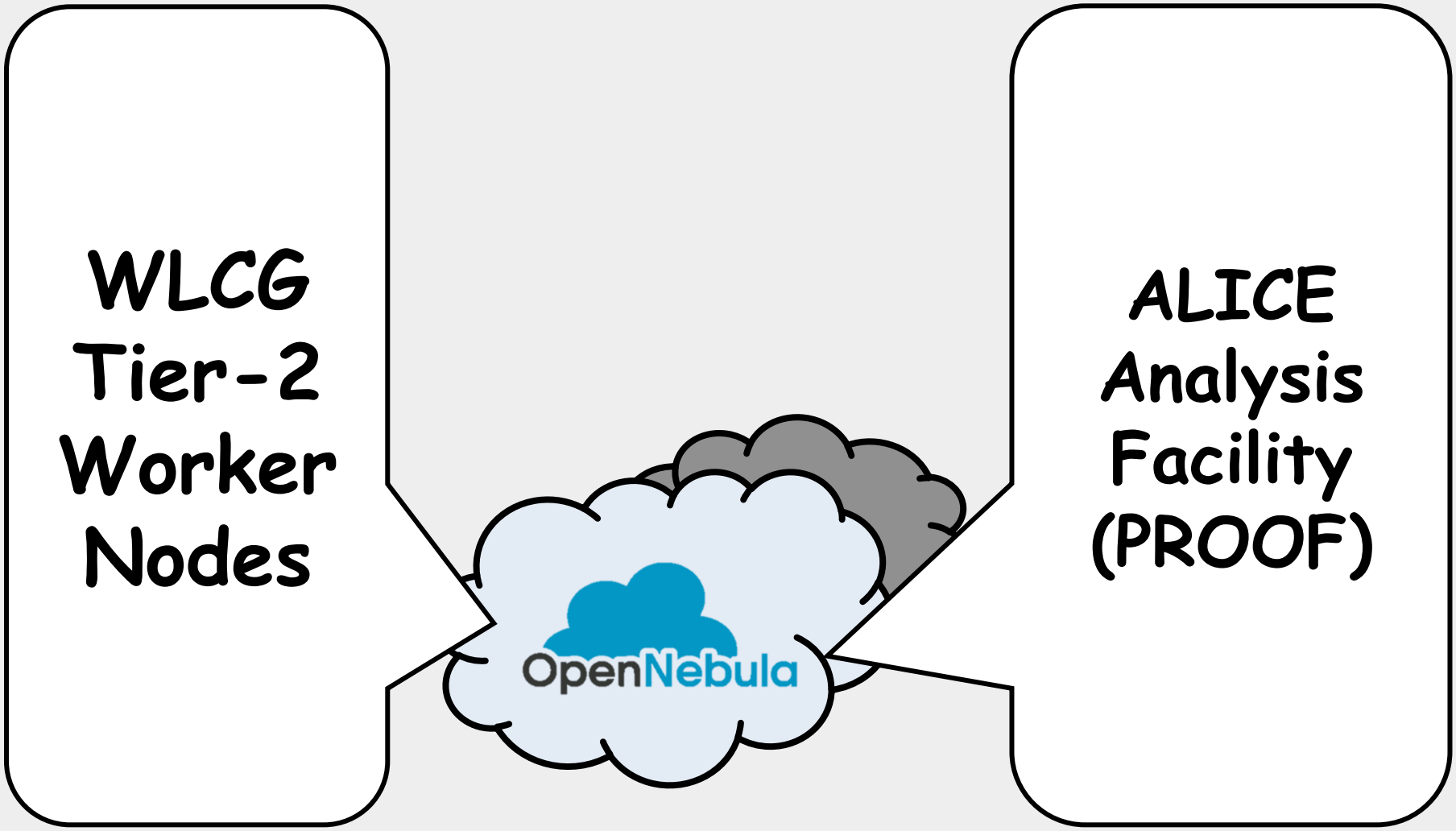
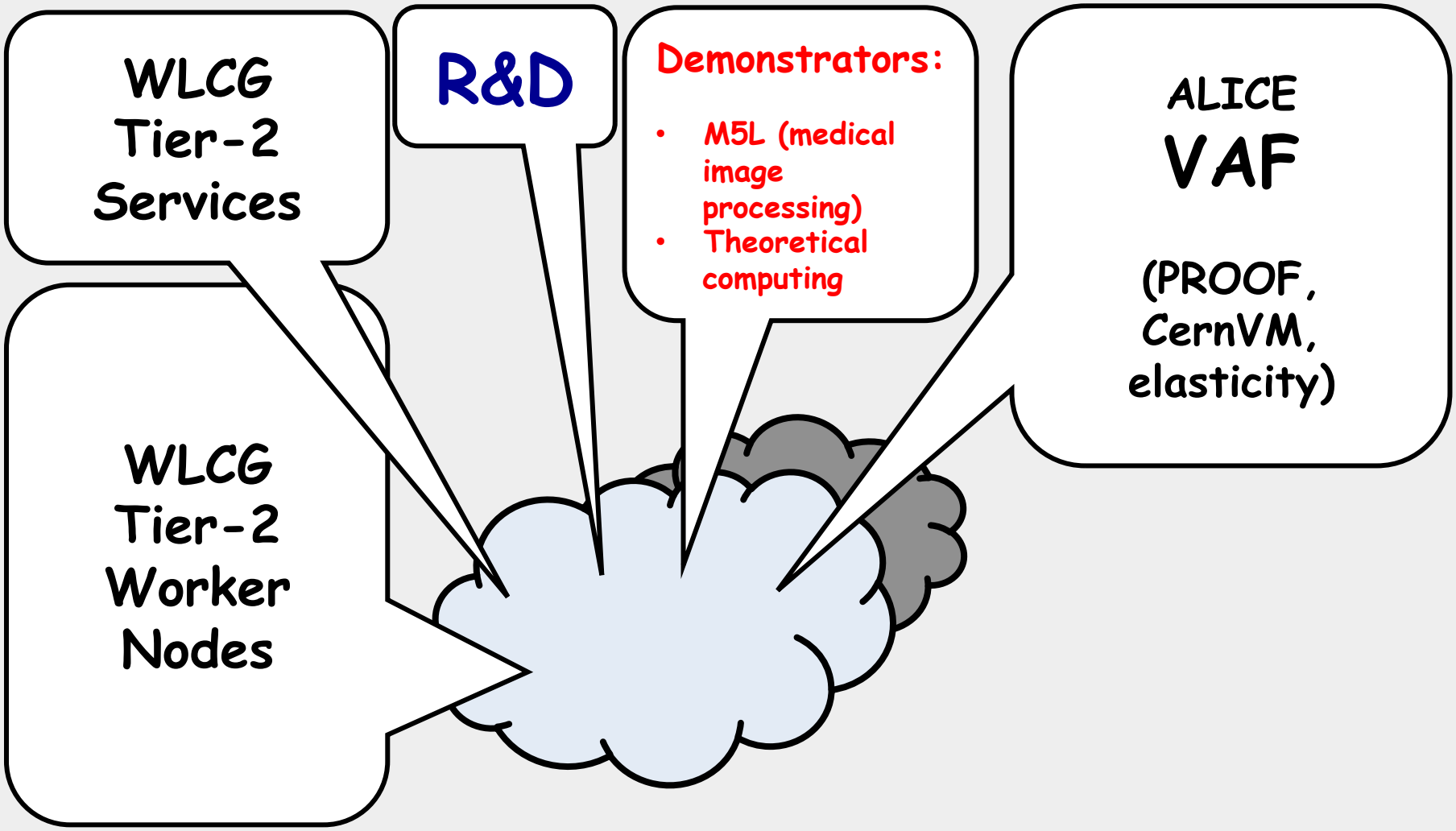
