Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

ICSC
Centro Nazionale di Ricerca in HPC,
Big Data and Quantum Computing

# Moving a data center keeping availability at the top

**D.Lattanzio***, A.Pascolini, A.Chierici, D.Michelotto, D.Cesini, G.Sergi

**Conference on Computing in High Energy and Nuclear Physics. October 19-25, 2024 - Kraków**

# Outline



- Introduction
- What we moved
- How we moved it
- Lesson learned

# *Introduction*

# *Introduction*

- The focus of this presentation is on computing (farming) and cloud infrastructures

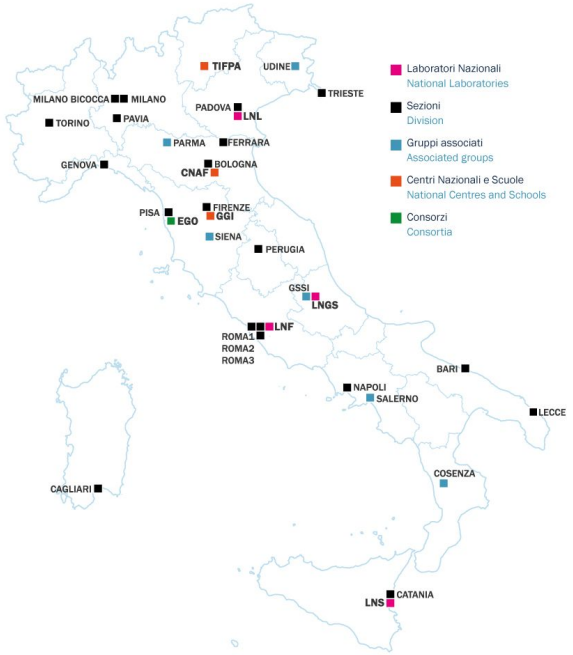
- We will describe how we were able to move all the resources without interrupting the service provided to users

# CNAF - Centro Nazionale Analisi Fotogrammi

INFN CNAF

National Centre of INFN (Institute for Nuclear Physics) for the information and communication technologies

- Supporting researchers in using available computing resources
- Hosting the **italian Tier-1** data center for the **WLCG** collaboration
- Representing a key computing facility for many experiments, not only for LHC and the physics field

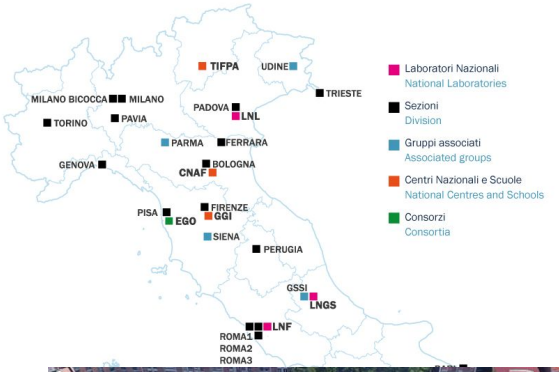


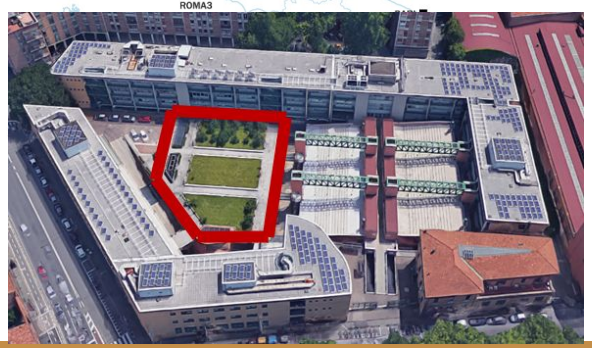
# CNAF - Centro Nazionale Analisi Fotogrammi

INFN CNAF

National Centre of INFN (Institute for Nuclear Physics) for the information and communication technologies

- Supporting researchers in using available computing resources
- Hosting the **italian Tier-1** data center for the **WLCG** collaboration
- Representing a key computing facility for many experiments, not only for LHC and the physics field



