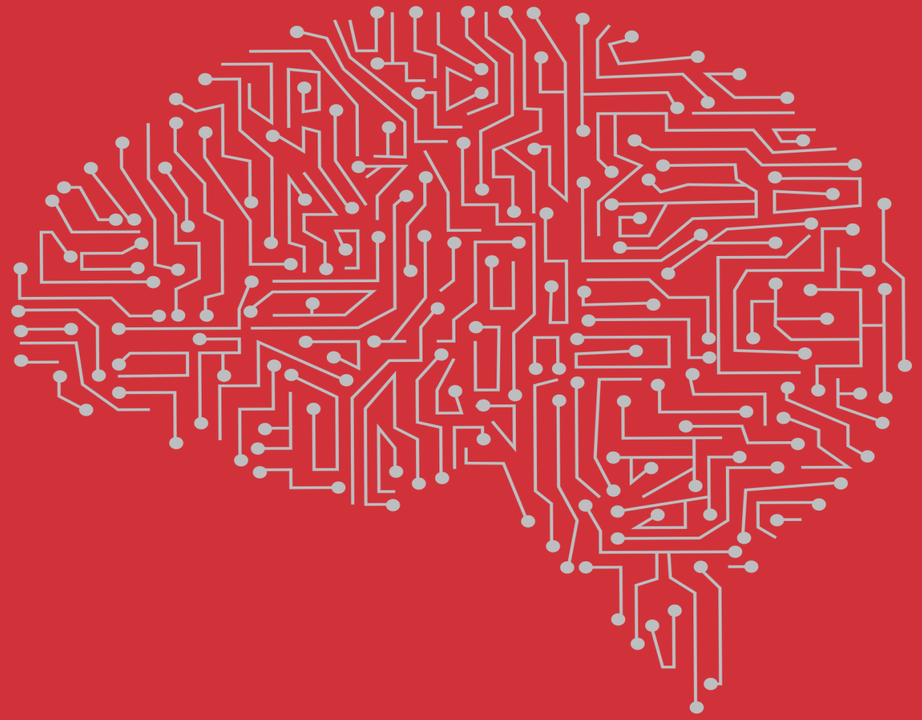


# Implementations of NVIDIA enterprise solutions in the CEE region

Petr Plodik, M Computers

M computers®

NVIDIA ELITE PARTNER

# Implementations of NVIDIA enterprise solutions in the CEE region

Learn about the IT architectures and designs of supercomputers that M Computers has delivered based on NVIDIA enterprise solutions – Volta and Ampere accelerators and DGX systems in the Central and Eastern Europe (CEE) region. We will describe our experiences, best practices and lessons learnt from real-life implementations.

During the session we will cover 3 approaches:

1) Single NVIDIA DGX System
2) NVIDIA reference architectures
3) Tailored solutions with NVIDIA accelerators

Each topic is accompanied by customer references and drawings of each IT architecture.

Summary of 7 steps how to successfully implement HPC and AI projects.

# M Computers profile

- NVIDIA ELITE partner based in the Czech Republic
- Coverage: Czech Republic, Slovakia, rest of CEE
- System integrator of HPC and AI infrastructure
- Focused on the design and delivery of:
    - HPC and AI supercomputers
    - high performance and large-scale storage
    - networking
    - data center integration
    - full software stack, management, scheduling, containers, …
- Consultations on HPC and AI topics.
- Part of NVIDIA test drive program


NVIDIA ELITE PARTNER

# 1. Single NVIDIA DGX System

Products:

- DGX Station, DGX A100,
- or older DGX-1, DGX-2


NVIDIA DGX Station

All-in-one solution for HPC and AI applications:

- compute resources (CPU and GPU)
- fast local storage (SSD/NVMe)
- connectivity to network
- DGX software stack