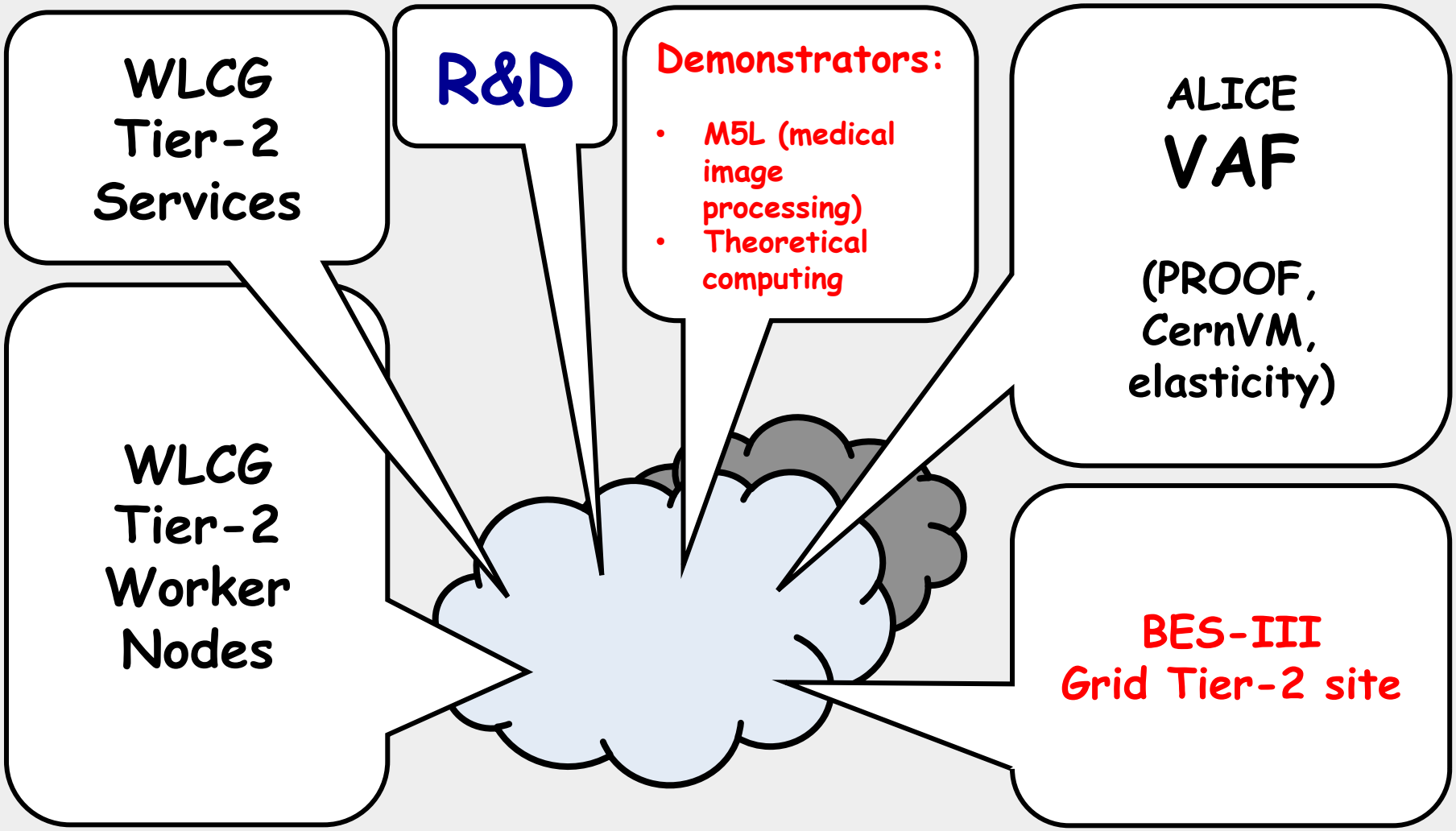
CLOUD COMPUTING ACTIVITIES IN TORINO