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

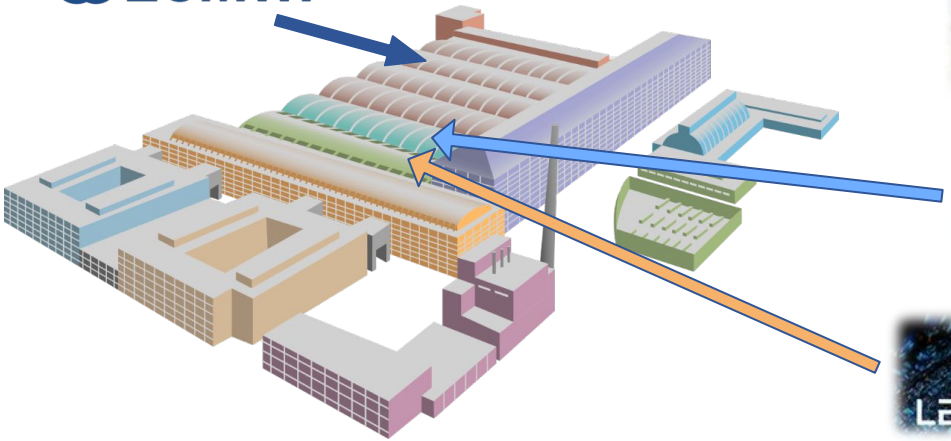

# The new INFN Data Center at Bologna Tecnopolo

# *INFN-T1 new data center*

INFN CNAF

- INFN-T1 has a new building, part of the «Bologna Technopole»

