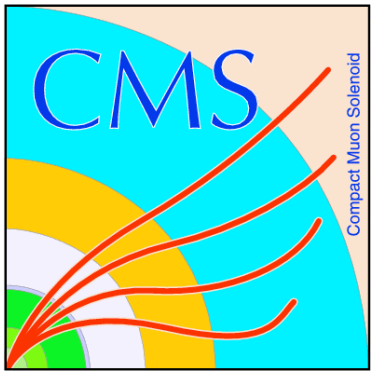


Elastic Computing from Grid sites to External Clouds

GIUSEPPE CODISPOTI, RICCARDO DI MARIA

C. AIFTIMIEI, D. BONACORSI, P. CALLIGOLA, V. CIASCHINI, A. COSTANTINI, S. DAL PRA, D. DEGIROLAMO, C. GRANDI, D. MICHELOTTO, M. PANELLA, G. PECO, V. SAPUNENKO, M. SGARAVATTO, S. TANEJA, G. ZIZZI

13-18 MARCH 2016, ACADEMIA SINICA, TAIPEI, TAIWAN

Abstract

Scientific experiments nowadays require ever-**increasing computing resources**

- Often they work in a ***"burst"*** modality:
  - peak usage periods where resource usage greatly increases with respect to periods of "normal“ usage

Traditional scientific (non-commercial) computing centres (CC) may find it difficult to size themselves

- They cannot be sized for peak usage
- They cannot acquire extra resources on demand
- Finally, they cannot absorb the peak usage of the experiments without generating excessively long queues.

Dynamic resources provisioning:

Access Cloud resources provided by external partners or commercial providers

- INFN- along the lines followed by other WLCG sites – is exploring the opportunity to access Cloud resources in order to cope with the request peaks

Bologna Tier3 Proof of Concepts

Dynamic resources provisioning

- Cloud Bursting: extending an existing site's LAN based batch system to:
  - Bologna Tier3 inside CNAF Tier 1 domain
  - Bologna Tier3 outside Tier 1 environment: Openstack

Access a newly defined Cloud-based WLCG site through experiments Workload Management Tools

