Status of Volunteer Computing at CERN

HEPiX spring meeting Oxford 23/3-2015

Nils Høimyr, IT/PES on behalf the BOINC service team

Outline

- Why Volunteer Computing?

- Infrastructure / Middleware – BOINC

- BOINC Compute Power

- Virtualisation with BOINC

- CERN BOINC projects

- BOINC Service at CERN

- Use cases for HEP sites

- Summary

- Questions

Why Volunteer Computing?

| Target	Deployment	Benefit
Volunteers	Uncoordinated, opportunistic	• Get additional, "free" compute cycles • Engage with communities outside HEP: **outreach** and **publicity** for HEP and science
Institute desktops	Coordinated, opportunistic	• Get additional, "free" compute cycles
Small to midsize server farms	Coordinated, pledged	• Easier to deploy than complete Grid middleware

Infrastructure / Middleware

- Most commonly used middleware: BOINC
  - Other choices: XtremWeb, HTCondor, …
  - Other initiatives based on virtualisation and clouds (ref: CernVM web-api talk at recent CernVM workshop)

- LHC@home uses BOINC
  - Focus on that for the rest of this presentation

BOINC

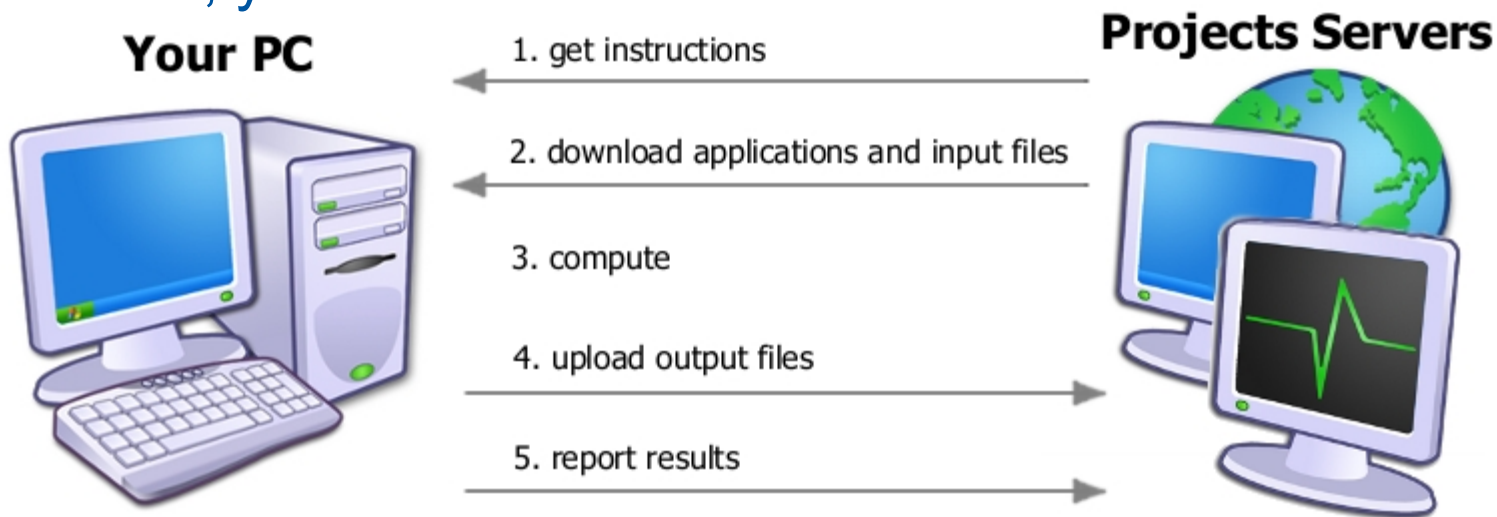## "Berkeley Open Infrastructure for Network Computing"

- 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
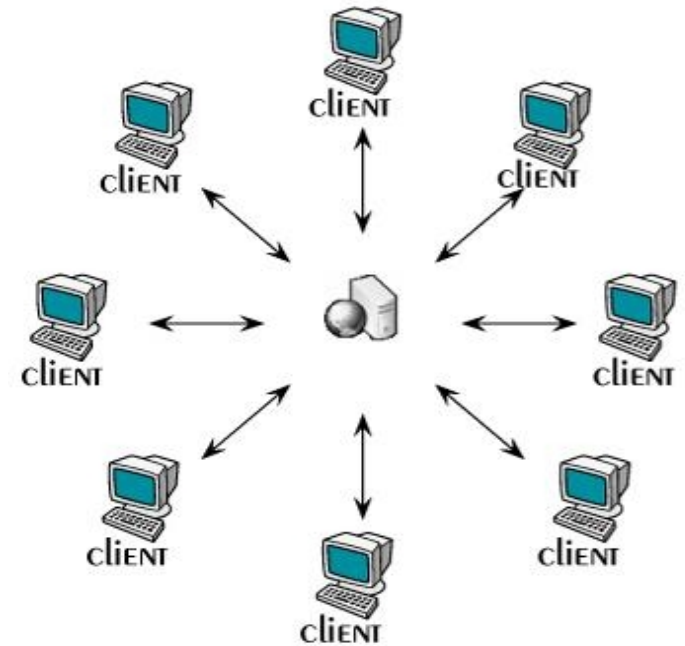
Volunteer Perspective of BOINC

- Download and run BOINC software
- Choose a project
- Enter an email address and password in the BOINC Manager
  - You can also make a silent connection with a key from the BOINC client
- That's it, your are done!

Your PC

1. get instructions
2. download applications and input files
3. compute
4. upload output files
5. report results

Projects Servers

BOINC Compute Power

Project	Average power
Seti@home	695 TFlops
Einstein@home	680 TFlops
World Community Grid	504 TFlops
LHC@home -classic	32 TFlops
Virtual LHC@home	3.4 TFlops

According to BOINCstats.com 4.3.2015

BOINC monitoring dashboard
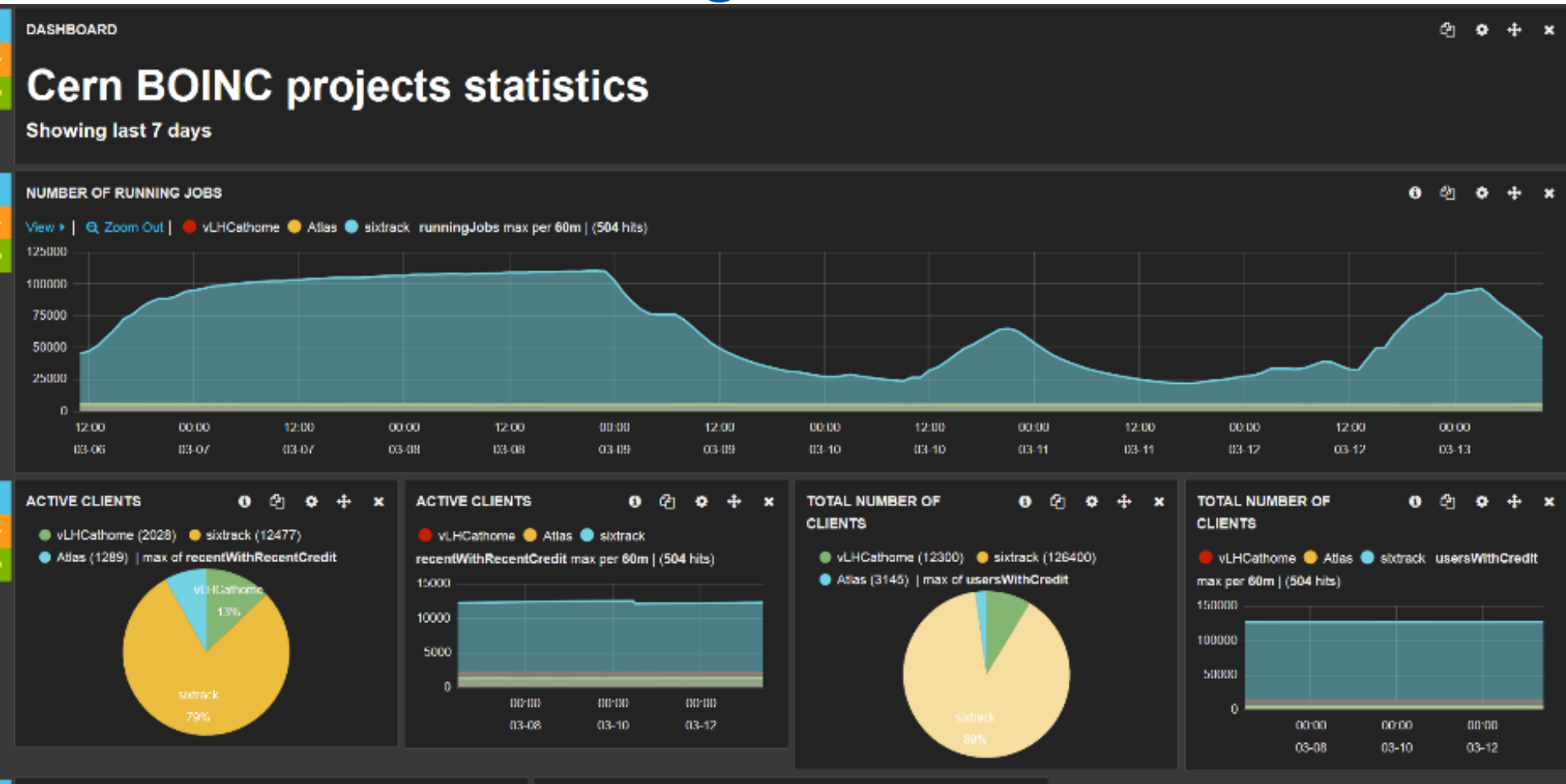
- http://cern.ch/go/9nRz
- See also: http://boincstats.com to compare with other projects

BOINC and Virtualisation - 1

- Pioneered at CERN by Test4Theory and the CernVM team in PH/SFT 2010-2011

- Later brought into BOINC mainstream code as "Vboxwrapper"
  - Ref: *http://boinc.berkeley.edu/trac/wiki/VboxApps*

  - BOINC developers very helpful with improvements

- BOINC projects currently deploying Virtualisation:
  - RNAword, Climateprediction.net, CAS@home

  - CERN (Theory, Atlas, CMS, LHCb)

BOINC and Virtualisation - 2
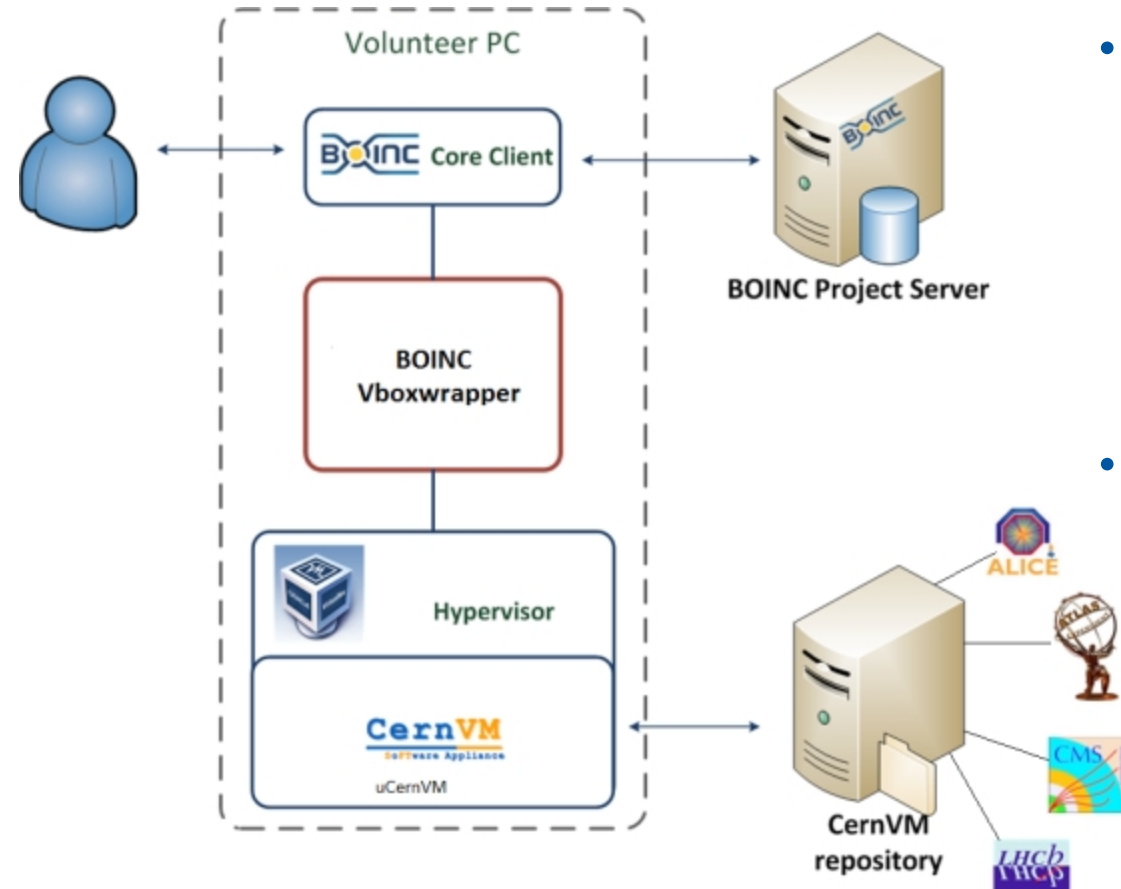
- BOINC distributes VMs to client machines along with a wrapper application

- The BOINC client installation for Windows now includes Virtual Box

Classic BOINC vs. Virtualisation

| Classic BOINC	Virtualisation
• Applications are native binaries • Unknown environment • Multitude of OS • Application building/testing and result verification is very labour-intensive	• Applications run in a VM • Typical hypervisor; VirtualBox (installed with BOINC on some OS) • Application to be built for one environment only
• BOINC takes care of job management • Local application framework must be integrated with BOINC	• BOINC takes care of distributing VM image • External job manager possible

LHC@home: Sixtrack

- Started as outreach project for CERN's 50th anniversary 2004, used for Year of Physics (Einstein Year) 2005
  - Based on experience from the Compact Physics Screen Saver (CPSS), which ran SixTrack on desktop computers at CERN

- Calculates stability of proton orbits in the LHC accelerator

- Written in FORTRAN, simulates particle trajectories

- Uses the classic BOINC approach

- Client runs on Linux, Mac and Windows platforms

- Renewed effort for LHC upgrade studies (HL-LHC)

- Total 118'000 volunteers, about 20'000 active recently

- Compute power: Peak 45 TFlops, average 13 TFlops

LHC@home Applications: Test4Theory

- Launched 2011 in partnership with the Citizen Cybercience Centre – CCC

- Theoretical fitting of all past experimental data (including LHC) using Monte Carlo simulation based on Standard Model

- Pioneered use of Virtualisation with BOINC

- Job reads data from CernVMFS

- External job management: CoPilot

- CernVM, CernVMFS, CoPilot: developed by CERN PH-SFT

- Wide range of potential (physics) applications
  - Project changed name in 2014 to **Virtual LHC@home**

Virtual LHC@home – jobs pr. month
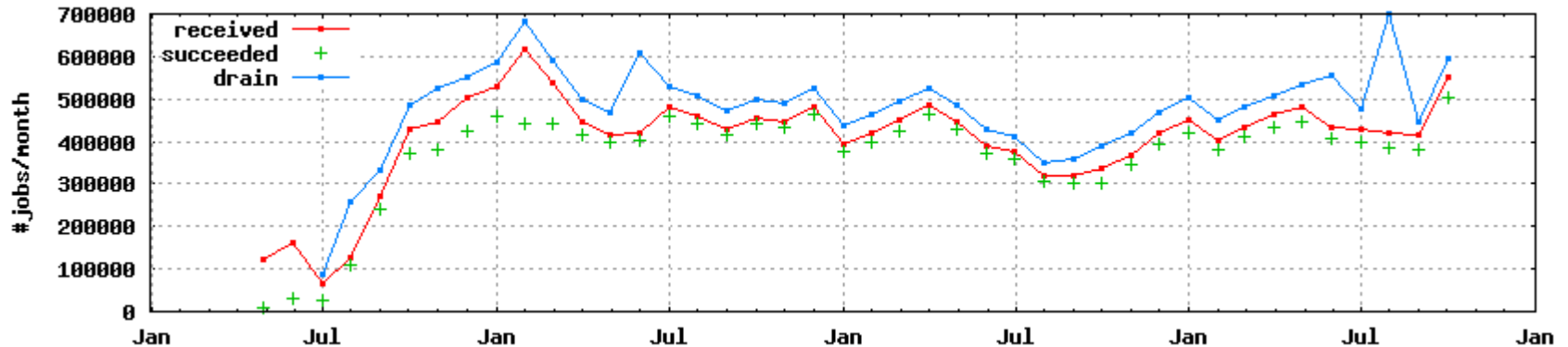
- Total of 1.7 trillion events simulated since 2011
- Source: MC Plots (http://mcplots-dev.cern.ch/production.php)
- See also: http://cern.ch/go/9nRz

LHC@home Applications: ATLAS

- Started as pilot within ATLAS early 2014, now public
  - http://atlasathome.cern.ch

    - *Development by D. Cameron, A Filipic, E. Lancon, W. Wu, C.A. Bourdarios et al.*
  - Also using CernVM and virtualisation
    - Evolved from CernVM to uCernVM last year
  - ARC CE integrated with BOINC, using PanDA for job management
- Supports simulations and potentially other types of ATLAS jobs
  - Job size and 64bit image limits to "hardcore" volunteers
  - Already significant CPU contribution (ref PanDA monitor)
- Integrated with LHC@home environment
  - BOINC server hosted by IT BOINC service
  - ARC-CE and BOINC sharing data via NFS

ATLAS@home jobs

Running jobs
662 Hours from 2015-02-18 to 2015-03-18 UTC
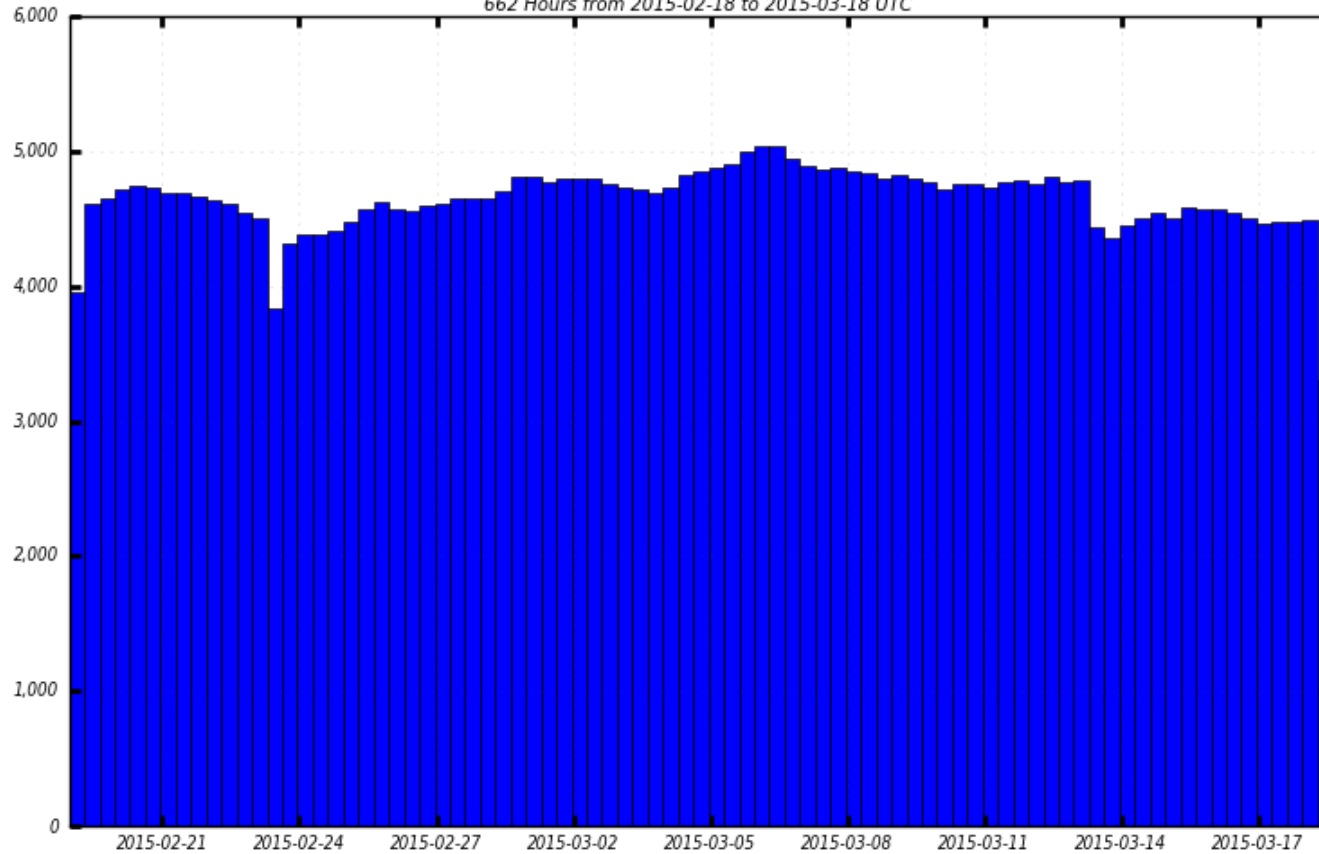
■ MC Simulation   ■ Others

Maximum: 5,039 , Minimum: 0.00 , Average: 4,589 , Current: 3,297

BOINC ATLAS contribution

BOINC is the 2nd largest simulation site

Running 4-5k parallel jobs

http://

Legend:
- BNL-ATLAS
- UKI-LT2-BRUNEL
- MWT2
- UKI-NORTHGRID-MAN-HEP
- BU_ATLAS_TIER2
- WUPPERTALPROD
- IFIC-LCG2
- UNI-FREIBURG
- CERN-P1
- BOINC
- CERN-PROD
- INFN-NAPOLI-ATLAS
- UKI-SOUTHGRID-RALP
- SIGNET
- UKI-NORTHGRID-LANC
- UTA_SWT2
- GRIF-IRFU
- GRIF-LPNHE
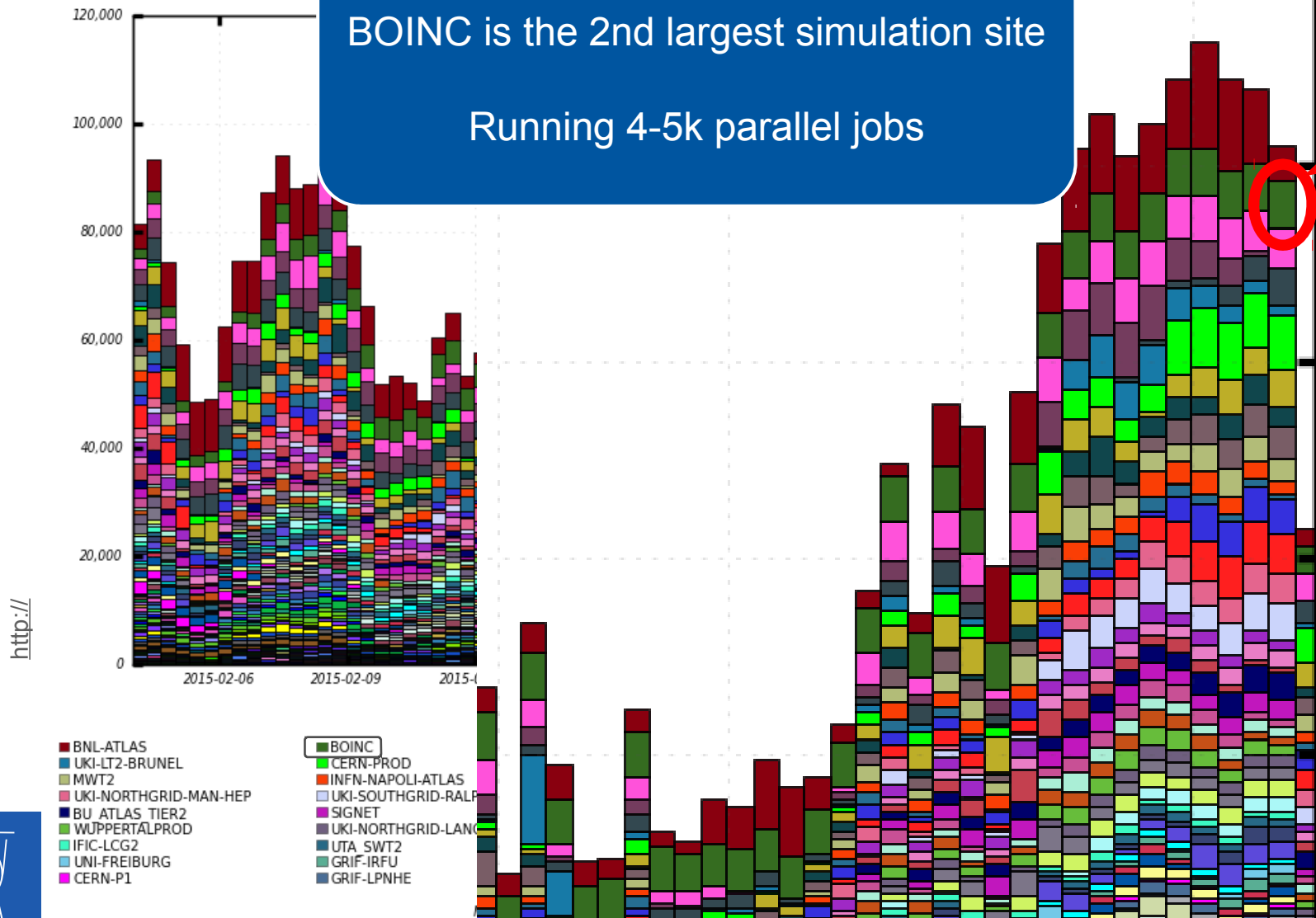
LHC@home Applications: Other

- ## LHCb: Beauty
  - Since 2012, still in test phase
  - Volunteers from within LHCb collaboration
  - Job management: DIRAC
  - Vboxwrapper application with CernVM image
    - uCernVM

- ## CMS
  - Started work in summer 2014
  - Prototype running, rapidly gaining experience
  - Job management: Crab linked to "Data Bridge" messaging service
  - To be added as beta application on Virtual LHC@home once stable

- ## A pilot project with EPFL for Human Brain project
  - Also using uCernVM

CMS@Home Infrastructure

CRAB3

TaskWorker
Plugin

Message
Queue

DataBridg

DynaFed

S3

Agent
Wrapper

VM

VBoxwrapper

FTS

Async stage-out

Grid

http://svnweb.cern.ch/trac/lcgdm/wiki/Dynafeds
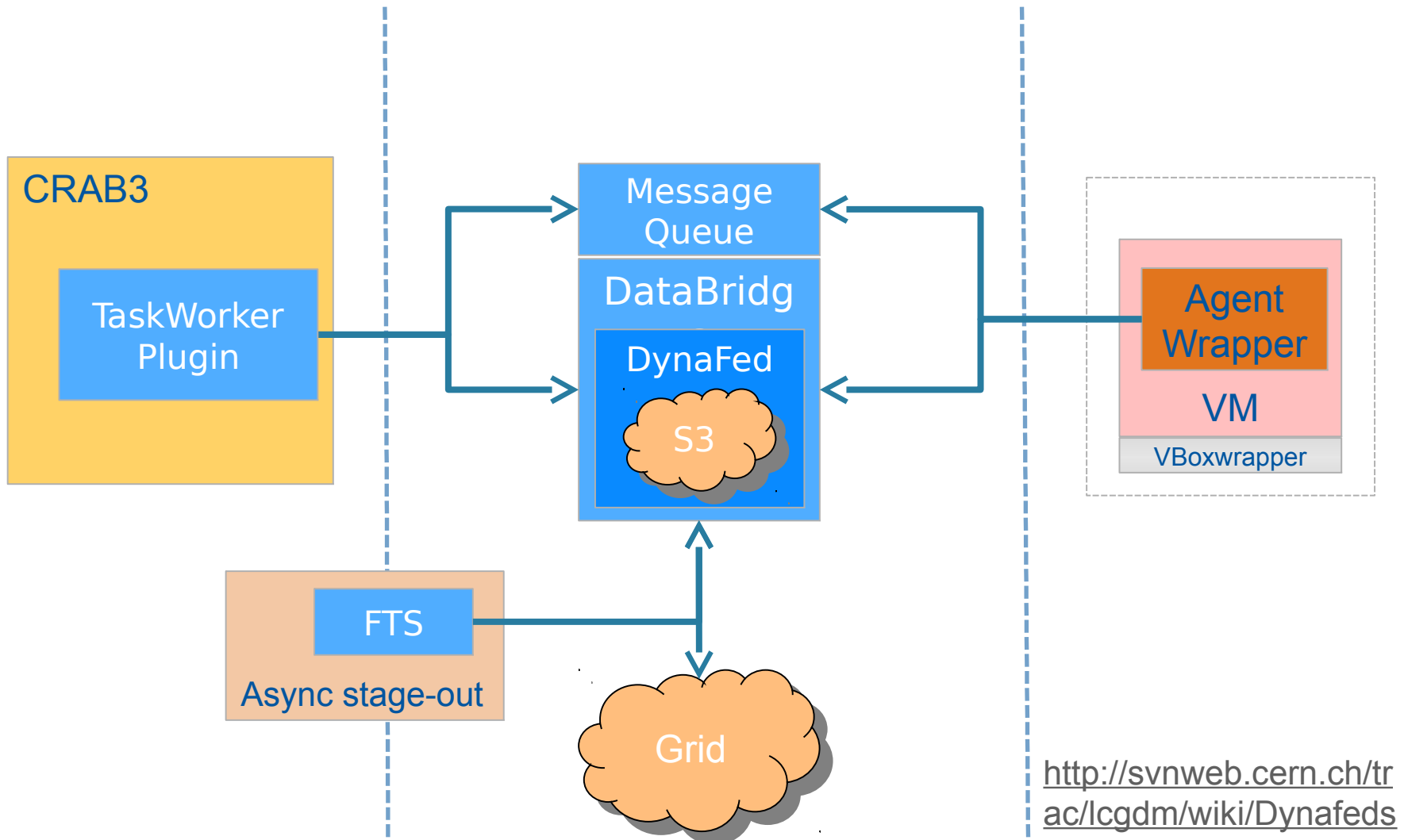
BOINC Service at CERN

- BOINC server cluster
  - LHC@home servers
    - Sixtrack, Theory, ATLAS
    - Test servers (CMS, LHCb, project with EPFL, dev environments)
- BOINC server application support
  - Configuration, monitoring
  - MySQL database server back-end
  - BOINC server application configuration and updates
- *Handled by the user project teams:*
  - *Porting of applications to BOINC*
  - *Application specific job management framework*
  - *Communication with users about scientific projects*
  - *Content of forums and portal*

BOINC at CERN – recent developments

- Service Consolidation
  - Drupal portal for lhcathome.cern.ch
  - OpenStack VMs used for the BOINC server cluster
    - Puppet managed machines.
    - Allows the quick creation of servers and even clients for testing purposes
  - Use of centralised MySQL service
    - BOINC server code modified by Tomi Asp (Graduate student at CERN for 1 year)
  - Use of centralised NFS disk space
    - Take advantage of the central services features and support

NFS backend for BOINC project

Puppet managed BOINC Server

Drupal Portal for all projects

Centralized MySQL database
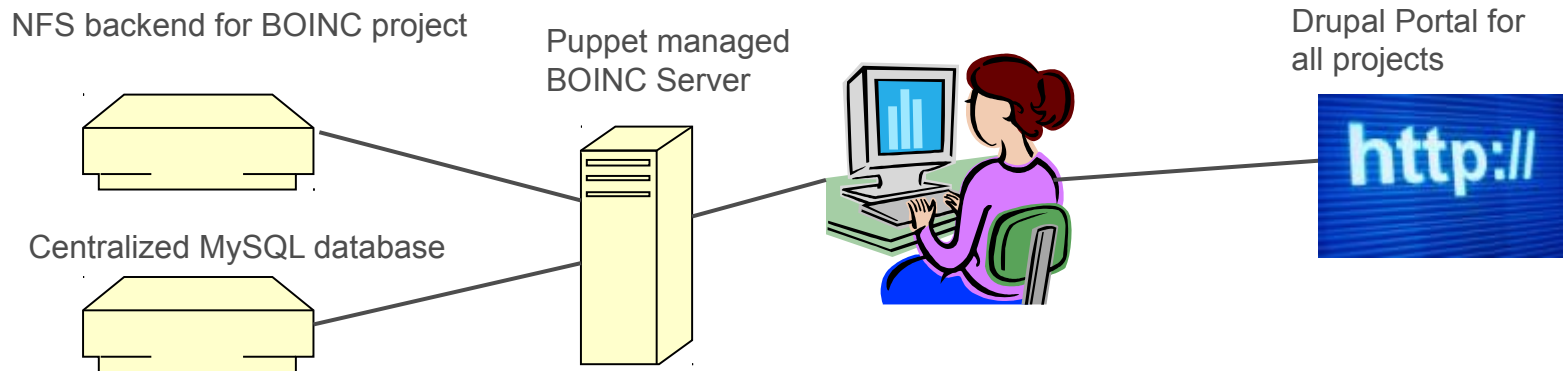
BOINC service evolution

- VM applications that report back to a local job management framework can be part of the Virtual LHC@homeBOINC project

- Other projects (Sixtrack, ATLAS) currently hosted on separate servers to avoid I/O bottleneck
  - A distributed server setup with separate servers for upload/download would allow for a single project

- Use Drupal portal as common entry point for all BOINC projects and applications

- Aim for standardisation on a volunteer cloud common job management solution (Data Bridge)

BOINC client - SLC6

- 3 steps and your machine is crunching:

```
#yum install boinc-client-7.2.33-3.git1994cc8.el6.x86_64

#yum install VirtualBox-4.3-4.3.12_93733_el6-1.x86_64.rpm

bash$ boinc_client --attach_project
http://lhcathome2.cern.ch/vLHCathome   a25c5d7c0887ef8faf5f9e46d3ca54fc
```

- Additionally "boincmgr" can be used on desktops for interactive use (attach to projects, suspend tasks, settings etc.)

HEP BOINC use cases

- Desktop: BOINC client with BOINC manager
  - Individual BOINC user like for volunteers among the general public
  - Or generic institute BOINC user for central deployment (desktop grid)

- Small clusters: BOINC client and virtual box
  - Install RPMs, provide startup script to run the BOINC client, generic BOINC user

- Larger clusters (e.g. small Tier-2 centers)
  - Like above, configured centrally, e.g. with Puppet

- Tier-2 with local grid storage
  - No grid credentials on VMs distributed with BOINC to access local storage.
  - Launching VMs with VAC or VCycle may be more appropriate

Conclusions

- Volunteer computing offers a lightweight way to distribute jobs
- BOINC is the de-facto standard middleware for volunteer computing
- Thanks to virtualization support, BOINC is now suitable for a wider range of HEP applications
- Applications running under CernVM and getting data from CernVMFS can hosted as part of LHC@home
- The size of the application data sets is a bottle neck
- Outreach and communication is essential to get contributions from the general public
- Desktops and other opportunistic local resources offer capacity that can be exploited

Acknowledgements

- BOINC service: Pete Jones, Tomi Asp, Alvaro Gonzalez

- Also Miguel Marquina, Helge Meinhard, Manuel Guijarro, Ignacio Reguero

- Test4Theory: Ben Segal, Peter Skands, Jakob Blumer, Ioannis Charalampidis, Artem Harutyunyan, Predrag Buncic, Daniel Lombrana Gonzalez, Francois Grey et al

- Sixtrack: Eric McIntosh, Riccardo de Maria, Massimo Giovannozi, Igor Zacharov et al

- ATLAS: David Cameron, Andrej Filipic, Eric Lancon, Wenjing Wu

- CMS: Laurence Field, Hendrik Borras, Daniele Spiga, Hassan Riahi

- LHCb: Federico Stagni, Joao Medeiros et al

- BOINC: David Anderson, Rom Walton

- All our IT colleagues offering a layered service, DB on Demand, Openstack, Puppet, AFS, NFS filers, Linux, network... :-)

References

- http://cern.ch/lhcathome

- http://boinc.berkeley.edu

- Contact the BOINC service team at CERN