

Summary of Volunteer Computing pre-GDB 11/11-2014

Nils Høimyr, CERN IT/PES with input and notes from Michel Jouvin and other participants

Meeting agenda & facts

- Agenda: <https://indico.cern.ch/event/272793/>
- Summary notes:
<https://twiki.cern.ch/LCG/GDBMeetingNotes20141111>
- Attendance:
- ~ 30 local (less before lunch)
- ~ 10 remote
- ATLAS and LHCb well represented
- CMS via IT/SDC
- No activity reported from ALICE

Objective of meeting

- Follow-up of GDB presentation of Volunteer Computing in January
- Share experience between LHC experiments on volunteer computing
 - BOINC middleware and volunteer computing
 - Also Xtremweb at LAL
 - Learn from LHC@home experience with BOINC
- Discuss the technology, experience and outreach aspects with the aim to move towards a strategy for volunteer computing within WLCG

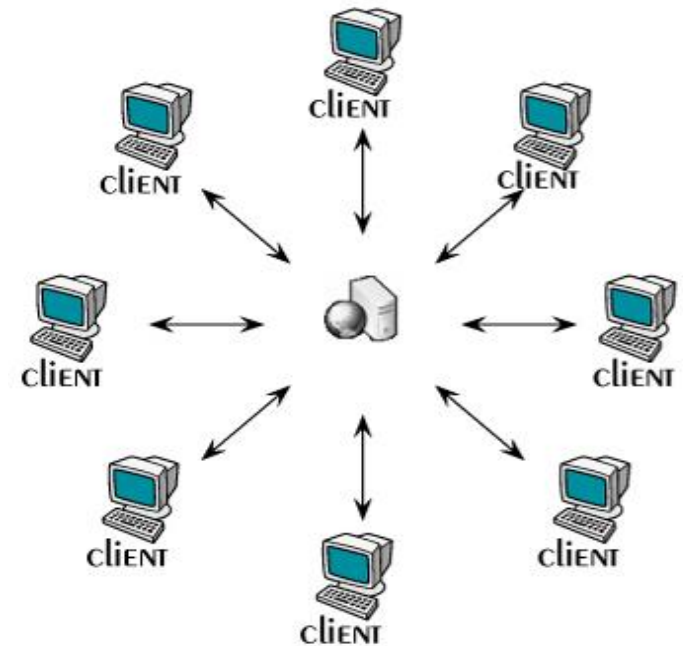
Why Volunteer Computing?

Target	Deployment	Benefit
Volunteers	Uncoordinated, opportunistic	<ul style="list-style-type: none">• Get additional, “free” compute cycles• Engage with communities outside HEP: outreach and publicity for HEP and science
Institute desktops	Coordinated, opportunistic	<ul style="list-style-type: none">• Get additional, “free” compute cycles
Small to midsize server farms	Coordinated, pledged	<ul style="list-style-type: none">• Easier to deploy than complete Grid middleware

BOINC

“Berkeley Open Infrastructure for Network Computing”

- Software platform for distributed computing using volunteered computer resources
- Client – server architecture
- Free and open source
- Used for
 - SETI@home
 - Climateprediction.net
 - Einstein@home
 - LHC@home
 - ...



<http://boinc.berkeley.edu>

BOINC Compute Power

Project	Average power
SETI@home	653 TFlops
Einstein@home	637 TFlops
World Community Grid	421 TFlops
LHC@home -classic	31 TFlops
Virtual LHC@home	2.6 TFlops

