Advances in software management tools

Jakob Blomer, CERN, EP-SFT
LHCP 2021
9 June 2021
LHC Experiment Software Stacks and Targets

Daily builds of hundreds of interdependent packages

- Analysis
- Simulation
- Reconstruction

HEP Libraries:
- ROOT, Geant4, DD4hep, Generators, ...

Data Access:
- XRootD, Davix, Arrow ...

External Libraries:
- BLAS, Tensorflow, ...
- C++ Compiler, Python

Distribution to $\mathcal{O}(1\text{million})$ heterogeneous cores

- Workstations
- Web (Jupyter)
- Supercomputers

WLCG & HLT
Daily builds of hundreds of interdependent packages

Distribution to $\mathcal{O}(1\text{million})$ heterogeneous cores

Analysis, Simulation, Reconstruction

Experiment Software Framework: EDM, Geometry, Conditions, ...

Data Access: XRootD, Davix, Arrow ...

External Libraries: BLAS, Tensorflow, ...

C++ Compiler, Python

Example from LCG software stack combinatorics:
500 packages $\times$ 5 operating systems $\times$
5 compilers $\times$ Python$\{2,3\} \times \{\text{opt, dbg}\}$
Compared to run 1-2, we now find

- Multiple target architectures: x86_64 micro-architectures (e.g. AVX512), AArch64, Power, GPUs
- A growing Python software ecosystem, in particular for machine learning tasks
- More agile software development: automated integration builds, nightly builds
- Generally we tend to add code and externals more often than removing software

My estimate: the software management problem for HL-LHC grows by a factor of 3-5.
Performance engineering example: improving CernVM-FS write performance with Ceph/S3 (vCHEP’21)

- Investment needed in the **performance** of software build, test, and distribution tools

- Constructing software stacks becomes an **engineering discipline** ("software librarian")
  - Tendency to build complete stack including OS layer
    → independent from target node OS
  - Dedicated projects for turnkey stacks
    - LCG
    - Key4HEP
  - Significant effort is going into adopting Spack package manager for maintaining HEP software stacks

- Preserving stacks gets harder: preservation tools are available but we risk to "capture the mess"
[jblomer@lxplus]$ source \\
/cvmfs/sft.cern.ch/lcg/views/LCG_100/x86_64-centos7-gcc10-opt/setup.sh
[jblomer@lxplus]$ root --version
ROOT Version: 6.24/00
Built for linuxx8664gcc on Apr 14 2021, 14:33:50
From tags/v6-24-00@v6-24-00
[jblomer@lxplus]$ python -V
Python 3.8.6
[jblomer@lxplus]$ python -c "import tensorflow"
...
Software Distribution: Status of CernVM-FS

Software, containers, conditions data for LHC, LIGO, EUCLID, LSST, EESSI, and many others

Available in the default configuration:
\sim 1.4 \text{B files} \\
\sim 150 \text{repositories}
Software Distribution: Status of CernVM-FS

Software, containers, conditions data for LHC, LIGO, EUCLID, LSST, EESSI, and many others

All software anywhere anytime

Available in the default configuration:
~ 1.4 B files
~ 150 repositories

jblomer@cern.ch

Software Management / LHCP’21
Containers for LHC Experiment Applications

- Decouples application from node operating system
- Solves the performance overhead problem of VMs
- Convenient to compose new container images based on existing layers → excellent tool to explore new software
- Allows for capturing a reproducible software environment

- Poorly addresses the distribution problem at scale
- Facilitates a “black box” approach that impedes a proper understanding of the software stack inside

A smart approach to containers combines their isolation capabilities with the distribution efficiency and well-maintained content of CernVM-FS
CernVM-FS as a Container Hub

- Centralized image management
- Vulnerability scan
- OCI artifact signing
- Aims at service deployment

Current whitelist & pull model for publishing images on CernVM-FS to be replaced by automatic image conversion based on webhook triggers.

- Efficient storage & distribution
- Enables large-scale image deployment
- Aims at scientific apps

push image

webhook notification

unpack image

jblomer@cern.ch

Software Management / LHCP’21
CernVM-FS as a Container Hub

<table>
<thead>
<tr>
<th>/cvmfs/unpacked.cern.ch</th>
<th>/cvmfs/singularity.opensciencegrid.org</th>
</tr>
</thead>
<tbody>
<tr>
<td>• &gt; 1000 images</td>
<td>• &gt; 630 images</td>
</tr>
<tr>
<td>• &gt; 6.5 TB</td>
<td>• &gt; 2.5 TB</td>
</tr>
<tr>
<td>• &gt; 95 M files</td>
<td>• &gt; 60 M files</td>
</tr>
</tbody>
</table>

