Arvados - Feature #4456

[Workbench] Provide more feedback about when a queued job is likely to start.

11/06/2014 09:30 PM - Brett Smith

Status:	New		Start date:	11/06/2014
Priority:	Normal		Due date:	
Assigned To:	Tom Clegg		% Done:	0%
Category:	Workbench		Estimated time:	0.00 hour
Target version:				

Description

From discussion in IRC, users often feel like the time their jobs start is effectively random. It would help to provide more visibility about what the cluster is doing.

Currently, if you look at the page for a queued job, it says, "There are N jobs in the queue ahead of this one." This is a good start, but it's still not sufficient information for users to understand what's happening. It would also help them to know whether or not there are nodes free right now, and whether or not a node will boot to run their job (see #4446).

We should display information about this on the Dashboard. For example, "Your next job in the queue is <link>. There are N jobs in the queue ahead of it. <text about node freeness>"

Subtasks:

Task # 5860: Provide detail about expected information/presentation	**New**

Related issues:

Related to Arvados - Feature #4446: [Workbench] Provide feedback on dashboard...	**Closed**	
Related to Arvados - Feature #5513: Node manager should always have one node ...	**Resolved**	**03/19/2015**
Related to Arvados - Feature #3605: [Workbench] improved dashboard page	**Closed**	**09/15/2014**

History

#1 - 11/06/2014 09:33 PM - Brett Smith

- Subject changed from [Workbench] Provide more visibility about state of jobs to [Workbench] Provide more visibility about state of queued jobs

#2 - 11/06/2014 09:43 PM - Nancy Ouyang

Well, just pulling out to the Workbench homepage "there are N jobs ahead" would be a great start, I didn't think to look inside the pipeline. Or job.

My specific issue is I'm trying to learn Arvados by running short pipelines, and I just want to know if I should expect my pipeline to take longer than a minute to start running, in which case I'll go do something else. After hitting run I'll go to the homepage and stare expectantly at the pipeline (the page does a good job of conveying that it is actively tracking the job and I don't need to hit refresh)

So my question is actually "Why is my pipeline not running" and I try to estimate it by "How long did the previous ones take to start" and I can't because it seems random to me.

Other possible solutions: A help topic, "Why isn't my job running" that explains the possible reasons.

At the core, being able to play around and get results quickly will make it more pleasant to learn Arvados. Sandboxes? Arv run? Arvados-like local docker? Immediately running pipelines? are all addressing this.

#3 - 11/07/2014 06:34 PM - Ward Vandewege

- Target version changed from Bug Triage to Arvados Future Sprints

#4 - 11/14/2014 06:43 PM - Tom Clegg

- Story points set to 1.0

#5 - 11/19/2014 06:43 AM - Tom Clegg

- Tracker changed from Bug to Feature

- Subject changed from [Workbench] Provide more visibility about state of queued jobs to [Workbench] Provide more feedback about when a queued job is likely to start.

#6 - 04/08/2015 08:20 PM - Tom Clegg

- Target version changed from Arvados Future Sprints to 2015-05-20 sprint

#7 - 04/29/2015 07:40 PM - Tom Clegg

- Assigned To set to Tom Clegg

#8 - 05/05/2015 02:56 PM - Tom Clegg

Some system states that could be translated to a start-time prediction:

- All worker nodes are busy. Nodemanager will probably do something about this in 10(?) seconds.
- Some worker nodes are bootstrapping or idle, but they'll be consumed by jobs ahead of yours. Nodemanager will probably do something about this in 10(?) seconds.
- After accounting for jobs ahead of yours in the queue, #nodes needed by your job are bootstrapping. Of the workers needed, the most recently started is X seconds old; bootstrapping is usually done in Y seconds according to workbench config (or according to API? a configured or computed ETA could be offered in the nodes#index API response).
- There are enough idle nodes to run your job now. Your job will probably start in 10(?) seconds.

#9 - 05/05/2015 08:20 PM - Tom Clegg

Possible ways/places to present this information:

- Pipeline instance #show → Components
  - If component state is Queued, show ETA / summary of system state as it relates to this job ("worker nodes spinning up, ETA 2m")
- Dashboard → Active pipelines
  - (?)Show nearest ETA of any queued job (if any)
- Dashboard → Compute and job status
  - (?) Show table of nodes (name, state, cores, ram, scratch)
  - Replace "submitted" column with ETA

If nodemanager isn't running:

- If there are enough idle nodes to run all jobs in the queue up to & including this one, ETA is 10(?) seconds.
- Hard to say much otherwise. Number of jobs ahead of yours?
- (How does Workbench know nodemanager isn't running?)

#10 - 05/19/2015 10:01 PM - Brett Smith

- Target version changed from 2015-05-20 sprint to Arvados Future Sprints

#11 - 07/06/2021 09:26 PM - Peter Amstutz

- Target version deleted (Arvados Future Sprints)