NVIDIA DGX A100

Applications ready to run minutes after delivery

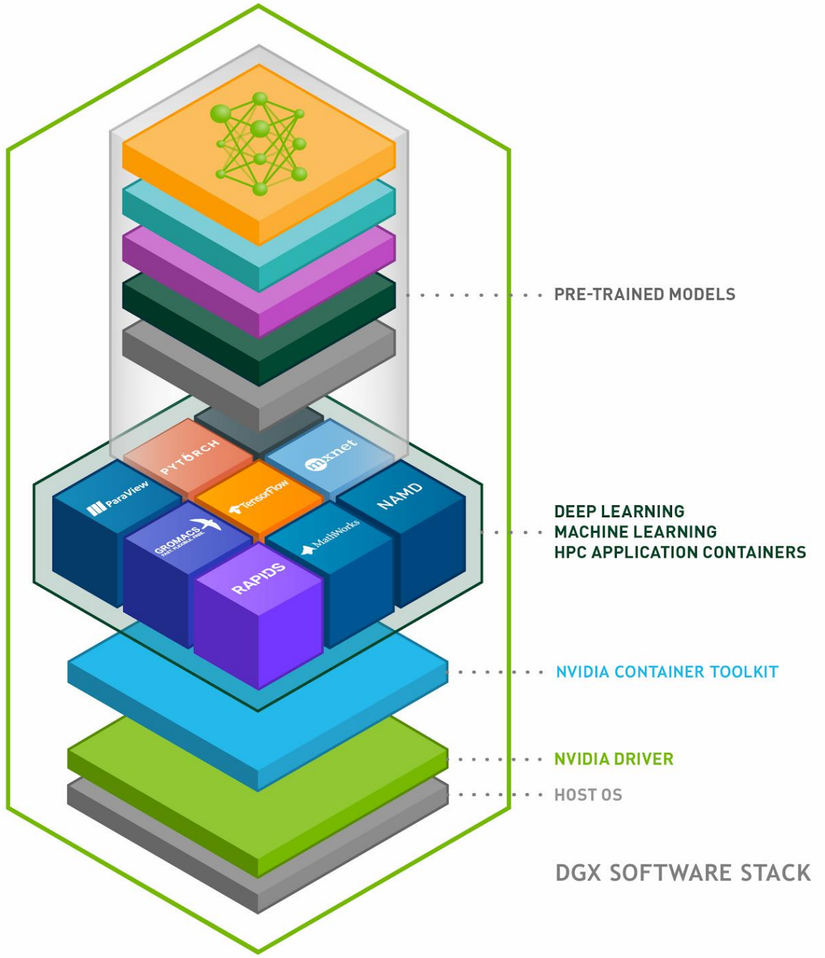
# 1. Single NVIDIA DGX System

NVIDIA DGX Software Stack
NVIDIA NGC Cloud (hub for GPU-optimized software)

