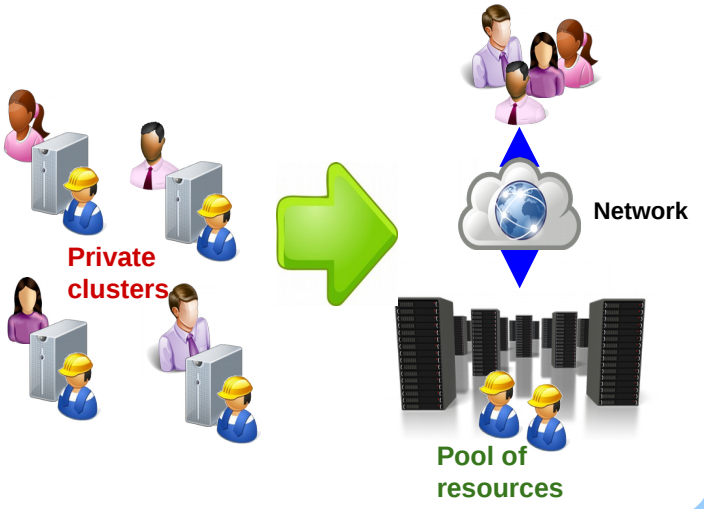
Lisa Zangrando - INFN Padova

On behalf of the ‘Cloud Area Padovana’ team

The Cloud Area Padovana: from pilot to production

Cloud Area Padovana

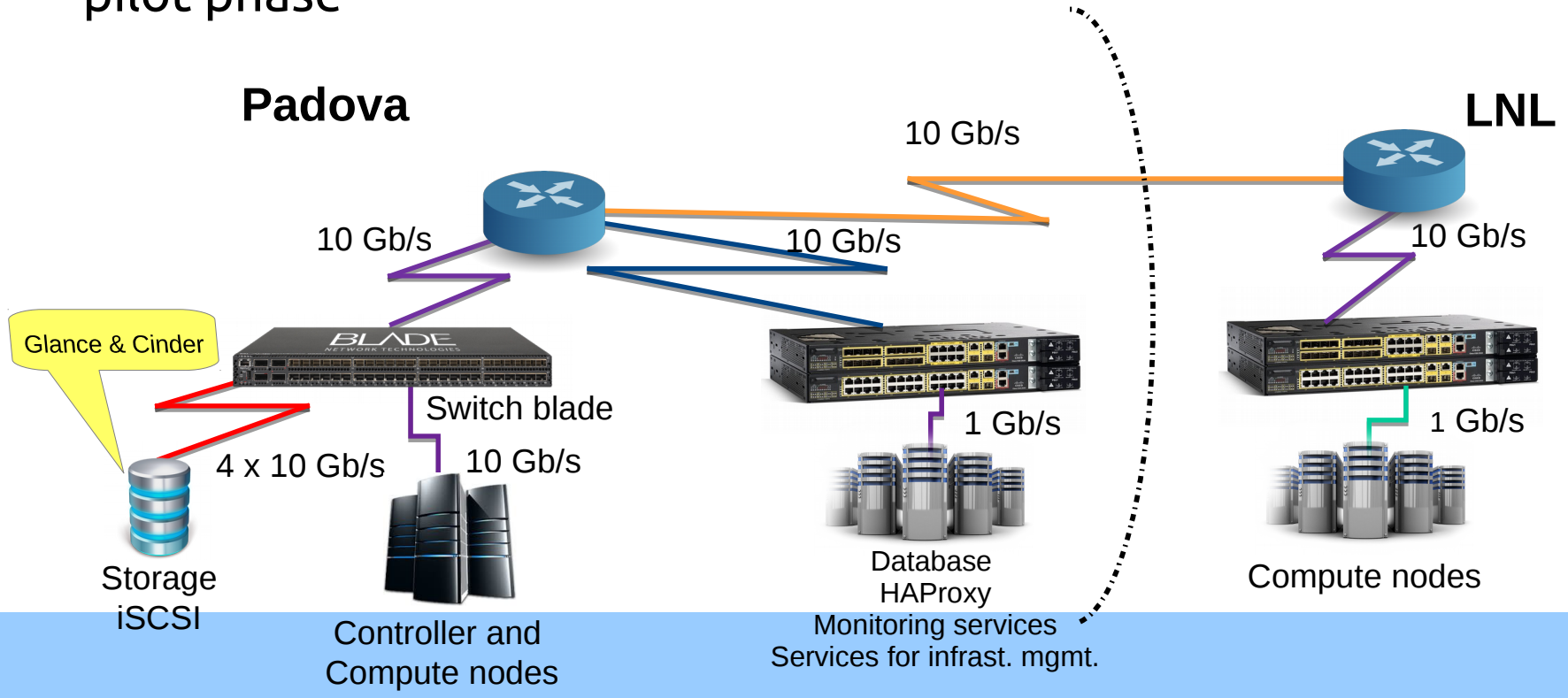
- Project started at the end of 2013 for realizing a Cloud infrastructure targeted to local physics experiments and services
- Main goal: improve the overall computing resources usage and decrease their maintenance costs
- Resources distributed between two INFN sites: Padova and INFN Legnaro National Labs (10 km far away)
  - the same data centers hosting the WLCG Tier-2



