Introduction to KISTI-5 Supercomputer NURION

Supercomputing Infrastructure Center, KISTI

JunWeon Yoon
jwyoon@kisti.re.kr

24 April 2019, Pragma 36

History of KISTI Supercomputers

1st Supercomputer in Korea

1st Large-scale self-made cluster in KISTI

Cray-2S
1988.11 ~ 1993.10
1st Supercomputer

Cray C90
1993.11 ~ 2001.5
2nd Supercomputer

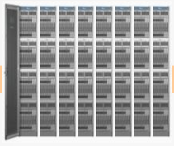
Cray T3E
1997.6 ~ 2002.12

PLUTO(128cpus)
2002. ~ 2008

IBM e1350 cluster (512cpus)
2004.1 ~ 2008.9

Tachyon

Sun C48 cluster
(3072 CPU cores)
2008.9 ~ 2016.4



Sun C48 cluster
(25,600 CPU cores)
2010.9 ~

Nurion

Cray CS500 cluster
(570,020 CPU cores)
2018.10 ~
5th Supercomputer

NEC SX-5
2001.5 ~ 2008.9
3rd Supercomputer

NEC SX-6
2003.2 ~ 2009.9
4th Supercomputer

HP SMP
2000.5 ~ 2006.12

IBM p690 (4nodes)
2002.1 ~ 2008.9

IBM p690+ (17nodes)
2003.7 ~ 2008.9

Nobel

Gaia

IBM p595 (10nodes)
2008.1 ~ 2015.12

IBM p595 (24nodes)
2009.10 ~ 2015.12

GPU Cluster
(1096 CPU cores,
44 GPUs)
2017.10 ~

1988	1993	1997	2000	2001	2002	2003	2008	2009	2018
2GF	16GF	131GF	242GF	786GF	5.2TF	8.6TF	29.9TF	360TF	26.2PF

1st CAVE in Korea

CAVE & SGI Onyx3400
2002.4 ~

Visualization Cluster
2004.10 ~

HP visualization cluster
(892 CPU cores, 109 GPUs)
2008 ~

KISTI-5 Procurement History

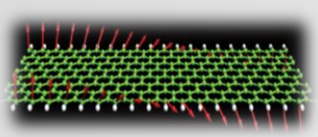
'15.06	▶ Data Center Building construction completed
'15.07.07	▶ Preliminary feasibility study approved
'16.03	▶ RFI and BMT announcement
'16.07	▶ RFP complete and submitted to PPS (Public Procurement Service)
'16.10~12	▶ Bidding
'17.01	▶ No Response in bidding & RFP modification
'17.02~05	▶ 2nd Bidding
'17.06~07	▶ Cray Inc. won bid and Tech/Price Negotiation
'17.08	▶ Contract finalized (49M USD)
'17.11	▶ Pilot system(16nodes) delivered
'17.12~'18.04	▶ Main system delivery and deployment
'18.05~	▶ BMT(HPL) and Performance Test
'18.07~	▶ Early Access
'18.12~	▶ Production service started

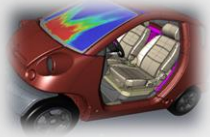
Efficiency: PUE < 1.35

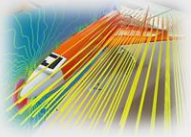
Performance: 25.7PFlops

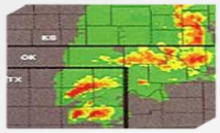
KISTI-5 Design

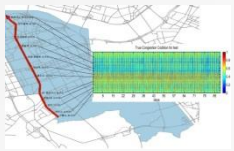
Providing Bigdata and Machine Learning User Environment
as well as Traditional HPC service


Molecular Dynamics
Material Design


Structure Design


Fluid Analysis


Weather/Climate Prediction




Machine Learning/Deep Learning


Computation Intensive
Computing

Cloud
Technology


Visualization

Data Intensive
Computing

Service Flexibility
(Logical Partitioning / Multi Platform Support)

