

## Arvados - Story #10679

### [Workbench] Update job stats to be more useful to end user

12/07/2016 04:37 PM - Tom Morris

<b>Status:</b>	New	<b>Start date:</b>	12/07/2016
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assigned To:</b>		<b>% Done:</b>	0%
<b>Category:</b>		<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	Arvados Future Sprints		

#### Description

As was highlighted during the grooming [#10516](#), our job stats could be more useful to user. This needs to be refined, but represents a first cut.

Here's an example of a current CWL job:

-----

This pipeline started at 9:34 AM 12/7/2016. It failed after 57m at 10:31 AM 12/7/2016. It ran for 6m(51mqueued) and used 6m of node allocation time (1.0 scaling).

created\_at: 9:34 AM 12/7/2016  
started\_at: 9:34 AM 12/7/2016  
finished\_at: 10:31 AM 12/7/2016

cwl-runner

Failed 6m / 6m (1.0) Output of cwl-runner

This job started at 10:25 AM 12/7/2016. It failed after 6m at 10:31 AM 12/7/2016.

It ran for 1m(5mqueued) and used 1m of node allocation time (1.0 scaling).

-----

Some things that I think we could improve:

- don't call the time the job was submitted a "started" time
- don't report top level CWL job separately from the overall workflow since the user basically considers them both part of the system machinery

Stats that are of interest to user:

- total wall clock time
- total core hours
- total cost

(for above three, both actuals for this run & totals with stats from previously run & reused jobs)

- queuing time until the first job in the workflow started
- queuing & overhead (separately?) time during the execution of the workflow
- ? maximum width / parallelism ?
- for individual jobs: keep cache hit rate & utilization, maximum CPU utilization, & other useful stats from crunchstat-summary

#### History

#1 - 12/07/2016 09:01 PM - Tom Morris

- Subject changed from Update job stats to be more useful to end user to [Workbench] Update job stats to be more useful to end user