CNAF and HPC

3d ML_INFN Hackathon

Nov. 20 2022

Stefano Dal Pra
dalpra@infn.it

Centro Nazionale Analisi Fotogrammi (CNAF) Yesterday

Centro Nazionale Analisi Fotogrammi (CNAF) Today

All digital

CNAF is the national center of INFN (Italian Institute for Nuclear Physics) dedicated to Research and Development on Information and Communication Technologies

https://cds.cern.ch/record/1541893

CNAF in a nutshell

- **WLCG Tier-1**: CNAF hosts a Worldwide LHC Computing Grid site.
- **Production:** It provides support to researchers in using the available computing tools and in software development.
- **R&D**: It investigates and develops innovative IT solutions aimed at improving the usability and the efficiency of the computing center and at enabling the use of geographically distributed systems.
- **Services:** It provides IT services of general utility for INFN
- **Partnership:** It collaborates with private companies and public administrations to share knowledge and expertise.
- We are about 60 people

CNAF Tomorrow: the Data Valley Hub

Supercomputing facilities of **ECMWF**, **CINECA** and **INFN**

- The Italian and Emilia-Romagna Region's largest investment in Big Data, Supercomputing and Research Infrastructure
- Hosting 80% of total computing capacity in Italy
- It will host important italian and international research institutes
- The move will begin in spring of 2023

HTC vs HPC

high-throughput

high-performance

HTC vs HPC

high-throughput

FLOPY

$$\left(\frac{\sum Completed\ Job\ Runtime}{Wall\ Time}\right)^{*}$$

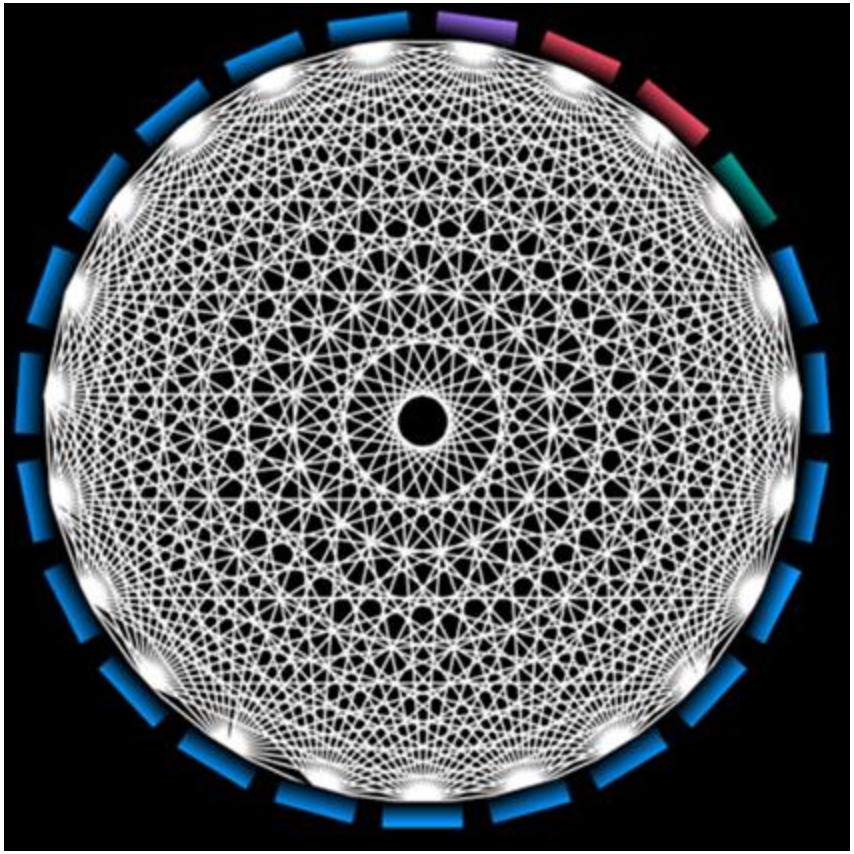
- Several indepenent "single node" programs
- Independent Computing nodes

