



Software Preservation – Introduction

Jakob Blomer

DPHEP Workshop 2017
March 13th, 2017



- Porting and validation for running experiments
- **Freezing** for decommissioned experiments
- Freezing technologies
 - Emulation: bridging hardware architectures
e.g. might become relevant again for x86_64 on Power / ARM
 - Virtualization + Containers: sandboxing execution environments
e.g. SLC 4 on CentOS 7
- Confident to bridge 15+ years with freezing
 - Demonstrated Aleph SLC 4 software in CernVM
 - CMS 2010 SLC5 CernVM
- Current work focused on
 - Identifying and removing external dependencies (e.g. databases)
 - **Understanding relevant constituents** of software stack;
how to move beyond “capture the mess”



CernVM (Container or VM)

- Curated Linux platform with all dependencies to LHC applications
- All base packages and build recipes archived
- Strongly versioned (CernVM-FS)
- Graphical (development environment) and batch flavors

CernVM-FS

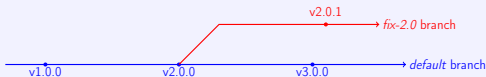
- Production system for LHC software distribution
- DP capabilities included: **versioning and snapshotting**

```
↑ Jakob — cernvm@cernvm002:~ — ssh — 109x36
cernvm@cernvm002:~$ sudo cvmfs_server ltags cernvm-prod.cern.ch
NAME | HASH | SIZE | REVISION | TIMESTAMP | CHANNEL | DESCRIPTION
cernvm-system-3.1.0.0 | fb17e39ca21729a9509fe836fc7f30d26cae1c82 | 14kB | 11 | 28 Jan 2014 14:31:17 | 0 |
cernvm-system-3.1.1.0 | d855c3c05e4fcd9d5c6f1d0b08c74094f4f5008 | 14kB | 13 | 30 Jan 2014 00:11:10 | 0 |
cernvm-system-3.1.1.1 | 3a06202aad3b3163b9c5bd36f48b25744f3f204 | 14kB | 16 | 5 Feb 2014 21:03:00 | 0 |
cernvm-system-3.1.1.2 | fc2faf3bc87a2f74da7db22525189b5c582975de | 14kB | 18 | 16 Feb 2014 13:01:32 | 0 |
cernvm-system-3.1.1.3 | fc0d2515c9e79f9fd3cf8b01eac0a16746f4f6cb | 14kB | 20 | 4 Mar 2014 09:26:27 | 0 |
cernvm-system-3.1.1.4 | 314d93015ce473d9a6c99a7365dd4ce38b4e7b13 | 14kB | 22 | 17 Mar 2014 11:07:02 | 0 |
HEAD | 314d93015ce473d9a6c99a7365dd4ce38b4e7b13 | 14kB | 22 | 17 Mar 2014 11:07:10 | 0 |
```



New features under construction

- 1 **Exposed Access to Snapshots.** A new virtual directory that provides access snapshot access within a single mountpoint, like `/cvmfs/alice.cern.ch/.cvmfs/snapshots/version-1.0`
- 2 **Branching.** Support for hotfixes of historic execution environments



- 3 **Snapshot Diffs.** Show change set between any to snapshots

```
$ cvmfs_server diff alice.cern.ch v1.0 v2.0
M /changelog (File)
M /latest (Link)
A /v2.0 (Directory)
R /externals/unused-lib (Directory)
```



Today's focus: Umbrella

Backup Slides



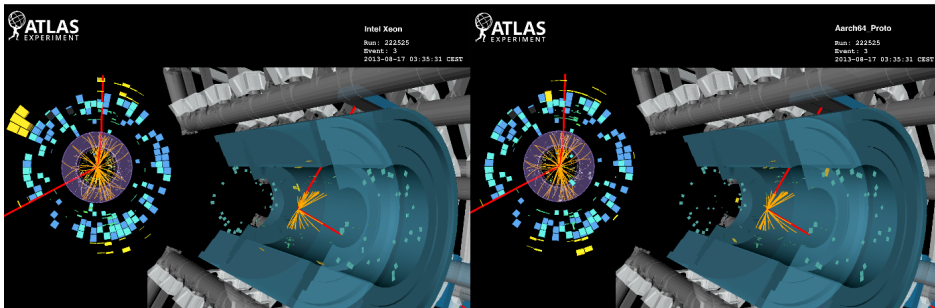
Use Cases for a Preserved Software Environment

- ① Processing of legacy data
 - Software implicitly encodes knowledge about the correct interpretation of the data
 - **After substantial upgrades** and modifications of the detector, the new software might lose this legacy knowledge
 - **After experiment decommission**, porting and validation of software is likely to end
 - Porting and validation will at some point become prohibitively expensive or just impossible
- ② Validation of new software versions
 - Comparison with historic version provides input for validation
- ③ Stable environment for education (cf. CMS Open Data Pilot)
 - Stable operating system and experiment software version accompanies “open data” set and well-defined analysis tasks
 - **Driver for data preservation:**
 - Opportunity to streamline data format and documentation
 - Disentangle from grid environment



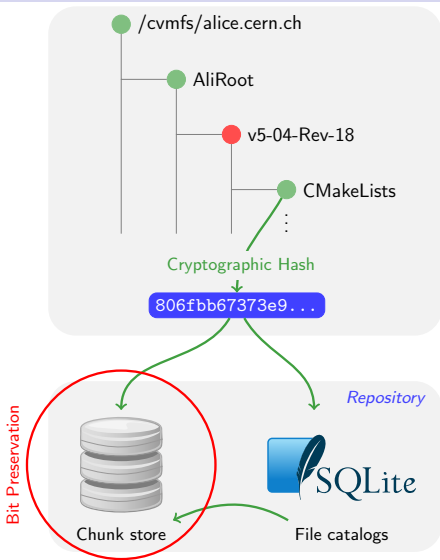
INTEL XEON

AARCH64_PROTO





Versioning and Snapshots in CernVM-FS



Data Store

- Eliminates duplicates
- Archiving

File Catalog

- Directory structure, symlinks
- Content hashes of regular files
- Digitally signed
- Plain files

Reduces preservation of software environment to bit preservation



ALEPH software on CernVM / SL4

```
pb-d-128-141-134-74:~ jakob$ ssh -X aleph@cernvm-aleph01
aleph@cernvm-aleph01's password:
[aleph@cernvm-aleph01 ~]$ source setaleph.sh
[aleph@cernvm-aleph01 ~]$ cd test/ALPHA/
[aleph@cernvm-aleph01 ALPHA]$ sh alpha.sh
*****
*****          ALPHA RUN          **** 11.6 ****
*****
*****

Wed Mar 19 16:10:27 CET 2014

*****
***   Compilation and creation of the makefile 6lep.mk
*****
gmake -f /home/aleph/test/ALPHA/6lep.mk
gmake: `6lep' is up to date.
```



CMS Open Data Pilot on CernVM / SL5