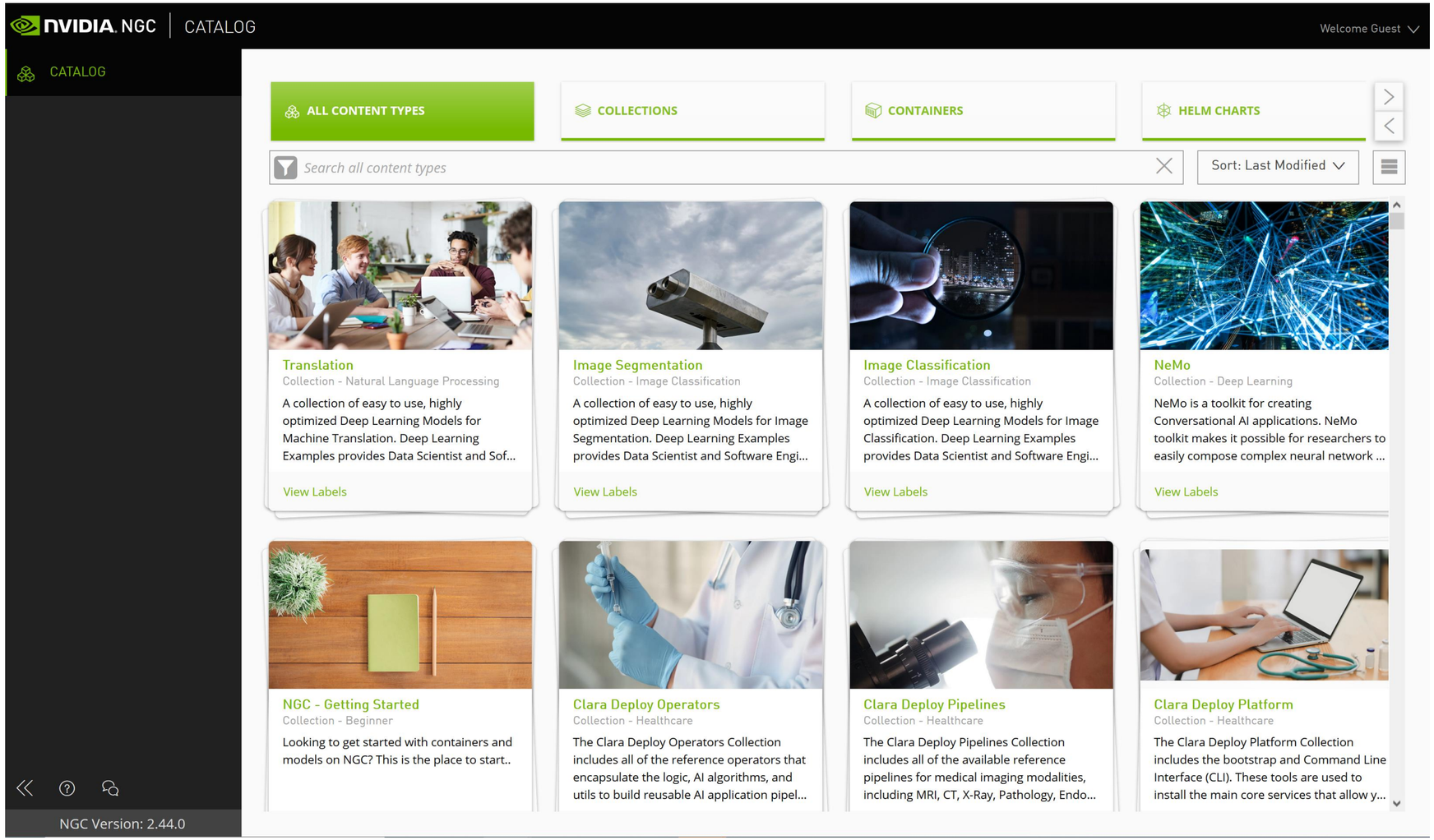
A great solution for:

- testing of HPC and AI applications
- department / workgroup workload
- required performance is within a server

DGX A100 could be divided into 56 GP GPU partitions
NVIDIA Support covers hardware and software stack
Easy migration to larger infrastructure or cloud



# 1. NVIDIA NGC




https://ngc.nvidia.com/

# 1. Single NVIDIA DGX System

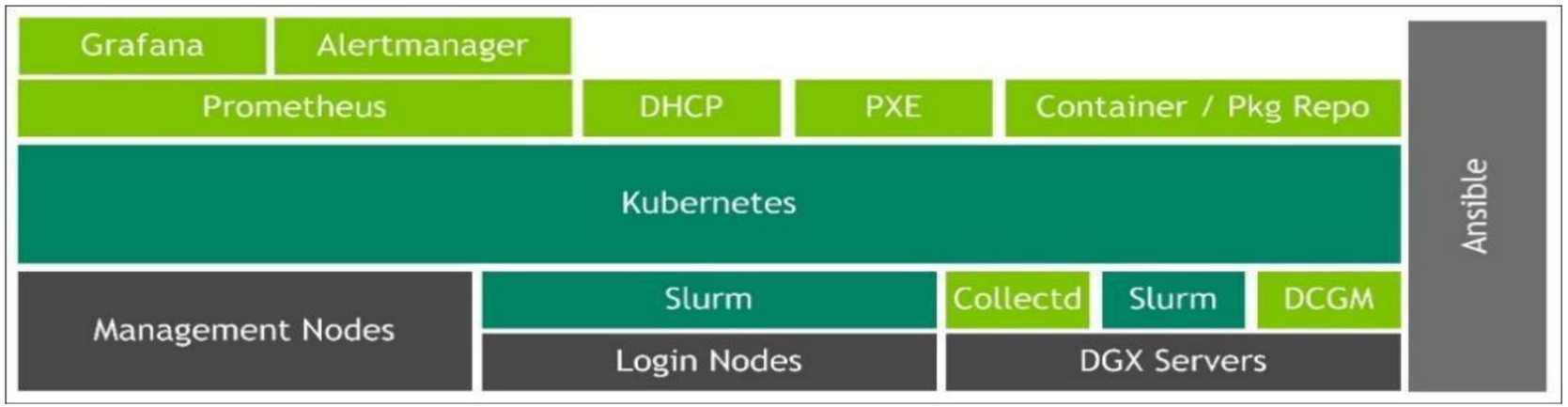
Our references in the CEE region:

