Western Norway
University of
Applied Sciences

Grid services in a box

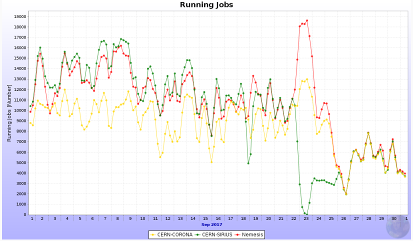
Container management in ALICE

Maxim Storetvedt

June 22, 2018

Using containers for site-services at ALICE

- This talk will focus on the initial experiences with managing containers for VOBOX use
    - Multiple deployed within ALICE as a pilot project
- Also planned for worker nodes
    - For more on this topic, see the talk by Miguel Martinez Pedreira on JAliEn



Containerised VOBOXes running production jobs

Using containers for site-services at ALICE (2)

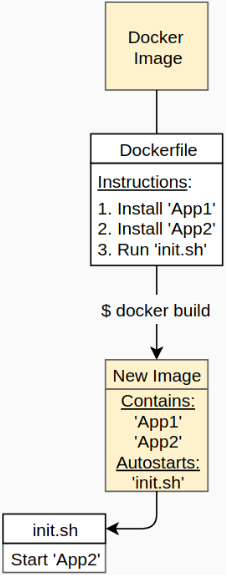
- Containers can provide several benefits over using virtual machines (VMs) for VOBOXes
  - Less overhead
  - Less use of storage
  - One-click deployment
- Container setup for VOBOXes is very different from VMs
- The next slides are dedicated to examining
  - Configuration
  - Downtime prevention
  - Performance

Selected VOBOX Container platform

- **Docker** used within ALICE for site-service containers
- Other container platforms available
  - **Singularity** quickly gaining ground within HPC
- Site-services, like VOBOXes, need a full networking stack
  - Not currently available in Singularity
  - Available in platforms like Docker and Rkt

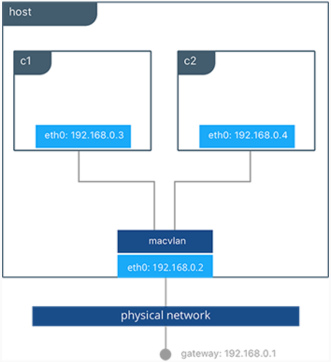
ALICE VOBOX image configuration

- We need automatic startup of VOBOX services at container launch
- Dockerfiles
  - Scripts composed of various commands to perform on a base image
  - End result is a new, customised, image
- An image must be rebuilt to apply changes to a Dockerfile
  - Since this is a pilot project, changes are frequent
  - Results in downtime
- Solved by pointing to a script within the container – e.g /etc/init.sh



ALICE VOBOX Network Configuration

- MACVLAN – A reverse VLAN
  - A VLAN maps an OS side of a networking interface to multiple virtual networks on its network side (one-to-many)
  - A MACVLAN maps a network side of an interface to multiple virtual interfaces, with each their own MAC address (many-to-one)
  - Traffic sent from the virtual interfaces is sent directly to the underlying network, and identified by the assigned MAC address.
- VOBOX containers networked using MACVLAN
  - Allows containers to appear as normal machines on the network



MACVLAN architecture

ALICE VOBOX host configuration

- VOBOXes need many files open simultaneously
  - Will quickly reach default system limit for maximum open files when more than two VOBOX containers run on a single host
  - Causes services to freeze or terminate
  - System limit must be increased to avoid these issues
- Autofs (for CVMFS mounting) disabled on all hosts
  - Otherwise known to cause problems for containers. Having it disabled requires less manual interaction

ALICE VOBOX host configuration (2)

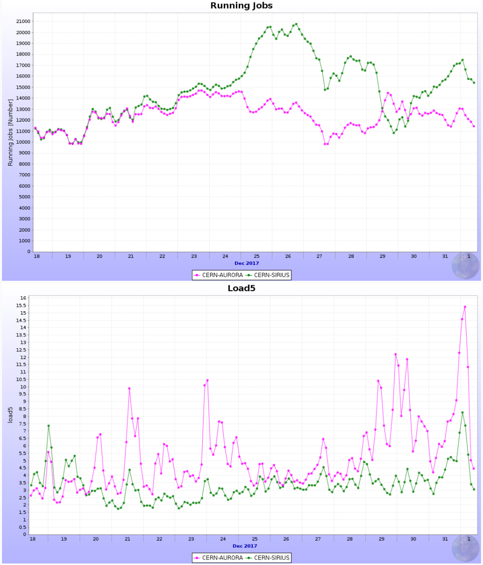
- Host connectivity
  - The host and its containers can not reach/ping each other
    - Specific to how MACVLAN works
  - Create a Docker bridge between the host/containers if connectivity is needed
- Kernel access privileges
  - Containers have limited access privileges by default
    - Several tools and services may fail to launch
    - Most networking tools are affected
  - Full privileges granted for VOBOXes
    - Limited risk for this purpose, as VOBOXes are handled by sysadmins

Preventing containerised VOBOX downtime

- The ALICE containerised VOBOXes use the Live Restore feature
    - Allows containers to run without the Docker service
    - Useful for system updates → avoid downtime
    - Containers must still reconnect with the service at some point
        - Will otherwise eventually fail due to log-buffer overflow
- Container management tool (Swarm) also available
    - Not used for VOBOXes within ALICE
    - Not efficient when having few containers

Performance

- Tests on both load and performance show similar results to that of VMs
  - With less overhead and a smaller storage footprint
  - Tested with the two main storage drivers for Docker – AUFS and Overlay2



Above: Running production jobs, alongside a VM (Overlay2)
Below: Load, alongside a VM (Overlay2)

Performance (2)

- Performance decreases when the number of storage layers increases
    - Common for copy-on-write filesystems
    - All changes to a container are stored on a separate storage layer
        - New layer added for each commit
    - Flattened images used during testing
        - All additional layers merged into one

Conclusion

- Containerised VOBOXes are in pilot production, as shown in this presentation
  - Positive results in terms of load/performance
  - Stability
- More to be deployed
  - See also the talk by Miguel Martinez Pedreira on JAliEn, for their use on worker nodes

End

Thank you

Questions or comments?

E-mail: msto@hvl.no