

Arvados - Support #3787

[Documentation] Add a section in the User Guide about how to create, upload, and use a custom docker image

09/02/2014 02:07 PM - Ward Vandewege

Status:	Resolved	Start date:	10/10/2014
Priority:	Normal	Due date:	
Assigned To:	Peter Amstutz	% Done:	100%
Category:	Documentation	Estimated time:	0.00 hour
Target version:	2014-10-29 sprint		
Description			
Subtasks:			
Task # 4224: Review 3787-docker-docs			Resolved
Task # 4147: Write new documentation page			Resolved
Related issues:			
Blocked by Arvados - Support #4188: [OPS] Automatically upload arvados/jobs i...			Resolved

Associated revisions

Revision a22d6e6b - 10/20/2014 07:36 PM - Peter Amstutz

Merge branch '3787-docker-docs' closes #3787

History

#1 - 09/02/2014 02:07 PM - Ward Vandewege

- Story points set to 0.5

#2 - 09/02/2014 03:52 PM - Ward Vandewege

- Category set to Documentation

#3 - 10/08/2014 05:50 PM - Ward Vandewege

- Target version changed from Arvados Future Sprints to 2014-10-29 sprint

#4 - 10/08/2014 07:20 PM - Peter Amstutz

- Assigned To set to Peter Amstutz

#5 - 10/10/2014 12:54 PM - Peter Amstutz

- Status changed from New to In Progress

#6 - 10/20/2014 01:52 PM - Brett Smith

Reviewing [4c0c769](#)

I like everything on this page. Thanks for writing it up. I understand this is intended to be introductory documentation and not a comprehensive reference, but even with that in mind, I think a few additions would be very helpful to beginners:

- A very brief (like, two or three sentences) explanation of what Docker is, with a link to its own documentation
- An explanation why this tutorial starts from the arvados/jobs image (i.e., note any Docker image you use has to include the SDKs for your Crunch script)
- Brief mention of `arv keep docker's --project-uuid` and `--name` options (doesn't even need to explain them in detail; punting with "run arv keep docker --help for details" would be fine by me)
- Mentioning the different ways to specify the runtime constraint, depending on how much flexibility you want: Docker name, Collection UUID, or portable data hash

Beyond that, a few small notes about the intro bits:

- I think it'd be nice to have the word "Docker" somewhere in the title, so somebody who's heard that they need to use Docker can Ctrl+F it on the user guide and find it. Something like "Creating a custom Crunch environment with Docker."
- The standard blurb about having a usable runtime environment should be extended to mention that Docker itself must be installed on the system and usable from the user's account. (Maybe link to Docker's installation documentation for more?)

- In the ordered list of "what this page covers," move "how to" from the first bullet point to the leading text, for grammatical consistency.

And then one last organizational thing: I understand the rationale behind adding links to the Job and PipelineTemplate documentation from this guide, and I support making this easier to navigate, but this approach feels like a rough fit with our current organization. This document is billed as the User Guide (i.e., non-reference material); adding a "Reference" section to it suggests that we're getting a little away from ourselves. And if nothing else, having the terse titles "Job" and "PipelineTemplate" in the sidebar is kind of awkward; new users won't understand what's on those pages. I think a nicer way to tackle this quickly would be to add links to those pages from the appropriate guide pages (e.g., "For complete documentation about pipeline templates, refer to the (link) PipelineTemplate model documentation in the API reference."). I realize that's not as fast for navigating, though. I'm open to other ideas.

Thanks.

#7 - 10/20/2014 06:09 PM - Peter Amstutz

Brett Smith wrote:

Reviewing [4c0c769](#)

I like everything on this page. Thanks for writing it up. I understand this is intended to be introductory documentation and not a comprehensive reference, but even with that in mind, I think a few additions would be very helpful to beginners:

- A very brief (like, two or three sentences) explanation of what Docker is, with a link to its own documentation

Fixed.

- An explanation why this tutorial starts from the arvados/jobs image (i.e., note any Docker image you use has to include the SDKs for your Crunch script)

Fixed.

- Brief mention of arv keep docker's --project-uuid and --name options (doesn't even need to explain them in detail; punting with "run arv keep docker --help for details" would be fine by me)

Added a "Share Docker images" section that demonstrates --project-uuid. Discussing the --name option seems potentially confusing since it specifies the name of the collection (which I think already has a reasonable default) and not the repository name. I'll add something in if you still think it is important, though.

- Mentioning the different ways to specify the runtime constraint, depending on how much flexibility you want: Docker name, Collection UUID, or portable data hash

Fixed.

Beyond that, a few small notes about the intro bits:

- I think it'd be nice to have the word "Docker" somewhere in the title, so somebody who's heard that they need to use Docker can Ctrl+F it on the user guide and find it. Something like "Creating a custom Crunch environment with Docker."