Workshop Commissione Calcolo e Reti | Napoli, Dec 16, 2014

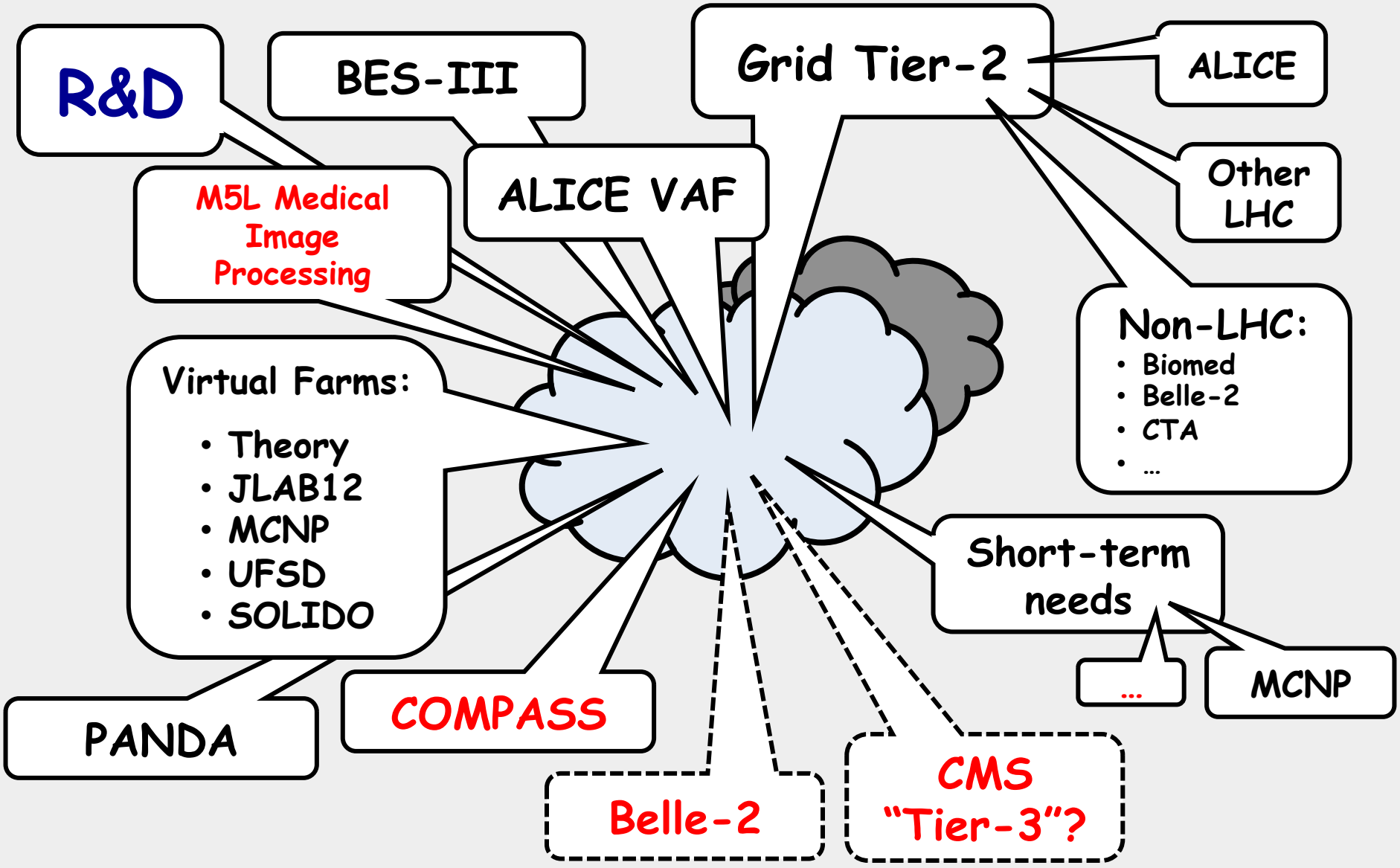
IN THE BEGINNING...




WLCG Tier-2 Services
R&D
Demonstrators:
• M5L (medical image processing)
• Theoretical computing
ALICE VAF
(PROOF, CernVM, elasticity)
WLCG Tier-2 Worker Nodes


WLCG Tier-2 Services
R&D
Demonstrators:
• M5L (medical image processing)
• Theoretical computing
ALICE VAF
(PROOF, CernVM, elasticity)
WLCG Tier-2 Worker Nodes
BES-III Grid Tier-2 site

...NOW!


R&D
BES-III
Grid Tier-2
ALICE
M5L Medical Image Processing
ALICE VAF
Other LHC
Virtual Farms:
• Theory
• JLAB12
• MCNP
• UFSD
• SOLIDO
Non-LHC:
• Biomed
• Belle-2
• CTA
• ...
Short-term needs
...
MCNP
PANDA
COMPASS
Belle-2
CMS "Tier-3"?

A TALE OF TWO CLUSTERS

VMs providing **critical services**:

- in- & out-bound connectivity
- public & private IP
- live migration
- no special I/O requirements

VMs providing **computing workforce**:

- example: Grid WNs
- private IP only
- high storage I/O performance

CC BY NC ND

A TALE OF TWO CLUSTERS

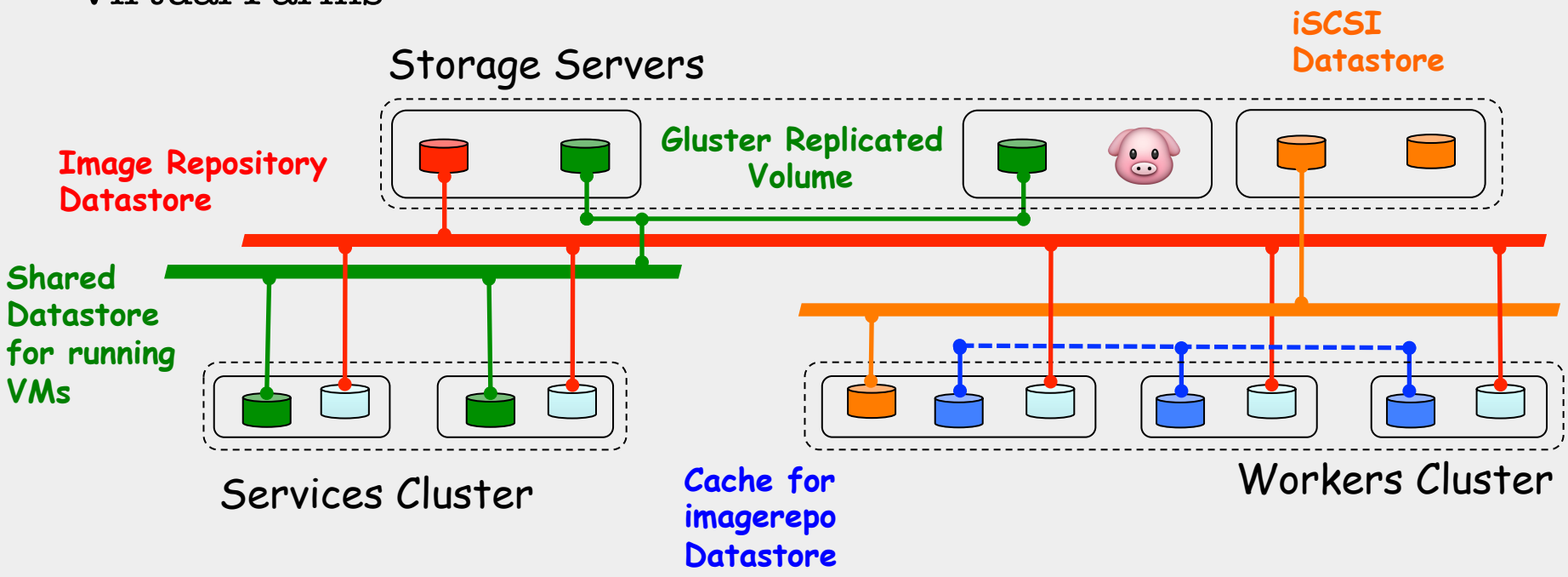
- Server-class hardware
- Shared image repository
- Resiliency-optimized FS for shared system disks
- Live migration possible