- CNAF Openstack and CMS Production Workload Management Infrastructure

Cloud Bursting:
extending existing site's queues

Cloud Bursting from an LSF site

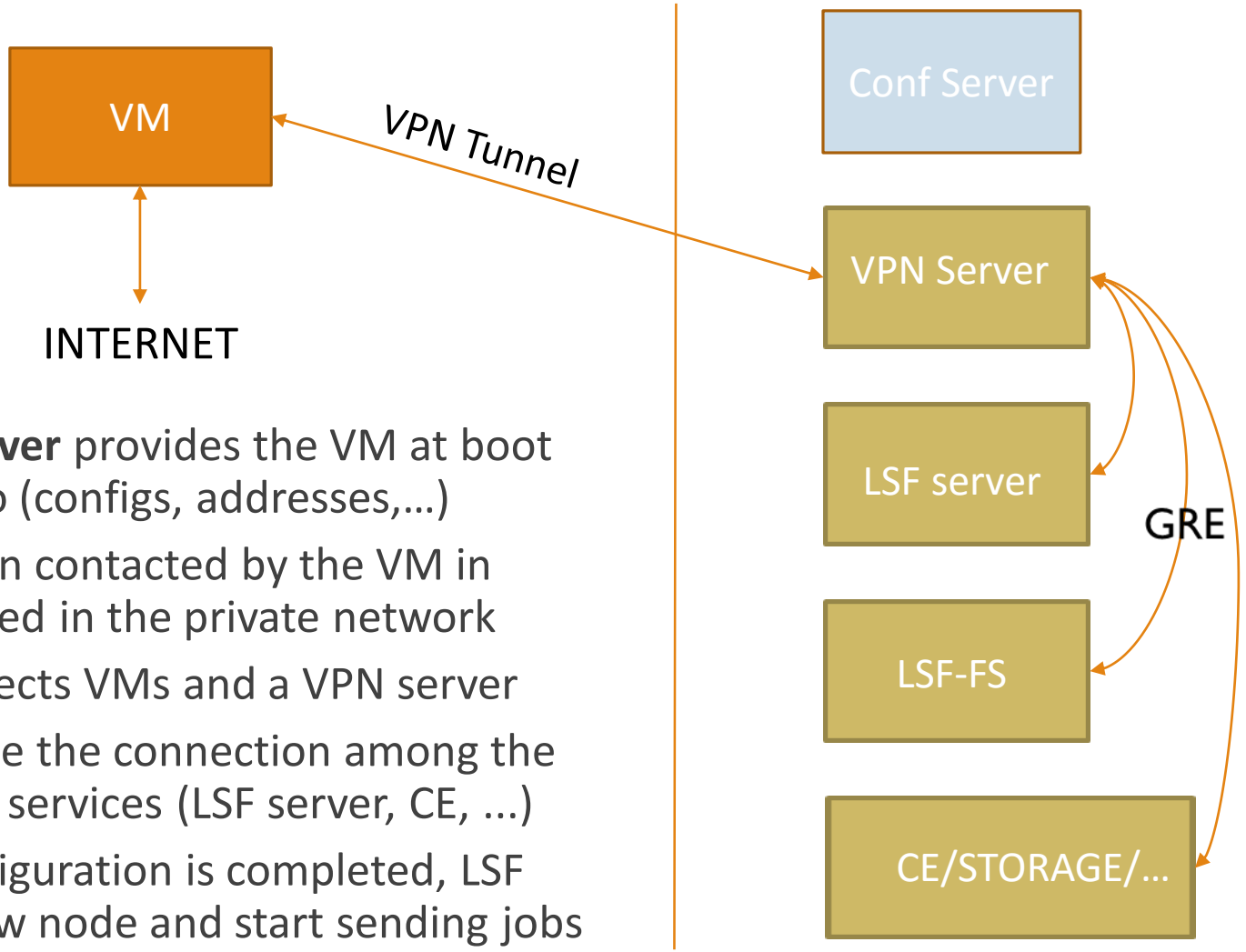
The idea: enabling dynamic extension of batch system working inside a LAN (e.g. LSF)

Problem: LSF does not support dynamic extensions of the batch queues

Solution: a VPN where new nodes (VMs) can be added dynamically

- No requests on the hypervisor
  - VMs can be run everywhere, Cloud providers included
- Only request on the VM is the installation of 2 additional rpms
- The network traffic is reduced to the minimum

The VPN configuration



- A **configuration server** provides the VM at boot with all needed info (configs, addresses,…)
- A **VPN server** is then contacted by the VM in order to be registered in the private network
- A **VPN tunnel** connects VMs and a VPN server
- A **GRE tunnel** enable the connection among the VMs and necessary services (LSF server, CE, ...)
- As soon as the configuration is completed, LSF Master sees the new node and start sending jobs

Dynamic Extension of the Bologna Local Farm

Step 0: *Virtualization*

- Used custom, «lightweight» images
  - Relying where possible on remote service (CVMFS)
  - No EMI Grid Software
- Where possible, reused existing Tier3 configurations
- Tested access to Tier-3 GPFS storage

