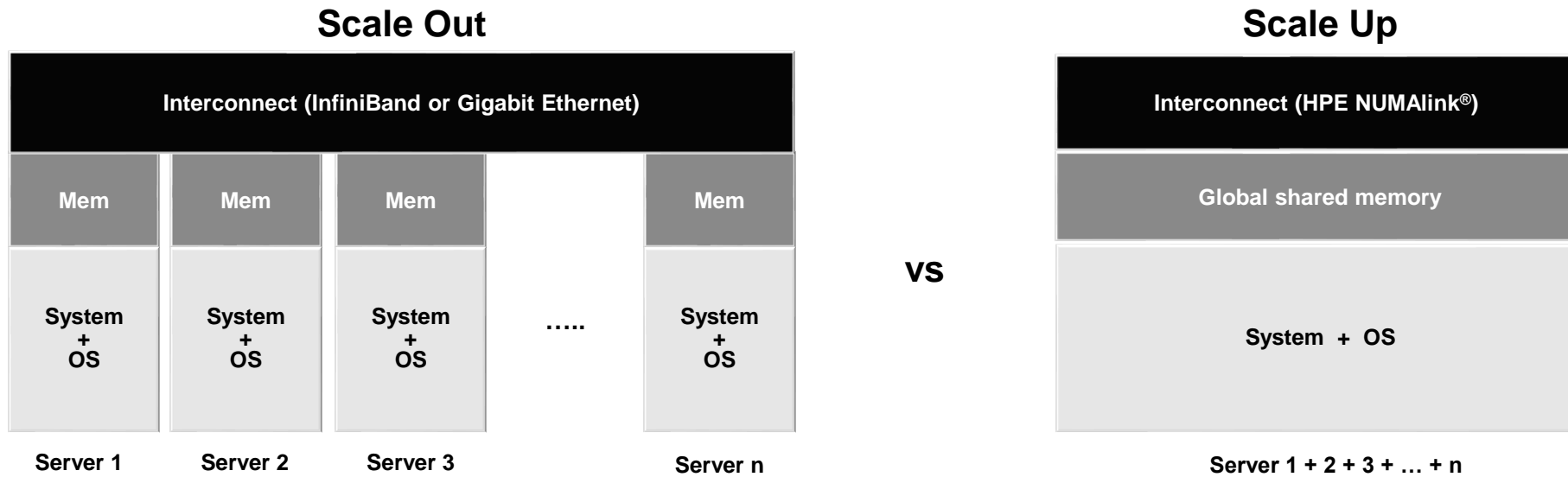
연구소	바이오/생명 공학, 신소재, 화학, 에너지 등의 기초과학 연구
항공, 우주	항공, 우주 연구 개발 프로젝트
기상, 기후	기상 예측, 기후 변화 연구
자동차	완성차 업체와 1차/2차 부품사들의 신제품 개발
중공업	선박, 플랜트 설계
전자	반도체, 휴대폰, 컴퓨터 등 신제품 개발
화학	화학 공정 시뮬레이션, 신물질 개발
금융	파생상품 거래, 알고리즘 기반 거래, Pricing, 위험관리 등
미디어	특수 효과, 애니메이션, 포스트 프로세싱 등 영상 제작

3. HPC 아키텍처

3. HPC 아키텍처

HPC 구성 아키텍처 (계산 노드 기준)

Scale Up vs. Scale Out



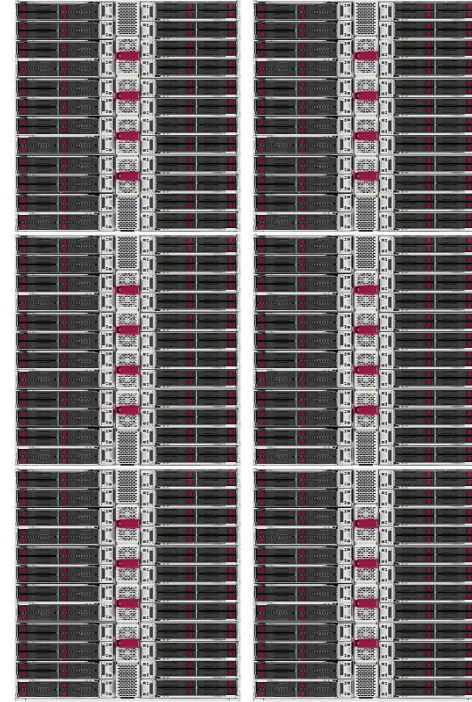
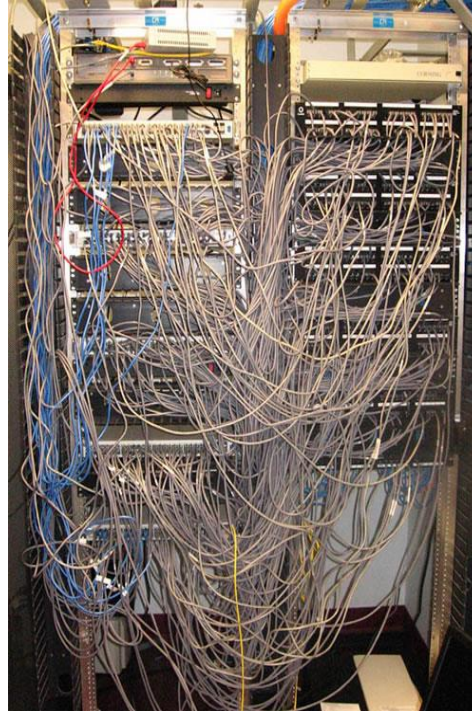
3. HPC 아키텍처

