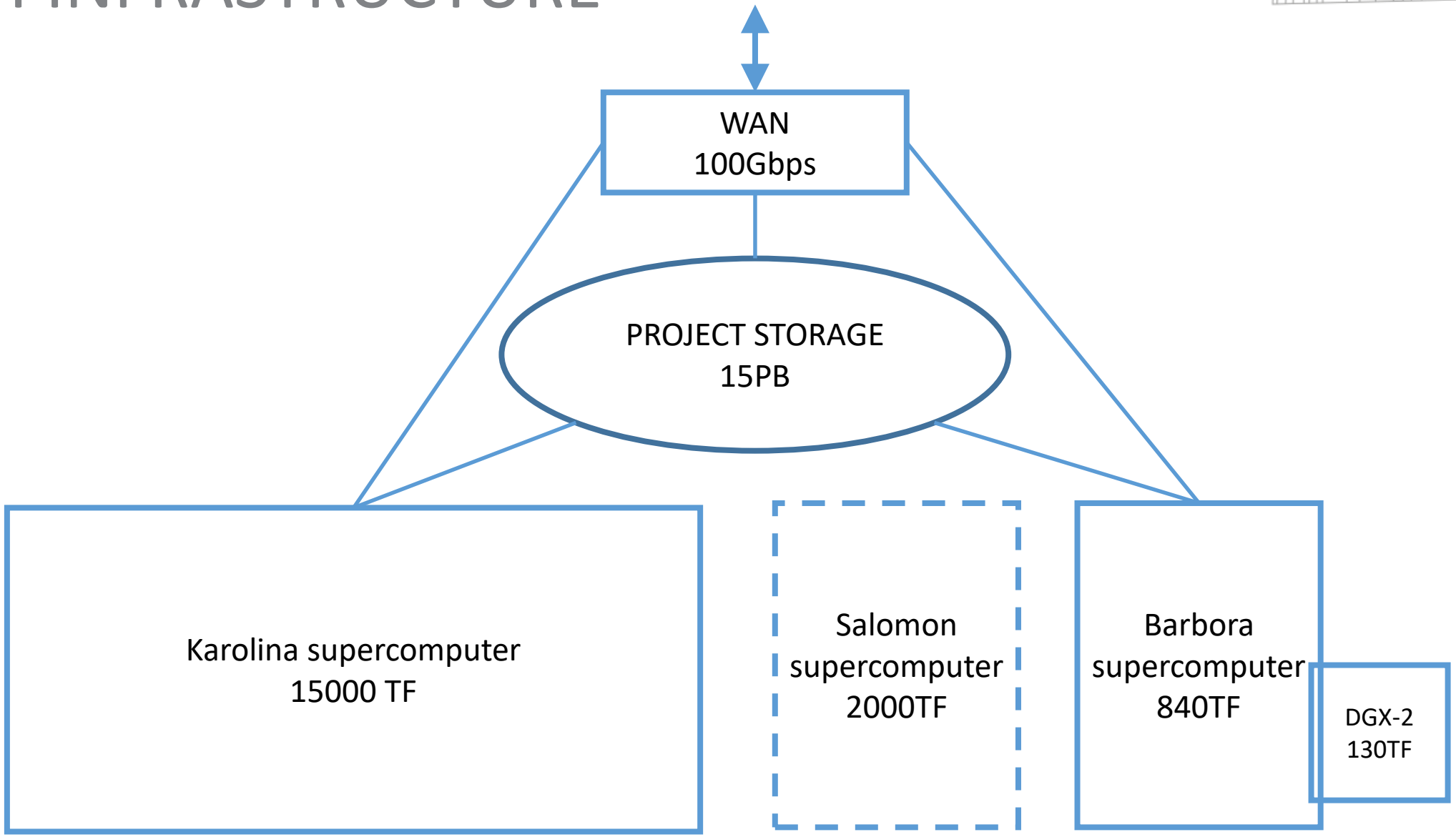
VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

