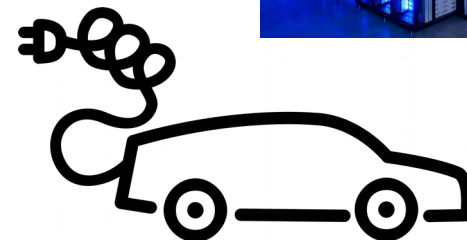


CVMFS On Volunteer Machines

Laurence Field

Volunteer Computing

- Computer owners donate computing capacity
 - To a cause or project
- Spare cycles on computers
 - Away from office/desk, unused cores
- Other opportunistic resources
 - Idle machines in data centers
 - IoT?
 - Opportunistic Energy?
 - Cars e.g. V2G?



Platform

- Berkeley Open Infrastructure for Network Computing
 - Started in 2002
 - By a team based at the Space Sciences Laboratory
 - University of California, Berkeley
 - Led by David Anderson
 - Now an open source project
- Provides the toolkit for volunteer computing
 - Client (Mac, Windows, Linux, Android)
 - Graphical User Interface
 - Application runtime system
 - Server software
 - Project Web site



Motivation

- Free* resources
 - 100K hosts achievable for large projects
 - Actual job slot count (number of cores) 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 costs associated with their use



Volunteer
computing
for the LHC

Search

[HOME](#)

[ABOUT](#) ▾

[PROJECTS](#) ▾

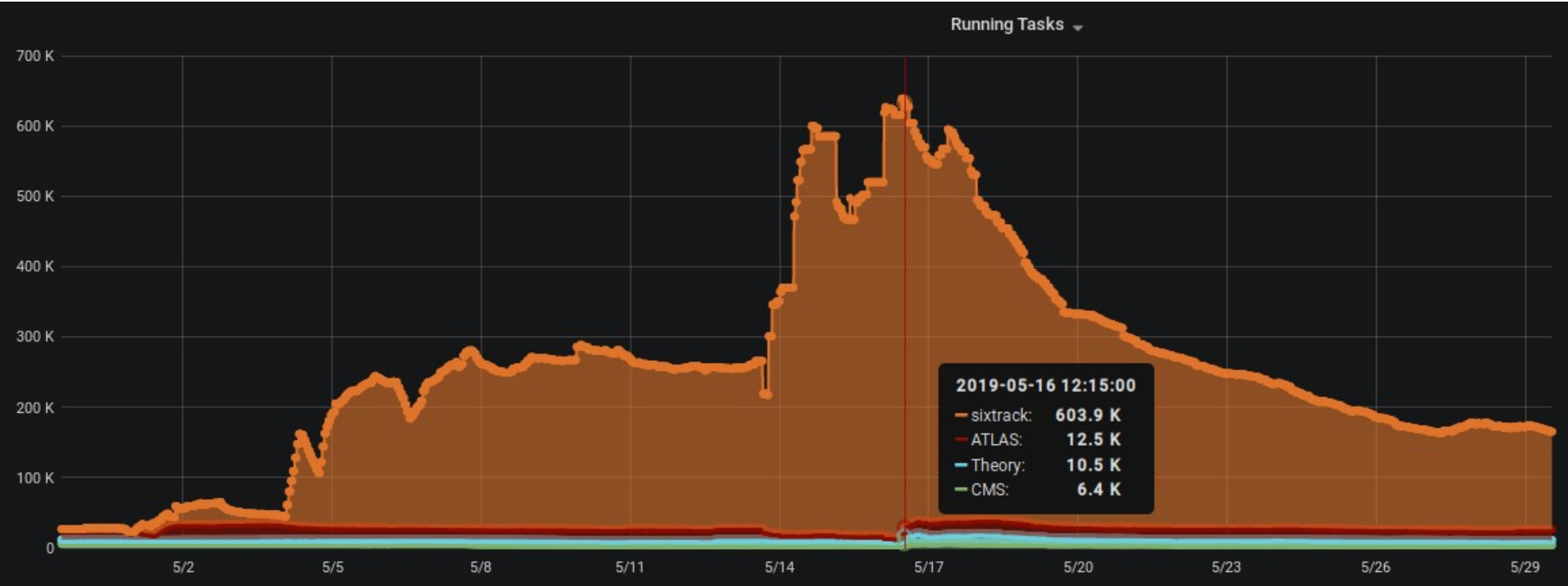
[JOIN US!](#)

