Arvados - Bug #9442

[CWL] runtime.outdir is not set to task.outdir by arvados-cwl-runner

06/20/2016 06:35 PM - Sarah Guthrie

Status:	Resolved		**Start date:**	06/20/2016
Priority:	Normal		**Due date:**	
Assigned To:	Peter Amstutz		**% Done:**	100%
Category:			**Estimated time:**	0.00 hour
Target version:	2016-07-06 sprint			

Description

In pipeline instance: https://workbench.qr1hi.arvadosapi.com/pipeline_instances/qr1hi-d1hrv-3h20r1mg6tp6nan

Jobs:

- https://workbench.qr1hi.arvadosapi.com/jobs/qr1hi-8i9sb-85b5vxu829r4e0u
- https://workbench.qr1hi.arvadosapi.com/jobs/qr1hi-8i9sb-rso53wu1aa00h3o

qr1hi-8i9sb-85b5vxu829r4e0u: valueFrom: $(runtime.outdir)/RNAqc.txt gets set to "/tmp/tmpzBkLFb/RNAqc.txt"
qr1hi-8i9sb-rso53wu1aa00h3o: valueFrom: $(runtime.outdir)/RNAqc.txt gets set to "/tmp/tmp2x83tZ/RNAqc.txt"

Both jobs failed with similar errors:

```
crunchrunner: $(task.outdir)=/tmp/crunch-job-task-work/compute3.1/outdir
...
Exception in thread "main" htsjdk.samtools.SAMException: Cannot write file: /tmp/tmp2x83tZ/RNAqc.p
df. Neither file nor parent directory exist.
```

Subtasks:

Task # 9462: Review 9442-cwl-runtime-outdir	**Resolved**
Task # 9475: Investigate	**Resolved**

Associated revisions

Revision 212c20e5 - 06/28/2016 07:50 PM - Peter Amstutz

Merge branch '9442-cwl-runtime-outdir' closes #9442

History

#1 - 06/21/2016 05:27 PM - Brett Smith

- Target version set to 2016-07-06 sprint

#2 - 06/22/2016 07:09 PM - Peter Amstutz

- Assigned To set to Peter Amstutz

#3 - 06/22/2016 07:23 PM - Peter Amstutz

- Story points set to 0.5

#4 - 06/28/2016 02:07 PM - Peter Amstutz

9442-cwl-runtime-outdir has fix.

When running a workflow, the outdir= option sets output directory for the workflow as a whole, not the output directory for each individual tool. This fix sets the outdir= option for individual tool jobs.

#5 - 06/28/2016 02:44 PM - Brett Smith

Reviewing 966acb6

Will this code behave well when self.work_api == 'containers'? Since $(task.outdir) is specific to run-command, which I don't think we're using for container requests, I'm worried setting this string unconditionally will cause trouble for container requests.

Does it make sense to set this as a default value, rather than always overwriting it? (kwargs.setdefault('outdir', '$(task.outdir)')) That would make this

class more extensible in the future, although I admit it's hard to say how important that is right now.

It would be great to see tests for this, for both jobs and containers work APIs.

Thanks.

#6 - 06/28/2016 02:46 PM - Sarah Guthrie

Trying to run arvados-cwl-runner without specifying the api flag (since I was unsure what it was for) caused this error:

```
Traceback (most recent call last):
6/28/2016 10:41:31 AM      compute0    1      task-print    0      File "/tmp/crunch-job/src/crunch_scripts/cwl
-runner", line 50, in <module>
6/28/2016 10:41:31 AM      compute0    1      task-print    0      runner = arvados_cwl.ArvCwlRunner(api_client
=arvados.api('v1', model=OrderedJsonModel()), work_api="jobs")
6/28/2016 10:41:31 AM      compute0    1      task-print    0      TypeError: __init__() got an unexpected keyw
ord argument 'work_api'
```

#7 - 06/28/2016 07:47 PM - Peter Amstutz

9442-cwl-runtime-outdir updated with containers fixed, testing, and a fix for the cwl-runner crunch script.

#8 - 06/28/2016 07:47 PM - Brett Smith

The branch at 8ffcc8c looks good to me. A couple of parting nits:

In __init__.py, "# todo: autodetect API to use." - It could be nice to extend this comment to suggest that we should check the local environment to try to see whether we're currently running as an Arvados job or container, and set work_api to match, and then detect the API server's preference if that turns up nothing. A strategy like that would keep the Crunch script working as-is, whereas a strategy that only asks the API server would break it.

In "%s/%s" % (self.outdir, self.stdout), using os.path.join would be a little cleaner and safer.

Thanks.

#9 - 06/28/2016 07:55 PM - Peter Amstutz

- Status changed from New to Resolved

- % Done changed from 50 to 100

Applied in changeset arvados|commit:212c20e55437a5e67482e1db971d0d826591e856.