ECMWF



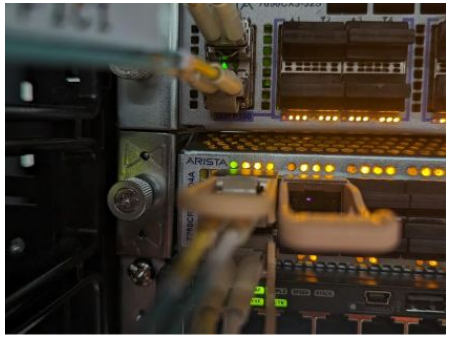
INFN CNAF **Our new DC building**

LEONARDO

BOLOGNA TECHNOPOLE the factory of the FUTURE

Switch on of the first WN rack in production at Technopole 19/03/24

INFN CNAF



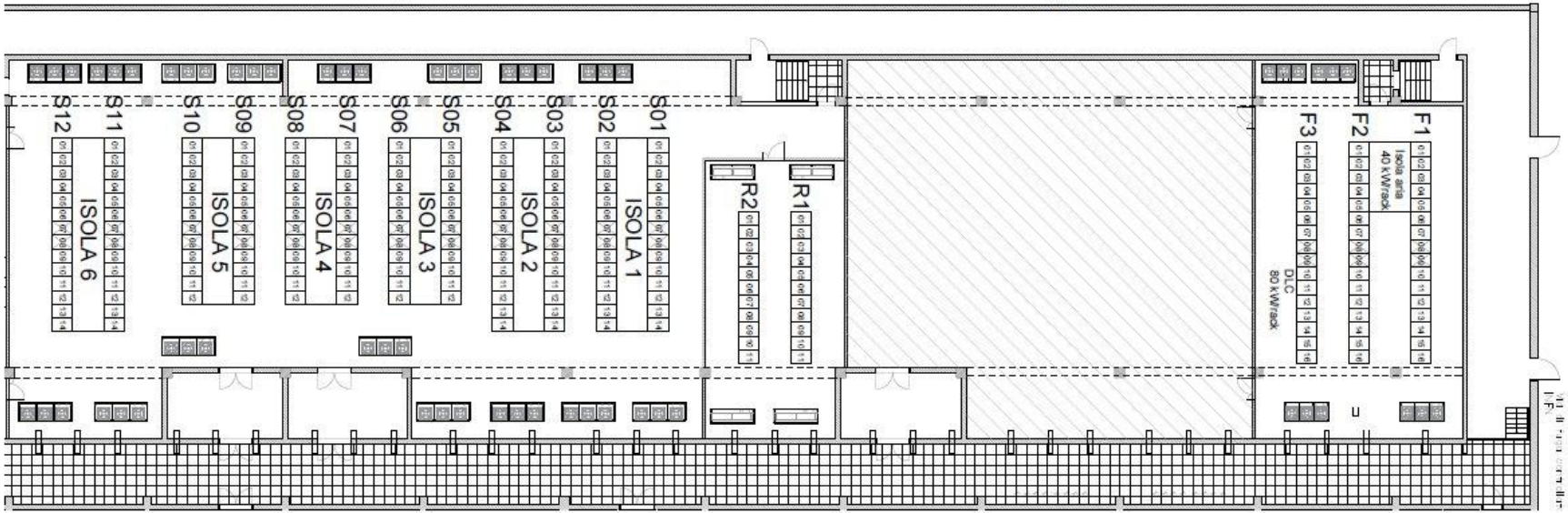


May 10th, 2024

# Layout of the new location

INFN CNAF

Storage | Computing and cloud | Network | Expansion | DLC zone



# *What we moved*

# *Computing situation before the transfer*

INFN CNAF

- Resources provided by our DC: 662k HS06
  - 112k HS06 are provided by old systems to be decommissioned (5 racks)
  - 139k HS06 are provided by rather new hardware (3 racks)
  - 411k HS06 are provided by nodes hosted at CINECA (7 racks)
  - 60k HS06, 35 GPU, 10 FPGA are provided by hypervisors nodes (3 racks)

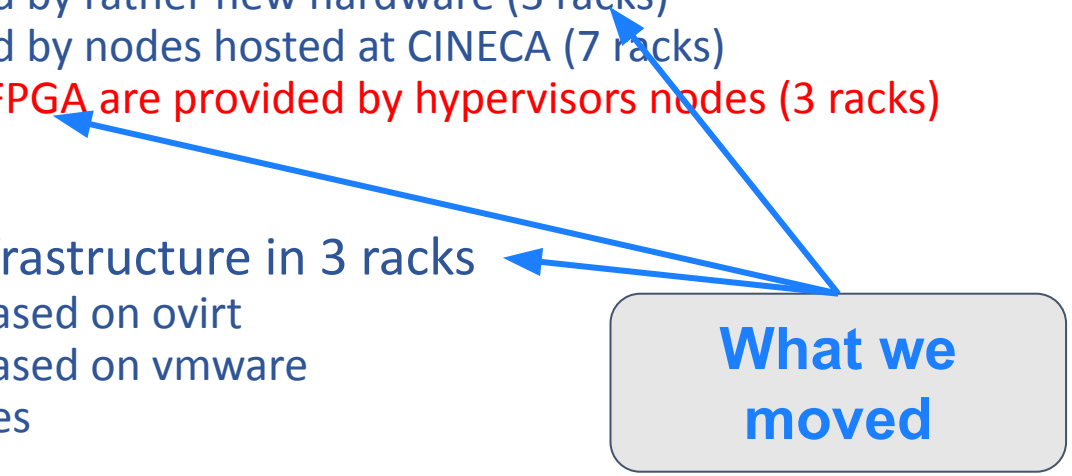
- We host our services infrastructure in 3 racks
  - Virtualization cluster based on ovirt
  - Virtualization cluster based on vmware
  - Some stand-alone nodes

# *Computing situation before the transfer*

INFN CNAF

- Resources provided by our DC: 662k HS06
  - 112k HS06 are provided by old systems to be decommissioned (5 racks)
  - 139k HS06 are provided by rather new hardware (3 racks)
  - 411k HS06 are provided by nodes hosted at CINECA (7 racks)
  - 60k HS06, 35 GPU, 10 FPGA are provided by hypervisors nodes (3 racks)