[HELP & FAQ](#)

[CONTACT](#)



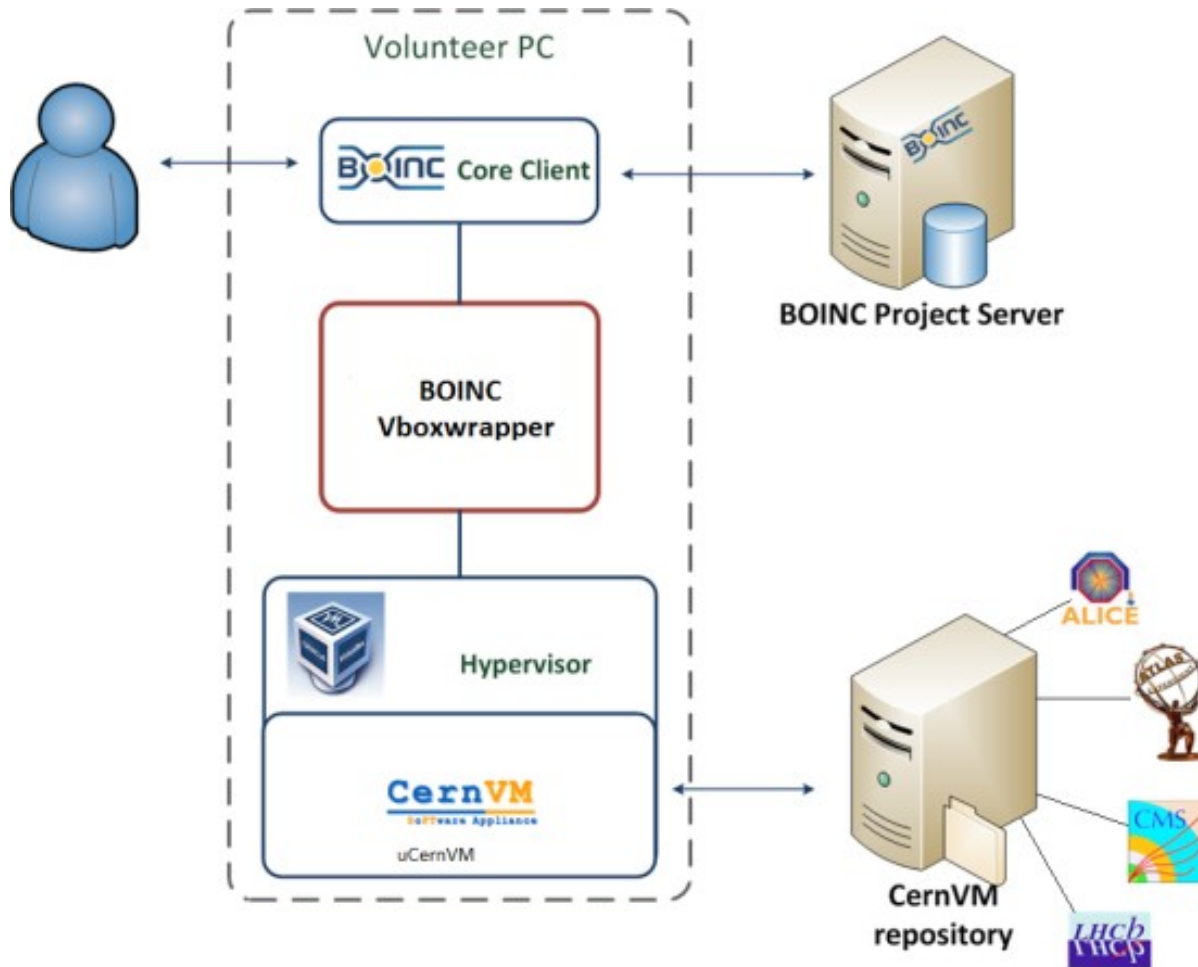
2019 Pentathlon



Challenges

- The cost of using the free resources
 - Initial integration requires investment
- Operations and Maintenance
 - Public facing support on all levels
 - Lowered by community supports
- Attracting and retention of volunteers
 - Advertisement and engagement
 - Communications cost for capacity building
- Low level of trust
 - Anyone can register as a volunteer
 - Not the same as internally owned systems
- Running scientific software on Windows
 - ~80% of the resources