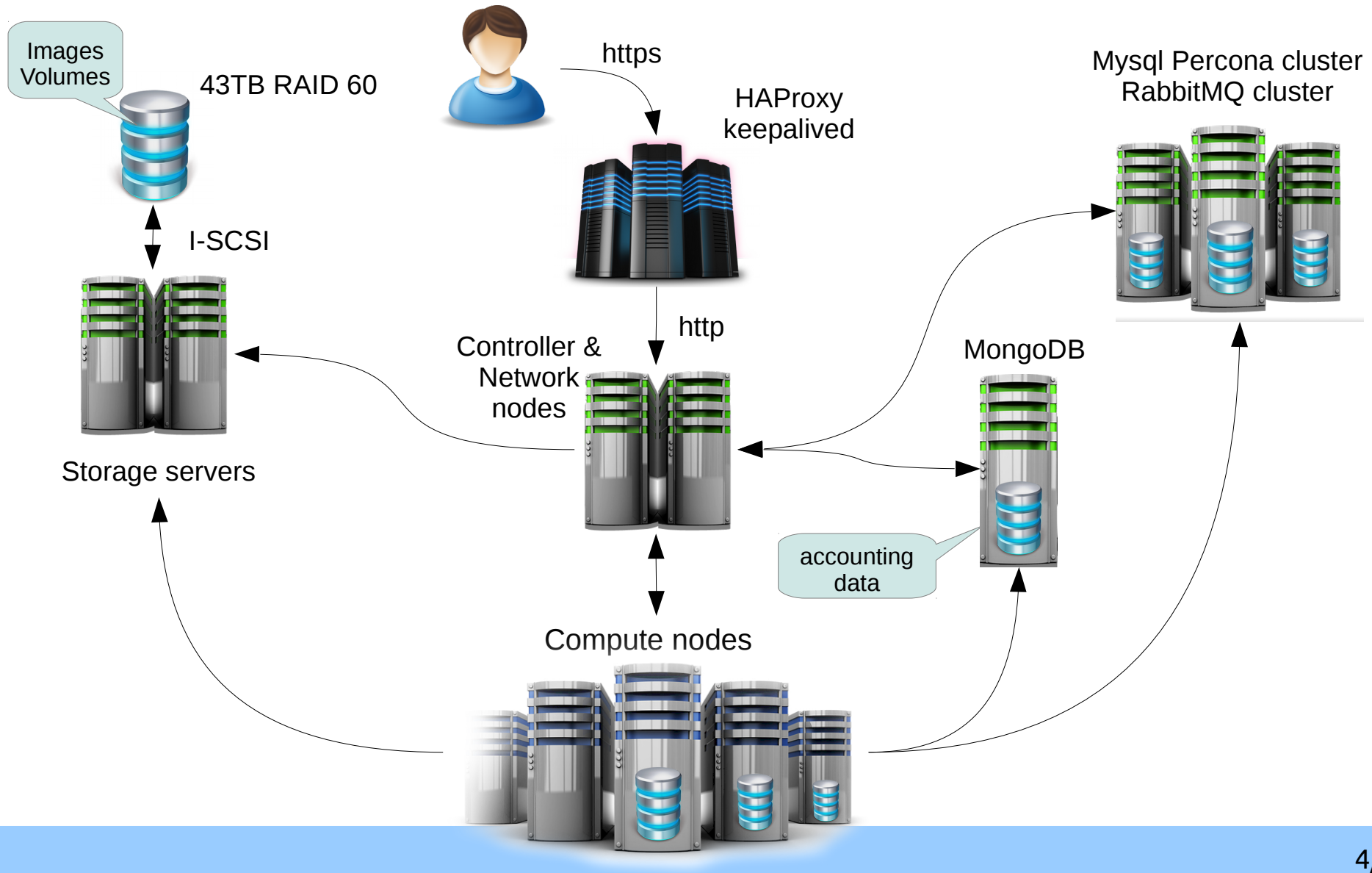
Cloud Area Padovana: current status

- Single OpenStack based IaaS
  - all services are deployed just in Padova and configured in High Availability mode
