

Arvados - Feature #17726

[singularity] add documentation

05/25/2021 06:27 PM - Ward Vandewege

Status:	Resolved	Start date:	08/05/2021
Priority:	Normal	Due date:	
Assigned To:	Tom Clegg	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:	2021-08-18 sprint		
Description			
Document singularity support from the admin perspective (configuration settings, etc)			
Document singularity support from the user perspective (what changes are needed in my CWL, if anything, when my cluster uses singularity instead of docker?).			
Document any limitations or differences in behavior when using singularity vs docker.			
As part of the refactoring required for Singularity support, we have a slight behavior change in the loading of docker images from Keep (cf. https://dev.arvados.org/issues/17296#note-34):			
<ul style="list-style-type: none">slight behavior change in loading docker images: before, the .tar file of the image had to be the first file in the collection, and that was loaded. Any subsequent .tar files were ignored. Now, it is an error if there is more than one .tar file. This change should be mentioned in the upgrade docs (though it's unlikely someone would run into it).			
Subtasks:			
Task # 17974: Review 17726-singularity-doc			Resolved
Related issues:			
Blocks Arvados Epics - Story #16305: Singularity support		In Progress	01/01/2021 09/30/2021

Associated revisions

Revision e99f026d - 08/12/2021 08:05 PM - Tom Clegg

Merge branch '17726-singularity-doc' into main

closes #17726

Arvados-DCO-1.1-Signed-off-by: Tom Clegg <tom@curii.com>

History

#1 - 05/25/2021 06:27 PM - Ward Vandewege

- Blocks Story #16305: Singularity support added

#2 - 05/25/2021 06:31 PM - Ward Vandewege

- Description updated

#3 - 07/06/2021 05:07 PM - Peter Amstutz

- Target version deleted (To Be Groomed)

#4 - 08/03/2021 01:52 PM - Peter Amstutz

- Target version set to 2021-08-18 sprint

#5 - 08/04/2021 03:52 PM - Peter Amstutz

- Assigned To set to Tom Clegg

#6 - 08/05/2021 03:21 PM - Tom Clegg

- Status changed from New to In Progress

- slight behavior change in loading docker images: before, the .tar file of the image had to be the first file in the collection, and that was loaded. Any subsequent .tar files were ignored. Now, it is an error if there is more than one .tar file. This change should be mentioned in the upgrade docs (though it's unlikely someone would run into it).

This is already mentioned at <https://doc.arvados.org/main/admin/upgrading.html> ("Multi-file docker image collections") but it's in the v2.2.0 section. Needs to move up to "development main".

Typically a docker image collection contains a single .tar file at the top level. Handling of atypical cases has changed. If a docker image collection contains files with extensions other than .tar, they will be ignored (previously they could cause errors). If a docker image collection contains multiple .tar files, it will cause an error at runtime, "cannot choose from multiple tar files in image collection" (previously one of the .tar files was selected). Subdirectories are ignored. The arv keep docker command always creates a collection with a single .tar file, and never uses subdirectories, so this change will not affect most users.

#7 - 08/05/2021 07:20 PM - Tom Clegg

17726-singularity-doc @ [1c3fcbc97db70680109aba244287df0231285648](https://github.com/1c3fcbc97db70680109aba244287df0231285648)

adds a page about switching from Docker to Singularity, and moves the note about docker image collections from 2.2 to main where it belongs.

singularity support from the user perspective

Not sure what to say about this, or where... thoughts?

#8 - 08/06/2021 07:49 PM - Peter Amstutz

- Although most people reading the documentation probably already have heard of singularity, linking to <https://sylabs.io/singularity/> would be useful.
- I would also add a sentence or two about why you might choose to use Singularity.
- "Each container will have access to all memory on the host where it runs." -- this should probably be qualified, the scheduler (slurm/LSF) might be configured to enforce limits.
- "Programs running in containers may behave differently due to differences between Singularity and Docker, e.g., the root (image) filesystem is read-only in a Singularity container." let's be precise about the known differences not leave it open ended:
 - Programs running in containers may behave differently due to differences between Singularity and Docker
 - The root (image) filesystem is read-only in a Singularity container. Programs that attempt to write outside the designated output or temporary directory are likely to fail.
 - Docker ENTRYPOINT is ignored

"Make sure Singularity is installed" should link to Singularity's install instructions. Possibly also mention you need squashfs-tools .

It mentions SLURM/LSF but only gives crunch-dispatch-slurm or arvados-dispatch-cloud as examples on the last line.

Users probably do need to be aware of points 2, 3, and 4 of the notes (lack of native SIF support, memory limits, and behavior differences). I'm also not sure of the best place to locate that in the documentation.

#9 - 08/09/2021 02:52 PM - Tom Clegg

17726-singularity-doc @ [b74bc91079b0c2b5491e08f8ff677451c6d7e60e](https://github.com/b74bc91079b0c2b5491e08f8ff677451c6d7e60e)

I would also add a sentence or two about why you might choose to use Singularity.

