Arvados - Story #15759

[arvados-dispatch-cloud] deploy/run correct version of crunch-run binary on worker nodes

10/24/2019 06:14 PM - Tom Clegg

Status:	Resolved		**Start date:**	12/30/2019
Priority:	Normal		**Due date:**	
Assigned To:	Tom Clegg		**% Done:**	100%
Category:			**Estimated time:**	0.00 hour
Target version:	2020-01-02 Sprint			

Description

arvados-dispatch-cloud should automatically deploy a suitable crunch-run binary to each worker node, instead of expecting someone else to install it as part of the worker's OS image or boot script.

Currently, arvados-dispatch-cloud assumes the configured worker image includes a compatible version of crunch-run. This means the sysadmin typically builds/updates a custom worker image and updates the cluster configuration each time arvados-dispatch-cloud is installed/upgraded. Even if this is done correctly, results may be unpredictable when worker nodes are still alive and running the old image after an upgrade.

To avoid version mismatches and (in some cases) eliminate the need for custom worker images entirely, arvados-dispatch-cloud should

- have the ability to run as "crunch-run" (refactor crunch-run as a library so arvados-server can import it)
- load its own executable (perhaps via /proc/self/exe)
- copy itself to each worker node as part of the booting/readiness process
- use the copied version instead of relying on the worker to have a matching version

The "instance set ID" already ensures that a given worker is only accessed by a single dispatch process, so it shouldn't be necessary to accommodate races between dispatchers. However, for some extra insurance, crunch-run should accept an "expected version" hash on the command line, and error out if that doesn't match the hash of its own executable.

Subtasks:

Task # 15950: Review 15759-deploy-crunch-run	**Resolved**

Related issues:

Related to Arvados - Story #13908: [Epic] Replace SLURM for cloud job schedul...	**Resolved**	
Related to Arvados - Bug #15734: [a-d-c] needs to populate node.json in the c...	**Resolved**	10/22/2019

Associated revisions

Revision 5bc5b8c1 - 12/31/2019 04:43 PM - Tom Clegg

Merge branch '15759-deploy-crunch-run'

closes #15759

Arvados-DCO-1.1-Signed-off-by: Tom Clegg <tom@tomclegg.ca>

History

#1 - 10/24/2019 06:14 PM - Tom Clegg

- Related to Story #13908: [Epic] Replace SLURM for cloud job scheduling/dispatching added

#2 - 10/24/2019 06:15 PM - Tom Clegg

- Related to Bug #15734: [a-d-c] needs to populate node.json in the container log collection added