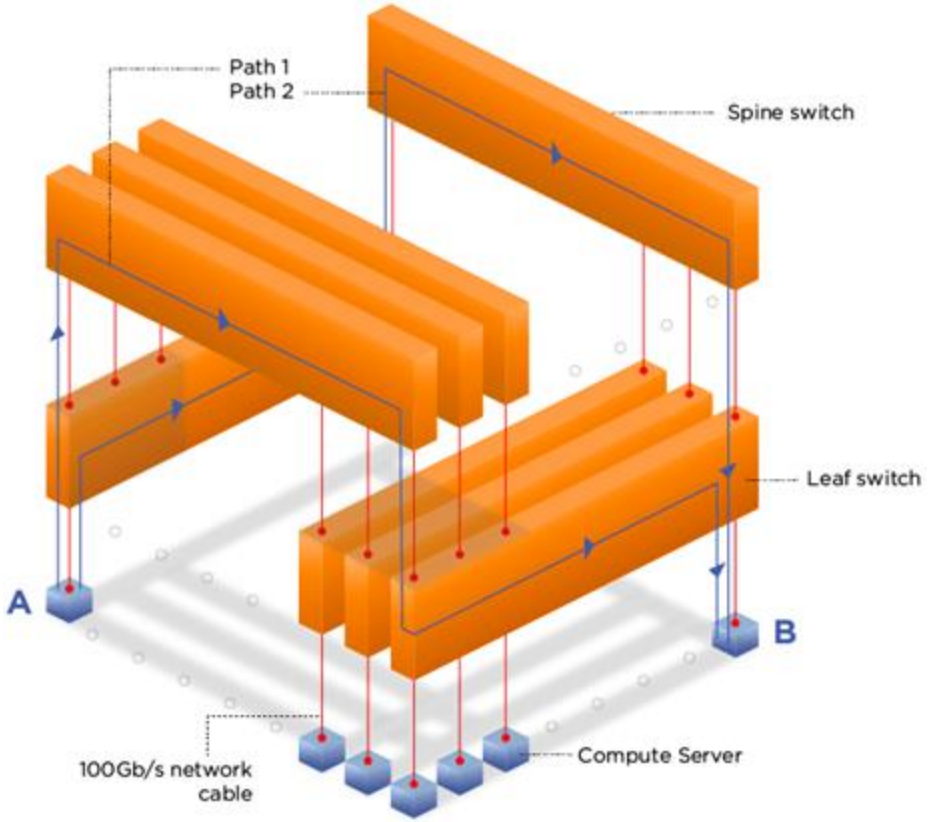
high-performance

FLOPS

- Huge multithread programs
- Span through several nodes
- The whole system has to be thought of as a single "computer" with an internal fabric joining it all up.

HPC and Interprocess comunication





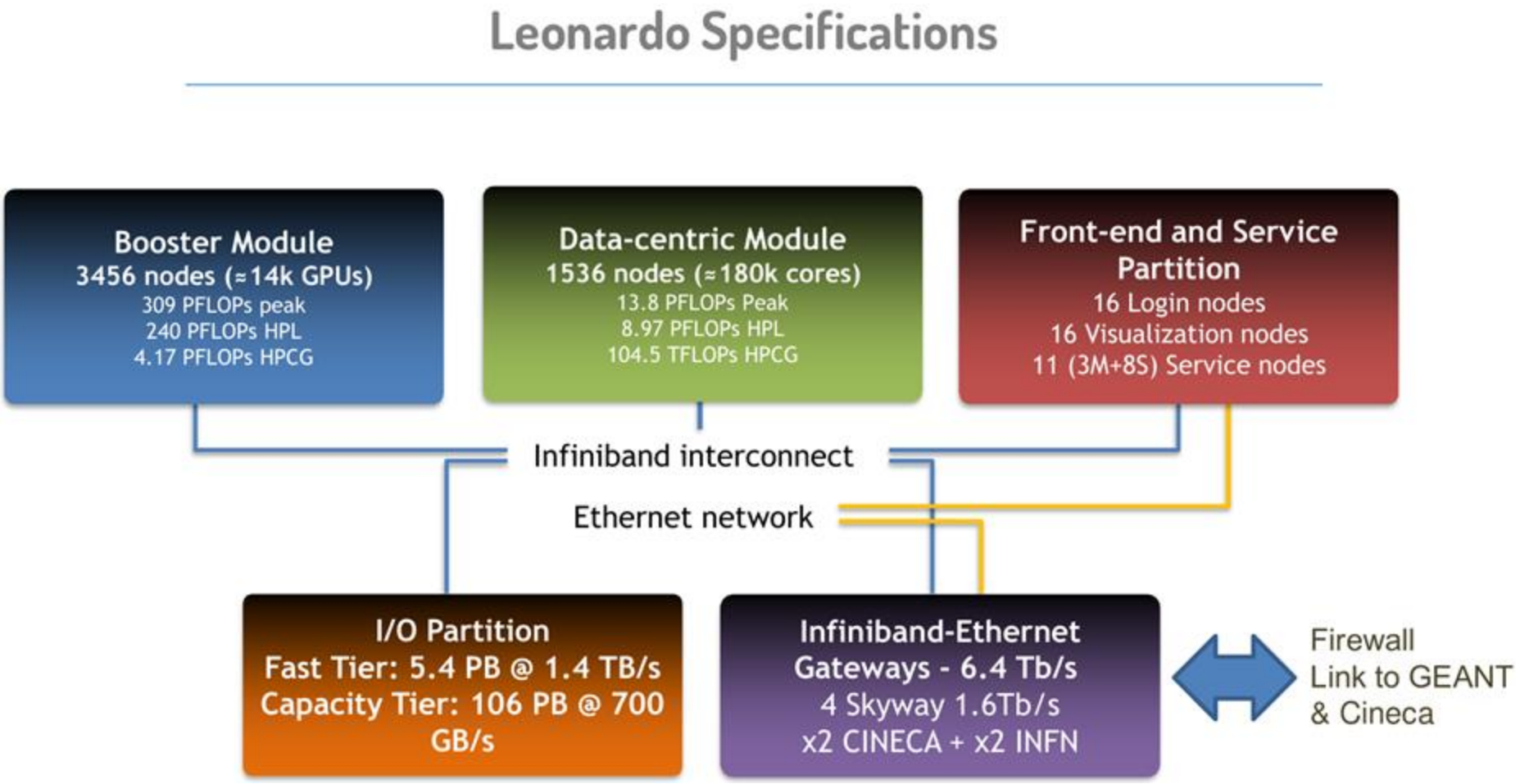
- Conceptually: direct connection
- Fabric: 2 layer Fat tree

