
21st International Conference on Computing in High Energy and Nuclear Physics **CHEP2015** Okinawa Japan: April 13 - 17, 2015

# Towards a production volunteer computing infrastructure for HEP

Nils Høimyr (CERN), Miguel Marquina (CERN)
Tomi Juhani Asp (University of Jyvaskyla, Finland),
Laurence Field (CERN), Alvaro Gonzalez (CERN), Pete Jones (CERN)

Presentation on behalf the CERN BOINC Service Team

# Why Volunteer Computing?

| Target | Deployment | Benefit |
|---|---|---|
| Volunteers | Uncoordinated, opportunistic | • Get additional, “free” compute cycles<br>• 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 (e.g: CernVM web-api talk at recent CernVM workshop)

CERN has adopted BOINC for VC projects (LHC@home)

# BOINC

## Berkeley Open Infrastructure for Network Computing

http://boinc.berkeley.edu

- Software platform for distributed computing using volunteered computer resources
- Client – server architecture
- Free and open source
- Used for:

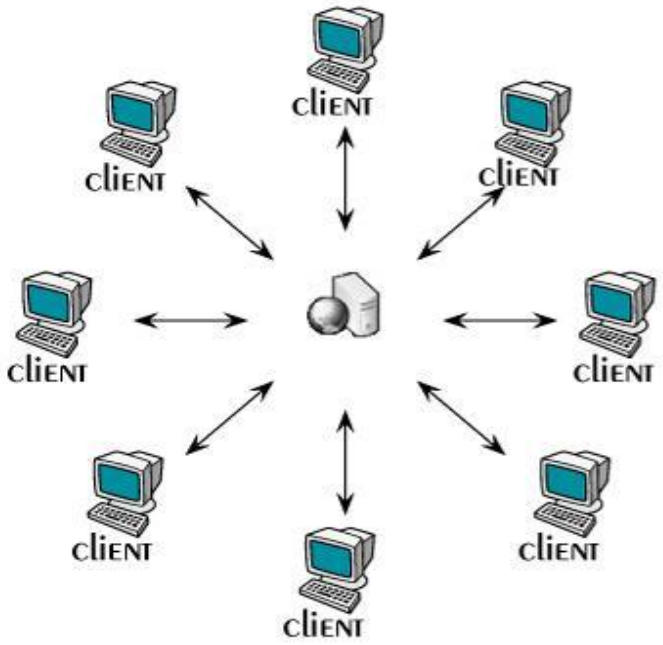
  SETI@home

  Climateprediction.net

  Einstein@home

  LHC@home

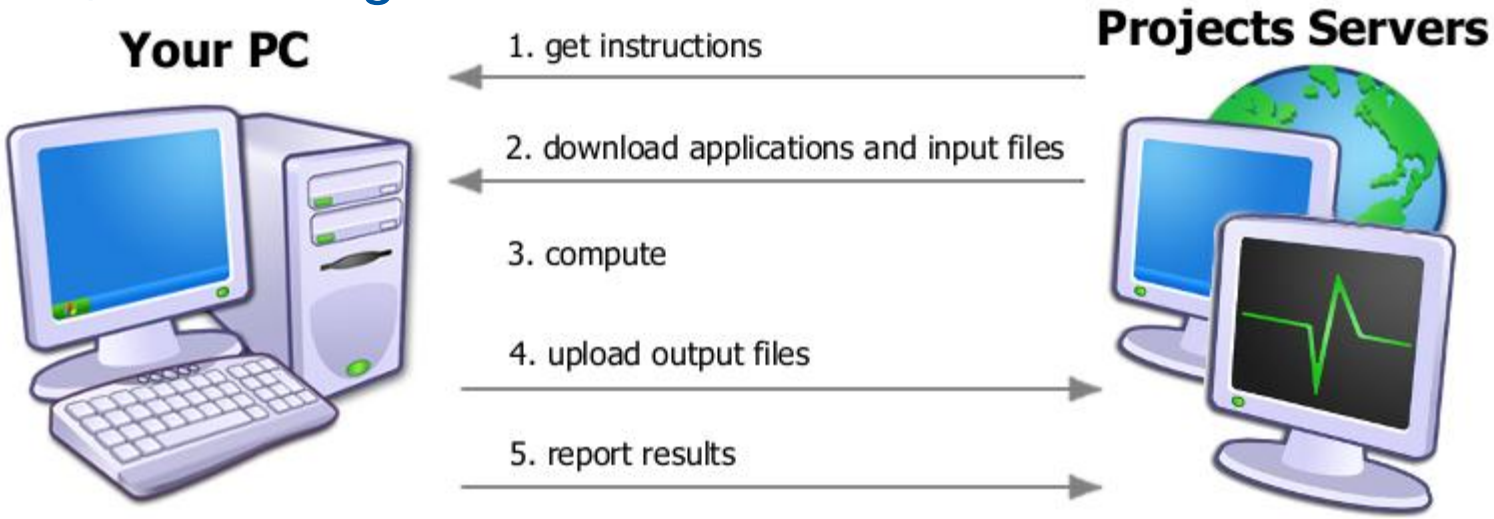
  …



# BOINC – Volunteer view

- Download and run BOINC client
- Choose a project
- Provide email and password to the BOINC Manager

  (alternatively make a silent connection with a key from the BOINC client)
- Done, crunching can start!



