CernVM@LHCb

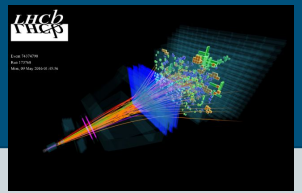
CernVM User Workshop

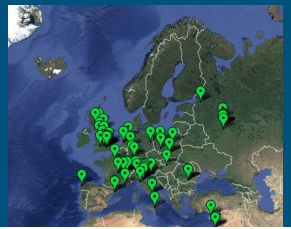
Ben Couturier (for the LHCb Computing team)

LHCb Collaboration

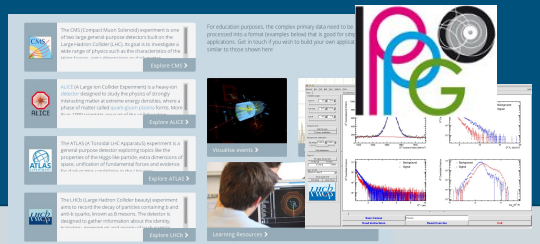
1174 Members,
69 Institutes,
16 Countries

Where do we use CernVM ?





DIRAC
THE INTERWARE



Trigger

Production

User Analysis

Data Preservation
Masterclasses

CernVM-FS Repositories

Production releases repository

/cvmfs/lhcb.cern.ch

- Software releases (Online/Offline)
  One release/day in avg
- Conditions (SQLite files)
  Released every hour

Stratum-0 managed by CERN-IT

No garbage collection

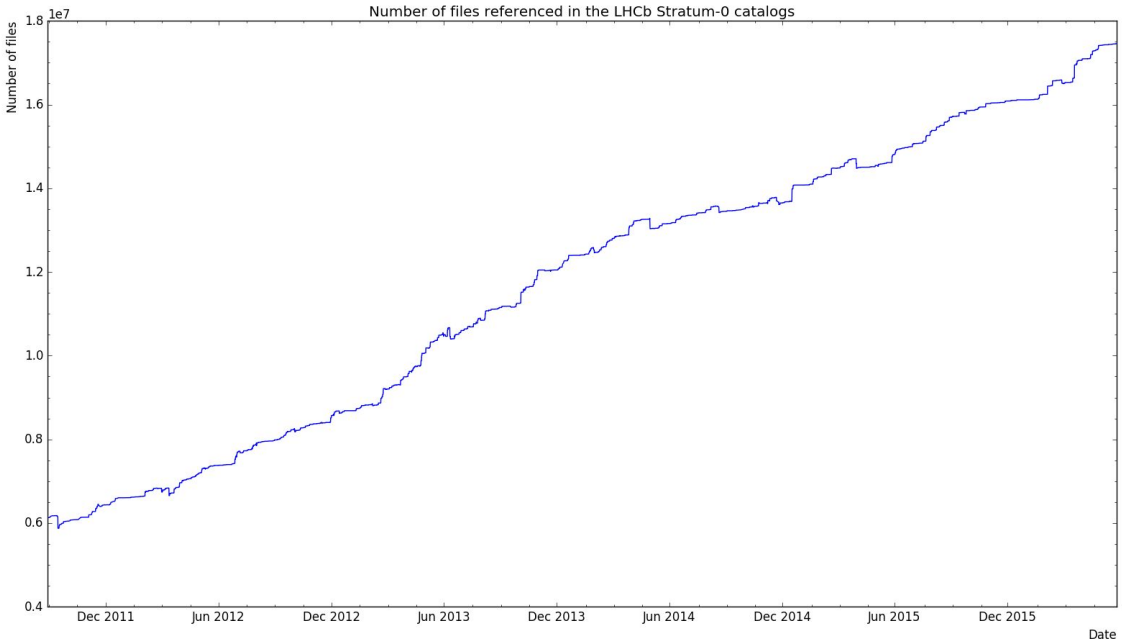
Repository for long term software preservation

Deduplication useful

Regular files	17560484
Directories	2415724
Symlinks	117468
Total file size	861.4 GB
Chunked files	13984
Chunked file size	231.9 GB
Number of chunks	56164
Nested catalogs	2263

* As of 30/05/2016

Production releases repository history



Tests with IT have shown we have a large margin

Continuous builds repository

/cvmfs/lhcbdev.cern.ch

- Continuous builds

Stratum-0 managed by CERN-IT

Garbage collected every night
(total size automatically under check..)