- Working-class hardware ☺
- Cached image repository
- Access to perfomance-optimized FS for data needs
- Runs about 200 VMs

CC BY NC ND

MULTIPURPOSE STORAGE: GLUSTERFS

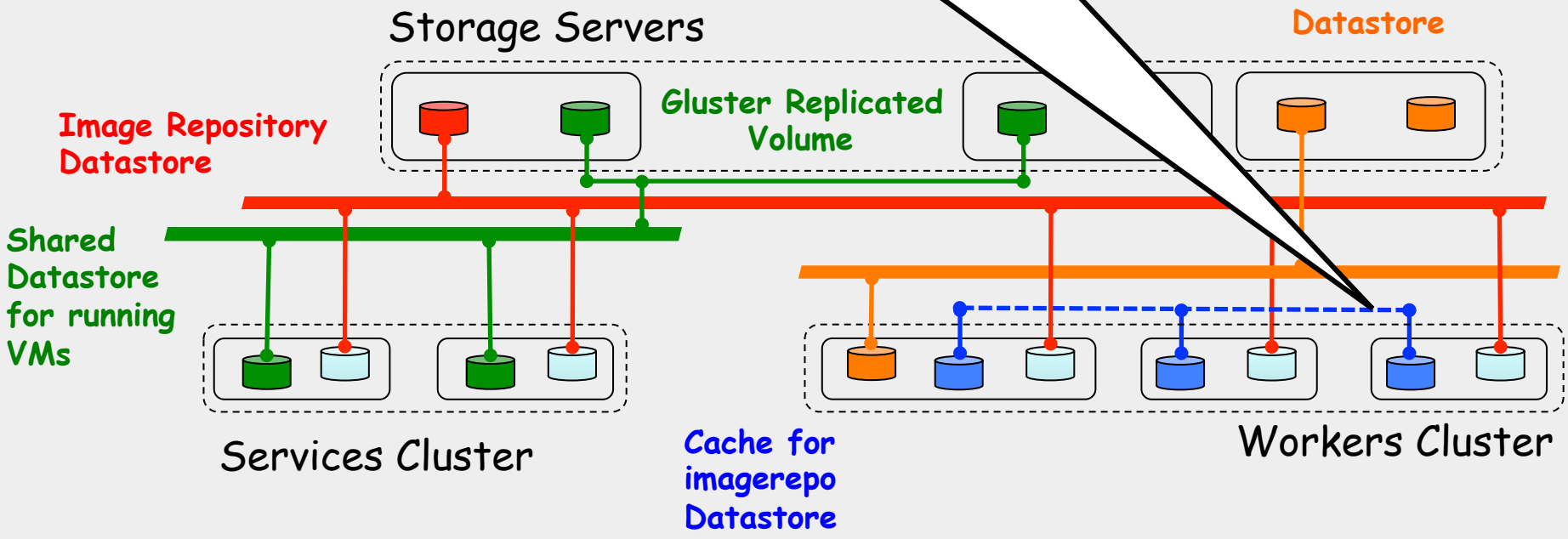
- Services System Datastore is **shared** to allow live migration
- Workers System Datastore is **local** to the hypervisors to increase I/O capacity. Images repository is locally **cached** on each hypervisor to reduce startup time
- Persistent Space Datastore is mounted on the relevant hypervisors using the **iSCSI** Transfer Manager to provide persistent storage to Virtual Farms



