Containers WG status

GDB 10 January 2018
Gavin McCance

Recent meetings

- Long summer gap after initial kick-off at pre-GDB
- Have agreed now bi-weekly meetings to iterate faster
- Two held so far, with focus on Singularity deployment
- Aim to come to an initial baseline recommendation for sites
    - CMS have requested sites install Singularity in Q1 2018 (for SL6 and CC7)

CMS request

- Request from CMS (WLCG broadcast) to sites to install Singularity for SL6 and Centos7 hosts by March 2018
  - 2018 production will be running inside Centos7 containers. CMS will no longer require gLexec.
    - Critical SAM test in March
- Default configuration of Singularity (from RPM, with setuid) is fine
- Details: https://twiki.cern.ch/twiki/bin/view/CMSPublic/FacilitiesServicesDocumentation#Singularity

Singularity Packaging

- Currently is available in EPEL but maintainer has lots of unreferenced patches and is slow to respond
  - Trying to get in touch - we'd prefer to use EPEL!
- Currently built by OSG and distributed using WLCG repo
  - Sites should use this for now instead of version in EPEL
  - singularity-2.3.2-0.1.1.osg34 is current
  - 2.4 in preparation, but still some issues

Issues tracked

- Some issues found between Singularity, CVMFS and overlayfs:
  - Bind-mount dropping bug has been addressed for a specific RH7 config with a fix to CVMFS (https://sft.its.cern.ch/jira/browse/CVM-1423)
  - Some murkiness between overlayfs / cvmfs / singularity being investigated (when overlayfs can't create subdirectories properly)
    - Tracked in https://sft.its.cern.ch/jira/browse/CVM-1434 though not clear where the bug is

Image distribution

- For normal production, everyone is agreed on unpacked CVMFS
    - LHCb just re-use CVMFS-hosted CernVM image as-is
        - Question of image building (how, who and what does it) is still open - central WLCG, per experiment?
- For HPC production, image files are likely to be needed
- For analysis, still unclear (see later...)

(Non)-setuid mode

- One of the long-term advantages of Singularity (from Traceability and Isolation WG) was the potential path to a non-setuid isolation solution

- Singularity still uses setuid by default, and is the only option on SL6

- RH7.4 now has unprivileged user namespace preview (can enable at boot and sysctl) which allows Singularity to run non-setuid

..but with non-setuid

1. Unpacked CVMFS chroot work well, but image files do not work at all - issue for HPC
2. Overlayfs isn't available in non-setuid mode

    - ..which means for bind-mounts, the top-level target directories in the container need exist and be known in advance (and pre-created in the image or CVMFS chroot)

..no overlayfs

- CMS payload $PWD is always bind-mounted to /srv inside container but this took quite some effort
- ATLAS payload not (yet?) there - quite some work would be needed
  - Also some sites require their POSIX local store to be mounted - at varied mount-points
  - Hard to pre-make an single image with all possible mount-points
  - Some workarounds being discussed -> one now requested upstream
- ALICE and LHCb are confident that they can live with mount-point restriction, though probably require mods to DIRAC/Alien to handle it

Analysis

- Limited discussion here, though lots of work in experiments
- Distribution question - CVMFS-hosted image files, CVMFS unpacked chroots, or WLCG Singularity repo service
  - Site concern about fast image turnaround with fat image files and caching them... and would also require site to use Singularity with setuid mode
  - Analysis group concern about the grid-wide deployment speed if the CVMFS-unpacked mechanism is required - notably for image development
- How fast is needed?

Baseline ongoing

- No unified baseline recommendation yet
- Aiming at 90% solution for "standard sites"
- For RH7, we believe the non-setuid approach is preferable in the long-run, but less clear in the 2018 timeframe. Sites should have the option (setuid or not) and we recommend experiments be able to function with both approaches
    - CMS done, Alice and LHCb look OK, Atlas to be determined
    - For SL6 setuid is the only option anyway
- CMS have a site recommended config for Q1 2018. Question of unified site config is still open until we understand ATLAS use
- Version 2.3 soon 2.4 - site's should use the WLCG repo, not EPEL