Ministry of Electronics & Information Technology

सुस्वागतम्
நல்வரவு
ସୁସ୍ୱାଗତମ
సుస్వాగతం
સુસ્વાગતમ
সুস্বাগতম
ಸುಸ್ವಾಗತ
സുസ്വാഗതം
সুস্বাগতম
ꯁꯨꯁꯥꯒꯤꯗꯃ
सुस्वागतम्
خوش آمدید

NPSF User Workshop 2023

Presentation On

NPSF & AIRAWAT-PSAI

Pankaj Dorlikar, C-DAC

npsfhelp@cdac.in

NPSF
National PARAM Supercomputing Facility

24th Jul, 2023

NVIDIA

Inception of NPSF

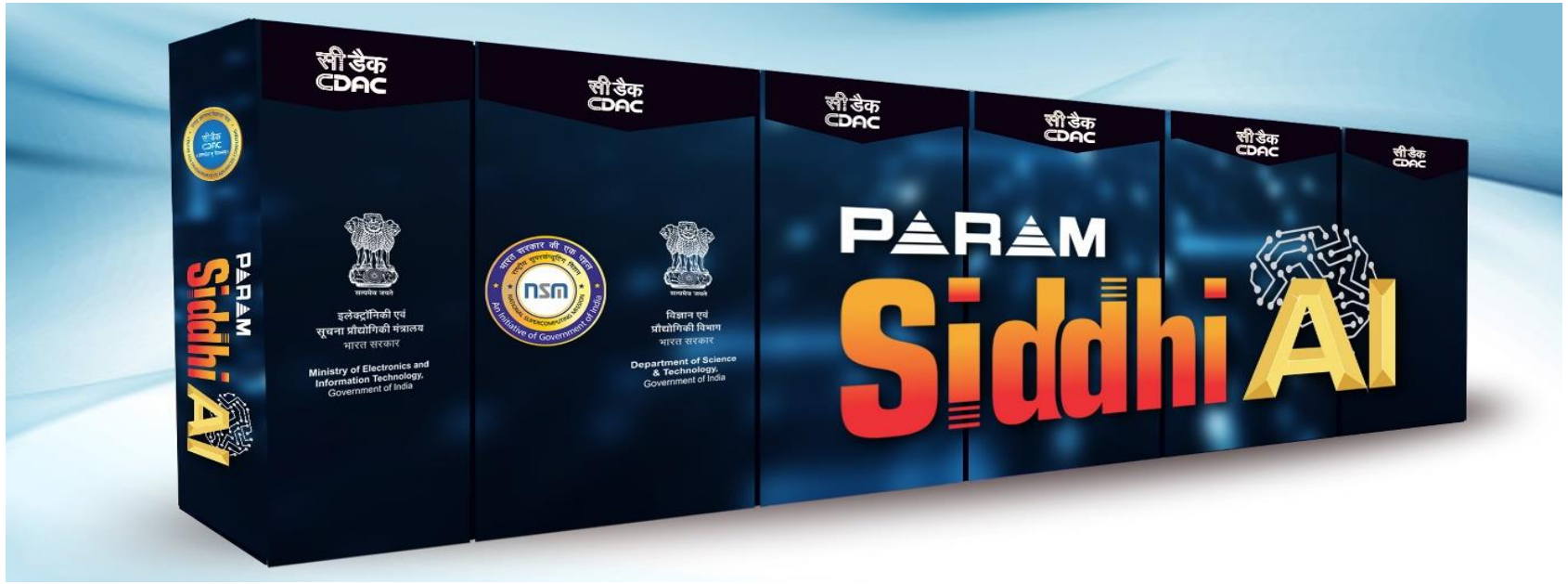
National PARAM Supercomputing Facility

C-DAC established National PARAM Supercomputing Facility (NPSF) with a mandate :

- To offer state-of-the-art High Performance Computing systems to various institutions and industries that need such a facility to process their diverse applications and resources

- Also to help them with the know-how and usage of such systems and proliferate HPC awareness in the country.

PARAM Siddhi-AI



- PARAM Siddhi - AI of 5.26 Petaflops (210 AI Petaflops) was the fastest Supercomputer in India and ranked at No. 62 position in 'TOP500 Supercomputer List – November 2020' declared at Supercomputing Conference 2020 (SC 20) at United States.
- Being made available to MSMEs and Start-ups

AIRAWAT PoC

- Proof of Concept (PoC) AI Research Analytics and Knowledge Dissemination Platform (AIRAWAT)
- 200AI Petaflops Mixed Precision peak compute capacity

AIRAWAT-PSAI

top500.org/lists/top500/list/2023/06/

75	AIRAWAT - PSAI - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Infiniband HDR, Netweb Technologies Center for Development of Advanced Computing (C-DAC) India	81,344	8.50	13.17

topsc.cdacb.in/filterdetails?slug=July2023

TOP SUPERCOMPUTERS INDIA

Home Login Supercomputer List Contact Us Summary

July2023

Following is the ranking of the systems in terms of Rmax (Linpack Benchmark performance).

- C - CPUs, G-GPUs, ICO - Intel Co-processors. Cores (4th column) that are not qualified with C/G/ICO refer to CPUs.
- OEM - Original Equipment Manufacturer, SI - System Integrator.

Rank	Site	System	Core/Processor/Socket/Nodes	Rmax (TFlops)	Rpeak (TFlops)
1 NEW	Center for Development of Advanced Computing (C-DAC),PUNE	AIRAWAT - PSAI is a NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHZ, NVIDIA A100, INFINIBAND HDR2	81344/2/82	8500	13176

- The AIRAWAT PoC of 200 AI Petaflops integrated with PARAM Siddhi – AI of 210 AI Petaflops gives a total peak compute of 410 AI Petaflops Mixed Precision (13.17 PF DP) and sustained compute capacity of 8.5 Petaflops (Rmax) Double Precision.
- Al Supercomputer ‘AIRAWAT-PSAI’, installed at C-DAC, Pune has been ranked 75th in the world.
- It was declared so in the 61st edition of Top 500 Global Supercomputing at the International Supercomputing Conference (ISC 2023) in Germany

Top 500 Certificate

TOP 500 The List. CERTIFICATE

AIRAWAT - PSAI - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Infiniband HDR

Center for Development of Advanced Computing (C-DAC), India

is ranked

No. 75

among the World's TOP500 Supercomputers

with 8.50 PFlop/s Linpack Performance

in the 61st TOP500 List published at the ISC23

Conference on June 01, 2023.

