Extending the farm on external sites: the INFN Tier-1 experience

Luca dell'Agnello

INFN-CNAF

CHEP 2016

S. Francisco, October 10 2016

INFN

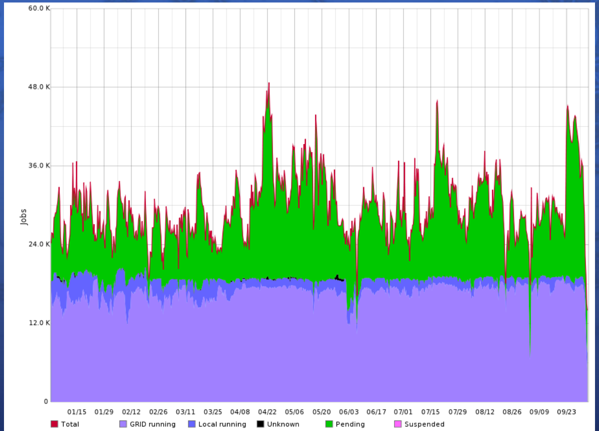
- **National Institute for Nuclear Physics** (INFN) is funded by Italian government
- Main mission is the research and the study of elementary particles and physics laws of the Universe
- Composed by several units
  - ~ 20 units dislocated at the main Italian University Physics Departments
  - 4 Laboratories
  - 3 National Centers dedicated to specific tasks
- CNAF is a National Center dedicated to computing applications

The Tier-1 at INFN-CNAF

- WLCG Grid site started as computing center for LHC experiments (ATLAS, CMS, LHCb, ALICE)
  - Nowadays provides services and resources to ~30 other scientific collaborations
- ~1.000 WNs , ~21.500 computing slots, ~220 kHS06
  - LSF as current Batch System, HTCondor migration foreseen
  - Also small (~33 TFlops) HPC cluster available with IBA
- 22 PB SAN disk (GPFS), 43 PB on tape (TSM) integrated as an HSM
  - Also supporting LTDP for CDF experiment
- Dedicated network channel (60 Gb/s) for LHC OPN + LHC ONE
  - 20 Gb/s reserved for LHC ONE
  - Upgrade to 100 Gb/s connection in 2017

Computing resource usage

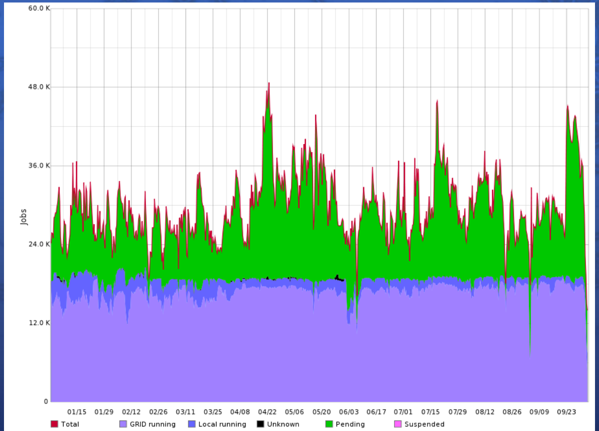
- Computing resources always completely used at CNAF with a large amount of waiting jobs (~50% of the running jobs)



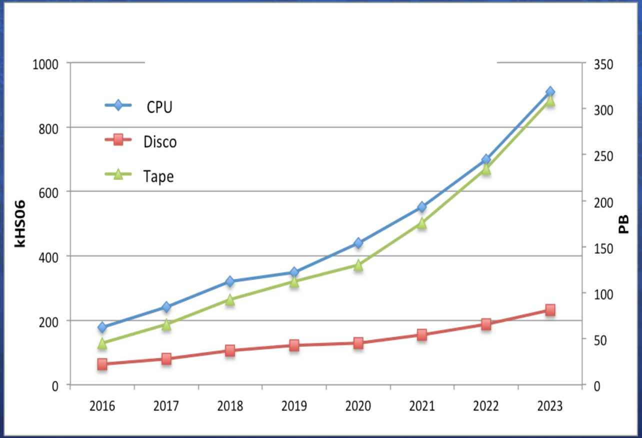
INFN Tier-1 farm usage in 2016

Computing resource usage

- Computing resources always completely used at CNAF with a large amount of waiting jobs (~50% of the running jobs)
- Expected huge resource increase in the next years mostly coming from LHC experiments



INFN Tier-1 farm usage in 2016



INFN Tier-1 resource increase

Toward a (semi-)elastic Data Center?

- Planning to upgrade the Data Center to host resources at least until the end of LHC Run 3 (2023)
- A complementary solution could be (dynamic) extension on remote farms
- Cloud bursting on commercial provider
  - Tests of opportunistic computing on Cloud providers
- Static allocation of remote resources
  - First production use case: part of 2016 pledged resources for WLCG experiments at CNAF are in Bari-ReCaS
- Also participating to HNSCicloud EU PCP project

Toward a (semi-)elastic Data Center?