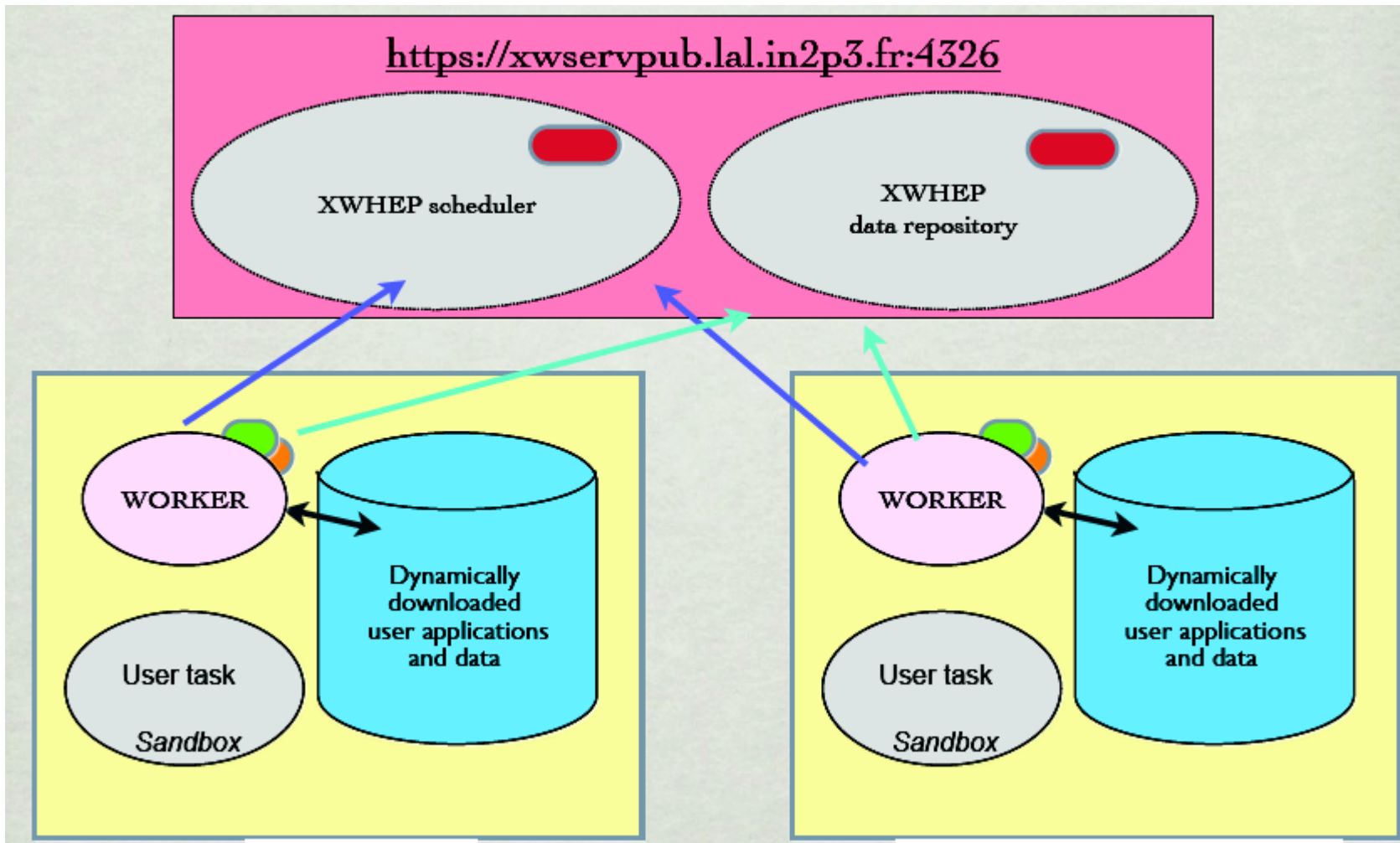
According to BOINCstats.com 10.11.2014

BOINC service in CERN IT

- Can host experiment applications using CernVM on the [Virtual LHC@home](#) BOINC platform
- Also includes [Drupal portal](#) as common entry point, where experiments can provide outreach material
- Handles the operation of the BOINC service layer
 - MySQL db on DB on Demand, managed servers with NFS
- BOINC application porting, forum discussions and outreach is the responsibility of the experiments using the service
- For more info, please contact the [BOINC service](#) team at CERN

Xtremweb (O. Lodginsky)

- Framework designed to run computing at opportunistic resources
 - Desktop grids
 - Server clusters
- Now Xtremweb-HEP (fork in 2006)
 - Includes virtualisation support
- Open-ID used for authentication
- Uses SSH tunnels for communication
- Deployed at LAL and some other sites
 - Grid5000
 - Quarnot computing (heating houses with CPU-heat :-)

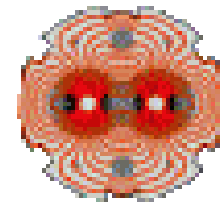


- <http://www.xtremweb.net/>
- www.desktopgridfederation.org

Accelerator beam simulations under LHC@home (M. Giovannozzi)

- User point of view, simulations with SixTrack to improve the LHC beam quality
- Very interesting description of the Dynamic Aperture (DA) and beam dynamics simulations of the LHC and the future HL LHC upgrade
- Challenges related to higher intensities requiring a very high number of high CPU/low I/O simulations with different parameters to achieve optimal beam quality
- SixTrack ported to BOINC under Linux, Windows and MAC, using the classic BOINC approach
 - Over the years since 2004 the team has well run in procedures for compiling on different architectures
- Future challenges includes improving code, I/O buffer, more apps

LHC@home: users



Active users are about 10% of total users

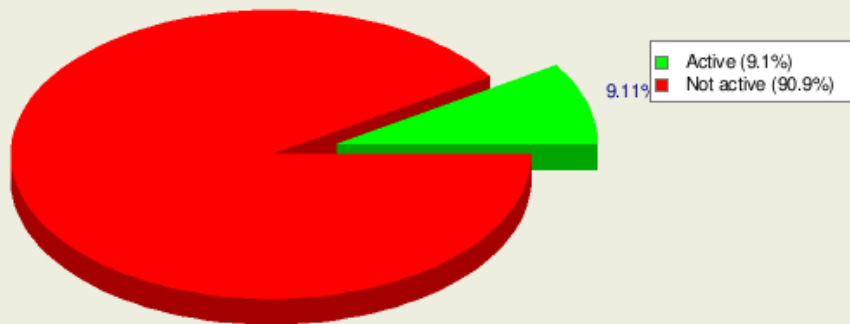
LHC start-up

Total number of users (last months)



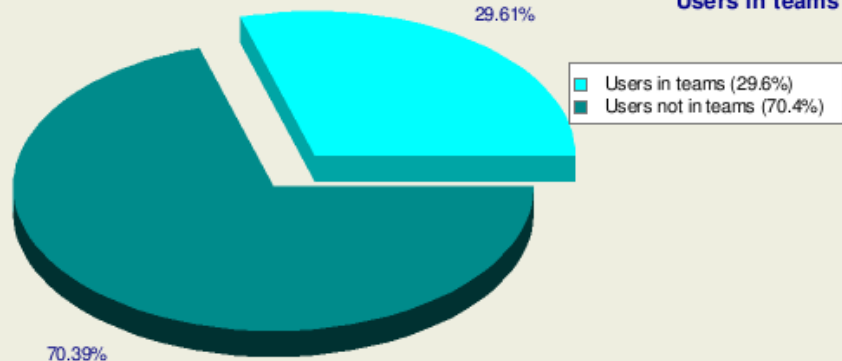
boincstats.com

Active users



boincstats.com

Users in teams



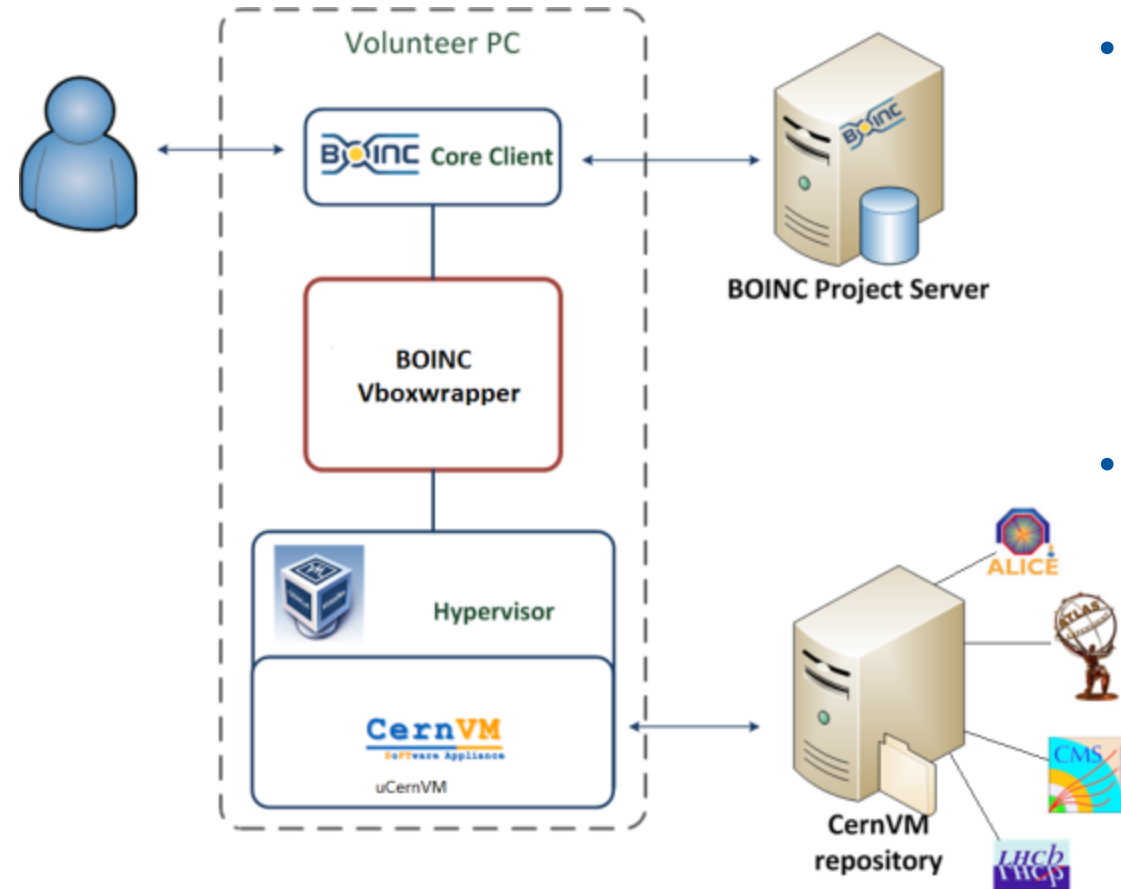
boincstats.com

Ma

BOINC and CernVM – Test4Theory (B. Segal)

- Challenged in 2006 to do High Energy Physics simulations under LHC@home
 - Looking into virtualisation to avoid porting of applications to Windows, Mac etc
- CernVM development in PH/SFT enabled the use of VMs for real HEP software
- Development with many students and short term personell from 2007-2011, with public launch of LHC@home 2.0 Test4Theory
- Design considerations for use of Virtualisation with BOINC
- 1.5 billion MC events simulated for the Theory group
- Project changed name to “Virtual LHC@home” with the view to host applications from experiments in addition to Theory

BOINC and Virtualisation

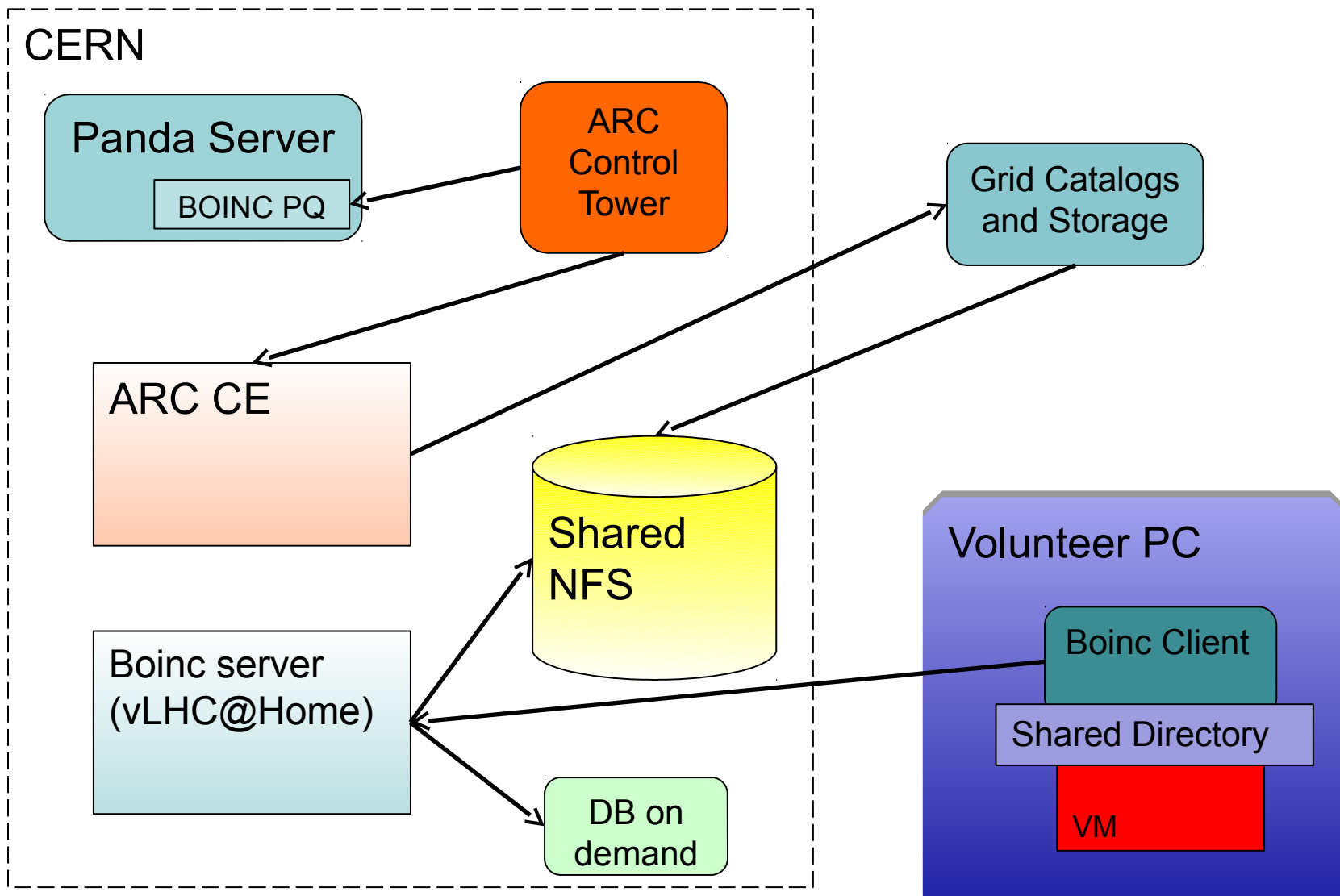


- BOINC distributes VMs to client machines along with a wrapper application
- The BOINC client installation for Windows now includes Virtual Box

ATLAS@Home (D. Cameron)

- Volunteer computing project within ATLAS
 - Focus on non-urgent Monte-Carlo simulations
 - Virtualisation used for ATLAS SW environment CernVM and Cernvmfs (initial VM download 500Mb, job size ~100Mb)
 - ARC-CE and BOINC pragmatic solution, jobs can be submitted from PanDA and is fully integrated in the ATLAS computing environment
 - Proxy certificate stays on ARC-CE server, no credentials sent to volunteers
 - Started in January with a test server integrating BOINC and ARC-CE (Discovered by volunteers :-)
 - Now hosted at CERN, ATLAS in charge of ARC-CE and IT of BOINC server layer

Current ATLAS@Home Setup



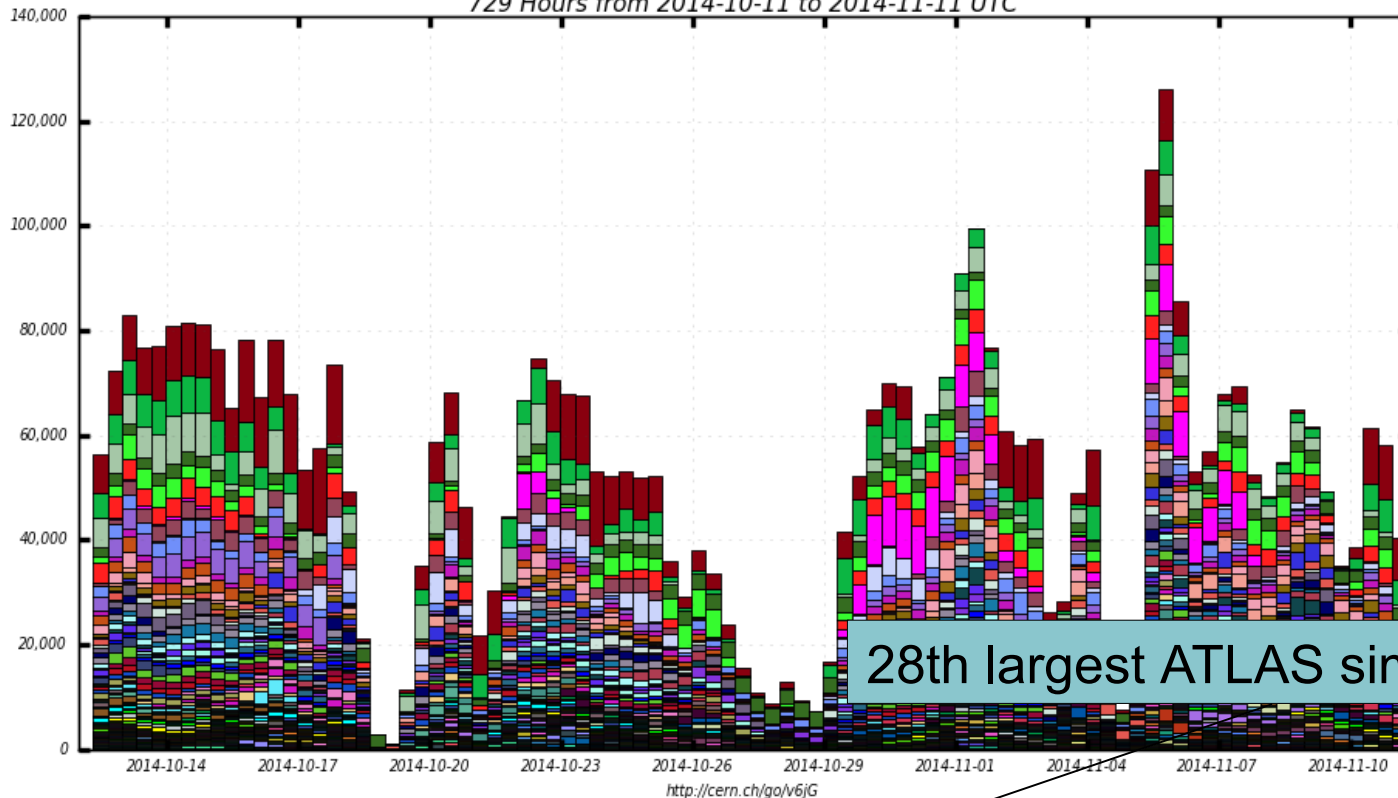
ATLAS – experience so far

- Public outreach effort in progress
 - Web site explaining physics
 - Plans for badges for volunteers etc
 - In spite of no active publicity 2000 users have signed up!
- Interaction with volunteers takes some effort
 - Many volunteers never run a job, as their computer is not powerful enough (64bit required, with 4Gb available for BOINC)
 - Still ~1100 hosts – like a Tier-2 site
 - Forum interaction, users have questions and expect answers
 - Issues with the VM wrapper “vboxwrapper” on some OS
- Current plateau of ~3000 running jobs
 - Plan tuning to scale better before making active publicity

Scale of ATLAS@Home



Slots of Running Jobs
729 Hours from 2014-10-11 to 2014-11-11 UTC



28th largest ATLAS simulation site

- | | | | | |
|-------------------|-----------------------|-------------------------|----------------|-----------------------|
| BNL-ATLAS | MWT2 | RAL-LCG2 | BOINC | WT2 |
| AGLT2 | CERN-P1 | TRIUMF-LCG2 | UNI-FREIBURG | NDGF-T1 |
| IN2P3-CC | TAIWAN-LCG2 | UKI-NORTHGRID-LANCS-HEP | INFN-T1 | UKI-SCOTGRID-GLASGOW |
| INFN-NAPOLI-ATLAS | FZK-LCG2 | SIGNET | LRZ-LMU | UKI-NORTHGRID-MAN-HEP |
| BU ATLAS_TIER2 | INFN-MILANO-ATLASC | UKI-LT2-QMUL | CSCS-LCG2 | INFN-ROMA1 |
| DESY-HH | SWT2_CP8 | UNIBE-LHEP | UTA SWT2 | WUPPERTALPROD |
| CA-SCINET-T2 | TOKYO-LCG2 | PIC | HU ATLAS_TIER2 | IN2P3-LAPP |
| IN2P3-CPPM | PRAGUELCG2 | UKI-SOUTHGRID-OX-HEP | IAAS | NIKHEF-ELPROD |
| SARA-MATRIX | UKI-NORTHGRID-LIV-HEP | LUCILLE | UKI-LT2-BRUNEL | ... plus 70 more |

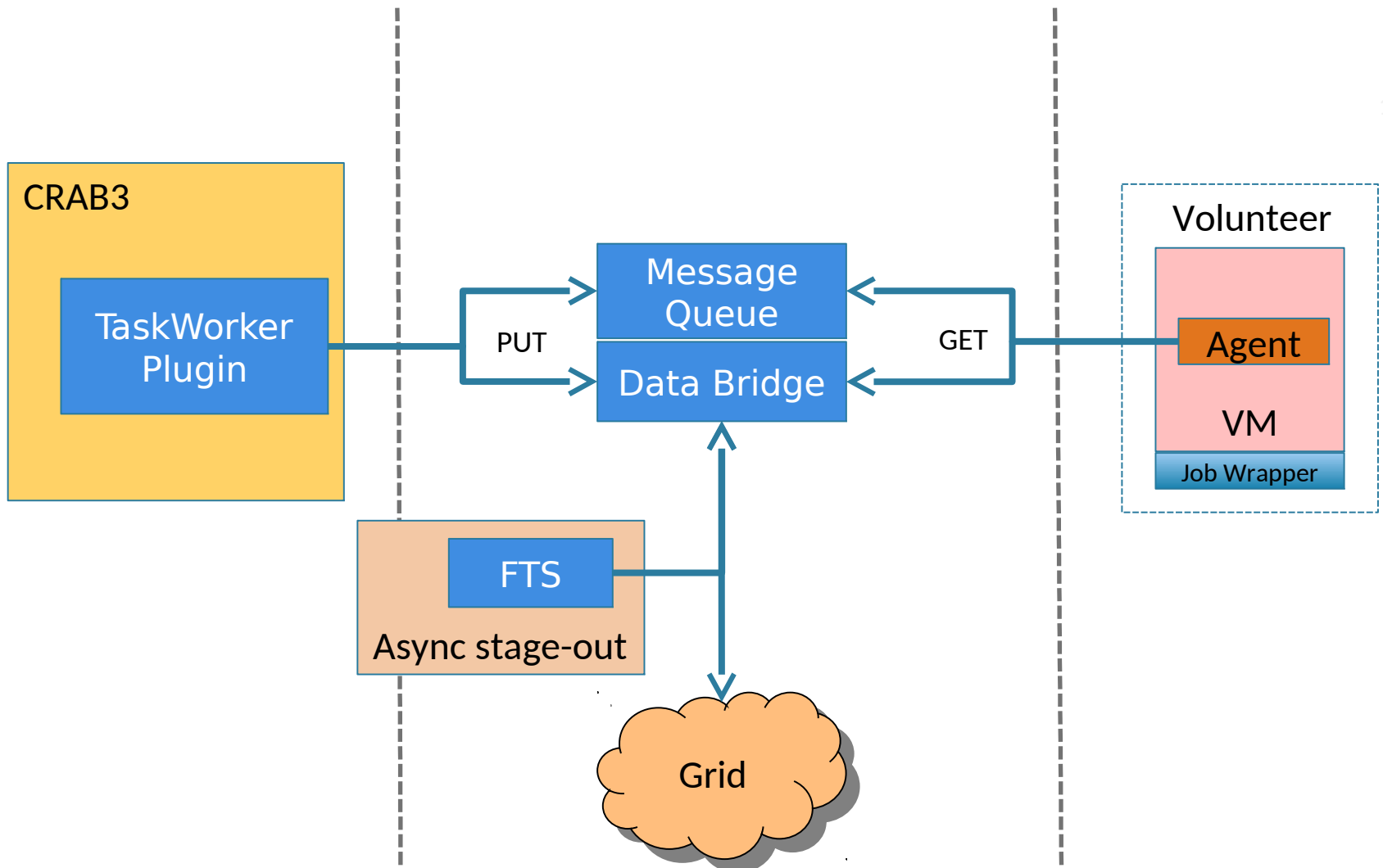
Maximum: 126,056 , Minimum: 0.00 , Average: 52,149 , Current: 40,314

CMS@home (L. Field)

- Summer student project to investigate feasibility and build a prototype
 - Benefit from experience in IT/SDC (CRAB3, Messaging, FTS3)
 - Use CernVM, prior experience from Test4Theory
- New “Data Bridge” component to span authentication domains
 - BOINC user credential and Grid x509 credentials
 - Aim to make this component sharable with other BOINC projects
- CMS test project in operation on BOINC test server, jobs from CRAB

CMS@home Architecture

19



LHCb “Beauty” pilot (F. Stagni)

- Started summer 2013 with a summer student
 - Use CernVM and Virtualisation with BOINC (T4T model)
 - Pause due to lack of manpower
 - Ported to CernVM3 in February, then a pause again
 - DIRAC used directly to manage jobs (Volunteers by invitation only in test phase)
- Issues solved:
 - Focus on short MC jobs
 - Adopt job length to resources
 - These modifications applied everywhere with DIRAC

LHCb “Beauty” challenges and next steps

- CernVMFS
 - Population of local VM cache, bandwidth issues with volunteers
- Resource requirements:
 - Like ATLAS, need 64bit and 4Gb of memory
- Security: Solve AuthN problem
 - Open to advice
 - Data Bridge could be a possible solution

Discussion

- Outreach – share and coordinate outreach efforts
- Volunteer Computing versus other opportunistic resources
- Technology
 - CernVM and CernVMFS common denominator
 - Areas for more work within HEP
 - CernVM image / CernVMfs, pre-cached images etc (Cloud working group)
 - Share experience on common components, e.g. Data Bridge

Opportunistic resources

- Volunteers among the general public
- Desktops in labs
 - Cost of opportunistic capacity versus extra servers in the data centre?
- For some small Tier 2s Virtualisation with BOINC may be an interesting alternative
- If local storage is used at the Tier2, other Cloud technology like e.g. VAC may be more appropriate

Volunteer vs other cloud?

- Volunteer Computing should be seen as an extension of the cloud model of the experiments
 - BOINC used as a vehicle to distribute CernVM images
 - This is also a good outreach channel, currently up to each experiment to exploit
- Desktop grids and some volunteer resources come with minimal additional cost
- Let's use these resources before turning to commercial cloud providers
 - Adopting our code to commercial clouds also has a cost

Conclusions

- We have seen promising developments around Volunteer Computing
- Let's learn from active projects and share!
 - Opportunistic resources and technology:
project-lcg-gdb-clouds-wg@cern.ch
 - BOINC, outreach and LHC@home Issues:
project-lhcathome@cern.ch