Congratulations from the TOP500 Editors

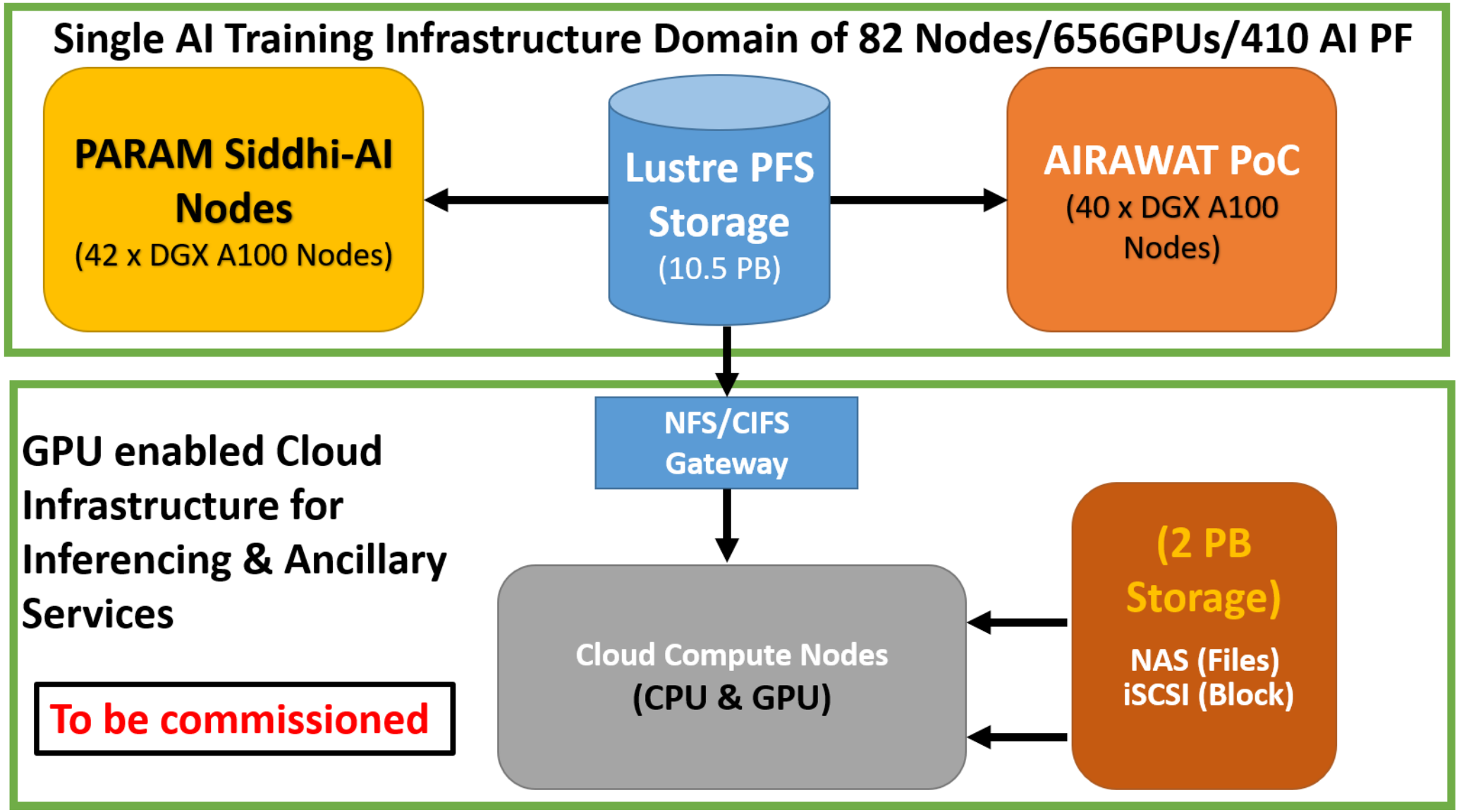
Erich Strohmaier
NERSC/Berkeley Lab

Jack Dongarra
University of Tennessee

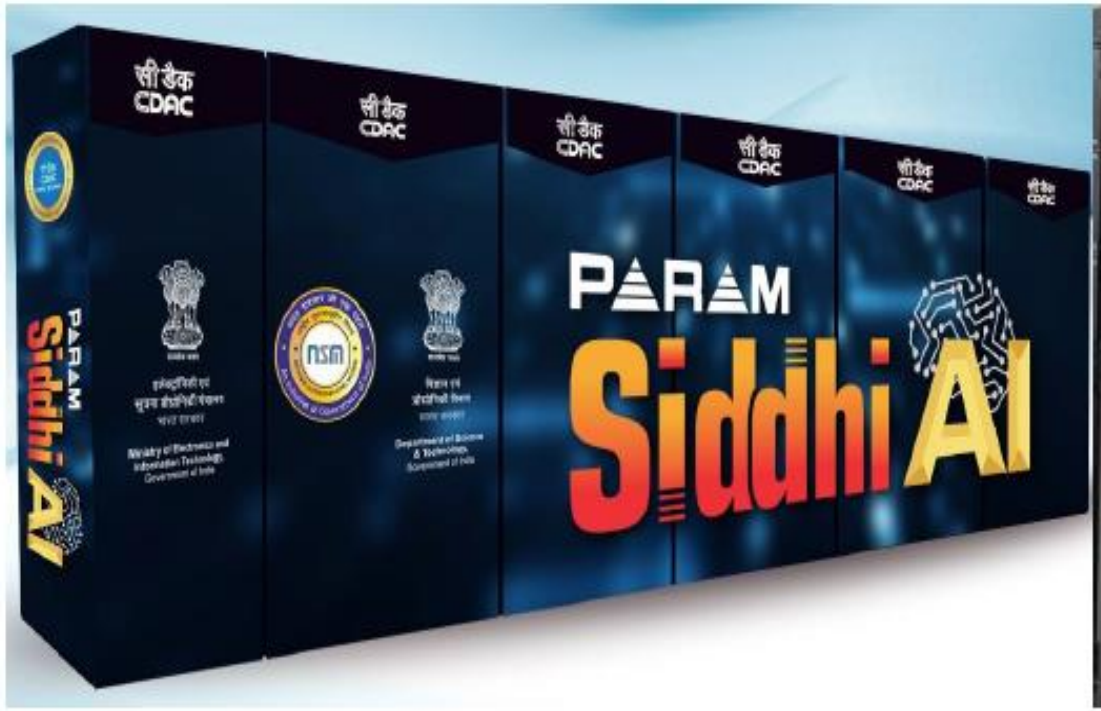
Horst Simon
NERSC/Berkeley Lab

Martin Meuer
Prometeus

AIRAWAT – PSAI: Logical Overview



NPSF : AIRAWAT-PSAI Stats



629 Users

176 Projects

114 Institutes/Organizations

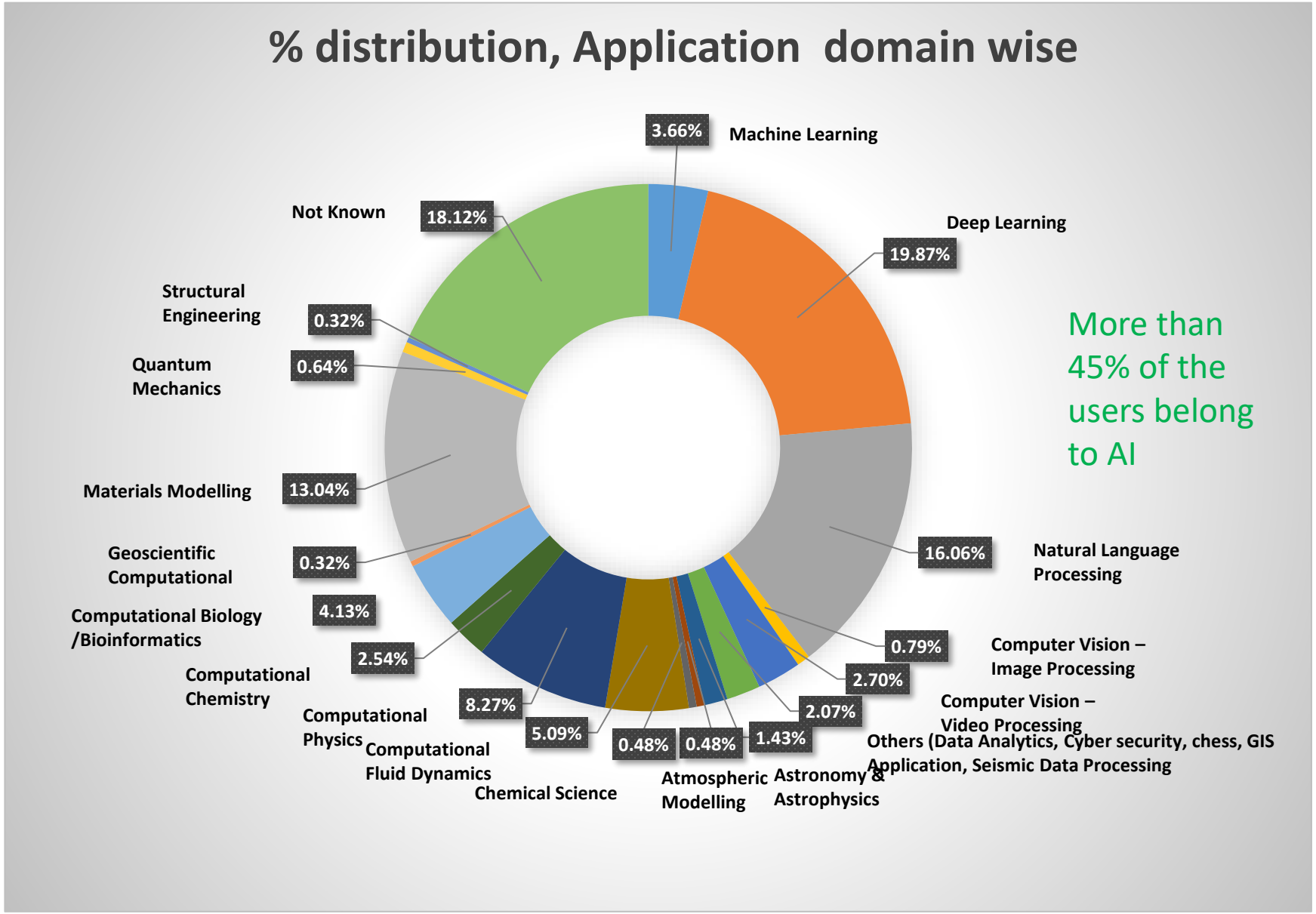
150+K Jobs

14 Start-ups

8.5 PFlop/s HPL (Rmax)

13.17 PFlop/s HPL (Rpeak)

NPSF Userbase



AIRAWAT-PSAI : Major System Components

82 Nos of DGX A100 Compute servers

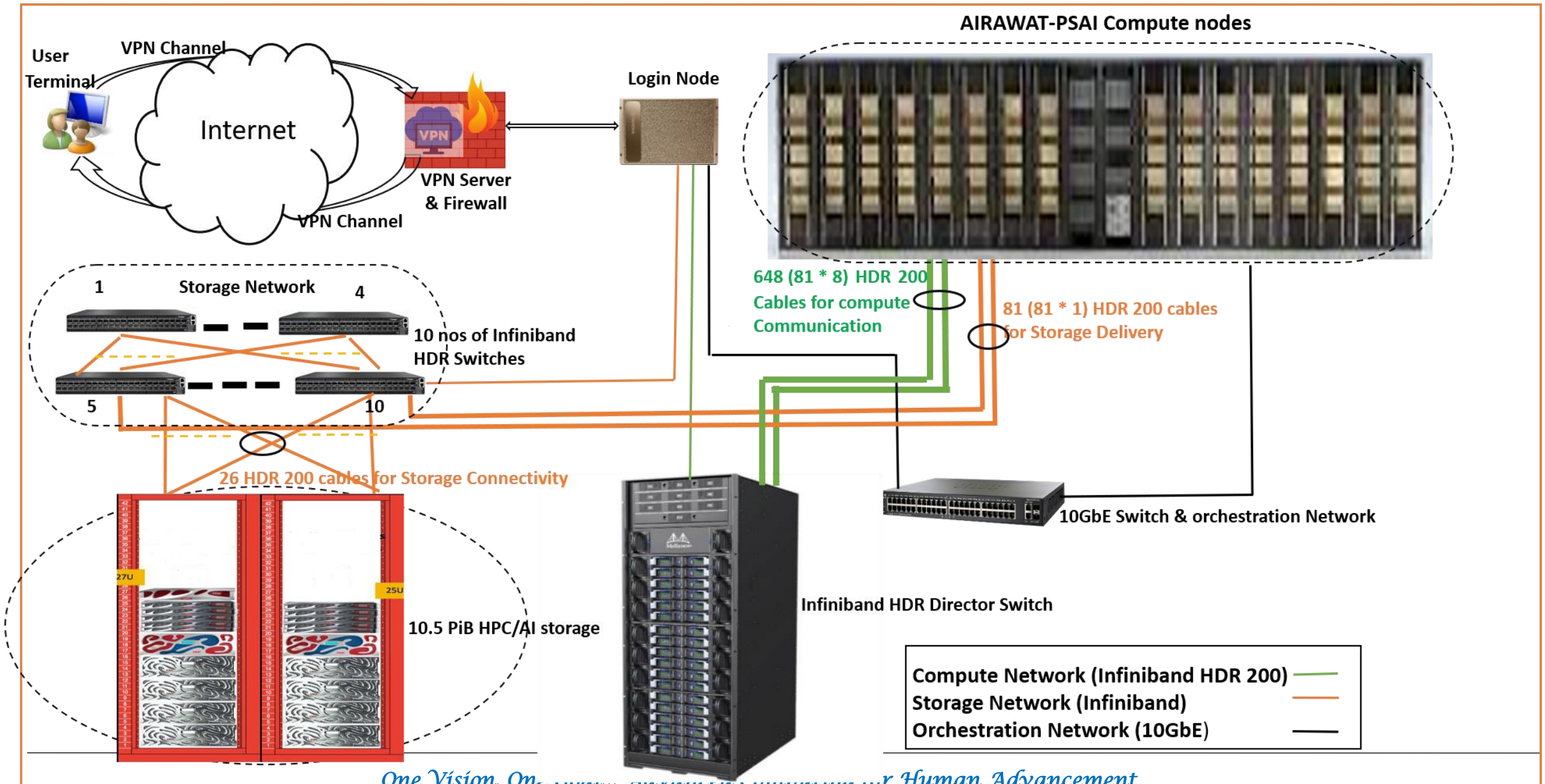
InfiniBand HDR200 Director Switch and IB Cables for Compute Communication