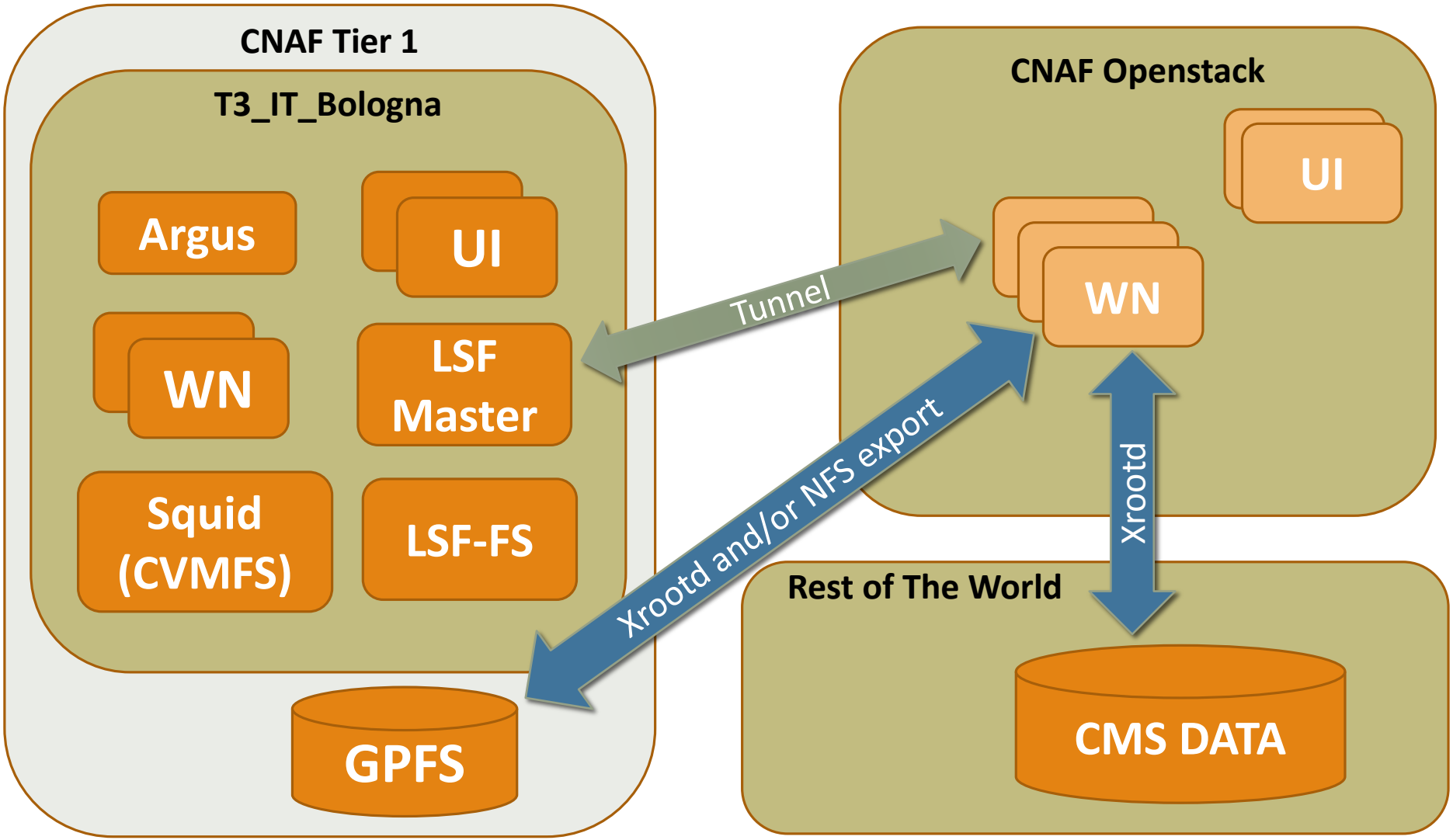
Step 1: *extending Bologna Local Farm*

- E.g.: adding static nodes to the farm and accessing them through a test LSF queue
- Testing the LSF dynamic extension

Step 2: *extending Tier3 Grid site to CNAF Openstack*

- E.g.: plug VM instantiated on Openstack into a Grid production queue

The Extended Tier3 configuration



Bologna Tier3 Cloud Bursting

Successfully performed the Bologna Tier-3 extension

- Tested over the CNAF Openstack infrastructure (both Havana and Juno)
- Experimented the access to the local storage (gpfs) through nfs export
  - Not ideal solution: bottleneck for the VMs and the whole gpfs system
  - Switched back to remote data access (xrootd, srm)

