Arvados - Feature #14714

[keep] keep-balance uses cluster config file

01/10/2019 06:47 PM - Peter Amstutz

Status:	Resolved	**Start date:**	09/26/2019
Priority:	Normal	**Due date:**	
Assigned To:	Eric Biagiotti	**% Done:**	100%
Category:		**Estimated time:**	0.00 hour
Target version:	2019-10-09 Sprint		

Description

Should include:

- use lib/service to do generic service things like set up logging and a Prometheus registry, find appropriate port to listen on, start http server (along the lines of source:services/health/main.go@13647-keepstore-config and source:lib/dispatchcloud/cmd.go@13647-keepstore-config)

Should not include:

- Load keep services list from config file instead of arvados/v1/keep_services endpoint (seems unsafe to do this until we can be assured the config is fully migrated; meanwhile, the keep_services endpoint is safe to use throughout the migration)
- Rendezvous by volume instead of by server (see #15641)

Subtasks:

Task # 15480: Review 14714-keep-balance-config	**Resolved**

Related issues:

Related to Arvados - Story #13648: [Epic] Use one cluster configuration file ...	**Resolved**	
Related to Arvados - Feature #9255: [keep] drain mode for a keepstore service	**New**	**05/23/2016**
Related to Arvados - Story #15641: [keep-balance] [SDKs] rendezvous by volume...	**New**	

Associated revisions

Revision 52b7b293 - 09/30/2019 04:31 PM - Eric Biagiotti

Merge remote-tracking branch 'origin/master' into 14714-keep-balance-config

refs #14714

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

Revision 30db189f - 10/03/2019 03:09 PM - Eric Biagiotti

Merge branch '14714-keep-balance-config'

refs #14714

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

Revision 80c1c0c5 - 10/03/2019 07:33 PM - Eric Biagiotti

Merge branch '14714-keep-balance-config'

refs #14714

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

History

#1 - 01/10/2019 06:47 PM - Peter Amstutz

- Related to Story #13648: [Epic] Use one cluster configuration file for all components added

#2 - 01/10/2019 06:47 PM - Peter Amstutz

- Tracker changed from Bug to Feature

#3 - 01/23/2019 05:54 PM - Lucas Di Pentima

Based on documentation at https://doc.arvados.org/install/install-keep-balance.html

As I believe there's only one instance of keep-balance per cluster, I think it would be appropriate to add its specific configs inside a NodeProfile instead of having a separate section as the dispatchers on #14713

```
Timers:
  KeepbalanceRunPeriod: 10m
NodeProfiles:
  keep:
    Keepbalance:
      Listen: :9005
      ManagementToken: xyzzy
      CollectionBatchSize: 100000
      CollectionBuffers: 1000
      KeepServiceTypes:
        1: disk
```

#4 - 02/06/2019 07:18 PM - Tom Morris

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

- Story points set to 1.0

#5 - 07/17/2019 02:54 PM - Tom Morris

- Target version changed from Arvados Future Sprints to 2019-07-31 Sprint

#6 - 07/17/2019 02:54 PM - Lucas Di Pentima

- Assigned To set to Lucas Di Pentima

#7 - 07/31/2019 02:52 PM - Lucas Di Pentima

- Target version changed from 2019-07-31 Sprint to 2019-08-14 Sprint

#8 - 07/31/2019 03:13 PM - Eric Biagiotti

- Assigned To changed from Lucas Di Pentima to Eric Biagiotti

#9 - 08/13/2019 08:17 PM - Eric Biagiotti

- Status changed from New to In Progress

#10 - 08/14/2019 01:28 PM - Eric Biagiotti

- Target version changed from 2019-08-14 Sprint to 2019-08-28 Sprint

#11 - 08/16/2019 01:50 PM - Eric Biagiotti

- Status changed from In Progress to New

#12 - 08/28/2019 02:28 PM - Eric Biagiotti

- Target version changed from 2019-08-28 Sprint to 2019-09-11 Sprint

#13 - 08/30/2019 06:57 PM - Eric Biagiotti

Some questions/comments about keep-balance flags and config options.

Config

- **KeepServiceTypes**: In the config wiki KeepServiceTypes is mapped to Volumes. I'm assuming this is meant to map to Volume.Driver types? This seems contingent on keepstore cluster config work.