Short term interest

***Very* high level of duplication**

Release as soon as the builds are available

Regular files	26669728
Directories	4980768
Symlinks	2558
Total file size	2.1 TB
Chunked files	97131
Chunked file size	1.5 TB
Number of chunks	364864
Nested catalogs	294

* As of 30/05/2016

CernVM-FS Usage

CernVM-FS @LHCb Online

~1.6k hosts at LHC pit 8

LHCb trigger takes releases from cvmfs

But

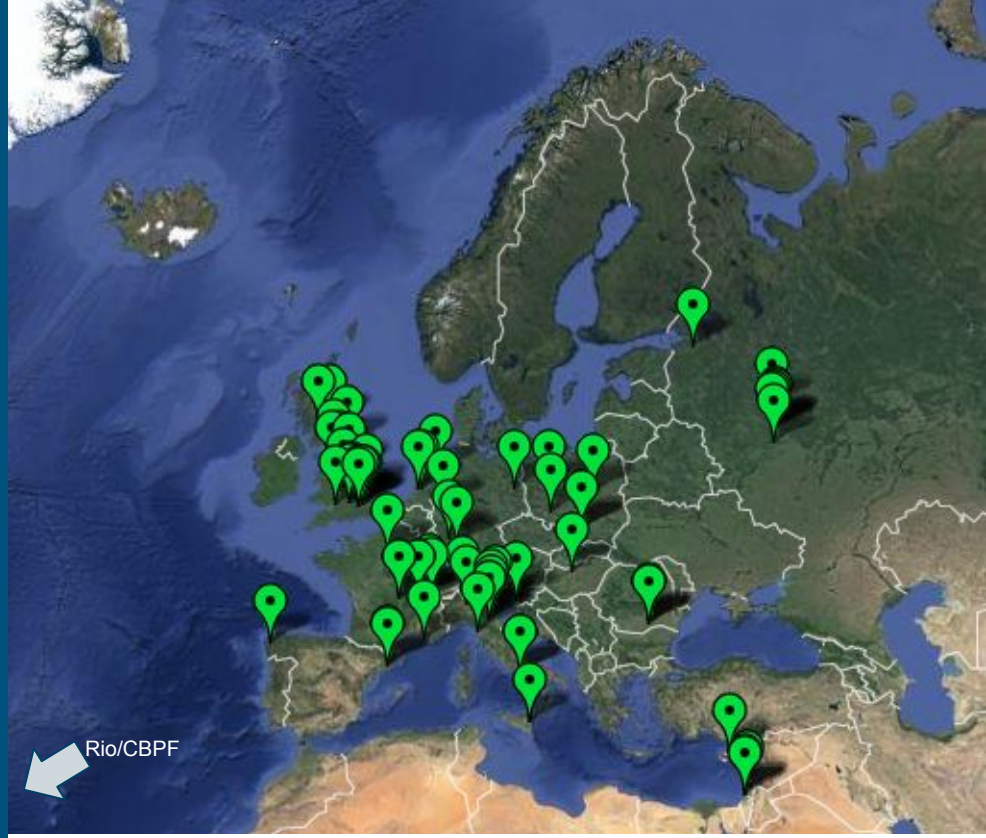
- NFS used for override
- Trigger images distributed via bittorrent

CernVM-FS
For Grid Production

CernVM-FS at all grid sites

(since 2013)

- Critical component for LHCb Production Data processing
- DIRAC Pilots can start from CernVM-FS (or download directly missing version)
- Used to propagate detector conditions (SQLite files)



DIRAC
THE INTERWARE

Feedback from production team?

VERY happy with the reliability of the system...

Feedback from Librarian

Happy but...

Still issues with the release process on stratum-0

Releasing is very time consuming
(exact bottlenecks to be investigated)

- ➔ Limiting factor for the nightly builds: some transactions take a couple of hours to publish in the morning
- ➔ This can be a problem when trying to get files to developers as early as possible

These points are becoming even more relevant due to upcoming AFS retirement at CERN

Monitoring & Control

Debugging problems is not always easy: all tools/tricks are welcome
E.g. tracing cache behaviour on the clients

For use on the trigger, we want to make sure we have no intempestive clear/reload of the cache
E.g. using features planned for HPC

