

Arvados - Feature #4027

[Crunch] Accept SDK version as a runtime constraint. Install SDK into the docker container before running tasks.

09/29/2014 07:54 PM - Ward Vandewege

Status:	Resolved	Start date:	11/11/2014
Priority:	Normal	Due date:	
Assigned To:	Brett Smith	% Done:	100%
Category:	Crunch	Estimated time:	0.00 hour
Target version:	2014-12-10 sprint		
Description			
During jobs.create and jobs.update, runtime_constraints['arvados_sdk_version'] provided by the client should be resolved to a commit hash and stored in arvados_sdk_hash.			
Example:			
In jobs.create request:			
<ul style="list-style-type: none">• <pre>"arvados_sdk_version": "master"</pre>			
In jobs.show response:			
<ul style="list-style-type: none">• <pre>"runtime_constraints": { "arvados_sdk_version": "master"</pre>			
During the "install source" phase of crunch-job (i.e., at least once per container before any task runs there), crunch-job should do something like this:			
<ul style="list-style-type: none">• <pre>git reset --hard {arvados_sdk_hash} cd sdk ./install.sh</pre>			
The sdk/install.sh program should install Python, Ruby, and Perl client libraries to the system package directories where they can be imported by Crunch tasks.			
Subtasks:			
Task # 4532: Review 4027-sdk-constraint-reuse-wip			Resolved
Task # 4557: Review arvados-dev branch 4027-sdk-debian-packages-wip			Resolved
Task # 4629: Review 4027-api-sdk-requires-docker-wip			Resolved
Task # 4627: Make sure crunch user can clone arvados git repository			Resolved
Task # 4209: crunch-job installs the SDK at the `arvados_sdk_hash` version if present			Resolved
Task # 4626: API validation rejects jobs with arvados_sdk_version without docker_image			Resolved
Task # 4208: jobs.create and jobs.update resolve runtime_constraints['arvados_sdk_versi...			Resolved
Task # 4489: Review 4027-api-sdk-version-wip			Resolved
Task # 4667: Review 4027-crunch-sdk-install-wip			Resolved
Related issues:			
Related to Arvados - Story #3126: [API] Support use of anonymous git url (lik...		Resolved	04/09/2015
Copied to Arvados - Bug #4750: [Crunch] crunch-dispatch's Bundler environment...		New	12/09/2014

Associated revisions

Revision 809b122e - 11/13/2014 01:57 PM - Brett Smith

Merge branch '4027-api-sdk-version-wip'

Refs #4027. Closes #4489.

Revision adb6ea84 - 11/17/2014 09:47 PM - Brett Smith

Merge branch '4027-sdk-debian-packages-wip'

Refs #4027. Closes #4557.

Revision adb6ea84 - 11/17/2014 09:47 PM - Brett Smith

Merge branch '4027-sdk-debian-packages-wip'

Refs #4027. Closes #4557.

Revision 14b762d2 - 11/18/2014 02:21 PM - Brett Smith

Merge branch '4027-sdk-constraint-reuse-wip'

Refs #4027. Closes #4532.

Revision 1a2e18ce - 11/20/2014 04:25 PM - Brett Smith

Merge branch '4027-api-sdk-requires-docker-wip'

Refs #4027. Closes #4626, #4629.

Revision d980949a - 12/09/2014 03:14 PM - Brett Smith

Merge branch '4027-crunch-sdk-install-wip'

Closes #4027, #4667.

Revision 63f3b606 - 12/09/2014 03:21 PM - Brett Smith

4027: Update arvados-cli in API server bundle.

Refs #4027.

Revision 366ba84c - 12/11/2014 03:54 PM - Brett Smith

4027: Document arvados_sdk_version's virtualenv requirement.

Refs #4027.

