

**Italian initiative for the optimisation of access to LHC data
and for the technology transfer towards other research areas**

(ATLAS activities)

Miguel Villaplana
Università degli Studi & INFN Milano

IVICFA's Fridays: COMPUTATION in Physics

Valencia, October 17th, 2014

Introduction & Objectives

PRIN: Scientific research program of relevant national interest

- Italian institutions hosting a LHC Tier
- Optimise data access
- Enable efficient interactive analysis
- Technology transfer to other research areas



Consistent with the objectives of the Horizon 2020
“Excellence Science” plan

ATLAS Activities

- EventIndex (see D. Barberis' talk)
- Cloud Computing
- HTCondor
- PoD
- LHCONE and Distributed Tier2

Cloud Computing

2 systems (production and development)

- OpenStack 2014.1 (Icehouse)
- VLAN-based
- Network node 10Gbps
- WN 1Gbps
- Storage facility (Gluster 3.4)
 - Shared Rome1-Napoli, with synchronous replica via WAN
- Machine management via Foreman/Puppet

Cloud Computing

Status

- ATLAS calibration facility migrated to cloud
- Integration of Frontier's squid
 - Test OpenStack LBaaS service
- First tests of autoscaling feature

Plans

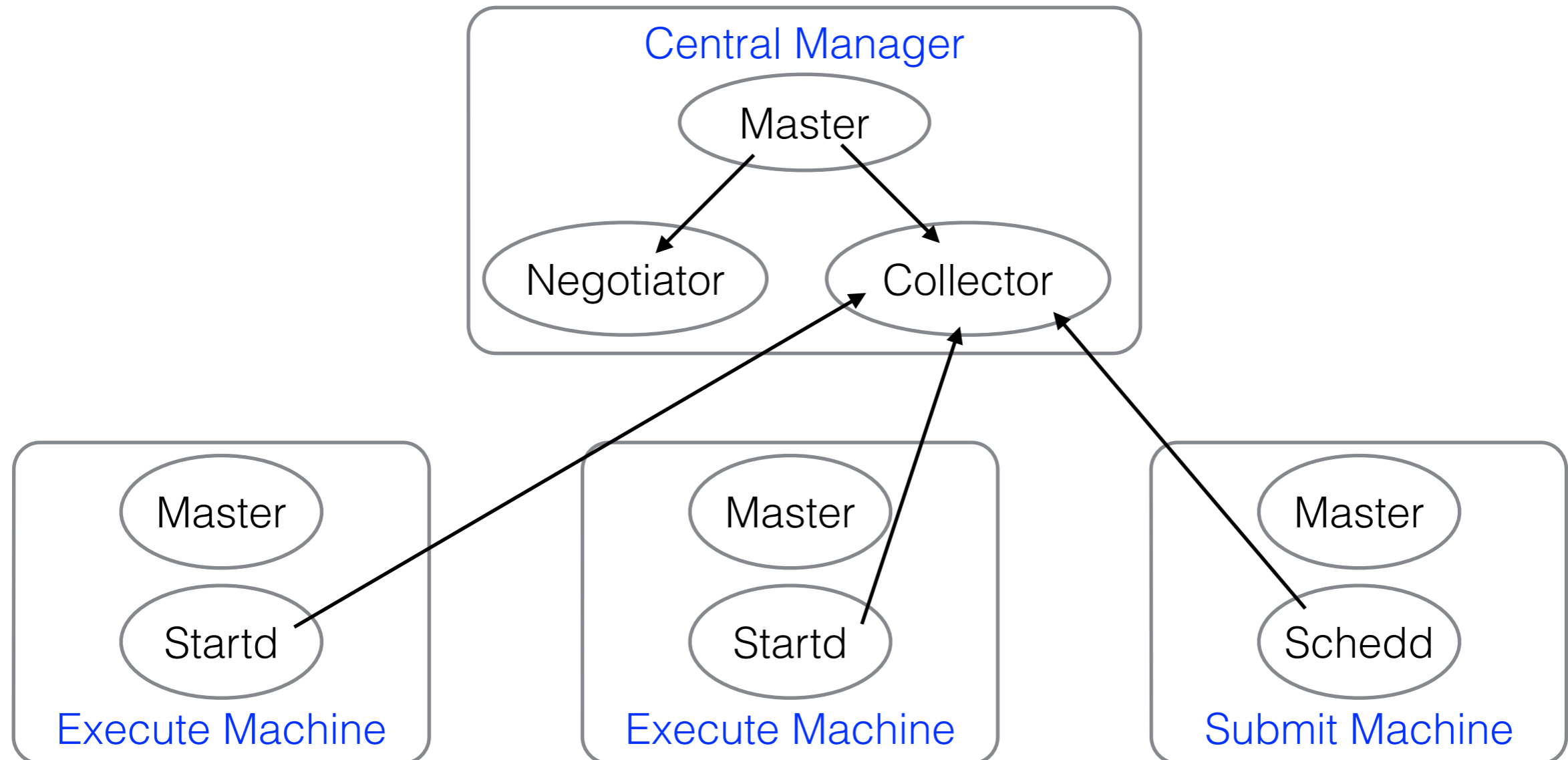
- Further testing of LBaaS and Autoscaling
- Application of the concept to real services (squid, DB cluster nodes, web services, etc.)
- Extension of the cloud cell to the distributed Tier2
- Extension to a multi-cell environment