- In production since the end of 2014, after a pilot phase

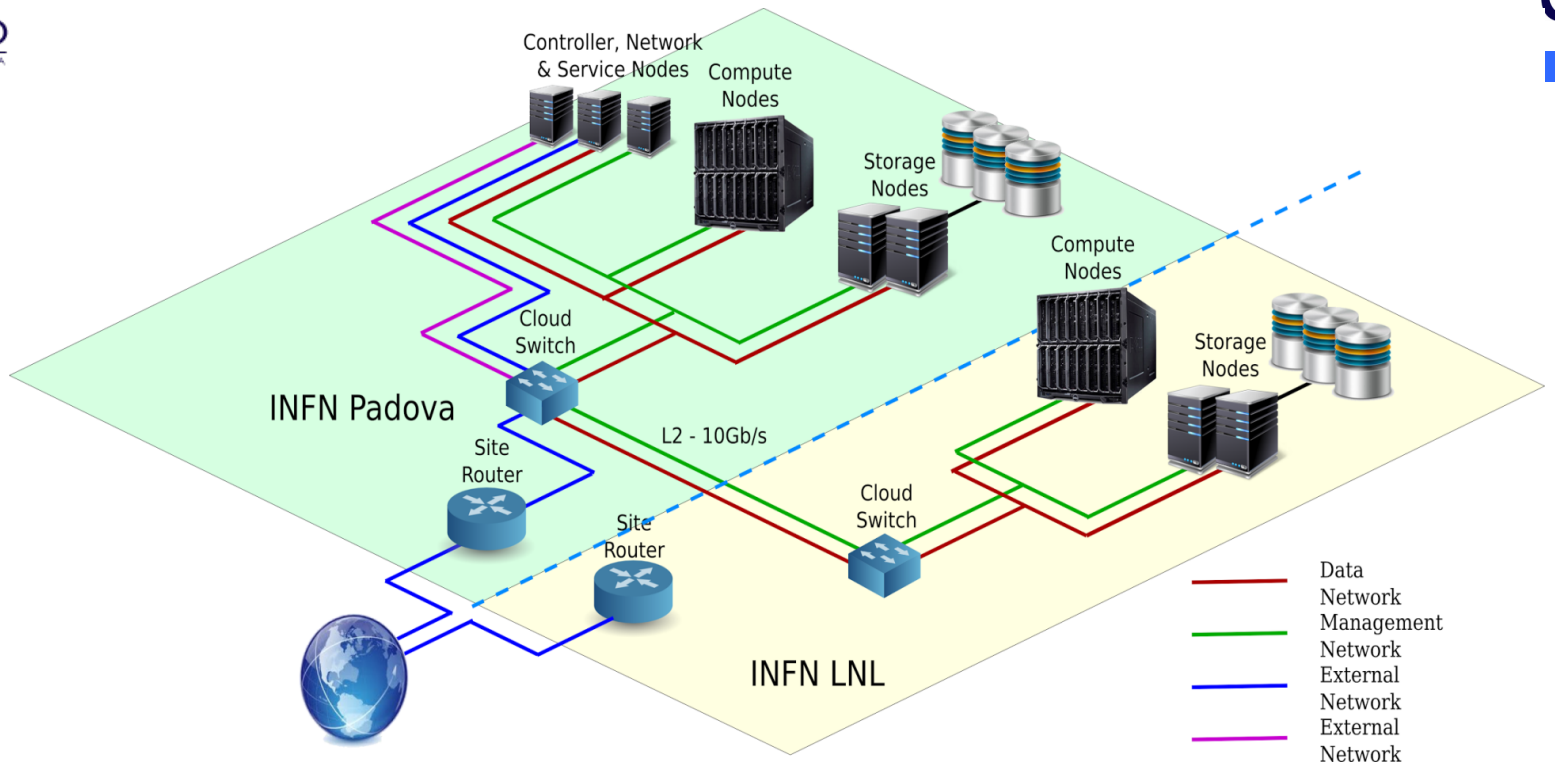
	Compute Nodes	Cores (in HT)	Storage (TB)
Padova	15	656	43 (images, volumes)
LNL	13	416	
Total	**28**	**1072**	**43**