History

#1 - 09/29/2014 07:56 PM - Ward Vandewege

- Subject changed from *Add sdk_version as a job parameter. Takes a git hash. Crunch should install that sdk version into the docker image before running the task.* to *[Crunch] Add sdk_version as a job parameter. Takes a git hash. Crunch should install that sdk version into the docker image before running the task.*

#2 - 09/29/2014 08:28 PM - Tom Clegg

- Subject changed from *[Crunch] Add sdk_version as a job parameter. Takes a git hash. Crunch should install that sdk version into the docker image before running the task.* to *[Crunch] Accept SDK version as a runtime constraint. Install SDK into the docker container before running tasks.*

- Description updated

#3 - 10/08/2014 05:50 PM - Ward Vandewege

- Target version changed from *Arvados Future Sprints* to *2014-10-29 sprint*

#4 - 10/08/2014 07:55 PM - Tim Pierce

- Category set to *Crunch*

- Assigned To set to *Tim Pierce*

#5 - 10/29/2014 04:27 PM - Tim Pierce

- Target version changed from *2014-10-29 sprint* to *Arvados Future Sprints*

#6 - 10/29/2014 05:44 PM - Ward Vandewege

- Target version changed from Arvados Future Sprints to 2014-11-19 sprint

#7 - 10/29/2014 07:19 PM - Brett Smith

- Assigned To changed from Tim Pierce to Brett Smith

Proposed implementation. After the normal Docker image install, use docker run to install the exact specified version in the downloaded image, then run the job in the generated image.

#8 - 11/11/2014 08:39 PM - Brett Smith

4027-api-sdk-version-wip is up for review. It implements the API server changes necessary to make it go. Once it's merged, I'll work on Crunch to finish the story.

One small discrepancy with the description: I named the new column `arvados_sdk_version` instead of `arvados_sdk_hash`. I didn't want to commit us for eternity to identifying SDK versions by commit hashes, so I made the name more generic. (It would be pretty great if we outlived git!)

#9 - 11/12/2014 06:37 PM - Tom Clegg

At [db8f547](#)...

The only complaint I can find is that the arvados repository fixture ("owned by All Users group") doesn't quite match what we suggest in the [install docs](#) ("owned by root, readable by All Users group"). (The latter sounds better to me.) Possible remedies:

- Make the fixture match the docs (less scope creep, probably better for now)
- Create it automatically at runtime instead, like `system_user` et al., and perhaps clone the repo automatically during install (maybe a better long term plan?)

Mostly out of curiosity: is the long form migration better? I thought "rails g migration" would normally produce a short form for this, like

- ```
AddArvadosSdkVersionToJobs < ActiveRecord::Migration
 def change
 add_column :jobs, :arvados_sdk_version, :string
 end
end
```

Nice fixes in `git_test_helper`. (Looks like someone didn't know about `tar -C`!)

**#10 - 11/12/2014 10:13 PM - Brett Smith**

Tom Clegg wrote:

At [db8f547](#)...

The only complaint I can find is that the arvados repository fixture ("owned by All Users group") doesn't quite match what we suggest in the [install docs](#) ("owned by root, readable by All Users group").

I looked at `4xphq`, where you own the arvados repo, and figured it didn't matter too much. :) But I'm happy to follow best practices. Updated to be owned by root and add a `can_read` permission.

Mostly out of curiosity: is the long form migration better?

Don't read too much into it. At one point, I had trouble rolling back the short form migration, but that was a little while ago and I can't rule out the possibility that I had another problem. At this point I'm admittedly copying the last migration I did and adapting it as necessary.

Now at [be3c438](#)

**#11 - 11/12/2014 11:21 PM - Tom Clegg**

Brett Smith wrote:

Don't read too much into it. At one point, I had trouble rolling back the short form migration, but that was a little while ago and I can't rule out the possibility that I had another problem. At this point I'm admittedly copying the last migration I did and adapting it as necessary.

Gotcha. In case it's helpful next time "bundle exec rails g migration AddArvadosSdkVersionToJobs arvados\_sdk\_version:string" should produce a reasonable migration, possibly quicker than copy&paste.

(Next we need "rails g fixture"...)

Now at [be3c438](#)

LGTM, thanks.

**#12 - 11/14/2014 02:31 PM - Peter Amstutz**

In order to support common workflow language tool descriptions, it will be essential to be able to take Docker images containing applications and configure them to run under Arvados. This means there is no assumption that mount points like /keep exist, or that there is a user named "crunch". These things need to be provisioned before the container can run crunch scripts.

