

JobAgent/JobWrapper status update

Progress and site readiness for a containerised JobAgent

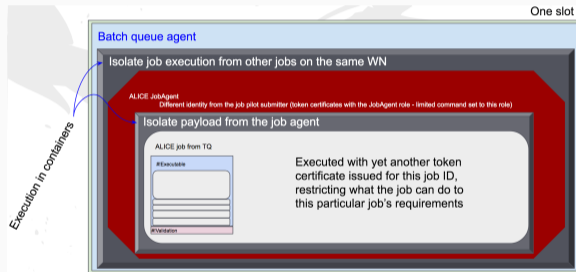
Maxim Storetvedt

`msto@hvl.no`

May 14, 2019

Recap: the two layer job pilot

- jAliEn moving towards a split, two-layer, job pilot
- Allows for improved isolation from other jobs on the same WN, and from the JobAgent
 - The JobAgent runs with a token certificate for the JobAgent role
 - The payload will be executed with another token certificate specific for the job ID
 - Can be run in containers for additional isolation

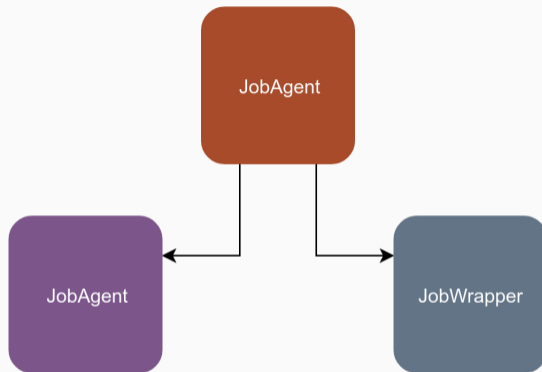


Concept presented by Miguel M. Pedreira at WLCG Containers WG meeting in December 2017



- JobAgent class split in two:

Split pilot - implementation



- JobAgent class split in two:
 - JobAgent – Gets matched job, launches JobWrapper
 - JobWrapper – Responsible for job execution

Split pilot - implementation summary

- JobWrapper has no other responsibility other than running each specific job
 - Everything else is handled by the JobAgent
 - JobWrapper sends messages to JobAgent with updates on status of the job
 - Status for the job is then changed within JobAgent
- Messages are handled by a listener process within the JobAgent and JobWrapper
 - Each received message is echoed back to confirm
- Logging options for the JobAgent are also applied to the JobWrapper, if available



- "Feature-complete":
 - Both pilot layers implemented (JobAgent, JobWrapper)
 - Both components run in separate JVMs, using separate identities
 - "Container ready" pipes for data and message passing
 - Safe passage of credentials
 - JDL and env info
 - Updates on jobstatus
 - Tested on numerous sites/configurations
- Merge request submitted
 - Awaiting approval

Next step: enabling container support

- The JobWrapper layer can be run in a container for improved isolation
 - The split JobAgent was conceived and implemented with this in mind
- Mainly two things must be changed to fully enable container support:
 - **Launch command generation**
 - Check for Singularity, then launch the JobWrapper in a container if available¹
 - Otherwise continue without containerisation
 - **Monitoring**
 - PID may change upon launching JobWrapper / payload in container
 - Monitor would no longer find process, throwing a NullPointerException → crash
 - Solved by having the JobWrapper pipe over the PID to JobAgent after start

¹Initial focus on Singularity, but can work with any container platform

Container support - availability

- Container support already available (separate branch on GitLab)
 - Currently undergoing testing to check for site specific issues / exceptions
- In other words, jAliEn is "*almost*" ready for running payload in containers

But what about the individual computing sites?

Container support - site requirements

- The ideal site would have:
 - **Singularity**, with
 - **Underlay** enabled
- This can be achieved by:
 - Installing a Singularity build (with underlay support) using the RPM
 - Available OSG builds for Singularity 2.X
 - Included in official builds starting with 3.0
 - Running Singularity through CVMFS
 - Requires an EL 7 system with the 7.6 kernel
 - User namespaces must be enabled on sites (requires root)



- **Underlay** allows the bind-mounting of directories that do not exist in the container image
 - As sites have different directory layouts and naming conventions, this is a must!
- This can also be achieved using **overlay** (more common), but it has limitations:
 - Does not work with EL 6
 - Has issues with images distributed over CVMFS
 - More prone to vulnerabilities