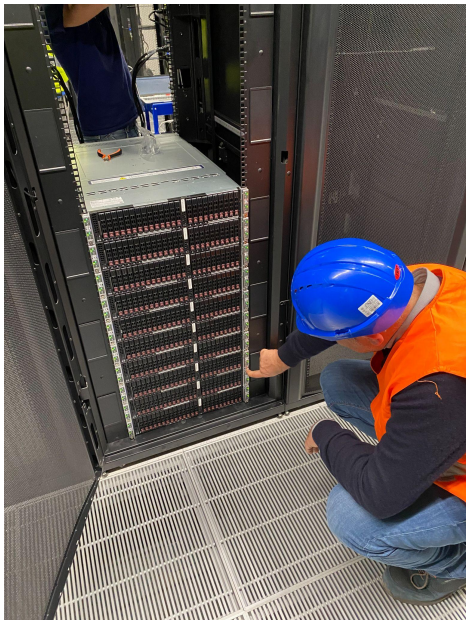
- We host our services infrastructure in 3 racks
  - Virtualization cluster based on ovirt
  - Virtualization cluster based on vmware
  - Some stand-alone nodes



# *How we moved it*

# Moving company

- We selected the moving company with a public tender
- We agreed to organize the move on a specific number of days
  - we could not switch off and move everything at once like people do with furniture




# Redundancy in first place

INFN CNAF

- It may sound obvious, but to avoid downtime it's fundamental to implement redundant services
  - HTCondor
  - Spread VMs across different virtualization systems
  - Rely on highly available hardware
  - Openstack, controller node and network node

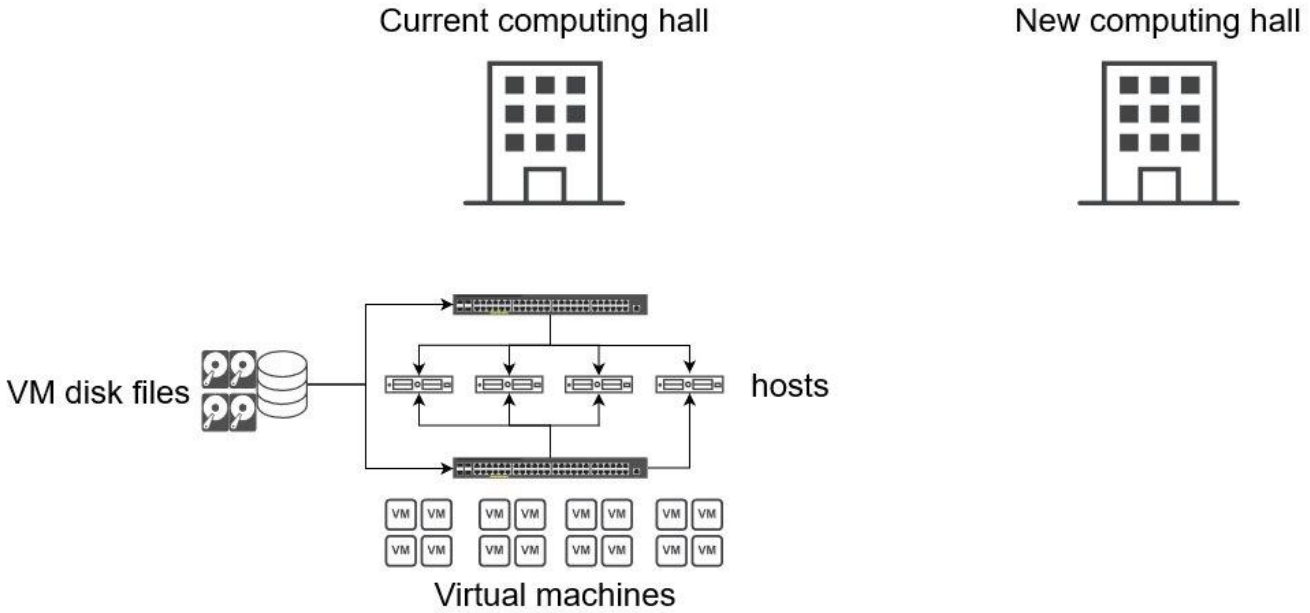
- This situation was already in place to keep site availability at the top on a daily basis and proved to be effective also during this «challenge»