# 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 |

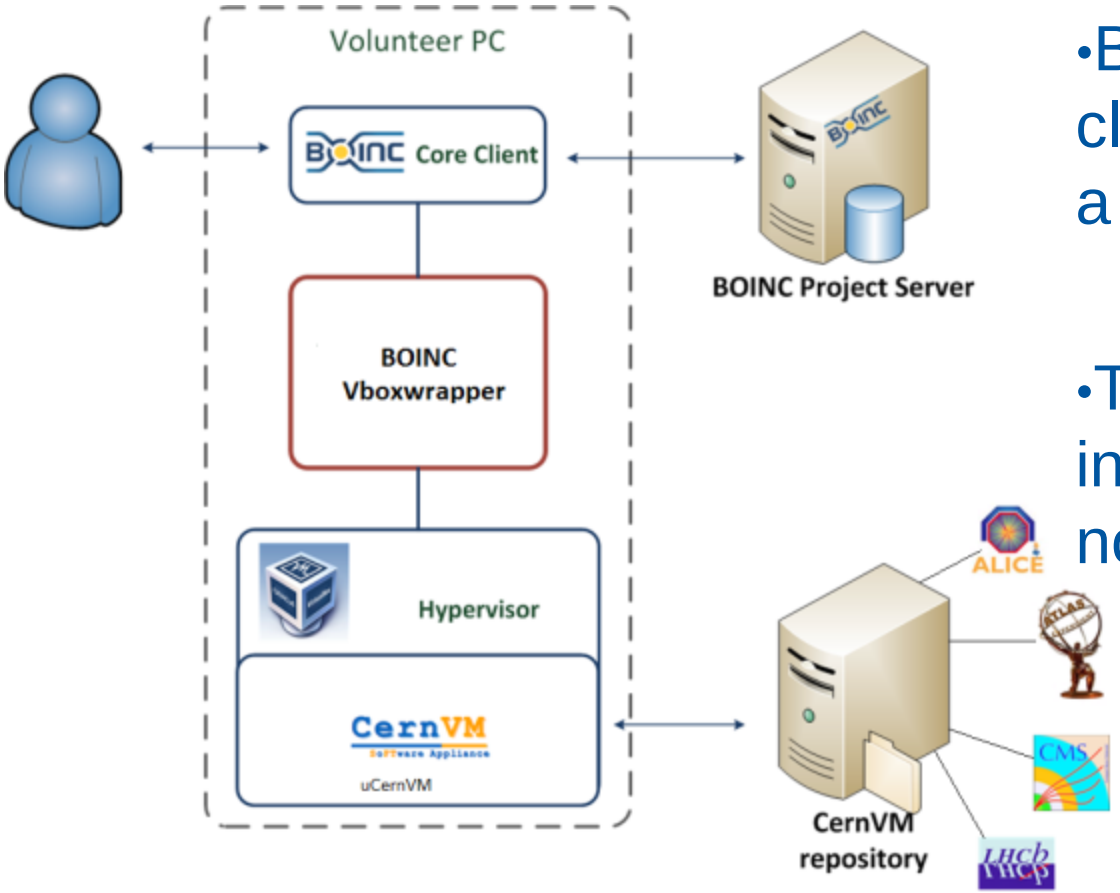
*According to BOINCstats.com 4.3.2015*

# Virtualisation in BOINC - 1

- Pioneered at CERN in 2010-2011 by Test4Theory and the CernVM team in PH/SFT
- Later brought into BOINC mainstream code as "Vboxwrapper"
  - Ref: *http://boinc.berkeley.edu/trac/wiki/VboxApps*
  - BOINC developers very helpful with improvements
- Besides CERN (Theory, Atlas, CMS, LHCb) there are several other BOINC projects now deploying Virtualisation:

  RNAword, Climateprediction.net, CAS@home

# Virtualisation in BOINC - 2




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

# BOINC: Classic vs Virtualisation

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