- Planning to upgrade the Data Center to host resources at least until the end of LHC Run 3 (2023)
- A complementary solution could be (dynamic) extension on remote farms
- Cloud bursting on commercial provider
  - Tests of opportunistic computing on Cloud providers
- Static allocation of remote resources
  - First production use case: part of 2016 pledged resources for WLCG experiments at CNAF are in Bari-ReCaS
- Also participating to HNSCicloud EU PCP project

Opportunistic computing on Aruba (1)

- One of the main Italian commercial resource providers
  - Web, host, mail, cloud …
  - Main datacenter in Arezzo (near Florence, ~140 km from CNAF)
- Goal
  - Transparently use these external resources “as if they were” in the local cluster, and have LSF dispatching jobs there when available
- Small scale test
  - 10x8 cores VM (160 GHz) managed by VMWare
- Use of idle CPU cycles
  - When a customer requires a resource used by us, the frequency of CPU of “our” VMs is lowered down to a few MHz **(not destroyed!)**
- Tied to CMS-only specifications
  - No storage on site: remote data access via Xrootd
  - Use of GPN (no dedicated NREN infrastructure)

Opportunistic computing on Aruba (2)

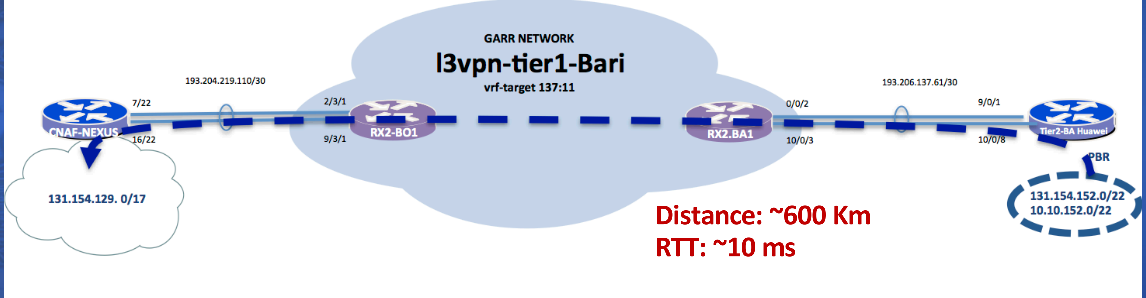
- The remote VMs run the very same jobs delivered to CNAF by GlideinWMS (CMS)
  - Ad hoc configuration at GlideIN could specialize delivery for these resources
- Job efficiency (CPT/WCT) depends on type of job
  - very good for certain type of jobs (MC)
  - Low on average (0.49 vs. 0.80)
- Guaranteed network bandwidth and/or cache system could improve these figures
  - We foresee additional tests with a cache system

Remote extension to Bari ReCaS

- 48 WNs (~26 kHS06) and ~330 TB of disk allocated to Tier-1 farm for WLCG experiments in Bari-ReCaS data center
  - Bari-ReCaS hosts a Tier-2 for CMS and Alice
- ~10% of CNAF total resources, ~13% of resources pledged to WLCG experiments
- Goal: direct and transparent access from CNAF
- Similar to CERN/Wigner extension