HTCondor

- Specialised **workload management system** for compute-intensive jobs.
- Provides a job queueing mechanism, scheduling policy, priority scheme, resource monitoring, and resource management.
- In many circumstances, can **checkpoint** and migrate a job to a different machine when needed.
- **Flocking** technology allows multiple HTCondor pools to work together.

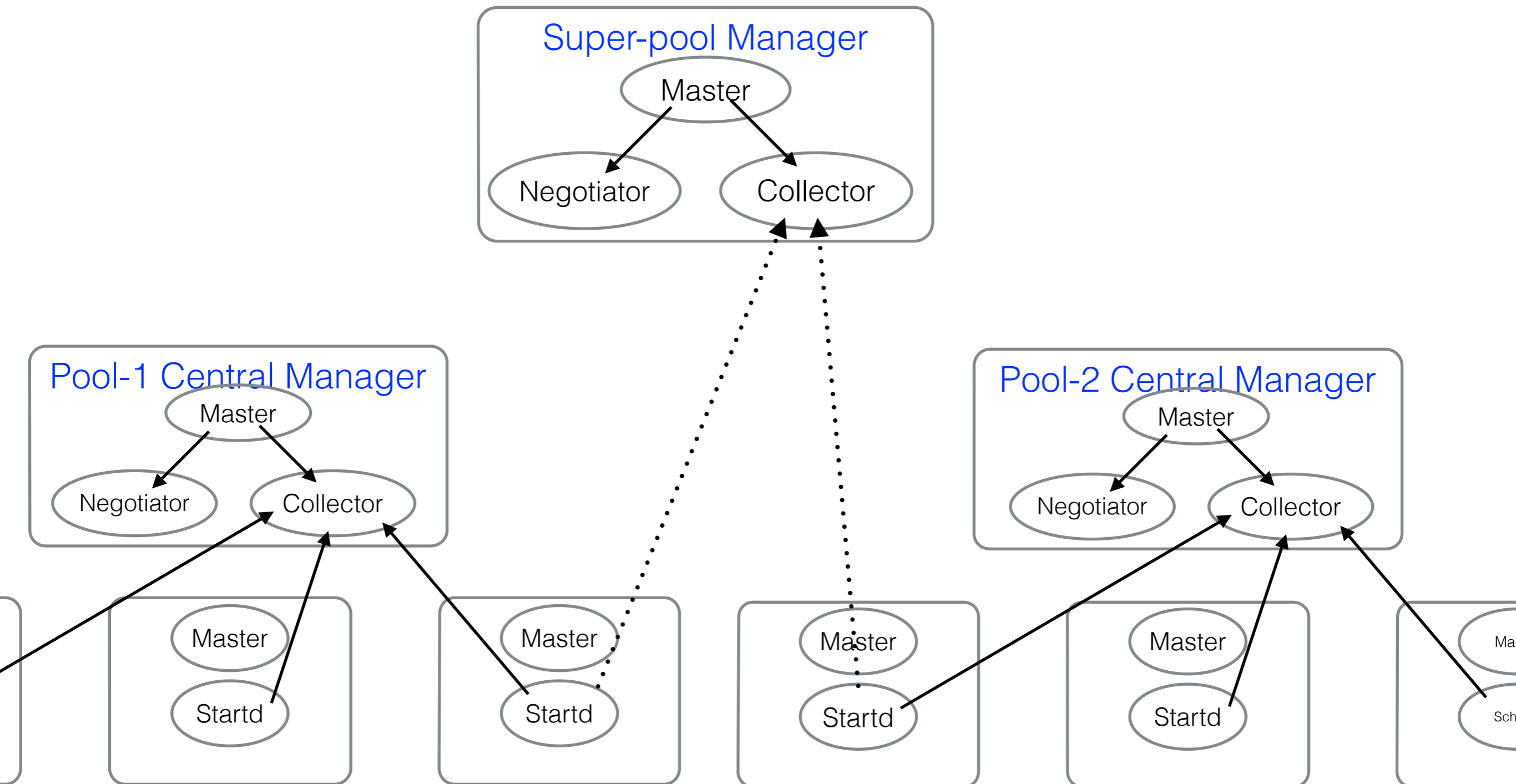
Share resources across groups with HTCondor