# Moving virtualization systems

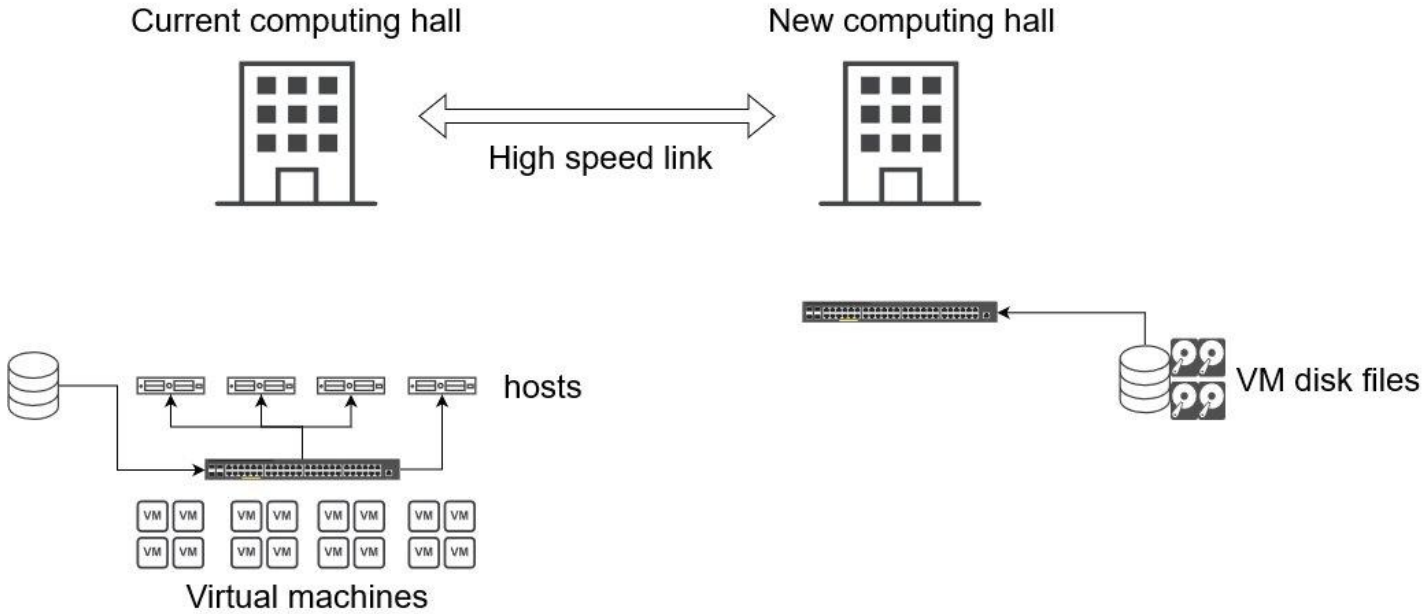
INFN CNAF

- Our infrastructures rely on an iscsi storage and several hosts
- Mandatory to have an extra storage
- How to implement the move
  - Move one of the redundant switches to the new computing hall
  - Install and make available the additional iscsi storage on the new computing hall
  - Live migration of disk image of each VM to the new storage
  - Move half of the hosts to the new computing hall
  - Live migrate all the VMs on the hosts in the new location
  - Move the rest of the hosts and switch to the new computing hall