Architecture



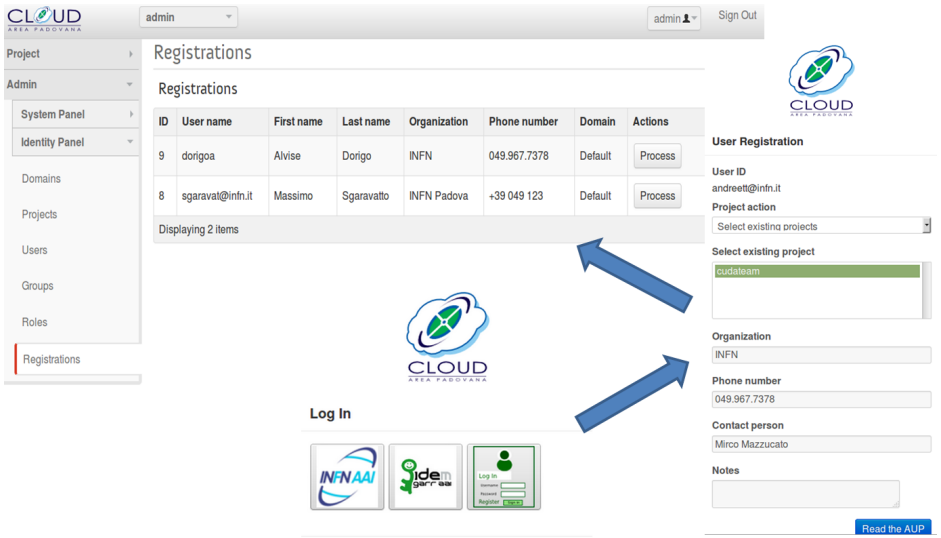
Networking

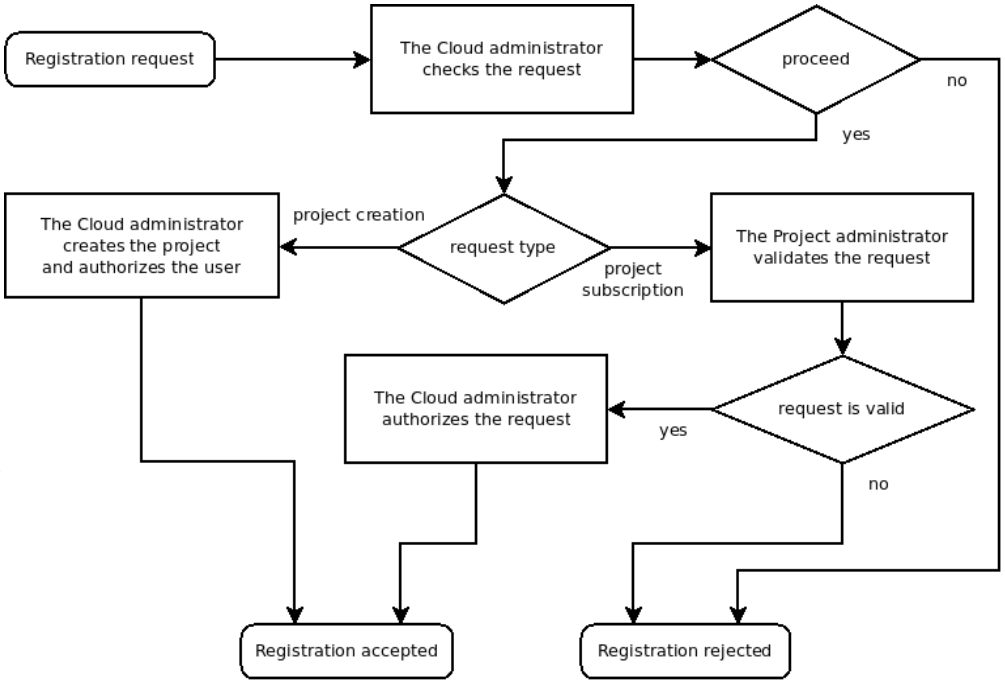


- Neutron with Open vSwitch/GRE configuration
- Two virtual routers with external gateways on public and LAN networks
  - this allows accessing the Cloud VMs from our LANs even if they don't have a floating IP
- GRE tunnels among Compute nodes and Storage servers to allow high performance storage access (via e.g. NFS) from VMs

Identity and access management

- OpenStack Keystone Identity service and Horizon Dashboard were extended:
  - to interact with INFN-AAI Identity Provider
  - to manage user and project registrations
    - requests for registration pass through a workflow (involving the Cloud administrator and the project manager) in order to be approved or rejected