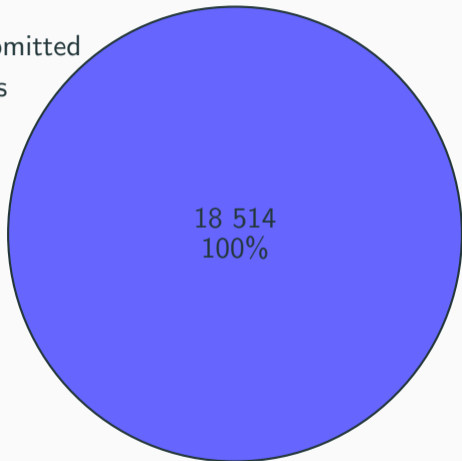
Goal: determine current site support for running payload in containers

Container support - overview

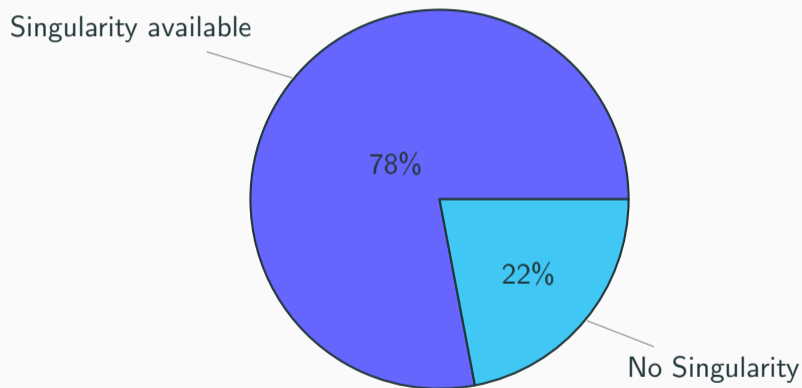
18,514 test jobs submitted

10,090 unique nodes

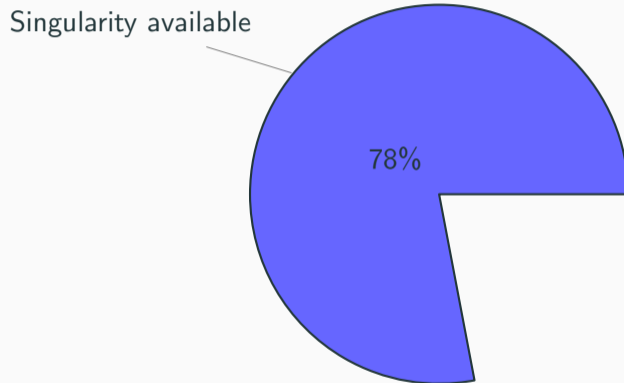
33 unique sites



Container support - overview

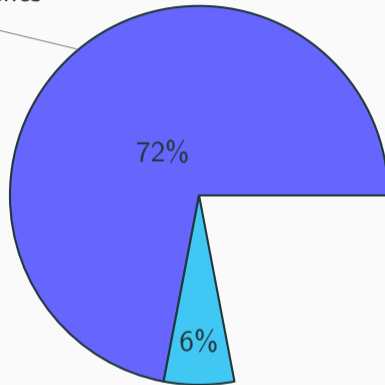


Container support - overview



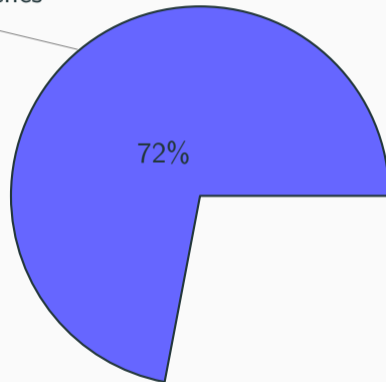
Container support - overview

Singularity successfully launches



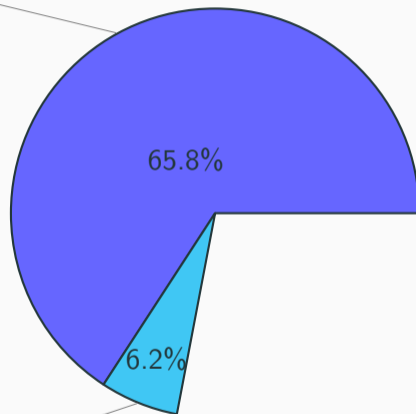
Singularity does not launch

Singularity successfully launches



Container support - overview

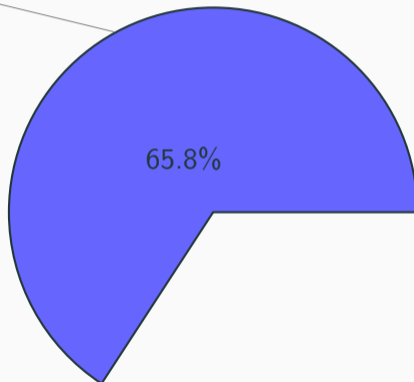
Loopback device mounting disabled



Loopback devices enabled

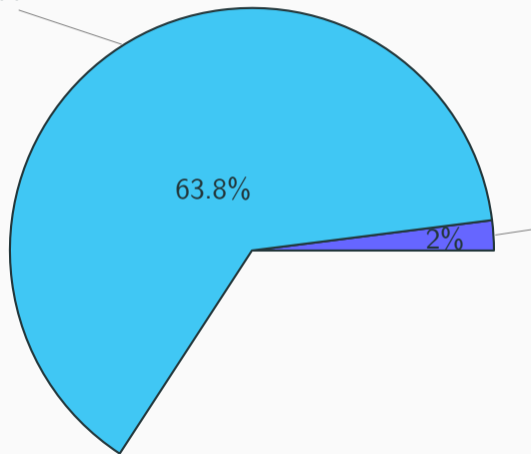
Container support - overview

Loopback device mounting disabled



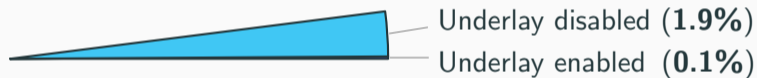
Container support - overview

Underlay not supported

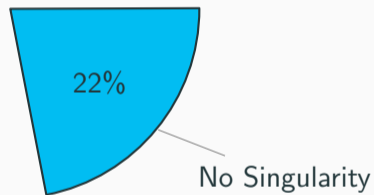


Underlay supported

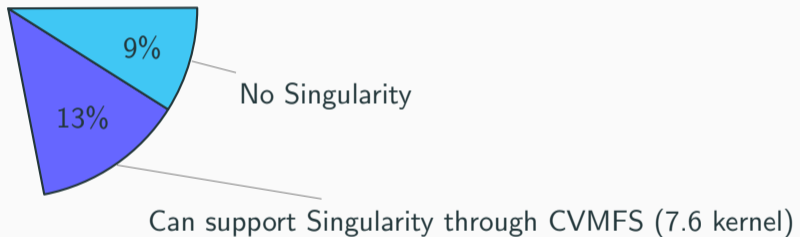
Container support - overview



Container support - overview



Container support - overview



