Arvados - Story #8654

[CWL] Incorporate cwl-runner into arvados/jobs

03/07/2016 04:46 PM - Brett Smith

Status:	Resolved		**Start date:**	03/07/2016
Priority:	Normal		**Due date:**	
Assigned To:	Peter Amstutz		**% Done:**	100%
Category:	Crunch		**Estimated time:**	0.00 hour
Target version:	2016-03-30 sprint			

Description

We want to do this to make it as easy as possible to run CWL workflows in Arvados. Adding it to arvados/jobs makes that simple because we already have infrastructure to keep it up-to-date on clusters.

- Update the arvados/jobs image to be based on Debian jessie.
- Add the cwl-runner package to arvados/jobs.
- Add a cwl-runner crunch_script to the Arvados repository to call arvados_cwl by filling in inputs from the current job's script_parameters.
- Add arvados-cwl-runner --submit option that submits a job that runs cwl-runner

Acceptance criteria:
- User can use arvados-cwl-runner --submit tool.cwl input.json to submit a CWL job, including uploading any workflow and input dependencies that are not already present on the target Arvados instance. This will create a job to execute the CWL workflow and submits additional jobs that do the actual work so that the workflow can run unattended similarly to existing Arvados pipelines.

Subtasks:

Task # 8777: Pass CWL conformance tests with --submit	**Resolved**
Task # 8735: Add cwl-runner --submit	**Resolved**
Task # 8736: Review	**Resolved**

Related issues:

Related to Arvados - Story #8815: [Crunch] crunch-job bind mounts crunchrunne...	**Resolved**	**03/29/2016**

Associated revisions

Revision cdcddca6 - 03/31/2016 03:35 PM - Peter Amstutz

Merge branch '8654-arv-jobs-cwl-runner' closes #8654

History

#1 - 03/07/2016 04:47 PM - Brett Smith

- Description updated

#2 - 03/08/2016 07:14 PM - Brett Smith

- Subject changed from [CWL] Verify cwl-runner runs inside an Arvados job to [CWL] Incorporate cwl-runner into arvados/jobs

- Description updated

- Story points set to 1.0

#3 - 03/08/2016 07:14 PM - Brett Smith

- Target version set to Arvados Future Sprints

#4 - 03/08/2016 07:41 PM - Brett Smith

- Description updated

#5 - 03/16/2016 05:18 PM - Brett Smith

- Target version changed from Arvados Future Sprints to 2016-03-30 sprint

- Status changed from New to In Progress

- Assigned To set to Peter Amstutz

#6 - 03/16/2016 07:42 PM - Peter Amstutz

- Description updated

#7 - 03/28/2016 05:02 PM - Peter Amstutz

- Description updated

Implementation notes:

- docker/jobs/Dockerfile:

Updated to include "nodejs" and "arvados-cwl-runner" packages.

- crunch_scripts/cwl-runner

Crunch script which bootstraps running the workflow runner in a crunch job. Workflow inputs are provided in "script_parameterS".

- sdk/python/arvados/commands/run.py

Tweak the "upload_files" feature imported from arv-run.

- sdk/cwl

Add support for "RunnerJob" which submits the cwl-runner job, and optionally (with --wait) waits for the workflow to complete.

#8 - 03/29/2016 03:45 PM - Radhika Chippada

Review comments:

- crunch_scripts/cwl-runner:
    - This print statement "print cwltool.main.versionstring()" : not sure it's something you forgot to remove. If not, can it be converted into a log statement with some description?
    - Can you please add a detailed description / comment about this script at the top of the file? Most importantly describing how the relationship between this and sdk/cwl/arvados-cwl-/__init__.py works would be good to have here.
    - In "np = argparse.Namespace()" : I could not tell why it is named as "np". May be args (if it does not conflict with other things or jobArgs or something …

- sdk/cwl/arvados_cwl/__init__.py
- Can you add a few comments, most importantly for the newly added RunnerJob? I also noticed that several others such as ArvadosJob and ArvCwlRunner could use some brief descriptions

- sdk/python/arvados/commands/run.py
    - typo in comment at line 113 "pathes"

- There were many places where variable names could use camel casing and "_" to separate words, but your call

- Would it be too hard to write some tests at least for the updates in *init*.py ?

- Does this need any documentation updates?

#9 - 03/30/2016 08:30 PM - Peter Amstutz

Radhika Chippada wrote:

> Review comments:
>
> - crunch_scripts/cwl-runner:
>   - This print statement "print cwltool.main.versionstring()" : not sure it's something you forgot to remove. If not, can it be converted into a log statement with some description?

That's to print the versions of several key packages to help with provenance/debugging. Converted to a logging statement.

> - Can you please add a detailed description / comment about this script at the top of the file? Most importantly describing how the relationship between this and sdk/cwl/arvados-cwl/__init__.py works would be good to have here.

Added comment.

> - In "np = argparse.Namespace()" : I could not tell why it is named as "np". May be args (if it does not conflict with other things or jobArgs or something …

Renamed to "args".

- sdk/cwl/arvados_cwl/__init__.py
- Can you add a few comments, most importantly for the newly added RunnerJob? I also noticed that several others such as ArvadosJob and ArvCwlRunner could use some brief descriptions

Added brief documentation strings to all the classes in that module.

- sdk/python/arvados/commands/run.py
  - typo in comment at line 113 "pathes"

Fixed.

- There were many places where variable names could use camel casing and "_" to separate words, but your call

I don't want to do this right now because of risk that in could introduce typos and hard-to-find bugs.

- Would it be too hard to write some tests at least for the updates in *init*.py ?

I had tests but failed to commit them. Thanks for that. Test added.

- Does this need any documentation updates?

Probably, but not in this story. I think this needs to get merged and squared away before hitting the documentation. Also as discussed I'm continuing to work on the CWL user guide this sprint.

Thanks.

#10 - 03/31/2016 01:44 PM - Radhika Chippada

Reviewed at [f7e07c3a28d4f4e1adbd957749bd825e44a2e523](#)

- Thanks for the comments. Made it easier to understand the code.

- In sdk/cwl/arvados_cwl/__init__.py :
  - The argument description for —local is a bit confusing. Can we skip "and submits jobs" here?
  - In the argument description for —submit, can we clarify that the job is submitted to (arvados?) server?

Tweaked the help text to try and make it a bit clearer.

- versionstring(): Is the ordering of strings as intended here? (for the last 4) do we want it to be: "arvados-python-client", arvpkg[0].version, "cwltool", cwlpkg[0].version instead?

You're right I messed that up. Thanks for catching that.

- sdk/cwl tests are failing (failures=1, errors=2) with run-tests script

Fixed.

- (If comments are allowed) I think adding a link in tests/*/*.cwl files briefly explaining what they are intended for might be helpful

Added comments.

- Can you please rename "tests/inp" as "tests/input" or "tests/inputs" ?

Done.

- (Since python is still new to me, I urge) Can you please revisit the indentations in test_submit.py at lines 37, 38, and 42? I am thinking in api.collections().create.assert_has_calls there are 5 mock.call(…) statements but they are not all lined up and I had difficulty understanding at first. Thanks.

Done.

Now at [9d035125ea1f8b6bd32fa6f862b32ba9308ebc66](#)

#11 - 03/31/2016 03:40 PM - Peter Amstutz

- Status changed from In Progress to Resolved

- % Done changed from 67 to 100

Applied in changeset arvados|commit:cdcddca613c896fd8395a4045c858945451c3fa0.