Scale Out 확장

- 네트워크를 통해 수십, 수백, 수천 노드로 연결 확장



·
·
·



3. HPC 아키텍처

Scale Up 확장

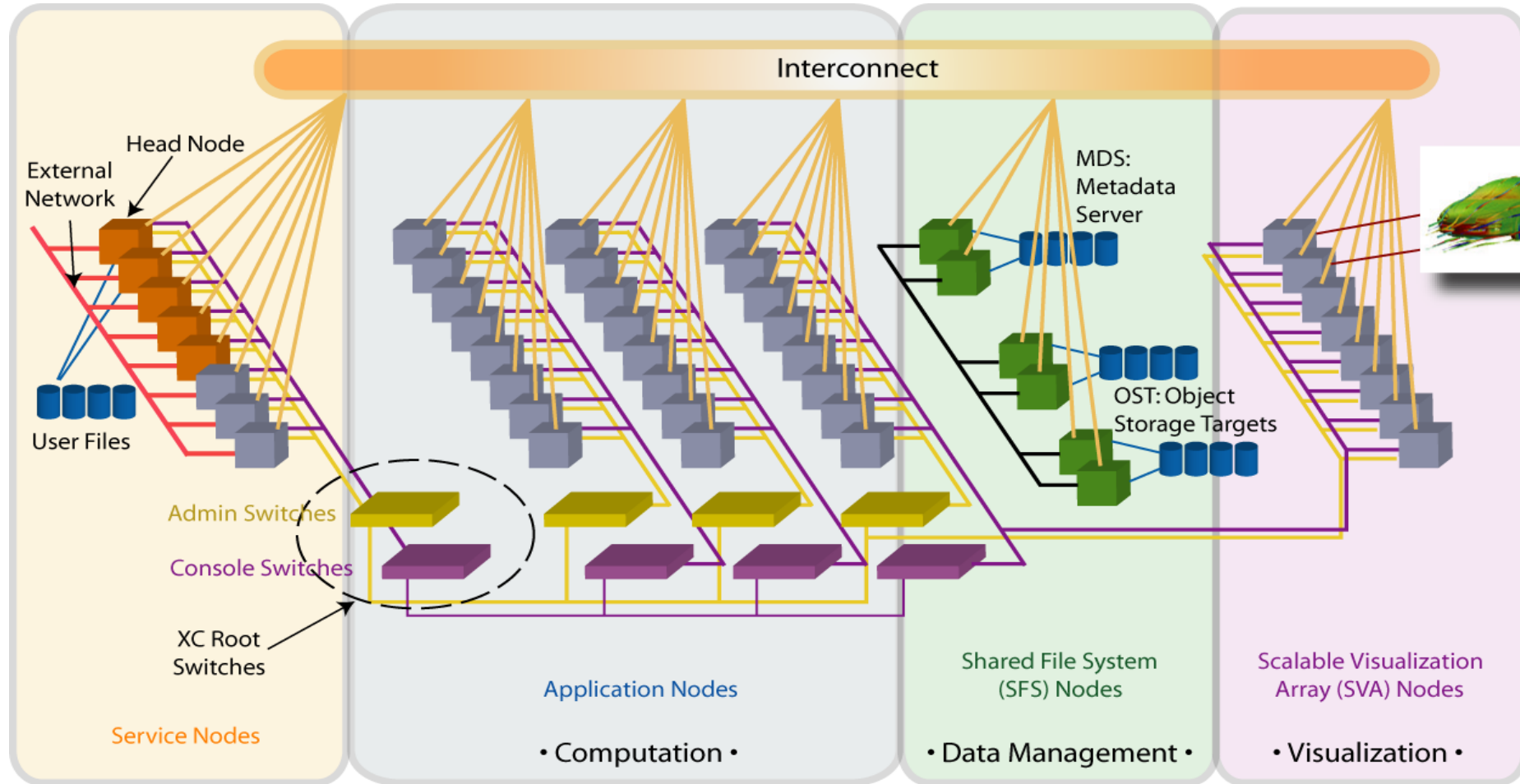
단일 서버 시스템으로서 다수의 노드를 연결 확장



확장성 제한되고, 고가의 시스템으로 매우 특정한 영역 이외에는 활용되지 않음

3. HPC 아키텍처

HPC 구성도



3. HPC 아키텍처

HPC S/W Stack

System Management	HPE Performance Cluster Management HPE Insight Cluster Management Utility		
	Bright Computing Bright Cluster Manager		
Workload Management	Altair PBS Pro, LSF Adaptive Computing Moab		
	SLURM, Open PBS		
Software Development	HPE Performance MPI, Open MPI, Intel MPI	Intel Parallel Studio XE	
	Mellanox HPC-X: OpenMPI, OpenSHMEM, Berkeley UPC, MXM, FCA NVIDIA CUDA Toolkit, OpenCL, OpenACC	PGI Compilers Rogue Wave TotalView Allinea Forge	
Visualization	NICE Software DCV, EnginFrame HPE High Performance Remote Visualization Solution		
Data Management	Tiered Data Management	Parallel File Systems	
	HPE Data Management Framework (DMF)	Lustre	Spectrum Scale (GPFS) CXFS
Fabric Software	Mellanox OFED	Intel Omni-Path Software: Intel Fabric Suite	
System Software	Commercially Supported Linux	Self-supported Linux	
	Red Hat Enterprise Linux SUSE Linux Enterprise Server	CentOS	

■ HPE developed software
 ■ Open Source software
 ■ Third party software

4. HPC 시스템 설계

HPC 시장 & 고객의 특성

- IT Driven이 아닌 **End-User Driven Market**
- HPC 시스템 구성, System & App S/W 등, **사용 목적에 대한 이해 및 지식 필요**
- 시스템 규모와 상관 없이 **기본적 HPC 구성 기술 지원 필요**
- 계산 노드, 스토리지, 인터커넥트, 아카이빙, 지원 S/W에 **대한 통합적 지식 및 경험 필요**
- 고객 사용 요건에 따른 **사전 컨설팅 중요**
- 프로젝트 **초기 협의** 및 지원 중요
- HPC User Community 형성, **지속 발전 변화 정보 공유**
- 최종 사용 목적에 맞는 **Application의 성능 중요** -> **BMT 필요성 증가**
- 각 Segment별로 **Reference 확보** 시 유사 고객에 **파급 효과** 있음
- **고객의 언어**로 말할 수 있는 것이 중요

