

Arvados - Story #14645

[arv-mount] Report FUSE operation latency

12/19/2018 03:24 PM - Peter Amstutz

Status: Resolved	Start date: 03/14/2019
Priority: Normal	Due date:
Assigned To: Eric Biagiotti	% Done: 100%
Category:	Estimated time: 0.00 hour
Target version: 2019-03-27 Sprint	
Description Record request time for FUSE operations in arv-mount. Determine if prometheus_client Python library is useful to aggregate counters/averages. Report total time spent in FUSE operations and last interval in crunchstat format.	
Subtasks: Task # 14663: Review 14645-fuse-operations-reporting Resolved	
Related issues: Blocks Arvados - Story #14952: [crunchstat-summary] Add arv-mount FUSE ops time New	

Associated revisions

Revision 5b863886 - 03/13/2019 05:49 PM - Eric Biagiotti

Merge remote-tracking branch 'origin/master' into 14645-fuse-operations-reporting

refs #14645

Arvados-DCO-1.1-Signed-off-by: Eric Biagiotti <ebiagiotti@veritasgenetics.com>

Revision 215d15bc - 03/21/2019 06:03 PM - Eric Biagiotti

Merge remote-tracking branch 'origin/master' into 14645-fuse-operations-reporting

refs #14645

Arvados-DCO-1.1-Signed-off-by: Eric Biagiotti <ebiagiotti@veritasgenetics.com>

Revision 762b54b8 - 03/21/2019 07:52 PM - Eric Biagiotti

Merge branch '14645-fuse-operations-reporting'

refs #14645

Arvados-DCO-1.1-Signed-off-by: Eric Biagiotti <ebiagiotti@veritasgenetics.com>

History

#1 - 12/19/2018 03:24 PM - Peter Amstutz

- Status changed from New to In Progress

#2 - 12/19/2018 03:26 PM - Peter Amstutz

- Subject changed from [arv-mount] Report GET/POST latency to [arv-mount] Report GET/PUT latency

- Description updated

- Status changed from In Progress to New

#3 - 12/19/2018 04:44 PM - Peter Amstutz

- Description updated

#4 - 12/19/2018 07:10 PM - Peter Amstutz

- Subject changed from [arv-mount] Report GET/PUT latency to [arv-mount] Report FUSE operation latency

- Description updated

#5 - 12/19/2018 07:20 PM - Peter Amstutz

- Description updated

#6 - 12/19/2018 07:20 PM - Peter Amstutz

- Story points set to 1.0

#7 - 12/19/2018 07:20 PM - Peter Amstutz

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

#8 - 01/02/2019 04:52 PM - Tom Morris

- Target version changed from Arvados Future Sprints to 2019-01-16 Sprint

#9 - 01/02/2019 04:53 PM - Lucas Di Pentima

- Assigned To set to Lucas Di Pentima

#10 - 01/16/2019 04:12 PM - Lucas Di Pentima

- Target version changed from 2019-01-16 Sprint to 2019-01-30 Sprint

#11 - 01/16/2019 04:41 PM - Tom Morris

- Assigned To changed from Lucas Di Pentima to Eric Biagiotti

#12 - 01/30/2019 04:21 PM - Eric Biagiotti

- Target version changed from 2019-01-30 Sprint to 2019-02-13 Sprint

#13 - 01/30/2019 04:25 PM - Eric Biagiotti

- Story points changed from 1.0 to 3.0

#14 - 02/13/2019 01:27 PM - Eric Biagiotti

- Target version changed from 2019-02-13 Sprint to 2019-02-27 Sprint

#15 - 02/27/2019 01:57 PM - Eric Biagiotti

- Target version changed from 2019-02-27 Sprint to 2019-03-13 Sprint

#16 - 03/01/2019 07:33 PM - Tom Morris

- Release set to 15

#17 - 03/07/2019 06:32 PM - Eric Biagiotti

Is our goal to make use of prometheus' ability to serve stats, or just its aggregation functionality? Is the requirement to provide this data in crunchstat format to improve output or do we already aggregate data based on this output format?