New nodes seen as “normal” Tier-3 nodes from Grid submission

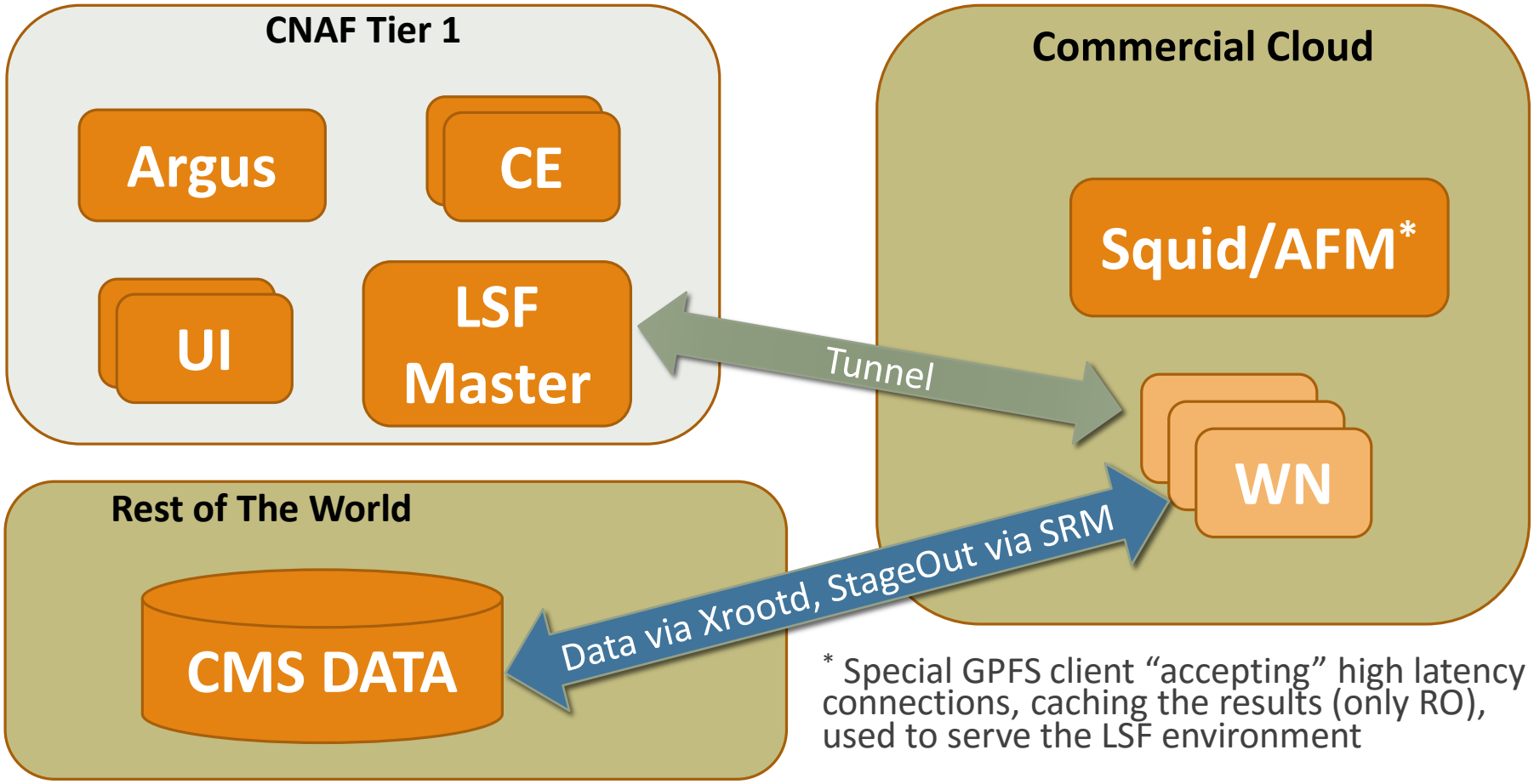
Completely transparent for the CMS tools:

- **used in the production system!**

Submitted >3000 jobs

- Using standard CMS workflow for final Analysis Objects creation
- Jobs spread among physical nodes and plugged VMs
  - About 5% reached the virtual nodes
- No failure observed

Extending over commercial resources: an example



* Special GPFS client "accepting" high latency connections, caching the results (only RO), used to serve the LSF environment

For the ***real application***, see "*Elastic CNAF DataCenter extension via opportunistic resources*" presented by Dr. Stefano DAL PRA on 18 Mar 2016 at 10:00