KAROLINA SUPERCOMPUTER

CZECH REPUBLIC EUROHPC JU

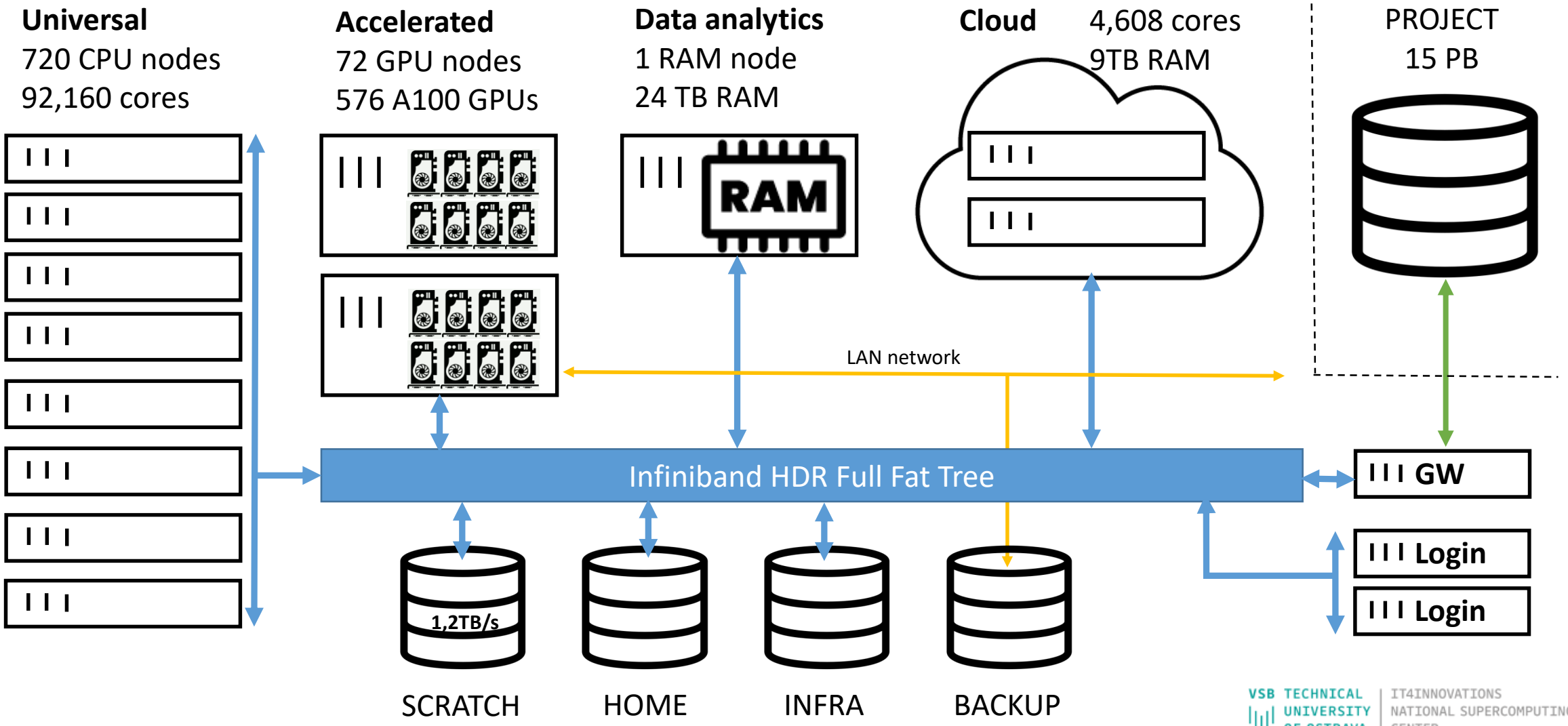
Branislav Jansík

IT4I INFRASTRUCTURE



KAROLINA SUPERCOMPUTER

KAROLINA ARCHITECTURE



UNIVERSAL PARTITION

- **720x HPE Proliant XL225n server**
- **2x AMD EPYC 7H12, 2x64 cores**
- **256GB RAM DDR4**
- **100Gb/s (HDR100)**
- **CentOS 7**
- **3816 TF Peak**

UNIVERSAL NODE PERFORMANCE

Processor performance

2x AMD EPYC 7H12, 2.60GHz

F64 FMA 5 317 Gflop/s

Memory performance

32GB RAM DDR4
25 GB/s

8x32 = 256 GB RAM DDR4
200 GB/s

NUMA Matrix

node	0	1	2	3	4	5	6	7
0:	10	12	12	12	32	32	32	32
1:	12	10	12	12	32	32	32	32
2:	12	12	10	12	32	32	32	32
3:	12	12	12	10	32	32	32	32
4:	32	32	32	32	10	12	12	12
5:	32	32	32	32	12	10	12	12
6:	32	32	32	32	12	12	10	12
7:	32	32	32	32	12	12	12	10

UNIVERSAL NODE PERFORMANCE

Processor performance

2x AMD EPYC 7H12, 2.60GHz

F64 FMA 5 317 Gflop/s

Memory performance

32GB RAM DDR4
25 GB/s

8x32 = 256 GB RAM DDR4
200 GB/s

NUMA Matrix (measured)

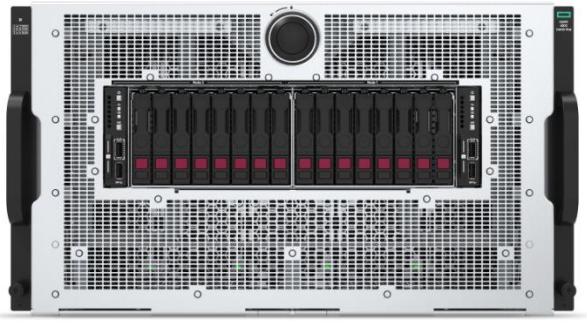
node	0	1	2	3	4	5	6	7
0:	10	11	11	11	13	13	13	13
1:	11	10	11	11	13	13	13	13
2:	11	11	10	11	13	13	13	13
3:	11	11	12	10	13	13	13	13
4:	13	13	13	13	10	11	11	11
5:	13	13	13	13	11	10	11	11
6:	13	13	13	13	11	11	10	11
7:	13	13	13	13	11	11	11	10

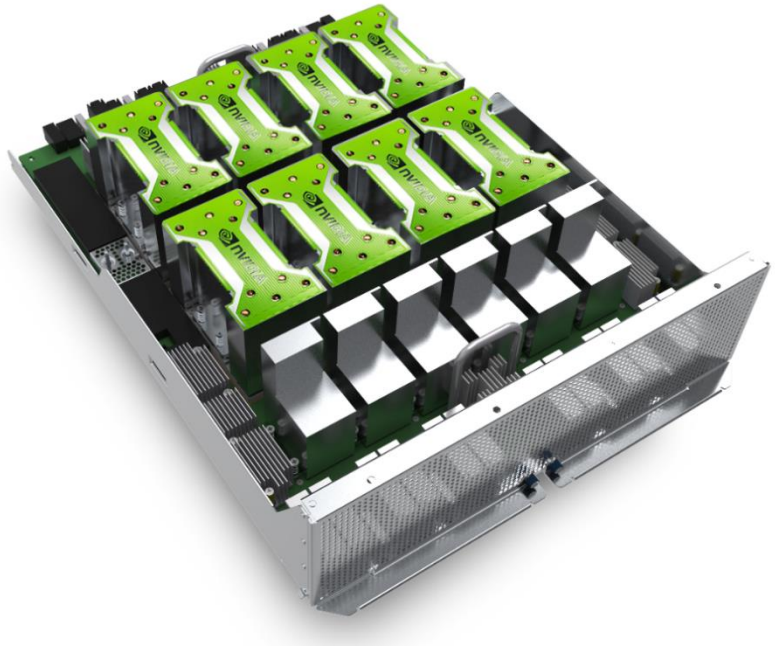
CLOUD PARTITION

- **36x HPE Proliant XL225n server**
- **2x AMD EPYC 7H12, 2x64 cores**
- **256GB RAM DDR4**
- **100Gb/s (HDR100)**
- **<u>10Gb/s Ethernet</u>**
- **<u>2x480 NVMe</u>**
- **CentOS 7**
- **190 TF Peak**