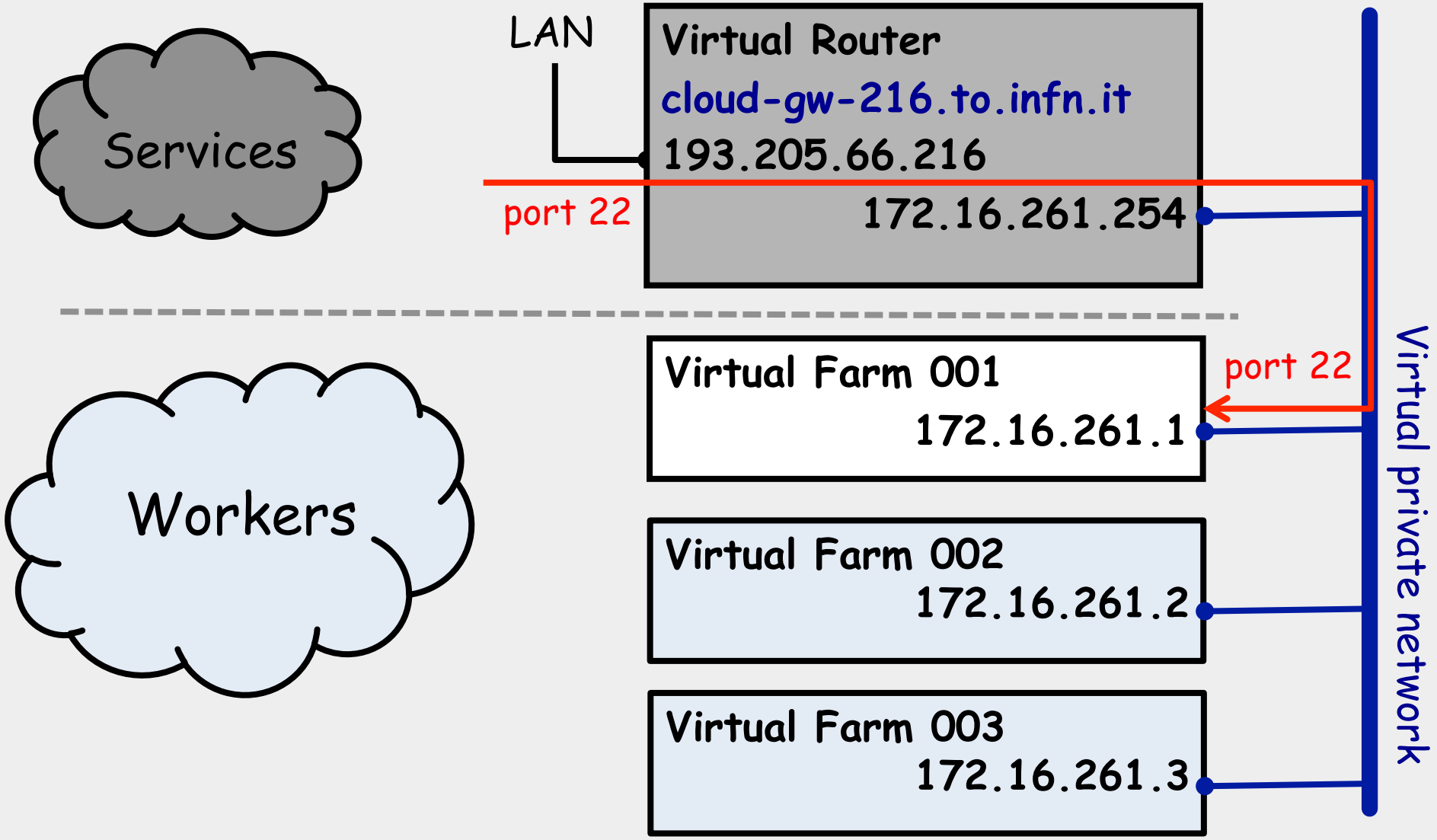
MULTIPURPOSE STORAGE: GLUSTERFS

- Services System Datastore is **shared** to allow live migration
- Workers System Datastore is **local** to the hypervisors to increase I/O capacity. Images repo … each hypervisor to reduce startup time
- Persistent Space Datas … vant hypervisors using the **iSCSI** Transf … istent storage to Virtual Farms

An ad-hoc script synchronizes the local copies using a custom "torrent-like" tool (scpWave + rsync) when new versions of the images are saved



BASIC VIRTUAL FARM PROVISIONING MODEL

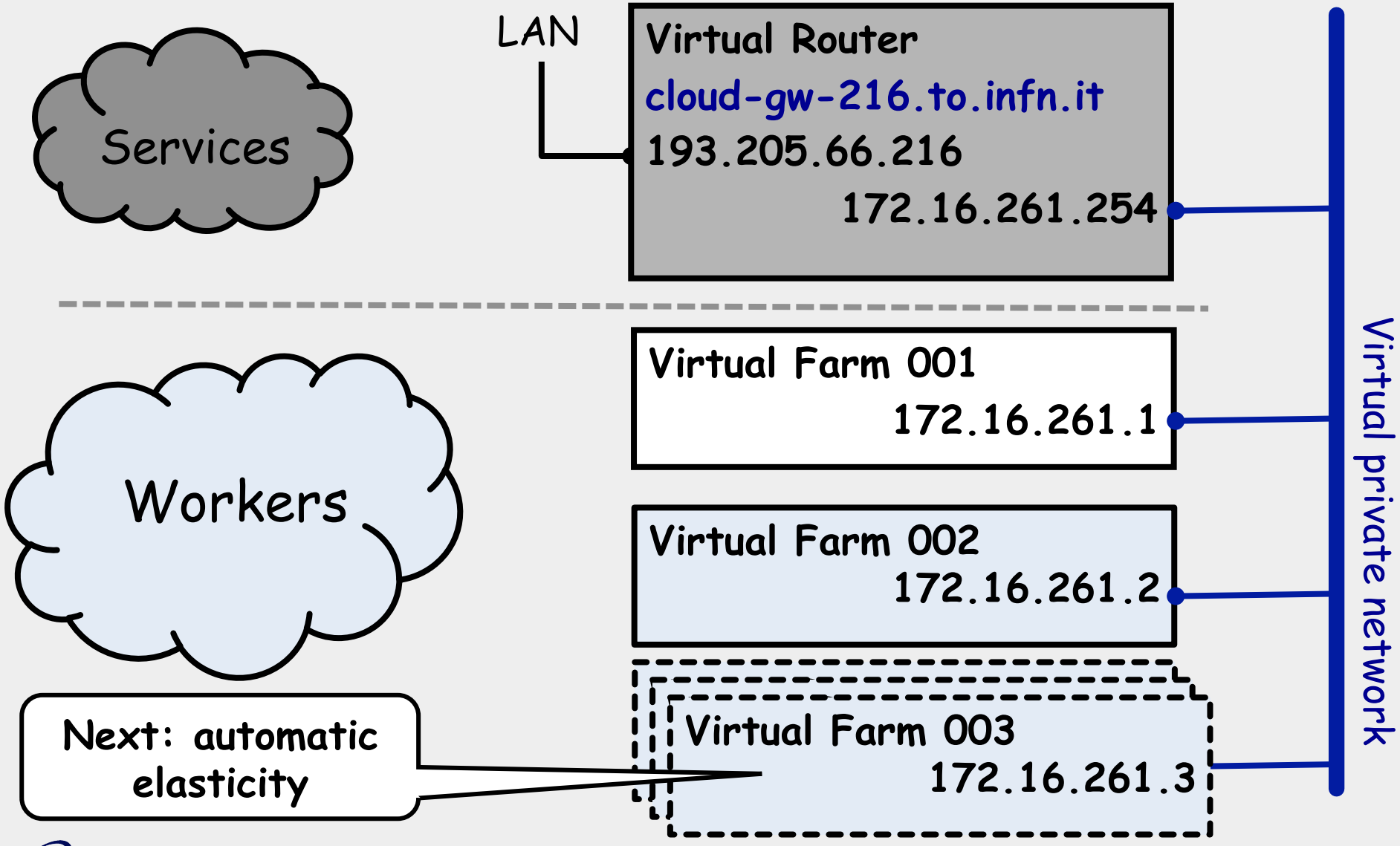


VIRTUAL FARM TOOLKIT

- ONeUser & quotas (CPU, memory,...)
- VRouter (OpenWRT small VM)
- OS Images (CentOS, Ubuntu,...)
- Private network (e.g. 172.16.216.0/24)
  - With DHCP, DNS, NAT,...
- "Elastic" public IP (e.g. 193.205.66.216)
- Volatile & persistent storage
  - iSCSI datastore
- Context file templates (Cloud-config)
  - E.g. HTCondor head node & workers
- EC-2 CLI access and Sunstone dashboard