- Sites with (some) fully compatible nodes:
 - infn.it
- Sites with compatible nodes, but underlay disabled in conf:
 - gridpp.rl.ac.uk
 - cea.fr
 - infn.it (plenty)
 - jinr.ru
 - itep.ru
 - kfi.hu
 - in2p3.fr

- Nodes without Singularity can run via CVMFS:
 - alice-cray
 - liu.se
 - ornl.gov
 - sdfarm.kr
- No support for containers:
 - uniba.sk

Options:

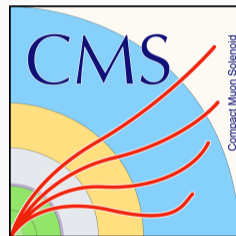
- Wait for sites to enable underlay support
- Wait for sites to upgrade to 7.6 kernel
- Create workaround

Options:

- ~~Wait for sites to enable underlay support~~
- ~~Wait for sites to upgrade to 7.6 kernel~~
- Create workaround*

***Conclusion from the discussion during the ALICE Computing Week in April**

- The "CMS way":
 - Bind site workdir and tmpdir to preexisting directories in the container
 - No underlay or overlay required, as the directories will always be in the container image
 - Used in production at CMS



Creating a workaround - current progress



- Changes applied to the relevant branch on GitLab:
 - Always bind site workdir to **/workdir**
 - Always bind site TMPDIR to **/tmp**
 - Adjustments to logging and LFN generation

