
CERN

Singularity

LHCb Computing Workshop
Ben Jones

Agenda

- What problems are we interested in solving?
- What is Singularity?
- Features we like
- Usage
- Deployment at CERN
- Future & Contacts

What problems are we solving?

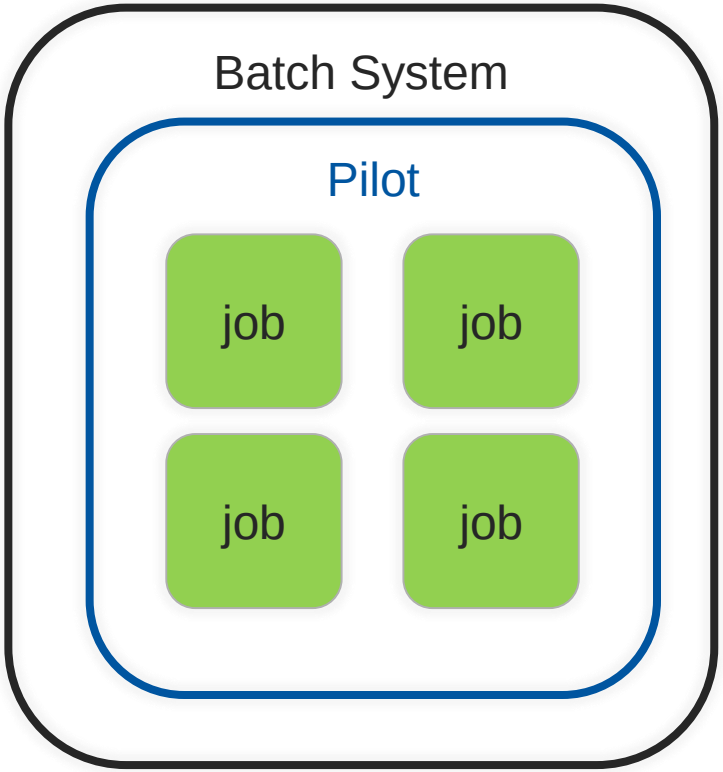
- Isolation
  - Isolation & Traceability working group and others interested in this area
  - Isolation of payload executed by pilots
  - File isolation: pilot determines what payload can r/w
  - Process isolation: payload can only affect its own processes
- Deco glexec
  - glexec more complex
  - But: glexec also provides traceability
- Decouple host OS from job OS environment

What is Singularity?

- "Singularity is a container solution created by necessity for scientific and application driven workloads”
- Made for HPC use case at Berkeley Lab
  - http://singularity.lbl.gov/
- Now going partially commercial
  - http://singularity.lbl.gov/2017-singularity-llc
- Simple container solution – possible to have single image and single rpm

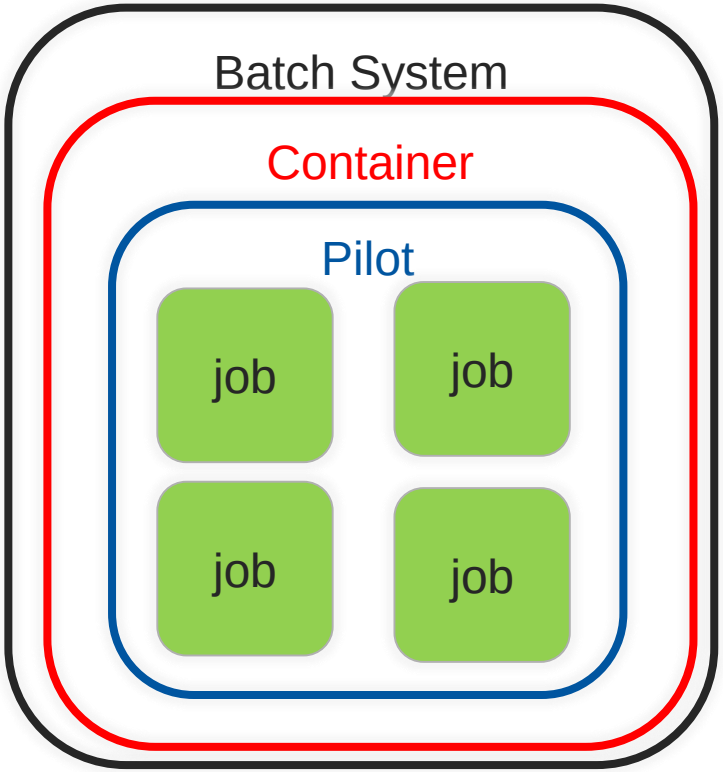
Containers all the way down

- Typical grid jobs are pilots, which have a payload of one or more jobs
- Batch system whether physical or virtual provides OS
- Infrequent OS upgrades