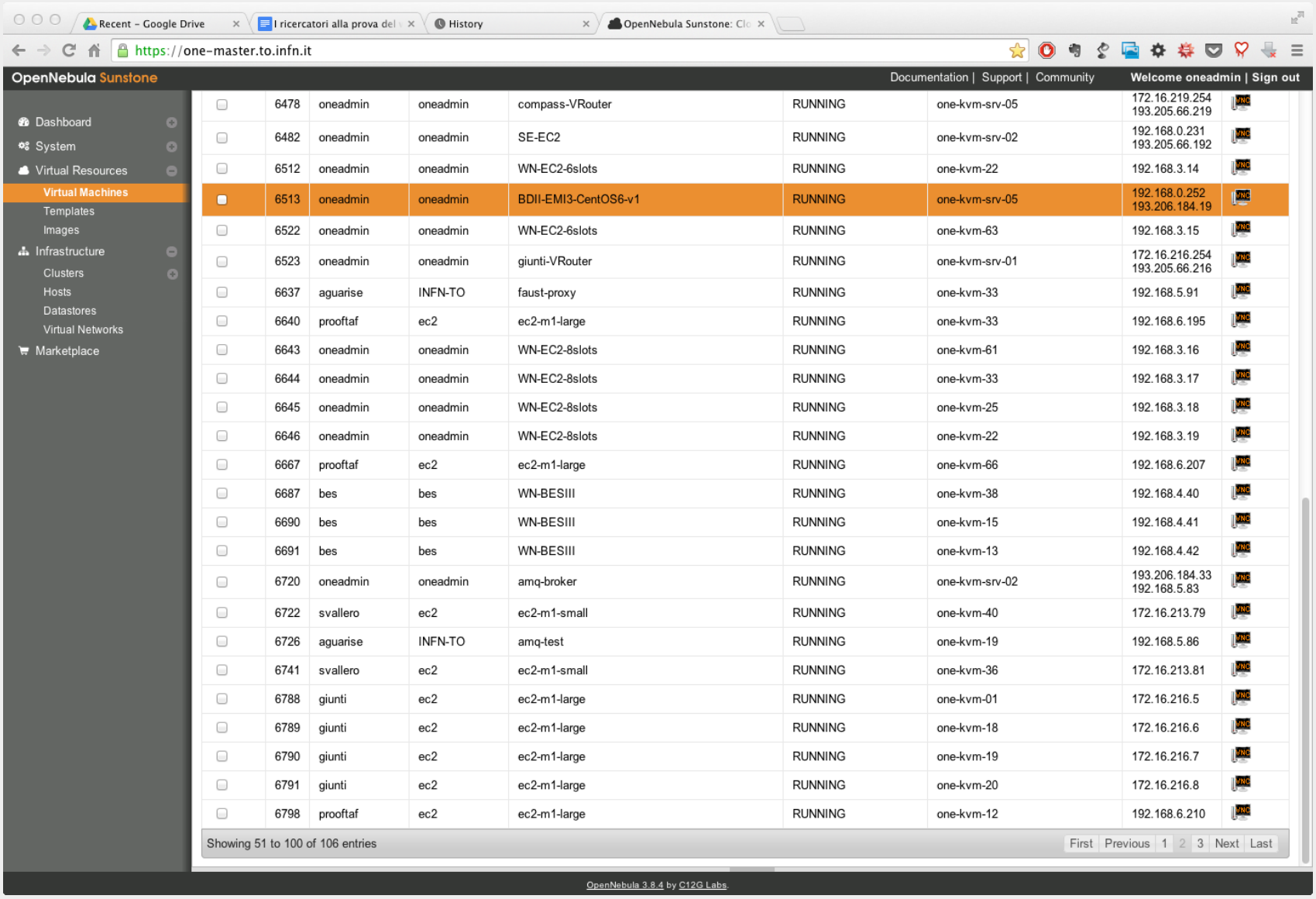
CC BY NC ND

VIRTUAL FARM PROVISIONING MODEL

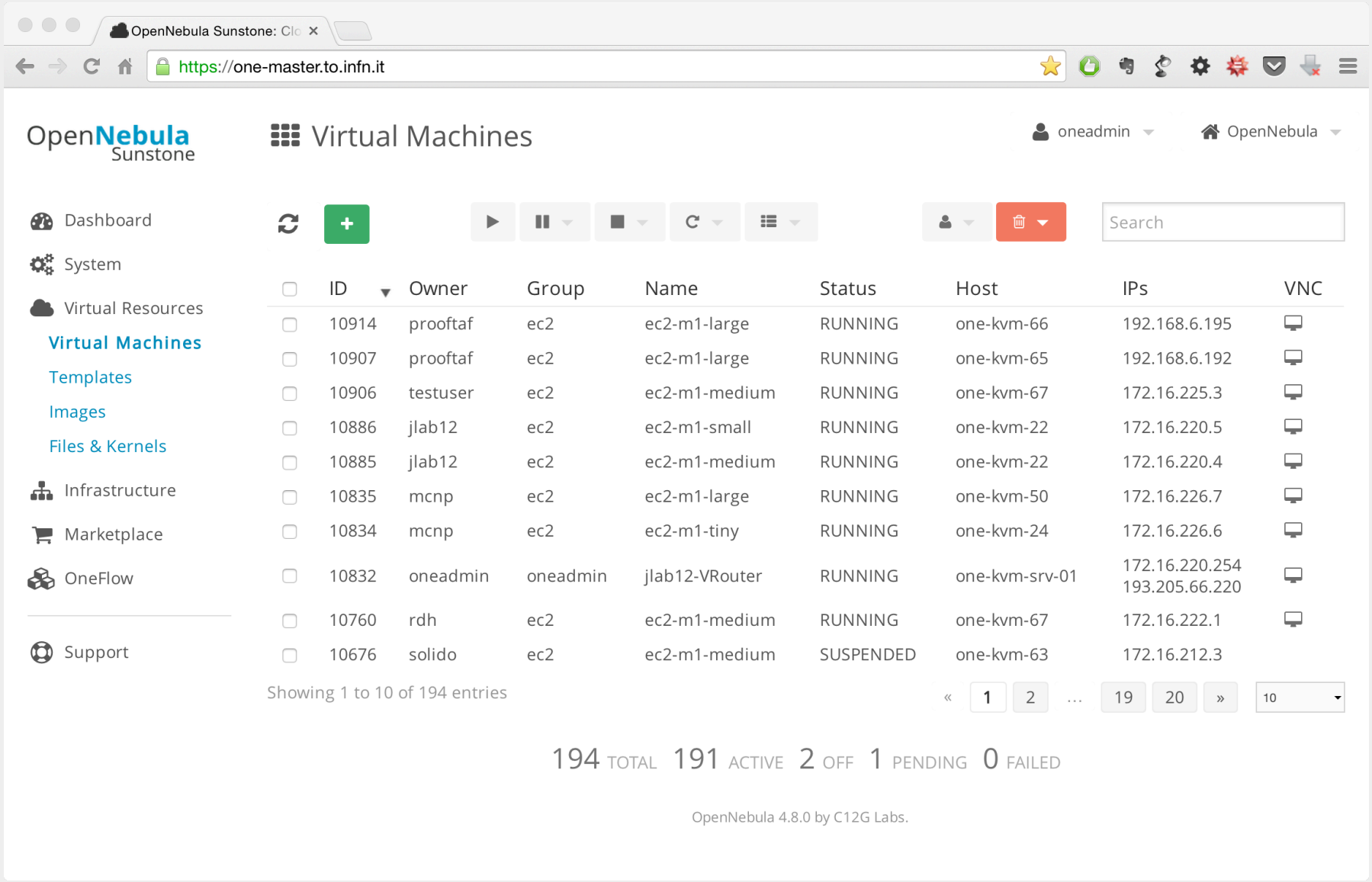


TENANT DASHBOARD

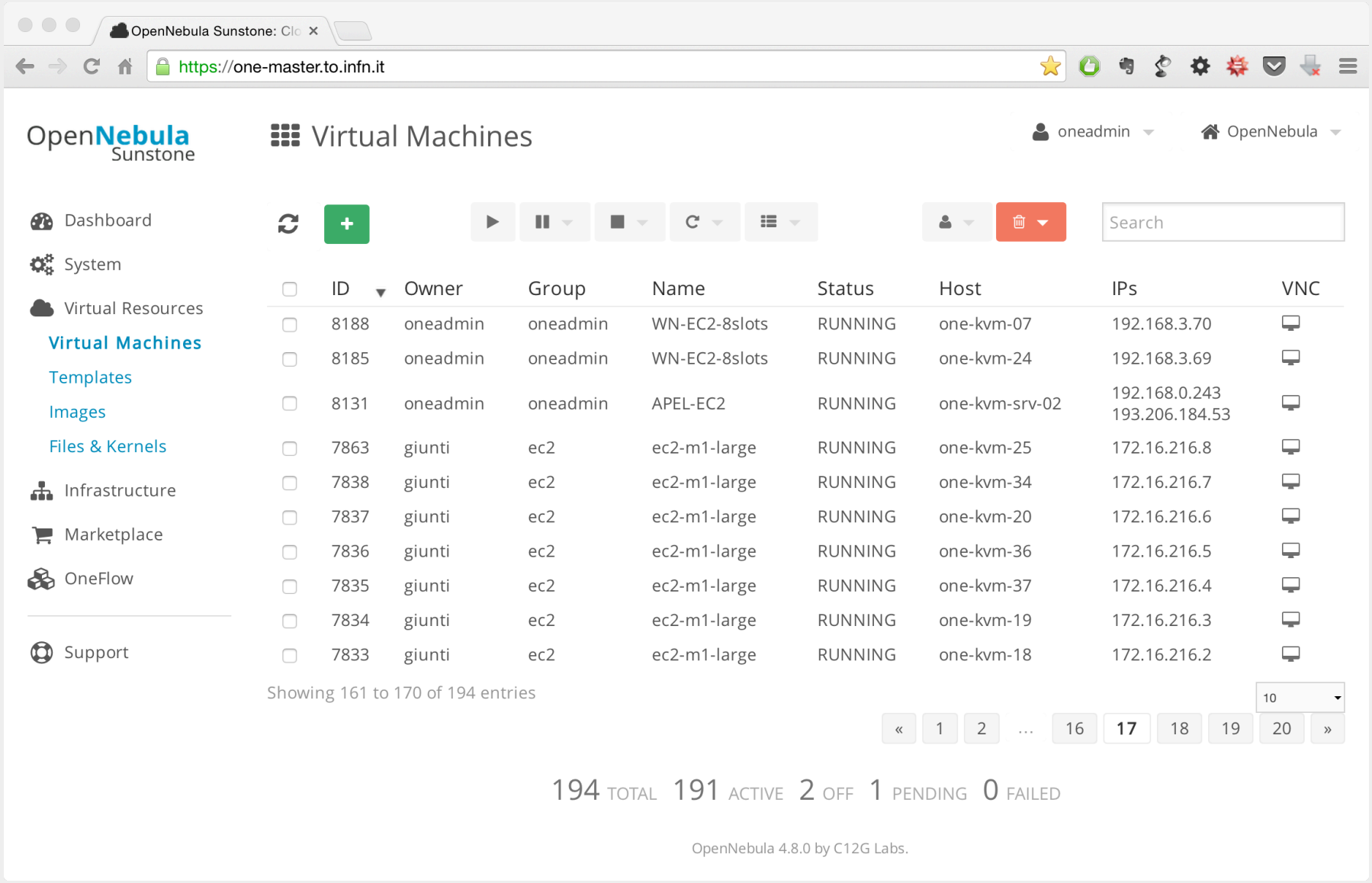
SUNSTONE DASHBOARD



SUNSTONE DASHBOARD



SUNSTONE DASHBOARD



RECENT AND CURRENT ACTIVITY 1

- Upgrade to OpenNebula 4.8 was not completely painless
  - Several things changed
  - We have a few hacks (to support the elastic IP model)
- Making plans for infrastructure 2.0
  - Big overhaul and rationalization planned for next year
  - iSCSI storage upgrade
  - fabric monitoring infrastructure
  - network redesign around 10Gb servers
  - database service consolidation?