InfiniBand HDR200 Edge Switches and IB Cables for Storage Communication

10.5 PiB Storage

NVIDIA DGX-A100 Compute Nodes	82 (20992 CPU cores)
NVIDIA A100-40GB Tensor Core GPUs	656 (82 nodes * 8 CPUs per node)
Mellanox 200G HDR InfiniBand Switch (Compute)	800 Ports (20 Leafs *40 ports per Leaf)
Mellanox 200G HDR InfiniBand Switches (Storage)	400 Ports (10 Switches * 40 ports per switch)
PFS based storage @250 GB/Sec, 4M IOPS	10.5 PiB (2 Tier Storage)

Architecture

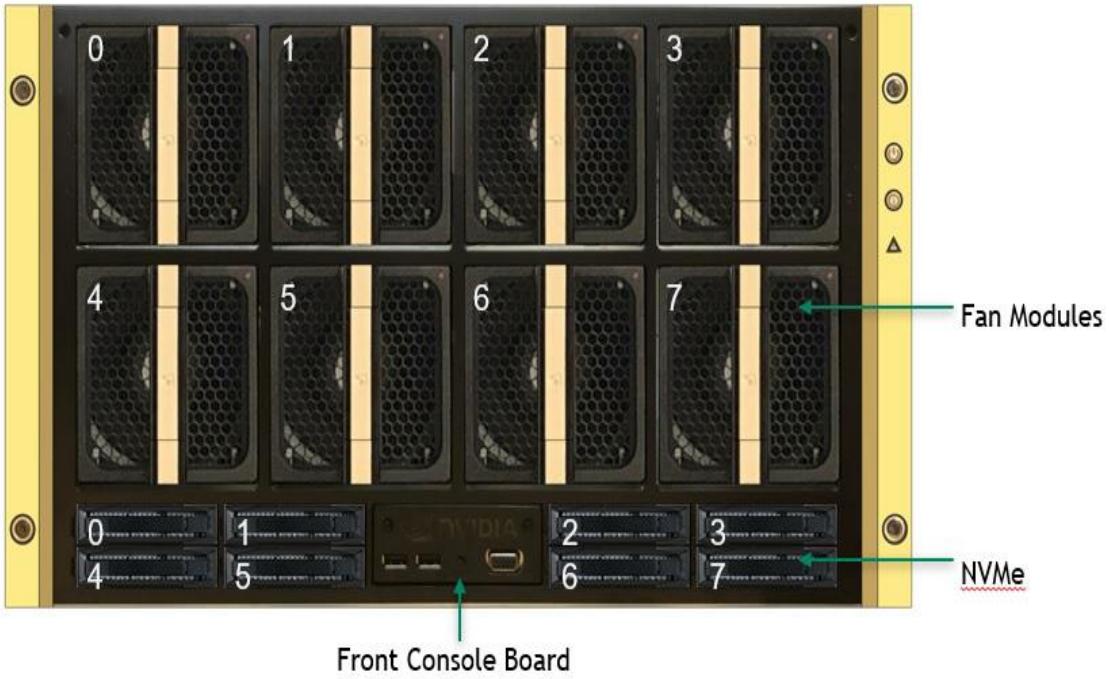


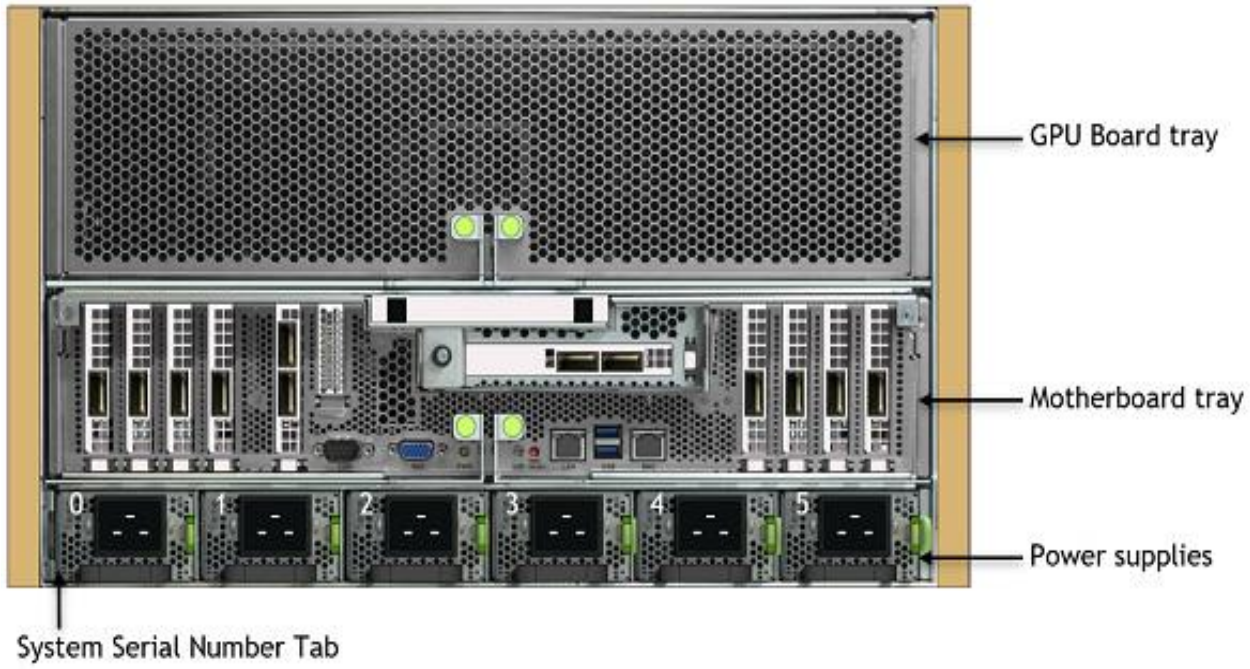
PARAM SIDDHI-AI : DGX-A100 SYSTEM INSIGHT

SYSTEM SPECIFICATIONS

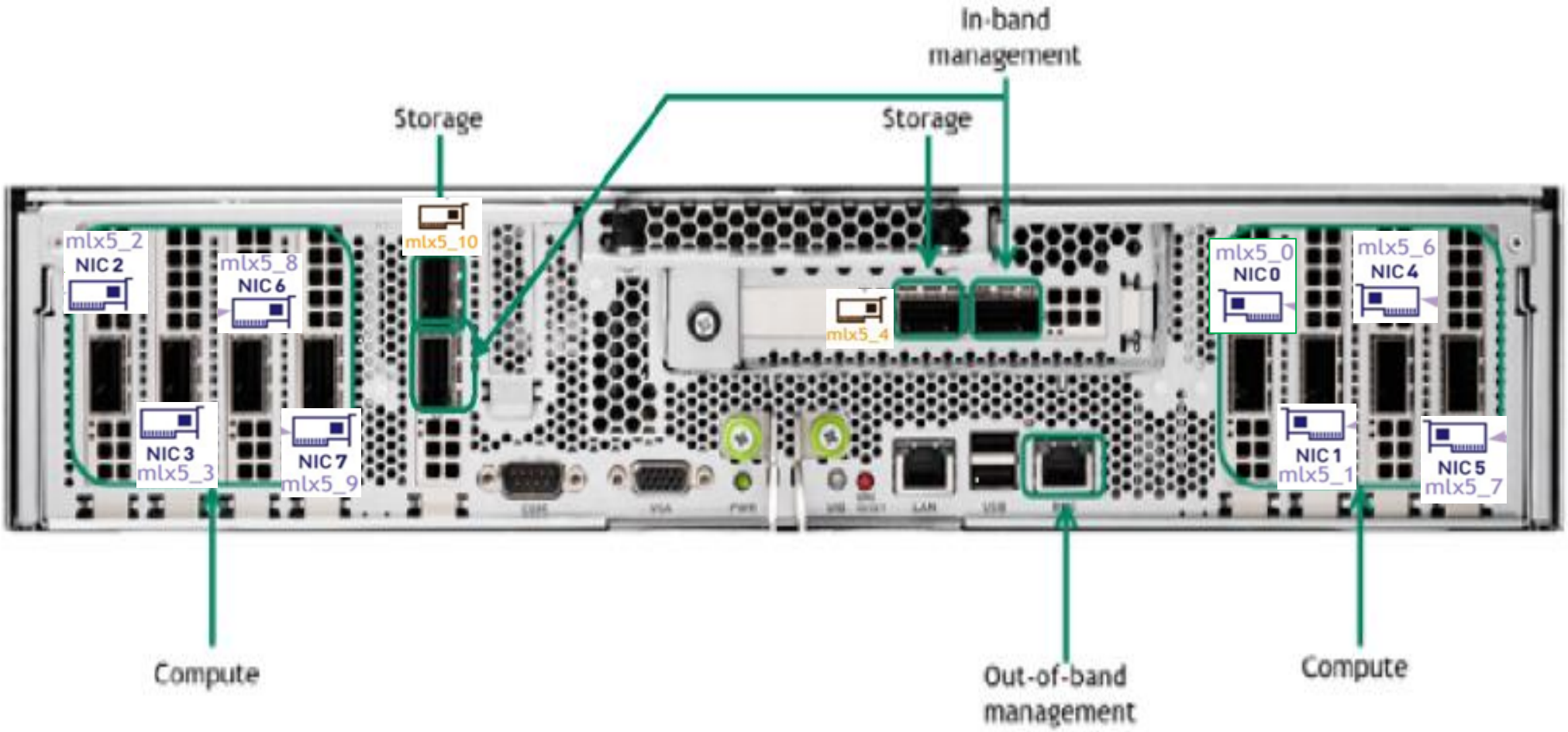
GPUs	8x NVIDIA A100 Tensor Core GPUs
GPU Memory	320 GB total
Performance	5 petaFLOPS AI 10 petaOPS INT8
NVIDIA NVSwitches	6
System Power Usage	6.5kW max
CPU	Dual AMD Rome 7742, 128 cores total, 2.25 GHz (base), 3.4 GHz (max boost)
System Memory	1TB
Networking	8x Single-Port Mellanox ConnectX-6 VPI 200Gb/s HDR InfiniBand 1x Dual-Port Mellanox ConnectX-6 VPI 10/25/50/100/200Gb/s Ethernet
Storage	OS: 2x 1.92TB M.2 NVME drives Internal Storage: 15TB (4x 3.84TB) U.2 NVME drives
Software	Ubuntu Linux OS
System Weight	271 lbs (123 kgs)
Packaged System Weight	315 lbs (143kgs)
System Dimensions	Height: 10.4 in (264.0 mm) Width: 19.0 in (482.3 mm) MAX Length: 35.3 in (897.1 mm) MAX
Operating Temperature Range	5°C to 30°C (41°F to 86°F)