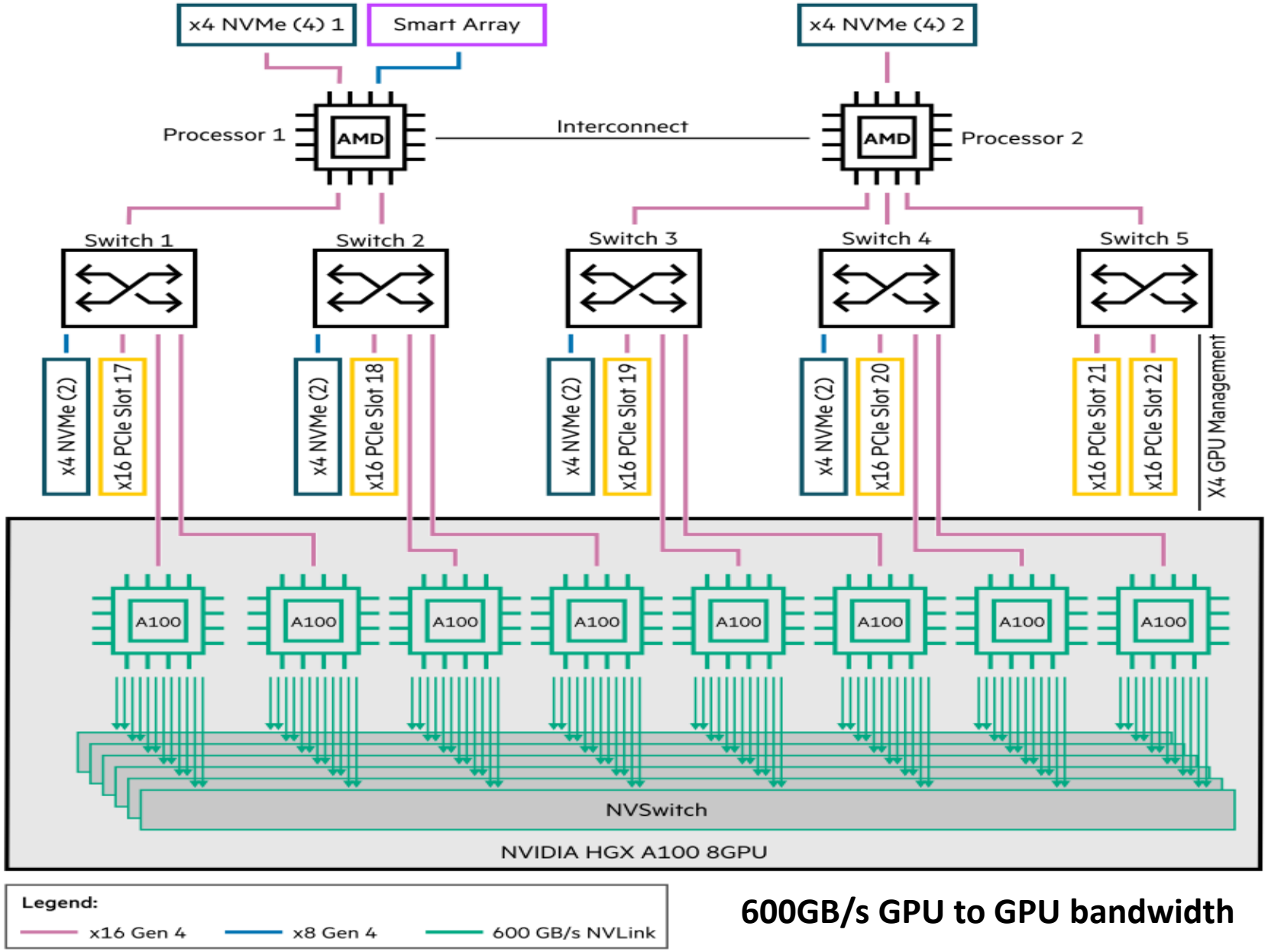
ACCELERATED PARTITION

- **72x HPE Apollo 6500 G10+**
- **2x AMD EPYC 7763, 2x64 cores**
- **8x Nnvidia A100, 40GB HBM2**
- **1024GB RAM DDR4**
- **4x200Gb/s HDR**
- **CentOS 7**
- **11088 TF Peak**





NVIDIA A100 GPU



600GB/s GPU to GPU bandwidth

KAROLINA ACN UNIFIED ADDRESS SPACE

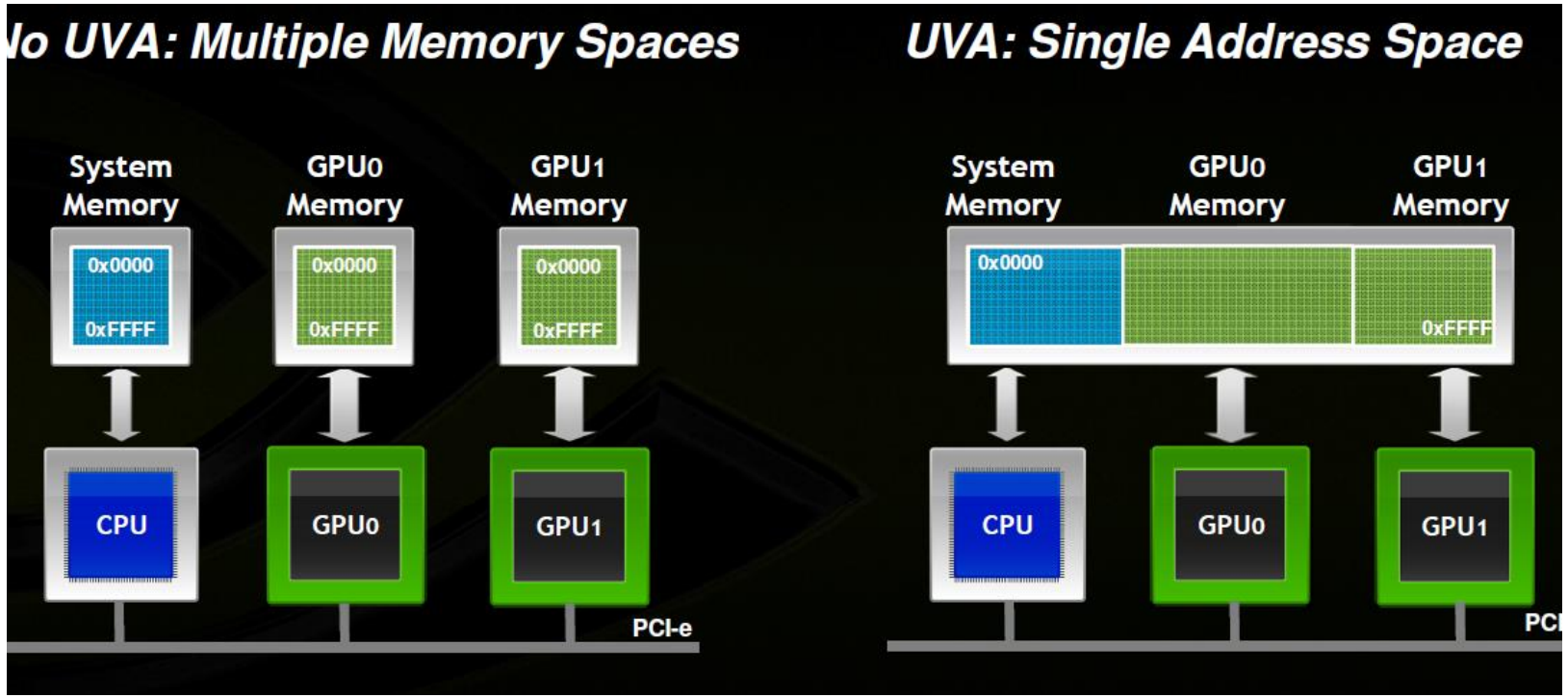
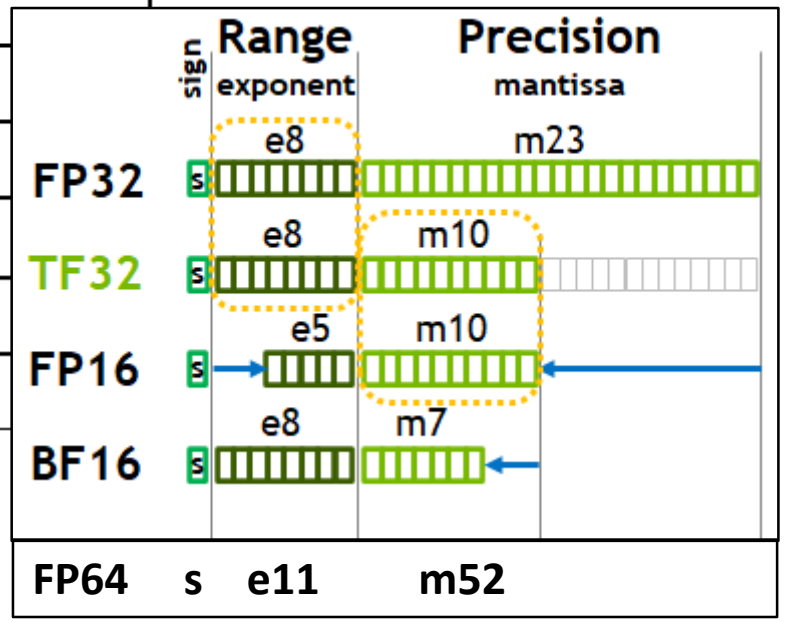
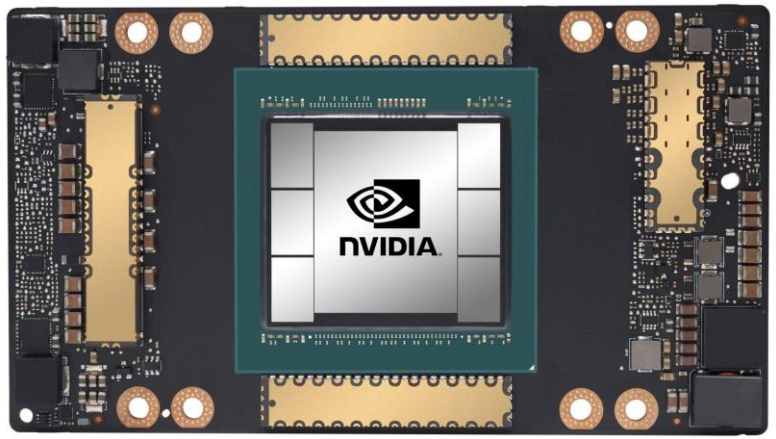