Manycore based System

x86 based System

High Performance Interconnect

Large-scale Storage

High Performance Storage (Burt Buffer)

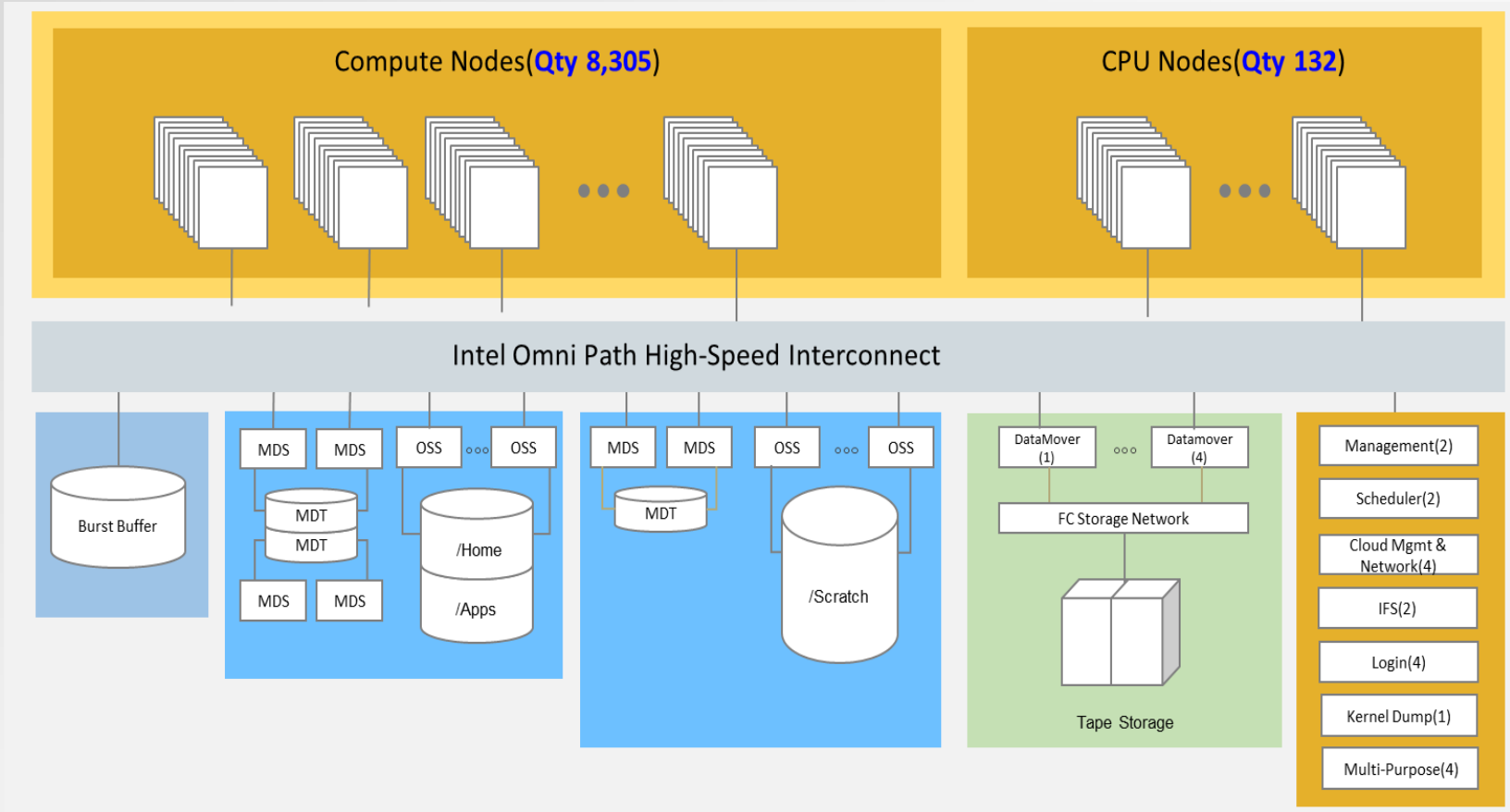
※ Power Consumption, Performance/Price, Service Continuance for User Adaptation