BARI – CNAF connectivity

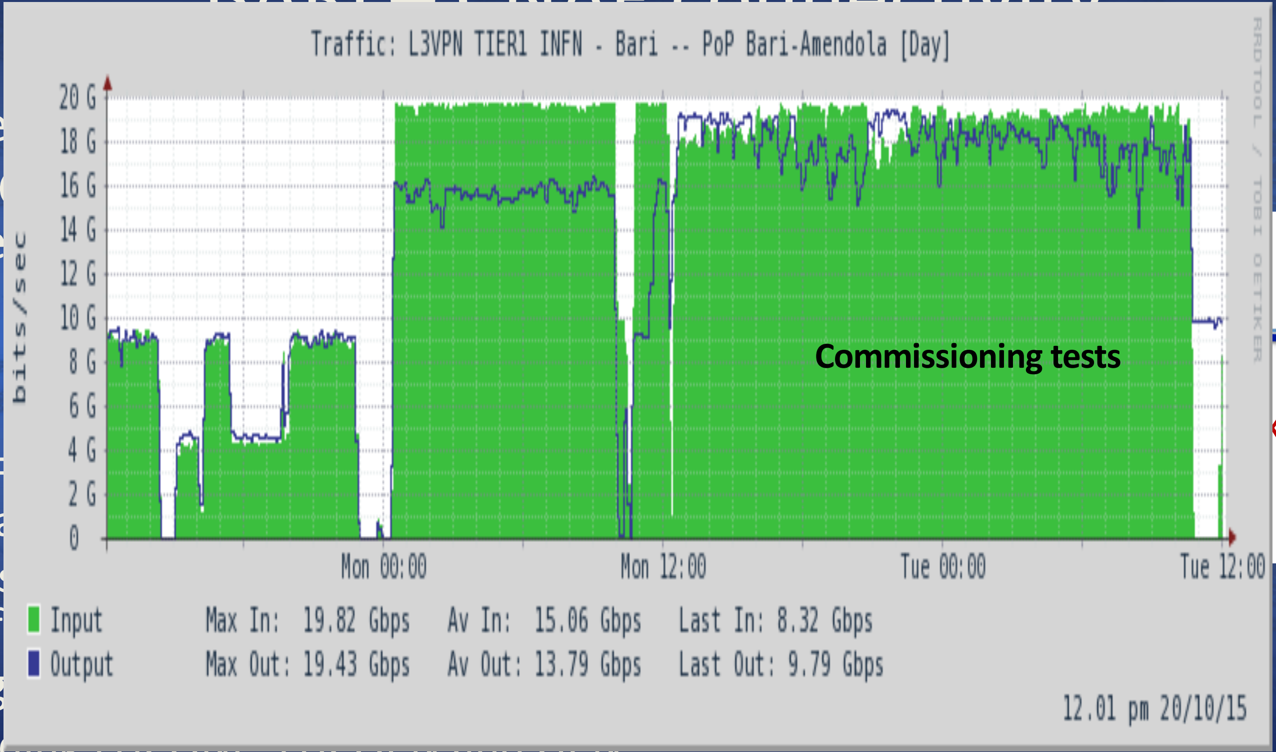
- Requirement: link CNAF-ReCaS at least 10 Gbit/s for 1000 cores
- Bari-ReCaS WNs to be considered as on CNAF LAN
- VPN L3 configured
  - 2x10 Gb/s, MTU=9000
  - CNAF /22 subnet allocated to BARI WNs
  - Also service networks (i.e. for WN management) accessible
- Routing through CNAF also for BARI WN
  - Including LHCONE, LHCOPN and GPN

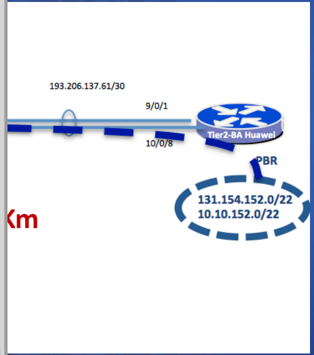


Layout of CNAF-BARI VPN

BARI – CNAF connectivity

- Require
- Bari-Re
  conside
  LAN
- VPN L3
  - 2x10
  - CNAF
    alloca
  - Also
    (i.e. f
- Routing
  - Including LHCONE, LHCOPN and GPN





Farm extension setup

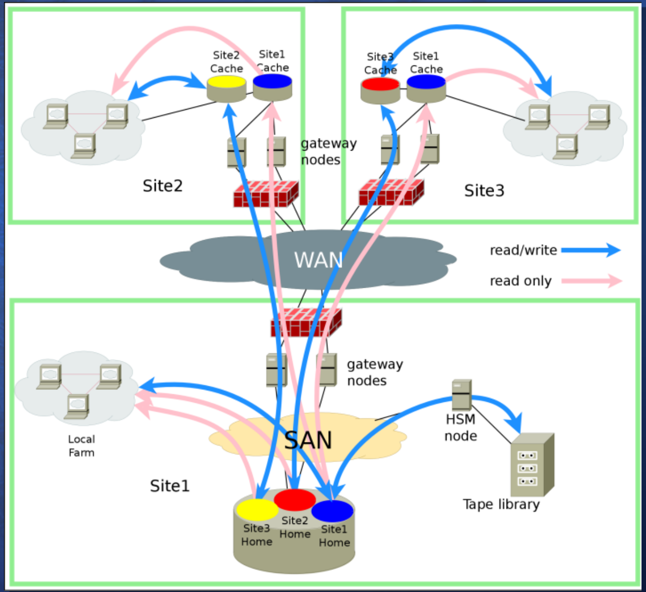
- Goal: transparent access from CNAF farm
  – Should be indistinguishable for users
- CNAF LSF master dispatches jobs also to Bari-ReCaS WNs
  – BARI WNs considered as local resources
- CEs (grid entry points for farm) at CNAF
- Auxiliary services installed in Bari-ReCaS
  - CVMFS Squid servers (for software distribution)
  - Frontier Squid servers (used by ATLAS and CMS for condition db)