4. HPC 시스템 설계

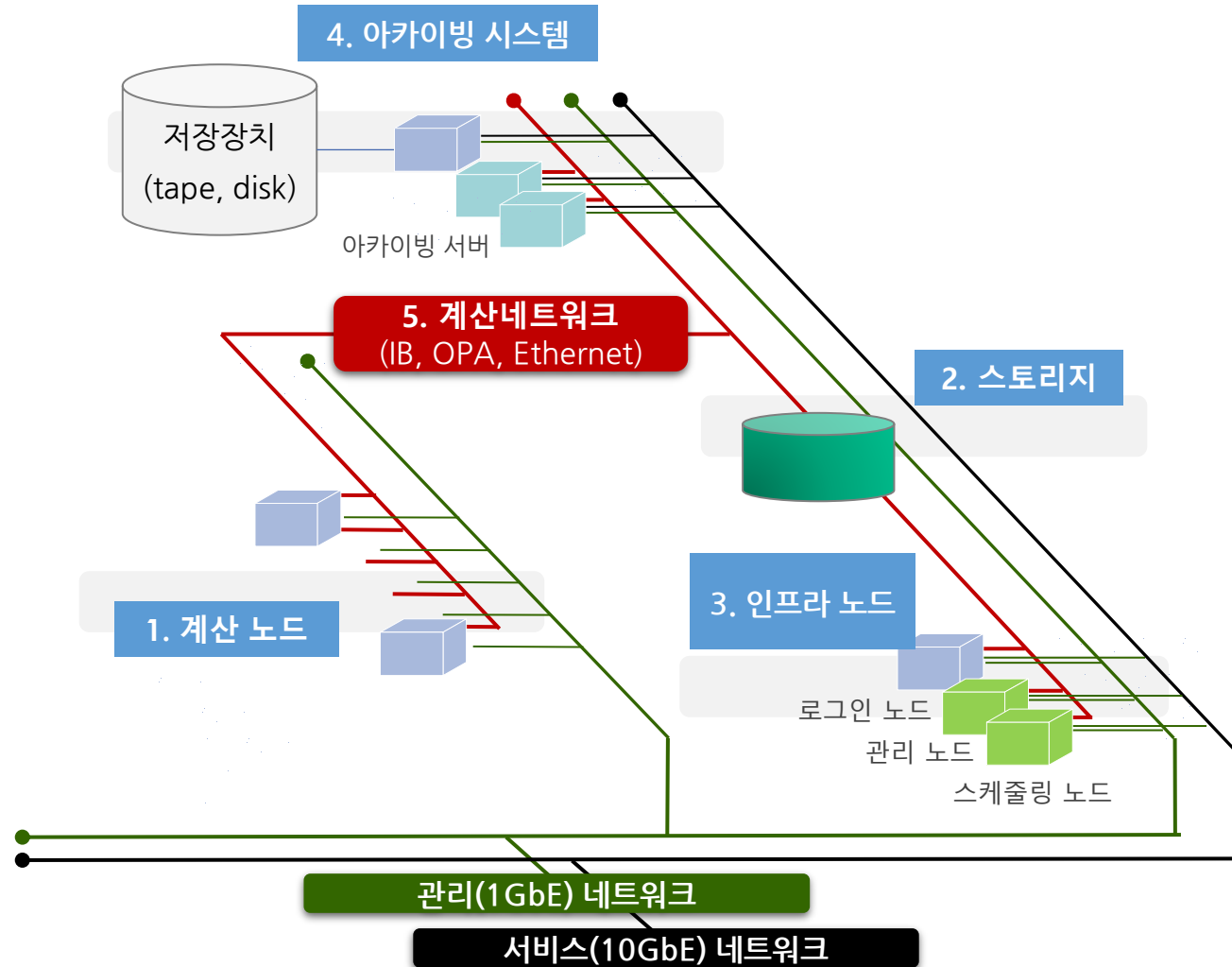
HPC 구성 고려 사항

항 목	내 용
계산 노드	• CPU, memory, disk drive, form factor, chassis, air/water cooling, etc.
스토리지 (파일시스템)	• data 용량, IO 속도, 파일 시스템(PFS, NFS, Scale-out NFS), disk drive(HDD, flash)
인터커넥트	• InfiniBand, OPA, Ethernet, topology
아카이빙	• data 크기, 생성/삭제 주기, 필요 용량, IO 속도, 저장 장치(tape, disk), HSM 기능
하이브리드 노드	• app 특성 고려, GPU option
인프라 노드	• HPC 관리/서비스 위한 노드(관리, 로그인, 스케줄러), HA, 적정 댓수
S/W	• 관리, 스케줄러, 라이브러리, Application

