A complete, Multilanguage Analysis Ecosytem

E. Guiraud, D. Piparo (CERN, EP-SFT)

▶ The problem we are trying to solve

▶ The LCG releases

- Usage by experiments, experimental communities, engineers

▶ Creation of a HEP analysis ecosystem

Objectives:
1) Describe a proposal to serve an analysis ecosystem, complete and easy to use, to scientists
2) Collect feedback, e.g. wrt other packaging systems or solutions

- ▶ Originated from the Software Process and Infrastructure project

  - Dates early days of LCG Application Area (~2002)

- ▶ A collection of ~400 packages, coherently built: see lcginfo.cern.ch

- ▶ Several languages: Java, Python (2&3), C++, Fortran

  - Examples: ROOT, Spark, Pandas, Herwig, PyTimber

- ▶ Different os+compiler combinations (platforms) supported

  - E.g. gcc7-dbg-cc7, gcc5-opt-slc6 …

- ▶ Distribution vector: CVMFS - `/cvmfs/sft.cern.ch/`

*See P. Mendez, Building, testing and distributing common software for the LHC experiments, Hall 3 12th July 11:00 !

▸ Releases identified by a number: LCG_91, LCG_92, LCG_93

- ● A "global tag"

▸ 4-5 / year

▸ "Nightly releases": every night

- ● A new package (version) is agreed: 1-2 days after it's available

▸ Content, timescales, requests, issues discussed at the *Librarians and Integration Meeting*

- ● Every 2 weeks

Description:

Description:	Daily publication of the 93c nightly build
Release date:	Jun 13, 2018
Platforms:	• x86_64-slc6-gcc62-opt • x86_64-slc6-gcc62-dbg • x86_64-centos7-gcc62-opt • x86_64-centos7-gcc7-dbg • x86_64-slc6-gcc7-dbg

Release Notes

Packages:

Databases

cx_oracle	5.1.1
genshi	0.7
mysql	5.7.20
mysql_python	1.2.3
oracle	11.2.0.3.0
sqlite	3210000

Graphics

coin3d	3.1.3p2
graphviz	2.28.0

lcgdmcommon	1.8.5-1
lcginfosites	3.1.0-3
voms	2.0.9-1

IO

CASTOR	2.1.13-6
COOL	3_2_0
CORAL	3_2_0
Frontier_Client	2.8.19
xrootd	4.8.2

Math

AIDA	3.2.1
blas	20110419
fastjet	3.3.0
fftw	3.3.4
GSL	2.1
lapack	3.5.0
libsvm	2.86
sympy	1.1.1

[…]

5

How Can LCG Releases be Used Today?

▶ From LXPLUS/LXBATCH at CERN

- Source a simple script

▶ From SWAN

- CERN's Jupyter Notebook Service: swan.cern.ch
- Lightweight, single container image, all software coming from LCG_Releases

▶ From any machine, provided that the os is supported and it mounts cvmfs

- This is a limitation, see following slides.

- ▶ Experiments: Atlas and LHCb
  - Base of their production stack
  - Packages versions are fixed during data taking
- ▶ CERN Beams Dept.:
  - Especially for the Python packages
- ▶ SWAN users
  - Beam physicists, particle physicists, accelerator engineers, IT specialists
  - Future LHC logging system, NXCALS, based on LCG releases
  - LCG also used for environment on Spark clusters accessible via SWAN
- ▶ "Individuals", e.g.
  - Students in experiments (no need to compile, full suite, evolving more rapidly than experiment's data taking stack)
  - Scientists with little computer knowledge under heavy time pressure (e.g. machine development periods)

Can we Increase the Pool of Users?

Reach the HEP analysis community with LCG releases

Potential advantages for analysers

▸ No need to install/compile
- Just source a script
- Stop fiddling around with self installed software

▸ Latest versions of most useful packages (e.g. ROOT, Py*)
- Much newer than versions provided in stable stacks used for data taking

▸ Reproducibility
- "Source this environment, run this python script"
- Build analysis projects based on the same stack