- **CollectionBatchSize**/**CollectionBuffers**: These are both mapped to API.MaxItemsPerResponse on the wiki, but it seems like we would be removing potentially useful resource usage tweaking. Are we sure we want to simplify this? See keep-balance/usage.go for more info.

- **LostBlockFile**: Not on the wiki, but Collections.KeepBalanceLostBlockFile would be a good place unless we want to make a new KeepBalance section in the config.

Flags

I plan on keeping all the flag options since keep-balance can be run once instead of as a service.

- **KeepServiceList**: Right now this is only a command line option. Unless we think its worth specifying a set of keep services to balance in the config, this will stay the same.

- **commit-pulls/commit-trash**: These are mapped to Collections.BlobReplicateConcurrency/Collections.BlobTrashConcurrency respectively, but we might want to consider settings these to false by default if keep-balance is run with --once to avoid a one-time op accidentally committing changes. We could also require these to be set explicitly if --once is used.

Docs

There is lots of good info in keep-balance/usage.go. Was planning on putting most of it in the install doc, but maybe a user guide page is more appropriate?

#14 - 08/30/2019 07:00 PM - Eric Biagiotti

- Status changed from New to In Progress

#15 - 09/03/2019 03:45 PM - Tom Clegg

> - **KeepServiceTypes**: In the config wiki KeepServiceTypes is mapped to Volumes. I'm assuming this is meant to map to Volume.Driver types? This seems contingent on keepstore cluster config work.

KeepServiceTypes supports filtering by service_type in the keep_services table, typically "disk" or "proxy". KeepServiceList supports using a specified (cached/fake/customized) set of keep_services rows.

I think we still need to support KeepServiceTypes until everyone has migrated their keepstore configs. After that,

- keepstore/keepproxy server addresses will be listed separately in the Services section typical installs won't need to specify KeepServiceTypes.
- for debugging/special situations, the list of servers can be controlled by using an altered version of the cluster config file.

We should check with ops to see whether there's still a need for specifying a subset or alternate list of services. If so, it should probably be done with per-volume flags (enable pull/trash) rather than per-server.

> - **CollectionBatchSize/CollectionBuffers**: These are both mapped to API.MaxItemsPerResponse on the wiki, but it seems like we would be removing potentially useful resource usage tweaking. Are we sure we want to simplify this? See keep-balance/usage.go for more info.

CollectionBatchSize might not be especially useful (keep-balance uses much less memory than apiserver for a given page size anyway). CollectionBuffers is hard to use effectively (anything far enough from 0 to affect performance uses arbitrary amounts of memory, and performance impact is minimal anyway).

That said, yes, at least for now let's just move these to Collections.BalanceCollectionBatch and Collections.BalanceCollectionBuffers.

> - **LostBlockFile**: Not on the wiki, but Collections.KeepBalanceLostBlockFile would be a good place unless we want to make a new KeepBalance section in the config.

Added Collections.BlobMissingReport to wiki.

> I plan on keeping all the flag options since keep-balance can be run once instead of as a service.
>
> - **KeepServiceList**: Right now this is only a command line option. Unless we think its worth specifying a set of keep services to balance in the config, this will stay the same.

Currently KeepServiceList can also be given literally in the keep-balance config file though, right? (see above re keeping/dropping)

> - **commit-pulls/commit-trash**: These are mapped to Collections.BlobReplicateConcurrency/Collections.BlobTrashConcurrency respectively, but we might want to consider settings these to false by default if keep-balance is run with --once to avoid a one-time op accidentally committing changes. We could also require these to be set explicitly if --once is used.

If we were starting fresh I'd say a "-n" (dry run) flag would be good -- but we're not, and changing the default from "don't commit" to "commit" seems iffy. Perhaps some input from ops?

> There is lots of good info in keep-balance/usage.go. Was planning on putting most of it in the install doc, but maybe a user guide page is more appropriate?

Sure, it looks like it can be split between config.default.yml, install doc, and ... a new page on the admin guide?