KISTI 5th Supercomputer NURION

- **128 Racks of Cluster Components**
  - ✓ 8 rows, with 16 Cabinets in each Row
  - ✓ 8,305 Compute Nodes
  - ✓ 132 Xeon Skylake CPU Nodes

- **12 Racks of DDN Storage**
  - ✓ 21 PB of Scratch Storage
  - ✓ 1.2 PB of Home and App Storage
  - ✓ 900 TB Burst Buffer
- **TS-4500 Tape Library**
  - ✓ 10PB / 1,700Medias
- **Interconnection Network**
  - ✓ Intel Omni-Path Architecture (100Gbps)

Proposed/Accepted System's Outline



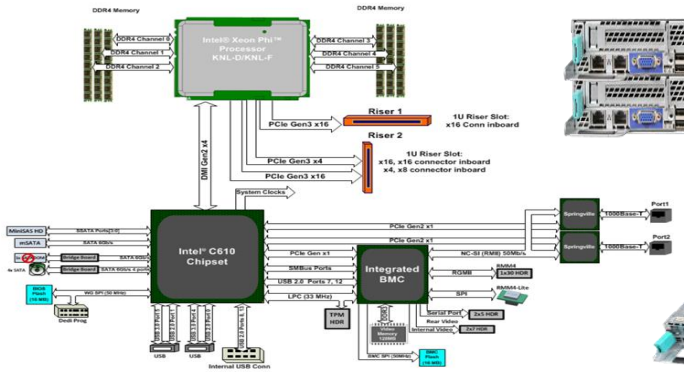

국가 슈퍼컴퓨팅센터
슈퍼컴퓨터 5호기 수송중

Computing nodes

The Largest KNL/OPA based Cluster-type Supercomputer

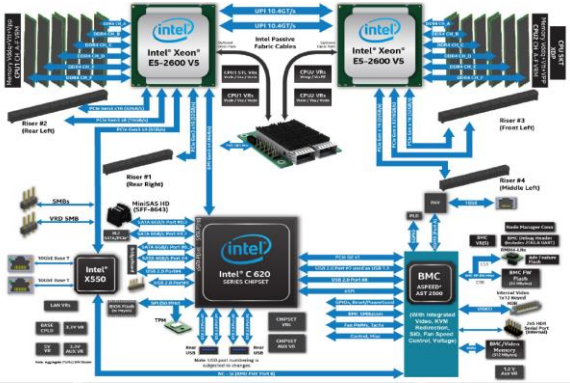
Compute nodes — Cray 3112-AA000T(2U enclosure), 8,305 KNL Computing modules

- 1x Intel Xeon Phi KNL 7250 processor
- 96GB (6x 16GB) DDR4-2400 RAM
- 1x Single-port 100Gbps OPA HFI card
- 1x On-board GigE (RJ45) port