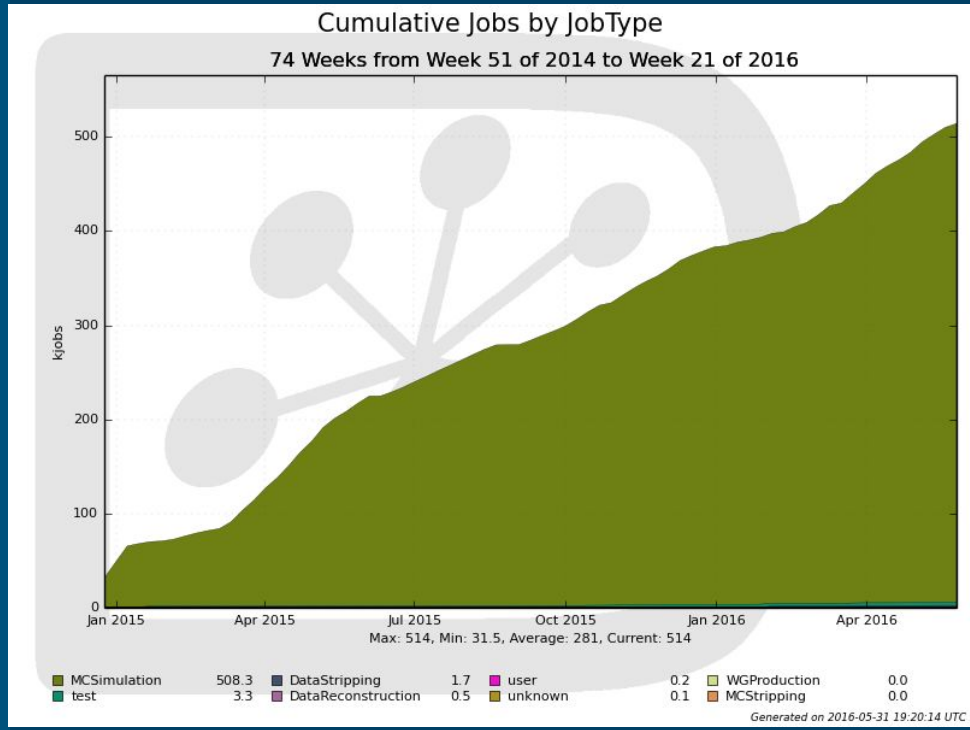
CernVM Usage

CernVM in Production

Used via VAC/VCycle
(c.f. http://iopscience.iop.org/article/10.1088/1742-6596/664/2/022030)

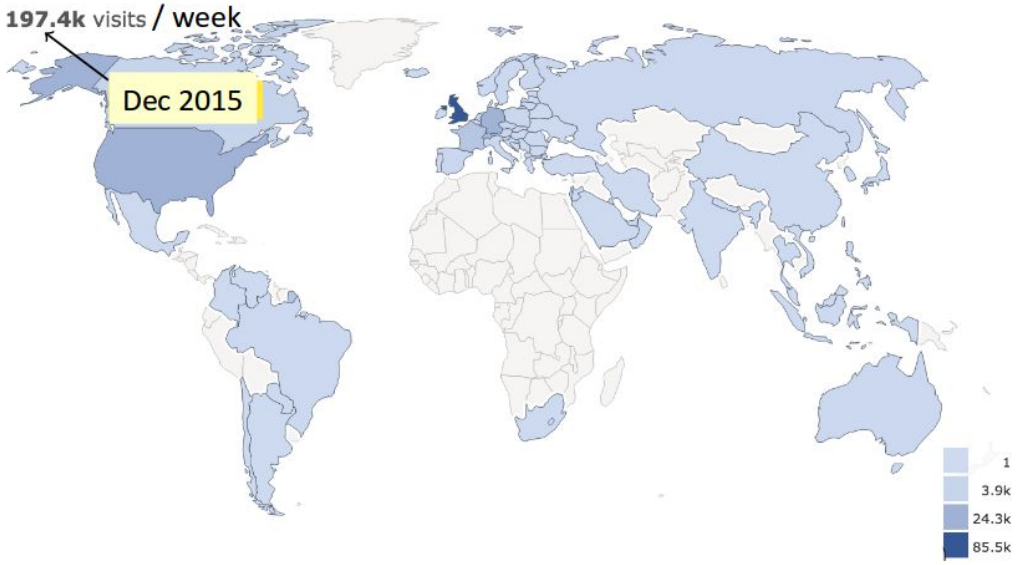
Small percentage of all jobs
(~1.5% since 01/2015)
...But growing