Image courtesy Tim C. Shroeder, NVIDIA, https://developer.download.nvidia.com/CUDA/training/cuda_webinars_GPUDirect_uva.pdf

NVIDIA A100 GPU, 108SM

Peak FP64[1]	9.7 TFLOPS
Peak FP64 Tensor Core[1]	19.5 TFLOPS
Peak FP32[1]	19.5 TFLOPS
Peak FP16[1]	78 TFLOPS
Peak BF16[1]	39 TFLOPS
Peak TF32 Tensor Core[1]	156 TFLOPS | 312 TFLOPS[2]
Peak FP16 Tensor Core[1]	312 TFLOPS | 624 TFLOPS[2]
Peak BF16 Tensor Core[1]	312 TFLOPS | 624 TFLOPS[2]
Peak INT8 Tensor Core[1]	624 TOPS | 1,248 TOPS[2]
Peak INT4 Tensor Core[1]	1,248 TOPS | 2,496 TOPS[2]





ACCELERATED NODE PERFORMANCE

Processor performance

8xNvidia A100-SXM4-40GB

F64 FMA 77 888 Gflop/s
F64 WMMA 155 443 Gflop/s

F16 WMMA 2482 995 Gflop/s

Memory performance

40GB RAM HBM2
1330 GB/s

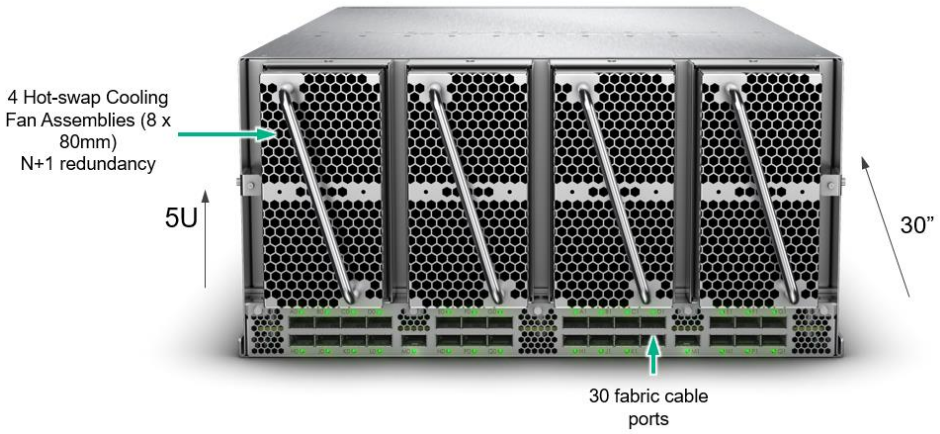
8x40 = 320 GB RAM HBM2
10640 GB/s

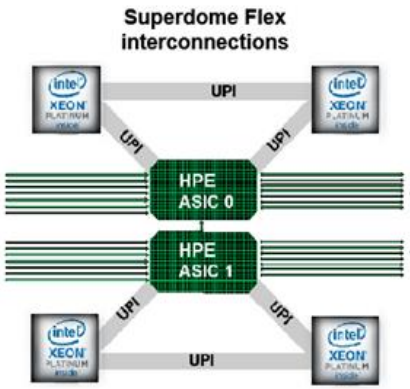
NUMA Matrix (measured)