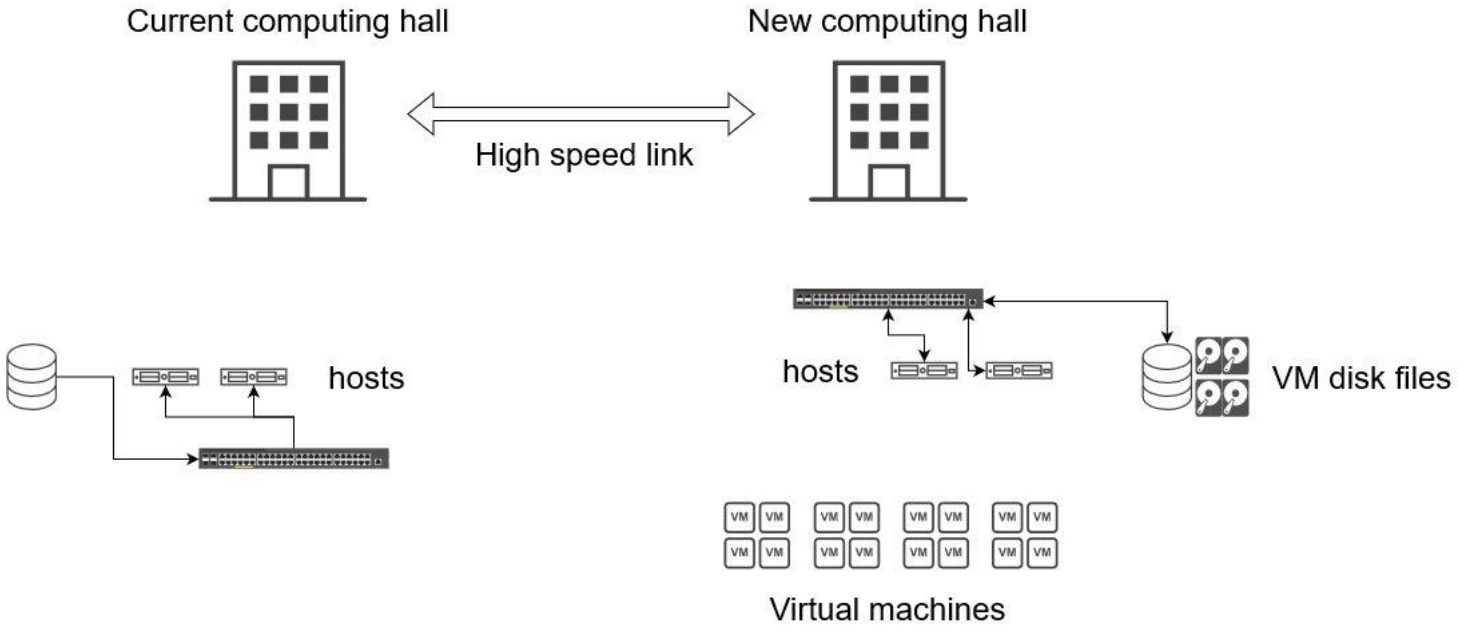
# Moving virtualization systems: initial situation



