Arvados - Feature #15000

[controller] publish safe config

03/20/2019 02:09 PM - Peter Amstutz

Status:	Resolved	**Start date:**	06/07/2019
Priority:	Normal	**Due date:**	
Assigned To:	Tom Clegg	**% Done:**	100%
Category:	API	**Estimated time:**	0.00 hour
Target version:	2019-06-19 Sprint		

Description

Controller should publish at a well-known API endpoint a subset of the configuration which is not sensitive, which can be accessed without authorization (similar to the discovery doc). It should have the same shape as the full configuration, but only include whitelisted keys/sections.

Sections that should be public:

- "ExternalURL" for "Services"
- Collections
    - DefaultReplication
    - DefaultTrashLifetime
    - CollectionVersioning
    - BlobSigningTTL
- Containers
    - SupportedDockerImageFormats
    - DefaultKeepCacheRAM
    - ~~MaxDispatchAttempts~~
    - MaxRetryAttempts
    - UsePreemptibleInstances
    - ~~Logging (all)~~
- RemoteClusters
- ~~Workbench~~

Subtasks:

Task # 15330: Review 15000-config-api	**Resolved**

Related issues:

Related to Arvados - Story #13648: [Epic] Use one cluster configuration file ...	**Resolved**	
Blocks Arvados - Story #14813: [Workbench2] Use cluster config	**Resolved**	**07/31/2019**

Associated revisions

Revision 5c1c5e34 - 06/14/2019 06:28 PM - Tom Clegg

Merge branch '15000-config-api'

refs #15000

Arvados-DCO-1.1-Signed-off-by: Tom Clegg <tclegg@veritasgenetics.com>

History

#1 - 03/20/2019 02:09 PM - Peter Amstutz

- Status changed from New to In Progress

#2 - 03/20/2019 02:09 PM - Peter Amstutz

- Status changed from In Progress to New

#3 - 03/20/2019 02:50 PM - Peter Amstutz

- Blocks Story #14813: [Workbench2] Use cluster config added

#4 - 03/20/2019 03:17 PM - Peter Amstutz

- Description updated

#5 - 03/20/2019 03:35 PM - Tom Morris

- Story points set to 2.0

#6 - 04/03/2019 02:07 PM - Tom Morris

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

#7 - 06/05/2019 03:13 PM - Ward Vandewege

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

#8 - 06/05/2019 03:13 PM - Peter Amstutz

- Blocks Feature #14812: [Workbench1] Load configuration from cluster config file added

#9 - 06/05/2019 03:43 PM - Tom Clegg

- Category set to API

- Assigned To set to Tom Clegg

- Target version changed from Arvados Future Sprints to 2019-06-19 Sprint

#10 - 06/07/2019 05:06 PM - Tom Clegg

15000-config-api @ 68211a7ead5f3cbe1a90b1b7769118a2b3543211 -- https://ci.curoverse.com/view/Developer/job/developer-run-tests/1292/

- added Workbench section to config.default.yml
- everything in config.default.yml is now represented in the Go config struct, too (and there's a test to prevent us from adding other stuff without updating the struct)
- controller exports the safe parts of the config struct at /arvados/v1/config
- every config has to be explicitly classified as safe/unsafe in export.go (there's a test that spits out some helpful Go code if anything is missing)

We still need to figure out how Workbench will get its secrets in #14812, since they can't be included in this unauthenticated endpoint. I don't think that needs to hold up this branch, though.

~~The gofmt test fails because I saved this with gofmt 1.12. I could use a useless-comment hack like we've done in other places, but I'm hoping Jenkins gets updated to 1.12 soon and makes that unnecessary.~~ Jenkins is on Go 1.12. Make sure you have Go 1.12 to pass gofmt tests.

#12 - 06/10/2019 02:15 PM - Ward Vandewege

- Release set to 22

#13 - 06/12/2019 08:40 PM - Peter Amstutz

It appears that a duration without a suffix is now a fatal error?

2019-06-12_20:38:29.42932 {"error":"transcoding config data: duration must be given as a string like \"600s\" or \"1h30m\"","level":"info","msg":"exiting","time":"2019-06-12T20:38:29.429230820Z"}

Unfortunately it didn't tell me what key has the problem.