Data Access

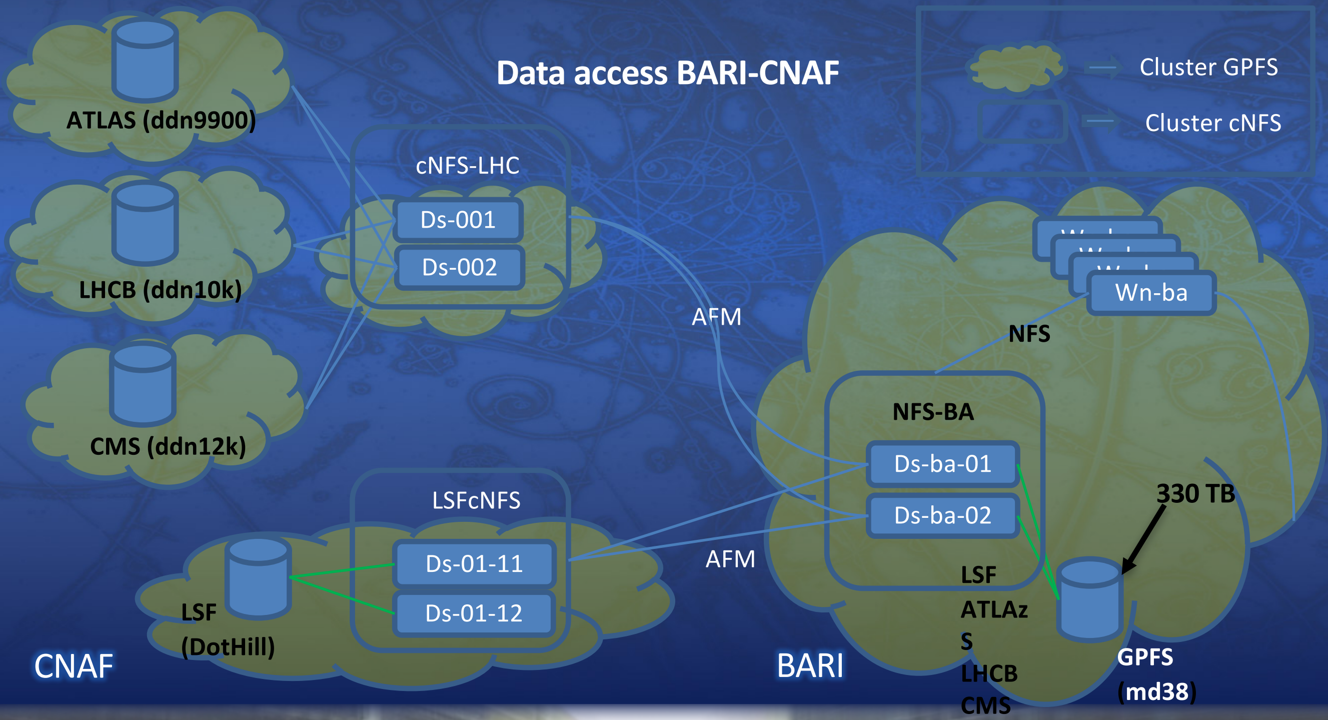
- Data at CNAF are organized in GPFS file-systems
  - Local access through Posix, Gridftp, Xrootd, and http
  - Remote fs mount from CNAF unfeasible (x100 RTT)
- Jobs expect to access data the same way as at CNAF
  - Not all experiments able to use a fallback protocol
- Local (@ Bari-ReCaS) Posix cache for data needed
  - Alice uses Xrootd only (no cache needed)
- Cache implemented with AFM (GPFS native feature)

Remote data access via GPFS AFM

- GPFS AFM
  - A cache providing geographic replica of a file system
  - Manages RW access to cache
- Two sides
  - Home - where the information lives
  - Cache
  - Data written to the cache is copied back to home as quickly as possible
  - Data is copied to the cache when requested
- AFM configured as RO for Bari-ReCaS
- Several tunings and reconfigurations required!
- In any case decided to avoid submission of high throughput jobs in Bari (possible for Atlas)