예 산

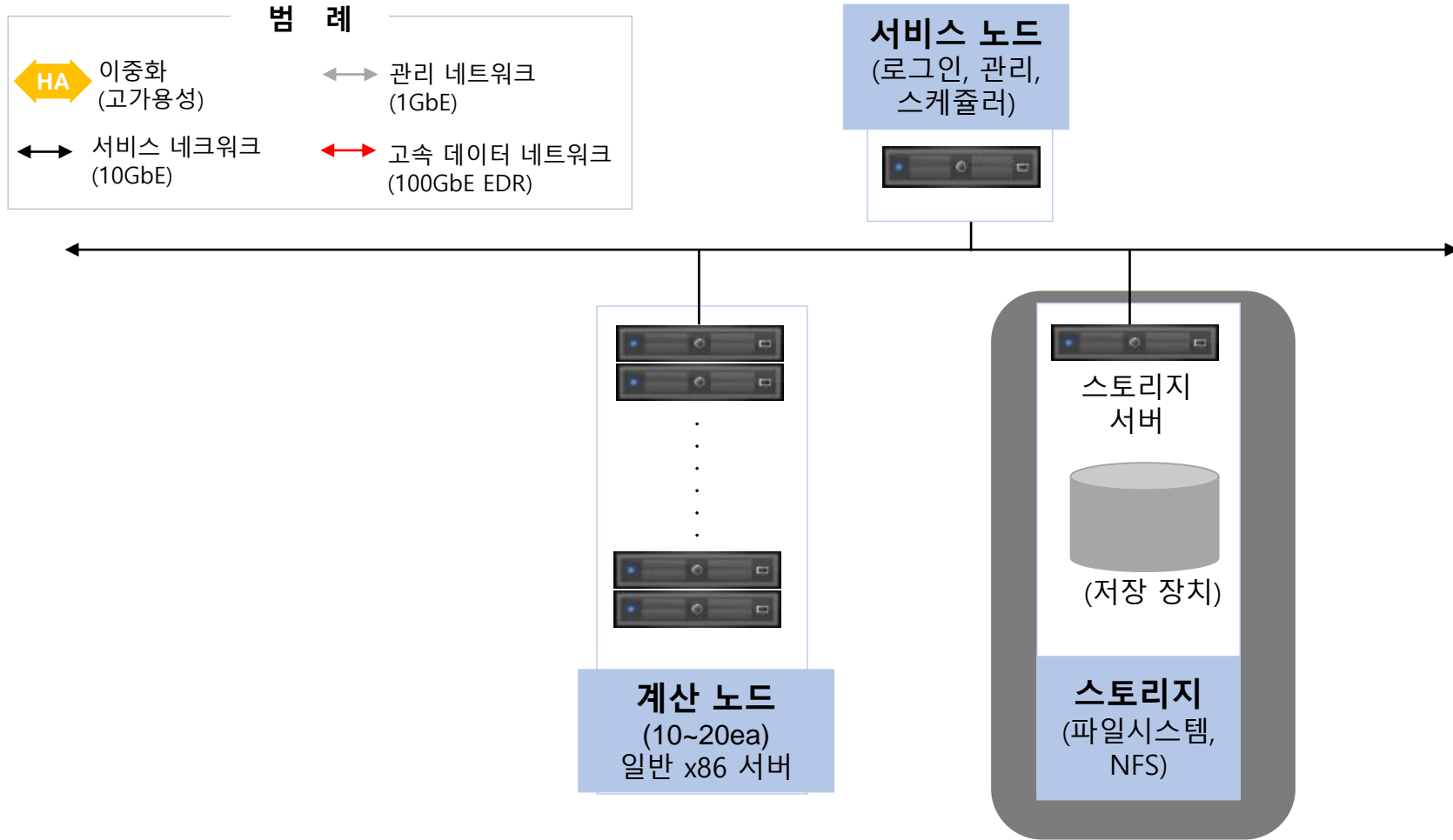
4. HPC 시스템 설계

HPC 시스템 설계 주요 부문