#14 - 06/12/2019 08:46 PM - Tom Clegg

Peter Amstutz wrote:

> It appears that a duration without a suffix is now a fatal error?

s/now/still/; cluster config has always required units for durations.

#15 - 06/13/2019 01:34 PM - Tom Clegg

Updated error to include the bad value, since that's easy. Including the key/line/column from the original YAML would be ideal, but beyond scope here.

15000-config-api @ eaf26ebd285952ec6695270f0eda04ea7bf7e6c6 -- https://ci.curoverse.com/view/Developer/job/developer-run-tests/1303/

#16 - 06/13/2019 01:37 PM - Peter Amstutz

Tom Clegg wrote:

> Peter Amstutz wrote:
>
>> It appears that a duration without a suffix is now a fatal error?

> s/now/still/; cluster config has always required units for durations.

Got it. Then probably this configuration key wasn't previously read by the Go code.

#17 - 06/13/2019 02:42 PM - Tom Clegg

Peter Amstutz wrote:

> Got it. Then probably this configuration key wasn't previously read by the Go code.

Ah, yes.

(Aside -- perhaps add to #14812? -- now that the keys are all there, we should make RailsAPI call config-dump instead of loading the cluster config file directly, so it gets the same checks/munges as everything else.)

#18 - 06/13/2019 06:39 PM - Peter Amstutz

Noticed this in config.default.yml

```
        ImageID: ami-01234567890abcdef
```

I don't think it is a good idea to have bogus values in the defaults file.

Also the VM image id feels like an internal implementation detail that maybe shouldn't be exposed without a management token.

I don't think the public config should include Containers.CloudVMs, Containers.JobsAPI or Containers.SLURM.

Git.Repositories and Workbench.RepositoryCache both have filesystem-local paths that don't need to be public.

I see Containers.DispatchPrivateKey ummmmmm

I don't think the "Users" and "Mail" sections should be published.

PostgreSQL includes "ConnectionPool"

TLS.Certificate I'm not sure what is supposed to go there, but if it is a path to a file it shouldn't be published.

#19 - 06/13/2019 07:46 PM - Tom Clegg

Peter Amstutz wrote:

> I don't think it is a good idea to have bogus values in the defaults file.

Agreed. Changed to "".

Now that we've established this won't be used by system components at all, I've moved the threshold from "non-secret configs needed by components" to "configs needed by clients".

15000-config-api @ 3f45203069cda0b906ceeaf64d1ac8146a085895 -- https://ci.curoverse.com/view/Developer/job/developer-run-tests/1306/

#20 - 06/13/2019 07:49 PM - Tom Clegg

- Description updated

#21 - 06/13/2019 07:49 PM - Tom Clegg

- Blocks deleted (Feature #14812: [Workbench1] Load configuration from cluster config file)

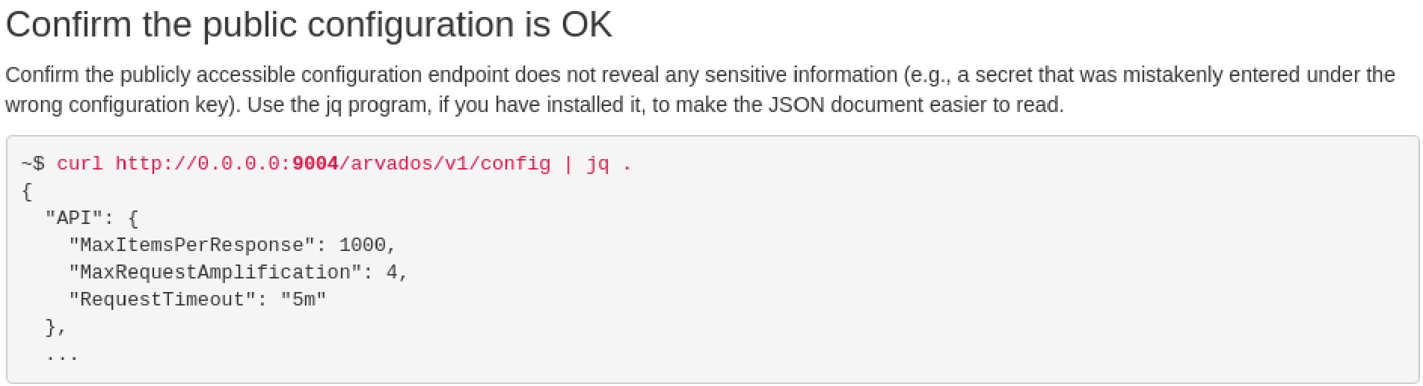
#22 - 06/14/2019 02:23 PM - Peter Amstutz

