Arvados - Feature #8558

[CWL] arvados-cwl-runner propagates ResourceRequirement

02/26/2016 04:58 PM - Peter Amstutz

Status:	Resolved		**Start date:**	03/08/2016
Priority:	Normal		**Due date:**	
Assigned To:	Radhika Chippada		**% Done:**	100%
Category:			**Estimated time:**	0.00 hour
Target version:	2016-03-16 sprint			

Description

Propagate draft-3 ResourceRequirement (cores/ram) to crunch runtime_constraints (min_cores_per_node, min_ram_mb_per_node) in arvados-cwl-runner.

References:

https://common-workflow-language.github.io/draft-3/CommandLineTool.html#ResourceRequirement

http://doc.arvados.org/api/schema/Job.html

In arvados/sdk/cwl/arvados_cwl/__init__.py ArvadosJob.run() you can acces self.builder.resources which is a dict containing keys for "cores", "ram", "tmpdirSize" and "outdirSize". These need to be translated to the appropriate arvados keys in runtime_constraints.

Subtasks:

Task # 8611: Review branch: 8558-cwl-propagate-resource-req   **Resolved**

Associated revisions

Revision be8fb727 - 03/04/2016 04:10 PM - Peter Amstutz

Self test script for running CWL conformance tests on arvados-cwl-runner
against Arvados arvbox instance. refs #8558

Revision 4bd414c0 - 03/04/2016 04:19 PM - Peter Amstutz

Bugfix for CWL conformance test script, refs #8558

Revision e11ee6c4 - 03/04/2016 04:23 PM - Peter Amstutz

CWL test select config, refs #8558

Revision 2fe15804 - 03/10/2016 04:39 PM - Radhika Chippada

refs #8558
Merge branch '8558-add-cwl-jenkins'

Revision 2fe15804 - 03/10/2016 04:39 PM - Radhika Chippada

refs #8558
Merge branch '8558-add-cwl-jenkins'

Revision a2d5d6b2 - 03/11/2016 02:46 PM - Radhika Chippada

closes #8558
Merge branch '8558-cwl-propagate-resource-req'

History

#1 - 02/29/2016 06:25 PM - Brett Smith

- Target version set to 2016-03-16 sprint

#2 - 03/01/2016 07:37 PM - Peter Amstutz

- Subject changed from [CWL] arvados-cwl-runner propagates ResourceRequirements to [CWL] arvados-cwl-runner propagates ResourceRequirement

- Description updated

#3 - 03/01/2016 07:42 PM - Brett Smith

- Story points set to 0.5

#4 - 03/02/2016 02:27 PM - Peter Amstutz

- Description updated

#5 - 03/02/2016 02:28 PM - Peter Amstutz

- Description updated

#6 - 03/02/2016 03:40 PM - Brett Smith

- Assigned To set to Radhika Chippada

#7 - 03/04/2016 01:16 PM - Radhika Chippada

- Added coresMin and ramMin from builder resources into runtime_constraints. Per the documentation, these can be long or string typed and hence I converted them into integer before adding to runtime_constraints (our documentation says we use integers).

- In our documentation, I did not find the equivalents of tmpdirSize and outdirSize. Hence, I did not handle these.

- I verified any syntax errors in the code I added, but could not test. We have absolutely no automated tests!! And, I could not figure out how to at least test manually from
  https://dev.arvados.org/projects/arvados/wiki/Running_Common_Workflow_Language_%28CWL%29_workflows_on_Arvados

#8 - 03/04/2016 02:23 PM - Radhika Chippada

- Status changed from New to In Progress

#9 - 03/04/2016 03:26 PM - Brett Smith

Radhika Chippada wrote:

- In our documentation, I did not find the equivalents of tmpdirSize and outdirSize. Hence, I did not handle these.

The constraint min_scratch_mb_per_node is pretty closely analogous to tmpdirSize. Both specify how much working disk space the job should be able to write to while it runs.

We don't have anything like outdirSize, so that shouldn't be handled, you're right.

#10 - 03/08/2016 03:25 PM - Peter Amstutz

The resources dict contains keys "cores", "ram", "tmpdirSize" and "outdirSize". This is generated by evalResources:
https://github.com/common-workflow-language/cwltool/blob/master/cwltool/process.py#L305

min_scratch_mb_per_node should be the sum of tmpdirSize and outdirSize because currently we use the same partition for both scratch space and output staging.

Python style note, if "cores" in resources is a more idiomatic way of writing if "cores" in resources.keys() and probably a bit more efficient (resources.keys() returns a list).

#11 - 03/09/2016 07:44 PM - Radhika Chippada

- Arvados branch 8558-cwl-propagate-resource-req:
  - Updated to use the correct names (tmpdirSize instead of tmpdirMin etc) and other suggested updates.
  - Updated a couple typos in the tests. Also, updated the tests to expect fields from default resources when one or more are omitted.

- Added branch 8558-add-cwl-jenkins in arvados-jenkins

#12 - 03/10/2016 04:03 PM - Peter Amstutz

```
            if "cores" in resources:
                try:
                    runtime_constraints["min_cores_per_node"] = int(resources["cores"])
                except:
                    runtime_constraints["min_cores_per_node"] = None
```

This should be written more compactly (the "get" method returns None if the key is unavailable):

```
  runtime_constraints["min_cores_per_node"] = resources.get("cores")
```

Similarly (the optional 2nd parameter to the "get" method is the value to return if the key is unavailable):

```
  runtime_constraints["min_scratch_mb_per_node"] = resources.get("tmpdirSize", 0) + resources.get("outdirSize"
, 0)
```

#13 - 03/10/2016 04:08 PM - Peter Amstutz

Actually, it's a bad idea to set runtime_constraints["min_cores_per_node"] to None, it should probably have a default value as well, eg runtime_constraints["min_cores_per_node"] = resources.get("cores", 1)

(Although in practice the "cores" key should always be there so the default doesn't really matter.)

#14 - 03/11/2016 02:50 PM - Radhika Chippada

- Status changed from In Progress to Resolved

- % Done changed from 0 to 100

Applied in changeset arvados|commit:a2d5d6b24dde2de391831aa122cfda8e2cb759e0.