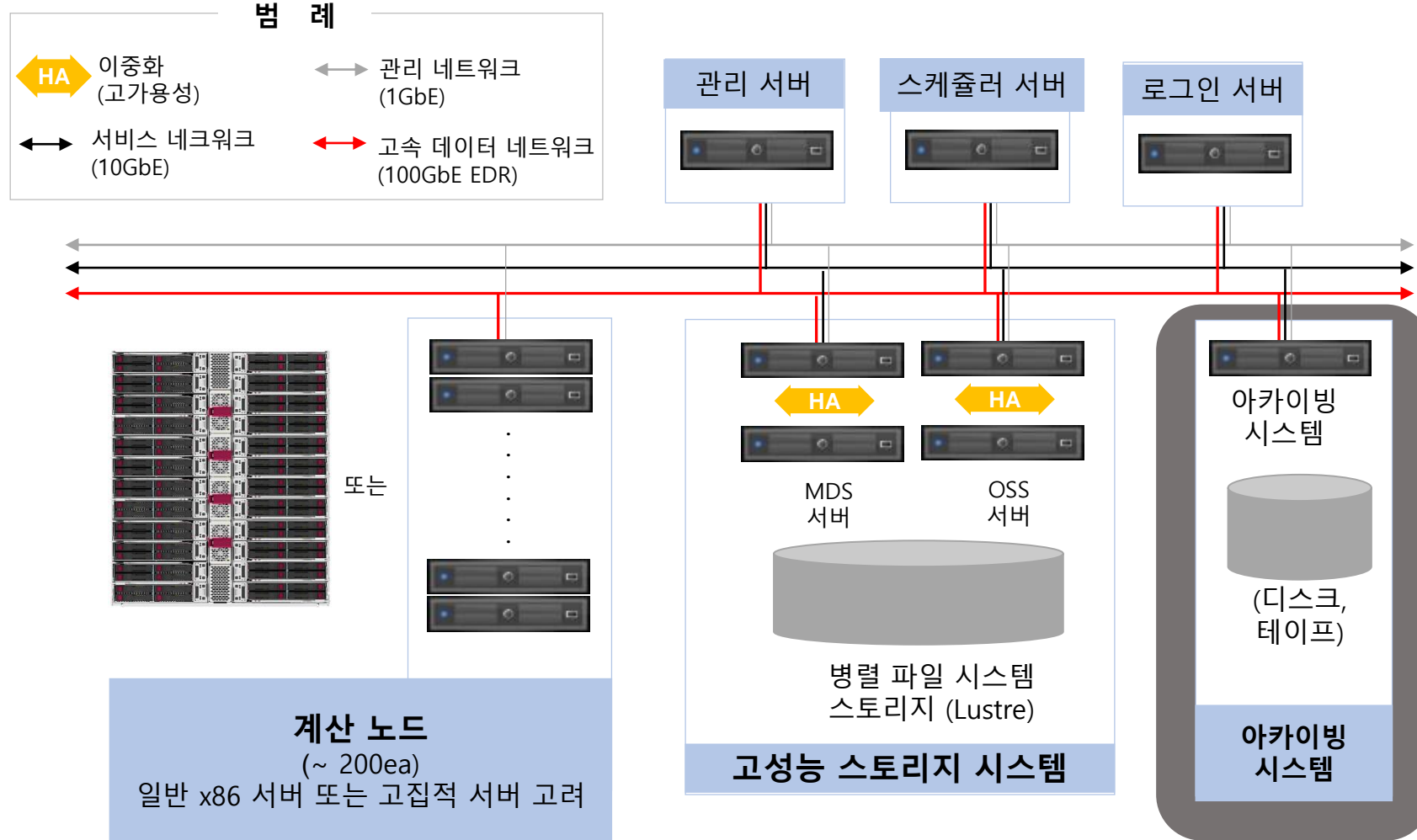
4. HPC 시스템 설계

HPC 시스템 구성 예 (소규모 시스템)