COMPUTE NODE

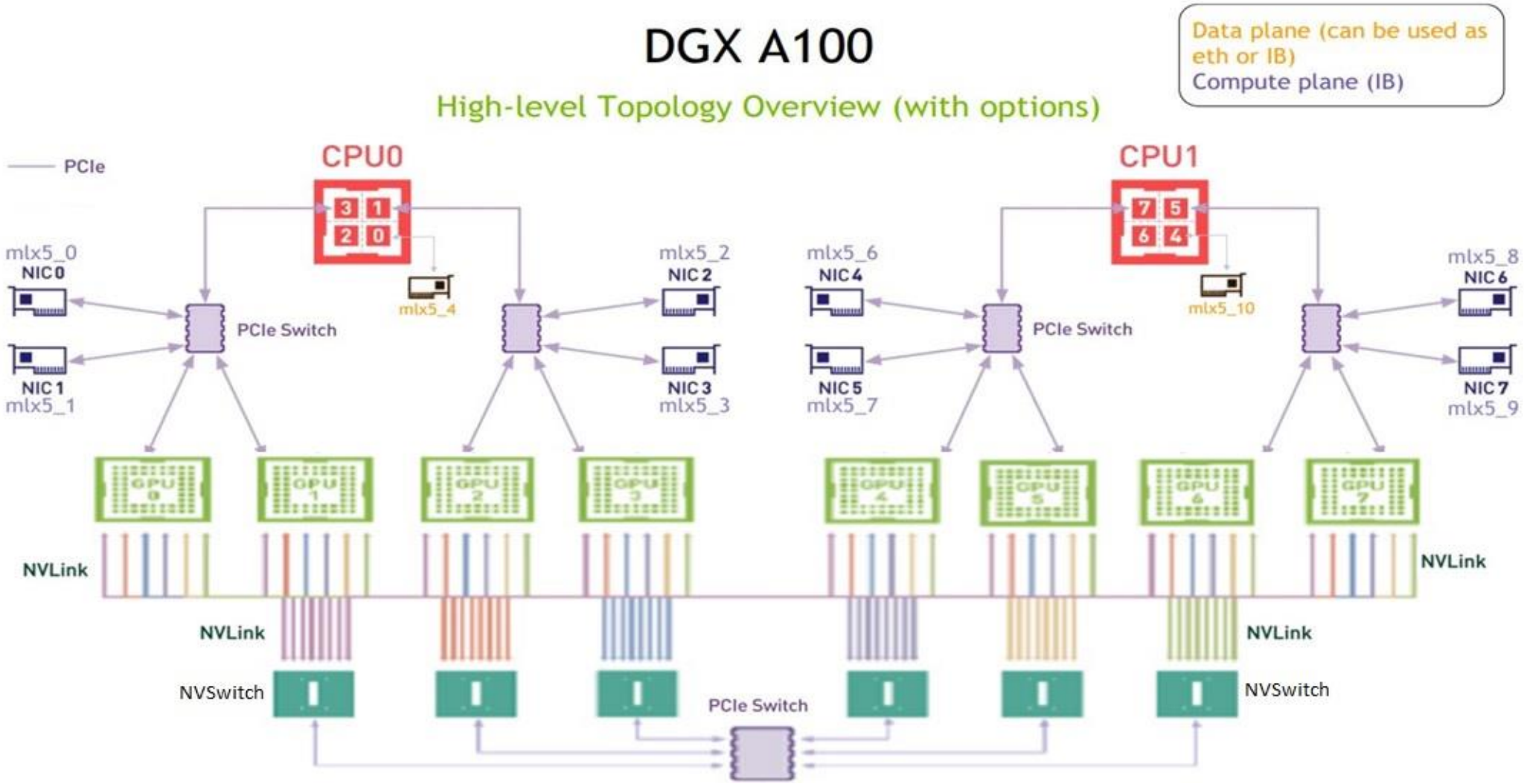




Compute node Networking



DGX A100 SYSTEM TOPOLOGY



GPU and IB HCA Affinity

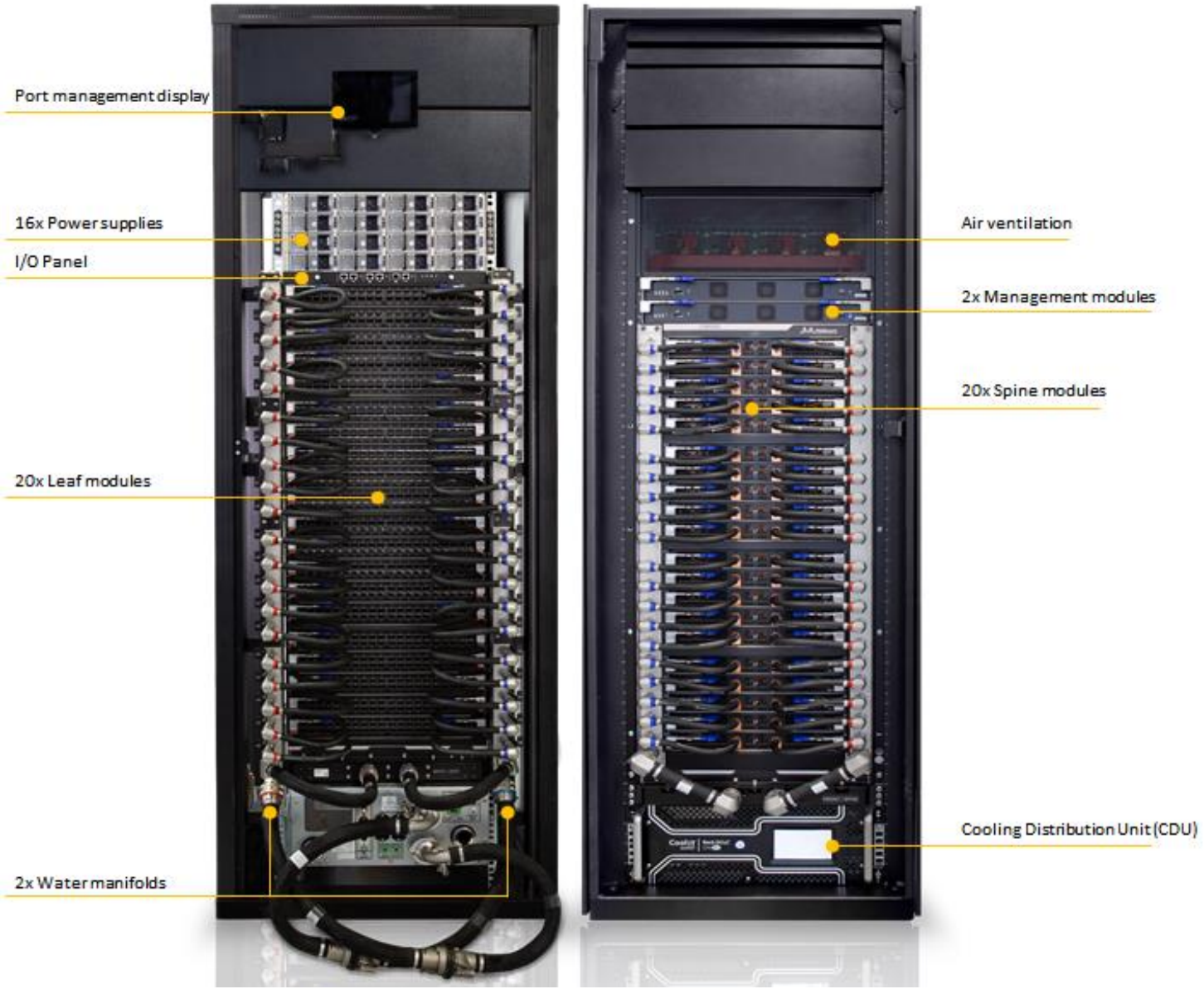
```
pankajd@scn1-mn:~$ nvidia-smi topo -m
        GPU0    GPU1    GPU2    GPU3    GPU4    GPU5    GPU6    GPU7    mlx5_0  mlx5_1  mlx5_2  mlx5_3  mlx5_4  mlx5_5  mlx5_6  mlx5_7  mlx5_8  mlx5_9  mlx5_10 mlx5_11 CPU Affinity    N
UMA Affinity
GPU0     X      NV12    NV12    NV12    NV12    NV12    NV12    NV12    PXB     PXB     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     48-63,176-191   3
GPU1    NV12     X      NV12    NV12    NV12    NV12    NV12    NV12    PXB     PXB     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     48-63,176-191   3
GPU2    NV12    NV12     X      NV12    NV12    NV12    NV12    NV12    SYS     SYS     PXB     PXB     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     16-31,144-159   1
GPU3    NV12    NV12    NV12     X      NV12    NV12    NV12    NV12    SYS     SYS     PXB     PXB     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     16-31,144-159   1
GPU4    NV12    NV12    NV12    NV12     X      NV12    NV12    NV12    SYS     SYS     SYS     SYS     SYS     SYS     PXB     PXB     SYS     SYS     SYS     SYS     112-127,240-255 7
GPU5    NV12    NV12    NV12    NV12    NV12     X      NV12    NV12    SYS     SYS     SYS     SYS     SYS     SYS     PXB     PXB     SYS     SYS     SYS     SYS     112-127,240-255 7
GPU6    NV12    NV12    NV12    NV12    NV12    NV12     X      NV12    SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     PXB     PXB     SYS     SYS     80-95,208-223   5
GPU7    NV12    NV12    NV12    NV12    NV12    NV12    NV12     X      SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     PXB     PXB     SYS     SYS     80-95,208-223   5
mlx5_0  PXB     PXB     SYS     SYS     SYS     SYS     SYS     SYS      X      PXB     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS
mlx5_1  PXB     PXB     SYS     SYS     SYS     SYS     SYS     SYS     PXB      X      SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS
mlx5_2  SYS     SYS     PXB     PXB     SYS     SYS     SYS     SYS     SYS     SYS      X      PXB     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS
mlx5_3  SYS     SYS     PXB     PXB     SYS     SYS     SYS     SYS     SYS     SYS     PXB      X      SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS
mlx5_4  SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS      X      PIX     SYS     SYS     SYS     SYS     SYS     SYS
mlx5_5  SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     PIX      X      SYS     SYS     SYS     SYS     SYS     SYS
mlx5_6  SYS     SYS     SYS     SYS     PXB     PXB     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS      X      PXB     SYS     SYS     SYS     SYS
mlx5_7  SYS     SYS     SYS     SYS     PXB     PXB     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     PXB      X      SYS     SYS     SYS     SYS
mlx5_8  SYS     SYS     SYS     SYS     SYS     SYS     PXB     PXB     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS      X      PXB     SYS     SYS
mlx5_9  SYS     SYS     SYS     SYS     SYS     SYS     PXB     PXB     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     PXB      X      SYS     SYS
mlx5_10 SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS      X      PIX
mlx5_11 SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     SYS     PIX      X

Legend:

  X    = Self
  SYS  = Connection traversing PCIe as well as the SMP interconnect between NUMA nodes (e.g., QPI/UPI)
  NODE = Connection traversing PCIe as well as the interconnect between PCIe Host Bridges within a NUMA node
  PHB  = Connection traversing PCIe as well as a PCIe Host Bridge (typically the CPU)
  PXB  = Connection traversing multiple PCIe bridges (without traversing the PCIe Host Bridge)
  PIX  = Connection traversing at most a single PCIe bridge
  NV#  = Connection traversing a bonded set of # NVLinks
pankajd@scn1-mn:~$
```