#16 - 09/03/2019 04:51 PM - Nico César

Tom Clegg wrote:

> > - **KeepServiceTypes**: In the config wiki KeepServiceTypes is mapped to Volumes. I'm assuming this is meant to map to Volume.Driver types? This seems contingent on keepstore cluster config work.
>
> KeepServiceTypes supports filtering by service_type in the keep_services table, typically "disk" or "proxy". KeepServiceList supports using a specified (cached/fake/customized) set of keep_services rows.
>
> I think we still need to support KeepServiceTypes until everyone has migrated their keepstore configs. After that,
>
> - keepstore/keepproxy server addresses will be listed separately in the Services section typical installs won't need to specify KeepServiceTypes.
> - for debugging/special situations, the list of servers can be controlled by using an altered version of the cluster config file.
>
> We should check with ops to see whether there's still a need for specifying a subset or alternate list of services. If so, it should probably be done with per-volume flags (enable pull/trash) rather than per-server.

For most clusters the approach for wait-for-the-migration of keepstore configs is great.
Listing separate keepstore and keepproxy sounds good.

Per volume flags sounds great. Specially in scenarios that "we need to migrate a volume" like switching them to read only, or future expansions of draining volumes feature.

As an special case that some cluster have: sometimes, specially on prem, we have a keepstore service running on the compute nodes. How does global configuration affects this? I'm just pointing out a potential problem, maybe I'm over thinking.

> That said, yes, at least for now let's just move these to Collections.BalanceCollectionBatch and Collections.BalanceCollectionBuffers.

From the Ops perspective: In the future I think this configuration knobs should have a recommended value at run-time based on the data available (an also auto select that value if needed and reporting the value in the logs). Specially useful with clusters that we don't have access to the keepstore servers, but we know that they could use resources more efficiently

> > I plan on keeping all the flag options since keep-balance can be run once instead of as a service.
> >
> > - **KeepServiceList**: Right now this is only a command line option. Unless we think its worth specifying a set of keep services to balance in the config, this will stay the same.
>
> Currently KeepServiceList can also be given literally in the keep-balance config file though, right? (see above re keeping/dropping)
>
> > - **commit-pulls/commit-trash**: These are mapped to Collections.BlobReplicateConcurrency/Collections.BlobTrashConcurrency respectively, but we might want to consider settings these to false by default if keep-balance is run with --once to avoid a one-time op accidentally committing changes. We could also require these to be set explicitly if --once is used.
>
> If we were starting fresh I'd say a "-n" (dry run) flag would be good -- but we're not, and changing the default from "don't commit" to "commit" seems iffy. Perhaps some input from ops?

I don't mind personally having changes in keep-balance flags. As long as we document it well and we have a version that shows it as deprecated.

> > There is lots of good info in keep-balance/usage.go. Was planning on putting most of it in the install doc, but maybe a user guide page is more appropriate?
>
> Sure, it looks like it can be split between config.default.yml, install doc, and ... a new page on the admin guide?

I like a new page in the admin page. Also adding a section "before you begging... think about your storage layer" that explains why you should have several keepstore servers. Talk a little about throughput and N-to-M connections. The audience for this section should be sysadmins that they've been managing NFS servers and RAIDs or similar technology and replication in Arvados is a hard concept to grasp.

#17 - 09/03/2019 08:51 PM - Tom Clegg

- Related to Feature #9255: [keep] drain mode for a keepstore service added

#18 - 09/11/2019 03:03 PM - Tom Morris

- Target version changed from 2019-09-11 Sprint to 2019-09-25 Sprint

#19 - 09/16/2019 08:15 PM - Tom Clegg

- Description updated

#20 - 09/19/2019 02:20 PM - Tom Clegg

- Related to Story #15641: [keep-balance] [SDKs] rendezvous by volume UUID instead of server UUID added

#21 - 09/25/2019 03:05 PM - Eric Biagiotti

- Target version changed from 2019-09-25 Sprint to 2019-10-09 Sprint

#22 - 09/26/2019 02:37 PM - Eric Biagiotti

- Target version deleted (2019-10-09 Sprint)

Latest at ac74894758f1b56c449022810b45f833adcfefa7

- Now uses cluster config.
- Reorganizes the Server struct to use lib/service to do generic service things.
- Removes dumpconfig flag.
- Removes the options to specify a keep service list or type. Keep-balance will now balance all keep services of type disk reported by the keep_services endpoint.
- Debug flag removed. Uses SystemLogs.LogLevel from cluster config instead.

Didn't add keep-balance start-up in CheckHealth or ServeHttp (like dispatcher does). Unless I am missing something, it doesn't seem to make sense to start keep-balance that way. By the time command.RunCommand calls CheckHealth, it should have a setup keep-balance via newHandler.