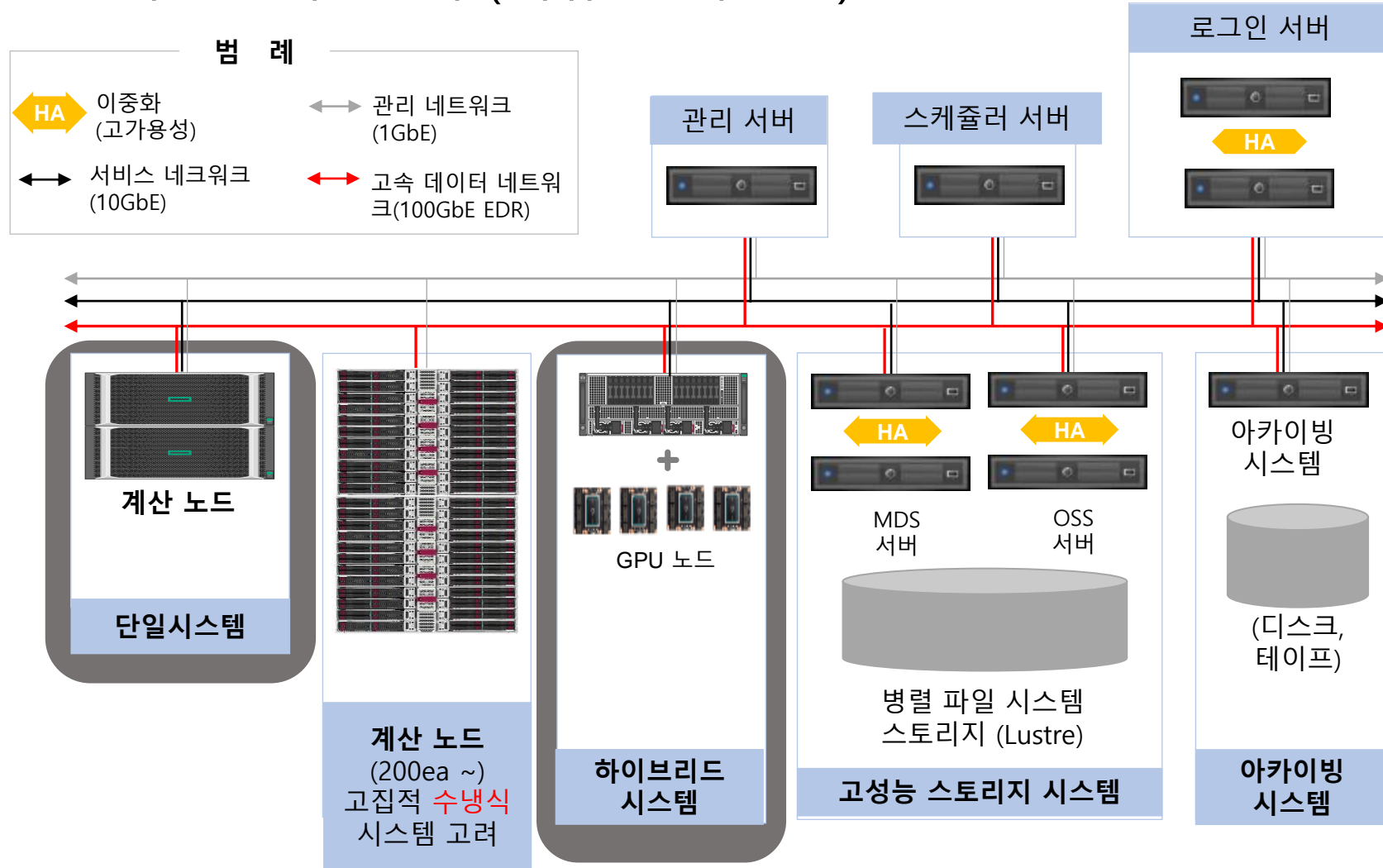
4. HPC 시스템 설계

HPC 시스템 구성 예 (중규모 시스템)



4. HPC 시스템 설계

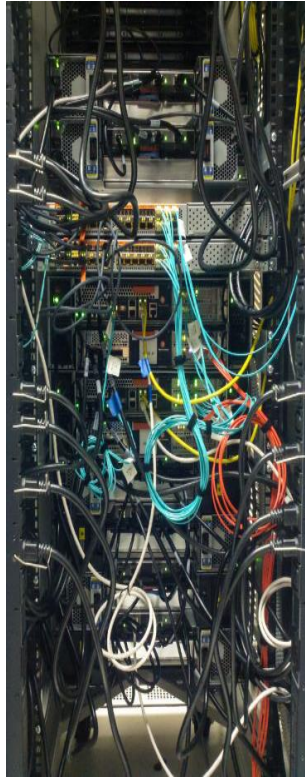
HPC 시스템 구성 예 (대규모 시스템)