I noticed that some services have {ExternalURL: "-"} which I assume means it is an internal service that isn't supposed to have an external URL (as opposed to the services with {ExternalURL: ""} where the external URL is missing from the configuration...), how hard would it be to just filter those out? What if an admin fills in ExternalURL by accident?

#23 - 06/14/2019 06:26 PM - Tom Clegg

- File 15000-docs.png added

15000-config-api @ d768d4e5b5b61949aeb2bb2e473c4ceec93957f1 adds a step to the install docs to check for accidentally-published stuff:



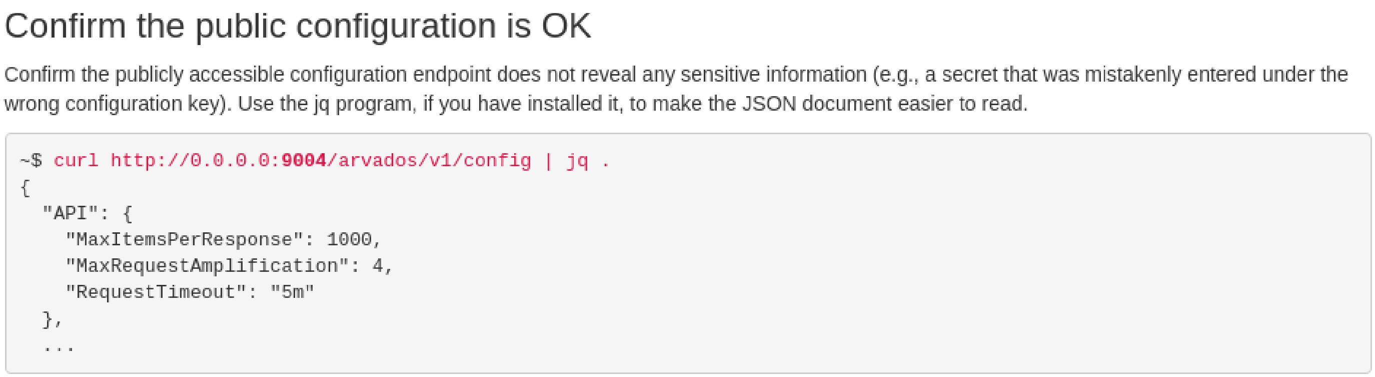
Confirm the public configuration is OK

Confirm the publicly accessible configuration endpoint does not reveal any sensitive information (e.g., a secret that was mistakenly entered under the wrong configuration key). Use the jq program, if you have installed it, to make the JSON document easier to read.

```
~$ curl http://0.0.0.0:9004/arvados/v1/config | jq .
{
  "API": {
    "MaxItemsPerResponse": 1000,
    "MaxRequestAmplification": 4,
    "RequestTimeout": "5m"
  },
  ...
```

#24 - 06/14/2019 06:28 PM - Peter Amstutz

Tom Clegg wrote:

> 15000-config-api @ d768d4e5b5b61949aeb2bb2e473c4ceec93957f1 adds a step to the install docs to check for accidentally-published stuff:
>
> 
>
> ## Confirm the public configuration is OK
>
> Confirm the publicly accessible configuration endpoint does not reveal any sensitive information (e.g., a secret that was mistakenly entered under the wrong configuration key). Use the jq program, if you have installed it, to make the JSON document easier to read.
>
> ```
> ~$ curl http://0.0.0.0:9004/arvados/v1/config | jq .
> {
>   "API": {
>     "MaxItemsPerResponse": 1000,
>     "MaxRequestAmplification": 4,
>     "RequestTimeout": "5m"
>   },
>   ...
> ```

Thanks, LGTM.

#25 - 06/19/2019 01:11 PM - Tom Clegg

- Status changed from New to Resolved

Files

15000-docs.png	37.5 KB	06/14/2019	Tom Clegg