#18 - 03/07/2019 06:43 PM - Peter Amstutz

Eric Biagiotti wrote:

Is our goal to make use of prometheus' ability to serve stats, or just its aggregation functionality?

Just the aggregation functionality, if it saves us any effort (someday we might want to export the metrics to a prometheus server, but that's not this story).

Is the requirement to provide this data in crunchstat format to improve output or do we already aggregate data based on this output format?

arv-mount already produces crunchstat format logs and the crunchstat-summary tool already knows how to consume crunchstat logs, so this store is just adding timing metrics.

#19 - 03/13/2019 01:05 PM - Eric Biagiotti

- Target version changed from 2019-03-13 Sprint to 2019-03-27 Sprint

#20 - 03/13/2019 01:11 PM - Eric Biagiotti

- Status changed from New to In Progress

#21 - 03/13/2019 04:03 PM - Eric Biagiotti

- Related to Story #14952: [crunchstat-summary] Add arv-mount FUSE ops time added

#22 - 03/13/2019 04:04 PM - Eric Biagiotti

- Related to deleted (Story #14952: [crunchstat-summary] Add arv-mount FUSE ops time)

#23 - 03/13/2019 04:04 PM - Eric Biagiotti

- Blocks Story #14952: [crunchstat-summary] Add arv-mount FUSE ops time added

#24 - 03/14/2019 03:08 PM - Eric Biagiotti

Ready for review at [97432e0e9d32a41c48a5298a0dbfd70b310e20e7](https://ci.curoverse.com/view/Developer/job/developer-run-tests/1125/)

- Added fuse operation timing using prometheus_client. Added a single Summary metric and individual labels for each operation. This should make querying for total fuse ops timing easier if we use prometheus to serve stats, while also making it possible to look at individual ops.
- Made arvmount crunchstat writing more generic. It can now accommodate different number of stats and types. (i.e. just a single float stat like ops time)

Potential improvements:

- Increase granularity. Wrap the specific llfuse calls, not the operator function?
- In __init__ update ops counters (write_ops_counter, read_ops_counter) to use prometheus instead of arvados.keep.Counter?

To Test:

- Set-up environment, ARVADOS_API_TOKEN, etc.
- Set-up a virtualenv
- \$ (cd arvados/services/fuse/ && python setup.py install)
- \$ mkdir test-dir
- \$ arv-mount --crunchstat-interval=1 --foreground --read-write --mount-tmp=tmp test-dir
- In another terminal, cd into the mounted directory and play around. You should see fuseopstime increasing.

Unit Tests: <https://ci.curoverse.com/view/Developer/job/developer-run-tests/1125/>

#25 - 03/14/2019 09:18 PM - Peter Amstutz

```
crunchstat: fuseopstime 117.273453 seconds -- interval 3.0000 seconds 0.000000 seconds
```

Format isn't quite right. The format should be:

```
crunchstat: <category> <value0> <name0> (<value1> <name1>...) --- interval N seconds <deltavalue0> <name0> (<deltavalue1> <name1>...) "
```

(I don't know why the name comes after the stat value)

I think we want to log all the individual stats. Some operations interact with the API server, some interact with keep servers, some are cached, some are not. So to diagnose performance problems we need to be able to distinguish where it is spending its time.

I noticed that a single very long operation won't be recorded until it completes. When graphed, the stats may be very spiky -- in other words an operation that takes 40 seconds won't show up as four 10-second intervals of 10 seconds of operation time each, it will show up as 40 seconds of operation time in one interval. I don't know if that can be helped.

#26 - 03/19/2019 08:35 PM - Eric Biagiotti

Latest at [0e0cbc43912e4bd8ad3fda616e552380d61c3f8d](https://ci.curoverse.com/view/Developer/job/developer-run-tests/1138/)

Unit tests: <https://ci.curoverse.com/view/Developer/job/developer-run-tests/1138/>

Peter Amstutz wrote:

[...]