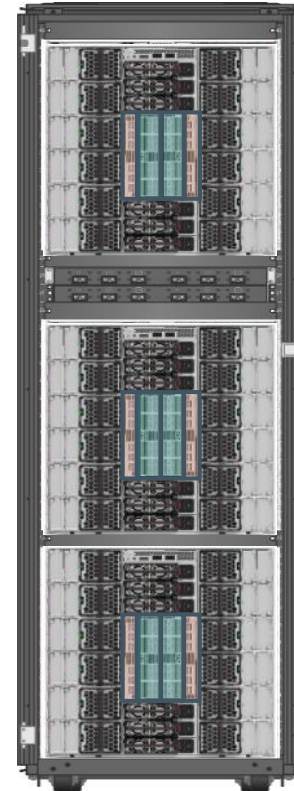
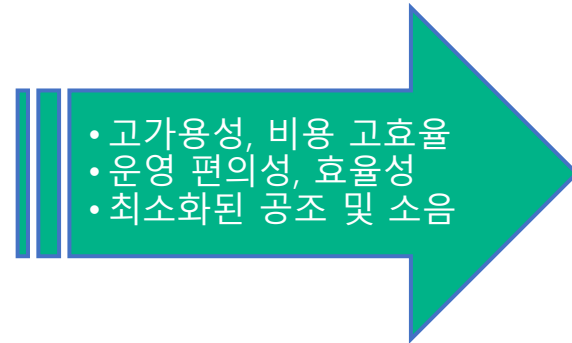
4. HPC 시스템 설계

고집적 계산 노드 시스템 (공냉식)

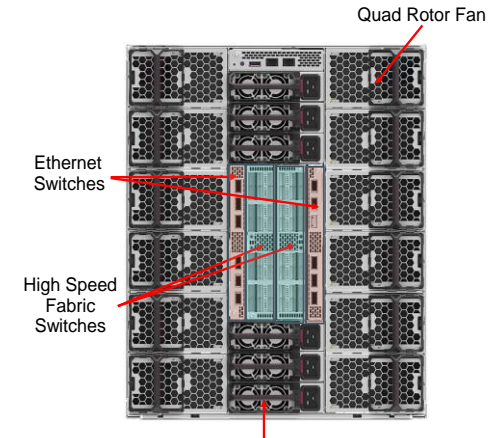
일반 랙 서버 대비 인프라 효율 개선



약 38~40노드/랙 (42U 랙)



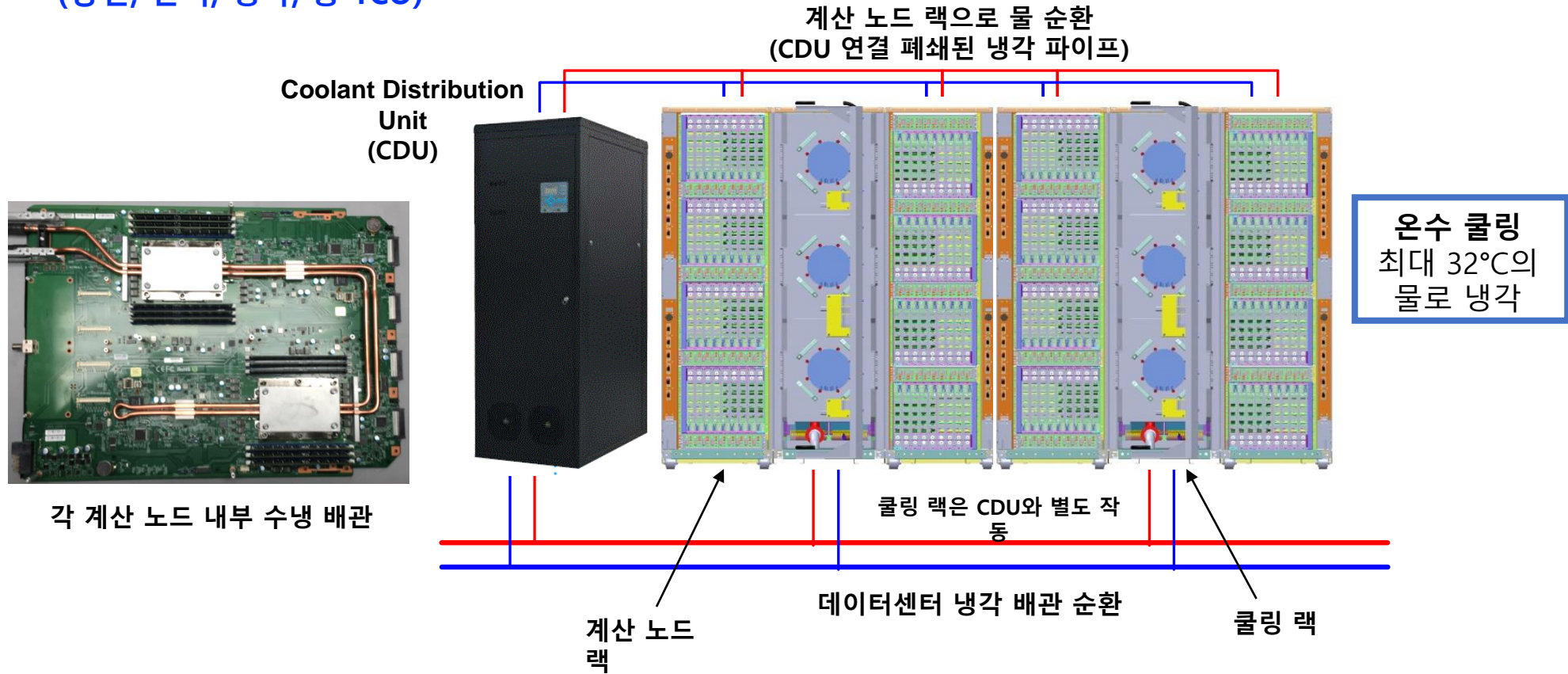
72~80노드/랙 (42U 랙)