The Bologna Tier 3 as a Pure Cloud Site

A CMS Site «as a Service»

Using the Bologna Tier3 setup we created a brand new, decoupled CMS site «as a service» in Openstack

- Registered as T3_IT_BolognaCloud inside CMS WM system

Access the new Site through the standard CMS WM tools

- CRAB: analysis jobs submission tool
- GlideIn-WMS: job management tool
- (WMAgent: production jobs submission tool)

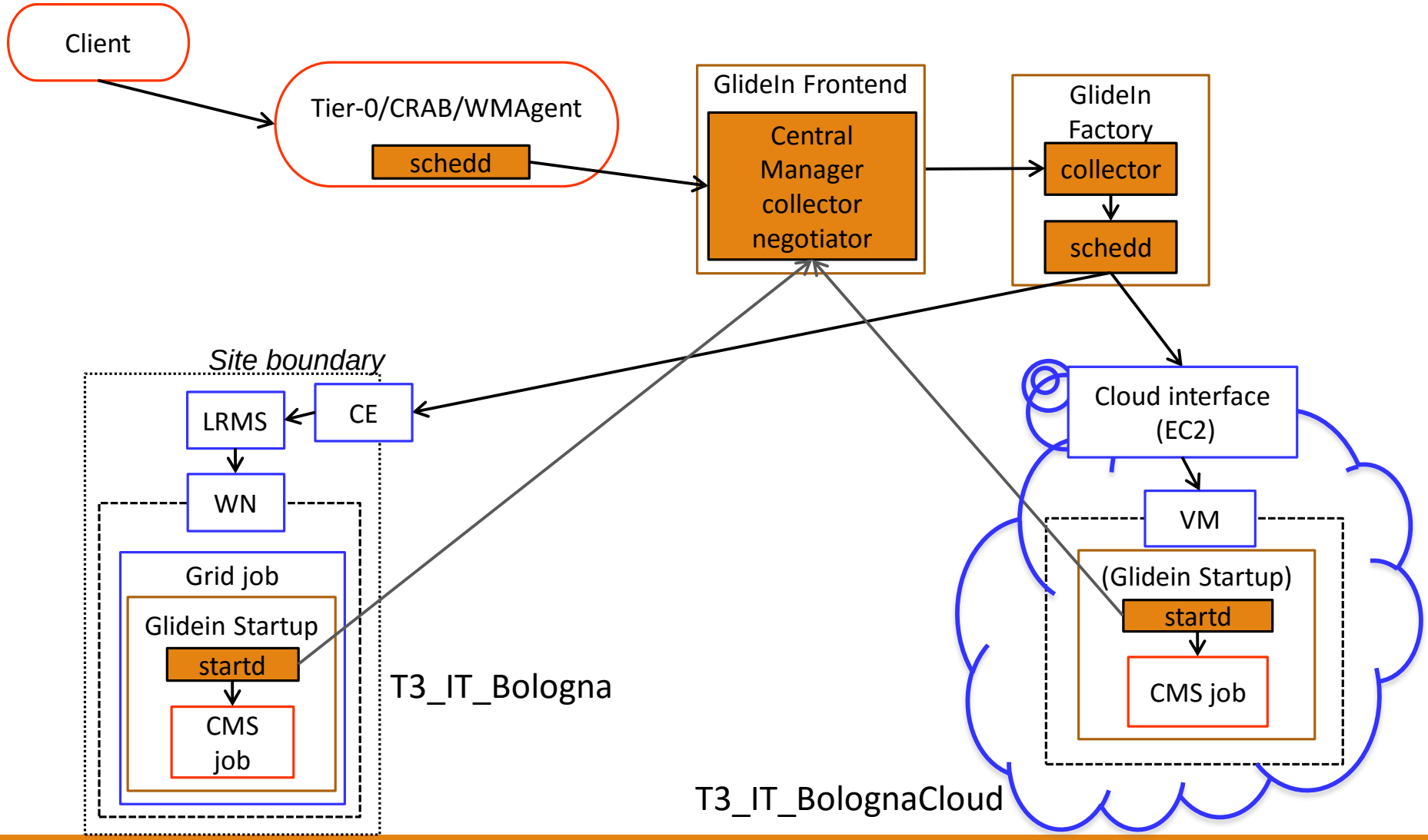
The first step:

- Re-exercising the CRAB v2 + Custom GlideIn-WMS + Openstack Havana chain
  - CRAB v2 is a pure command line tool, in our case tuned to work with custom GlideIn-WMS

The new CMS wide test:

- Using CRAB v3 + CERN Integration Test Bed (ITB) GlideIn-WMS + Openstack Juno
  - CRAB3 is a client-server infrastructure: needed to use production tools!
  - ITB GlideIn-WMS is a central infrastructure for pre-production tests
  - This exercise was one of the first main test for the newly Openstack Juno CNAF installation

Tier3 as a Service at CNAF Openstack



CNAF Openstack as a CMS site

We used Bologna Tier3 setup in order to create a new site “as a service”

- Custom, “lightweight” images
- No need to create a new site
  - Only GlideIn-WMS need to be aware: high flexibility in instantiating sites on-demand

Successfully used crab3 + ITB GlideIn-WMS+ Openstack Juno

- Using standard CMS workflow for final Analysis Objects creation
  - VM flavor: Quadcore, 8GB RAM VMs
- Also tested more memory+CPU intensive tasks (reconstruction jobs for upgrade scenario)
  - VM flavor: Quadcore, 12GB RAM VMs
- Submitted 4 task made of 200 jobs each
  - Using a production infrastructure, our resource slot was limited
- No failures observed
- Good job efficiency (peak at about 98%)
  - Efficiency = (CPU Time)/(Wall Clock time)

Next Steps and Plans for the Future

We used custom lightweight images reproducing Bologna Tier-3 existing configuration

- Next step could profit of the µCernVM images
  - Generic ISO image 12 MB sized with OS entirely on CVMFS
  - Faster to instantiate, easier to keep up-to-date
- Already used by other LHC experiments and recently adapted also by CMS

We tested the dynamic extension for the CMS only case

- Bologna Tier3 is a multi-VO environment
- We are working to extend the usability to ATLAS

Using µCERNVM + Parrot under Docker, the requests on the host system can be further reduced:

- No kernel privileged access

The Bologna Tier3 is a local users facility

- It can be costly in terms of maintenance and manpower
- We can profit of the exercise to turn the Site into a purely Cloud site if the others customer VO's are able to adapt
- We can opportunistically expand over non-used CNAF Tier1 resources

Summary

Using Bologna Tier3 setup we realized two prototypes for the extension of a Grid site into Cloud resources

- We tested the dynamic extension of a CMS site over external resources with production tools
  - Bologna Tier3 over CNAF Openstack
- We tested the access of a purely Cloud instantiated CMS Site with production tools

The proof of concept realized are already serving bigger scale projects

- Extending the CNAF Tier1 over commercial resources (ARUBA, Italian Cloud provider)
  - see "*Elastic CNAF DataCenter extension via opportunistic resources*" presented by Dr. Stefano DAL PRA on 18 Mar 2016 at 10:00
- Looking forward the opportunity to define the Tier3 as a pure Cloud instanciated site inside the CNAF Tier1 Openstack infrastructure to reduce mantainance costs

Credits

C. Aiftimiei[3,4], D. Bonacorsi[1,2], P. Calligola[1] , V. Ciaschini[3] ,

G. Codispoti[1,2], A. Costantini[3], S. Dal Pra[3], D. DeGirolamo[3],

R. Di Maria[1,2,a], C. Grandi[1], D. Michelotto[3], M. Panella[3], G. Peco[1],

V. Sapunenko[3], M. Sgaravatto[5], S. Taneja[3], G. Zizzi[3]

[1]INFN Bologna, Bologna, Italy
[2]Physics and Astronomy, University of Bologna, Bologna, Italy
[3]CNAF, Bologna, Italy
[4]IFIN-HH, Magurele, Romania
[5]INFN, Padova, Italy
[5]now at Imperial College, London, UK