

Arvados - Story #8884

[Docs] Pipeline author guide gives a basic demonstration of writing a pipeline template

04/05/2016 06:06 PM - Sarah Guthrie

Status:	Resolved	Start date:	04/05/2016
Priority:	Normal	Due date:	
Assigned To:	Bryan Cosca	% Done:	100%
Category:	Documentation	Estimated time:	0.00 hour
Target version:	2016-04-27 sprint		
Description			
Write a new wiki page describing how to write a pipeline template (wrapping a git repository and docker image together):			
<ul style="list-style-type: none">• How to wrap a git repository containing a crunch script and a docker image into a component<ul style="list-style-type: none">◦ Link to "Git Strategy for Pipeline Development" wiki page• How to write script_parameters<ul style="list-style-type: none">◦ dataclasses and their actual meanings ("Collection" passes in the pdh string, etc)• How to choose runtime_constraints<ul style="list-style-type: none">◦ The actual meaning of min_nodes◦ Setting max_tasks_per_node != 1			
Subtasks:			
Task # 8984: Review			Closed
Task # 8968: Review			Resolved

History

#1 - 04/12/2016 05:37 PM - Sarah Guthrie

- Description updated

- Story points set to 1.0

#2 - 04/13/2016 05:01 PM - Sarah Guthrie

- Target version changed from Pipeline Future Sprints to 2016-04-27 sprint

#3 - 04/13/2016 07:08 PM - Bryan Cosca

- Assigned To set to Bryan Cosca

#4 - 04/19/2016 02:57 PM - Bryan Cosca

- Status changed from New to In Progress

#5 - 04/19/2016 08:40 PM - Bryan Cosca

Ready for review: https://dev.arvados.org/projects/arvados/wiki/Pipeline_template_development

#6 - 04/20/2016 08:24 PM - Sarah Guthrie

We should not include "arvados_sdk_version" in the main example given the reported bugs in installing it. Mentioning it and stating that it's a good idea to require the arvados sdk to be installed in the docker image.

It would be awesome to connect the script_parameters (required, default, dataclass) with the effects they have on the workbench view (from a user's perspective).

#7 - 04/21/2016 02:59 PM - Bryan Cosca

I've updated the wiki to remove arvados_sdk_version and added screenshots for script_parameters.

#8 - 04/21/2016 04:16 PM - Sarah Guthrie

Feel free to link to this page if it will help you at all: https://dev.arvados.org/projects/arvados/wiki/Writing_a_Script_Calling_a_Third_Party_Tool

We currently are missing a description of "components" in this page, which are a fairly large part of pipeline templates. Simply adding a paragraph saying that pipeline templates are composed of a dictionary of components and that each component maps to a job would be helpful. We can then state that the rest of the document describes the specific pieces of a component.

The main example still has `arvados_sdk_version` defined.

The introductory paragraph under "Writing script parameters" should talk about the "required" flag so it doesn't come as a surprise later.

"yields this example" is fairly vague. Saying that a particular pipeline template yields a pipeline instance is more specific and accurate.

How does a dataclass "File" influence the view? What about the dataclass "number"?

The following is inconsistent enough to be confusing:

The inputs tab in the pipeline instance page shows all the required parameters.

...

For the 'additional_params' parameter, since its not required, its in the 'Components' tab, where you can set it:

Maybe try?

The "Inputs" tab in the pipeline instance page shows all the required parameters.

...

The "Components" tab in the pipeline instance page shows all the parameters. Thus it is the only place where non-required parameters, such as 'additional_params' may be set.

Why is "read_group" highlighted in red in the Components tab, but nothing else is? That will likely be distracting to a new viewer.

One runtime constraint is `docker_image`.

What does `docker_image` control? You describe hints for it, but don't say what it actually does. Is it required? What needs to be in the docker image?

The `max_tasks_per_node` parameter will allow you to allocate more computations on your node

What are "computations"? We need to use more specific language here. What happens if `max_tasks_per_node` is equal to 2? When will multiple jobs be scheduled on that node? When will multiple tasks be scheduled on that node?

#9 - 04/21/2016 07:14 PM - Bryan Cosca

Why is "read_group" highlighted in red in the Components tab, but nothing else is? That will likely be distracting to a new viewer.

I don't know. I could hack it by making the template set it as false, but this is what it looks like when a Text/number parameter is 'required'.

Everything else has been added.

#10 - 04/21/2016 07:45 PM - Sarah Guthrie

Alright, I'm happy with this. Brett is up next.

#11 - 04/29/2016 02:36 PM - Bryan Cosca

- Status changed from *In Progress* to *Resolved*

#12 - 04/29/2016 02:36 PM - Bryan Cosca

resolving due to time constraints and brett does not have time to review and sally reviewed already