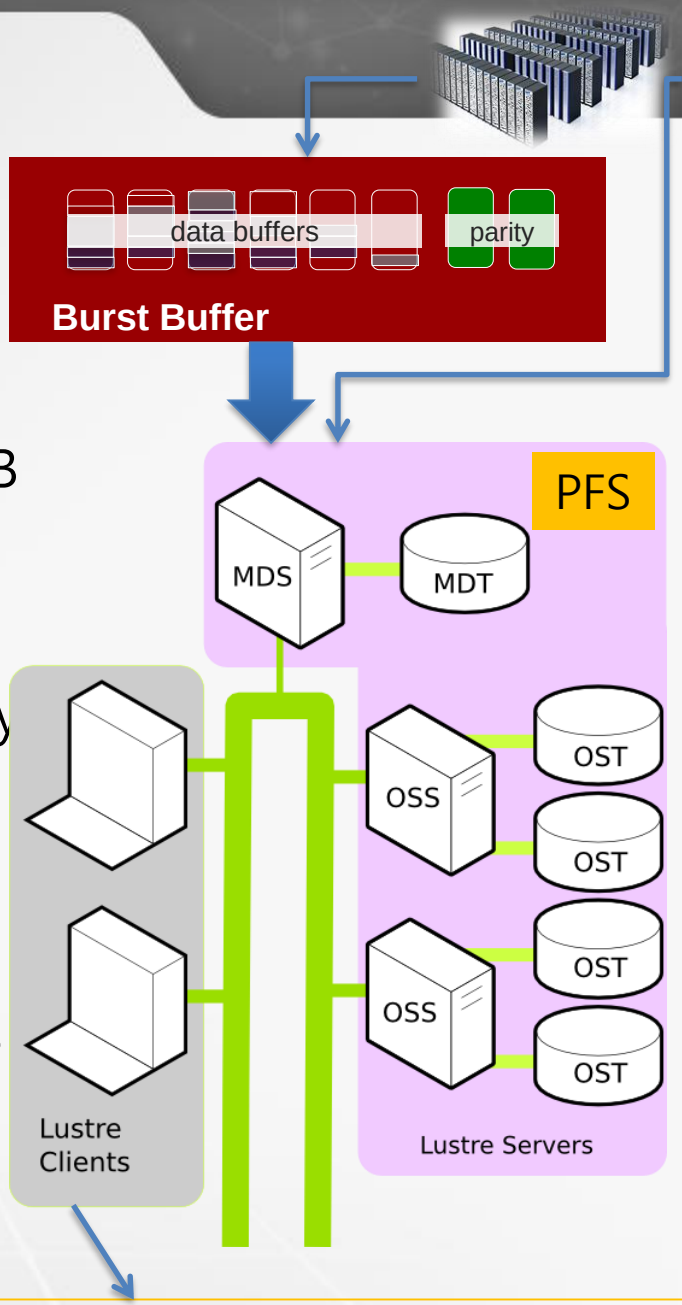
CPU-only nodes — Cray 3111-BA000T(2U enclosure), 132 Skylake Computing modules

- 2x Intel Xeon SKL 6148 processors
- 192GB (12x 16GB) DDR4-2666 RAM
- 1x Single-port 100Gbps OPA HFI card
- 1x On-board GigE (RJ45) port



