

## Arvados - Bug #12933

### [crunch2] add equivalent of cloud\_node line

01/09/2018 06:49 PM - Ward Vandewege

<b>Status:</b>	Resolved	<b>Start date:</b>	01/11/2018
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assigned To:</b>	Tom Clegg	<b>% Done:</b>	100%
<b>Category:</b>	Crunch	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	2018-01-17 Sprint		
<b>Description</b>			
In crunchv1, a line is logged in every job that looks like this: <pre>{"cloud_node":{"size":"Standard_D3_v2","price":0.229},"total_cpu_cores":4,"total_scratch_mb":204695,"total_ram_mb":14023}</pre>			
We need the equivalent in crunchv2. The node-info output is useful, but the string above has several other things going for it: a) it has cloud node information (node type, price, actual ram) b) it is machine parsable			
The format doesn't need to be exactly like the above ("size" seems a misnomer for "instance_type", e.g.), but it needs to have all relevant cloud node information and it should be machine parsable.			
<b>Subtasks:</b>			
Task # 12946: Review 12933-log-node-properties			<b>Resolved</b>
<b>Related issues:</b>			
Related to Arvados - Feature #12746: [crunch2] Add I/O (and other?) stats to ...		<b>Resolved</b>	<b>01/26/2018</b>
Related to Arvados - Bug #12465: [crunchv2] Improve crunch-run environment re...		<b>New</b>	

### Associated revisions

#### Revision 453f922b - 01/11/2018 07:04 PM - Tom Clegg

Merge branch '12933-log-node-properties'

refs #12933

Arvados-DCO-1.1-Signed-off-by: Tom Clegg <[tclegg@veritasgenetics.com](mailto:tclegg@veritasgenetics.com)>

### History

#### #1 - 01/09/2018 06:49 PM - Ward Vandewege

- Related to Feature #12746: [crunch2] Add I/O (and other?) stats to crunch-run added

#### #2 - 01/09/2018 07:29 PM - Tom Clegg

This is how we do it in crunch1, in [source:sdk/cli/bin/crunch-job](#) (where @node is a list of node names obtained from `sinfo [...]` `--nodes=$SLURM_NODELIST`)

```
my $resp = api_call(
  'nodes/list',
  'filters' => [['hostname', 'in', \@node]],
  'order' => 'hostname',
  'limit' => scalar(@node),
);
for my $n (@{$resp->{items}}) {
  Log(undef, "$n->{hostname} $n->{uuid} ".JSON::encode_json($n->{properties}));
}
```

#### #3 - 01/09/2018 08:34 PM - Tom Clegg

In crunch2 we can add this to the node-info logger: If `$SLURMD_NODENAME` is not empty, call `/arvados/v1/nodes?filters=[[hostname,=$nodename]]` and print the uuid and properties hash of the returned item (if any).

#### #4 - 01/09/2018 09:59 PM - Ward Vandewege

Since a container is like a unix process, i.e. it runs exactly once, it would sure be nice if this information was captured in the container object. Then we don't even need to add it to the logs.

#### #5 - 01/10/2018 07:59 AM - Tom Clegg

- Status changed from New to In Progress

- Assigned To set to Tom Clegg

12933-log-node-properties @ [e128fc5885c553c9e9b55f2529d0ea6937e5a6b7](#)

#### #6 - 01/10/2018 03:31 PM - Peter Amstutz

- Related to Bug #12465: [crunchv2] Improve crunch-run environment reporting added

#### #7 - 01/11/2018 02:45 PM - Tom Clegg

12933-log-node-properties @ [5469772c43759b8bde77c3d78450658e266b9cf0](#)

This version saves a node.json file in the log (analogous to container.json), with the admin-only "info" field removed. This should be easy to json.Unmarshal into the new Node type in the Go SDK.

#### #8 - 01/11/2018 02:51 PM - Tom Clegg

- Target version changed from To Be Groomed to 2018-01-17 Sprint

#### #9 - 01/11/2018 06:14 PM - Lucas Di Pentima

- File services/crunch-run/crunchrun.go
  - Line 749: The comment seems to need an update
  - Lines 741 & 750: I think the params argument should be passed to the CallRaw call, right?

#### #10 - 01/11/2018 06:19 PM - Tom Clegg

Fixed both issues.

12933-log-node-properties @ [813f5f4aad5da71c4fcfe6639c9010e1056acf1f](#)

#### #11 - 01/11/2018 06:52 PM - Lucas Di Pentima

LGTM, thanks!

#### #12 - 01/13/2018 04:51 PM - Ward Vandewege

This works but only when the workflow is run without --local.