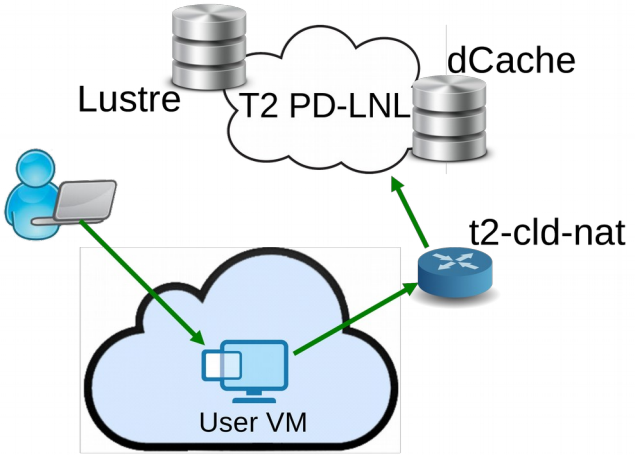


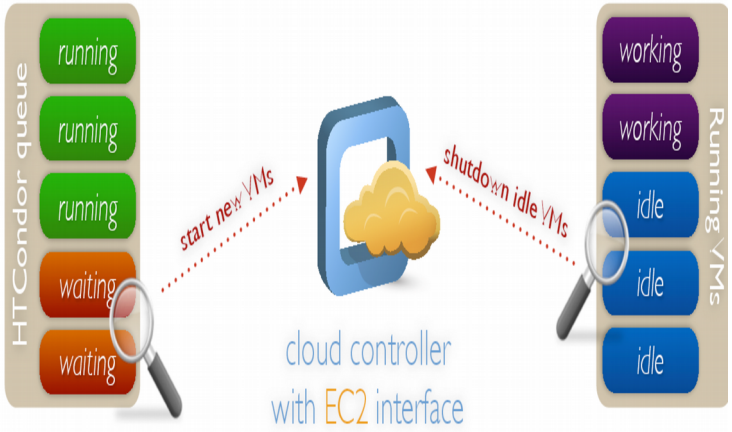
Usage

- 90+ users
- ~25 projects
- resources used for satisfying different use cases
  - interactive analysis
  - batch processings
  - services

Some use cases: CMS

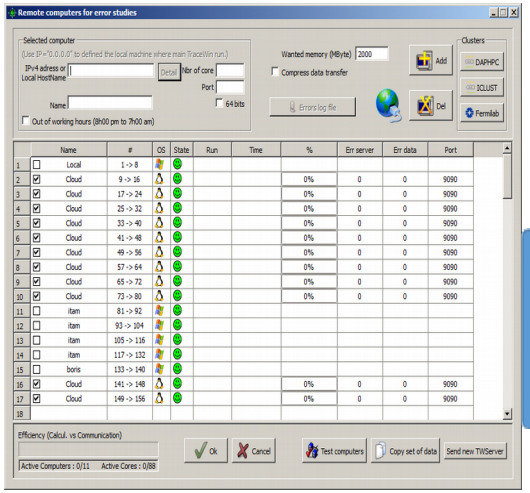
- Interactive usage
  - code development and build, ntuple productions, end-user analysis, Grid User Interface
    - each user instantiates his own VM and destroys it when not needed anymore
  - Cloud is integrated with the local Tier-2
    - access to dCache and Lustre storage
- Batch usage
  - elastic batch cluster (HTCondor) managed by elastiq
    - new VMs are created when there are jobs waiting in the queue
