Arvados - Feature #18298

Feedback when container can't be scheduled on LSF

10/26/2021 06:05 PM - Peter Amstutz

Status:	Resolved		**Start date:**	11/18/2021
Priority:	Normal		**Due date:**	
Assigned To:	Tom Clegg		**% Done:**	100%
Category:			**Estimated time:**	0.00 hour
Target version:	2021-11-24 sprint			

Description

If a container requests excessive resources that cannot be fulfilled, it should provide visible feedback to the user via the Arvados logs, instead of remaining queued forever.

Subtasks:

Task # 18360: Review 18298-lsf-no-suitable-hosts	**Resolved**

Associated revisions

Revision bfdecdca - 11/19/2021 10:29 PM - Tom Clegg

Merge branch '18298-lsf-no-suitable-hosts'

refs #18298

Arvados-DCO-1.1-Signed-off-by: Tom Clegg <tom@curii.com>

Revision 8529a9c5 - 11/22/2021 06:45 PM - Tom Clegg

Merge branch '18298-lsf-no-suitable-hosts'

refs #18298

Arvados-DCO-1.1-Signed-off-by: Tom Clegg <tom@curii.com>

Revision a4890844 - 11/22/2021 06:58 PM - Tom Clegg

Merge branch '18298-lsf-no-suitable-hosts'

refs #18298

Arvados-DCO-1.1-Signed-off-by: Tom Clegg <tom@curii.com>

Revision 82691f82 - 11/22/2021 06:58 PM - Tom Clegg

Merge branch '18298-lsf-no-suitable-hosts'

refs #18298

Arvados-DCO-1.1-Signed-off-by: Tom Clegg <tom@curii.com>

History

#1 - 10/26/2021 06:05 PM - Peter Amstutz

- Status changed from New to In Progress

#2 - 10/26/2021 06:06 PM - Peter Amstutz

- Category set to Crunch

- Description updated

- Subject changed from Feedback when container can't be scheduled to Feedback when container can't be scheduled on LSF

#3 - 10/26/2021 06:06 PM - Peter Amstutz

- Status changed from In Progress to New

- Category changed from Crunch to 0

#4 - 10/26/2021 06:06 PM - Peter Amstutz

- Release set to 45

- Category deleted (0)

#5 - 10/27/2021 02:49 PM - Peter Amstutz

- Target version changed from 2021-11-10 sprint to 2021-11-24 sprint

#6 - 11/09/2021 08:25 PM - Peter Amstutz

- Tracker changed from Bug to Feature

#7 - 11/10/2021 04:22 PM - Peter Amstutz

- Assigned To set to Tom Clegg

#8 - 11/12/2021 04:05 PM - Tom Clegg

Hm.

Here is `bjobs -UF` ("unformatted") for a job that will run when another job finishes:

```
Job <22701>, Job Name <aaaaa-aaaaa-aaaaaaaaaab>, User <tom>, Project <default>, Status <PEND>, Queue <normal>,
 Command <sleep 120>
Fri Nov 12 15:54:26: Submitted from host <9tee4.arvadosapi.com>, CWD <$HOME>, Requested Resources <rusage[mem=
8000.00] span[hosts=1]>;
 PENDING REASONS:
 Job requirements for reserving resource (mem) not satisfied: 2 hosts;

 SCHEDULING PARAMETERS:
           r15s   r1m  r15m   ut      pg    io    ls    it    tmp    swp    mem
 loadSched   -     -     -     -       -     -     -     -     -      -      -
 loadStop    -     -     -     -       -     -     -     -     -      -      -

 RESOURCE REQUIREMENT DETAILS:
 Combined: select[type == local] order[r15s:pg] rusage[mem=8000.00] span[hosts=1]
 Effective: -
```

Here is a job that will never run because there are no nodes with this much memory:

```
Job <22735>, Job Name <aaaaa-aaaaa-aaaaaaaaaag>, User <tom>, Project <default>, Status <PEND>, Queue <normal>,
 Command <sleep 120>
Fri Nov 12 15:57:44: Submitted from host <9tee4.arvadosapi.com>, CWD <$HOME>, Requested Resources <rusage[mem=
8000000.00] span[hosts=1]>;
 PENDING REASONS:
 Job requirements for reserving resource (mem) not satisfied: 2 hosts;

 SCHEDULING PARAMETERS:
           r15s   r1m  r15m   ut      pg    io    ls    it    tmp    swp    mem
 loadSched   -     -     -     -       -     -     -     -     -      -      -
 loadStop    -     -     -     -       -     -     -     -     -      -      -

 RESOURCE REQUIREMENT DETAILS:
 Combined: select[type == local] order[r15s:pg] rusage[mem=8000000.00] span[hosts=1]
 Effective: -
```