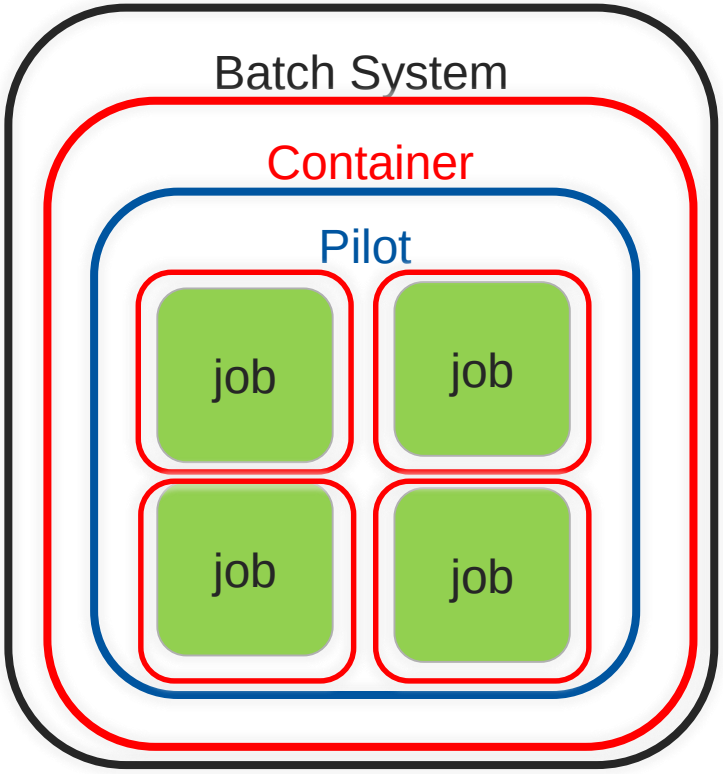
Containers all the way down

- Useful for sites to put Pilot inside a container
- Allows host OS to decouple from that required by Pilot
- Not all hardware supports rhel6 now
- Non wlcg workloads may require newer OS



Containers all the way down

- Pilots running payload in containers isolates jobs from each other
- Also decouples from the host OS. Whether or not a site has container OS, can run desired OS for job



Support for our current platforms

- Singularity can run on both SLC6 & CentOS7
- No OverlayFS support in SLC6
- Some SUID required on SLC6 & current CentOS7 for bind mounts
  - CC7.4 should remove need for SUID
- Can run cc7 on slc6 (and of course vice versa)
- Important as not everyone can run slc6 binaries on cc7

Simplicity

- A simple wrapper to bind-mount, isolate pid and chroot
- No daemon to run
- No config file to edit*
  - * ok, but it's a small config file and the defaults should be good enough for most use cases
- No resource management
  - The batch system does things like enforce memory usage

Singularity vs glexec

- glexec provides isolation by changing user context of payload job
- glexec also provides traceability, calling out to external services (ie ARGUS) to log context change
- Singularity provides payload isolation, but no traceability
- Rely on experiment frameworks to provide traceability
- HTCondor 8.6 release provides job logging framework

Security

- CERN Security audit of Singularity was very positive
  - Usual caveats that Control Technique doesn't mean your car won't break down…
- SUID not necessary in more modern kernels
- User in container is user outside container
- Namespaces are fairly new (hence RedHat refused to add in 7.3) so may still be bugs / security issues.

”Image” deployment

- Several container formats supported
  - Singularity container
  - Directory
  - Archives (tar[.{gz,bz2}], cpio[.gz])
  - Docker pull
- Docker url ok for testing but…
  - Container has to be downloaded each run
  - No cache
  - Reproducibility?
- Container big blob, no real benefits over filesystems (HPC aside)

Filesystem & CVMFS

- Singularity can just use a filesystem as a container.
- Flat filesystem deployed on CVMFS nice model
  - Cache on clients
  - Easy on the squids – only pull files you use
- OSG have tooling to sync Docker images to CVMFS
  - docker push to CVMFS in ~1h