Fixed.

- The standard blurb about having a usable runtime environment should be extended to mention that Docker itself must be installed on the system and usable from the user's account. (Maybe link to Docker's installation documentation for more?)

Fixed.

- In the ordered list of "what this page covers," move "how to" from the first bullet point to the leading text, for grammatical consistency.

Fixed.

And then one last organizational thing: I understand the rationale behind adding links to the Job and PipelineTemplate documentation from this guide, and I support making this easier to navigate, but this approach feels like a rough fit with our current organization. This document is billed as the User Guide (i.e., non-reference material); adding a "Reference" section to it suggests that we're getting a little away from ourselves. And if nothing else, having the terse titles "Job" and "PipelineTemplate" in the sidebar is kind of awkward; new users won't understand what's on those pages. I think a nicer way to tackle this quickly would be to add links to those pages from the appropriate guide pages (e.g., "For complete documentation about pipeline templates, refer to the (link) PipelineTemplate model documentation in the API reference."). I realize that's not as fast for navigating, though. I'm open to other ideas.

It is already the case that these pages are linked off of the relevant guide pages, but it's at the bottom and easy to miss. Instead of linking directly I

added a stub page with a link to the PipelineTemplate schema page so as give a slightly more informative title and not to jump directly from the user guide to the API section. Alternately I could just back this out because it's not directly related to documenting Docker.

I think there's a case to be made that the reference section should go into SDKs or some new section that documents specific tools and scripts, but we haven't built out that part of the documentation. There is a bigger reorganization story here.

#8 - 10/20/2014 07:03 PM - Brett Smith

Reviewing [ff76923](#)

Peter Amstutz wrote:

Brett Smith wrote:

- The standard blurb about having a usable runtime environment should be extended to mention that Docker itself must be installed on the system and usable from the user's account. (Maybe link to Docker's installation documentation for more?)

Fixed.

This is good, but it still doesn't mention anything about the user running the commands being allowed to send Docker commands (e.g., being in the docker system group). Because this is tripping people up (per recent IRC), I feel like it's important to mention very specifically.

It is already the case that these pages are linked off of the relevant guide pages, but it's at the bottom and easy to miss. Instead of linking directly I added a stub page with a link to the PipelineTemplate schema page so as give a slightly more informative title and not to jump directly from the user guide to the API section. Alternately I could just back this out because it's not directly related to documenting Docker.

I think there's a case to be made that the reference section should go into SDKs or some new section that documents specific tools and scripts, but we haven't built out that part of the documentation. There is a bigger reorganization story here.

Yeah, agreed. I'm okay with merging that in its current state.

From here, it's just little style consistency things:

- In the intro, there's a period at the end of the fourth point, but no others.
- "You may compile C programs or libraries" - may as well omit "C" since we already have users compiling code in other languages.
- Under "Install new packages," "docker run" should be monospace since it's a command.
- Under "Share Docker images," "UUID" in the text should be uppercase to match our other documentation.

Thanks.

#9 - 10/20/2014 07:25 PM - Peter Amstutz

Brett Smith wrote:

This is good, but it still doesn't mention anything about the user running the commands being allowed to send Docker commands (e.g., being in the docker system group). Because this is tripping people up (per recent IRC), I feel like it's important to mention very specifically.

Added some more text about it. This is tricky since we specifically don't want to write a "how to install and configure docker" guide.

From here, it's just little style consistency things:

- In the intro, there's a period at the end of the fourth point, but no others.
- "You may compile C programs or libraries" - may as well omit "C" since we already have users compiling code in other languages.
- Under "Install new packages," "docker run" should be monospace since it's a command.
- Under "Share Docker images," "UUID" in the text should be uppercase to match our other documentation.

Fixed.

#10 - 10/20/2014 07:33 PM - Brett Smith

Peter Amstutz wrote:

Brett Smith wrote:

This is good, but it still doesn't mention anything about the user running the commands being allowed to send Docker commands (e.g., being in the docker system group). Because this is tripping people up (per recent IRC), I feel like it's important to mention very specifically.

Added some more text about it. This is tricky since we specifically don't want to write a "how to install and configure docker" guide.

Yeah, I know there's a million ways to deal with the issue and I agree it's not our place to lay them all out. I wanted it to point out the requirement, but I don't think we need to go into detail about satisfying it. I like what you've come up with here. Just make it say "docker group" (monospace) both times, and this is good to merge. Thanks.

#11 - 10/20/2014 07:40 PM - Anonymous

- *Status changed from In Progress to Resolved*

- *% Done changed from 50 to 100*

Applied in changeset arvados|commit:a22d6e6bf03c6f274237f775a4c02122c9ce38ef.