

Arvados - Feature #6518

[Crunch] [Crunch2] Dispatch containers via slurm

07/08/2015 07:02 PM - Tom Clegg

| | | | |
|---|-------------------|------------------------|-------------------|
| Status: | Resolved | Start date: | 07/08/2015 |
| Priority: | Normal | Due date: | |
| Assigned To: | Radhika Chippada | % Done: | 100% |
| Category: | Crunch | Estimated time: | 0.00 hour |
| Target version: | 2016-03-16 sprint | | |
| Description | | | |
| When containers appear in the queue, use SLURM to execute them on worker nodes. | | | |
| For now, the queue is arvados.v1.containers.queue (much like the Crunch1 job queue). | | | |
| From Crunch2 dispatch : | | | |
| slurm batch mode | | | |
| <ul style="list-style-type: none">• Use "sinfo" to determine whether it is possible to run the container• Submit a batch job to the queue: "echo crunch-run --job {uuid} sbatch -N1"• When container priority changes, use scontrol and scancel to propagate changes to slurm• Use strigger to run a cleanup script when a container exits | | | |
| The cleanup script just has to deal with cases like the node dying before crunch-run has a chance to update the container record to state="Complete" | | | |
| Subtasks: | | | |
| Task # 8474: Review 6518-crunch2-dispatch-slurm | | | Resolved |
| Task # 8522: Implement crunch-dispatch-slurm | | | Resolved |
| Task # 8608: Review tests branch: 6518-crunch2-dispatch-slurm-tests | | | Resolved |
| Task # 8607: Add tests | | | Resolved |
| Related issues: | | | |
| Related to Arvados - Story #6282: [Crunch] Write stories for implementation o... | | Resolved | 06/23/2015 |
| Related to Arvados - Feature #7816: [Crunch2] Execute minimal container spec ... | | Resolved | 11/17/2015 |
| Related to Arvados - Feature #8128: [Crunch2] API support for crunch-dispatch | | Resolved | 04/28/2016 |
| Blocked by Arvados - Story #6429: [API] [Crunch2] Implement "containers" and ... | | Resolved | 12/03/2015 |

Associated revisions

Revision e407a1d4 - 02/18/2016 09:28 PM - Peter Amstutz

Run slurmctld and slurmd inside arvbox. refs #6518

Revision 2669dd05 - 02/18/2016 09:28 PM - Peter Amstutz

Run slurmctld and slurmd inside arvbox. refs #6518

Revision e407a1d4 - 02/18/2016 09:28 PM - Peter Amstutz

Run slurmctld and slurmd inside arvbox. refs #6518

Revision 7bb66fca - 03/16/2016 01:48 AM - Peter Amstutz

Merge branch '6518-crunch2-dispatch-slurm' closes #6518

History

#1 - 07/08/2015 07:03 PM - Tom Clegg

- Tracker changed from Bug to Feature

#2 - 11/17/2015 03:08 PM - Brett Smith

- Target version set to Arvados Future Sprints

#3 - 12/02/2015 02:39 PM - Peter Amstutz

Suggest writing crunch 2 job dispatcher as a new set of actors in node manger.

This would enable us to solve the question of communication between the scheduler and cloud node management ([#6520](#)).

Node manager already has a lot of the framework we will want like concurrency (can have one actor per job) and a configuration system.

Different schedulers (slurm, sge, kubernetes) can be implemented as modules similarly to how different cloud providers are supported now.

#4 - 12/02/2015 04:36 PM - Peter Amstutz

More ideas:

Have a "dispatchers" table. Dispatcher processes are responsible for pinging the API server similar to how it is done for nodes to show they are alive.

A dispatcher claims a container by setting "dispatcher" field to it's UUID. This field can only be set once and that locks the record so that only the dispatcher can update it.

If a dispatcher stops pinging, the containers it has claimed should be marked as TempFail.

Dispatchers should be able to annotate containers (preferably through links) for example "I can't run this because I don't have any nodes with 40 GiB of RAM".

#5 - 12/12/2015 02:03 AM - Peter Amstutz

If we go with the architecture described in [#8001](#), that will be is a prerequisite.

#6 - 12/13/2015 01:58 AM - Peter Amstutz

- *Description updated*

#7 - 12/14/2015 02:56 PM - Peter Amstutz

[#7816](#) is now the story for actually running containers

#8 - 12/15/2015 03:41 PM - Brett Smith

- *Target version deleted (Arvados Future Sprints)*

- *Release set to 11*

#9 - 01/05/2016 03:21 PM - Brett Smith

- *Story points set to 3.0*

#10 - 02/17/2016 03:10 PM - Peter Amstutz

I think we can narrow this down to a 1 point story that just submits to "sbatch" and possibly checks "squeue" for status updates.

#11 - 02/17/2016 08:38 PM - Peter Amstutz

- *Story points changed from 3.0 to 1.0*

#12 - 02/17/2016 08:54 PM - Peter Amstutz

- *Target version set to 2016-03-02 sprint*

#13 - 02/17/2016 08:55 PM - Peter Amstutz

- *Assigned To set to Peter Amstutz*

#14 - 02/17/2016 09:00 PM - Tom Clegg

- *Description updated*

#15 - 02/17/2016 09:03 PM - Tom Clegg

- *Description updated*

#16 - 03/01/2016 06:47 PM - Radhika Chippada

Review feedback for branch 6518-crunch2-dispatch-slurm

crunch-dispatch-slurm.go

- Comment for runQueuedContainers says “Invoke dispatchLocal for each ticker cycle”. Please update to say “Invoke dispatchSlurm ...” instead
- It would greatly improve readability of code if camel case is used consistently for names such as submiterr, stdinerr, similar to updateErr etc. There are several variables that could be updated as such.
- func strigger: can you please rename it say what it does “setup trigger for when job finishes” ?
- comment for func run (line 225): pl update it to say submit batch command etc. Current comment is not quite correct (applicable to crunch-dispatch-local)
- can you please add comments to submit and strigger funcs
- It would be helpful if we add tests that mock slurm commands. I searched a little and found a few links that may come handy.
 - This link says "helperCommand" can be used to mock a command.
<http://stackoverflow.com/questions/24286683/how-to-test-system-commands-in-go>
 - https://golang.org/src/os/exec/exec_test.go
- This comment “#uuid=\$(squeue --jobs=\$jobid --states=all --format=%j --noheader)” in the shell script seems to be out of sync?

#17 - 03/01/2016 07:25 PM - Peter Amstutz

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

#18 - 03/02/2016 08:13 PM - Brett Smith

- Assigned To changed from *Peter Amstutz* to *Radhika Chippada*

- Target version changed from *2016-03-02 sprint* to *2016-03-16 sprint*

#19 - 03/11/2016 04:34 PM - Radhika Chippada

Added tests in branch 6518-crunch2-dispatch-slurm-tests, derived from 6518-crunch2-dispatch-slurm at [bf3a2814843a8f7a78592e3fb4c629fc9f4819b9](https://github.com/6518-crunch2-dispatch-slurm-tests/commit/bf3a2814843a8f7a78592e3fb4c629fc9f4819b9)

#20 - 03/16/2016 01:55 AM - Peter Amstutz

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

- % Done changed from *50* to *100*

Applied in changeset arvaados|commit:7bb66fca9371232cc32dd6b365ceb33e926eb0e7.