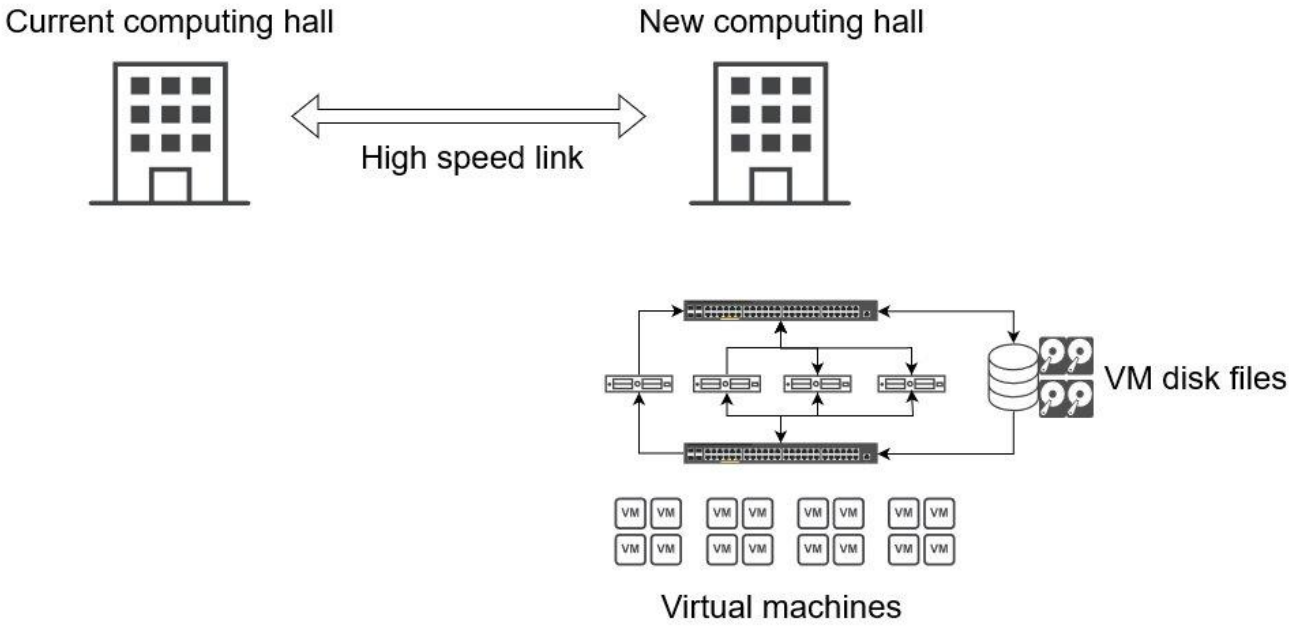
# Moving virtualization systems: step 1



# Moving virtualization systems: step 2



# Moving virtualization systems: step 3



# Moving computing resources

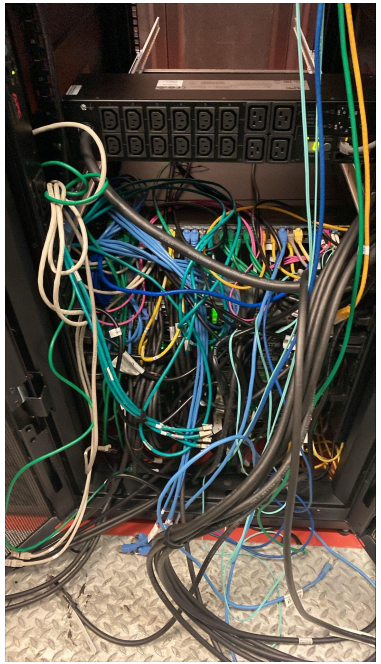
INFN CNAF

- As described in a previous slide, we just moved recent computing resources
  - Old resources will be decommissioned at the end of 2024
  - Resources hosted at CINECA DC will continue to run flawlessly
- Moving computing resources is straightforward, since the operation requires downtime
  - We moved one rack at a time, to reduce impact on HS06 provided by the center
- Moving the resources gave us the opportunity to solve the «spaghetti cabling» problem
  - Racks at new hall are wider
  - We labelled all the cables
  - We populated correctly our inventory system (based on OpenDCIM)