Nurion Storage Summary

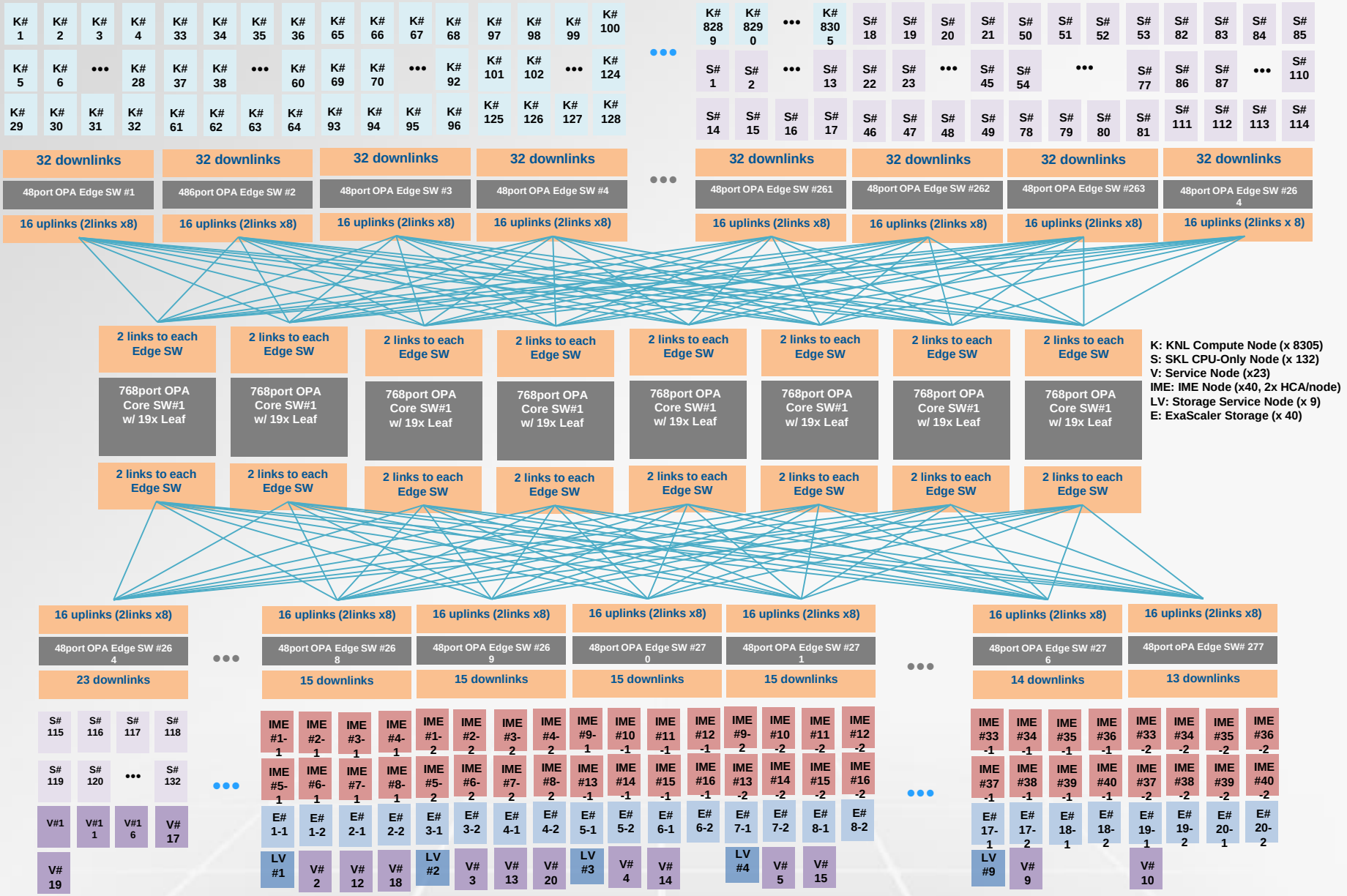
- Burst Buffer
  - 40x IME240
  - Performance : 800GB/s, Capacity : 900TB
- Scratch
  - 9x ES14KX (Lustre Embedded Storage Sy
  - Performance : 300GB/s, Capacity : 21PB
- Home/Apps
  - 1x ES14KX
  - Performance : 30GB/s, Capacity : 1.26PB



Computing nodes, Login nodes, Datamover, etc.