# LHC@home - Sixtrack

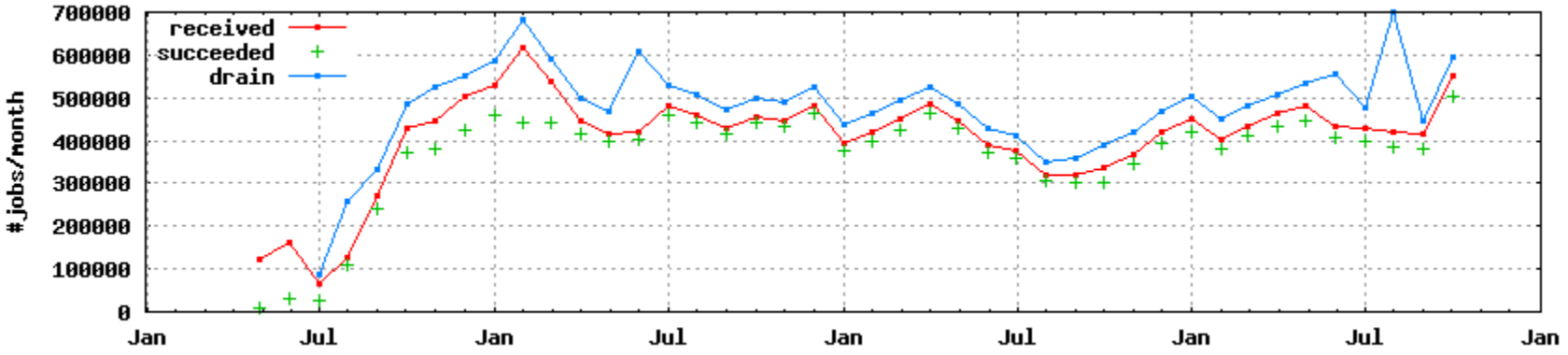

- Started as outreach project for CERN's 50th anniversary 2004, used for Year of Physics (Einstein Year) 2005
  - Based on experience from the Compact Physics Screen Saver (CPSS), which ran SixTrack on desktop computers at CERN
- Calculates stability of proton orbits in the LHC accelerator
- Written 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 - 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 (being phased out)
- 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




**vLHC@home jobs run per month (2011-2015)**

- 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 – LHC experiments

ATLAS

- started early 2014 as internal pilot, now public
- using µCernVM and virtualisation

