

Arvados - Feature #8123

[Crunch] Add crunchstat-summary --pipeline-instance, --format=html, and --include-child-jobs features

01/06/2016 04:03 PM - Tom Clegg

Status:	Resolved	Start date:	01/06/2016
Priority:	Normal	Due date:	
Assigned To:	Tom Clegg	% Done:	100%
Category:	Crunch	Estimated time:	0.00 hour
Target version:	2016-02-03 Sprint		
Description			
Subtasks:			
Task # 8255: Review 8123-crunchstat-graphs			Resolved

Associated revisions

Revision fc3b155f - 01/25/2016 09:08 PM - Tom Clegg

Merge branch '8123-crunchstat-graphs' closes #8123

Revision 1dbf2e03 - 02/03/2016 05:50 PM - Tom Clegg

8123: Install chart.js asset file.

...during "setup.py install" too, not just when installing via package.

refs #8123

History

#1 - 01/06/2016 04:17 PM - Tom Clegg

- Subject changed from [Crunch] Add crunchstat-summary --format=html and --include-child-jobs features to [Crunch] Add crunchstat-summary --pipeline-instance, --format=html, and --include-child-jobs features

8123-crunchstat-graphs @ [d24103c](#) has

- adjust RAM recommendations to account for bugs elsewhere:
 1. cloud providers' "node size" lists tend to advertise nodes as having 8192 MiB RAM, even though the nodes turn out to have a bit less than 8192 MiB RAM
 2. node manager doesn't expect this problem to happen, so if you ask for a number like 8192 that matches an advertised node size exactly, node manager brings up an 8192 MB node; when the node comes up, "node ping" reports the smaller amount; and crunch-dispatch does not run the job on that node, because it is slightly too small.
- --format=html makes a (rather large) html file that uses canvas.js to draw graphs
- --pipeline-instance=UUID generates summaries/graphs for each job in the given pipeline instance
- --include-child-jobs includes stats for "child jobs" submitted by a Queue-like job, if any are mentioned in the log
- --debug prints some debug logs as we go

#2 - 01/12/2016 06:26 PM - Sarah Guthrie

Fails on job <https://workbench.tb05z.arvadosapi.com/jobs/tb05z-8i9sb-khsk5rmf4xjdcbl>

#3 - 01/20/2016 06:01 PM - Brett Smith

- Target version set to 2016-02-03 Sprint

#4 - 01/25/2016 12:06 AM - Brett Smith

Reviewing [0d3411ab](#), and this is very nice. There's a small number of small things that need to get fixed before merge:

- setup.py needs to declare that tests_require mock now.
- The warning '%s: omitting %s (try --include-child-job)' typos the option name, and should say --include-child-jobs. It might be extra-nice if it explained the omission a hair more, like omitting stats from child job %s.
- We should escape any HTML entities in the chart label when we generate it. cgi.escape is the obvious method to do so.

Beyond that, questions are more subjective thoughts:

- I would encourage us to consider enabling `--include-child-jobs` by default. Is there any situation where the user wouldn't immediately rerun with it on after seeing the warning?
 - I'm having a lot of trouble following `_recommend_ram`. I especially don't understand why the "convert a real amount to an amount to request" function involves dividing by `AVAILABLE_RAM_RATIO` rather than multiplying, and why it involves dividing again by 1024. Could you please explain?
- Readability suggestions I might make to help make this easier to follow:
- Define a function to return `amount / AVAILABLE_RAM_RATIO / 1024`. A good name and maybe a docstring could help answer this question for future readers.
 - Variable names can help keep track of what conversions have taken place. e.g., when we convert the amount of used RAM into MiB, maybe that could be `used_ram_mb`?
 - 1024 seems easier to read and write than `(1<<10)`.
- `class PipelineSummarizer()` - Please inherit from something, at least object, to help avoid surprising future developers with old-style class behavior.
 - `if isinstance(job, str)` - Prefer `basestring` over `str` here to accept Unicode strings as well.
 - When I read the help for `--verbose` vs. `--debug`, I thought they toggled separate streams of logs. I was a little surprised to see `--debug` implies `--verbose`. Would accepting multiple `--verbose` options be closer to your intent? `dockercleaner` has an example of this.
 - I feel like the way existing `_constraints` is defined obscures what's going on a bit. The mutable class variable makes me think it's going to hold data shared across instances, but no, it's meant to be a fallback default for classes that don't define this instance variable. I feel like having `Summarizer.__init__` check if it's defined and define an instance variable if not would be clearer, and help prevent future bugs of modifying the default dictionary. If not that, a comment would be helpful to add.
 - `JobSummarizer.__init__` defines `self.label` twice.

Thanks.

#5 - 01/25/2016 06:44 AM - Tom Clegg

I don't see it mentioned in the docs but the playground at <http://canvasjs.com/docs/charts/basics-of-creating-html5-chart/title/> indicates HTML in the title text already gets escaped by canvasjs -- if we translate "<" to "<" then we'll see "<" on the screen.

For `"-v"` / `"-vv"` I used `WARNING-10*verbose` instead of `ERROR-10*verbose` as in `dockercleaner`. The Python default level is `WARNING` and I thought it would be weird if you had to use `"-v"` to see warnings. Perhaps add `"-q` for quiet" for hiding `warn/err/crit`? Change to `ERROR-10*verbose` to match `dockercleaner`? (It doesn't seem like a huge issue here either way, but if you have an opinion I'll take it.)

Everything else changed as suggested, thanks. →commit:fdd28d6

(I might have gone a little overboard on the RAM-calculation comment... hopefully that still counts as an improvement.)

#6 - 01/25/2016 03:44 PM - Brett Smith

Tom Clegg wrote:

I don't see it mentioned in the docs but the playground at <http://canvasjs.com/docs/charts/basics-of-creating-html5-chart/title/> indicates HTML in the title text already gets escaped by canvasjs -- if we translate "<" to "<" then we'll see "<" on the screen.

Sorry, I meant where it appears in the HTML title tag, not the chart JSON (I figured ChartJS should escape that). It gets that too?

For `"-v"` / `"-vv"` I used `WARNING-10*verbose` instead of `ERROR-10*verbose` as in `dockercleaner`. The Python default level is `WARNING` and I thought it would be weird if you had to use `"-v"` to see warnings. Perhaps add `"-q` for quiet" for hiding `warn/err/crit`? Change to `ERROR-10*verbose` to match `dockercleaner`? (It doesn't seem like a huge issue here either way, but if you have an opinion I'll take it.)

No strong opinion. I am comfortable with `crunchstat-summary` and `dockercleaner` having different default levels, since their roles and interaction methods are pretty different: one is a user tool and the other is a system daemon.

Everything else changed as suggested, thanks. →commit:fdd28d6

Looks good to me. Thanks.

#7 - 01/25/2016 10:15 PM - Tom Clegg

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

- % Done changed from 0 to 100

Applied in changeset `arvados|commit:fc3b155fb7ad82cf97cdb81a02a9cfaebe4389ff`.