- Each group has its own resources under a HTCondor pool.
- Machines report to the pool's central manager.



Share resources across groups with HTCondor

- Define a **super-pool**: one new central manager which all the execute machines **also** report to.



Share resources across groups with HTCondor

- Define a **super-pool**: one new central manager which all the execute machines **also** report to
 - This provides usage accounting across all of the resources together.
 - It serves as a top-level pool to submit jobs to when users want to access all possible resources.
- Group pools remain the default pool for job submission, but with the super-pool added to their **FLOCK_TO** list.
 - This way, users get the quality of service they were already enjoying from their own central manager, but excess jobs may be conveniently sent to all of the other resources.
- Give the group's negotiator priority over super-pool.
 - This guarantees high priority to group users on their own machines.

Status

- Super-pool setup tested using 6 eight-core machines recently decommissioned from Tier2.
- Users have been encouraged to run their jobs in the ATLAS HTCondor pool.
- All have been successful (with some help).
- Different analysis use-cases have been tested.
 - Real complete ATLAS physics analysis.
 - MC generation using software from CVMFS, or local, or both.
 - Nuclear physics calculations.
 - Astrophysics (open MPI).

Next steps

- Once non-ATLAS users are comfortable enough with HTCondor, we can help them organise their resources under HTCondor pools.
- Aggregate all pools under a super-pool.
- Might be interesting to add a common shared area.
- Study the possibility of adding extra resources from cloud when needed.