BOINC with Virtualization



Regenerates

- Every 24 hours

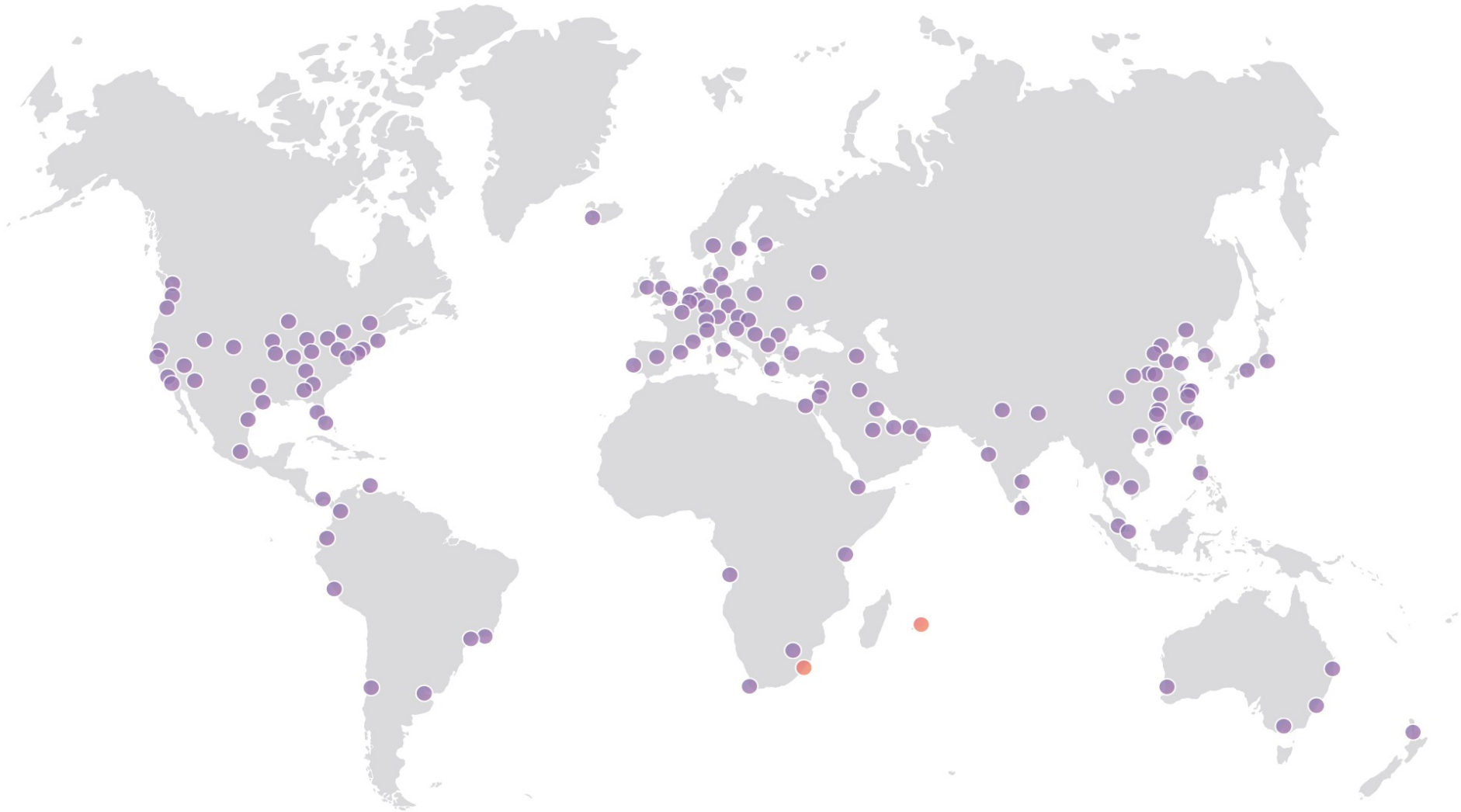
Bake image

- One-time download
- Auto-updates
- Cache decay

LHC@home Squids

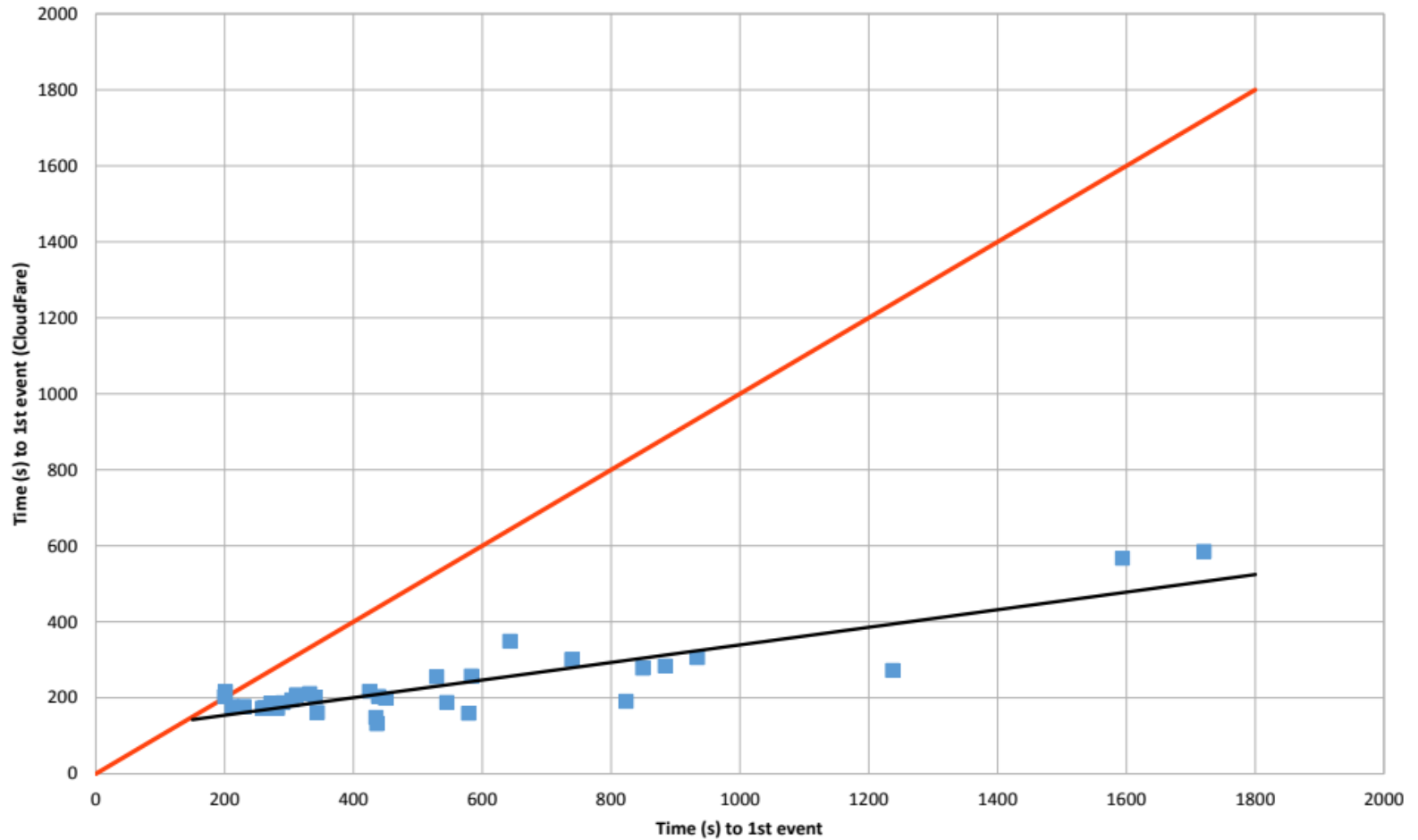
- Used to have two services
 - lhchomeproxy.cern.ch
 - lhchomeproxy.fnal.gov
 - Two machines behind each service
- Used CVMFS_PAC_URLS
 - <http://lhchomeproxy.cern.ch/wpad.dat>
- Geographical load balancing
 - Redirect to the nearest instance