▸ Interoperable with other interfaces (e.g. SWAN)
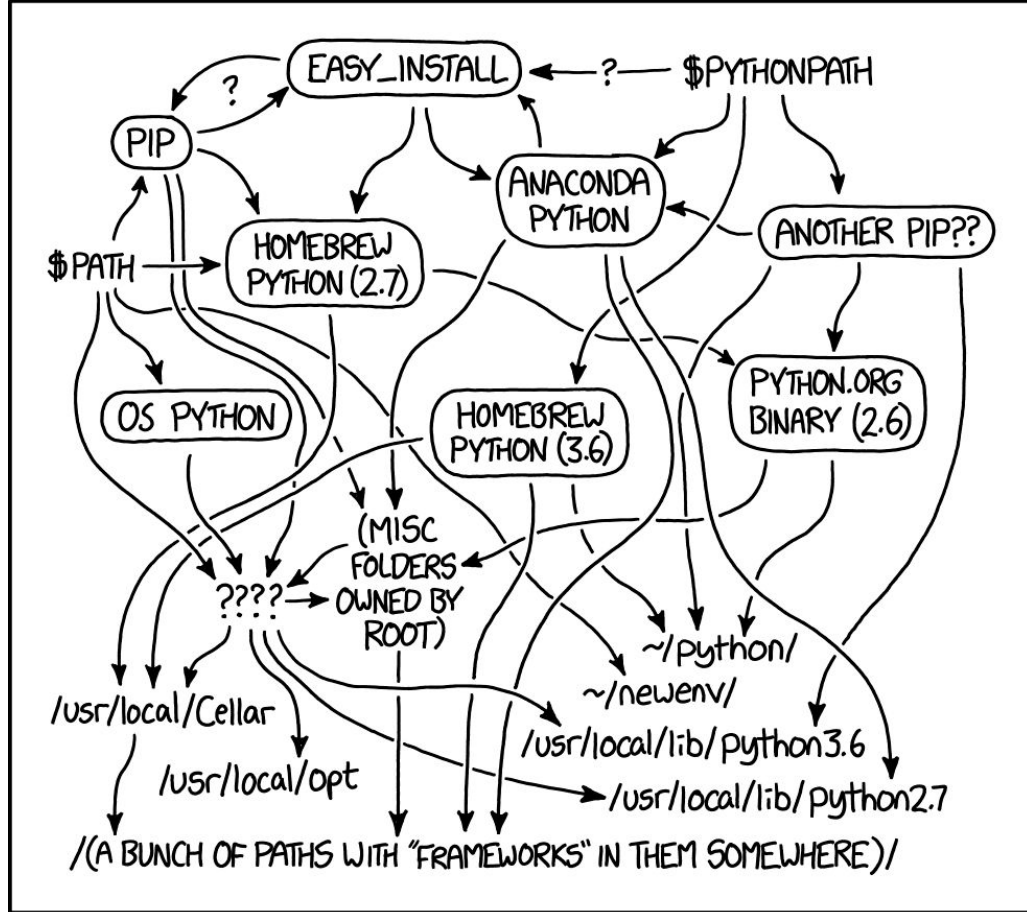
MY PYTHON ENVIRONMENT HAS BECOME SO DEGRADED
THAT MY LAPTOP HAS BEEN DECLARED A SUPERFUND SITE.

https://xkcd.com/1987/

Hurdles and Potential Difficulties

- ## A variety of platforms

  - Osx, Debian, Ubuntu, Fedora: not only SLC/CC

- ## Need to mount CVMFS

  - Widely varying computer skills, need to remove all obstacles

- ## Need network connection

  - Bad quality, travel

- ## Need to advertise these stacks

  - A few analysers are aware of LCG stacks

Hurdles and Potential *Solutions*

A variety of platforms

▸ Provide releases for OSx, Ubuntu (besides already supported platforms!)

▸ Rely on container technology: "HEP Software in a Box"

▸ 1 script to boot, mount local space, CVMFS managed transparently

or

▸ Provide 1 or 2 stack for OSx (current previous version)

▸ Provide 1 single RPATH-based stack for all Linux flavours*

Need to mount CVMFS

▸ One script to download and mount for aforementioned platforms

* See G. Amadio, Robust Linux Binaries, Hall 3 12th July 10:30 !

Need network connection

- Yes, like with packages via package managers.
- Normal operation of CVMFS: cache on disk what has been used
- Selective, user-driven caching (steered by simple scripts): equivalent to installation

Need to advertise these stacks

- More infrastructure: community based mattermost? Discourse? Talks during *Analysis* meetings/weeks of experiments?

- Container based solution: suitable for HT Condor

  - Isolate environment and submit it as is to the farm/Grid

- Inform users about the new stacks and the evolution of the bleeding edge

  - "Why we believe you might be interested in moving to the new stack"

- ▶ LCG releases and their distribution vector serve us well
  - ● Presently mainly limited to centrally managed systems and SWAN
- ▶ Proposal: expand the user base reaching analysers
- ▶ Clear advantages for scientists, no known blocking/difficult issues ahead of us
- ▶ Encourage even more reproducibility and preservation
  - ● A "global tag" identifying clearly a stack
- ▶ Coherent offer of *Software as a Service*
  - ● Scientific sw usable from interactive login service, batch, web based analysis *and personal laptops?*