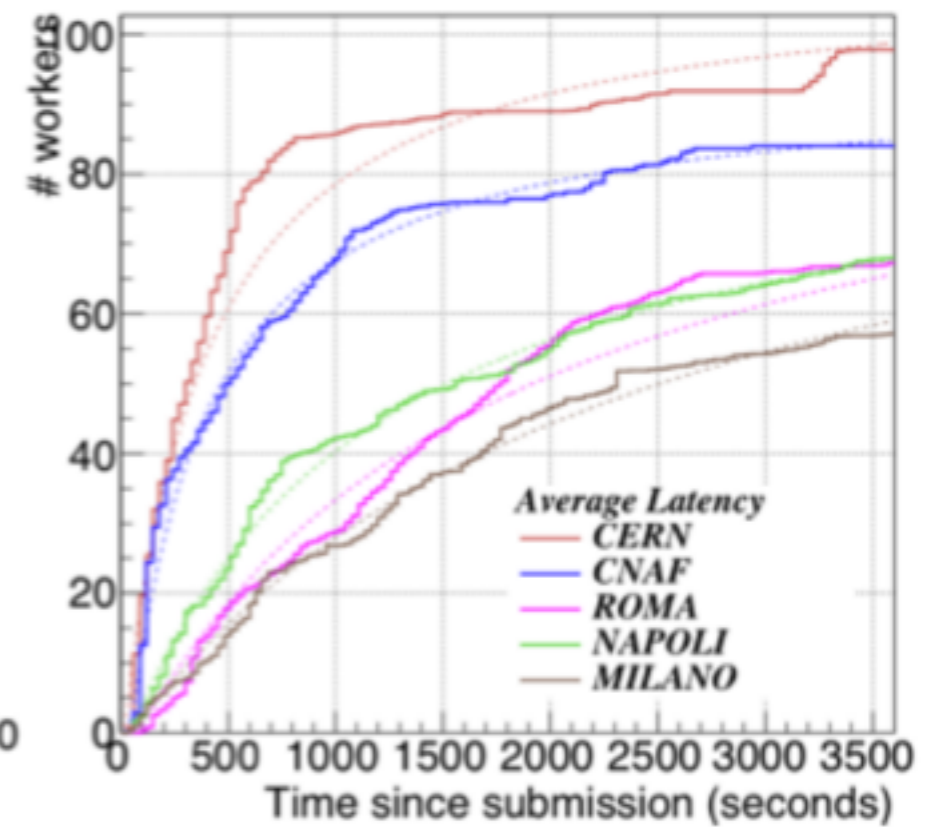
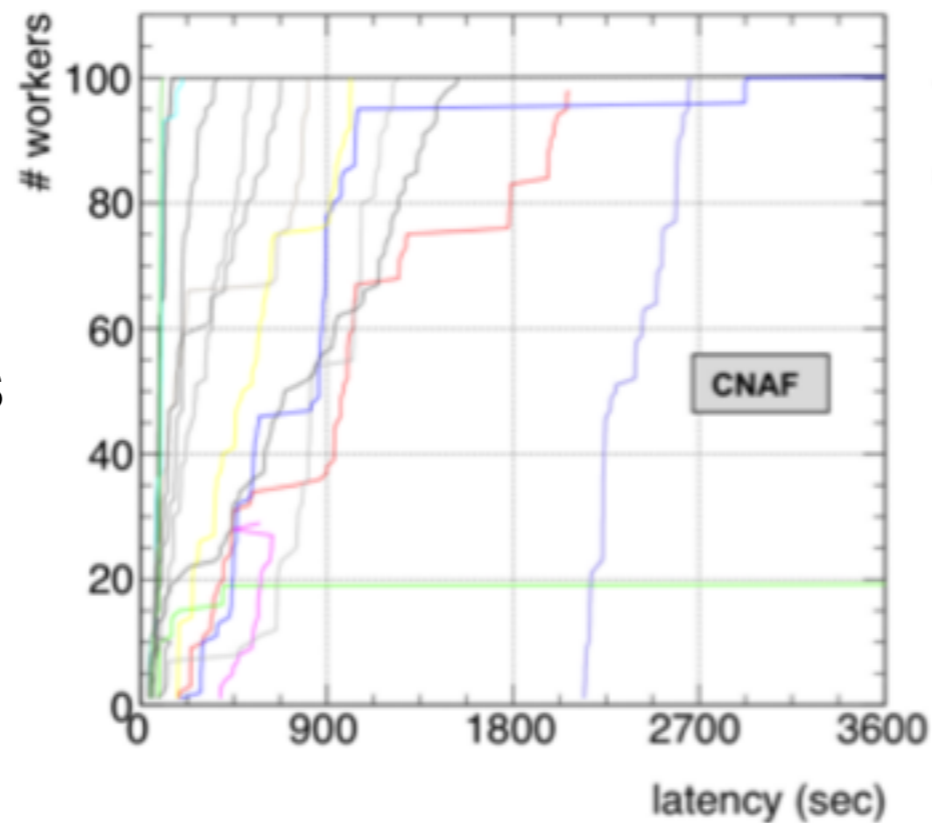
PoD

PROOF on Demand (PoD) is a tool-set developed at GSI, which sets up a PROOF cluster on any resource management system (<http://pod.gsi.de/>)

PoD

Tests (2014 J. Phys.: Conf. Ser. 513 032102)

- PoD in PanDa
 - Submit N jobs, which activate the xproofd process in the WN
 - Once the cluster is completely active the analysis is submitted directly from the PROOF master (ie, UI)
- ATLAS analyses
 - Different real analyses
 - Usual Tier activity
- Storage benchmark
 - DPM, StoRM/GPFS, EOS
- Xrootd on LAN, WAN (FAX)
- Startup latency
 - Site size/occupation



PoD

Next Steps

- More realistic tests (bigger # users)
- Repeat storage benchmarks
 - new Rucio redirector
 - data access saturation vs # workers
- Test dynamic addition of resources with PoD (D.Berzano)

LHCONE and Distributed Tier2

GARR-X is the project for a next-generation multi-service telecommunication network for the Italian Academic and Research community that shall gradually replace the existing network