I ran a test workflow with

```
cwl-runner --local download.cwl --bashScript download.sh --urlFile 2.txt
```

which resulted in

```
$ cwl-runner --local download.cwl --bashScript download.sh --urlFile 2.txt
2018-01-13 16:18:58 cwltool INFO: /usr/bin/cwl-runner 1.0.20171211211613 1.0.20171211211613, arvados-python-client 0.1.20171211211613, cwltool 1.0.20170928192020
2018-01-13 16:18:58 cwltool INFO: Resolved 'download.cwl' to 'file:///data-nvme1n1/home/wvandewege/downloader/download.cwl'
2018-01-13 16:18:59 arvados.arv-run INFO: Upload local files: "download.sh" "2.txt"
2018-01-13 16:18:59 arvados.arv-run INFO: Uploaded to 5elebe288e1daccf5744d1849610d292+71 (
dhhck-4zz18-0aog130ddkrk3yk)
2018-01-13 16:18:59 cwltool INFO: [workflow download.cwl] start
2018-01-13 16:18:59 cwltool INFO: [step readUrlList] start
2018-01-13 16:18:59 cwltool INFO: [step readUrlList] completed success
2018-01-13 16:18:59 cwltool INFO: [step downloadUrl] start
2018-01-13 16:18:59 arvados.cwl-runner INFO: [container downloadUrl] dhhck-xvhdp-9299d8y8q16fbfu
state is Committed
2018-01-13 16:18:59 cwltool INFO: [step downloadUrl] start
2018-01-13 16:19:00 arvados.cwl-runner INFO: [container downloadUrl_2] dhhck-xvhdp-a2zmgofa2d818dh
state is Committed
2018-01-13 16:19:29 arvados.cwl-runner INFO: [container downloadUrl] dhhck-xvhdp-9299d8y8q16fbfu is Final
2018-01-13 16:19:44 arvados.cwl-runner INFO: [container downloadUrl_2] dhhck-xvhdp-a2zmgofa2d818dh is Final
2018-01-13 16:19:44 cwltool INFO: [step downloadUrl] completed success
2018-01-13 16:19:44 cwltool INFO: [workflow download.cwl] completed success
2018-01-13 16:19:44 arvados.cwl-runner INFO: Overall process status is success
```

```
2018-01-13 16:19:44 arvdos.cwl-runner INFO: Final output collection 660188f814755c40f7a719f2b94d6f19+59 "Output of download.cwl (2018-01-13T16:19:44.827Z)" (dhhck-4zz18-j72zsyb8qxt9u)
{
  "out1": null
}
2018-01-13 16:19:44 cwltool INFO: Final process status is success
```

The log collections for the containers **do not** have the node.json file.

When I ran it without --local, like so:

```
cwl-runner download.cwl --bashScript download.sh --urlFile 2.txt
```

```
2018-01-13 16:26:33 cwltool INFO: /usr/bin/cwl-runner 1.0.20171211211613 1.0.20171211211613, arvdos-python-client 0.1.20171211211613, cwltool 1.0.20170928192020
2018-01-13 16:26:33 cwltool INFO: Resolved 'download.cwl' to 'file:///data-nvme1n1/home/wvandewege/downloader/download.cwl'
2018-01-13 16:26:34 arvdos.arv-run INFO: Upload local files: "download.sh" "2.txt"
2018-01-13 16:26:34 arvdos.arv-run INFO: Uploaded to 5elebe288eldaccf5744d1849610d292+71 (dhhck-4zz18-hhc7pht771wfx0)
2018-01-13 16:26:34 arvdos.cwl-runner INFO: [container download.cwl] submitted container dhhck-xvhdp-9gzxm1rcb9s4kne
2018-01-13 16:27:49 arvdos.cwl-runner INFO: [container download.cwl] dhhck-xvhdp-9gzxm1rcb9s4kne is Final
2018-01-13 16:27:49 arvdos.cwl-runner INFO: Overall process status is success
2018-01-13 16:27:49 arvdos.cwl-runner INFO: Final output collection 660188f814755c40f7a719f2b94d6f19+59
{
  "out1": null
}
2018-01-13 16:27:49 cwltool INFO: Final process status is success
```

The resulting log collections do have the node.json file.

Why is this?

#### #13 - 01/15/2018 09:03 PM - Tom Clegg

Probably because [dhhck-xvhdp-9299d8y8q16fbfu](#) ran on compute3, which had

```
2018-01-13T16:19:10.117547746Z crunch-run 0.1.20171212165144.296aa66 started
```

while [dhhck-xvhdp-9gzxm1rcb9s4kne](#) ran on compute1, which had

```
2018-01-13T16:26:36.644719254Z crunch-run 0.1.20180111190404.453f922 started
```

#### #14 - 01/17/2018 04:17 PM - Tom Clegg

- Status changed from In Progress to Resolved