GPU ID	CPU / Core Identifier (NUMA Affinity / Domain)	IB Device
0	3	mlx5_0
1	3	mlx5_1
2	1	mlx5_2
3	1	mlx5_3
4	7	mlx5_4
5	7	mlx5_5
6	5	mlx5_6
7	5	mlx5_7

Infiniband Switch

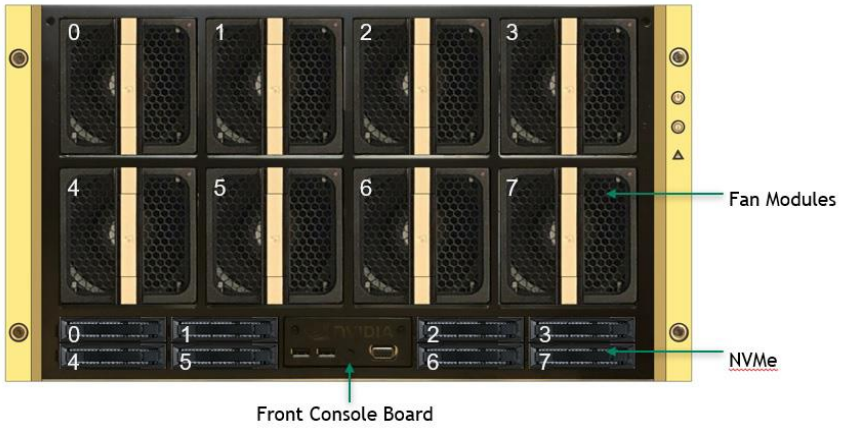


Switch	NVIDIA Quantum CS 8500 MODULAR SWITCH
Ports	800 HDR 200Gb/s ports 1,600 HDR100 100Gb/s ports
Performance	320Tb/s aggregate switch throughput
Cooling	Liquid-cooled
Weight	747 KG

Storage Types

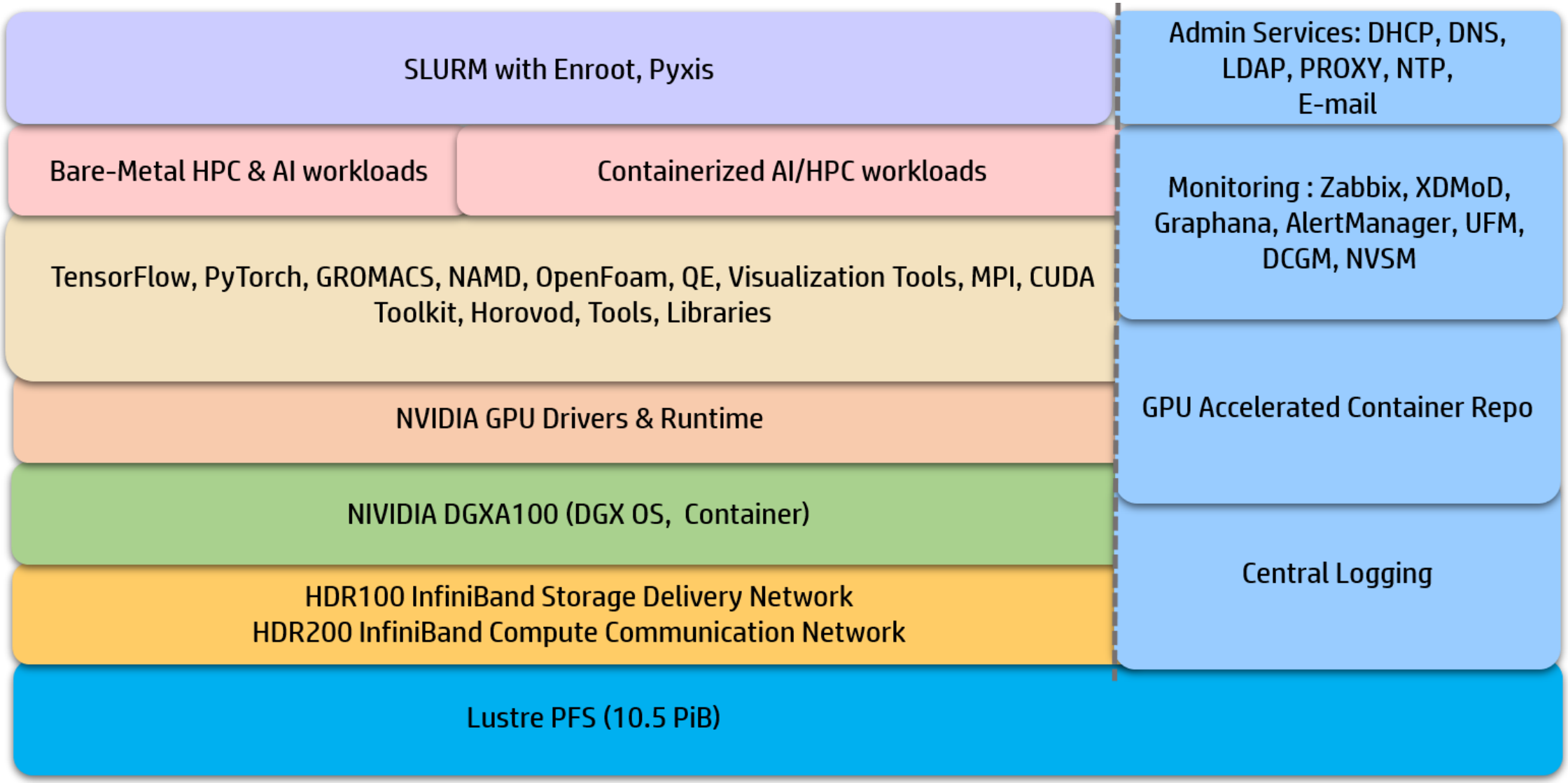
Network Attached Shared Storage

Local Storage



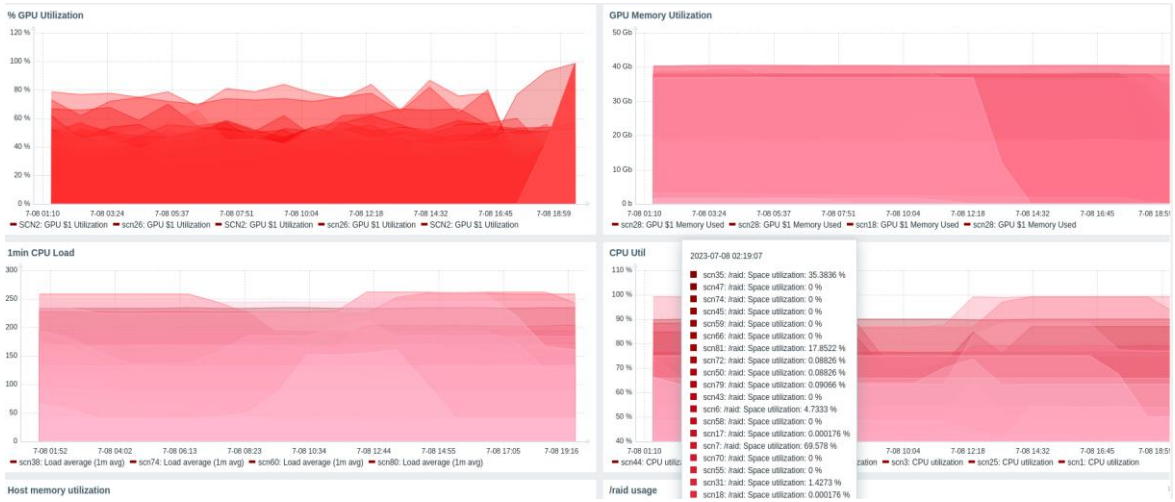
Type of Storage	Description	Performance	Capacity	Mountpoint
Shared Storage (Through network)	Fast parallel Lustre based storage (NLSAS based)	250 GB/Sec, 4 M IOPs	10.5 PB	**/nlsasfs**
Local Storage (Each compute node having 4 NVMe Drives in RAID 0)	Fast Storage	~25 GB/Sec Performance	14 TB	**/raid**