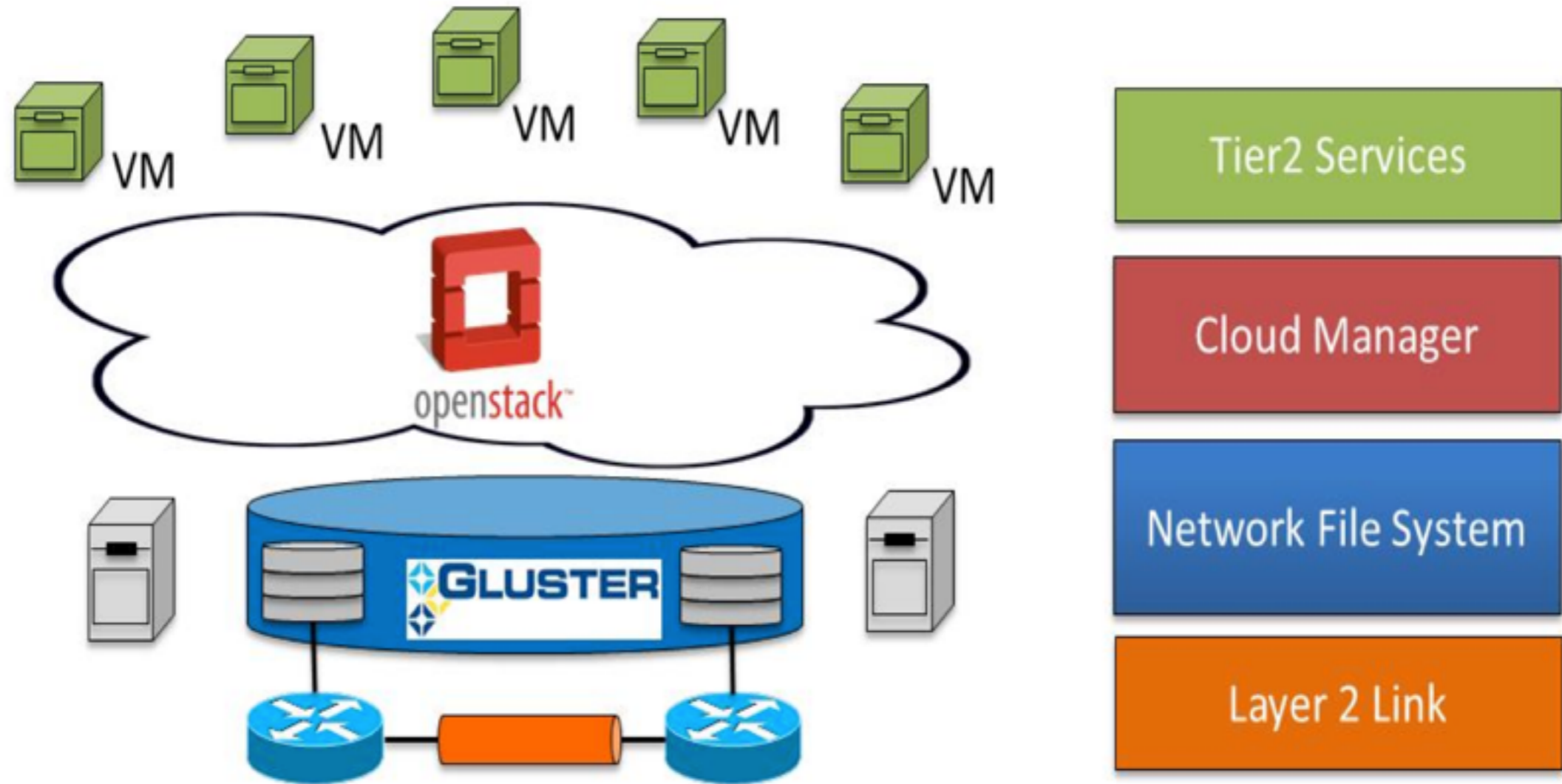
- Layer 2 geographic connexions between 2 sites
- Dedicated network
- QoS: latency, delay, jitter, ...



LHCONE and Distributed Tier2

- The **objective** is to create a geographically distributed computation facility
- Tests between INFN Napoli and Roma La Sapienza
 - Dedicated network of 1Gbps
 - Layer 2 geographic connection allows:
 - Network sharing (ie, WNs or internal storage networks)
 - Better performance wrt routing
 - Shared GlusterFS
 - Cloud Manager (OpenStack Icehouse)

LHCONE and Distributed Tier2



Status

- Successfully verified the possibility of using a layer 2 geographic link to create virtual machines
- First resilience tests show good behaviour
- Now testing:
 - Architecture response to latency variations due to interaction with GARR
 - GlusterFS vs CEPH
 - Implementation of a distributed cloud with federated cells
 - Test ATLAS distributed services, via grid in HA or a distributed cloud for the experiment

Summary

- Preliminary tests on cloud computing as a service done
 - Works, but not out of the box
 - Some stability and performance issues detected
- The possibility of sharing resources with other research areas using a solution based on HTCondor has been tested
 - Prototype shows good behaviour
- PoD has been thoroughly tested
 - Eager to see dynamic additions of resources at work
- Distributed Tier2
 - Layer2 link works fine
 - First resilience tests show goog behaviour