

Arvados - Bug #17339

[keepstore] unexplained memory use with new S3 driver

02/04/2021 05:41 PM - Ward Vandewege

Status:	In Progress	Start date:	
Priority:	Normal	Due date:	
Assigned To:	Ward Vandewege	% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:	2021-10-27 sprint		

Description

I'm using the new S3 driver on 2xpu4. This is keepstore 2.1.1:

```
keep0.2xpu4:~# keepstore -version
keepstore 2.1.1 (go1.13.4)
```

The keepstore daemon on keep0.2xpu4 seems to use more memory than expected.

The machine it runs on is modest, with just 3.7G of ram:

```
# free

```

	total	used	free	shared	buff/cache	available
Mem:	3827036	3219100	317440	1000	290496	3255540
Swap:	0	0	0			

I used our little formula to calculate a suitable value for MaxKeepBlobBuffers:

```
$ perl -e 'print 3827036/88/1024'
42.4697709517045
```

That was too many; keepstore kept getting killed by the OOM killer. I reduced it to 40, same problem. I reduced it to 20 and that made it happy, stable at an RSS use of 2999284 bytes, and suitably making clients wait when its 20 buffers are in use (writing roughly 600 MiB/sec into S3).

After an hour of being mostly idle, the keepstore RSS memory use is unchanged:

```
root      3366 31.2 78.3 4218032 2999284 ?      Ssl  16:22  25:05 /usr/bin/keepstore
```

Two surprising things/questions here:

a) why is it using ~2.86GiB of ram with only 20 buffers? I was expecting memory use to be 1.28GiB of ram (20 * 64 MiB) plus a bit of overhead.

b) why is the memory use not going down after a long period of idle?

Here is a snapshot of the metrics for the keepstore (also available in prometheus/grafana as a time series):

```
# HELP arvados_concurrent_requests Number of requests in progress
# TYPE arvados_concurrent_requests gauge
arvados_concurrent_requests 0
# HELP arvados_keepstore_bufferpool_allocated_bytes Number of bytes allocated to buffers
# TYPE arvados_keepstore_bufferpool_allocated_bytes gauge
arvados_keepstore_bufferpool_allocated_bytes 5.905580032e+09
# HELP arvados_keepstore_bufferpool_inuse_buffers Number of buffers in use
# TYPE arvados_keepstore_bufferpool_inuse_buffers gauge
arvados_keepstore_bufferpool_inuse_buffers 0
# HELP arvados_keepstore_bufferpool_max_buffers Maximum number of buffers allowed
# TYPE arvados_keepstore_bufferpool_max_buffers gauge
arvados_keepstore_bufferpool_max_buffers 20
# HELP arvados_keepstore_pull_queue_inprogress_entries Number of pull requests in progress
# TYPE arvados_keepstore_pull_queue_inprogress_entries gauge
arvados_keepstore_pull_queue_inprogress_entries 0
```

```

# HELP arvados_keepstore_pull_queue_pending_entries Number of queued pull requests
# TYPE arvados_keepstore_pull_queue_pending_entries gauge
arvados_keepstore_pull_queue_pending_entries 0
# HELP arvados_keepstore_trash_queue_inprogress_entries Number of trash requests in progress
# TYPE arvados_keepstore_trash_queue_inprogress_entries gauge
arvados_keepstore_trash_queue_inprogress_entries 0
# HELP arvados_keepstore_trash_queue_pending_entries Number of queued trash requests
# TYPE arvados_keepstore_trash_queue_pending_entries gauge
arvados_keepstore_trash_queue_pending_entries 0
# HELP arvados_keepstore_volume_errors Number of volume errors
# TYPE arvados_keepstore_volume_errors counter
arvados_keepstore_volume_errors{device_id="s3:///2xpu4-nyw5e-0000000000000000-volume",error_type="*
s3manager.multiUploadError 000 MultipartUpload"} 2
arvados_keepstore_volume_errors{device_id="s3:///2xpu4-nyw5e-0000000000000000-volume",error_type="s
3.requestFailure 404 NotFound"} 1927
# HELP arvados_keepstore_volume_io_bytes Volume I/O traffic in bytes
# TYPE arvados_keepstore_volume_io_bytes counter
arvados_keepstore_volume_io_bytes{device_id="s3:///2xpu4-nyw5e-0000000000000000-volume",direction="
in"} 2.63069164e+08
arvados_keepstore_volume_io_bytes{device_id="s3:///2xpu4-nyw5e-0000000000000000-volume",direction="
out"} 1.25395672825e+11
# HELP arvados_keepstore_volume_operations Number of volume operations
# TYPE arvados_keepstore_volume_operations counter
arvados_keepstore_volume_operations{device_id="s3:///2xpu4-nyw5e-0000000000000000-volume",operation
="get"} 28
arvados_keepstore_volume_operations{device_id="s3:///2xpu4-nyw5e-0000000000000000-volume",operation
="head"} 5857
arvados_keepstore_volume_operations{device_id="s3:///2xpu4-nyw5e-0000000000000000-volume",operation
="list"} 112
arvados_keepstore_volume_operations{device_id="s3:///2xpu4-nyw5e-0000000000000000-volume",operation
="put"} 5821
# HELP arvados_max_concurrent_requests Maximum number of concurrent requests
# TYPE arvados_max_concurrent_requests gauge
arvados_max_concurrent_requests 0
# HELP request_duration_seconds Summary of request duration.
# TYPE request_duration_seconds summary
request_duration_seconds_sum{code="200",method="get"} 118.02206067300001
request_duration_seconds_count{code="200",method="get"} 42
request_duration_seconds_sum{code="200",method="put"} 29078.17948122801
request_duration_seconds_count{code="200",method="put"} 3906
request_duration_seconds_sum{code="500",method="put"} 19.722034916
request_duration_seconds_count{code="500",method="put"} 4
# HELP time_to_status_seconds Summary of request TTFB.
# TYPE time_to_status_seconds summary
time_to_status_seconds_sum{code="200",method="get"} 106.74335901900001
time_to_status_seconds_count{code="200",method="get"} 42
time_to_status_seconds_sum{code="200",method="put"} 29075.862950967014
time_to_status_seconds_count{code="200",method="put"} 3906
time_to_status_seconds_sum{code="500",method="put"} 19.72120516
time_to_status_seconds_count{code="500",method="put"} 4