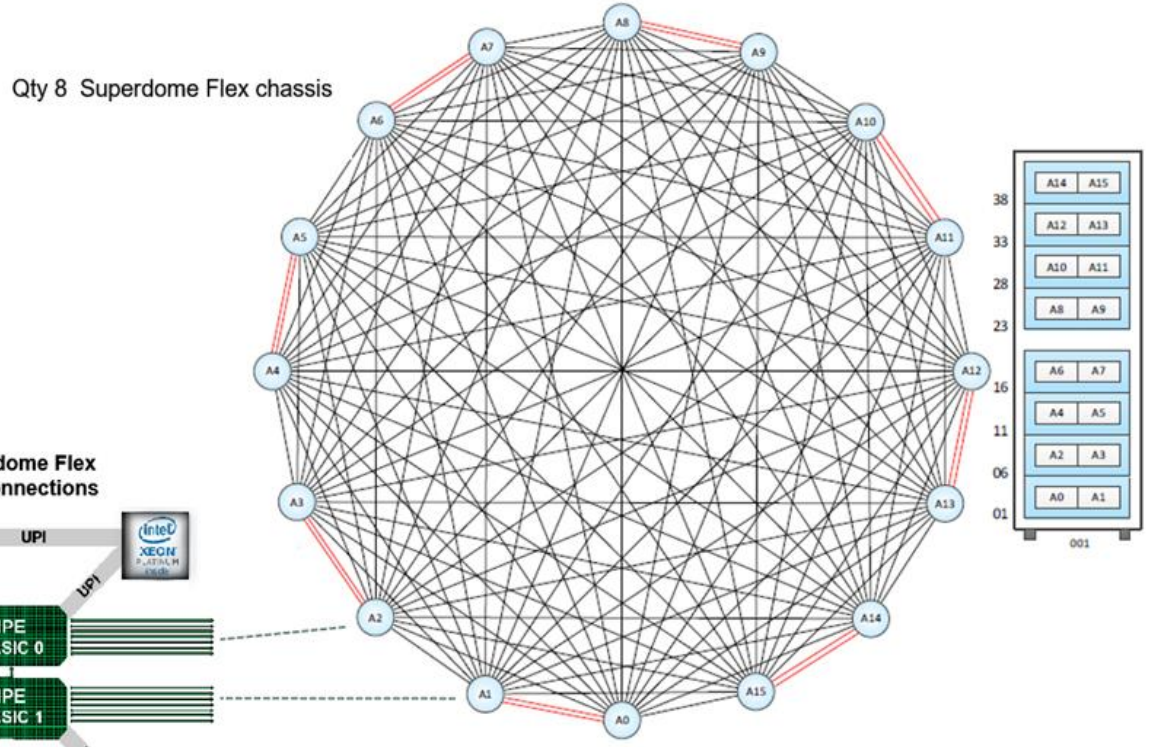
node	0	1	2	3	4	5	6	7
0:	10	39	39	39	39	39	39	39
1:	39	10	39	39	39	39	39	39
2:	39	39	10	39	39	39	39	39
3:	39	39	39	10	39	39	39	39
4:	39	39	39	39	10	39	39	39
5:	39	39	39	39	39	10	39	39
6:	39	39	39	39	39	39	10	39
7:	39	39	39	39	39	39	39	10

DATA ANALYTICS PARTITION

- **1xHPE Superdome Flex**
- **32x Intel Xeon 8268, 32x24 (768 cores)**
- **24576GB RAM DDR4**
- **2x200Gb/s HDR**
- **RedHat 7**
- **70 TF Peak**







UNIVERSAL PARTITION NODE PERFORMANCE

Processor performance

32x Platinum 8268 CPU, 2.90GHz

F64 FMA-512 60 000 Gflop/s

Memory performance

768GB RAM DDR4
80 GB/s

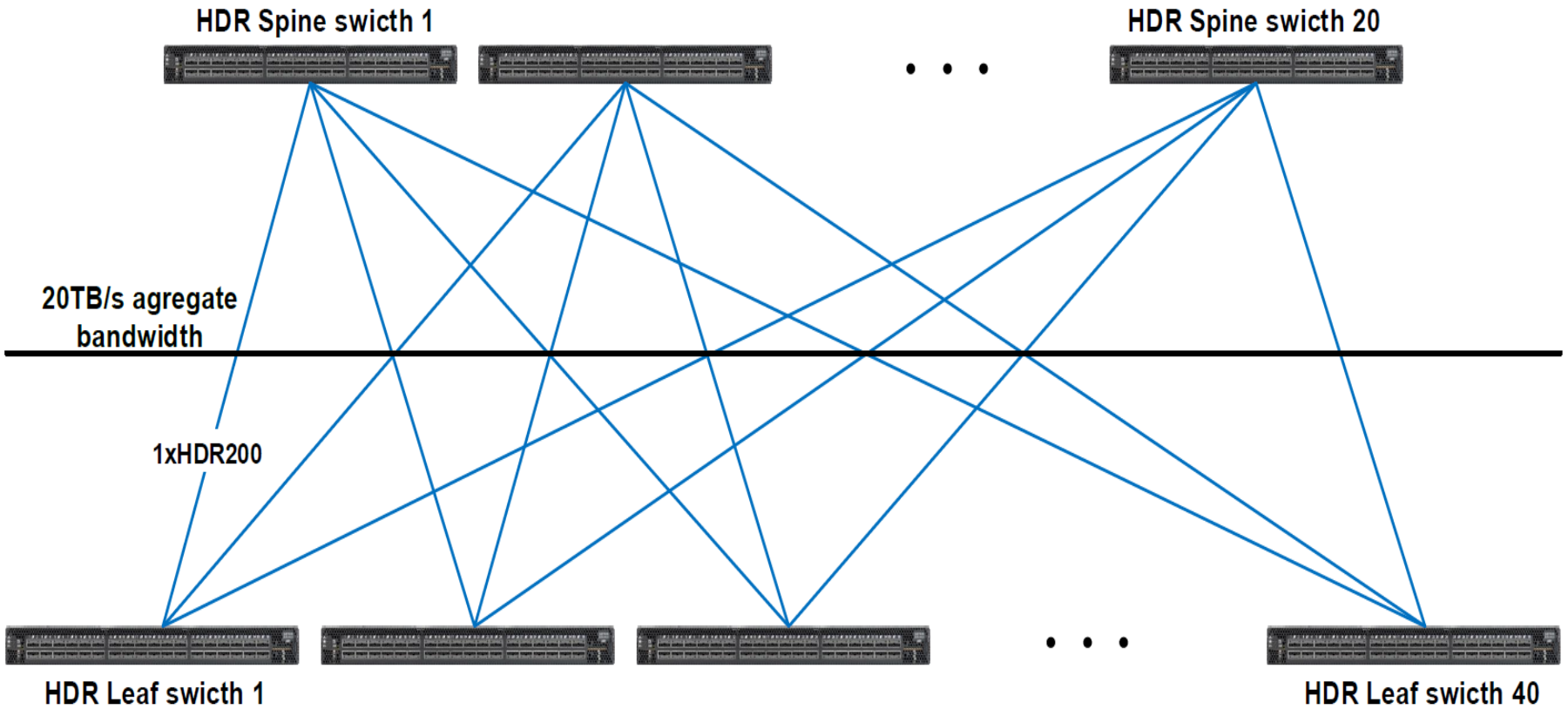
24576 GB RAM DDR4
2560 GB/s

NUMA Matrix

node	0	1	2	31
0:	10	16	24	43
1:	16	10	16	43
2:	24	16	10	43
31:	43	43	43	10

COMPUTE NETWORK

Technology: HDR
Toplogy: Non-Blocking Fat Tree
Throughput: 200Gb/s for HDR200 connection, 100Gb/s for HDR100 connection
Latency: Expected less than 3 microseconds