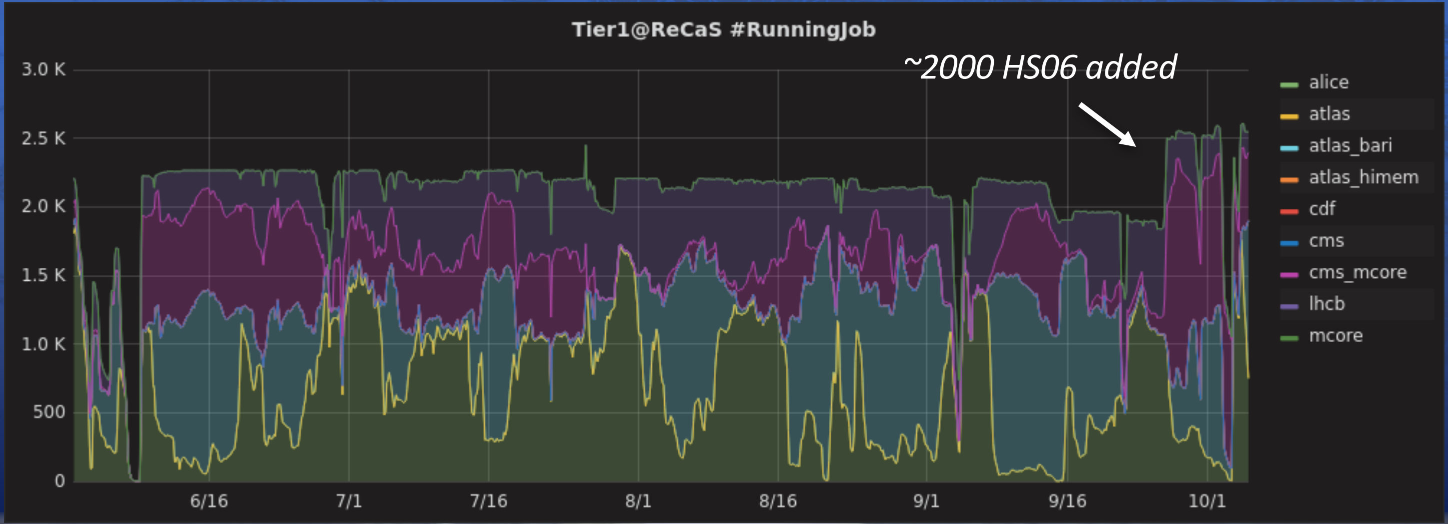


AFM cache layout



Results: Bari-ReCaS (1)

- Several issues has been addressed
  - Mainly cache reconfiguration and tuning
- At steady state since June 2016
  - ~550 k production jobs (~8% CNAF)



Results: Bari-ReCaS (2)

- Job efficiency @ Bari-ReCaS equivalent (or even better!)
  - In general jobs at CNAF use WNs shared among several VOs
    - during the Summer one of these (non WLCG), with misconfigured jobs, has affected efficiency of all the other VOs
  - Atlas submits only low I/O jobs on Bari-ReCaS
  - Alice uses only XrootD, no cache
    - "intense" WAN usage also from CNAF jobs
- Network was not an issue
  - We could work w/o cache for data using Xrootd
    - But probably we would need more than 20 Gb/s
  - Anyway cache needed for some experiments

Experiment	NJobs	Efficiency
Alice	105109	0,87
Atlas	366999	0,94
CMS	34626	0,80
LHCb	39310	0,92

Job efficiency @BARI

Experiment	NJobs	Efficiency
Alice	536361	0,86
Atlas	4956628	0,87
CMS	326891	0,76
LHCb	263376	0,88

Job efficiency @CNAF

Efficiency = CPT/WCT

Conclusions

- INFN Tier-1 is fully addressing computing requirements from experiments in which INFN is involved
- We are planning to upgrade the Data Center to host resources at least until the end of LHC Run 3 (2023)….
- … but testing (elastic) extension of our Data center is important
  - Infrastructure could not scale indefinitely
  - External extensions could address other use cases such as temporary peak requests in a cheaper way than with flat deployment

Conclusions

- INFN Tier-1 is fully addressing computing requirements from experiments in which INFN is involved
- We are planning to upgrade the Data Center to host resources at least until the end of LHC Run 3 (2023)….
- ... but testing (elastic) extension of our Data center is important
  - Infrastructure could not scale indefinitely
  - External extensions could address other use cases such as temporary peak requests in a cheaper way than with flat deployment
- The first results are more than promising …
- … but for a more general solution a cache system should be developed

Conclusions

- INFN Tier-1 is fully addressing computing requirements from experiments in which INFN is involved
- We are planning to upgrade the Data Center to host resources at least until the end of LHC Run 3 (2023)….
- ... but testing (elastic) extension of our Data center is important
  - Infrastructure could not scale indefinitely
  - External extensions could address other use cases such as temporary peak requests in a cheaper way than with flat deployment
- The first results are more than promising …
- … but for a general solution a cache system should be developed

Backup slides

HNSciCloud

- EU Project (call ICT 8a di H2020)
  - Approved (September 2015)
- “Pre-Commercial Procurement” to lease IaaS cloud services
  - 2/3 of funding from EU
- Goal: realize a prototype of “hybrid cloud” with commercial providers covering ~5% of all WLCG resources
- Involved CERN, most of EU Tier-1s, DESY, EGI, EMBL
- Still in the phase of writing the technical specifications for the tender.
  - Non negligible administrative effort ☹