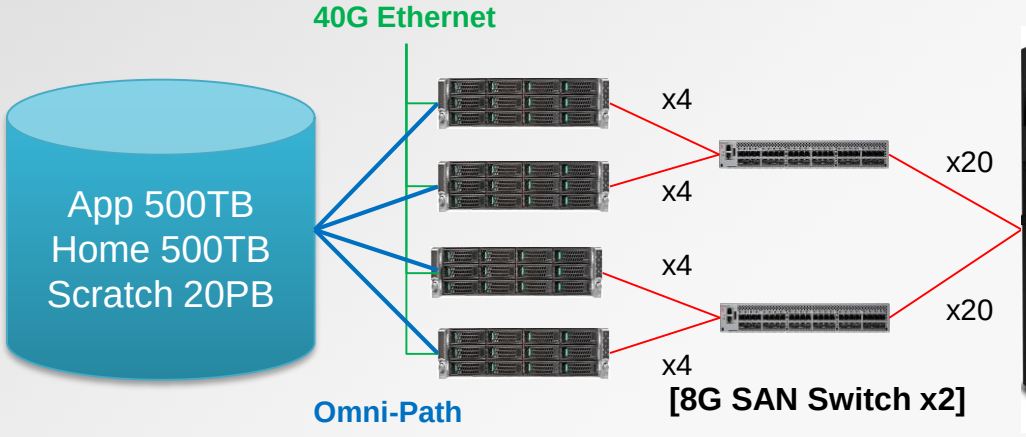
Storage Diagram Overview

2:1 Blocking OPA Interconnect

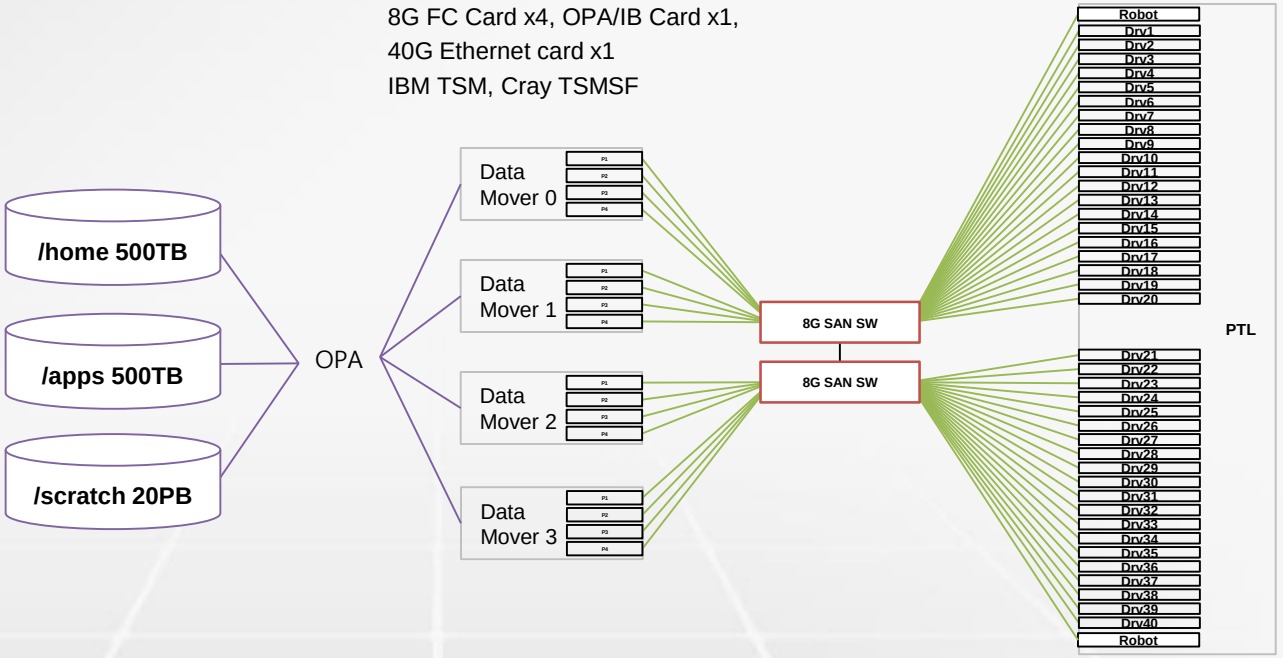


Tape based Archiving & Backup System

- TS 4500
  - LTO7 40 Drives
  - 6TB / Media
  - 10PB / 1700 Media
- SAN Switch
  - 2x 48port 8Gbps
- HSM Software
  - IBM TSM
  - Cray TSMSF (HSM)