- NVIDIA DGX-2 at **IT4Innovations** – National Supercomputing Center, Ostrava, CZ
  - development & testing of new large-scale algorithms for HPC and AI applications
- NVIDIA DGX Station at the **University of Defense**, Hradec Králové
  - HPC applications
- 5× NVIDIA DGX-1 MAXQ at **CIIRC, Czech Technical University**
  - AI algorithms for robotics applications



# 2. NVIDIA DGX reference architecture

- Best practices how to build HPC and AI infrastructure
- Compute nodes, fabrics, storage, system and management software stack
- ONTAP-AI = reference architecture based on NVIDIA and NetApp products
  - DGX systems, NetApp AFF A800, Cisco or Mellanox switches
- Includes data center considerations (power, cooling)
- Performance expectations of AI applications/frameworks
- Management stack is covered by NVIDIA support

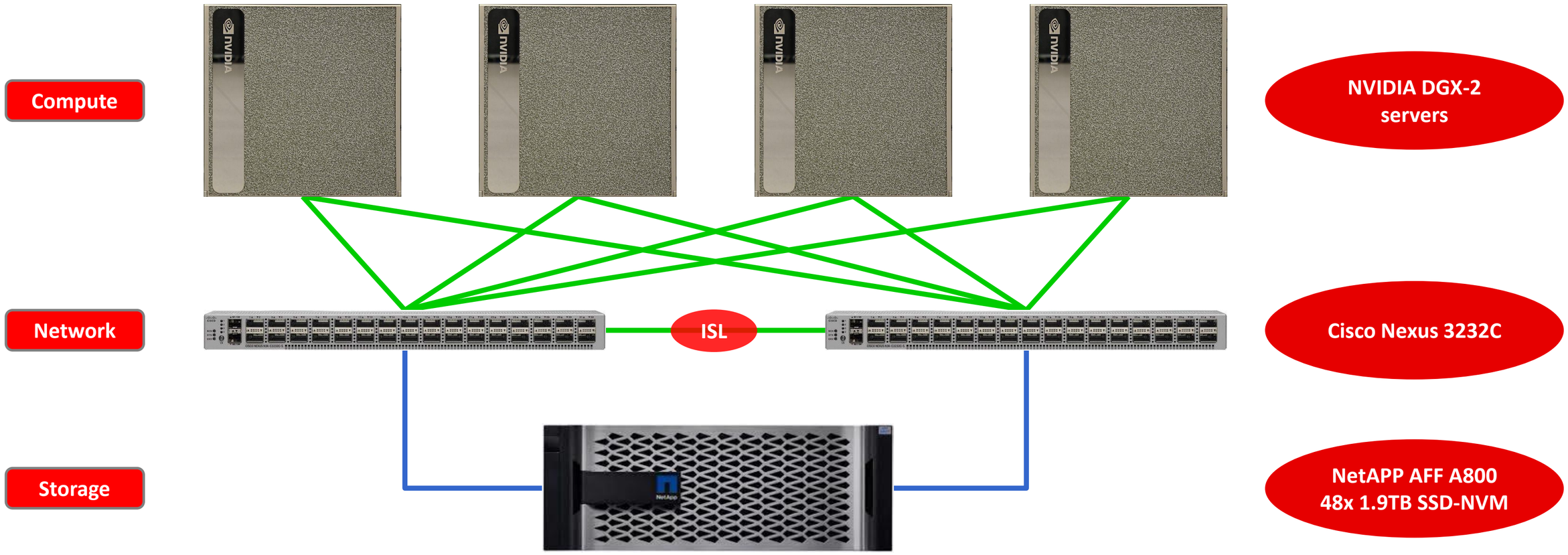




NetApp®

# 2. NVIDIA DGX reference architecture

We have delivered two projects with NVIDIA reference architectures (DGX-1 and DGX-2)