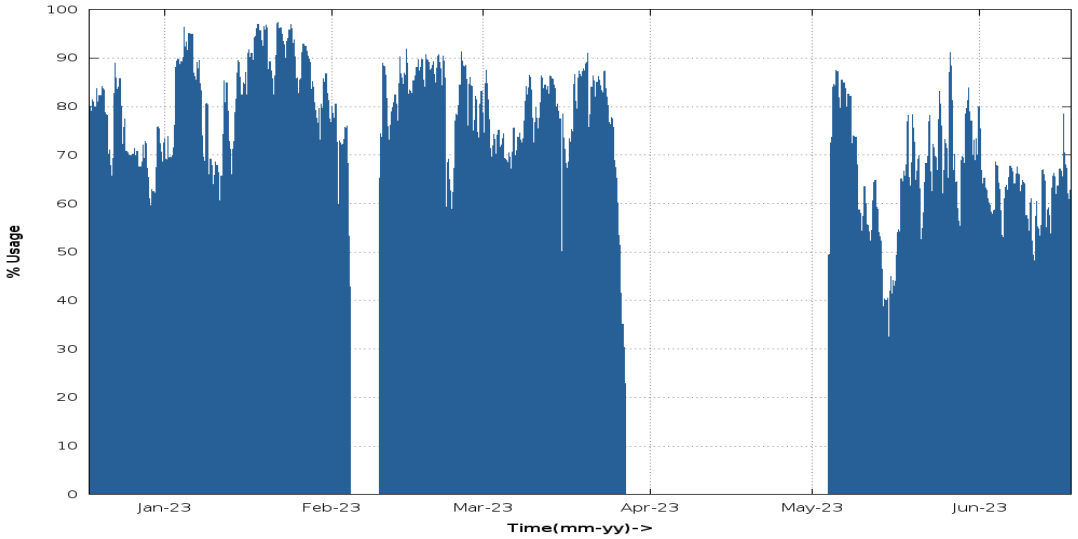
Software Stack

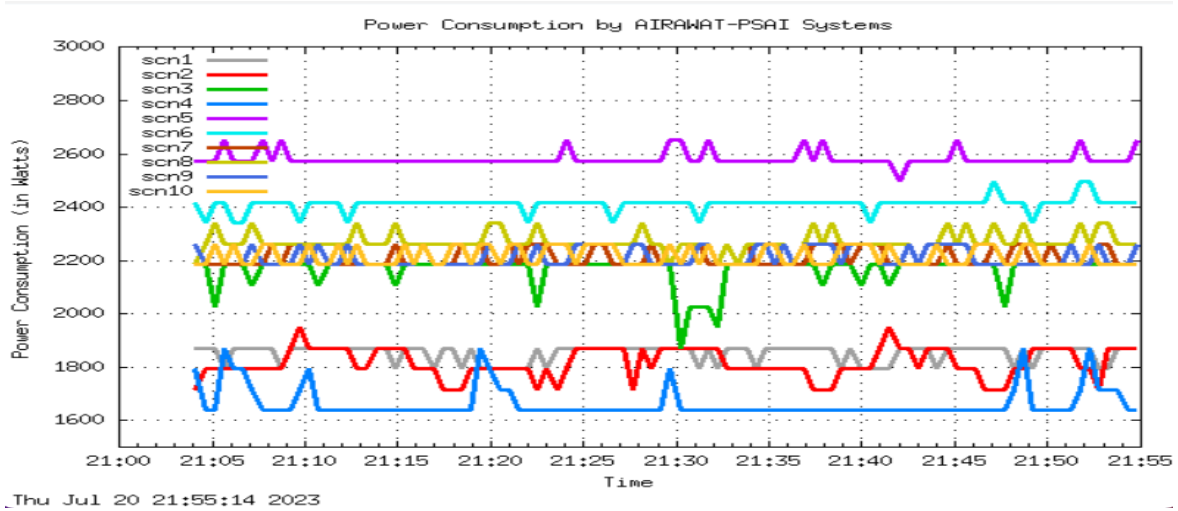


Monitoring: Compute

- System Stats, Health, Utilization, power / temperature

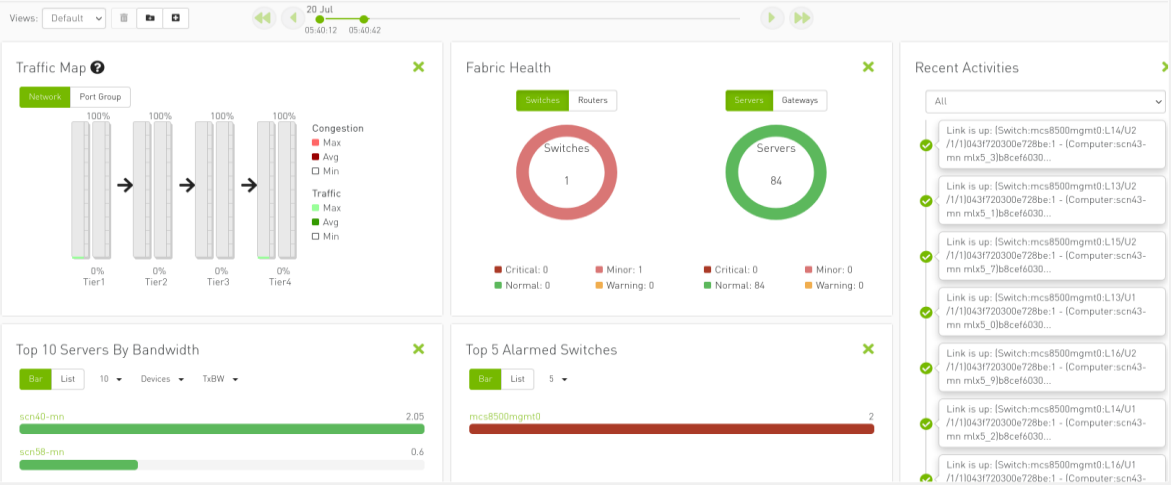




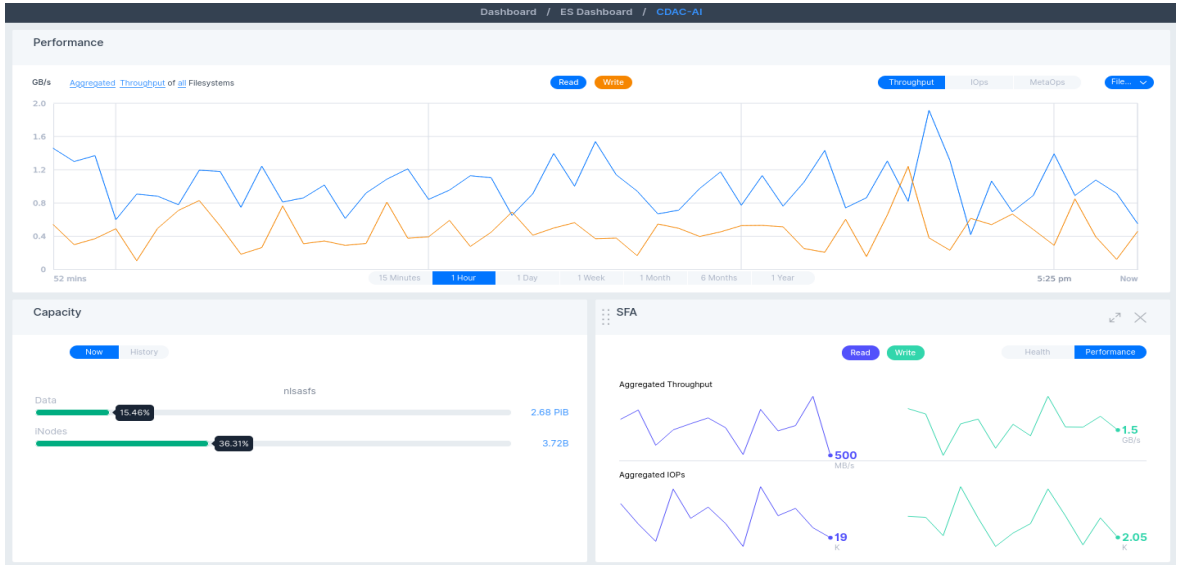


Monitoring : Network and Storage

Infiniband Director Switch Monitoring



Infiniband Fabric Management and Monitoring

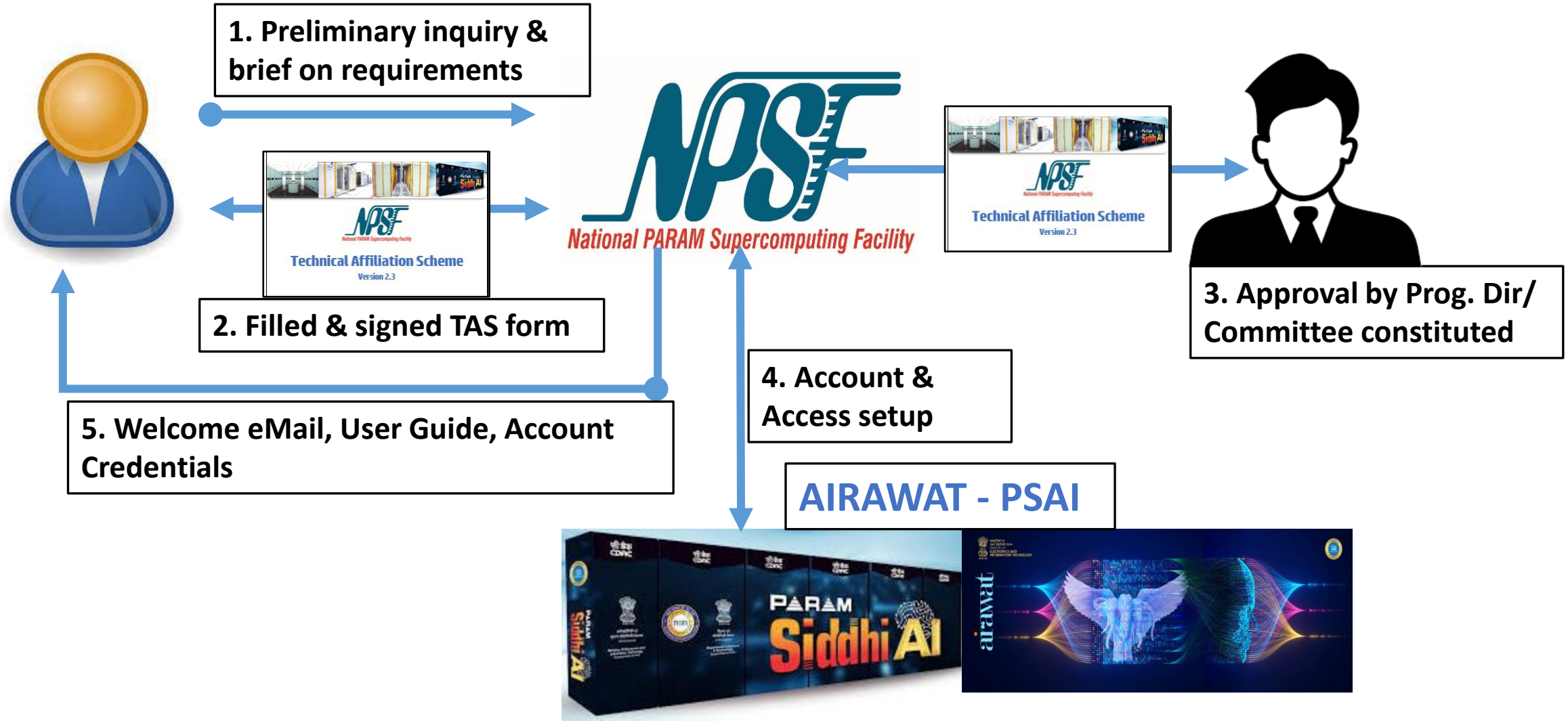


Storage Monitoring

Technical Affiliation Scheme (TAS)