Format isn't quite right. The format should be:

Fixed

I think we want to log all the individual stats. Some operations interact with the API server, some interact with keep servers, some are cached, some are not. So to diagnose performance problems we need to be able to distinguish where it is spending its time.

Done and added individual op counts also.

I noticed that a single very long operation won't be recorded until it completes. When graphed, the stats may be very spiky -- in other words an operation that takes 40 seconds won't show up as four 10-second intervals of 10 seconds of operation time each, it will show up as 40 seconds of operation time in one interval. I don't know if that can be helped.

I'll add a feature issue for this so we have it documented.

FYI, most of [daa4ee15fdf51a573f96b74c9fd01747b63bbeb0](#) was changed in the next commit. Keep that in mind if you are going commit by commit.

Also, here is an example of how the prometheus Summary metric is organized. It might be useful for following the logic.

```
Metric(arvmount_fuse_operations_seconds,
Time spent during FUSE operations, summary, , [
Sample(name='arvmount_fuse_operations_seconds_count', labels={'op': u'write'}, value=0.0, ...),
Sample(name='arvmount_fuse_operations_seconds_sum', labels={'op': u'write'}, value=0.0, ...),
Sample(name='arvmount_fuse_operations_seconds_created', labels={'op': u'write'}, value=0.0, ...),
Sample(name='arvmount_fuse_operations_seconds_count', labels={'op': u'read'}, value=0.0, ...),
Sample(name='arvmount_fuse_operations_seconds_sum', labels={'op': u'read'}, value=0.0, ...),
Sample(name='arvmount_fuse_operations_seconds_created', labels={'op': u'read'}, ...)]]\n"
```

#27 - 03/20/2019 07:54 PM - Peter Amstutz

The output format is not quite right, the 1st position is the resource being measured, followed by the resource's metrics. So instead of this:

```
crunchstat: fuseop:count 0 open -- interval 10.0000 seconds 0 open
crunchstat: fuseop:time 0.000000 open -- interval 10.0000 seconds 0.000000 open
```

It should be something like this:

```
crunchstat: fuseop:open 0 count 0.000000 time -- interval 10.0000 seconds 0 count 0.000000 time
```

#28 - 03/21/2019 03:14 PM - Eric Biagiotti

Peter Amstutz wrote:

The output format is not quite right, the 1st position is the resource being measured, followed by the resource's metrics. So instead of this:

[...]

It should be something like this:

[...]

Fixed in [fb8c51a195f2d9603cb129d8c3a21dc5d700aab3](#)

Unit tests: <https://ci.curoverse.com/view/Developer/job/developer-run-tests/1140/>

```
crunchstat: fuseop:flush 10 count 0.000100 time -- interval 1.0000 seconds 10 count 0.000100 time
crunchstat: fuseop:on_event 0 count 0.000000 time -- interval 1.0000 seconds 0 count 0.000000 time
crunchstat: fuseop:read 5 count 0.000292 time -- interval 1.0000 seconds 5 count 0.000292 time
crunchstat: fuseop:write 5 count 0.013155 time -- interval 1.0000 seconds 5 count 0.013155 time
...
```

I also manually added a few fuseop:flush lines to the crunchstat-summary test logs and ran the tests. They obviously break, but the new lines are being processed correctly. A few more decimal places would be nice, but that's for the follow up story [#14952](#).

```
--- 9,16 ----
cpu    sys      1.92    0.04    1.92
cpu    user     3.83    0.09    3.83
cpu    user+sys  5.75    0.13    5.75
+ fuseop:flush    count    10    10.00    10
+ fuseop:flush    time     0.00    0.00    0.00
fuseops  read     0     0     0
fuseops  write    0     0     0
keepcache hit     0     0     0
```

#29 - 03/21/2019 07:51 PM - Eric Biagiotti

Merged master into this branch to get some unit test fixes at [215d15bc03a38be1965a5d9df21417a3f7eae032](#)

Units tests: <https://ci.curoverse.com/view/Developer/job/developer-run-tests/1143/>

#30 - 03/21/2019 07:59 PM - Eric Biagiotti

- Status changed from In Progress to Resolved