Very convenient to use:
OS in CVMFS means transparent deployment of security updates



Vac / Vcycle

CernVM Virtual Appliance

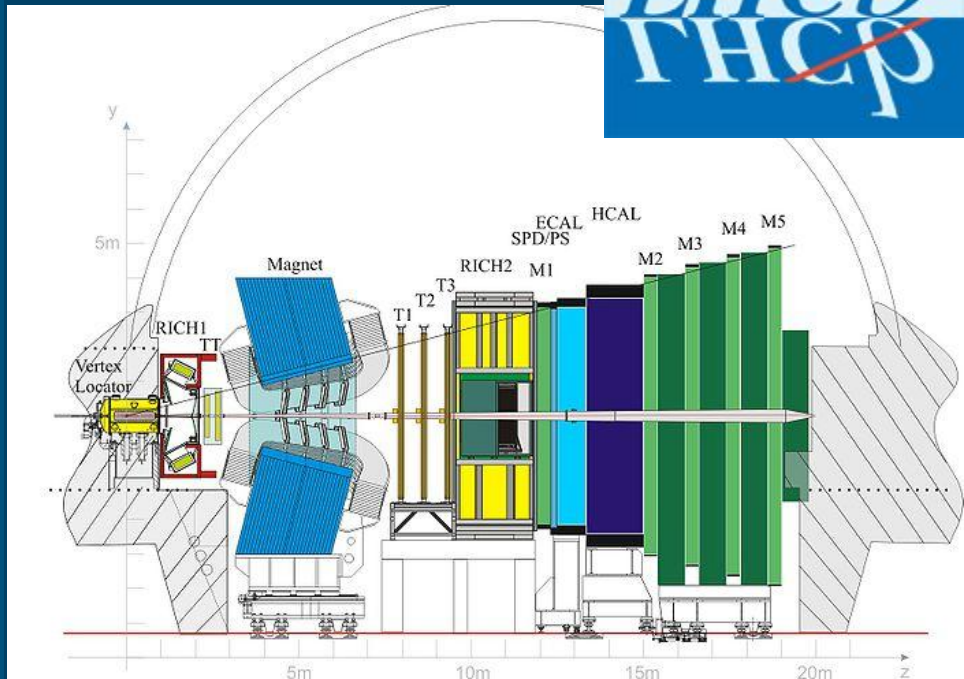


Boot of new CernVM
• ~30000 / day, ~75 countries (~40% UK)

Volunteer Computing

http://lhcbathome.cern.ch/Beauty/

- CernVM 3 with BOINC
- Same contextualization as production



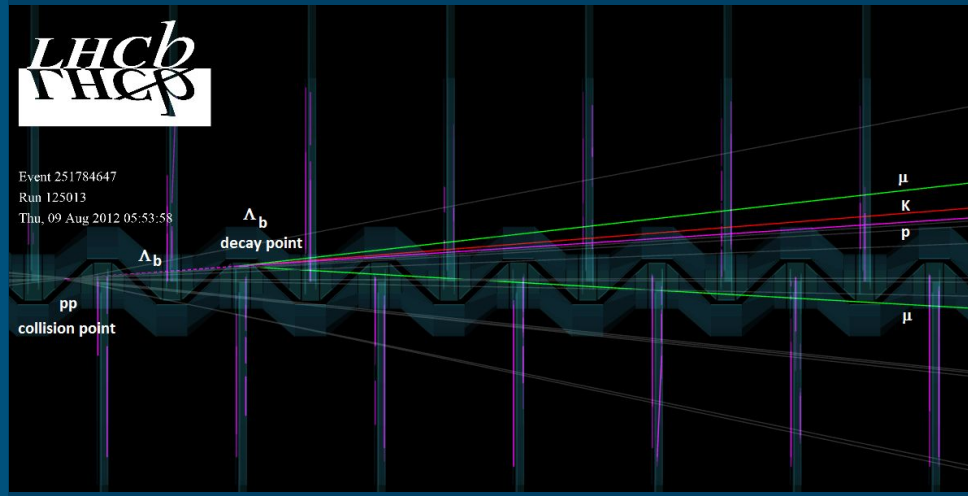
CernVM for Data Analysis & Development

Allows LHCb Users to:

- run the LHCb Data Analysis stack on their laptops (e.g. under MacOS)
- Develop new algorithms

Many happy testimonies, however:

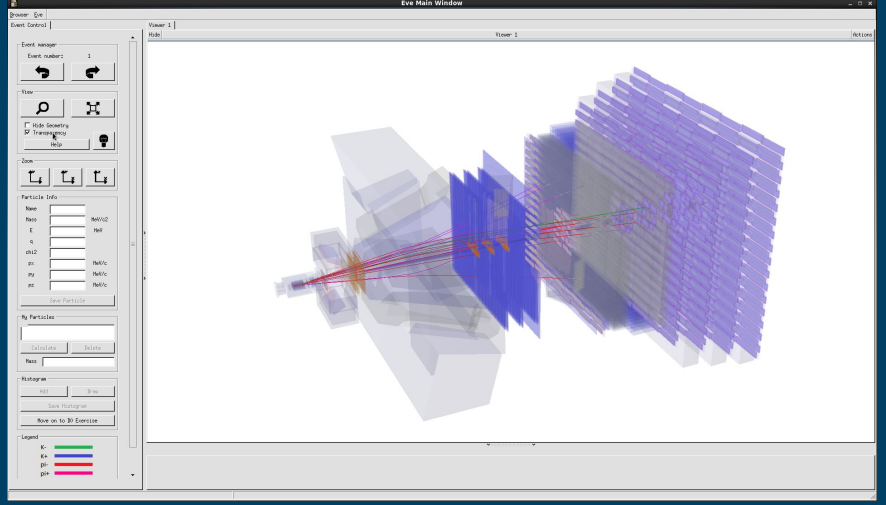
Integration with CERN Infrastructure (AFS/EOS) not so easy.
CernVM not in OpenStack default images...

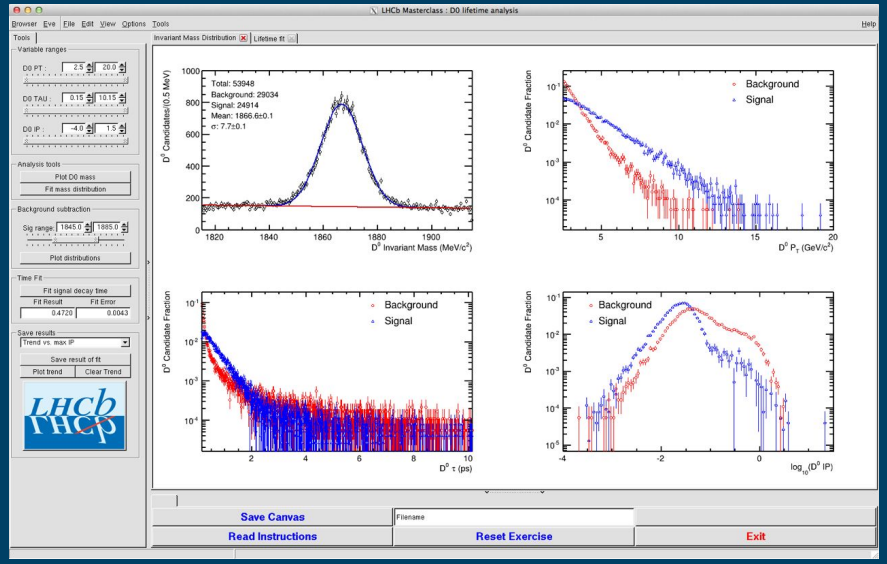


Outreach

- LHCb Masterclasses released in custom ova image (thanks Jakob)
- CernVM is the default platform for OpenData releases

http://opendata.cern.ch/VM/LHCb





Data Preservation

Will need to keep running:

- ➔ SLC5
- ➔ SLC6
- ➔ Centos7

CernVM ideal to fulfill that role

Conclusion

CernVM/FS crucial to LHCb **production** and **Data Preservation**.

LHCb Happy with production experience.

CernVM/FS are core tools for LHCb

CernVM-FS is part of the long term strategy for software distribution

And its role will be even more central once AFS is retired, meaning that improvements are needed to:

- allow for faster releases on stratum-0
- Speed-up propagation time through stratum-1

CernVM is an ideal tool for long term data preservation
(we need to make sure we can run our use cases)

Many thanks !

- To Jakob, Rene, Gerri and all involved in developments
- To Steve, Dan from CERN-IT for running the service