KAROLINA PERFORMANCE

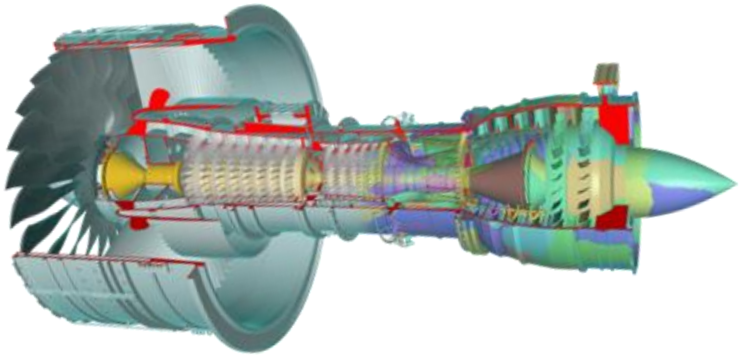
Performance stats:

- **R_Peak: 15.7 PFlop/s**
- **R_Max: 9.1 PFlop/s (LINPACK)**
- **R_AI: 350 PFlop/s (DeepLearning)**

- **Universal partition**: 2.3 PFlop/s (LINPACK) (720 nodes)
- **Accelerated partition**: 6.6 PFlop/s (LINPACK) (72 nodes)
  350 PFlop/s (DeepLearning)
- **Data analytics partitition**: 40 TFlop/s (LINPACK)
- **Cloud partitition**: 131 TFlop/s (LINPACK) (36 nodes)

Estimated TOP 500 ranking:

- **Ranking (1H2021): #69** (worldwide) **#19** (Europe)
- **Ranking #15** (worldwide) **Green 500**

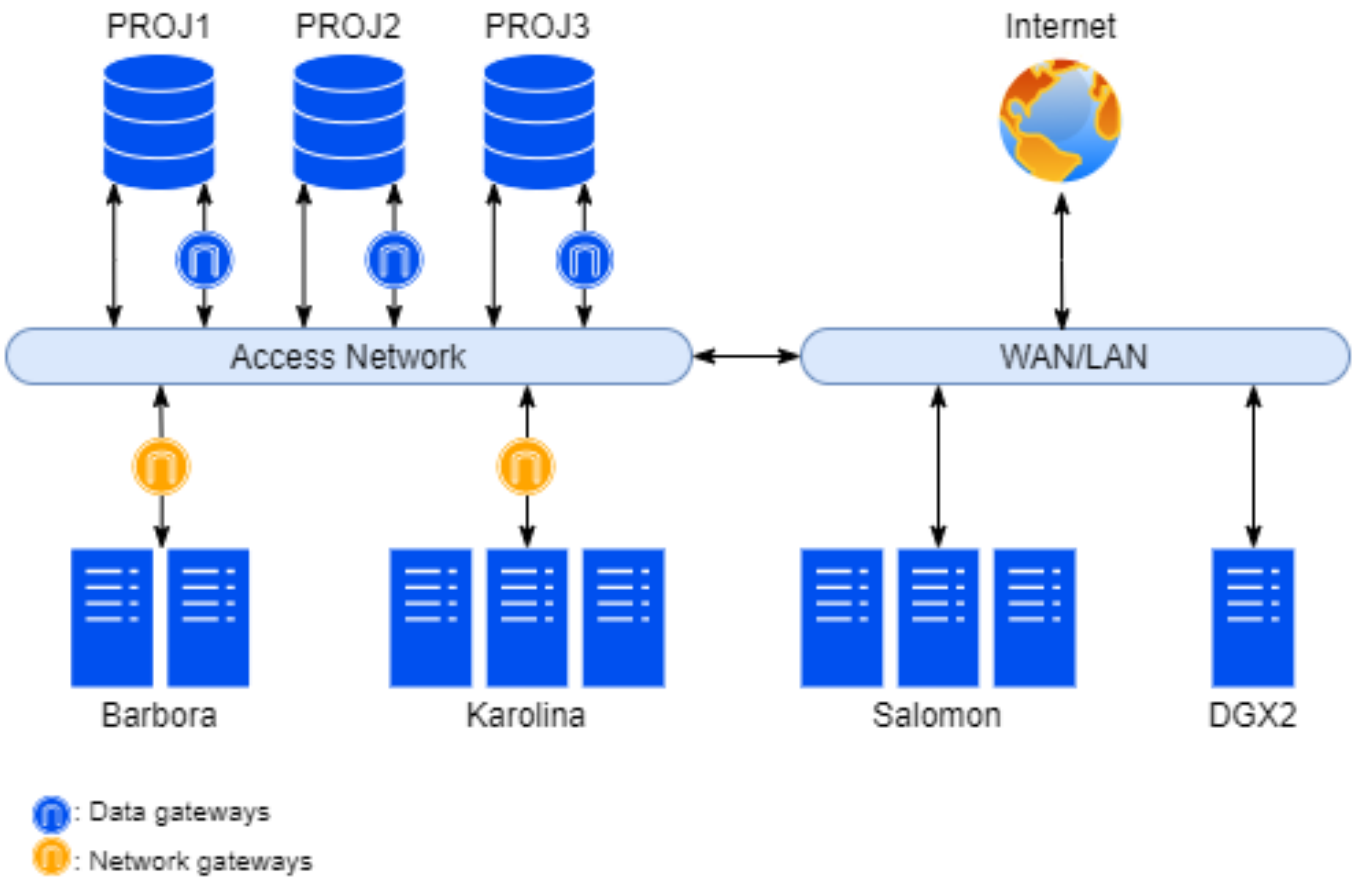


THE PROJECT STORAGE

- Independent
- Extendable
- Scalable
- Redundant

- 3x5 PB
- 39GB/s agregated
- NFS protocol
- Data gateways (GridFTP, RSYNC, etc)