    - VMs are destroyed when idle (i.e. queue empty)





Some use cases: SPES

- Simulations for the SPES experiment
  - to tune the “perturbated” parameters of the accelerator
  - necessary to have many simulations in a short time
- This is done on the client using a client-server framework called TraceWin



TraceWin Client with GUI on local desktop

Remote Server On Cloud
Remote Server
Remote Server
Remote Server
Remote Server
Remote Server
Remote Server
Remote Server
...

Some use cases: CMT

CMT: Cosmic Muon Tomography

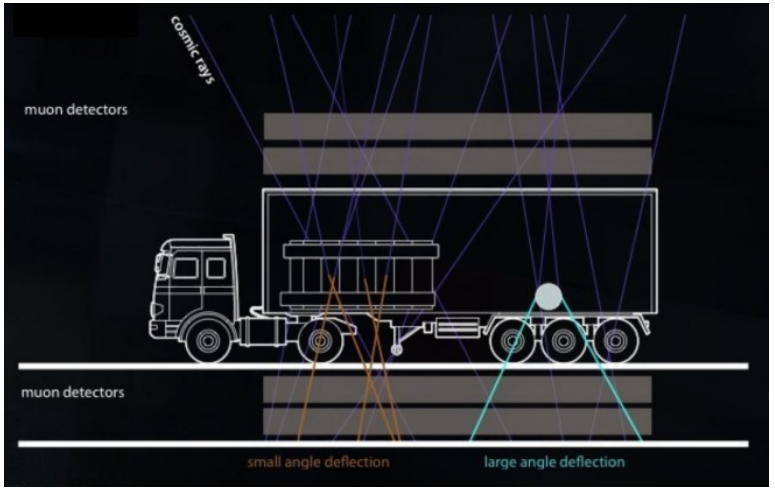
Imaging system to scan inaccessible volumes of materials with high atomic number…