CMS

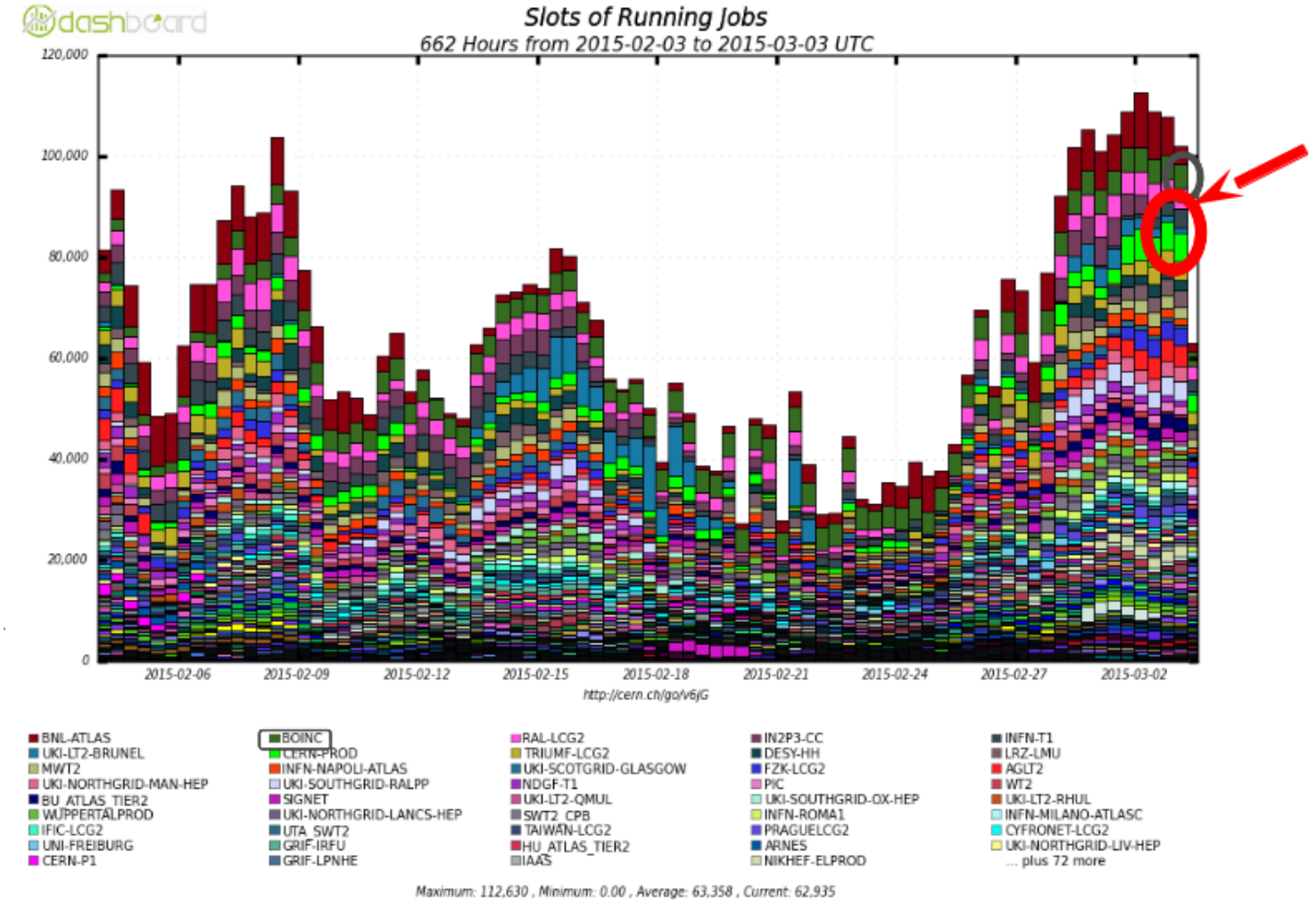
- started work in summer 2014
- prototype running, rapidly gaining experience

listen to the next two contributions in this track

LHCb (Beauty)

- prototyping started in 2012
- Currently fed by volunteers inside the collaboration