CDN: Cloudflare

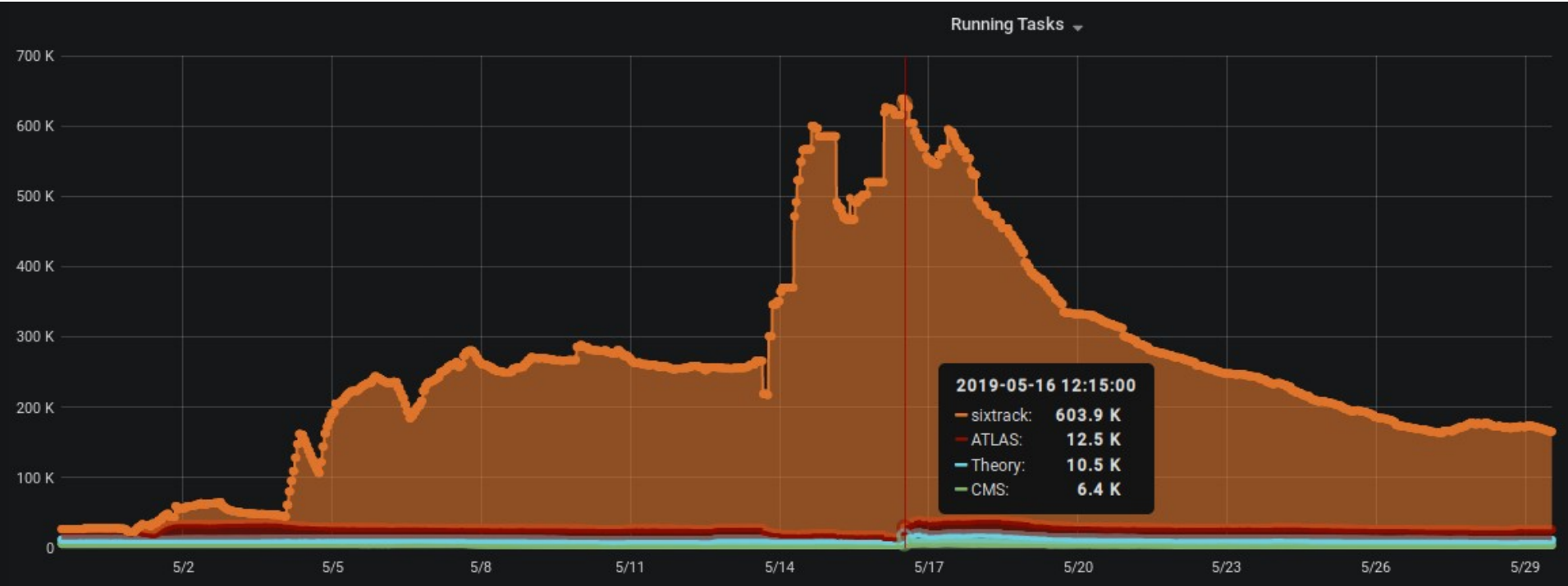


CMS@home Job Speedup

CloudFare



2019 Pentathlon



Native Application

- ATLAS using Singularity for Linux
 - Opportunistically on CERN AI and T2/3 sites
 - BOINC on Batch
- VM Capacity Gap
 - More Sixtack tasks being run than VM tasks
- More resources on Linux
 - Was 80%W, 10%L, 10%M
 - Now 67%W, 41%L, 2%M

Native Theory

- Install CVMFS locally
 - Contributed quickstart documentation
 - Volunteer Expert: Stefan Pavli
- Use runc rather than Singularity
 - One less dependency
- BOINC wrapper and Cranky
 - Used Bash over Python
 - Create filesystem
 - Links to CVMFS
 - Use runc from CVMFS
 - Checkpointing not working for rootless containers



Windows

- Still many resources
 - VMs required for experiment code
 - Not gaining traction
- Cache reset when rebooting VMs
 - Each task, assignment of credit
- What to do?
 - CVMFS on Windows?
 - Mount through VMs
 - Use cache file?
 - How to share between Vms
 - Local squid proxy?
 - Issues?
 - Other ideas?

Conclusion

- Volunteer Computing
 - Profit from spare computing capacity
- CernVM
 - Enables HEP applications to be run
- Cloudflare can be used as CDN
 - Improve performance for the volunteers
- Native Applications
 - CVMFS required for HEP applications
 - Containers working well
 - Checkpointing not there for rootless containers
- What to do about windows?