**#13 - 11/17/2014 06:54 PM - Radhika Chippada**

Review feedback:

- I could not tell if such a test already exists; if not, does it make sense to add a test that fails on job->find\_arvados\_sdk\_version by passing an invalid / incorrect sdk version?
- Wondering if we could add (probably an integration) test that creates a job with a specific version, read it and verify version, update it to a different version, read back and verify? It appeared that the functional tests were doing one step per test (as they should be).
- Does it make sense to add a job fixture with a specific version and test being able to read it and update it? I am thinking this would be similar to reading a job that was previously created with a sdk version.
- Does it make sense to add a workbench integration test to verify the sdk version of this job fixture?
- Not sure if it would be possible / meaningful to add tests for the update for the method "find\_docker\_image\_locator"?
- This is really a small issue, but was wondering if we really need 255 bytes for the newly added job attribute (sdk version)?
- Everything else looked good. Thanks.

**#14 - 11/17/2014 08:31 PM - Tom Clegg**

4027-sdk-debian-packages-wip @ commit:ab8b92d

This variable could probably be renamed to PATCH\_VERSION:

-   
+GIT\_HASH=\$(version\_from\_git)

The rest LGTM, with the caveat that I haven't actually tried running it yet to convince myself the Perl package builds...

**#15 - 11/17/2014 09:50 PM - Brett Smith**

Tom Clegg wrote:

This variable could probably be renamed to PATCH\_VERSION:

You're proving why it should be renamed: it *used* to be PATCH\_VERSION, but now it's the entire package version. Renamed to PKG\_VERSION and pushed at [adb6ea8](#). Thanks.

**#16 - 11/17/2014 10:38 PM - Brett Smith**

Radhika,

Thanks for the feedback. I believe many of the tests you're requesting are covered by unit tests in test/unit/job\_test.rb. Details below.

Radhika Chippada wrote:

Review feedback:

- I could not tell if such a test already exists; if not, does it make sense to add a test that fails on job->find\_arvados\_sdk\_version by passing an invalid / incorrect sdk version?

This should be covered by "creating job with SDK version '#{search}'" when search == "\_\_nonexistent\_\_".

- Wondering if we could add (probably an integration) test that creates a job with a specific version, read it and verify version, update it to a different version, read back and verify? It appeared that the functional tests were doing one step per test (as they should be).

Great suggestion. Added as a unit test in [fc1a628](#).

- Does it make sense to add a job fixture with a specific version and test being able to read it and update it? I am thinking this would be the similar to reading a job that was previously created with a sdk version.

After implementing the test in the last suggestion, I'm not sure what additional Arvados code this would test. If I wrote the test right, I don't think there should be much difference between "create a job with an SDK constraint, update the constraint, and check the update" and "load a fixture with an SDK constraint, update the constraint, and check the update"—it would just check that Rails loads fixtures correctly.

- Does it make sense to add a workbench integration test to verify the sdk version of this job fixture?

No Workbench functionality has been added in this ticket yet, so I'm not sure what it would test. I think the only place it would appear is in the job's "API response," which I believe is already tested generally.

- Not sure if it would be possible / meaningful to add tests for the update for the method "find\_docker\_image\_locator"?

This should already be covered by the various Docker-related tests in job\_test.rb. This method has existed for a while now; this branch just refactored it, so it could share common code with find\_arvados\_sdk\_version.

- This is really small issue, but was wondering if we really need 255 bytes for the newly added job attribute (sdk version)?

It's a fair point, but that has already been pushed and deployed on 4xphq. Unfortunately I think trying to roll it back now would be more trouble than it's worth.

Ready for another look at [fc1a628](#). Thanks.

#### #17 - 11/18/2014 12:25 PM - Radhika Chippada

Brett, sorry for the review confusion. I did start with reviewing the correct branch but made a mistake in the end and ended up reviewing wrong branch. So, now I reviewed the sdk branch also. Just a few comments below, none major. You can address or not, and merge without waiting on me anymore. Thanks.

- In job\_controller -> load\_job\_specific\_filters -> @filters.select! do |(col, op, val)|  
Elsewhere in our code we used the variable name "attr" instead of "col". Wondering if you would like to name it attr?
- Regarding the comment "# Using the HEAD default is fine." in the rescue block. I am sure you made sure this works :). Just making sure in this case it would default to HEAD?
- I think "previous\_job\_run\_with\_arvados\_sdk\_version" fixture can simply be "job\_with\_arvados\_sdk\_version"
- In JobReuseControllerTest, I am wondering if the variable names filter\_h and filter\_a can be more descriptive. I am thinking these stand for filter\_hash and filter\_added or something?
- And you already added the extra test that I suggested. Thanks.