Run a docker image

```
[bejones@lxplus042 bejones]$ singularity exec --home $PWD:/srv --pwd /srv --scratch /var/tmp/ --scratch /tmp/
--containall docker://python:latest python  -c 'import sys; print("Hello World: The Python version is %s.%s.%s
" % sys.version_info[:3])'
library/python:latest
Downloading layer: sha256:a3ed95caeb02ffe68cdd9fd84406680ae93d633cb16422d00e8a7c22955b46d4
Downloading layer: sha256:df67ecfc860bee4ca5110c9c4160acd37561ad05e8bfe237dded172fb835ed01
Downloading layer: sha256:a3ed95caeb02ffe68cdd9fd84406680ae93d633cb16422d00e8a7c22955b46d4
Downloading layer: sha256:6df6941e6752d66da0c4ebb4b0f5f2f3bca84c8e22624f2b9dcc338d546dd709
Downloading layer: sha256:9fb634154ace28722ced66d9272eef2d97eab3ab4c9452aaed3c1aff7c0eecdf
Downloading layer: sha256:a3ed95caeb02ffe68cdd9fd84406680ae93d633cb16422d00e8a7c22955b46d4
Downloading layer: sha256:a3ed95caeb02ffe68cdd9fd84406680ae93d633cb16422d00e8a7c22955b46d4
Downloading layer: sha256:ea8a37f1516109860d0de28779a53c0d67a36c51fc6aa6d083f21e36a7d2260e
Downloading layer: sha256:a3ed95caeb02ffe68cdd9fd84406680ae93d633cb16422d00e8a7c22955b46d4
Downloading layer: sha256:a3ed95caeb02ffe68cdd9fd84406680ae93d633cb16422d00e8a7c22955b46d4
Downloading layer: sha256:dbed9b09434efb583eb5f23173bf2aadb578cbfe2516e26226e3f7e458fac621
Downloading layer: sha256:9021b2326a1e3a942223c7e349a92203df184f2dcca45f5be7b0b80ac50e2ccf
Downloading layer: sha256:fb5937da9414eeab6d68ce06a7ff60d8be1e2c1518ac2588d5df135ab54a9801
Downloading layer: sha256:a3ed95caeb02ffe68cdd9fd84406680ae93d633cb16422d00e8a7c22955b46d4
Downloading layer: sha256:10a267c67f423630f3afe5e04bbbc93d578861ddcc54283526222f3ad5e895b9
/bin/bash: warning: setlocale: LC_ALL: cannot change locale (en_US.UTF-8)
Hello World: The Python version is 3.6.1
```

CVMFS

- $ singularity exec --containall --home /var/tmp/bejones:/srv -B /cvmfs /cvmfs/cernvm-prod.cern.ch/cvm3/ sh
  - Run a cernvm shell with /cvmfs bind mounted

```
[bejones@lxplus042 bejones]$ singularity exec --home $PWD:/srv --pwd /srv --scratch /var/tmp/ --scra
tch /tmp/ --containall /cvmfs/singularity.opensciencegrid.org/bbockelm/cms:rhel6/ python  -c 'import
 sys; print("Hello World: The Python version is %s.%s.%s" % sys.version_info[:3])'
Hello World: The Python version is 2.6.6
[bejones@lxplus042 bejones]$
```

```
[bejones@lxplus042 bejones]$ singularity exec --home $PWD:/srv --pwd /srv --scratch /var/tmp/ --scra
tch /tmp/ --containall /cvmfs/singularity.opensciencegrid.org/bbockelm/cms:rhel7/ python  -c 'import
 sys; print("Hello World: The Python version is %s.%s.%s" % sys.version_info[:3])'
Hello World: The Python version is 2.7.5
[bejones@lxplus042 bejones]$
```

Pilot Integration

- In principle pilot integration should be easy
- From:
  - job_wrapper.sh <args>
- To:
  - singularity exec /cvmfs/cernvm-prod.cern.ch/cvm3 job_wrapper.sh <args>
- The same userid is used in the container
- Lots of options for bind mounts

Deployment at CERN

- singularity deployed across htcondor batch pool
- Also installed on lxplus
- Running jobs:

Site: Descending	Count
t2_us_nebraska	50,863
t2_us_ucsd	47,530
t2_us_wisconsin	41,900
t2_us_purdue	31,163
t2_ch_cern	28,782
t2_us_caltech	27,660
t3_uk_scotgrid_gla	3,302
t2_tw_nchc	2,170
t3_us_osg	840

Future

- WLCG Working group: wlcg-containers
- Discussion around how bind mounts should be defined
  - Determined by VO or site?
- What Singularity isn’t:
  - Magic tool to make deployment trivial everywhere. Simpler than glexec but will still require some customisation
  - A tool to allow users to each run their payload in their own pre-configured environment: distributing/debugging/supporting said environments isn't reasonable on a large scale

Other resources

- aka Plagiarism Watch
- Singularity @SiGNET: https://indico.cern.ch/event/612601/contributions/2495598/attachments/1424187/2183998/Singularity_at_SiGNET.pdf
- CMS use case: https://indico.cern.ch/event/612601/contributions/2495602/attachments/1424434/2184476/Singularity-in-CMS-v2.pdf

Questions?