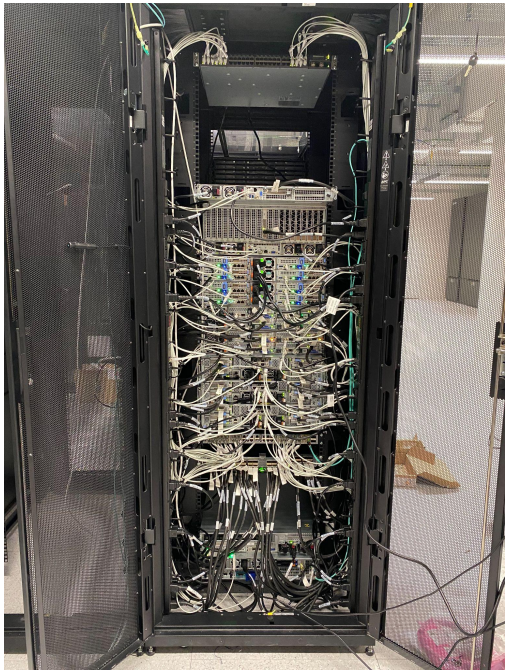
# HPC cluster RACK

INFN CNAF

**Before**



**After**



# Moving cloud resources

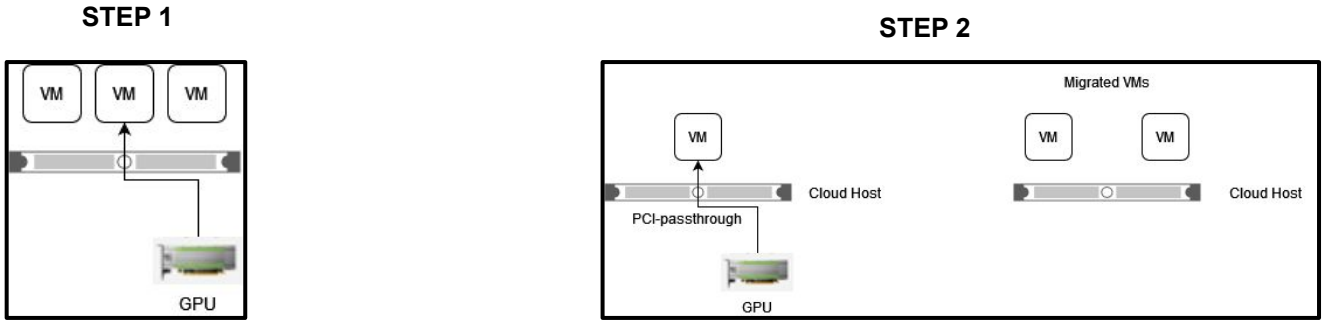
INFN CNAF

- Computing
  - VMs on cloud were moved in the same way as virtualization systems
  - Hypervisors with GPU and FPGA were moved shutting down the VMs using the directly attached hardware (30 of 700 VMs)

# Moving cloud resources

INFN CNAF

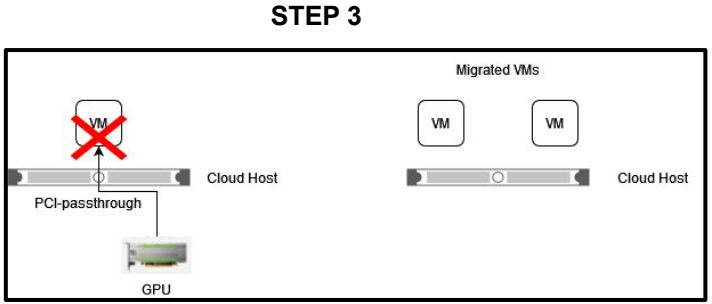
- Computing
  - VMs on cloud were moved in the same way as virtualization systems
  - Hypervisors with GPU and FPGA were moved shutting down the VMs using the directly attached hardware (30 of 700 VMs)

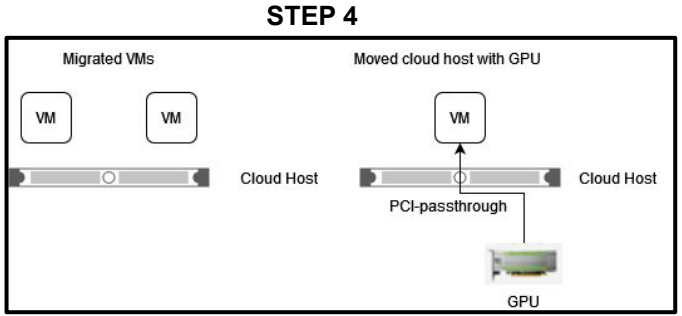


# Moving cloud resources

INFN CNAF

- Computing
  - VMs on cloud were moved in the same way as virtualization systems
  - Hypervisors with GPU and FPGA were moved shutting down the VMs using the directly attached hardware (30 of 700 VMs)



# Moving cloud resources

INFN CNAF

- Storage: Ceph cluster of 12 nodes with redundant network configuration
  - 1 out of 2 switches was moved to the new datacenter
  - 4 new nodes were installed on the new datacenter and joined to the cluster
  - 1 node at time was drained, moved to new datacenter and rejoined to the cluster (8 times)
  - the second switch was moved to the new datacenter
  - the redundant network connections of all 12 nodes were reestablished

# *Lesson learned*

INFN CNAF

- Moving a computing infrastructure is challenging
- We planned the activity with great attention and this prevented errors
- Network is the base, common layer that needs to be set-up in first place
  - Strong collaboration with network group is crucial
- To avoid downtime it's necessary to implement redundant services
  - but this is the base of a Tier-1 center like ours
- Selecting a moving company is not trivial


Thanks for the attention


ICSC
Centro Nazionale di Ricerca in HPC
Big Data and Quantum Computing
Supercomputing
shaping the future