#### #18 - 11/18/2014 02:10 PM - Brett Smith

Radhika Chippada wrote:

Brett, sorry for the review confusion. I did start with reviewing the correct branch but made a mistake in the end and ended up reviewing wrong branch.

Radhika,

No worries about that at all. I know this ticket has been confusing with multiple branches (there will be four by the time this is all over), and I'm happy to have the extra pair of eyes for the first one. Thanks for following up.

- In job\_controller -> load\_job\_specific\_filters -> @filters.select! do |(col, op, val)|  
Elsewhere in our code we used the variable name "attr" instead of "col". Wondering if you would like to name it attr?

You're absolutely right. I have changed all of these variable names to match what we use in related loops like ApiClientAuthorizationsController: attr, operator, operand.

- Regarding the comment "# Using the HEAD default is fine." in the rescue block. I am sure you made sure this works :). Just making sure in this case it would default to HEAD?

Yes, this is a little Ruby magic, but the git\_filters hash has a default value that's set the first time any item is accessed, and that sets "max\_version" => "HEAD". Refer to the git\_filters = Hash.new ... line early in the method; that's where the default value is defined. I updated the comment to try to clarify what's going on here; let me know if that's helpful.

- I think "previous\_job\_run\_with\_arvados\_sdk\_version" fixture can simply be "job\_with\_arvados\_sdk\_version"

We use the "previous\_job\_run" pattern for all the jobs that are in the completed state, which are primarily tested by the job reuse tests. This fixture is in the same boat, so I think the consistency is helpful, even if it is a little long.

- In JobReuseControllerTest, I am wondering if the variable names filter\_h and filter\_a can be more descriptive. I am thinking these stand for filter\_hash and filter\_added or something?

I meant filter\_hash and filter\_array, but you're right this is a little too brief. Changed to filters\_hash and filters (since the array corresponds to what literally gets passed in to the API request), respectively.

Ready for another look at [bc47a4a](#). Thanks.

#### #19 - 11/19/2014 07:19 PM - Brett Smith

- Target version changed from 2014-11-19 sprint to Arvados Future Sprints

#### #20 - 11/19/2014 07:20 PM - Brett Smith

- Target version changed from Arvados Future Sprints to 2014-11-19 sprint

#### #21 - 11/19/2014 08:08 PM - Brett Smith

- Target version changed from 2014-11-19 sprint to 2014-12-10 sprint

#### #22 - 11/19/2014 08:08 PM - Tom Clegg

- Status changed from New to In Progress

- Target version changed from 2014-12-10 sprint to 2014-11-19 sprint

#### #23 - 11/19/2014 08:18 PM - Brett Smith

- Target version changed from 2014-11-19 sprint to 2014-12-10 sprint

#### #24 - 11/20/2014 04:14 PM - Radhika Chippada

Review feedback for branch :

- Just one question. Should this test name in "unit/job\_test.rb" say "without Docker image" instead of "with Docker image"?  
test "can't specify SDK version with Docker image" do
- Everything else looks good. Thanks.

#### #25 - 11/26/2014 07:51 PM - Peter Amstutz

Reviewing 4027-crunch-sdk-install-wip [0e9c824](#)

- Thank you for refactoring the increasingly monolithic start\_jobs method.
- Since you're already refactoring the git code in crunch-dispatch, it would be better to run git with popen so we don't have to shell escape everything.
- A style tweak would make this logic clearer on line 271 of crunch-dispatch.rb (I realize I probably wrote this code originally ;-)

```
src_repo = File.join(@repo_root, repo_name + '.git')
```

to

```
src_repo = File.join(@repo_root, "#{repo_name}.git")
```

- In crunch-job, prefer "source" to "." on line 1825 for readability
- I get this error:

[4n8aq-8i9sb-i4rk2oysvvc9mik](#)

```
27772 0 stderr crunchstat: Running [/usr/bin/docker.io run --rm=true --attach=stdout --attach=stderr --attach=stdin -i --user=crunch --cidfile=/tmp/crunch-job-1001/4n8aq-ot0gb-i4f91shujvlgsk2.cid --sig-proxy --dns 10.13.4.125 --dns 172.17.42.1 --volume=/tmp/crunch-job-1001/src:/tmp/crunch-job-1001/src:ro --volume=/tmp/crunch-job-1001/opt:/tmp/crunch-job-1001/opt:ro --volume=/tmp/crunch-job-1001/task/localhost.1.keep:/keep:ro --volume=/tmp --env=JOB_PARAMETER_COMMAND=ARRAY(0x32c2cd8) --env=CRUNCH_WORK=/tmp/crunch-job-1001/work --env=TASK_QSEQUENCE=0 --env=CRUNCH_SRC_COMMIT=b77fb69914fda0aedecddf2e62076cb404265647 --env=ARVADOS_API_TOKEN=3ehs221j7jcdw7igofn0vztkwhblir8lqd4wsomn6ds05d9iz3 --env=TASK_UUID=4n8aq-ot0gb-i4f91shujvlgsk2 --env=CRUNCH_INSTALL=/tmp/crunch-job-1001/opt --env=CRUNCH_JOB_UUID=4n8aq-8i9sb-i4rk2oysvvc9mik --env=CRUNCH
```

```
_NODE_SLOTS=4 --env=ARVADOS_API_HOST=peterel:3001 --env=CRUNCH_SRC=/tmp/crunch-job-1001/src --env=TASK_TMP
DIR=/tmp/crunch-job-task-work/localhost.1 --env=JOB_SCRIPT=run-command --env=CRUNCH_TMP=/tmp/crunch-job-10
01 --env=TASK_KEEPMOUNT=/keep --env=JOB_UUID=4n8aq-8i9sb-i4rk2oysvvc9mik --env=TASK_SLOT_NODE=localhost --
env=ARVADOS_API_HOST_INSECURE=true --env=JOB_WORK=/tmp/crunch-job-work --env=TASK_WORK=/tmp/crunch-job-tas
k-work/localhost.1 --env=CRUNCH_SRC_URL=/home/peter/work/_arvados_internal.git --env=TASK_SEQUENCE=0 --env
=TASK_SLOT_NUMBER=1 --env=CRUNCH_REFRESH_TRIGGER=/tmp/crunch_refresh_trigger --env=CRUNCH_JOB_BIN=/home/pe
ter/work/arvados/services/crunch/crunch-job --env=HOME=/tmp/crunch-job-task-work/localhost.1 777ef687a8811
f22fdd7c615be9356a92ce5f2150ff481bd368def31eaelbcl15 stdbuf --output=0 --error=0 perl - /tmp/crunch-job-100
1/src/crunch_scripts/run-command]
```

```
2014-11-26 14:42:38 -0500 -- pipeline_instance 4n8aq-dlhrv-cmildf07v92ddee
command 4n8aq-8i9sb-i4rk2oysvvc9mik {:failed=>0, :done=>0, :running=>1, :todo=>0}
4n8aq-8i9sb-i4rk2oysvvc9mik
27772 0 stderr crunchstat: reading stats from /sys/fs/cgroup/memory/memory.stat
4n8aq-8i9sb-i4rk2oysvvc9mik
27772 0 stderr crunchstat: mem 5393547264 cache 0 swap 8059 pgmajfault 4882997248 rss
4n8aq-8i9sb-i4rk2oysvvc9mik
27772 0 stderr crunchstat: reading stats from /sys/fs/cgroup/cpuacct/cpuacct.stat
4n8aq-8i9sb-i4rk2oysvvc9mik
27772 0 stderr crunchstat: reading stats from /sys/fs/cgroup/cpuset/cpuset.cpus
4n8aq-8i9sb-i4rk2oysvvc9mik 27772 0 stderr crunchstat: cpu 245540.6500 user 12745.2200 sys 4 cpus
4n8aq-8i9sb-i4rk2oysvvc9mik
27772 0 stderr crunchstat: reading stats from /sys/fs/cgroup/blkio/blkio.io_service_bytes
4n8aq-8i9sb-i4rk2oysvvc9mik 27772 0 stderr crunchstat: blkio:8:32 15360 write 79734784 read
4n8aq-8i9sb-i4rk2oysvvc9mik 27772 0 stderr crunchstat: blkio:8:16 270336 write 72303616 read
4n8aq-8i9sb-i4rk2oysvvc9mik 27772 0 stderr crunchstat: blkio:8:0 188539445248 write 4412491776 read
4n8aq-8i9sb-i4rk2oysvvc9mik
27772 0 stderr crunchstat: reading stats from /sys/fs/cgroup/cpuacct/cgroup.procs
4n8aq-8i9sb-i4rk2oysvvc9mik 27772 0 stderr crunchstat: net:docker0 1296 tx 2192 rx
4n8aq-8i9sb-i4rk2oysvvc9mik 27772 0 stderr crunchstat: net:eth1 996622231 tx 1353980473 rx
4n8aq-8i9sb-i4rk2oysvvc9mik
27772 0 stderr python: /lib/x86_64-linux-gnu/libc.so.6: version `GLIBC_2.15' not found (required by pytho
n)
4n8aq-8i9sb-i4rk2oysvvc9mik
27772 0 stderr python: /lib/x86_64-linux-gnu/libc.so.6: version `GLIBC_2.14' not found (required by pytho
n)
4n8aq-8i9sb-i4rk2oysvvc9mik 27772 0 child 27865 on localhost.1 exit 1 success=
4n8aq-8i9sb-i4rk2oysvvc9mik 27772 0 failure (#1, permanent) after 2 seconds
```