Images are readily available to run with singularity, including **base operating systems**, **experiment software stacks**, **explorative tools** (ML etc.), **user analyses**, and **special-purpose containers** such as **folding@home**

```bash
[jblomer@lxplus.cern.ch]$ singularity exec
'/cvmfs/unpacked.cern.ch/registry.hub.docker.com/library/debian:stable'

  cat /etc/issue

Debian GNU/Linux 10
```

`jblomer@cern.ch`
## Container Runtime Integration

<table>
<thead>
<tr>
<th>Runtime</th>
<th>Type</th>
<th>CernVM-FS Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singularity</td>
<td>flat (+ layers)</td>
<td>native</td>
</tr>
<tr>
<td>podman</td>
<td>layers</td>
<td>native (use image storage from /cvmfs)</td>
</tr>
<tr>
<td>containerd¹ / k8s</td>
<td>layers</td>
<td>plugin ▶️ remote snapshotter</td>
</tr>
<tr>
<td>docker</td>
<td>layers</td>
<td>“graph driver” image storage plugin²</td>
</tr>
</tbody>
</table>

Documentation chapter on containers & CernVM-FS:

¹ Requires containerd version >= 1.4
² Expected to be replaced by containerd as foundation of docker
De-duplication works properly only on file-level granularity

CernVM-FS exploits that only a tiny fraction (~ 2%) of images are used at runtime
Integration of Supercomputers (HPC Sites)

During the past years, a number of CernVM-FS HPC add-ons facilitated HPC deployments:

- Preloader: work around compute nodes without Internet connectivity
- Tiered cache: work around diskless compute nodes
- Unprivileged mounting of /cvmfs with `cvmfsexec`
- Integration with HPC container workflows through singularity `fusemount` plugin

Upper cache layer in WN Memory

Lower cache layer on `/gpfS/`
During the past years, a number of CernVM-FS HPC add-ons facilitated HPC deployments:

- **Preloader**: work around compute nodes without Internet connectivity
- **Tiered cache**: work around diskless compute nodes
- **Unprivileged mounting of /cvmfs with cvmfsexec**
- **Integration with HPC container workflows through singularity fusemount plugin**

Onboarding a new HPC site may need some custom tweaks. There is experience in the CernVM-FS development team. Please get in touch.

Onboarding a new HPC site may need some custom tweaks. There is experience in the CernVM-FS development team. Please get in touch.

jblomer@cern.ch
Software Management / LHCP’21
SWAN: Jupyter Notebook Service for HEP

Data analysis environment available from all devices, anywhere
Computational resources in (academic) clouds
Great for teaching, tutorials, exploring
200-250 users per day on SWAN

2 Displaying graphics

We can now draw the histogram. We will first create a *TH1*, the entity which in ROOT holds graph primitives. Note that thanks to *Jupyter*, this is not a static but an interactive visualization. Try to play with it and save it as image when you are satisfied.

```python
In [5]:
c = ROOT.TH1 canvas()
    .Draw()
    .Save()
```

My Histos

- Data analysis environment available from all devices, anywhere
- Computational resources in (academic) clouds
- Great for teaching, tutorials, exploring
- 200-250 users per day on SWAN

jblomer@cern.ch
SWAN: Jupyter Notebook Service for HEP

- SWAN presents the familiar environment as a notebook
- A selection of software stacks readily available
- Data access via EOS home folder
- Can scale out computation to Spark
- Can be deployed on-premise with Science Box
• Like our other LHC computing problems, the problem of software stack construction and distribution is growing in scale
  
  • We need to invest in the engineering of software construction
  • We need to invest in the performance of the distribution (both publishing and reading)

• We should be smart about combining industry standard tools with HEP tailor-made tools (e.g. CernVM-FS and Docker or Jupyter and EOS)

• After several years of effort, harnessing HPC environments is now a reality; still some case-by-base adjustments needed

• Jupyter notebooks—integrated with the HEP ecosystem—have matured as very useful tools for teaching and outreach
Backup Slides
Container ImageSizes

Distribution of container images sizes in
/cvmfs/unpacked.cern.ch and /cvmfs/singularity.opensciencegrid.org

Entries

833

Mean

1.473

Std Dev

1.472

unpacked

4

−3 −2 −1 −0 1 2 3 4 5 6 (size [GB])

2

log

0

20

40

60

80

100

120

140

160

180

Number of images

128M

512M

2G

8G

32G

Entries

573

Mean

1.49

Std Dev

1.502

singularity

unpacked.cern.ch

singularity.opensciencegrid.org

128M

512M 2G 8G 32G