THE PROJECT STORAGE



- Independent
- Extendable
- Scalable
- Redundant

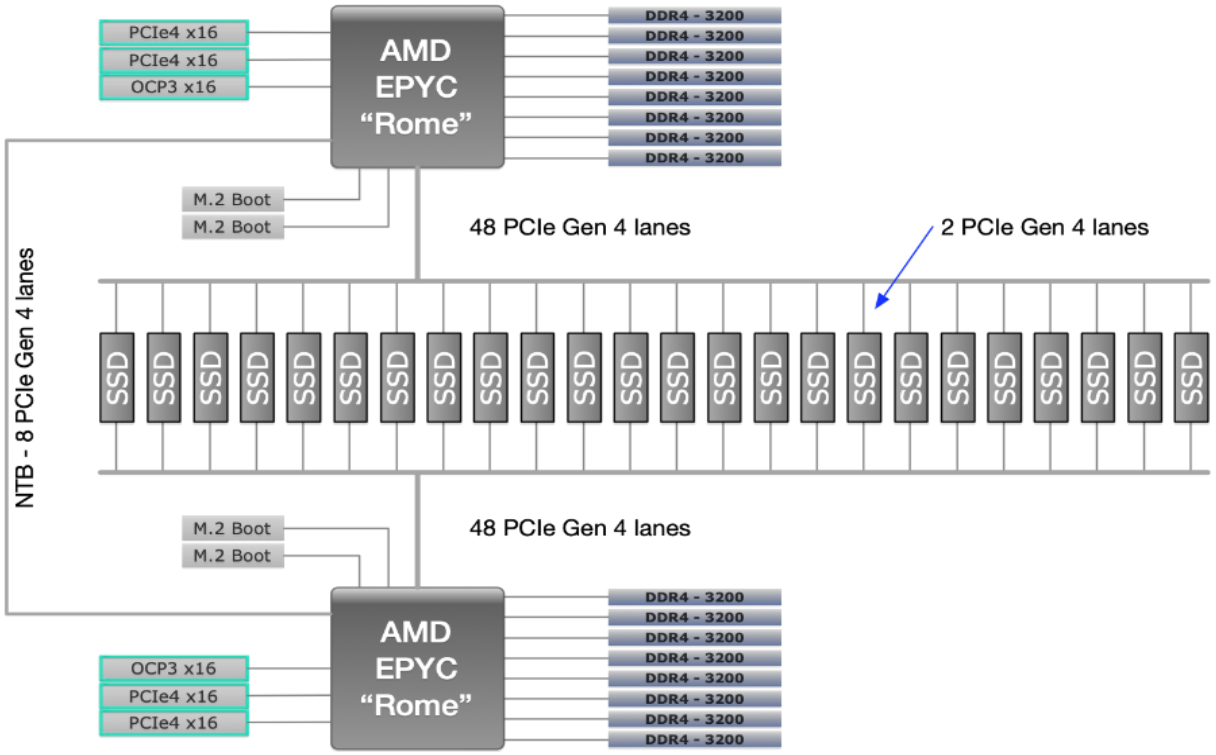
- 3x5 PB
- 39GB/s agregated
- NFS protocol
- Data gateways (GridFTP, RSYNC, etc)

THE PROJECT STORAGE

PROJECT	
Mountpoint	/mnt/proj{1,2,3}/PROJECT_ID
Capacity	15PB
Throughput	39GB/s
IO Performance	57kIOPS
Default project space quota	20TB
Default project inodes quota	20 mil.
Protocol	NFS

SCRATCH STORAGE

- **ClusterStor E1000 All Flash**
- **1xSMU (system mgmt)**
- **1xMDU (metadata ctl)**
- **24xSSU-F (storage unit)**
- **Size 1330TB**
- **Throughput 1200GB/s** All flash
- **LUSTRE Filesystem**




```
$ lfs check osts
$ lfs getstripe filename
$ lfs setstripe –s stripe_size –c stripe_count –o stripe_offset filename
```

SCRATCH STORAGE

SCRATCH STORAGE	
Mountpoint	/scratch/projects/PROJECT_ID/
Capacity	1361 TB
Throughput	730.9 GB/s write, 1198.3 GB/s read
PROJECT quota	20 TB
PROJECT inodes quota	20 M
Default stripe size	1 MB
Default stripe count	1
Protocol	Lustre

KAROLINA DOCUMENTATION

https://docs.it4i.cz/karolina/introduction

IT4Innovations Documentation | Search | sccs/docs.it4i.cz

General | Storage | Clusters | Software

Clusters

- Karolina
  - Introduction
  - Hardware Overview
  - Compute Nodes
  - Storage
  - Network
  - Visualization Servers
- Barbora
- NVIDIA DGX-2
- Salomon
- Archive

Introduction

Karolina is the latest and most powerful supercomputer cluster built for IT4Innovations in Q2 of 2021. The Karolina cluster consists of 829 compute nodes, totaling 106,752 compute cores with 313 TB RAM, giving over 15.2 PFLOP/s theoretical peak performance and is ranked in the top 10 of the most powerful supercomputers in Europe.

Nodes are interconnected through a fully non-blocking fat-tree InfiniBand network, and are equipped with AMD Zen 2, Zen3, and Intel Cascade Lake architecture processors. Seventy two nodes are also equipped with NVIDIA A100 accelerators. Read more in Hardware Overview.

The cluster runs with an operating system compatible with the Red Hat Linux family. We have installed a wide range of software packages targeted at different scientific domains. These packages are accessible via the modules environment.

The user data shared file-system and job data shared file-system are available to users.

The PBS Professional Open Source Project workload manager provides computing resources allocations and job execution.

Read more on how to apply for resources, obtain login credentials and access the cluster.

Table of contents

Actions

- Edit This Page
- Request Change
- Get Support

NS
ERCOMPUTING

SUMMARY

- IT4INNOVATIONS – Czech national supercomputing center

- **Karolina EUROHPC supercomputer** - 9.1 PFlop/s Linpack
- Massively accelerated - 8x Nvidia Ampere A100 per node (6.6 PFlop/s)
- Partial acceptance April 2021, full acceptance June 2021

Our supercomputers support science, industry, and society

Branislav Jansík
branislav.jansik@vsb.cz

IT4Innovations National Supercomputing Center
VSB – Technical University of Ostrava
17. listopadu 2172/15
708 00 Ostrava-Poruba, Czech Republic
www.it4i.cz

VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

IT4I ACCESS MECHANISMS

https://www.it4i.cz/en/for-users/computing-resources-allocation

VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

EUROCC E-INFRA CZ EXTRANET | CZ

ABOUT INFRASTRUCTURE RESEARCH INDUSTRY COOPERATION FOR USERS EDUCATION EVENTS

OPEN ACCESS COMPETITION

Institutions can apply for computational resources within Open Access Grant Competitions. The grant competition is announced 3 times a year (February, June, October) for employees of research, scientific and educational organizations.

RECENT OPEN CALL

You can also apply for projects in the form of "multi-year" open access, in which computing resources are provided for a period of 18, 27, or 36 months. The purpose is to support long-term scientific grants.