#### #26 - 12/01/2014 03:55 PM - Peter Amstutz

Per our discussion on IRC, the virtualenv is being set up based on the host's python rather than the docker container's python, which can fail in a variety of ways due to software mismatches.

Also, it appears that \$CRUNCH\_INSTALL is not being cleaned up in the "clean work dirs" step, which results in another error when it tries to set up a virtualenv where one already exists.

#### #27 - 12/01/2014 09:18 PM - Brett Smith

Peter Amstutz wrote:

- In crunch-job, prefer "source" to "." on line 1825 for readability

I agree completely about readability. Unfortunately, "." is better for portability. "source" is not POSIX and not implemented in Debian's dash.

I implemented all your other suggestions as of [48cec3a](#). The \$CRUNCH\_INSTALL issue is rendered moot by moving the install step to inside Docker.

#### #28 - 12/02/2014 07:35 PM - Peter Amstutz

Functionally the code looks good. However, I would like to see a little bit more logging in the \_\_DATA\_\_ bootstrap script in crunch-job. Right now it doesn't log anything about setting up the virtualenv unless there was some kind of error. I think at least one line saying "installing SDK version XYZ in virtualenv" would make it less mysterious.

Also, make sure you update the SDK version pin in services/api/Gemfile.

#### #29 - 12/04/2014 06:26 PM - Brett Smith

Peter Amstutz wrote:

Functionally the code looks good. However, I would like to see a little bit more logging in the \_\_DATA\_\_ bootstrap script in crunch-job. Right now it doesn't log anything about setting up the virtualenv unless there was some kind of error. I think at least one line saying "installing SDK

version XYZ in virtualenv" would make it less mysterious.

Added in [53a3b79](#), along with some other clean-ups to make output handling more consistent with previous versions. Because the virtualenv build happens inside Docker, I made it so the logging only happens during task 0, to prevent the same output from appearing over and over again in the logs.

Also, make sure you update the SDK version pin in services/api/Gemfile.

I know, but I believe I have to do this after I merge (but before I push). Jenkins will build gems based on the merge commit, and I've seen gem get grumpy when a Gemfile refers to gems that don't exist (which they won't if I try to refer to my intermediate commits).

**#30 - 12/05/2014 01:59 PM - Peter Amstutz**

[536f0f3](#)

- As discussed on IRC, it would be helpful if crunch-dispatch could log the version of the arvados-cli gem it is using to run crunch-job, for debugging and reproducibility.
- I got "Warning: Python SDK installed, but can't build virtualenv" which does not tell me what went wrong.

**#31 - 12/08/2014 03:48 PM - Brett Smith**

Peter Amstutz wrote:

- As discussed on IRC, it would be helpful if crunch-dispatch could log the version of the arvados-cli gem it is using to run crunch-job, for debugging and reproducibility.

Looking further, I don't think crunch-dispatch can do this, since it has no write access to the job log file. Added some basic diagnostics to crunch-job instead, reporting the full path (since crunch-dispatch accepts any crunch-job script defined in \$CRUNCH\_JOB\_BIN) and available Gem information. It ignores errors for the sake of local development runs.