Also, I haven't actually run keep-balance outside of its tests. Working on getting something working in arvbox.

#23 - 09/26/2019 07:53 PM - Eric Biagiotti

Latest at: 8a879fd6ffb38927150be85c63d926cd6a4c0d42

Rebased on master with the keepstore merge.

- Updated to use the prometheus registry passed in with newHandler
- Updated the test logging to use the ctxlog TestLogger.

Also kicked off tests:

https://ci.curoverse.com/view/Developer/job/developer-run-tests/1549/

#24 - 09/26/2019 07:56 PM - Eric Biagiotti

- Target version set to 2019-10-09 Sprint

#25 - 09/26/2019 09:03 PM - Tom Clegg

lib/config/config.default.yml: I think a better default for BalanceCollectionBatch might be 0 (effectively "same as MaxItemsPerResponse").

lib/config/config.default.yml: BlobMissingReport comment is due for a word wrap

lib/config/deprecated.go: "please update to the cluster config." → "please update the cluster config." Come to think of it, I think this needs a bit more detail. What kind of update would accomplish this ("balance specific keep services")? Would it be better to say something like "keep-balance operates on all configured volumes"?

lib/config/deprecated.go: "not longer supported" → "no longer supported"

lib/config/deprecated.go: if KeepServiceTypes is provided, rather than warn, I think it would be better to either ignore it silently (if it's the recommended value ["disk"], in which case we'd end up doing that anyway) or throw an error (if it's something else, in which case we'd end up doing something different than what was configured).

services/keep-balance/main.go: I think we should lose the "If using the legacy keep-balance.yml config, ..." bit. Maybe this whole error message could be "cannot start service: Collections.BalancePeriod is zero (if you want to run once and then exit, use the -once flag)"? Might help to emphasize normal usage first.

services/keep-balance/main.go: "debugf = log.Printf" bypasses the log formatting config -- how about replace the two uses of debugf() with bal.Logger.Debugf(), and remove debugf entirely?

other globals to get rid of:

- version -- we already have lib/cmd.version for this, except that it looks like we (I) didn't update the package scripts, or provide a way to log "starting X version Y" at startup (I can work on this)
- command -- seems like it could be deleted, and just call service.Command(...).RunCommand(...) at the end of runCommand()
- options -- if newHandler() becomes an inline func inside runCommand(), then options can be a local var in runCommand()

services/keep-balance/server.go: Start() is only called once from one place, and only has one LOC, so setupOnce and Start itself seem superfluous -- could newHandler just say "go srv.Run()"?

(tbc)

#26 - 09/30/2019 04:46 PM - Eric Biagiotti

Latest at 52b7b2934d5d74ee67ca13f8d1cc95f1379faddc
Dev tests running: https://ci.curoverse.com/view/Developer/job/developer-run-tests/1563/

Tom Clegg wrote:

> lib/config/config.default.yml: I think a better default for BalanceCollectionBatch might be 0 (effectively "same as MaxItemsPerResponse").
>
> lib/config/config.default.yml: BlobMissingReport comment is due for a word wrap
>
> lib/config/deprecated.go: "please update to the cluster config." → "please update the cluster config." Come to think of it, I think this needs a bit more detail. What kind of update would accomplish this ("balance specific keep services")? Would it be better to say something like "keep-balance operates on all configured volumes"?
>
> lib/config/deprecated.go: "not longer supported" → "no longer supported"
>
> lib/config/deprecated.go: if KeepServiceTypes is provided, rather than warn, I think it would be better to either ignore it silently (if it's the recommended value ["disk"], in which case we'd end up doing that anyway) or throw an error (if it's something else, in which case we'd end up doing something different than what was configured).
>
> services/keep-balance/main.go: I think we should lose the "If using the legacy keep-balance.yml config, ..." bit. Maybe this whole error message could be "cannot start service: Collections.BalancePeriod is zero (if you want to run once and then exit, use the -once flag)"? Might help to emphasize normal usage first.