#3 - 10/29/2019 01:52 PM - Tom Clegg

- Target version changed from Arvados Future Sprints to To Be Groomed

#5 - 11/13/2019 03:45 PM - Tom Morris

- Target version changed from To Be Groomed to Arvados Future Sprints

- Story points set to 3.0

#6 - 12/11/2019 04:06 PM - Tom Clegg

- Target version changed from Arvados Future Sprints to 2020-01-15 Sprint

- Assigned To set to Tom Clegg

#7 - 12/17/2019 04:39 PM - Tom Clegg

- Status changed from New to In Progress

#8 - 12/30/2019 05:05 PM - Tom Clegg

dd9367afefff5d0cd38d1549e32e2794e4614fb4-dev on su92l:

```
Started arvados-dispatch-cloud.
{"N":0,"PID":85591,"level":"info","msg":"loaded initial instance list","time":"2019-12-30T16:25:51.919157692Z"
}
{"PID":85591,"level":"info","msg":"FixStaleLocks finished (218.951827ms), starting scheduling.","time":"
2019-12-30T16:25:51.919257195Z"}
{"ContainerUUID":"su92l-dz642-c770f80we1flli6","InstanceType":"Standard_DS1_v2","PID":85591,"Priority":
1124322183683972,"State":"Queued","level":"info","msg":"adding container to queue","time":"
2019-12-30T16:25:58.908458085Z"}
{"ContainerUUID":"su92l-dz642-c770f80we1flli6","InstanceType":"Standard_DS1_v2","PID":85591,"level":"info","
msg":"creating new instance","time":"2019-12-30T16:25:59.009299767Z"}
{"Address":"10.28.64.17","IdleBehavior":"run","Instance":"
/subscriptions/3fa048dc-aa38-4820-85ba-68498da5f26b/resourceGroups/su92l/providers/Microsoft.Compute/virtualMa
chines/compute-f51710e302afe4aef4a97c634a7c2ed3-u0ng37y6mfppwlu","InstanceType":"Standard_DS1_v2","PID":85591,
"State":"booting","level":"info","msg":"instance appeared in cloud","time":"2019-12-30T16:26:52.060385944Z"}
{"Address":"10.28.64.17","Command":"/bin/ls /arvados-compute-node-boot.complete  \u003e/dev/null 2\u003e\u0026
1","Instance":"
/subscriptions/3fa048dc-aa38-4820-85ba-68498da5f26b/resourceGroups/su92l/providers/Microsoft.Compute/virtualMa
chines/compute-f51710e302afe4aef4a97c634a7c2ed3-u0ng37y6mfppwlu","InstanceType":"Standard_DS1_v2","PID":85591,
"level":"info","msg":"boot probe succeeded","stderr":"","stdout":"","time":"2019-12-30T16:27:04.145503815Z"}
{"Address":"10.28.64.17","Instance":"
/subscriptions/3fa048dc-aa38-4820-85ba-68498da5f26b/resourceGroups/su92l/providers/Microsoft.Compute/virtualMa
chines/compute-f51710e302afe4aef4a97c634a7c2ed3-u0ng37y6mfppwlu","InstanceType":"Standard_DS1_v2","PID":85591,
"cmd":"sudo sh -c 'set -e; dstdir=\"/var/lib/arvados/\"; dstfile=\"
/var/lib/arvados/crunch-run~70761fb034f6b8633803f649e6da8acc\"; mkdir -p \"$dstdir\"; touch \"$dstfile\"
; chmod 0755 \"$dstdir\" \"$dstfile\"; cat \u003e\"$dstfile\"'","hash":"70761fb034f6b8633803f649e6da8acc","
level":"info","msg":"installing runner binary on worker","path":"
/var/lib/arvados/crunch-run~70761fb034f6b8633803f649e6da8acc","time":"2019-12-30T16:27:04.149045430Z"}
{"Address":"10.28.64.17","Instance":"
/subscriptions/3fa048dc-aa38-4820-85ba-68498da5f26b/resourceGroups/su92l/providers/Microsoft.Compute/virtualMa
chines/compute-f51710e302afe4aef4a97c634a7c2ed3-u0ng37y6mfppwlu","InstanceType":"Standard_DS1_v2","PID":85591,
"ProbeStart":"2019-12-30T16:27:01.751685496Z","level":"info","msg":"instance booted; will try probeRunning","
time":"2019-12-30T16:27:04.610927564Z"}
{"Address":"10.28.64.17","Instance":"
/subscriptions/3fa048dc-aa38-4820-85ba-68498da5f26b/resourceGroups/su92l/providers/Microsoft.Compute/virtualMa
chines/compute-f51710e302afe4aef4a97c634a7c2ed3-u0ng37y6mfppwlu","InstanceType":"Standard_DS1_v2","PID":85591,
"ProbeStart":"2019-12-30T16:27:01.751685496Z","RunningContainers":0,"State":"idle","level":"info","msg":"
probes succeeded, instance is in service","time":"2019-12-30T16:27:04.632272459Z"}
{"Address":"10.28.64.17","ContainerUUID":"su92l-dz642-c770f80we1flli6","Instance":"
/subscriptions/3fa048dc-aa38-4820-85ba-68498da5f26b/resourceGroups/su92l/providers/Microsoft.Compute/virtualMa
chines/compute-f51710e302afe4aef4a97c634a7c2ed3-u0ng37y6mfppwlu","InstanceType":"Standard_DS1_v2","PID":85591,
"level":"info","msg":"crunch-run process started","time":"2019-12-30T16:27:05.634698188Z"}
{"Address":"10.28.64.17","ContainerUUID":"su92l-dz642-c770f80we1flli6","Instance":"
/subscriptions/3fa048dc-aa38-4820-85ba-68498da5f26b/resourceGroups/su92l/providers/Microsoft.Compute/virtualMa
chines/compute-f51710e302afe4aef4a97c634a7c2ed3-u0ng37y6mfppwlu","InstanceType":"Standard_DS1_v2","PID":85591,
"Reason":"state=Complete","level":"info","msg":"killing crunch-run process","time":"
2019-12-30T16:27:38.769700849Z"}
{"ContainerUUID":"su92l-dz642-c770f80we1flli6","PID":85591,"State":"Complete","level":"info","msg":"
dropping container from queue","time":"2019-12-30T16:27:39.746299838Z"}
{"Address":"10.28.64.17","ContainerUUID":"su92l-dz642-c770f80we1flli6","Instance":"
/subscriptions/3fa048dc-aa38-4820-85ba-68498da5f26b/resourceGroups/su92l/providers/Microsoft.Compute/virtualMa
chines/compute-f51710e302afe4aef4a97c634a7c2ed3-u0ng37y6mfppwlu","InstanceType":"Standard_DS1_v2","PID":85591,
"level":"info","msg":"crunch-run process ended","time":"2019-12-30T16:27:41.774051843Z"}
{"Address":"10.28.64.17","IdleBehavior":"run","IdleDuration":129.977703,"Instance":"
/subscriptions/3fa048dc-aa38-4820-85ba-68498da5f26b/resourceGroups/su92l/providers/Microsoft.Compute/virtualMa
chines/compute-f51710e302afe4aef4a97c634a7c2ed3-u0ng37y6mfppwlu","InstanceType":"Standard_DS1_v2","PID":85591,
"State":"idle","level":"info","msg":"shutdown worker","time":"2019-12-30T16:29:51.751756102Z"}
{"PID":85591,"level":"info","msg":"
Will delete compute-f51710e302afe4aef4a97c634a7c2ed3-u0ng37y6mfppwlu-nic because it is older than 20s","time":
"2019-12-30T16:30:52.596329941Z"}
{"Instance":"
/subscriptions/3fa048dc-aa38-4820-85ba-68498da5f26b/resourceGroups/su92l/providers/Microsoft.Compute/virtualMa
chines/compute-f51710e302afe4aef4a97c634a7c2ed3-u0ng37y6mfppwlu","PID":85591,"WorkerState":"shutdown","level":
```