- To leverage NPSF HPC-AI Infrastructure, a person needs to first become a Technical Affiliate of the National PARAM Supercomputing Facility. For this, the prospective user / affiliate have to fill up the account request form.
- The form is to be filled for every single project that the prospective affiliate wants to register for using PARAM Siddhi-AI system. The primary prospective affiliate / user who registers as the Chief Investigator is expected to be a Faculty member / Research Scientist/CEO/CTO/MD/Professor organization/institute/university or manager in the Industry / Startup.
- The Chief Investigator can request for additional accounts associated with this project co-workers as well as collaborators.
- Once the committee/Prog. Dir. approves the request, respective accounts are created and on-boarding concludes.

Technical Affiliation Scheme (TAS): Application Flow



Charging Model

	GPU Charges (NVIDIA A100)						Storage Charges	Registration Charges
Type of Organization	GPU	Hourly (INR/GPU/Hr)	1 Month Reserved	3 Month Reserved	6 Month Reserved	12 Month Reserved	One Month	One Year
R&D Govt./ Academia/PSU/ Startup	1XA100	₹160	₹70,080 ₹96 per hour (40% discount)	₹2,03,232 ₹92.8 per hour (42% discount)	₹3,85,440 ₹88 per hour (45% discount)	₹7,00,800 ₹80 per hour (50% discount)	Allocation of 1 TB @ Rs ₹350 per month	₹30,000
Industry	1XA100	₹170	₹74,460 ₹102 per hour (40% discount)	₹2,15,934 ₹98.6 per hour (42% discount)	₹4,09,530 ₹93.5 per hour (45% discount)	₹7,44,600 85 per hour (50% discount)	Allocation of 1 TB @ Rs ₹350 per month	₹30,000

Statement of account

NATIONAL PARAM SUPERCOMPUTING FACILITY

STATEMENT OF ACCOUNT

INVOICE No.: NPSF/Siddhi/2023/06/[redacted] acc

Chief Investigator : Mahesh Bhargava

Institute : [redacted]

Address : [redacted]

E-Mail ID : [redacted]

Contact No. : [redacted]

Project : [redacted] acc

System : **PARAM Siddhi-AI**

Invoice Date : **05-Jul-2023**

Invoice Period : **01-Jun-2023 to 30-Jun-2023**

Project Start Date : **21-Sep-2021**

Summary of Usage & Charges Applicable

Infrastructure Head	Units	Rate (₹)	Charges(in eMoney)*
NVIDIA A100 40GB GPU	25917.946 (in GPU Hours)	160/GPU/Hour	4146871.360
Additional Project Storage Quota	11 (in TeraBytes)	350/TB	3850.000
Total			**41,50,721.360**
Round Off			**-0.360**
Total (after Round Off)			**41,50,721.000**

Summary of Association & Jobs

5 Total No. of Users **196** Total No. of Jobs

For NPSF, C-DAC
(This is a computer generated invoice and no signature is required.)
Authorized Signatory

This in not a commercial invoice and does not carry any financial implications.

*1 eMoney is equivalent to 1 INR

Note :
1. Default storage quota is 1 TB, which is not chargeable
2. Rates are exclusive of any applicable discounts, subsidies, and taxes
3. The Statement of Account should not be treated as commercial invoice

NATIONAL PARAM SUPERCOMPUTING FACILITY

Itemized Usage For Jobs

Project : [redacted]-acc

Invoice Period : **01-Jun-2023 to 30-Jun-2023**

Sr. No.	Job ID	User	Start Date-Time (dd/mm/yyyy-hh:mm:ss)	End Date-Time (dd/mm/yyyy-hh:mm:ss)	GPUs Allocated	GPU Hours
1	[redacted]	[redacted]	01/06/2023-00:00:00	01/06/2023-22:20:34	8	178.742
2	[redacted]	[redacted]	01/06/2023-00:00:00	01/06/2023-10:22:36	8	83.013
3	[redacted]	[redacted]	01/06/2023-00:00:00	01/06/2023-10:37:48	16	170.080
4	[redacted]	[redacted]	01/06/2023-10:12:48	01/06/2023-10:14:35	16	0.476
5	[redacted]	[redacted]	01/06/2023-10:33:45	01/06/2023-10:40:10	8	0.856
6	[redacted]	[redacted]	01/06/2023-	01/06/2023-	[redacted]	[redacted]

NATIONAL PARAM SUPERCOMPUTING FACILITY

Project : [redacted]-acc

Invoice Period : **01-Jun-2023 to 30-Jun-2023**

Sr. No.	Job ID	User	Start Date-Time (dd/mm/yyyy-hh:mm:ss)	End Date-Time (dd/mm/yyyy-hh:mm:ss)	GPUs Allocated	GPU Hours
204	[redacted]	[redacted]	29/06/2023-22:12:08	29/06/2023-22:19:32	8	0.987
Total						**25917.946**

National PARAM Supercomputing Facility(NPSF),
Centre for Development of Advanced Computing(C-DAC),
Innovation Park, Panchavati, Pashan,
Pune-411008, India
e-mail: [redacted]

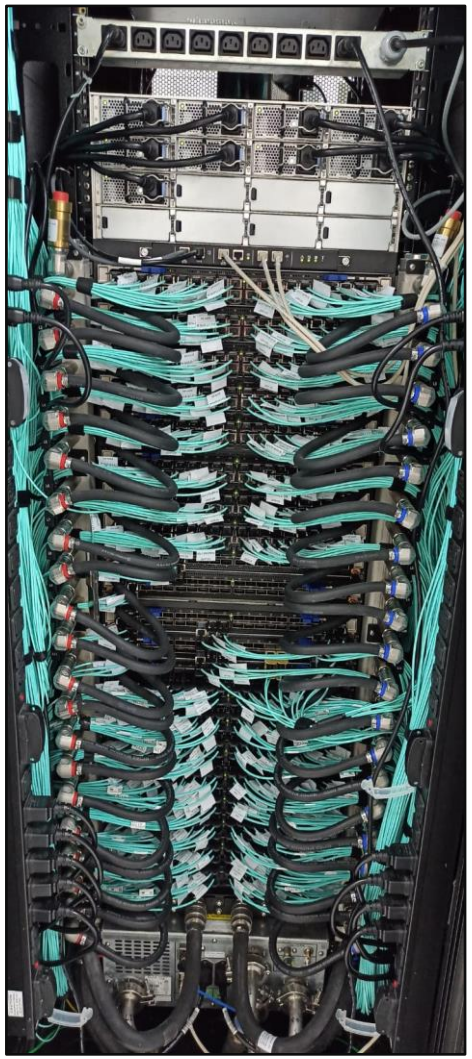
AIRAWAT-PSAI

PARAM Siddhi-AI

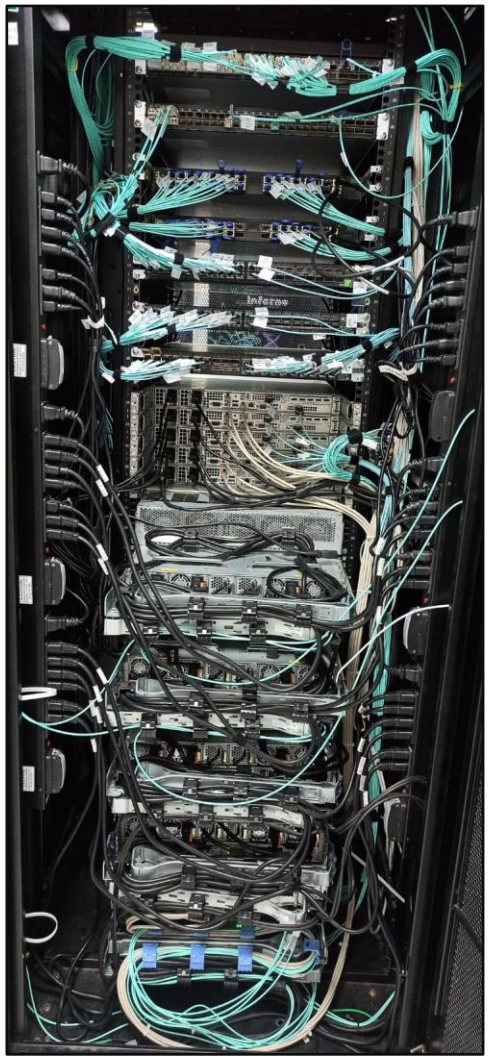
One Vision. One Goal… Advanced Computing for Human Advancement…