- I got "Warning: Python SDK installed, but can't build virtualenv" which does not tell me what went wrong.

It means virtualenv can't be run inside the Docker image—probably because it isn't installed, but it could also be caused by a bad \$PATH, bad permissions, etc. Clarified the log message. Now at [188a9dd](#).

**#32 - 12/08/2014 09:58 PM - Peter Amstutz**

1. It appears that virtualenv gets installed in the "shell" and "compute" Docker images, but not arvados/jobs ? That explains why I can't test this successfully, I am using a default arvados/jobs image from a few weeks ago.
2. For the log statement bikeshed, I suggest "Warning: virtualenv not found in Docker container default \$PATH. Cannot install Python SDK."
3. For reporting the gem versions, I get:

```
/home/peter/work/arvados/sdk/cli/bin/crunch-job with arvados-cli Gem version(s) 0.1.20141202211726, 0.1.20141118163122, 0.1.20141112222031, 0.1.20141017191701, 0.1.20141014201516
```

Telling me the path it is using is very helpful, telling me the gems that are installed without knowing which one is actually active, less so - consider taking that out.

**#33 - 12/08/2014 11:23 PM - Brett Smith**

Peter Amstutz wrote:

1. It appears that virtualenv gets installed in the "shell" and "compute" Docker images, but not arvados/jobs ? That explains why I can't test this successfully, I am using a default arvados/jobs image from a few weeks ago.
2. For the log statement bikeshed, I suggest "Warning: virtualenv not found in Docker container default \$PATH. Cannot install Python SDK."

Both done as of [046dc01](#).

1. For reporting the gem versions, I get:

[...]

Telling me the path it is using is very helpful, telling me the gems that are installed without knowing which one is actually active, less so - consider taking that out.

Fixed by having crunch-dispatch invoke crunch-job with bundle exec, so they see the same Ruby environment, and only one version should normally be listed. This is probably better for consistency anyway, and won't leak through to Docker so we don't have to worry too much about it messing up those jobs. It does affect non-Docker jobs, but we haven't been guaranteeing much about the Ruby setup there either, so I don't think that's an issue.

**#34 - 12/09/2014 02:30 PM - Brett Smith**

Brett Smith wrote:

Fixed by having crunch-dispatch invoke crunch-job with bundle exec, so they see the same Ruby environment, and only one version should normally be listed. This is probably better for consistency anyway, and won't leak through to Docker so we don't have to worry too much about it messing up those jobs. It does affect non-Docker jobs, but we haven't been guaranteeing much about the Ruby setup there either, so I don't think that's an issue.

I have more proof that this is OK to do, but also found an unrelated bug here. Check out [qr1hi-8i9sb-uddf0fqrj36ayui](https://github.com/1hi-8i9sb-uddf0fqrj36ayui), which ran in production and just prints its environment and tries to list active gems. It clearly shows that Bundler is active in the current environment. I believe this is happening because in production, we have sudo configured to preserve the environment for Crunch. It looks like that probably isn't set up in your development environment, or in Docker. But we're already running with it, so making it explicit with bundle exec is relatively safe, and would just serve to bring development environments and production closer together.

The bug here is that this theoretically messes things up for jobs. The script tried to output the list of arvados-cli gems. That generated no output, because the Bundler environment lived all the way through to the compute nodes, where it no longer exists. Ruby is limiting itself to a nonexistent Gem set. The proper way to fix this would be to have crunch-job clean the environment for non-Docker jobs. I'm not going to tackle that in this branch, because I'm not sure it's important enough to do under crunch-job specifically rather than wait for v2, or whatever.

**#35 - 12/09/2014 03:08 PM - Peter Amstutz**

I get "sudo: bundle: command not found" when I try to run a job now. Maybe use "which" to get the right path? However, since we're at the end of the sprint, I'm also okay with backing out this feature entirely.

Everything else LGTM.

**#36 - 12/09/2014 03:20 PM - Brett Smith**

- *Status changed from In Progress to Resolved*

- *% Done changed from 84 to 100*

Applied in changeset arvados|commit:d980949ac4c092a44f3b64fb7cbd4a27a49256fb.