8G FC Card x4, OPA/IB Card x1,
40G Ethernet card x1
IBM TSM, Cray TSMSF



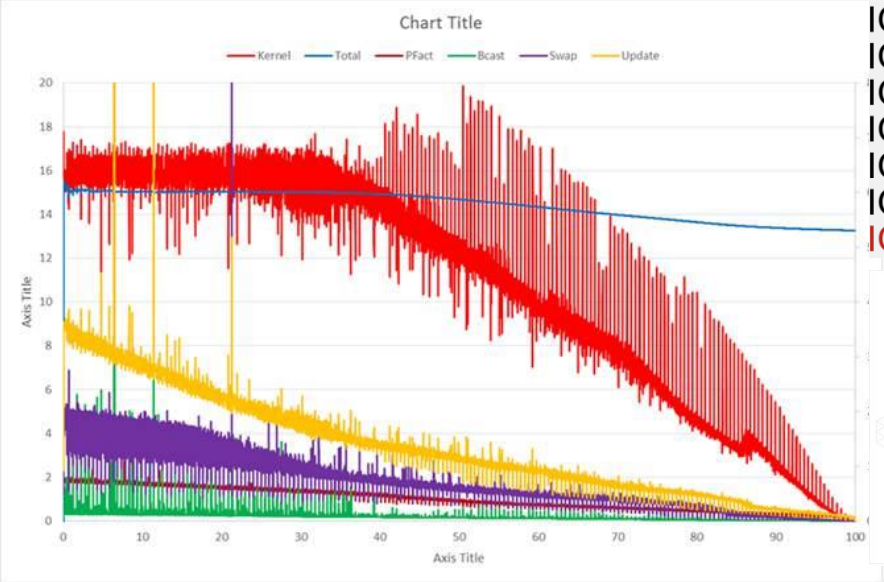
System Software

Scalability for +8,000 diskless, standard linux, compute nodes

Specification	Products	Version
Cluster Manager (Provisioing, Mgmt, Monitoring, OpenStack)	Bright Computing	8.0
Operating System	CentOS	7.4 ←~~7.3~~
Workload Manager	PBS Pro	14.2
Compilers	Intel	2017-17.0.2
	Cray PE	17.10
	Cray Compiler	8.6.3
	GNU	4.8.5
MPI Libraries	Intel MPI	2017.2
	Open MPI	1.10.3
	MVAPICH2	2.2rc1
Interconnect Software	Intel OPA	Driver 10.3.1.0 Fabric Manager 10.8 ←~~10.6~~
Parallel File System	Lustre	2.10.5 ← ~~2.7.21.3~~
	DDN ES	3.2
Burst Buffer	IME	1.2.1 ← 1.2
Debugger	Allinea DDT	18.0
Profiler	CrayPat Intel Vtune	-
User Account Management	LDAP	-