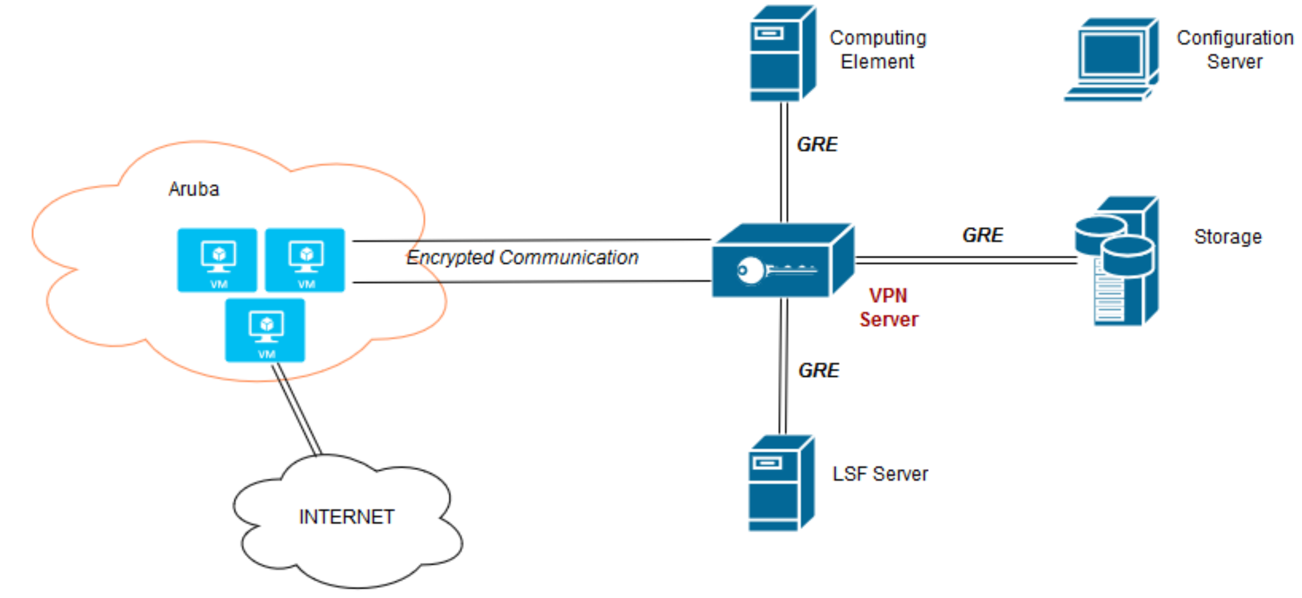
Dynfarm concepts

- The VM at boot connects to a OpenVPN based service at CNAF
  - It authenticates the connection (RSA)
  - Delivers parameters to setup a tunnel with (only) the required services at CNAF (LSF, CEs, Argus)
  - Routes are defined on each server to the private IPs of the VMs (GRE Tunnels)
  - Other traffic flows through general network

Dynfarm deployment

- VPN Server side, two RPMs:
  - dynfarm-server, dynfarm-client-server
    - In the VPN server at CNAF. First install creates one dynfarm_cred.rpm which must be present in the VMs
- VM side, two RPMs:
  - dynfarm_client, dynfarm_cred (contains keys to initiate connection and get authenticated by the VPN Server)
- Management: remote_control <cmd> <args>

Dynfarm workflow



Auxiliary services

- Cache system for other services to offload network link and speed-up response
  - CVMFS Squid servers (for software distribution)
  - Frontier Squid servers (used by ATLAS and CMS for condition db)
- Dedicated DNS servers at BARI
  - Offer different view to WNs respect to CNAF for application specific servers (e.g. Frontier squids)

[root@ba-3-8-01 ~]# host squid-lhc-01
squid-lhc-01.cr.cnaf.infn.it has address 131.154.152.38

[root@wn-206-08-21-03-a ~]# host squid-lhc-01
squid-lhc-01.cr.cnaf.infn.it has address 131.154.128.23

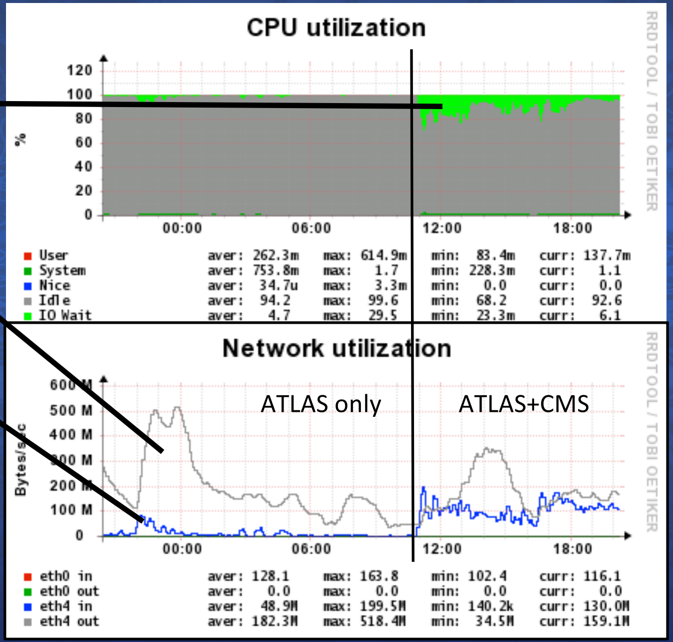
AFM deployment

- Cache storage GPFS/AFM
  - 2 server, 10 Gbit
  - 120 TB → 330 TB (Atlas, CMS, LHCb) as cache for data
- Alice experiment does not need cache
  - Remote Xrootd access to data in any case
- CMS able to fallback to Xrootd protocol in case of posix access failure
- (Small) AFM cache also for LSF shared fs
  - Decoupled from the cache for data to avoid interferences due to I/O intensive jobs