4. HPC 시스템 설계

고집적 계산 노드 시스템 (수냉식)

일반 랙 서버 대비 획기적 인프라 효율 개선
(상면, 전력, 냉각, 총 TCO)



4. HPC 시스템 설계

HPC 시스템 사이징 (예): 1 PF급 시스템 (단순 계산 시스템 기준)

Step 1	CPU 선택 • Price / Performance • Application 특성 고려(Clock?, Core?, Memory BW/Core?)
Step 2	필요 노드 수 = 1,000,000 GF / Node 당 GF 예를 들어, Xeon Gold 6258R(2.7GHz, 28core), 2 socket/node의 경우: $1,000,000 \text{ GF} / (2419.2 \text{ GF} \times 2 \text{ sockets}) = 206.6$ (약 207 nodes)
Step 3	메모리, local drive 요건 확인 • 최소 1 DIMM per Channel 구성 • App 특성 고려 노드당 memory 용량 선택 • local drive 필요 여부 및 drive type, 용량 선택
Step 4	계산 노드 네트워크 구성 • Interconnect (IB, OPA, Ethernet), 관리 네트워크, 서비스 네트워크
Step 5	데이터 센터 환경 확인 • 전력(용량, 규격), 냉각, 상면, 하중 등
Step 6	기술적 요건 및 예산 상황에 따라 제안 시스템 선정 (표준 랙 type, 고집적 type, 계산 노드 type)

4. HPC 시스템 설계

HPC 시스템 관리

시스템 관리 (Admin.)

- 시스템 정지, 부팅, 재부팅 등의 Cluster를 위한 일상적인 관리 작업 가능
- 노드 그룹을 통해서 전체나 일부 노드에 동일한 명령어들을 실행

시스템 복제 배포 (Provisioning)

- OS 배포 및 자동 설정 기능 제공
- 클러스터에서 모든 노드에 golden image를 빠른 시간 내에 deploy

시스템 관제 Monitoring

- 클러스터의 상태 변화 추이 신속 확인
- Computing, Utility 뿐만 아니라 GPU에 대한 상태 모니터링
- 노드의 상태가 바뀔 때 마다 알림

효율적인 원격 관리

The screenshot shows a management console with a list of nodes on the left and a context menu for a selected group of nodes on the right. The context menu includes options like 'Shutdown (using SSH)', 'Power Off (using BMC)', 'Boot (using BMC)', 'Reboot (using SSH)', 'Change UID LED Status', 'Multiple Windows Broadcast', 'PDSH (using cmudiff)', 'PDCP (Distributed Copy)', 'Backup (Capture Image)', 'Cloning (Deploy Image)', and 'Auto Install (Kickstart | AutoYaST | Preseed)'. Callout boxes highlight '다중노드 선택' (Multi-node selection), '전원제어' (Power control), '그룹 Command' (Group Command), '구성변경비교' (Configuration comparison), and 'OS 배포' (OS deployment).

전체 노드를 한눈에 확인할 수 있는 통합 관리

The screenshot shows a monitoring dashboard with a 3D visualization of server racks. The dashboard displays various performance metrics and status indicators for the nodes. Callout boxes highlight 'amb1_temp', 'cpu1_temp', and 'power'.

5. 파트너사 협력



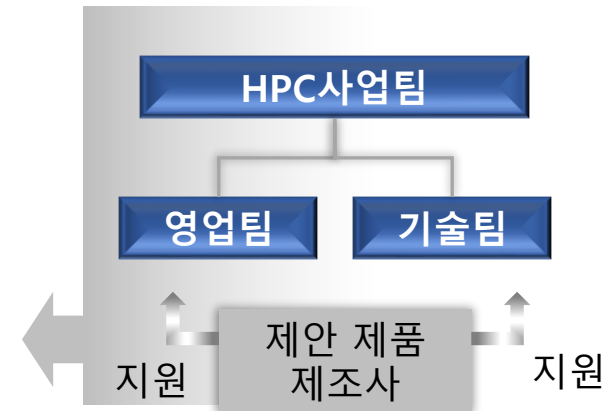
고객

파트너사



- 영업 초기 협의
- 고객 동행 방문
- 예산, 기술, 목표 시스템
- 협의 및 컨설팅
- 제안 및 기술 지원 협력
- 수행 지원 및 협력

(주)베이넥스



최적의 목표 시스템 설계, 구축, 서비스
"성공적 사업 협력"

감사합니다.