Any suggestions? Easier install/config? Reduce potential for boot-time complications related to docker daemon?

Users ... not sure of the best place to locate that

Maybe on <https://doc.arvados.org/main/user/topics/arv-docker.html> at the end of the "Upload your image" section, after the DockerRequirement bit, pointing out that depending on configuration it might end up being used in a singularity container rather than a docker container per se, with a link to the install page?

Maybe on https://doc.arvados.org/main/api/methods/container_requests.html in the "container_image" parameter description, along similar lines?

#10 - 08/09/2021 08:50 PM - Peter Amstutz

Tom Clegg wrote:

17726-singularity-doc @ [b74bc91079b0c2b5491e08f8ff677451c6d7e60e](https://github.com/b74bc91079b0c2b5491e08f8ff677451c6d7e60e)

I would also add a sentence or two about why you might choose to use Singularity.

Any suggestions? Easier install/config? Reduce potential for boot-time complications related to docker daemon?

I would say something about how Singularity is more likely to be containerization of choice on HPC systems, it is somewhat simpler / easier to manage due to not having a separate daemon, and that once a Singularity image is cached, the Singularity container may have less overhead than the equivalent Docker container.

Users ... not sure of the best place to locate that

Maybe on <https://doc.arvados.org/main/user/topics/av-docker.html> at the end of the "Upload your image" section, after the DockerRequirement bit, pointing out that depending on configuration it might end up being used in a singularity container rather than a docker container per se, with a link to the install page?

Maybe rename that page from "Working with Docker images" to "Working with container images" and then at the top of the page mention that both Docker and Singularity are supported, and add a section that repeats the the notes about how Singularity works differently.

Maybe on https://doc.arvados.org/main/api/methods/container_requests.html in the "container_image" parameter description, along similar lines?

I'm not sure what you're suggesting, we haven't changed the container request API for this?

#11 - 08/10/2021 09:01 PM - Tom Clegg

Peter Amstutz wrote:

I would say something about how Singularity is more likely to be containerization of choice on HPC systems, it is somewhat simpler / easier to manage due to not having a separate daemon, and that once a Singularity image is cached, the Singularity container may have less overhead than the equivalent Docker container.

Added this. I didn't include a part about "less overhead" because it seemed a bit awkward to explain (slower on first use by a given user, maybe faster after that, depends on how often the image is used) compared to the benefit being explained... is that OK?

Maybe rename that page from "Working with Docker images" to "Working with container images" and then at the top of the page mention that both Docker and Singularity are supported, and add a section that repeats the the notes about how Singularity works differently.

Added this. Linked to the install/config page instead of repeating the info, though. Maybe when singularity support is a bit more mature it will make more sense to move the "differences" bit to a separate user-facing page?

Maybe on https://doc.arvados.org/main/api/methods/container_requests.html in the "container_image" parameter description, along similar lines?

I'm not sure what you're suggesting, we haven't changed the container request API for this?

It was just an idea of a place to put the singularity info where users might find it. But I think "working with container images" is fine.

17726-singularity-doc @ [5bc8ad779b8d39c63df88c20e5a883f4fe15c6da](https://doc.arvados.org/main/api/methods/container_requests.html)

#12 - 08/12/2021 08:01 PM - Peter Amstutz

Tom Clegg wrote:

Peter Amstutz wrote:

I would say something about how Singularity is more likely to be containerization of choice on HPC systems, it is somewhat simpler / easier to manage due to not having a separate daemon, and that once a Singularity image is cached, the Singularity container may have less overhead than the equivalent Docker container.

Added this. I didn't include a part about "less overhead" because it seemed a bit awkward to explain (slower on first use by a given user, maybe faster after that, depends on how often the image is used) compared to the benefit being explained... is that OK?

Maybe rename that page from "Working with Docker images" to "Working with container images" and then at the top of the page mention that both Docker and Singularity are supported, and add a section that repeats the the notes about how Singularity works differently.

Added this. Linked to the install/config page instead of repeating the info, though. Maybe when singularity support is a bit more mature it will make more sense to move the "differences" bit to a separate user-facing page?

Maybe on https://doc.arvados.org/main/api/methods/container_requests.html in the "container_image" parameter description, along similar lines?

I'm not sure what you're suggesting, we haven't changed the container request API for this?

It was just an idea of a place to put the singularity info where users might find it. But I think "working with container images" is fine.

17726-singularity-doc @ [5bc8ad779b8d39c63df88c20e5a883f4fe15c6da](#)

LGTM

#13 - 08/12/2021 08:21 PM - Tom Clegg

- *Status changed from In Progress to Resolved*

Applied in changeset [arvados|e99f026d040c6020dfcc51c6d988cf18d325a530](#).