CNAF and HPC

- Small HPC cluster, local submission
- 44 nodes, recently moved LSF –> Slurm
- GPFS shared filesystem
- A few nodes having V100 GPUs
- ML_INFN: Accelerated HW via CLOUD:
- GPU: (Tesla 10 x T4, 5 x RTX5000, A30, 2 x A100)
- FPGA: 2 x Xilinx Alveo U50, 1 x Xilinx Alveo U250
- Coming soon:
- FPGA: 5 more Alveo U50,
- GPU: 2 more A100

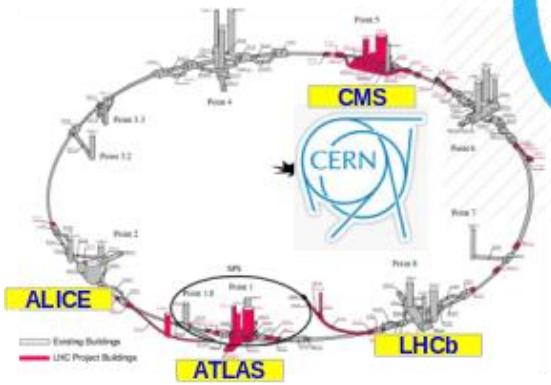
CINECA and HPC

CINECA and Leonardo



Evolution in HPC: GPUs and CPUs cohexisting

CNAF and HPC@CINECA



- **HPC center:** often has free resouces, **WLCG center:** always need more.
- Several works done and in progress to integrate HTC@CNAF with HPC@CINECA

HTC and HPC: integration challenges

A typical HPC machine:

- No outbound connectivity, no cvmfs (WLCG: relies on both)
- Local submission only (WLCG: submission to a CE)
- Whole node as minumum job size (WLCG: 1 or 8 core)
- 24h as lease time (WLCG: 48h or more)
- Access to registered identified users (WLCG: late binding, externally identified and authorized)
- Little RAM/CPU ratio (WLCG: 2GB/core, typical 3.5 or 4)
- Not only x86 arch, ppc also possible (Marconi-100) and GPUs

CNAF and HPC: Marconi A2

- connectivity: allowed outbound connections toward named CNAF and CERN; cvmfs: adopted by CINECA
- Submission: A HTCondor-CE at CINECA, managed by CNAF and routing jobs to the local Slurm batch system.
- Whole node: LHC users adapting their payloads
- lease time: agreement for extension to 48h
- Access: pool accounts (local dummy users running locally on behalf of remote Grid user)
- Storage: direct access from CNAF (dedicated 600Gb/s network channel), plus XROOTD cache at CINECA for remote access
- Cfr. CHEP2019: Extension of the INFN Tier-1 on a HPCsystem

CNAF and HPC: Marconi-100

- PowerPC architecture. Need suitable software
- 4 x V100 GPUs per node.
- Opportunistic model: use the when available

Sketch of solution:

- Slurm submission to a "CNAF" queue.
  - At start, the job run a Singularity container, Who run as a HTCondor Compute Node
  - It has credentials to register with CNAF and join its HTCondor pool.
  - It can run payloads submitted to the CNAF CEs
- Cfr. ACAT 2022: Transparent extension of INFN-T1 with heterogeneous computing architectures