Singularity & ATLAS Status

Andrej Filipcic

Containers

- The most efficient “virtualization” technique
  - Native performance as compared to true virtualization
  - Effectively using a custom set of OS libs and software apart from sharing the kernel with host OS (similar to chroot)
- Providing isolation
  - PID namespace
  - Network configuration
  - UID/GID
  - Filesystem, device access
- Better suited for batch execution than VMs

Docker and Singularity

- Docker:
  - Analogous of VMs
  - Docker instances can be long lived - service deployment model
  - Or application oriented - execution of complex workflows
  - Provisioning model similar to VM
- Singularity:
  - Designed for batch job execution, focusing on simplicity and minimal configuration
  - Eg: singularity <OS Image> <command>
  - Can also run in user space (no SUID) with limited functionality (eg no bind mounting)

ATLAS needs both

- Singularity:
  - Well suited to be used everywhere making the site SW specifics irrelevant to ATLAS
  - ATLAS jobs can be executed on every site regardless of the site OS and do not require any customization at the site. Site upgrades decoupled from ATLAS SW requirements
  - ATLAS can use several OS versions at the same time matching the ATHENA release requirements - eg Run-1 analysis on SLC5, SLC6 images, Run-2 on SLC6, CC7, ...
- Docker:
  - Currently the best way to encapsulate more complex tasks, such as software development/testing or analysis preservation

Singularity is better suited for job execution at sites, while docker requires more complex deployment and more privileges on a site

Key differences

- Docker is better suited for full encapsulation
  - Full software and environment stack
  - Arbitrary workflow or service execution
  - Instances can be long lived
- Singularity:
  - OS encapsulation but use as much as possible from host OS
  - lower initialization latency

But:

Docker and singularity can use identical images, the usage of either is purely context dependent.

Current deployment focus

- Singularity should be widely deployed on most of our computing resources both pledged an opportunistic
- Docker is for now not considered yet, although some proactive sites are already supporting it
  - To be addressed in 2018
- Several sites (~10) are using singularity already for ATLAS production, although in a way which is not controlled by ATLAS
  - Eg: forcing automatic execution of all ATLAS jobs in SLC6 containers

Execution modes and status

- Fully containerized jobs, implemented at the site (batch, CE) level
  - Good for now but ATLAS cannot choose the container to be used for a specific workflow/job
  - There are “simple” recipes for deployment on sites that really need it (eg ubuntu, gentoo, coreOS clusters) - works perfectly well on Tier-1 and Tier-2 sites
- Job step containerized execution:
  - Stagein, RunJob, stageout pilot steps are each executed in a separate container instance
  - Configured as singularity_options in AGIS catchall
  - Implemented in pilotcode and tested on few sites
  - Proof of concept for now, will need more pilotcode refactoring
- Launch the container in the wrapper and run the pilot fully in container
  - To be implemented soon - the easiest way to go for now

WLCG direction

- CMS is widely deploying singularity and using it in production on at least SLC7 sites
- WLCG gave a green light to singularity at the last workshop
  - Start with wider deployment and usage
  - Experiments should collect the experience, site specific requirements or configuration specifics in the next few months
  - Provide the input to WLCG at the next GDB in October
- Points to evaluate:
  - Can the experiments use the common images?
  - Are unprivileged containers enough (can run completely in user space)?
  - Are there custom configuration requirements on sites?

ATLAS direction

- Almost a magical solution to address job execution on ALL sites
  - Host OS is completely decoupled from execution environment
  - HPCs: a way to deliver ATLAS software to sites without cvmfs
- Software development and testing in controlled environment
- Analysis preservation:
  - a way to store the complete workflow of all analysis steps (env setup, preparation steps, data processing, post processing, assembly of the results)

Point to discuss

- Deploy the ATLAS-controlled singularity, use more sites
  - Wrapper? Time scale?
- Find a common solution for grid and HPCs (and the rest)
  - Common procedure/recipe to create images (and unpacked images)
- Make the image creation part of software release process
  - Store the images in cvmfs, fat images in datasets…
- Documentation and installation instructions for deployment