Arvados - Story #13048

Refactor crunch2 logging

02/07/2018 08:04 PM - Tom Clegg

Status:	New		**Start date:**	
Priority:	Normal		**Due date:**	
Assigned To:			**% Done:**	0%
Category:			**Estimated time:**	0.00 hour
Target version:	Arvados Future Sprints			

Description

Functionally, source:services/crunch-run is doing a reasonable job. However, the way it's implemented makes it difficult to make some of the changes we want.

Relevant issues

- #10181 save logs to keep periodically while a container is running (not just after it exits & saves staged outputs)
- #13005 timestamps are sometimes wrong/confusing because of throttle behavior
- #13100source:services/crunch-run and source:sdk/go/crunchrunner should drop their custom manifest-writing code, now that we have generalized write support in #12483
- The implementation is more complicated / harder to follow than it should be, given the low complexity of the problem it's solving

Proposed improvements

- Refactor the various functional aspects (add timestamps, throttle, write to apiserver) into modular parts that communicate through simple interfaces like io.Writer.
- Use io.MultiWriter from stdlib, instead of custom routing built into the processing modules.
- Use (*arvados.Collection)FileSystem() to open/write log files (and staged outputs? → delete upload*.go)
- Drop the pretense of splitting long lines (apparently this isn't needed; MaxLogLine seems to have been disconnected 2 years ago in b719ef57055ba2fd06c7a1377cc0d47ee5df935e)

Related issues:

Related to Arvados - Feature #10181: Crunch job output logging improvement st...	**Resolved**	**02/16/2017**
Related to Arvados - Bug #13005: [Crunch2] All stdout gets the same timestamp...	**New**	
Related to Arvados - Bug #13100: [crunch-run] Replace custom manifest-writing...	**Resolved**	**03/15/2018**

Associated revisions

Revision 95be914a - 04/10/2018 01:21 PM - Tom Clegg

Merge branch '13100-crunch-run-output'

fixes #13100
fixes #11583
fixes #12606
refs #13048

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

History

#1 - 02/07/2018 08:04 PM - Tom Clegg

- Related to Feature #10181: Crunch job output logging improvement stories added

#2 - 02/07/2018 08:04 PM - Tom Clegg

- Related to Bug #13005: [Crunch2] All stdout gets the same timestamp and other logging problems added

#3 - 02/07/2018 08:59 PM - Tom Clegg

- Description updated

#4 - 02/07/2018 09:15 PM - Tom Morris

- Tracker changed from Bug to Story

- Target version set to Arvados Future Sprints

- Story points set to 2.0

#5 - 03/14/2018 07:57 PM - Tom Clegg

- Related to Bug #13100: [crunch-run] Replace custom manifest-writing code with collectionFS added

#6 - 12/12/2018 04:45 PM - Tom Clegg

- Description updated