#9 - 11/12/2021 04:14 PM - Tom Clegg

If we add -R 'select[mem>8000000M]' to the bsub arguments, bjobs -UF says "New job is waiting for scheduling" for a while, then:

```
Job <22930>, Job Name <aaaaa-aaaaa-aaaaaaaaaah>, User <tom>, Project <default>, Status <PEND>, Queue <normal>,
 Command <sleep 120>
Fri Nov 12 16:11:05: Submitted from host <9tee4.arvadosapi.com>, CWD <$HOME>, Requested Resources < select[mem
>8000000.00] rusage[mem=8000000.00] span[hosts=1]>;
 PENDING REASONS:
 There are no suitable hosts for the job;

 SCHEDULING PARAMETERS:
           r15s   r1m  r15m   ut      pg    io    ls    it    tmp    swp    mem
 loadSched   -     -     -     -       -     -     -     -     -      -      -
 loadStop    -     -     -     -       -     -     -     -     -      -      -

 RESOURCE REQUIREMENT DETAILS:
 Combined: select[(mem>8000000.00) && (type == local)] order[r15s:pg] rusage[mem=8000000.00] span[hosts=1]
```

```
Effective: -
```

#10 - 11/18/2021 04:30 PM - Tom Clegg

- Status changed from New to In Progress

#11 - 11/18/2021 09:50 PM - Tom Clegg

18298-lsf-no-suitable-hosts @ [f6e8d7c2cada1570bac3e98f0712ad8651b8d9fd](#) -- [https://ci.arvados.org/view/Developer/job/developer-run-tests/2810/](#)

If LSF reports the job status is PEND and the reason contains the magic string "There are no suitable hosts for the job", cancel the container and copy the reason text into runtime_status["errors"].

#12 - 11/19/2021 09:34 PM - Peter Amstutz

Tom Clegg wrote:

> 18298-lsf-no-suitable-hosts @ [f6e8d7c2cada1570bac3e98f0712ad8651b8d9fd](#) -- [https://ci.arvados.org/view/Developer/job/developer-run-tests/2810/](#)
>
> If LSF reports the job status is PEND and the reason contains the magic string "There are no suitable hosts for the job", cancel the container and copy the reason text into runtime_status["errors"].

- Getting json structured output out of bjobs is nice.
- It seems unfortunate that in the first case it tells you what resource can't be reserved (mem), but not in the second case.
- Were you able to test this branch on 9tee4 already or should we plan to do a quick manual test it after it is merged and auto-deployed to 9tee4?

#13 - 11/19/2021 10:28 PM - Tom Clegg

Peter Amstutz wrote:

> - It seems unfortunate that in the first case it tells you what resource can't be reserved (mem), but not in the second case.

Yeah, it seems a bit backwards.

> - Were you able to test this branch on 9tee4 already or should we plan to do a quick manual test it after it is merged and auto-deployed to 9tee4?

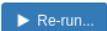I have not tested it with real LSF. Doing that after auto-deploy sounds worthwhile, yes.

#14 - 11/22/2021 04:59 PM - Tom Clegg

- File test18298-9tee4.png added

[9tee4-xvhdp-fjb1ctlvsbtn5dk](#) / [9tee4-dz642-u3hjqy0agtl3jya](#)

test 18298 - 2 ✎

(No description provided) ✎

▶ Re-run...   📦 Move container request...

| Status | Log | Provenance | Advanced |

This container is cancelled.   `Cancelled`

It has runtime of 0m 0s.

⚠ ▾ Error: There are no suitable hosts for the job;

| uuid: | 9tee4-xvhdp-fjb1ctlvsbtn5dk | priority: | 1 |

#15 - 11/22/2021 05:05 PM - Tom Clegg

Couple of caveats:

- Thanks to auto-retry after cancel, this has to happen 3 times before giving up on the container request
- If the admin has configured an explicit BsubArgumentsList based on an old config.default.yml file, the new arguments won't get passed to bsub and they will get the old "queued forever" behavior. Should we reconsider the previous (pre [#18290](#)) approach of appending the configured args to the defaults rather than replacing the defaults? Or just add an upgrade note?

#16 - 11/22/2021 06:04 PM - Peter Amstutz

Tom Clegg wrote:

> Couple of caveats:
>
> - Thanks to auto-retry after cancel, this has to happen 3 times before giving up on the container request

This seems inelegant but not actually a problem?

> - If the admin has configured an explicit BsubArgumentsList based on an old config.default.yml file, the new arguments won't get passed to bsub and they will get the old "queued forever" behavior. Should we reconsider the previous (pre #18290) approach of appending the configured args to the defaults rather than replacing the defaults? Or just add an upgrade note?

Upgrade note is fine. I don't know if anyone is actually using LSF yet. We can communicate about the upgrade to customers.

#17 - 11/22/2021 06:46 PM - Tom Clegg

- Status changed from In Progress to Resolved

> > - Thanks to auto-retry after cancel, this has to happen 3 times before giving up on the container request
>
> This seems inelegant but not actually a problem?

Right.

> Upgrade note is fine. I don't know if anyone is actually using LSF yet. We can communicate about the upgrade to customers.

Added upgrade note & merged.

Files

test18298-9tee4.png	29.6 KB	11/22/2021	Tom Clegg