https://www.netapp.com/us/media/nva-1135-design.pdf

# 2. NVIDIA DGX reference architecture

Why do we choose NetApp for NVIDIA reference architecture?

There are several storage vendors for NVIDIA reference architectures – NetApp, DELL EMC, IBM, Pure and DDN.

Based on market research we selected the NetApp solution as the best option for projects:

- NetApp is represented by ALEF Distribution in CEE region
- fast responses and support from both ALEF and NetApp teams
- 100Gb storage connectivity available from the beginning (NetApp AFF A800)
- seamless integration of Windows apps requested by some clients
- flexible and fast during price negotiation

During project implementation we appreciated:

- the technical support of HPC/AI solutions and consultations during implementation
- great support from the NVIDIA team – tuning scripts, DeepOps tool

# 3. Tailored solutions with NVIDIA accelerators

Tuned IT architecture for specific requirements

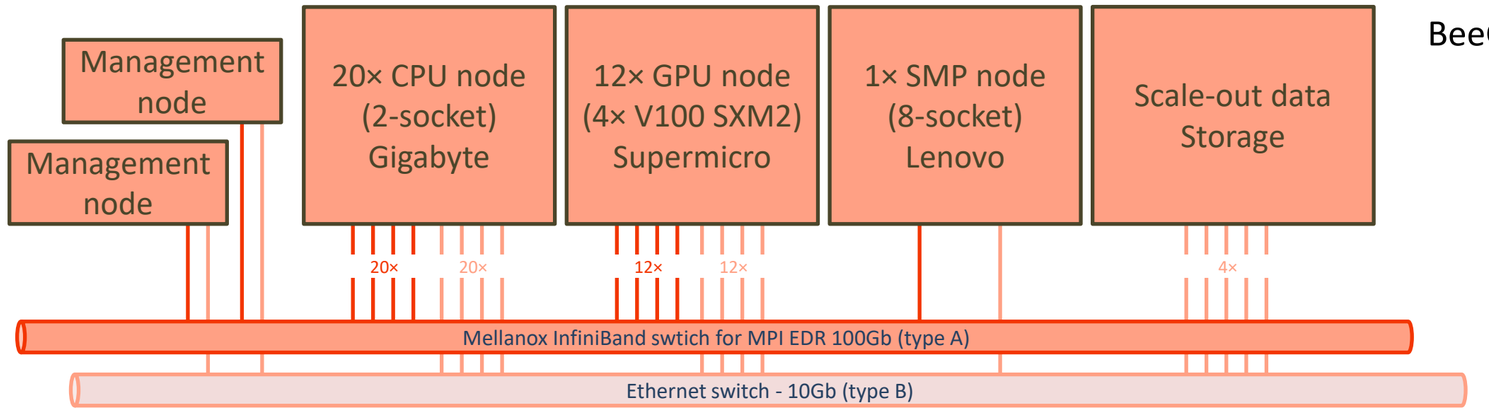
Flexibility of choice of hardware vendors, special platforms – e.g. Supermicro, Gigabyte

Combinations of different:

- CPU and GP GPU workloads
- different server architectures (cluster, SMP servers), form factors
- fabrics (Infiniband, Ethernet), bandwidth (1Gbps to 200Gbps)
- storage solutions (scratch, scale-out, object, SdS)
- management stack, workload scheduler, administration of container
- cooling (air, hot or cold liquid coolling)