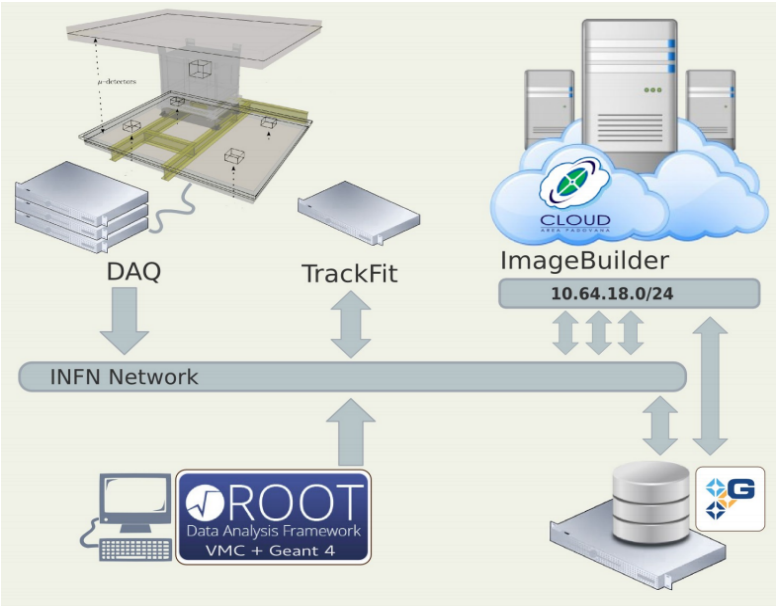
… by studying the scattering of muons after matter collisions and applying statistical and iterative algorithms

… to find the optimal density of the target material that fits the muon trajectories and then its volume and shape

The algorithms for data analysis require a huge computing power

Using ImageBuilder as reconstruction software



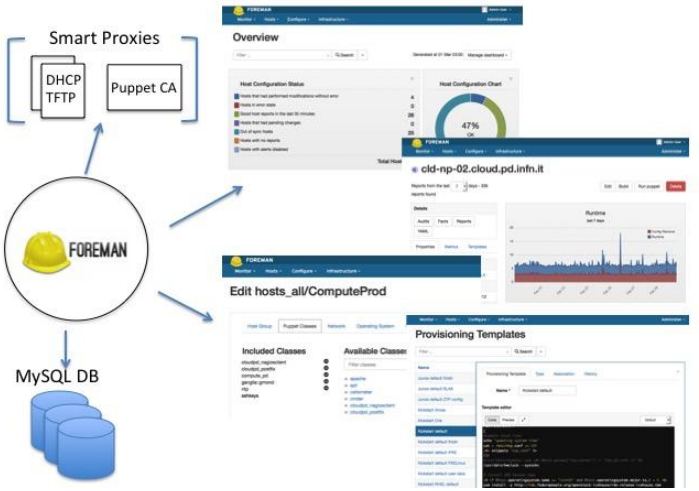


Some lessons learnt ...

- Properly evaluate where to deploy the services
  - in particular don't mix storage servers with other services
  - initial configuration:
    - 2 nodes configured as controller nodes
    - 2 nodes configured as network nodes + storage (Gluster) servers
  - current deployment:
    - 2 nodes configured as controller nodes + network nodes
    - 2 nodes configured as storage (Gluster) servers
- Database is a critical component
  - started with Percona deployed on 3 VMs, then moved to physical machines for performance reasons
  - using different primary servers for different services (e.g. glance, cinder)

Some lessons learnt … (continued)