CC BY NC ND

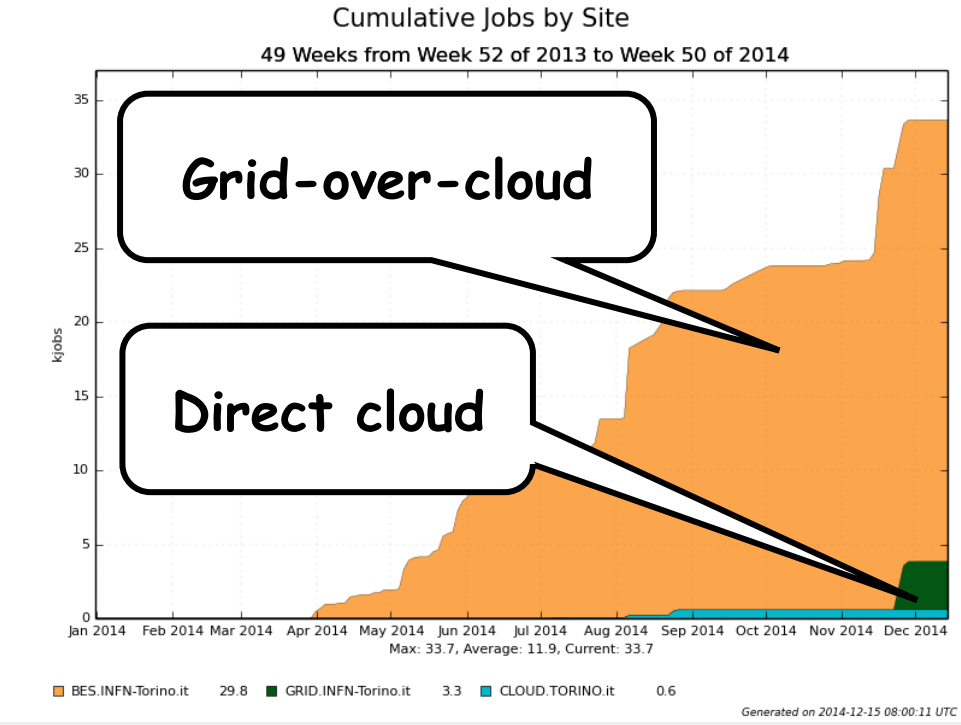
RECENT AND CURRENT ACTIVITY 2

- Coordination of the OCP Monitoring Working Group
  - See Sara's talk later
- Monitoring-as-a-service prototype in place
  - Also see Sara's talk later
- OpenStack test installation being built
  - Primarily for the OCP testbed
  - Will also serve as a test bench for interoperability
- Application integration and operations support
  - Lots thereof

CC BY NC ND

ACTIVITIES WITH IHEP

- Direct interface of DIRAC to the OpenNebula Cloud Manager
  - Done using the rOCCI interface instead of EC2 because DIRAC's rOCCI modules are more mature
  - This is on a separate test instance of OpenNebula
  - DIRAC server at IHEP can directly start and stop VMs according to the number of running jobs
- Several production and stress tests
  - Another large one planned for the coming weeks
  - This is done using the Grid infrastructure, that runs on top of the production Cloud instance



FORTHCOMING ACTIVITIES

- Monitoring-as-a-service and local accounting/billing
- Automatic elasticity where applicable
- Interoperability & integration
- Switch to new components, adopt new ONe tools
  - Use an HTCondor CE to simplify elasticity?
  - Replace homemade VRouters with OpenVSwitch?
  - Use OneFlow (or Cloudify or whatever) to orchestrate complete virtual farms?
- Even more application support...

CC BY NC ND

NEXT BIG CHALLENGE: C3S

- The University of Torino applied to a call by Compagnia di San Paolo for a grant to build an inter-departmental scientific computing facility next year
- Funding approval is not yet final but deemed “highly likely”, total amount at least 500 kEUR
- INFN-Torino is a partner in the project and will host the facility

CC BY NC ND

NEXT BIG CHALLENGE: C3S

- The “Centro di Competenza per il Calcolo Scientifico” (**C^{3}S**) will serve both as a production and an R&D facility
  - Diverse and manifold infrastructure: conventional nodes, “fat” nodes with 4-way servers and huge memory, GPU nodes
  - High-bandwidth, low-latency interconnection (InfiniBand?)
  - Two-tier storage
  - Will share the physical infrastructure with the existing cloud facility

- Will also help building a forum for scientific computing-related activities, and be a platform on which to host further projects

CC BY NC ND

Our ideas:

- Manage the "conventional" part of the facility with a cloud infrastructure like the existing one
  - But this is a true HPC cluster!
- Explore the opportunity of managing heterogeneous resources with a single tool
  - e.g.: GPU virtualization? Can containers help?
- Experiment with interoperability and cloudbursting

CC BY NC ND

CONCLUSIONS

- Gaining a lot of operational experience, solving issues as they present themselves
  - "Agile" development cycle
  - Need to keep growth under control (in terms of number of tenants)
- We will stick to OpenNebula unless somebody provides extremely good reasons
  - Originally adopted for historical reasons
  - However, proved itself ideal for our environment
- Many future developments still vague
  - Will be guided by coordination with INFN-wide strategy
  - C3S will be a challenge but also an opportunity

CC BY NC ND

REFERENCES

- S. Bagnasco, D. Berzano, R. Brunetti, S. Lusso, S. Vallero, "Managing a Tier-2 Computer Centre with a Private Cloud Infrastructure", Proceedings of ACAT2013, Beijing, *J. Phys.: Conf. Ser. 1742-6596* **523** 012012 (2014)
- S. Bagnasco, D. Berzano, R. Brunetti, S. Lusso, S. Vallero, "Integrating multiple computing needs via a Private Cloud infrastructure", proceedings of CHEP2013, Amsterdam, *J. Phys.: Conf. Ser. 1742-6596* **513** 032100 (2014)
- D. Berzano, J. Blomer, P. Buncic, I. Charalampidis, G. Ganis, G. Lestaris and R. Meusel, "PROOF as a Service on the Cloud: a Virtual Analysis Facility based on the CernVM ecosystem", proceedings of CHEP2013, Amsterdam, *J. Phys.: Conf. Ser. 1742-6596* **513** 032007 (2014)
- S. Vallero, A. Guarise, S. Lusso, D. Berzano, S. Bagnasco, "Monitoring of IaaS and scientific applications on the Cloud using the Elasticsearch ecosystem", poster presented at ACAT 2014, Prague.

CC BY NC ND