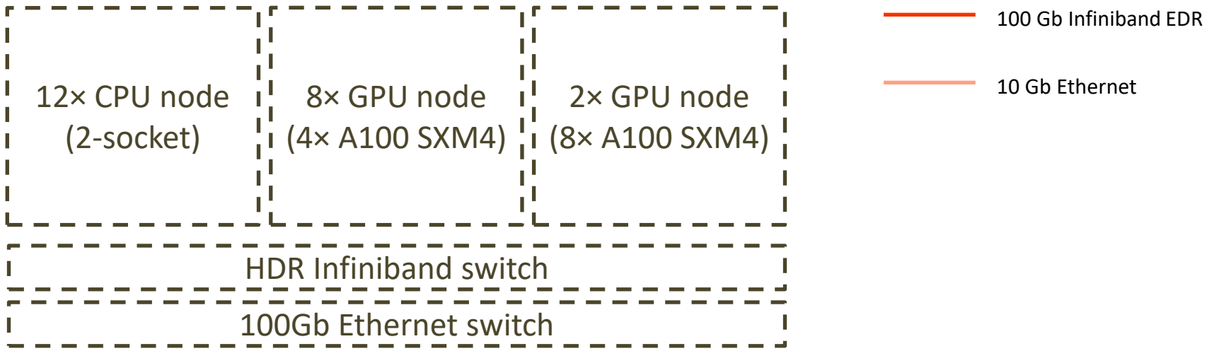
# 3. Tailored solutions with NVIDIA accelerators

## Architecture of RCI cluster at FEL, Czech Technical University in Prague




BeeGFS is used for scratch storage on top of local NVMe drives

Planned extension:







# 3. Tailored solutions with NVIDIA accelerators

Our references in the CEE region:

- RCI cluster at the **Faculty of Electrical Engineering, Czech Technical University**
  - development & testing of AI applications
- HPC cluster projects at the **Czech Academy of Sciences**
  - several projects for biochemistry, physics, data analysis, …
- HPC cluster projects at **Masaryk University** in Brno
  - CERIT SC projects for research in life sciences, AI, …
- AI supercomputer at **Tomas Bata University in Zlin**
- AI supercomputer at Czech startup **Quantasoft**
  - computer vision, biometrics, AI
- NVIDIA A100 accelerators for the **Slovak Academy of Sciences**
  - upgrade of systems for AI research

# Design and implement HPC and AI projects in 7 steps

## 1. First contact

- mcomputers.cz/en/nvidia
- obchod@mcomputers.cz
- **NVIDIA partner portal**

## 2. HPC/AI requirements

- discussions/consulting
- **NVIDIE ELITE partner**
- **NVIDIA tech guys**

## 3. Proof of concept (POC)

- **NVIDIA test drive**
- DGX station, A100, V100, T4, Jetson Nano, NX

## 4. Design

- **NVIDIA reference architecture**
- Tailored solutions

## 5. Implementation

- complete solution
- **NVIDIA software stack**
- **NVIDIA support**

## 6. Testing

- validation
- benchmarks
- **NVIDIA support**

## 7. Education

- **NVIDIA Deep Learning Institute (DLI)**
- Custom trainings

# References

NVIDIA Enterprise solutions micro-site: https://www.mcomputers.cz/en/nvidia/

NVIDIA reference architecture: https://www.nvidia.com/en-us/data-center/dgx-pod/

NVIDIA Deep Learning institutute: https://www.nvidia.com/en-us/deep-learning-ai/education/

NVIDIA NGC repository: https://ngc.nvidia.com/

DGX-2 at IT4Innovations: https://www.it4i.cz/infrastruktura/nvidia-dgx-2

CERIT Scientific Cloud: https://www.cerit-sc.cz/

RCI project at FEL, Czech Technical University in Prague: http://rci.cvut.cz/

RICAIP project at CIIRC, Czech Technical University in Prague: http://ricaip.eu/

Quantasoft, Czech AI start-up: https://www.quantasoft.com/

Tomas Bata University in Zlin:

https://fai.utb.cz/en/faculty/about-us/structure/departments/the-department-of informatics-and-artificial-intelligence/

NetApp ONTAP AI solution: https://www.netapp.com/us/products/ontap-ai.aspx

# Questions?

**Petr Plodik, M Computers**
Prague, Czech Republic
petr.plodik@mcomputers.cz
+420 737 264 480