ba-3-x-y: Feb 8 22:56:51 ba-3-9-18 kernel: nfs: server nfs-ba.cr.cnaf.infn.it not responding, timed out

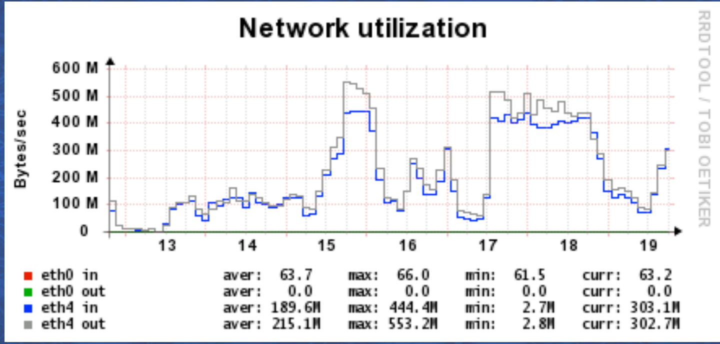
Cache issues

- Local cache access critical
  - Potential bottleneck
- First "incarnation" of cache
  - 120 TB of net disk space
  - Max 1 GB/s r or w
  - Concurrent r/w degrade performances to 100 MB/s
  - 20 TB-N/experiment
    - CMS fills space in 12h
    - Atlas, LHCb use only 10% of the space
- Very low efficiency for CMS jobs
  - Emergency solution: disable cache access
  - Xrootd fallback



Cache tuning (1)

- Enlargement of data cache (from 120 to 330 TB-N)
  - ~100 TB-N per experiment
  - > 50 TB-N CMS can easily accommodate datasets to be reprocessed
    - Avoid pass-through effect

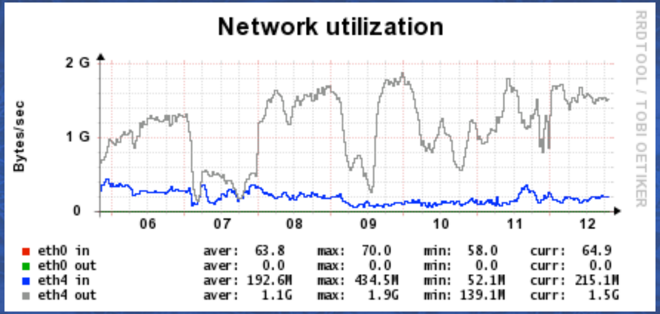


Cache throughput

- ... but performance limits still present
  - Increase of number of disks does not help in this case
- Investigation on GPFS/AFM configuration

Cache tuning (2)

- GPFS optimization normally based on supposition that 1 RAIDset =1 LU and is done on LU level
  - In our case 1 RAIDset contains 12 LU
  - we needed to lower number of processes (threads) working with each LU by factor of 10.
- Increase of fs block size from 1MB to 4MB has reduced I/O operations to get same throughput (and also reduced concurrent I/O on a specific RAIDset)



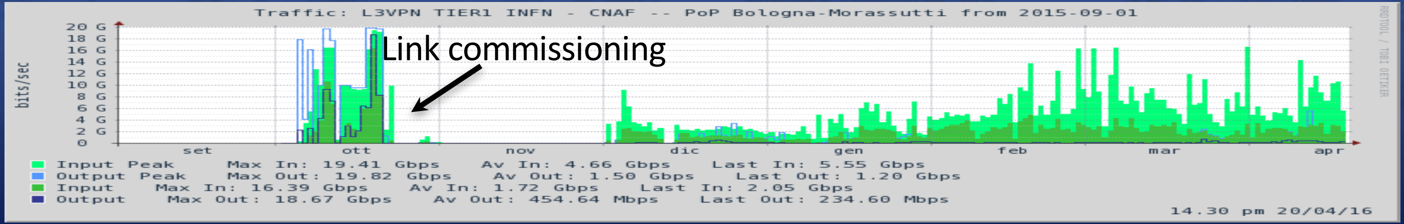
Cache throughput

Other issues

- Too high # of cores
  - An hw problem on a single WN affects up to 64 jobs
  - Mean job duration time: 3 days
  - Can cost 100 days of wasted CPU time
- I/O load on WN local disk
  - Due to large number of independent processes this can cause latency to access the local disks and hence be a bottleneck
- Suspect occasional problems with the power supplies
  - Too much power needed when WN fully loaded? Still unclear…