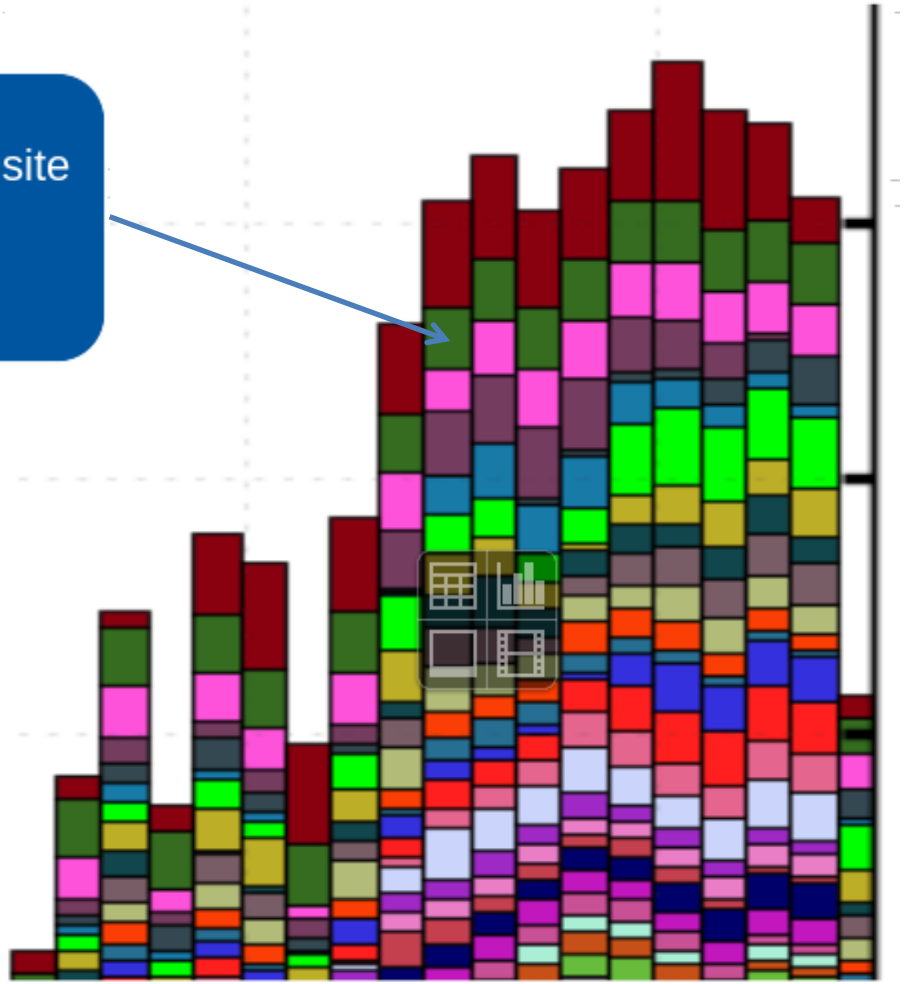
# BOINC contribution to ATLAS



# BOINC contribution to ATLAS

BOINC is the 2nd largest simulation site

Running 4-5k parallel jobs



BNL-ATLAS

BOINC

RAL-LCG2

IN2P3-CC

INFN-T1

UKI-LT2-BRUNEL

CERN-PROD

…

# CERN BOINC Service

## Server cluster

- LHC@home servers in production (Sixtrack, Theory, ATLAS)
- Test servers – used as dev/prototyping environments by the projects

## Server application support

- Configuration, monitoring
- MySQL database server back-end
- BOINC server application configuration and updates

## Website framework (http://cern.ch/LHCathome)

- common Drupal portal, as entry point for all BOINC projects and applications hosted at CERN

# CERN BOINC Service - 2

We are not involved in the R&D and outreach aspects specific to the projects. So the corresponding teams deal with:

- Porting of applications to BOINC
- Application-specific job management framework
- Communication/outreach with volunteers about science involved
- Management of user forums and project material in the portal

# CERN BOINC Service - 3

## Service evolution

- VM applications that report back to a local job management framework can be part of Virtual LHC@home
- Other (Sixtrack, ATLAS) are currently hosted on separate servers to avoid I/O bottleneck
- Aim for standardisation on a volunteer cloud common job management solution (Data Bridge, more at the CMS talk)

# BOINC - use cases for HEP

## 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 be hosted as part of LHC@home
- The size of the application data sets remains 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

# This is the joint work of many people…

- BOINC service: Nils Hoimyr, 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, Efrat Tal Hod, Wenjing Wu
- CMS: Laurence Field, Hendrik Borras, Daniele Spiga, Hassan Riahi, Ivan Reid
- LHCb: Federico Stagni, Joao Medeiros, Cinzia Luzzi et al
- BOINC: David Anderson, Rom Walton
- and many CERN colleagues offering the underlying layered services


CERN