AIRAWAT PoC

Cabling : Back side of the Racks

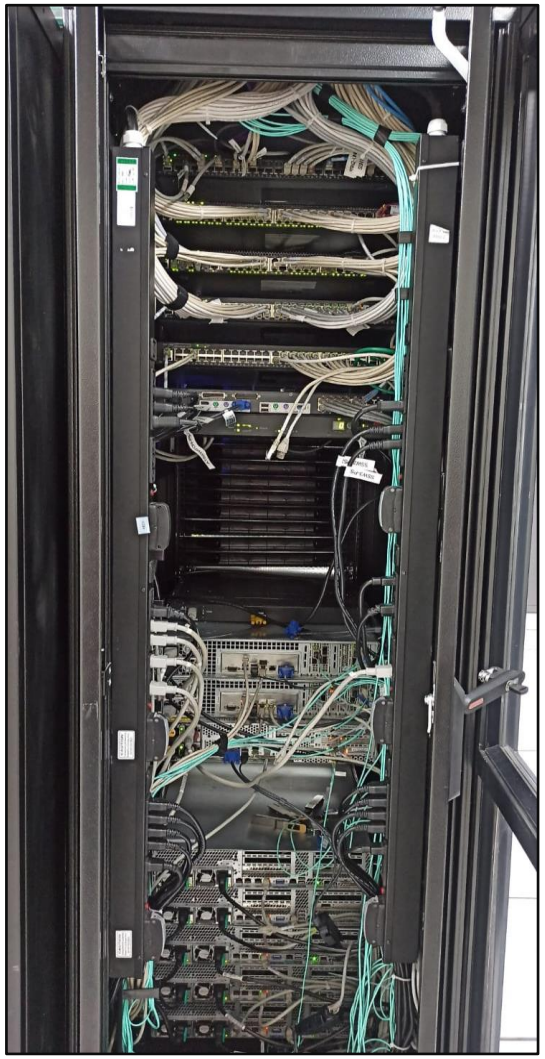


Director Switch



10.5 PiB Storage

DGX-A100 Compute Rack



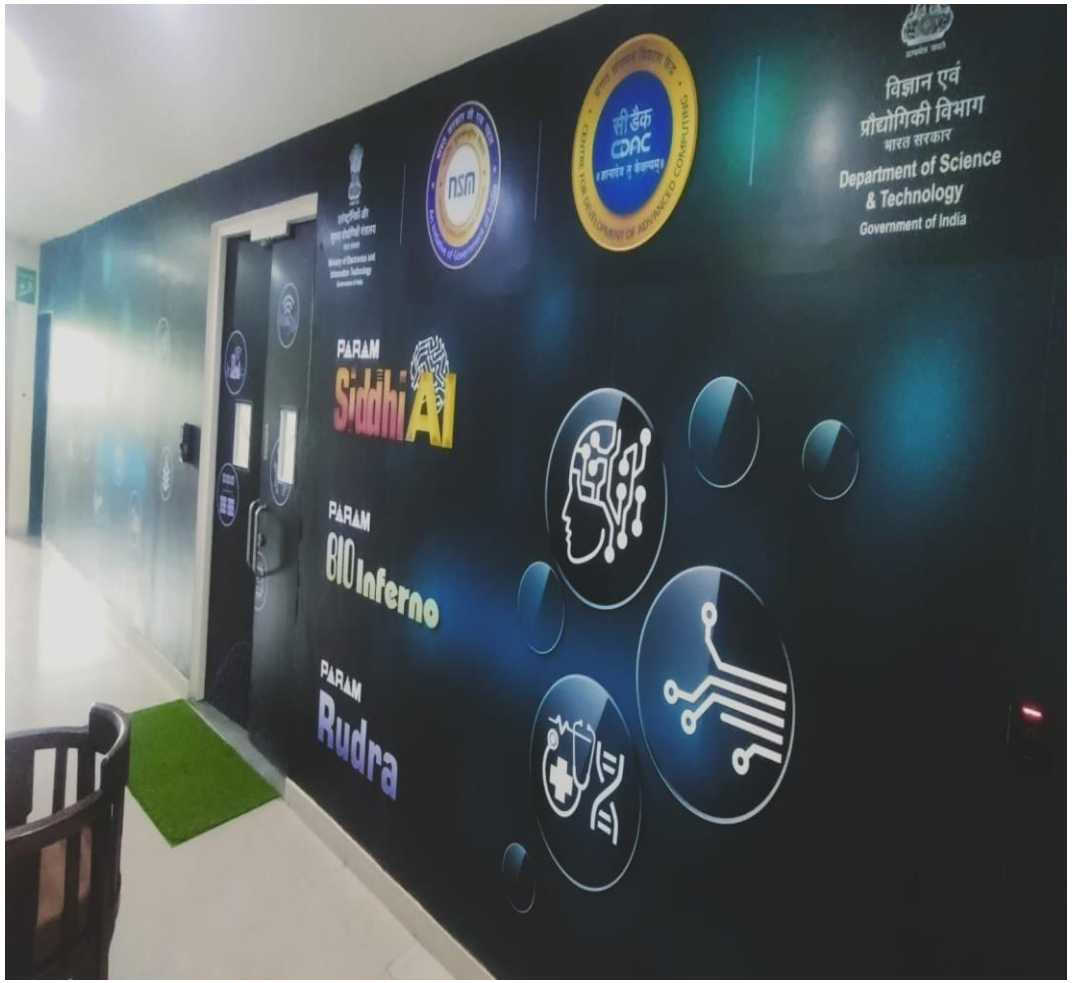
Services and Management network

Cooling Infra

- 40 KW Rack Density RDHx Units
- 17 nos. of RDHx enabled Racks across 2 DCs

- 40 TON Air Conditioners

NPSF : Data Center



Electrical Infra (Transformer, DG Set and UPS)

Transformer

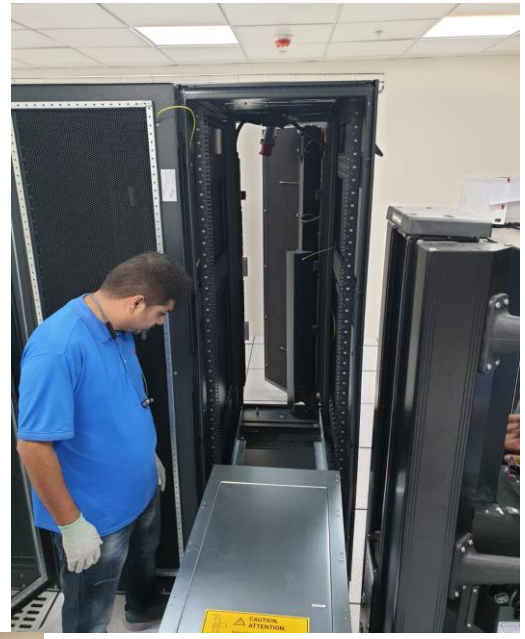
DG Set

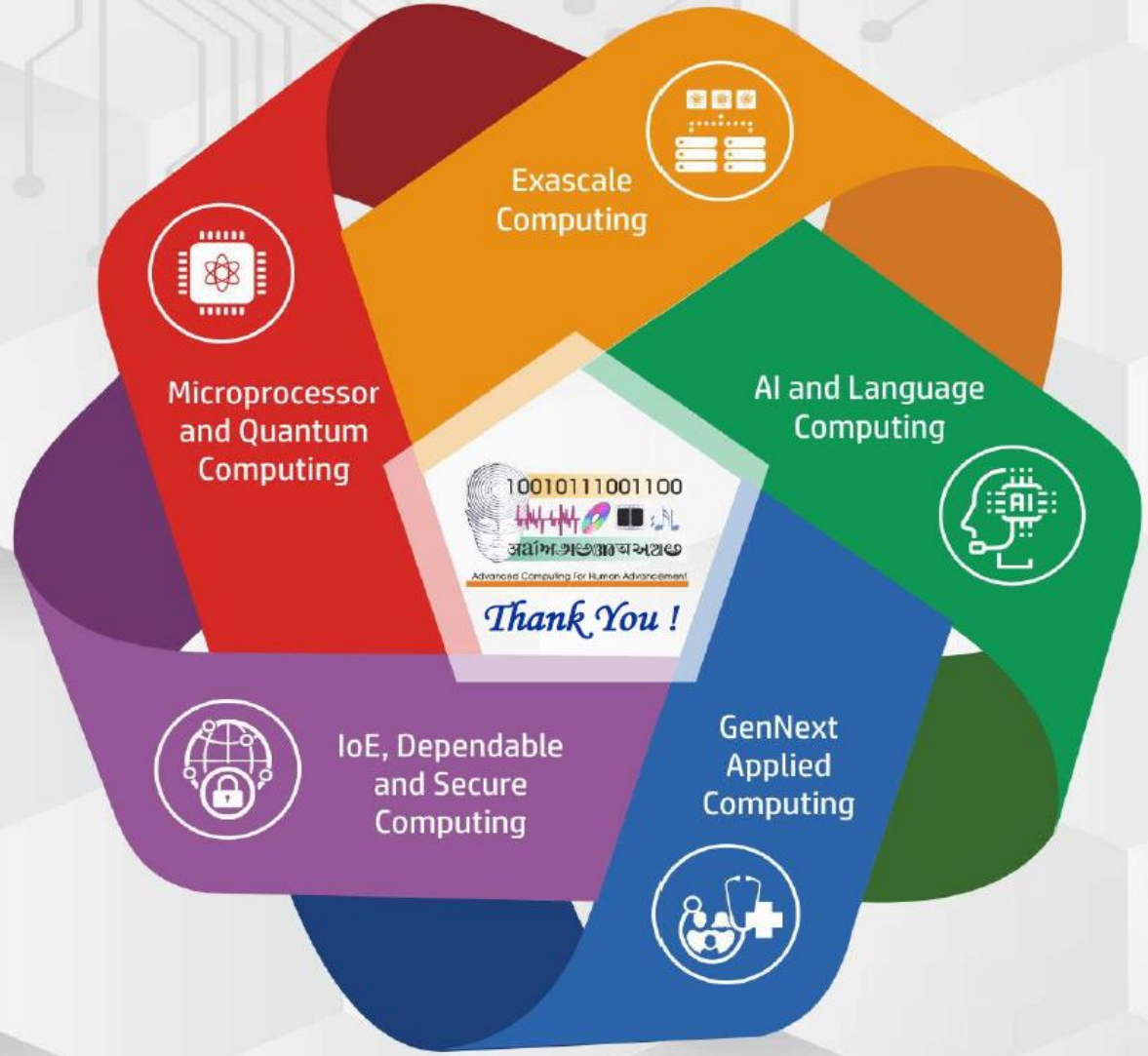
UPS

Transformer	2 MVA
Diesel Generator Set	2 MW
Modular UPS	800 KVA

Commissioning

Commissioning





One Vision. One Goal... Advanced Computing for Human Advancement...