"info","msg":"instance disappeared in cloud","time":"2019-12-30T16:30:52.673688560Z"}
{"PID":85591,"level":"info","msg":"Deleted NIC compute-f51710e302afe4aef4a97c634a7c2ed3-u0ng37y6mfppwlu-nic","
time":"2019-12-30T16:31:02.780823195Z"}
{"PID":85591,"level":"info","msg":"
Blob compute-f51710e302afe4aef4a97c634a7c2ed3-u0ng37y6mfppwlu-os.vhd is unlocked and not modified for 319.6311
56866 seconds, will delete","time":"2019-12-30T16:35:51.674151866Z"}
{"PID":85591,"level":"info","msg":"
Deleted blob compute-f51710e302afe4aef4a97c634a7c2ed3-u0ng37y6mfppwlu-os.vhd","time":"
2019-12-30T16:35:51.849445174Z"}

container log:

2019-12-30T16:27:06.292289754Z crunch-run dd9367afefff5d0cd38d1549e32e2794e4614fb4-dev (go1.13.4) started
2019-12-30T16:27:06.292985680Z Executing container 'su92l-dz642-c770f80we1flli6'
2019-12-30T16:27:06.293197858Z Executing on host 'compute-f51710e302afe4aef4a97c634a7c2ed3-u0ng37y6mfppwlu'
...

15759-deploy-crunch-run @ dd9367afefff5d0cd38d1549e32e2794e4614fb4 -- https://ci.arvados.org/view/Developer/job/developer-run-tests/1701/

#9 - 12/30/2019 06:18 PM - Tom Clegg

crunch-run should accept an "expected version" hash on the command line, and error out if that doesn't match the hash of its own executable.

As implemented, the dispatcher writes the binary to "/var/lib/arvados/crunch-run~${md5}". Given that, having crunch-run check its own md5sum seems superfluous, so I didn't bother adding that.

#10 - 12/31/2019 03:58 PM - Peter Amstutz

This needs a documentation update, should merge/rebase master since the new documentation is merged.

#11 - 12/31/2019 05:15 PM - Anonymous

- % Done changed from 0 to 100

- Status changed from In Progress to Resolved

Applied in changeset arvados|5bc5b8c150860a22d7a66b14aedddf30e270c7b6.

#12 - 12/31/2019 05:42 PM - Peter Amstutz

As discussed on gitter, don't want to complicate the "set up a compute node image" documentation.

LGTM.

#13 - 12/31/2019 05:48 PM - Peter Amstutz

- Target version changed from 2020-01-15 Sprint to 2020-01-02 Sprint

#14 - 01/21/2020 09:27 PM - Peter Amstutz

- Release set to 22