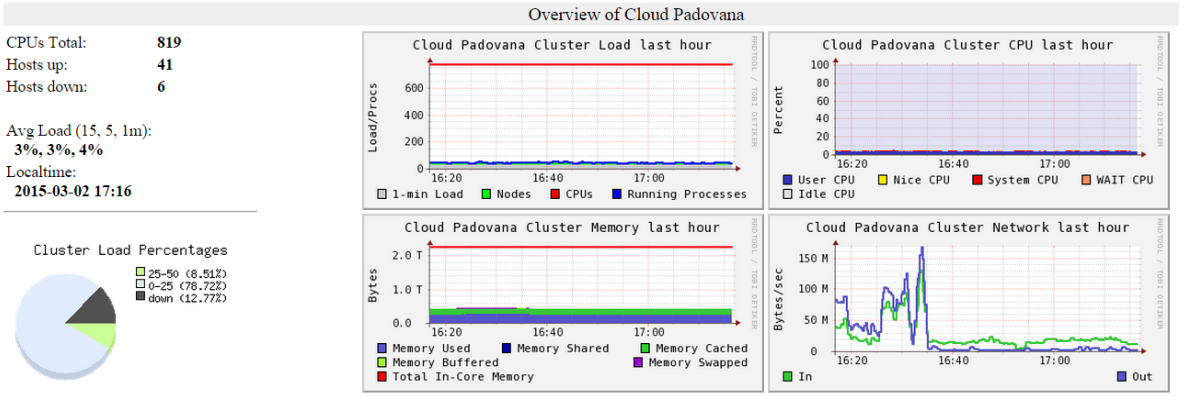
- Avoid any manual configuration
  - combined use of Foreman + Puppet as infrastructure manager
    - used not only to configure OpenStack, but also the other services (e.g. ntp, nagios probes, ganglia, etc)

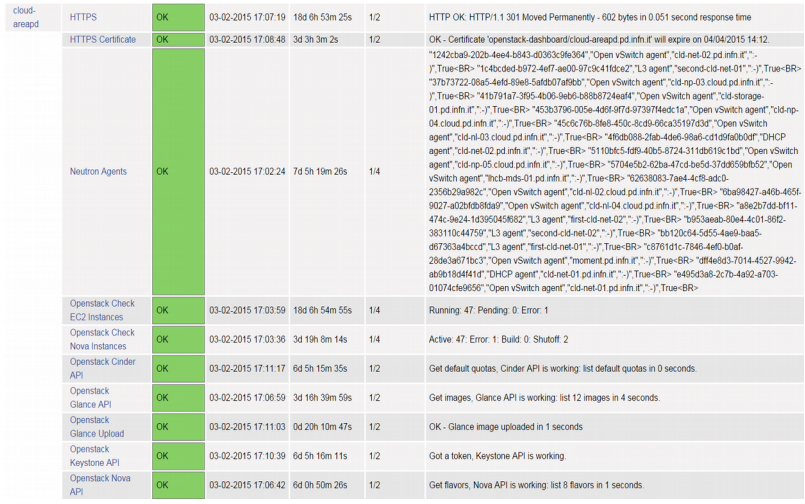


- Evaluate pros and cons of live migration
  - we experienced scalability problems by using a shared file system (GlusterFS) to enable live migration
    - live migration not really a requirement for many applications and it doesn't always work out of the box
  - now only a few compute nodes have a shared file system to host 'critical' VMs
    - the other nodes use the local storage

Some lessons learnt … (continued)

- Monitoring is very important
  - monitoring infrastructure based on Nagios, Ganglia and Cacti
  - in particular Nagios heavily used to prevent/early detect problems
    - e.g. nagios sensors that instantiate VMs on each compute node and test their network connectivity, etc.
  - developing a tool (CAOS) to collect and harmonize monitoring and accounting information





Updates

- We perform one OpenStack update per year (i.e. skipping one OpenStack release)
  - as balance between having latest features and fixes and the need of limiting the manpower
  - OpenStack LTS releases would be really appreciated!
- Every change (in particular the updates) are prepared and tested on a testbed environment before being deployed in production
  - small infrastucture which however simulates the production one (e.g. services deployed in High Availability)
- We are now running OpenStack kilo but almost ready with the update to Mitaka

Some future activities

- Deployment of the Indigo-DataCloud Synergy service
  - to assure an optimal resource usage, without static partitioning
  - see talk #367 and poster #357

- Finalize the implementation of the CAOS monitoring and accounting portal

- Integration with the University of Padova's Cloud, to implement a reference Cloud for our local communities

Questions?