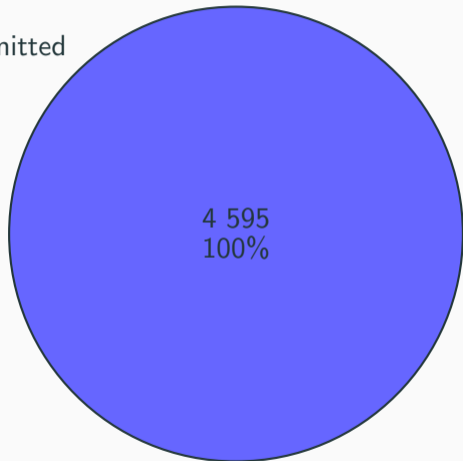
Now in early testing

Creating a workaround - initial results

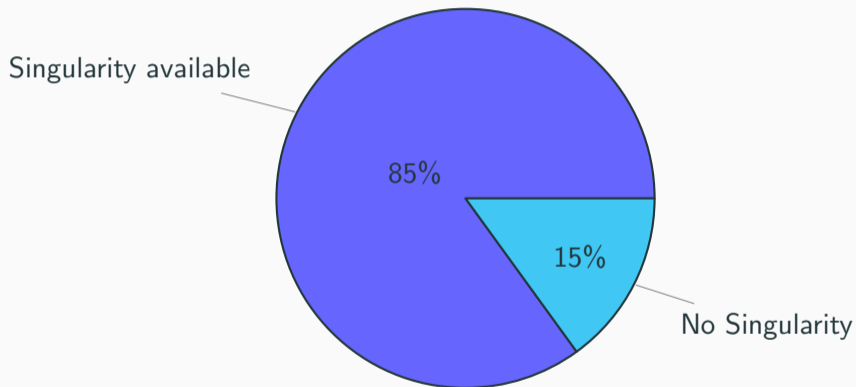
4,595 test jobs submitted

2,779 unique nodes

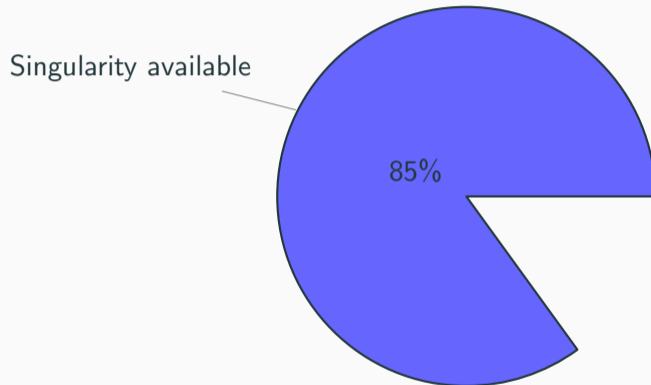
34 unique sites



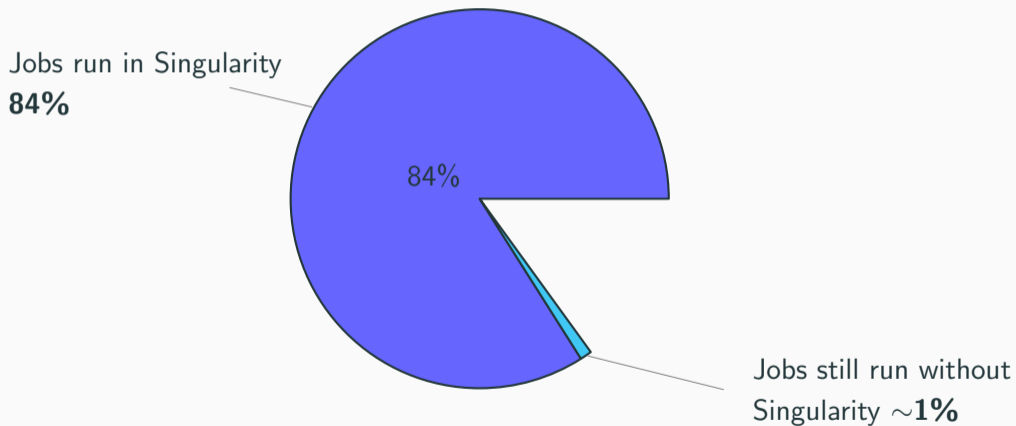
Creating a workaround - initial results



Creating a workaround - initial results



Creating a workaround - initial result



Summary and outlook

- jAliEn is steadily moving towards support for running job payload in containers
 - Split JobAgent being finalised
 - Containerised JobAgent in testing
 - But – are sites prepared for running jobs in containers?
 - If not, the split JobAgent will be run without
- A majority of sites have support for Singularity
 - Yet only a handful have an "ideal" configuration (<2%)
 - Can be improved by using the workaround (~75%)
- Outlook:
 - Further testing of the containerised JobAgent
 - Adding JDK and image to CVMFS

Thank you

Questions or comments?

E-mail: msto@hvl.no