Preliminary conclusions

- Several issues has been addressed
  - Not at steady state yet
  - We need to gain more experience to understand limits
- Network was not an issue ☺
  - We could work w/o cache for data using Xrootd
    - But (probably) we would need more than 20 Gb/s
  - Anyway cache needed for some experiments



Preliminary conclusions

- Several issues has been addressed
  - Not at steady state yet
  - We need to gain more experience to understand limits
- Network was not an issue ☺
  - We could work w/o cache for data using Xrootd
    - But probably we would need more than 20 Gb/s
  - Anyway cache needed for some experiments
- Is this model convenient?
  - Not clear....
    - Need to quantify costs due to efficienty loss, network etc...

The use-case

- Agreement CNAF - Aruba
  - Aruba has provided a small amount of Virtual resources (CPU cycles, RAM, DISK) out of a pool assigned to real customers
    - 10x8 cores VM (160 GHz) managed by VMWare
  - When a customer requires a resource used by us, the frequency of CPU of “our” VMs is lowered down to a few MHz **(not destroyed!)**
- Goal
  - Transparently join these external resources “as if they were” in the local cluster, and have LSF dispatching jobs there when available
  - Tied to CMS-only specifications
    - No data caching (hence Xrootd fallback)

Some configuration issues

- Remote Virtual WNs need read-only access to the cluster shared fs (/usr/share/lsf)
  - Use of GPFS/AFM cache as in Bari
- VMs have private IP, are behind NAT & FW, outbound connectivity only, but have to be reachable by LSF
  - Developed an ad hoc service at CNAF (**dynfarm**) to provide integration between LSF and virtualized computing resources
- LSF needs host resolution (IP ↔ hostname) but no DNS available for such hosts
  - Manually fixed in /etc/hosts
- Use of GPN (no dedicated link)
  - No problem for a small scale test-bed

“Comparative” Results

Queue	Nodetype	Njobs	Avg_eff	Max_eff	Avg_wct	Avg_cpt
Cms_mc	AR	2984	0,602	0,912	199,805	130,482
Alice	T1	98451	0,848	0,953	16,433	13,942
Atlas_sc	T1	1211890	0,922	0,972	1,247	1,153
Cms_mc	T1	41412	0,707	0,926	117,296	93,203
Lhcb	T1	102008	0,960	0,985	23,593	22,631
Atlas_mc	T1	38157	0,803	0,988	19,289	18,239
Alice	BA	25492	0,725	0,966	14,446	10,592
Atlas	BA	15263	0,738	,979	1,439	1,077
Cms_mcore	BA	2261	0,444	0,805	146,952	69,735
Lhcb	BA	13873	0,916	0,967	12,998	11,013
Mcore	BA	20268	0,685	0,878	24,378	15,658

"Comparative" Results

Queue	Nodetype	Njobs	Avg_eff	Max_eff	Avg_wct	Avg_cpt
Cms_mc	AR	2984	0,602	[illegible]	199,805	130,482
Alice	T1	98451	[illegible]	[illegible]	16,433	13,942
Atlas_sc	T1	1211890	[illegible]	[illegible]	1,247	1,153
Cms_mc	T1	[illegible]	[illegible]	[illegible]	117,296	93,203
Lhcb	T1	[illegible]	[illegible]	,985	23,593	22,631
Atlas_mc	T1	[illegible]	[illegible]	0,988	19,289	18,239
Alice	BA	[illegible]	25	0,966	14,446	10,592
Atlas	BA	[illegible]	0,738	,979	1,439	1,077
Cms_mcore	BA	[illegible]	0,444	0,805	146,952	69,735
Lhcb	BA	13873	0,916	0,967	12,998	11,013
Mcore	BA	20268	0,685	0,878	24,378	15,658

WARNING!
Unfair comparison: not homogeneous sets of observed jobs

Remote extension to Bari ReCaS

- 48 WNs (~26 kHS06) and ~330 TB of disk allocated to Tier-1 farm for WLCG experiments
  - 64 cores per mb (546 HS06/WN)
  - 1 core/1 slot, 4GB/slot, 8,53 HS06/slot
- ~10% of CNAF total resources, ~13% of resources pledged to WLCG experiments
- Goal: direct and transparent access from CNAF
- Similar to CERN/Wigner extension

The Bari ReCaS Data Center

- Common effort of INFN and Università degli Studi di Bari “Aldo Moro”
- Active from July 2015
- 128 WNs , 8192 (+4000 the old data center) computing slots, ~100k HS06
  - Small HPC Cluster (800 cores) with IBA
- 3.6 PB SAN of disk space, 2.5 PB of space on tape library
- INFN quota (~25 kHS06, 1.1 PB of disk) allocated to CMS and Alice Tier-2