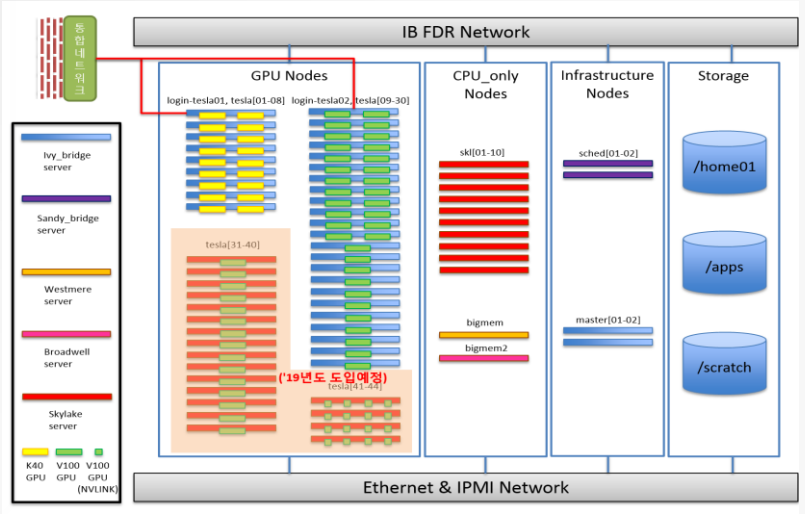
Benchmark Performance

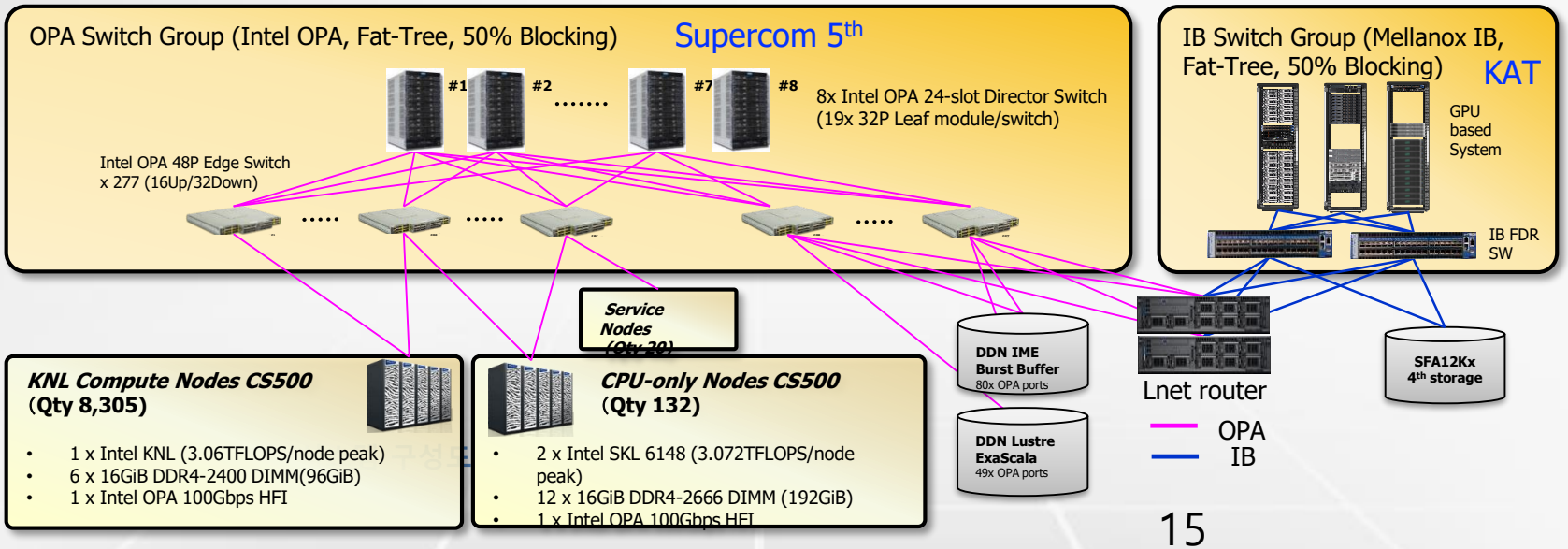
Category	Features	Score	World Ranking
HPL	Large-scale Dense Matrix Computation Used for Top500	13.93PF	11th (Jun 2018)
HPCG	Large-scale Sparse Matrix Computation Similar to normal user applications	0.39PF	8th (Nov 2018)
Graph500	Breadth-First Search, Interconnect Performance	1048.86GE	23rd (Nov 2018)
IO500	Various IO Workloads	160.67	2nd (Nov 2018)



Support for GPU resources

- Production Service starts in May 2019
- 59 Servers , 41 V100 GPU in this year
  - Peak Performance (Double Precision) : 600TF
- Parallel File system sharing with Nurion 5 via OPA-IB routing





Thank you

KISTI Supercomputing Center

http://www.ksc.re.kr

http://en.kisti.re.kr/supercomputing