```

Subtasks:

Task # 17411: Review

New

History

#1 - 02/04/2021 05:41 PM - Ward Vandewege

- Status changed from New to In Progress

#2 - 02/04/2021 05:49 PM - Ward Vandewege

- Description updated

#3 - 02/04/2021 05:50 PM - Ward Vandewege

- Description updated

#5 - 02/04/2021 05:51 PM - Ward Vandewege

- Subject changed from [keepstore] unexplained memory use to [keepstore] unexplained memory use with new S3 driver

#6 - 02/04/2021 06:21 PM - Ward Vandewege

- Description updated

#7 - 02/17/2021 02:05 PM - Ward Vandewege

- Target version changed from 2021-02-17 sprint to 2021-03-03 sprint

#8 - 03/03/2021 06:14 PM - Peter Amstutz

- Target version changed from 2021-03-03 sprint to 2021-03-17 sprint

#9 - 03/17/2021 03:58 PM - Ward Vandewege

- Target version changed from 2021-03-17 sprint to 2021-03-31 sprint

#10 - 03/31/2021 03:24 PM - Peter Amstutz

- Target version changed from 2021-03-31 sprint to 2021-04-14 sprint

#11 - 03/31/2021 03:31 PM - Peter Amstutz

- Target version changed from 2021-04-14 sprint to 2021-04-28 bughunt sprint

#12 - 04/28/2021 03:32 PM - Peter Amstutz

- Target version changed from 2021-04-28 bughunt sprint to 2021-05-12 sprint

#13 - 05/12/2021 03:54 PM - Peter Amstutz

- Target version changed from 2021-05-12 sprint to 2021-05-26 sprint

#14 - 05/26/2021 04:35 PM - Peter Amstutz

- Target version changed from 2021-05-26 sprint to 2021-06-09 sprint

#15 - 06/09/2021 03:50 PM - Peter Amstutz

- Target version changed from 2021-06-09 sprint to 2021-06-23 sprint

#16 - 06/19/2021 03:06 PM - Ward Vandewege

- Target version changed from 2021-06-23 sprint to 2021-07-07 sprint

#17 - 07/07/2021 03:36 PM - Peter Amstutz

- Target version changed from 2021-07-07 sprint to 2021-07-21 sprint

#18 - 07/21/2021 03:12 PM - Peter Amstutz

- Target version changed from 2021-07-21 sprint to 2021-08-04 sprint

#19 - 08/04/2021 03:28 PM - Peter Amstutz

- Target version changed from 2021-08-04 sprint to 2021-08-18 sprint

#20 - 08/18/2021 03:03 PM - Peter Amstutz

- Target version changed from 2021-08-18 sprint to 2021-09-01 sprint

#21 - 09/01/2021 03:15 PM - Peter Amstutz

- Target version changed from 2021-09-01 sprint to 2021-09-15 sprint

#22 - 09/15/2021 03:02 PM - Peter Amstutz

- Target version changed from 2021-09-15 sprint to 2021-09-29 sprint

#23 - 09/29/2021 03:22 PM - Peter Amstutz

- Target version changed from 2021-09-29 sprint to 2021-10-13 sprint

#24 - 10/13/2021 03:48 PM - Peter Amstutz

- Target version changed from 2021-10-13 sprint to 2021-10-27 sprint