All of the above fixed in c8bfd534fc6e33b0b37f9fed1ee6232159edb631

> services/keep-balance/main.go: "debugf = log.Printf" bypasses the log formatting config -- how about replace the two uses of debugf() with bal.Logger.Debugf(), and remove debugf entirely?
>
> other globals to get rid of:
>
> - version -- we already have lib/cmd.version for this, except that it looks like we (I) didn't update the package scripts, or provide a way to log "starting X version Y" at startup (I can work on this)
> - command -- seems like it could be deleted, and just call service.Command(...).RunCommand(...) at the end of runCommand()
> - options -- if newHandler() becomes an inline func inside runCommand(), then options can be a local var in runCommand()
>
> services/keep-balance/server.go: Start() is only called once from one place, and only has one LOC, so setupOnce and Start itself seem superfluous -- could newHandler just say "go srv.Run()"?
>
> (tbc)

I left in setupOnce in case I was missing something regarding CheckHealth; but yea, it doesn't need to be there. Fixed in 8db3a110976bbb9288214eeb3064d31cb12c1c82

I also updated docs and made a new admin keepbalance page. Seems like the keepstore install page "Notes on storage management" section has a good explanation of keep-balance. Maybe I should move that language into the keepbalance page and link to it from "Notes on storage management" instead of linking to the keepbalance install docs?

I also noticed that running with -once, keep-balance doesn't actually exit. Not sure how that would work with the new service lib, but I'll looking into it more. Let me know if you have some insight into this. Thanks!

#27 - 09/30/2019 07:59 PM - Tom Clegg

Eric Biagiotti wrote:

> I also noticed that running with -once, keep-balance doesn't actually exit. Not sure how that would work with the new service lib, but I'll looking into it more. Let me know if you have some insight into this. Thanks!

Maybe add an os.Exit() to (*Server)run():

```
         if err != nil {
                 srv.Logger.Error(err)
+                os.Exit(1)
+        } else {
+                os.Exit(0)
         }
```

#28 - 09/30/2019 08:33 PM - Tom Clegg

Now that keep-balance uses lib/service.Command(), it doesn't need to serve /metrics itself, so

- (*Server)setup() can go away (in fact it looks like it's already unused, except from tests -- I suppose the tests will need some other way to access the metrics registry, perhaps pass s.Metrics.reg to a helper based on https://github.com/prometheus/client_golang/blob/master/prometheus/registry.go#L554 that writes to a buffer instead of a file on disk)
- (*Server)httpHandler can be set to http.NotFoundHandler() when initializing in main.go since (currently) keep-balance doesn't have any APIs other than the ones handled by lib/service
- (*Server)httpHandler could be embedded as just "http.Handler" instead, and the ServeHTTP() forwarding func wouldn't be needed at all

#29 - 10/01/2019 03:00 PM - Tom Clegg

services/keep-balance/balance.go: could remove the now-overbuilt okTypes argument and "ok" map from DiscoverKeepServices() -- just test srv.ServiceType == "disk" in the loop.

#30 - 10/01/2019 08:30 PM - Eric Biagiotti

Latest at e86ee860d99036ff4ac61f6635b738f3a408e446

- Removes metrics serving
- Updates test gathering of metrics
- Simplifies balancer keep service discovery
- Moves keep-balance info from keepstore install docs to the new admin page
- Keepbalance now exits correctly when using -once

https://ci.curoverse.com/view/Developer/job/developer-run-tests/1574/

#31 - 10/03/2019 03:00 PM - Tom Clegg

Even the pared-down "Note on storage management" on the "install keepstore" guide seems like a non sequitur (why is this being mentioned now? what am I supposed to do about it?) ... otoh I'm not sure where it should be moved to, and this is already tangential to the issue at hand, so I'm inclined leave it here for now.

LGTM, thanks

(test failures are surely the ones already fixed in master)

#32 - 10/03/2019 03:13 PM - Eric Biagiotti

- Status changed from In Progress to Resolved

#33 - 10/03/2019 06:48 PM - Eric Biagiotti

- Status changed from Resolved to Feedback

Latest at 5ffded80f04bc1b38574b0a4eee54c6ceb9b9112

- Updated the error message for invalid legacy config options to point to the upgrade notes.
- Also updated the upgrade notes to be more clear about how keep-balance determines what to balance.

#34 - 10/03/2019 07:15 PM - Tom Clegg

> all volumes of service type disk

...should probably be "all keepstore servers with service_type disk" (that endpoint doesn't list volumes, only services)

otherwise LGTM, thanks

#35 - 10/03/2019 07:47 PM - Eric Biagiotti

- Status changed from Feedback to Resolved

#36 - 01/22/2020 02:46 PM - Peter Amstutz

- Release set to 22