MORE ABOUT THE MULTI-YEAR APPROACH

IT4I ACCESS MECHANISMS

- Open Access Competition
  - Employees of research, scientific and educational organizations
  - 3 times a year (February, June, October), 9 months utilization
  - Multiyear access: 18, 27, 36 months
- EuroHPC JU Grant Competitions
  - Researchers from academia, research institutes, public authorities and industry from Member State or a country associated to Horizon 2020
  - Extreme, Regular, **Benchmark, Development**, Fast track and Industry track, 12 months utilization
- PRACE Access
  - PRACE Preparatory, DECI, Regular Access
  - Open to academia & industry for Open R&D research purpose.
  - Multi year allocations possible for specific calls

IT4I ACCESS MECHANISMS

- Directors Discretion
  - Submit any time
  - Both commercial and non-commercial sectors can apply in case Open Access Grant Competitions cannot be used.
- Rental of computational resources
  - Standard allocation arranged for a specific period with a pre-agreed quota
  - Customized allocation
  - Pay per use

MULTI YEAR OPEN ACCESS

- **Period 18, 27 or 36 months**

- **Eligibility**
  - Success history (1 completed project, registered o outcome)
  - Long term Research project H2020, ERC, EuroHPC, TAČR, GAČR or other
  - Resource utilization plan

- **Rules of use**
  - Resources consumed according to plan
  - Progress report, in 9 months period, review by the Allocation committee

IT4I PARADIGM SHIFT

Switch from Core hours to **Node hours**

Number of node hours requested for each platform

a) Barbora CPU:
b) Barbora GPU:
c) Barbora FAT:
d) DGX-2:
e) Karolina CPU:
f) Karolina GPU:
g) Karolina FAT

KAROLINA EUROHPC ACCESS

Benchmark calls are designed for code scalability tests, the outcome of which is to be included in the proposal in a future EuroHPC Extreme Scale and Regular call. Users receive a limited number of node hours; the maximum allocation period is three (3) months.
Development calls are designed for projects focusing on code and algorithm development and optimisation. This can be in the context of research projects from academia or industry, or as part of large public or private funded initiatives as for instance Centres of Excellence or Competence Centres. Users will typically be allocated a small number of node hours; the allocation period is one (1) year and is renewable up to 2 times.

System/partition	Benchmark		Development	
	Node hours	Core hours	Node hours	Core hours
Karolina CPU	7000	896000	15000	1920000
Karolina GPU	1000	128000	3000	384000
Karolina analytics	9	6912	22	16896

Fixed resources allocated per project, 8 + 8 projects can be supported in each call

GETTING ACCESS

https://www.it4i.cz/en/for-users/computing-resources-allocation

IT4INNOVATIONS
NATIONAL SUPERCOMPUTING
CENTER

EUROCC E-INFRA CZ EXTRANET | CZ

ABOUT INFRASTRUCTURE RESEARCH INDUSTRY COOPERATION FOR USERS EDUCATION EVENTS

OPEN ACCESS COMPETITION

Institutions can apply for computational resources within Open Access Grant Competitions. The grant competition is announced 3 times a year (February, June, October) for employees of research, scientific and educational organizations.

RECENT OPEN CALL

You can also apply for projects in the form of "multi-year" open access, in which computing resources are provided for a period of 18, 27, or 36 months. The purpose is to support long-term scientific grants.

MORE ABOUT THE MULTI-YEAR APPROACH

LUMI SUPERCOMPUTER COMPUTATIONAL RESOURCES

The EuroHPC LUMI supercomputer, currently being implemented in Kajaani, Finland, will be one of the world's fastest computing systems with performance over 550 PFlop/s. The LUMI supercomputer is procured jointly by the EuroHPC Joint Undertaking and the LUMI

GETTING ACCESS

https://scs.it4i.cz

https://scs.it4i.cz

130%

IT4Innovations Information System

Self-service portal to manage your HPC resources

IT4Innovations offers HPC resources which are provided on project basis. You can apply for the resources and manage them here.

By signing up for and by signing in to this service you accept our:

- Acceptable Use Policy
- User´s duties.

Single Sign-On | External users

Sign-in with an IT4Innovations account, a Federation account or a public service account.

Sign in with **It4innovations** account

Sign in with **eduID.cz**

Sign in with **eduGAIN**

Sign-in with a social net if attached to your account.

Sign in with **GitHub**

Sign in with **LinkedIn**

Sign in with **Twitter**

GETTING ACCESS

https://scs.it4i.cz/project_requests/new

Agendas Accounting Requests Extranet PRACE Check lists

New Project request

* Name

Name of the project.

* Type

standard

* Call

Open Access (use Call number 24)

Additional resources:

- [ ] FAT nodes
- [ ] GPU nodes

* PI login

PI Salutation

* Organization

Select...

* Project area

Select...

Address

* Abstract

Include a popular abstract in a form which is immediately available for publication on the website or in newspapers etc., outlining the related research, the methods to be used, and the expected impact, in language appropriate for the general public.
Be concise; do not exceed 1500 characters in abstract

IT4I PROPOSAL EVALUATION

- Review Results
  - https://www.it4i.cz/en/for-users/open-access-evaluation.
  - Be clear of about your objectives
  - Justify the requested resources by an explicit calculation
  - Consider, how your project will contribute to public welfare

- Applicant History
  - Registered publications
  - Number of projects
  - We expect **1 registered publication** per project (in 3 year sliding window)

50:50

GETTING ACCESS

Sender: nobody@it4i.cz
Subject: SCS IT4I| Access to IT4I systems

Dear John Smith,

A project ID **OPEN-0-0** has been assigned to your project Neural 3D Scene Representation.
The 1024000 core-hours allocated to your project will be available from 2021-01-11 00:01:00 +0100 to 2022-06-24 23:59:00 +0200.

The allocation decision may be based on the review of your project. The possible review could be found at https://scs.it4i.cz/, Project requests section.

Instructions on how to proceed ….

GETTING ACCESS

To: support@it4i.cz
Subject: Access to IT4Innovations

Dear support,

Please open the user account for me and attach the account to OPEN-0-0
Name and affiliation: John Smith, john.smith@myemail.com, Department of Chemistry, MIT, US
I have read and accept the Acceptable use policy document (attached)

Preferred username: johnsm

Thank you,
John Smith
(Digitally signed)

GETTING ACCESS

You receive login credentials by protected email.

- username
- SSH private key and private key passphrase
- system password

The clusters are accessed by the **private key** and username.

Username and **password** are used for login to the information systems.

To be continued by Jakub Kropacek

Branislav Jansík
branislav.jansik@vsb.cz

IT4Innovations National Supercomputing Center
VSB – Technical University of Ostrava
17. listopadu 2172/15
708 00 Ostrava-Poruba, Czech Republic
www.it4i.cz

VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER