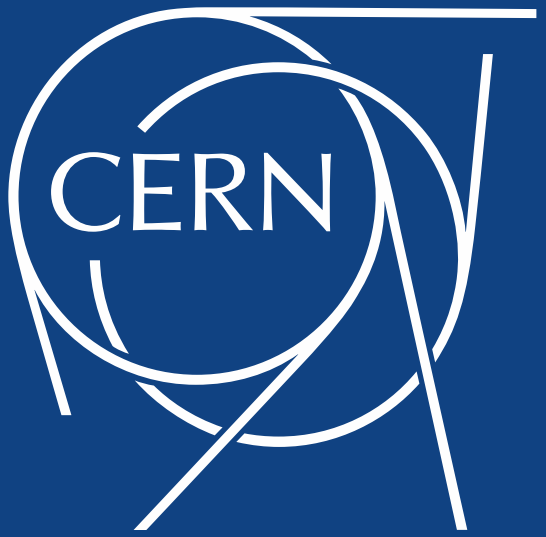

CERN

Volunteer Computing

Overview

- Volunteer Computing
  - BOINC
- Volunteer Computing For HEP
  - Virtualization
- Volunteer Computing @CERN
  - Towards a Common Platform

Volunteer Computing

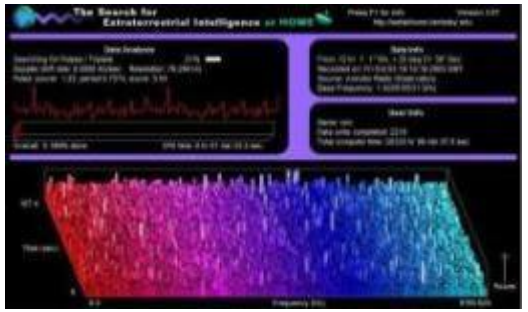
Volunteer Computing

- A type of distributed computing
- Origins in mid 1990s
- Computer owners *donate* computing capacity
  - To a *cause* or project
- Not necessarily only spare cycles on desktops
  - Idle machines in data centers
  - Home clusters
- SETI@home and Folding@home
  - Launched 1999

SETI HOME

- Search for Extra-Terrestrial Intelligence
- Analyses radio signals
  - Arecibo Observatory in Puerto Rico
- Supporting scientific work
  - Detection intelligent life outside Earth
    - Yielded no conclusive results
      - No evidence for ETI signals
- Viability and practicality of volunteer computing
  - 120K Active Users
  - 180K Active Hosts

BOINC STATS

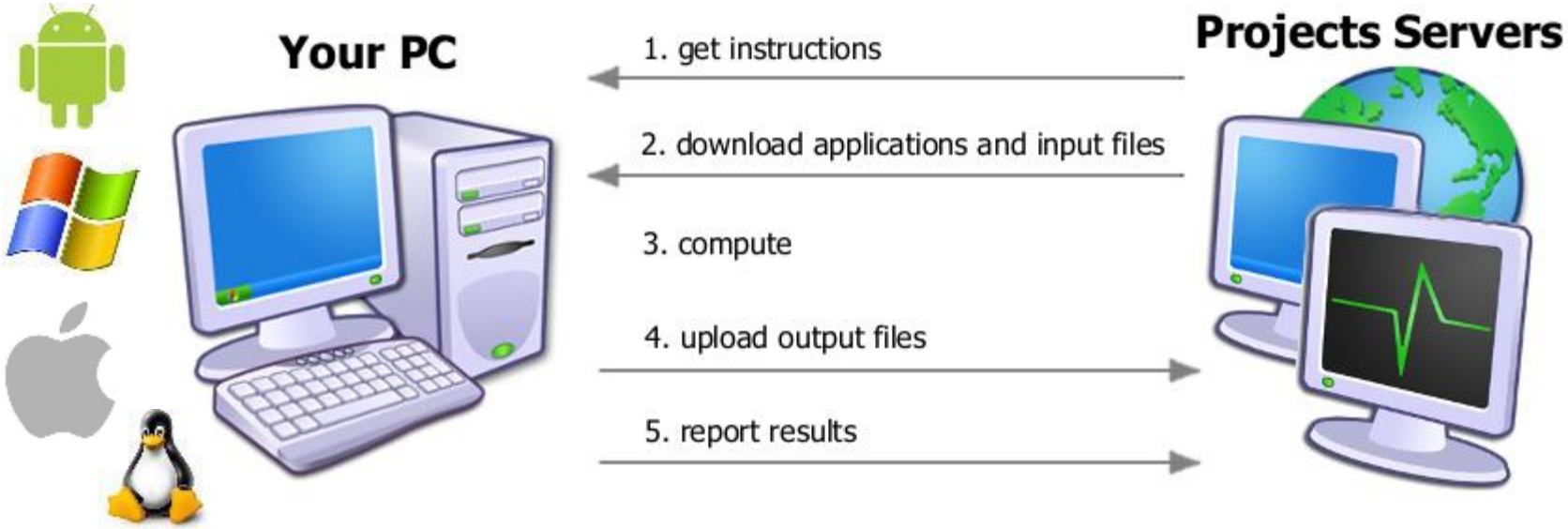


- Berkeley Open Infrastructure for Network Computing
  - Started in 2002
  - Funded by the National Science Foundation (NSF)
  - Developed by a team based at the Space Sciences Laboratory
    - University of California, Berkeley
    - Led by David Anderson
- Provides the middleware for volunteer computing
  - Client (Mac, Windows, Linux, Android)
  - GUI
  - Application runtime system
  - Server software
  - Project Web site

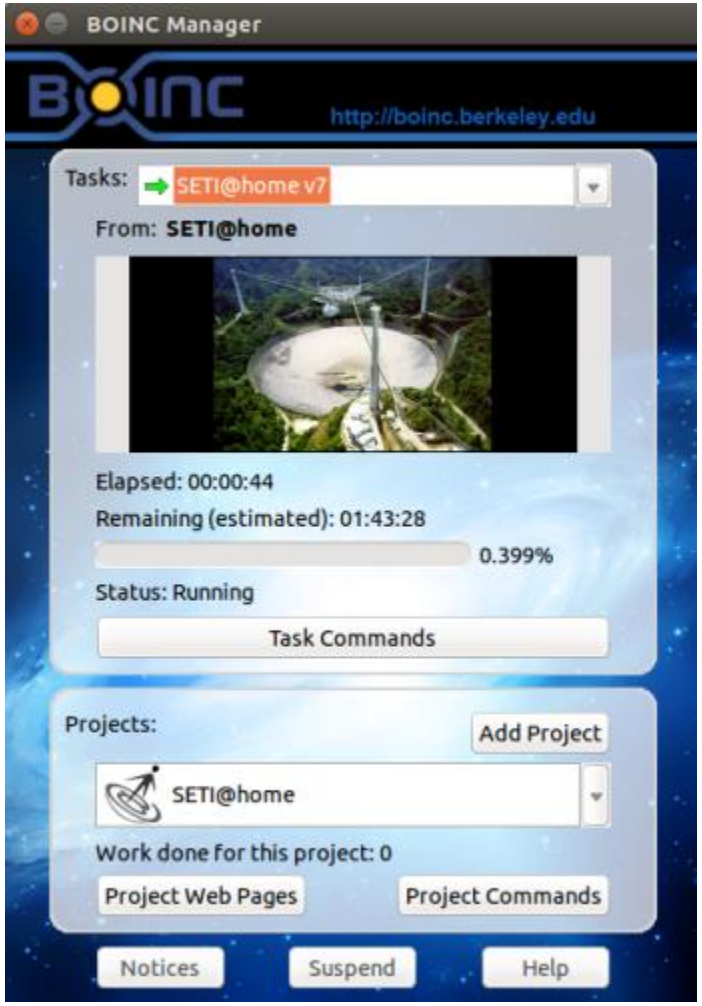
- The first project based on BOINC was Predictor@home
  - Predict protein structure from protein sequences
- Soon thereafter SETI@home and ClimatePrediction.net
  - CP.net: 12K Active Users, 16K Active Hosts
- Numerous other BOINC-based projects
  - Rosetta@home
    - Protein structure prediction
    - 97K Active Users, 115K Active Hosts
  - Einstein@home
    - Gravitational-waves (LIGO detectors)
    - 31K Active Users, 86K Active Hosts
- In 2007, IBM World Community Grid switched to BOINC
  - Multiple projects
  - 67K Active Users, 1.8M Active Hosts

Volunteer Perspective

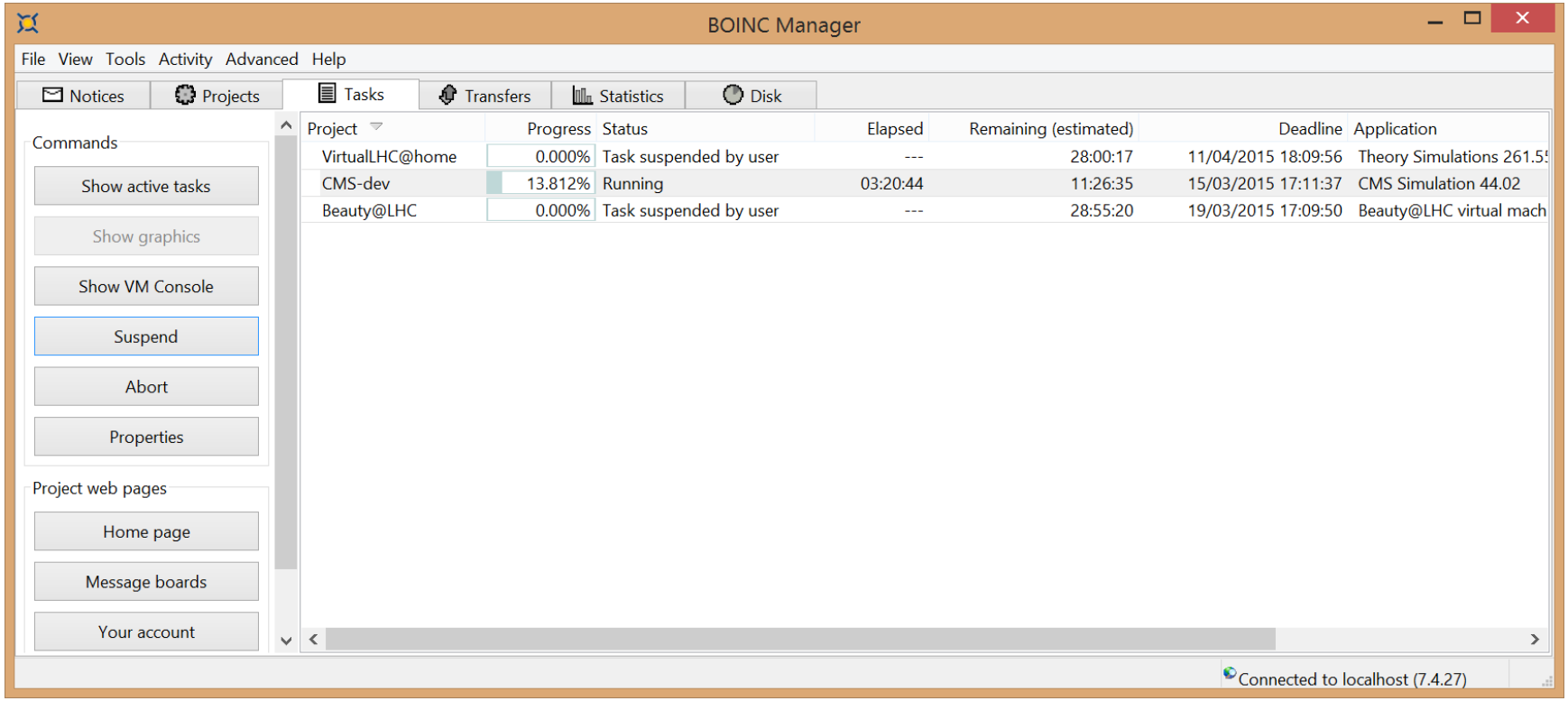
- Download and run BOINC software
- Choose a project
- Enter an email address and password
  - Or silent connection with a key
- Earn Credit



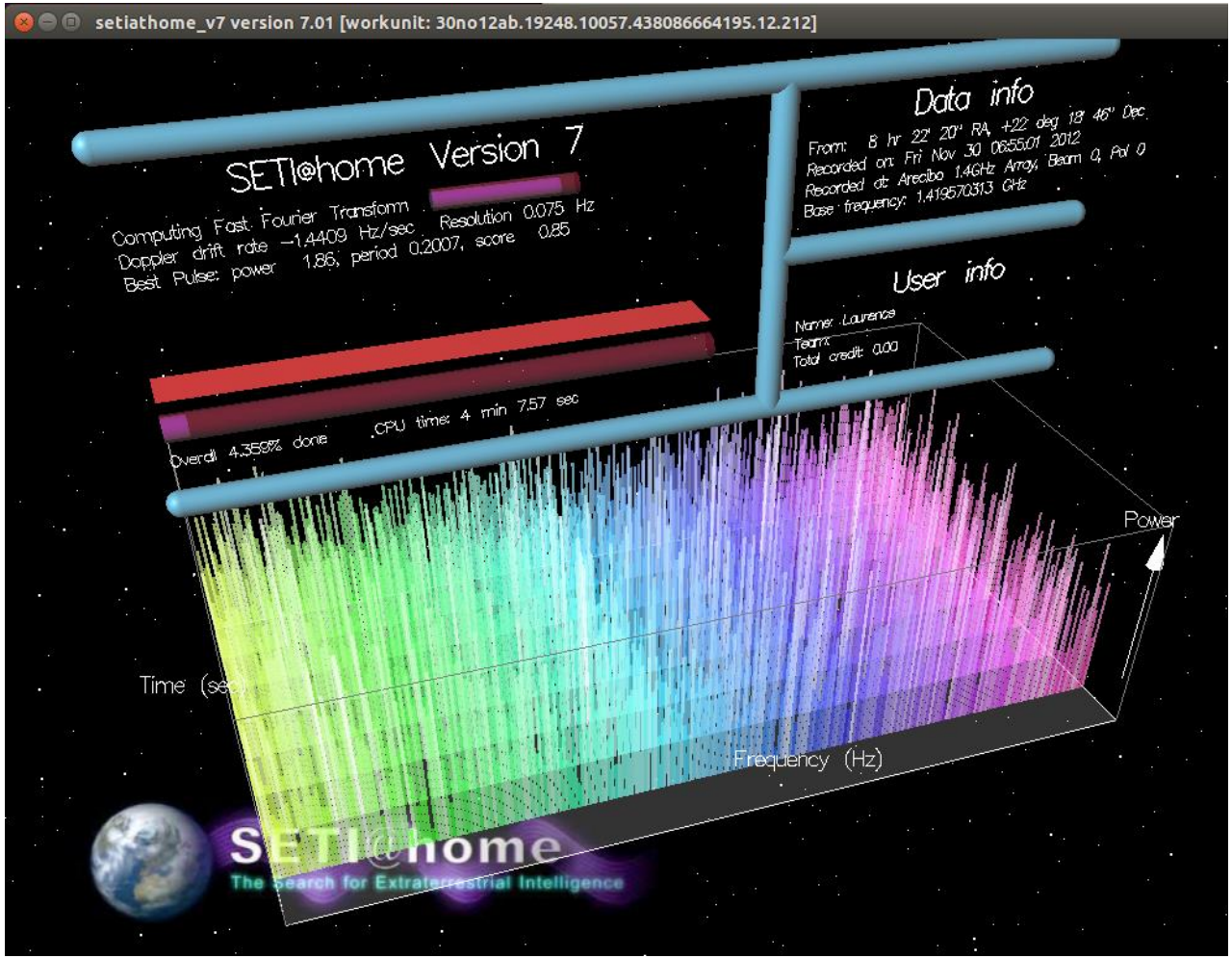
BOINC Manager



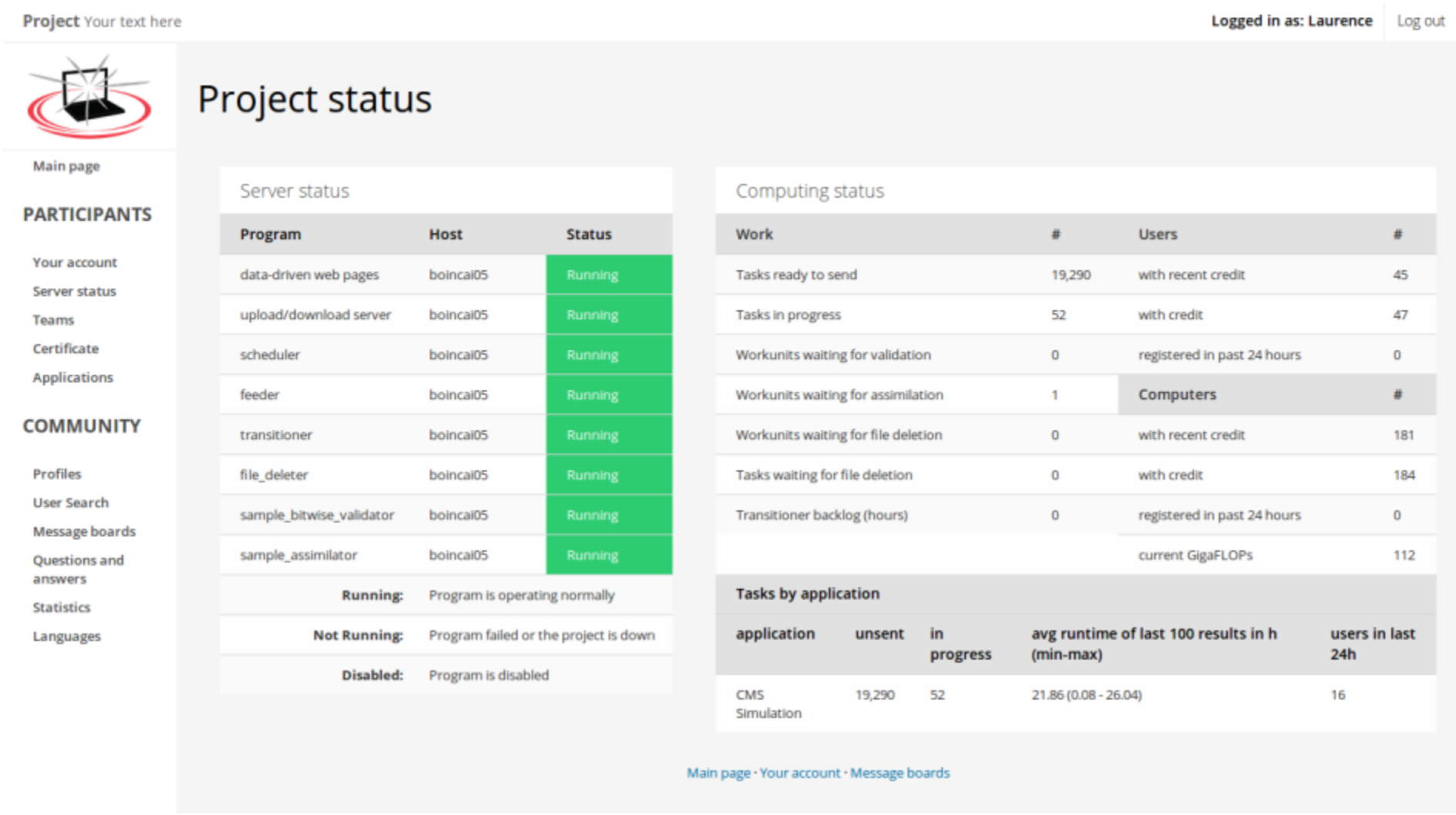
Advanced View



Eye Candy


setiathome_v7 version 7.01 [workunit: 30no12ab.19248.10057.438086664195.12.212]
SETI@home Version 7
Computing Fast Fourier Transform
Doppler drift rate −1.4409 Hz/sec Resolution 0.075 Hz
Best Pulse: power 1.86, period 0.2007, score 0.85
Data info
From: 8 hr 22' 20" RA, +22 deg 18' 46" Dec
Recorded on: Fri Nov 30 06:55:01 2012
Recorded at: Arecibo 1.4GHz Array Beam 0, Pol 0
Base frequency: 1.419570313 GHz
User info
Name: Laurence
Team:
Total credit: 0.00
Overall 4.359% done
CPU time: 4 min 7.57 sec
Power
Time (sec)
Frequency (Hz)
SETI@home
The search for Extraterrestrial Intelligence

BOINC Server



Volunteer Computing For HEP

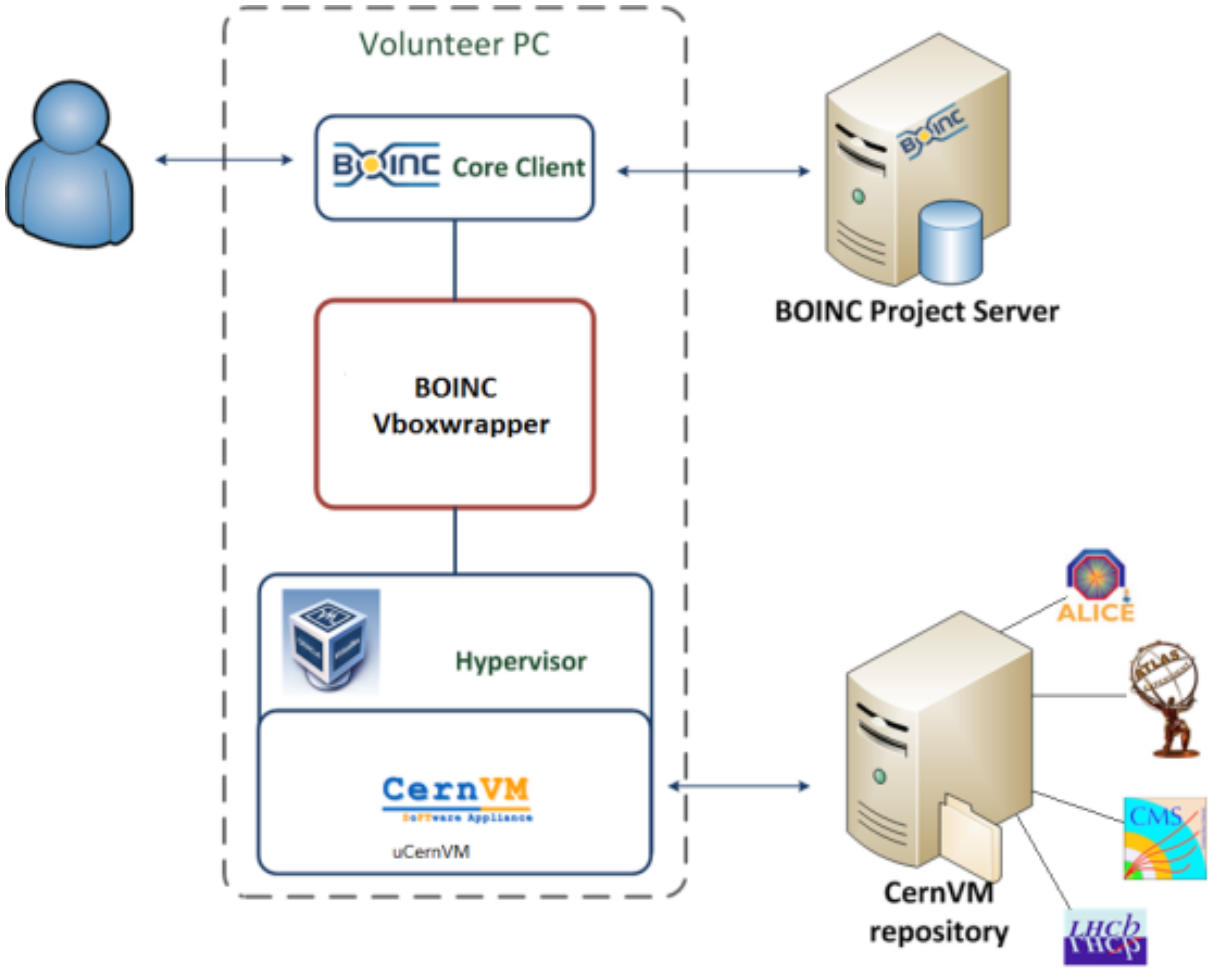
Motivation

- Free* resources
  - 100K hosts achievable for large projects
    - Actual core count is higher
- Community engagement
  - Outreach channel
    - Explaining the purpose and value of the science
  - Participation
    - Offering people a chance to contribute
      - Engagement forms a strong bond
    - Community support
- * There are cost associated with their use

Challenges

- The cost of using the free resources
  - Initial integration requires investment
- Operations and maintenance
  - Public facing support
    - Lowered by community support
- Attracting and retention of volunteers
  - Advertisement
  - Engagement
- Low Level of Assurance
  - Anyone can register as a volunteer
    - Not the same level of trust as with Grid authentication
- Running HEP software on Windows
  - 85% of the resources

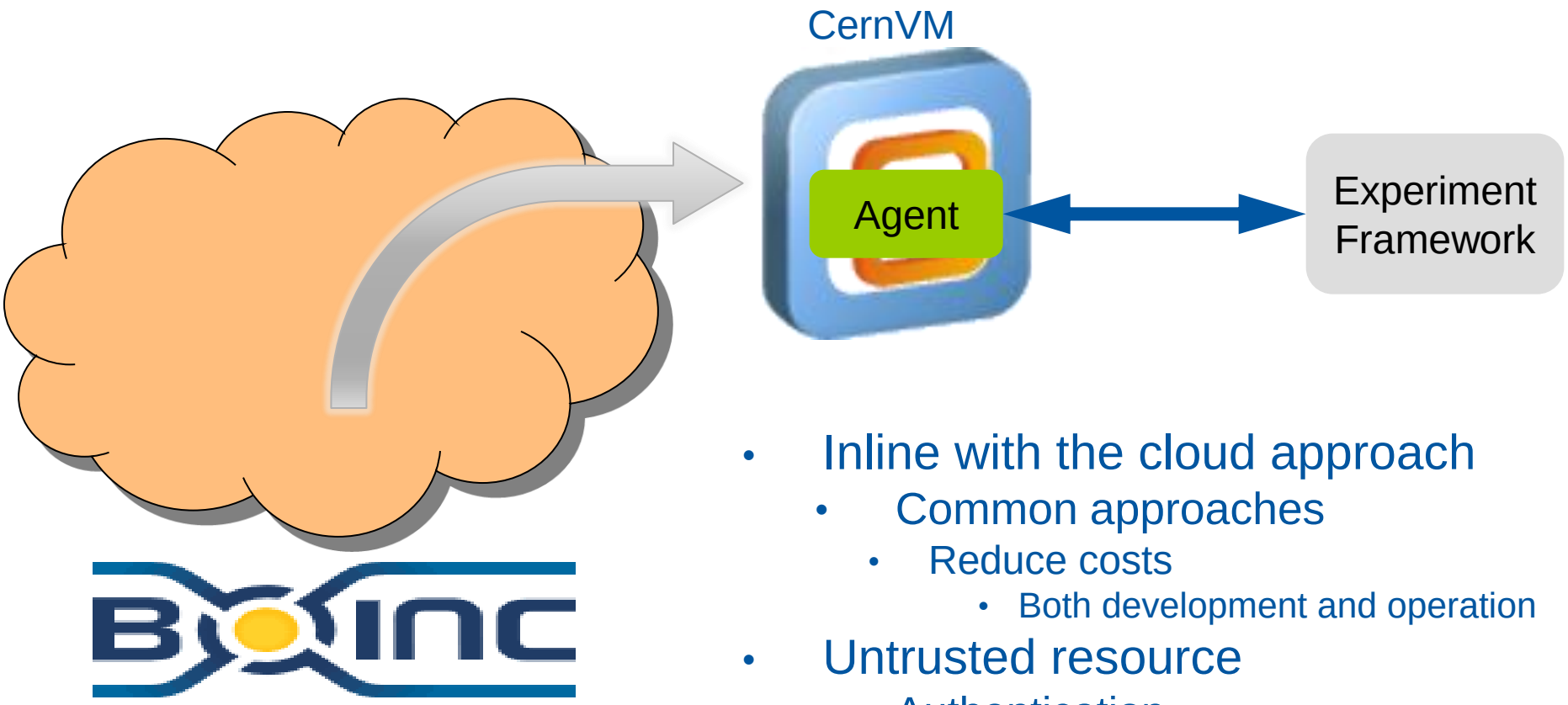
BOINC With Virtualization



Virtualization

- Pioneered with Test4Theory and CernVM 2010-2011
  - Included into the mainstream BOINC code
- Vboxwrapper
  - http://boinc.berkeley.edu/trac/wiki/VboxApps
  - BOINC developers very helpful with improvements
- BOINC projects currently deploying Virtualisation:
  - RNA World
  - Climateprediction.net
  - CAS@home
  - CERN (Theory, Atlas, CMS, LHCb)
- Heavy lifting is done with VirtualBox
- WebAPI, an example of alternative approaches
  - CERN 60 - Public Computing Challenge
  - http://test4theory.cern.ch/challenge/

The Vacuum Model



- Inline with the cloud approach
  - Common approaches
    - Reduce costs
      - Both development and operation
- Untrusted resource
  - Authentication
  - Validation

Volunteer Computing @CERN

BOINC Service @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 project teams:
  - Porting of applications to BOINC
  - Application specific job management framework
  - Communication with users about scientific projects
  - Content of forums and portal

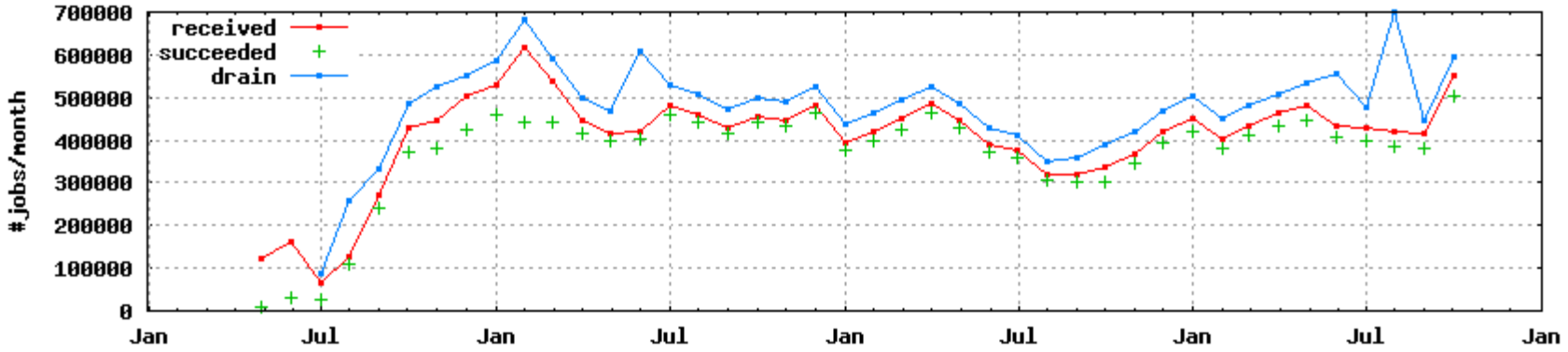
SixTrack (LHC machine)

- Original classic BOINC project for beam simulations
  - Calculates stability of proton orbits in the LHC accelerator
    - Simulates particle trajectories
- Based on experience from the Compact Physics Screensaver (CPSS)
  - Ran SixTrack on desktop computers at CERN
- Outreach project for CERN's 50th anniversary 2004
  - Also Year of Physics (Einstein Year) 2005
- Application written in FORTRAN
  - Runs on Linux, Mac and Windows platforms
- Renewed effort for LHC upgrade studies (HL-LHC)
  - 12K Active Users
  - 19K Active Hosts
  - 35 TeraFLOPS

Test4Theory

- Theoretical fitting of all past experimental data
  - Including from the LHC
  - Using Monte Carlo simulation based on Standard Model
- Launched 2011
  - In partnership with the Citizen Cyberscience Centre (CCC)
- Pioneered use of Virtualization with BOINC
- Uses recent developments from CERN's PH-SFT Group
  - CernVM
  - CernVMFS
  - CoPilot
- Wide range of potential (physics) applications
  - In 2014 changed name to Virtual LHC@home

Test4Theory Usage

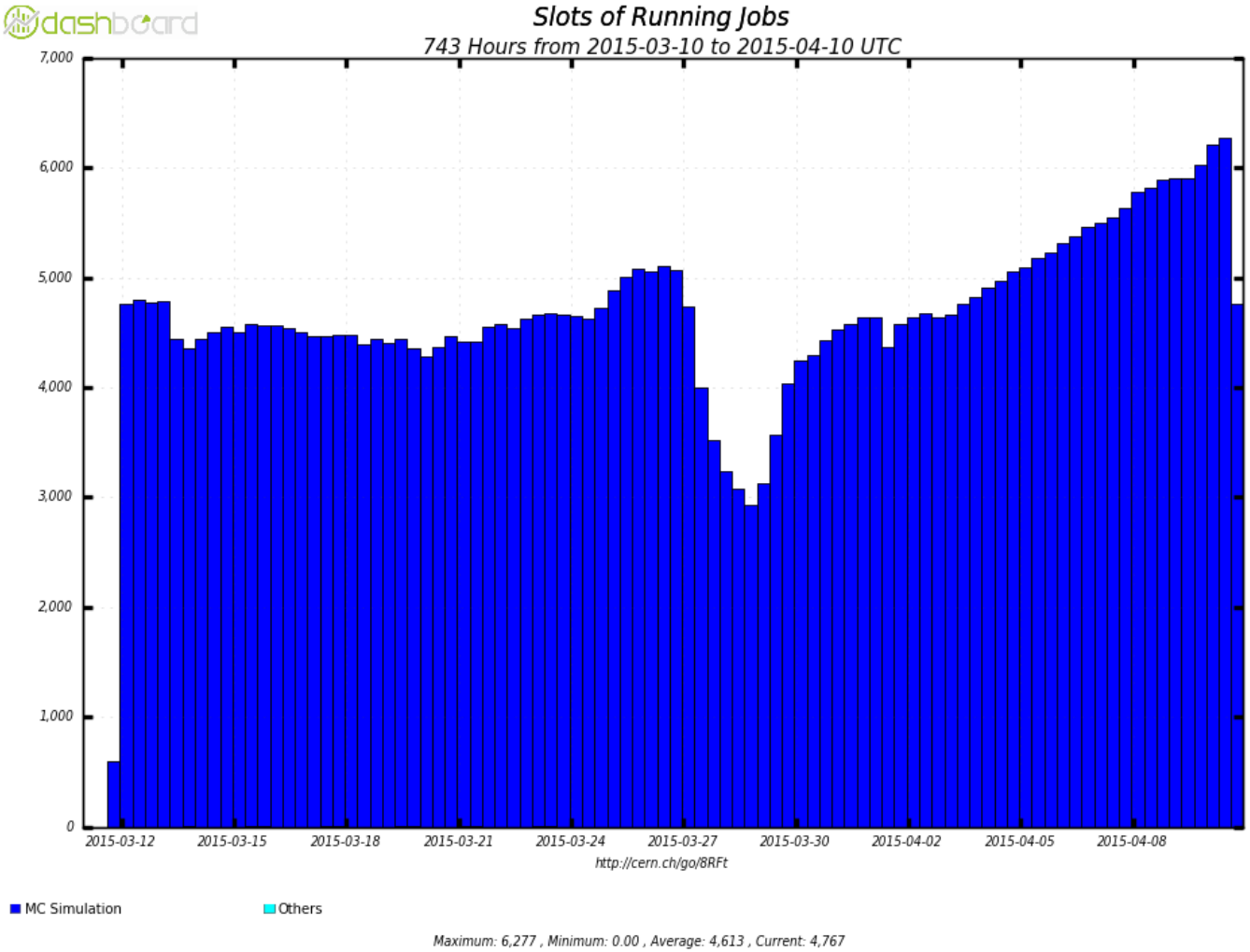


- Total of 1.7 trillion events simulated since 2011
- Source: MC Plots (http://mcplots-dev.cern.ch/production.php)
- See also: http://cern.ch/go/9nRz

ATLAS@home

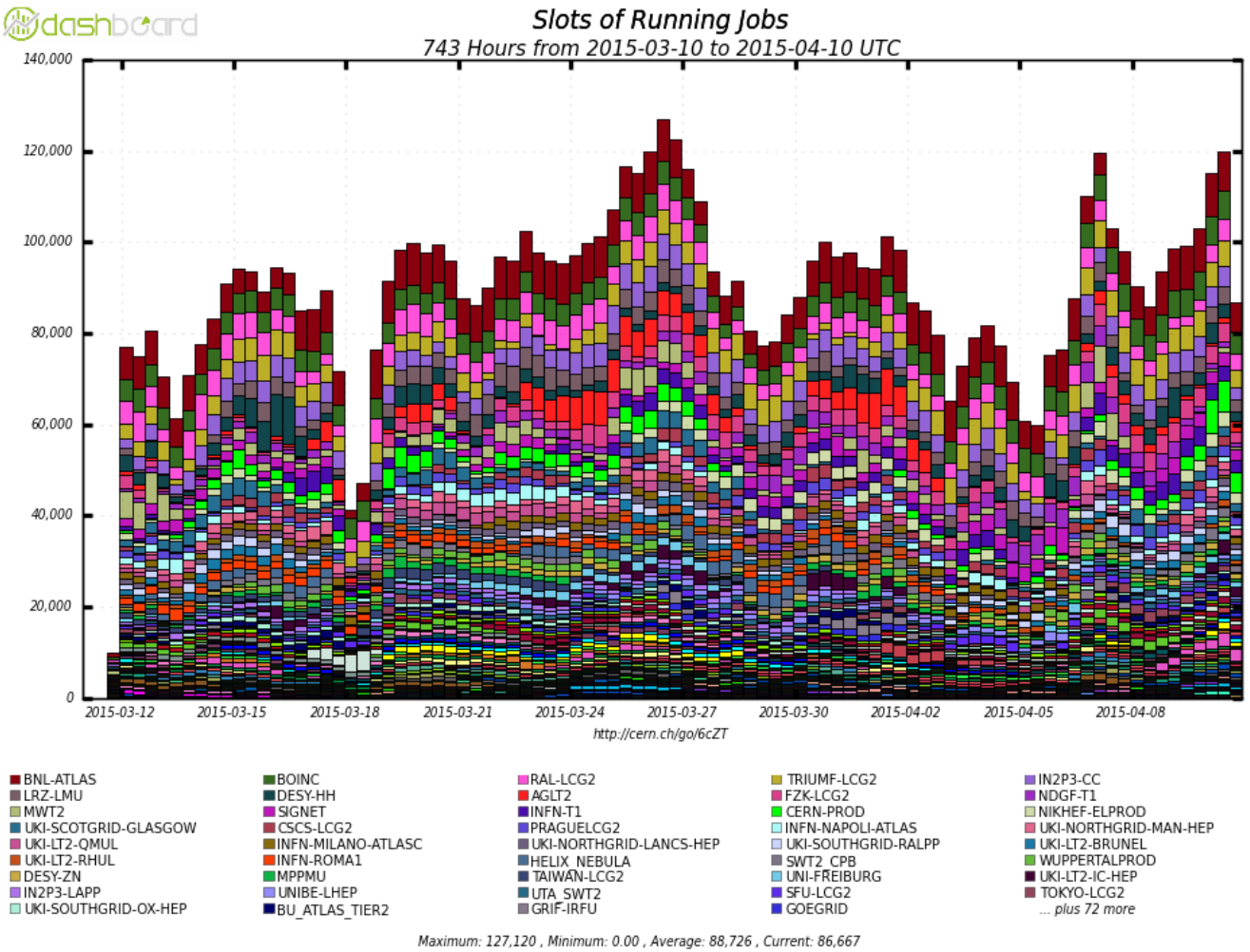
- Started as pilot beginning of 2014
  - Now open to the public
    - http://atlasathome.cern.ch
- Also using CernVM and virtualization
  - Classic BOINC model
- ARC CE used to interface with BOINC
  - PanDA for job management
- Supports simulations
  - Potentially other types of ATLAS workloads
- Job size and 64bit image limits to “hardcore” volunteers
  - Already significant CPU contribution
- Integrated with LHC@home environment
  - BOINC server hosted by CERN’s IT-PES group
  - ARC-CE and BOINC sharing data via NFS

ATLAS@home Usage

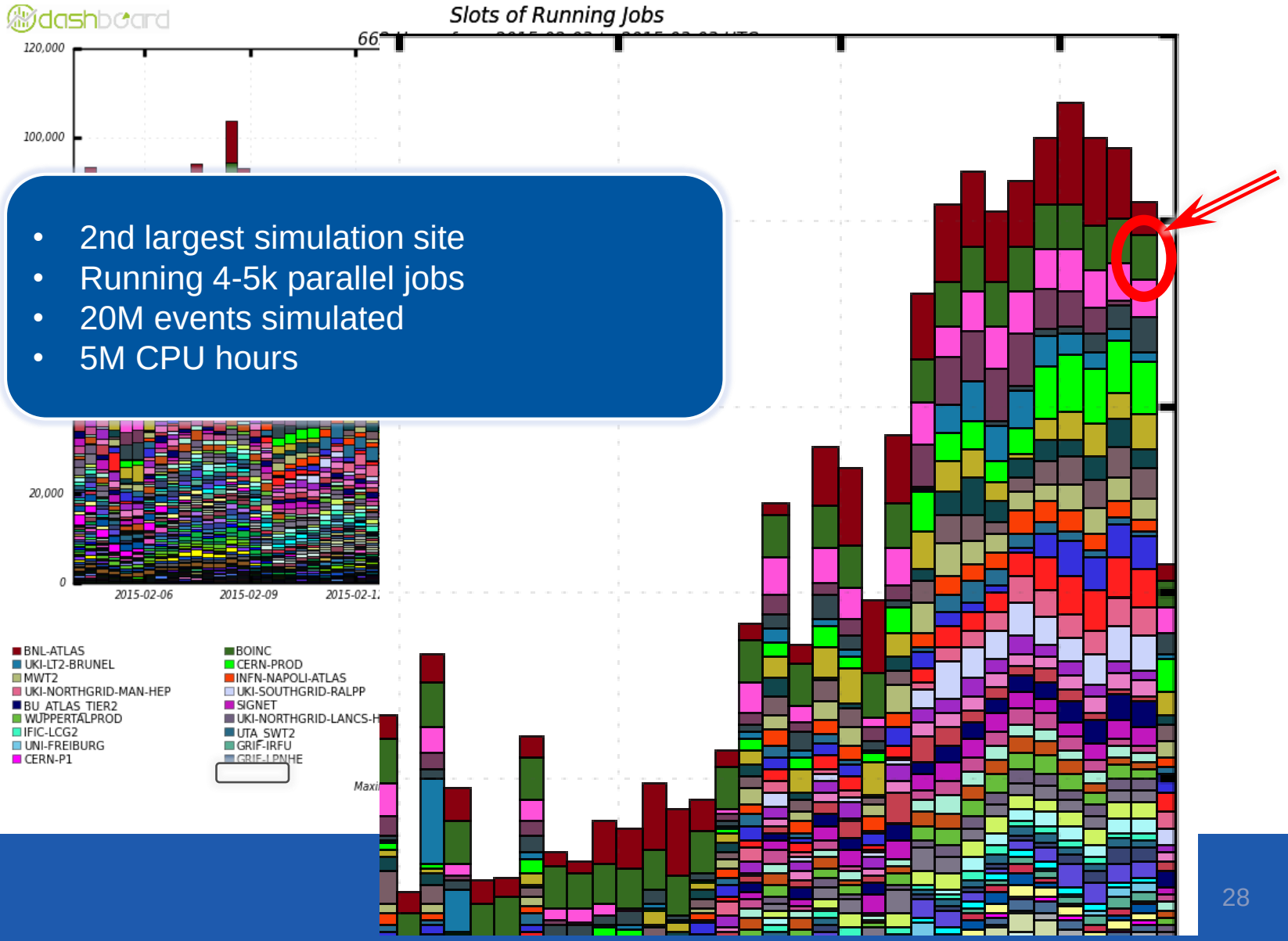


ATLAS@home Contribution

http://atlasathome.cern.ch/atlas_job.php



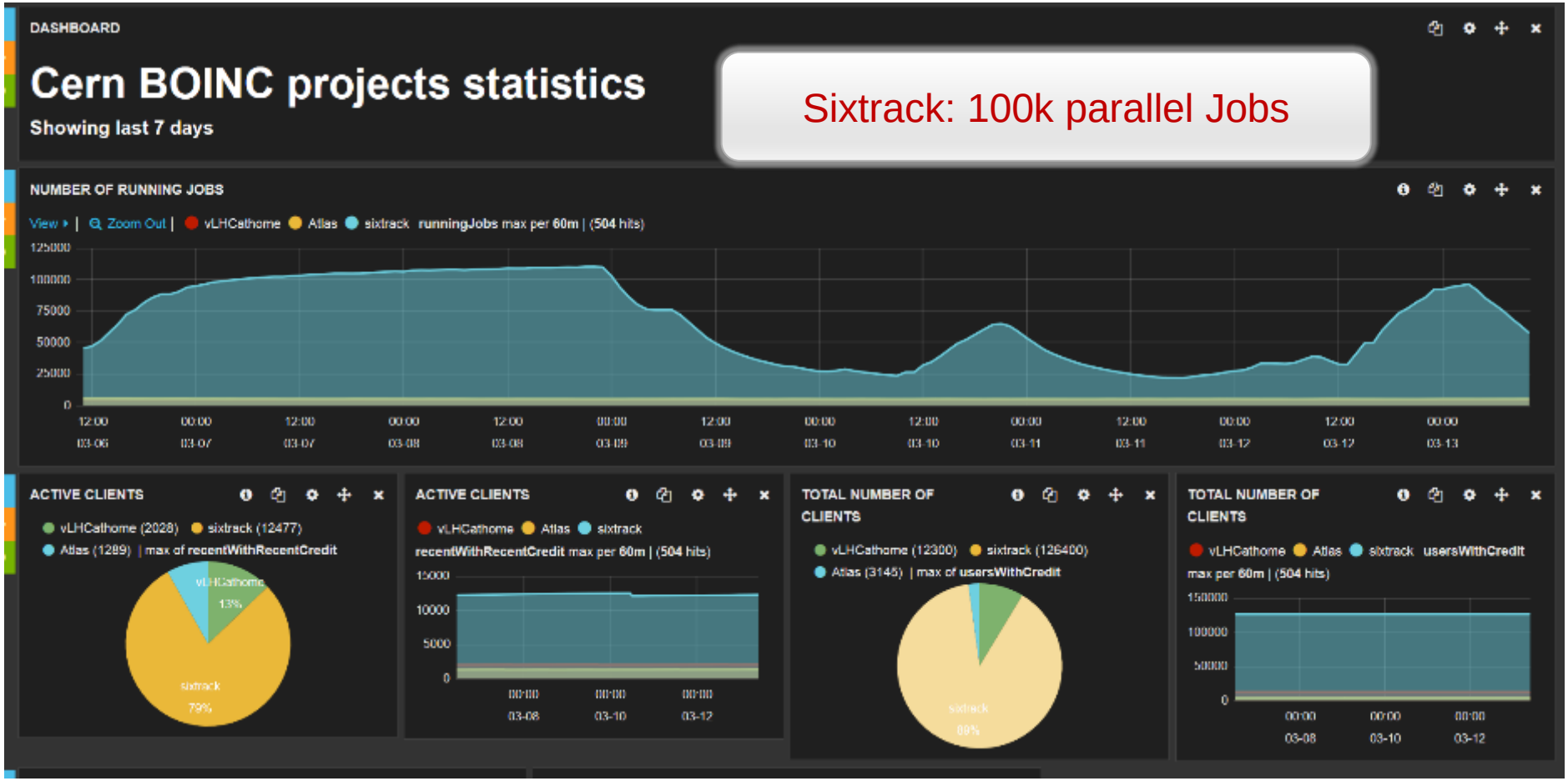
ATLAS@home Contribution



- 2nd largest simulation site
- Running 4-5k parallel jobs
- 20M events simulated
- 5M CPU hours

http://atlasathome.cern.ch/atlas_job.php

BOINC Service Monitor



IT-PES http://cern.ch/go/9nRz

Beauty@home

- In development since 2012
  - Requires x509 credential in the client VM
    - Volunteers from within LHCb collaboration
- Communicates directly with DIRAC
- Vboxwrapper application
  - Using uCernVM

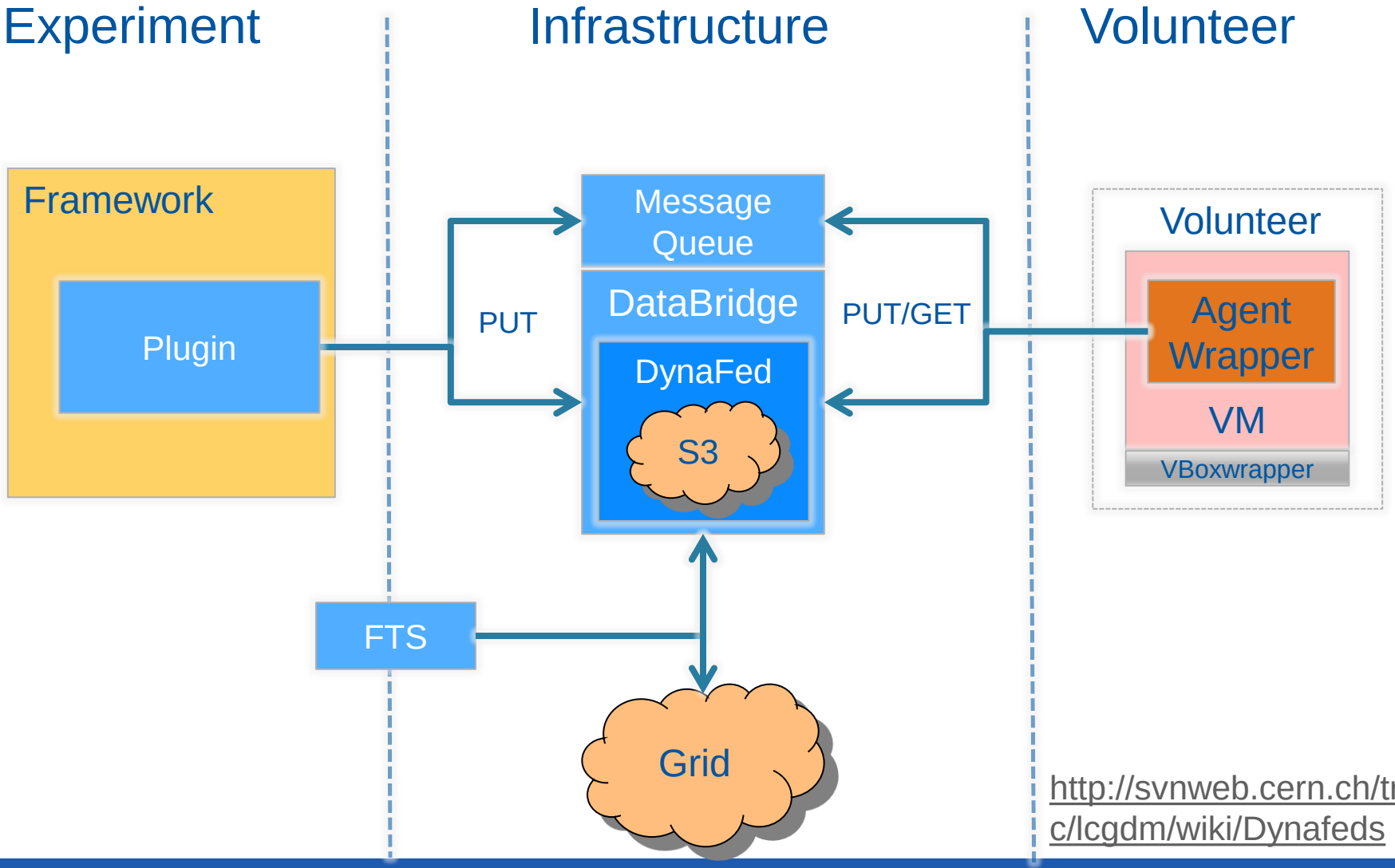
CMS@home

- In development summer 2014
  - Prototype service running
    - Rapidly gaining experience
- To be added as application in vLHC@home
  - Once stable

BOINC Service Evolution

- vLHC@home BOINC Project
  - Currently for applications following the vacuum model
- Separate servers
  - To avoid I/O bottleneck
  - e.g. Sixtrack, ATLAS
- A distributed server setup
  - For upload/download
  - Would allow for a single project
- Drupal portal as common entry point
  - For all BOINC projects and applications
- Aim for common solutions
  - To support the experiment frameworks

The DataBridge



http://svnweb.cern.ch/trac/lcgdm/wiki/Dynafeds

Towards A Common Platform

- Coordinated outreach efforts
  - Maximize the potential resource pool
    - Fair share the resources
      - Volunteers typically configure multiple projects
- Development, Maintenance and Operations
  - Share the costs
- Build upon a common approach
  - Reuse components and services
    - Provided centrally as an infrastructure
- Common platform for Volunteer Computing
  - BOINC
  - Web presence
  - Outreach
  - Databridge

Summary

- Volunteer Computing can and is providing
  - Significant additional computing resources
    - Potentially O(100K) machines
- Virtualization enables HEP applications
  - To run on multiple x86 platforms
    - Can therefore reach more volunteers
      - And hence resources
- The experiments are trying to exploit this opportunistic resource
  - Many @home project exist or are in development
- Requires investment
  - Initial integration
  - Attracting volunteers
  - Supporting volunteers via the forum
- Work towards a common platform
  - Share Development, Maintenance and Operations

Aknowledgements

- 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... :-)

www.cern.ch