

Arvados - Feature #16462

Expand arvados-controller to expose forecast features

05/20/2020 04:04 PM - Nico César

Status: New	Start date:
Priority: Normal	Due date:
Assigned To: Nico César	% Done: 0%
Category: API	Estimated time: 0.00 hour
Target version: To Be Groomed	
Description Arvados-controller forecast endpoints should be created based on the work done in <code>git.arvados.org:arvados-forecaster.git</code>	
Subtasks: Task # 16546: Review New	
Related issues:	
Related to Arvados Epics - Story #16517: Workflow runtime/cost visibility and...	In Progress 05/01/2020 11/30/2021
Related to Arvados - Story #16534: [controller] facilitate database access by...	Resolved 06/26/2020
Related to Arvados - Feature #17054: Custom naming for scatter steps	In Progress
Blocked by Arvados - Feature #17014: Add Container Requests endpoint to contr...	Resolved 11/10/2020

Associated revisions

Revision 3e8a3101 - 06/08/2020 08:35 PM - Nico Cesar

added type ContainerRequestList

refs #16462

Arvados-DCO-1.1-Signed-off-by: <nico@nicocesar.com>

Revision 96e01b6f - 08/18/2020 05:31 PM - Nico Cesar

initial go mod tidy

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision ad05fa2c - 08/18/2020 08:19 PM - Nico Cesar

Boilerplate for all the controller and federation needed for forecast

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision 3d4c3231 - 09/14/2020 04:25 PM - Nico Cesar

replaced `arvados.ForecastDatapointsOptions` for `arvados.GetOptions`

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision 86d6ed00 - 09/15/2020 02:38 PM - Nico Cesar

added generated code for CR, EndPoints for `/a/v1/container_requests`

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision fadbf1d8 - 09/15/2020 02:49 PM - Nico Cesar

copy and paste leftover

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision 5a2c3c48 - 09/15/2020 09:16 PM - Nico Cesar

CR route were missing in controller/router/router.go

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision f716696e - 09/17/2020 01:32 PM - Nico Cesar

Trying out sqlx options and StructScan()

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision 549c1548 - 09/18/2020 03:23 PM - Nico Cesar

Added all the ContainerRequest api call to APIStub

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision 0875fa95 - 09/18/2020 04:17 PM - Nico Cesar

First working datapoint endpoints for CR

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision d8d31995 - 09/21/2020 03:52 PM - Nico Cesar

Typos

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision 5743f9d6 - 09/21/2020 06:53 PM - Nico Cesar

Merge branch 'master' into 16462-forecast-wip

Once more having go.sum in my branch has lots of issues.
adding the one in master for now. I'll make a better commit once this
is about to be merge to master

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision 9cce9619 - 09/22/2020 08:08 PM - Nico Cesar

test are not happy. maybe the transaction gets rolledback?

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision 7f216ff3 - 09/28/2020 08:27 PM - Nico Cesar

adding fixtures for container and container_requests

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision 0c302f31 - 09/29/2020 03:52 PM - Nico Cesar

fixtures now are in rails app

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision 5846dac4 - 09/29/2020 07:01 PM - Nico Cesar

circular dependencies, trying to get a real connection

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision 9893aad5 - 10/13/2020 05:22 PM - Nico Cesar

owner of the fixtures should be the "active user" of the test

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision 0c0e656b - 10/13/2020 05:52 PM - Nico Cesar

Added context to use fixture user's credentials

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision 790e0d29 - 10/13/2020 06:31 PM - Nico Cesar

Merge branch 'master' into 16462-forecast-wip-ldp

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision 0dccb3d7 - 10/14/2020 06:42 PM - Nico Cesar

Merge branch 'master' into 16462-forecast-wip

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision d81436f3 - 10/14/2020 07:21 PM - Nico Cesar

Good progress! first tests pass

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision 74a8f47e - 10/15/2020 03:21 PM - Nico Cesar

added *golden to .licenseignore

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision 71733fbe - 10/15/2020 03:21 PM - Nico Cesar

First batch of passing tests with golden files

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision f6ccc08c - 10/15/2020 08:14 PM - Nico Cesar

Added /arvados/v1/container_requests/UUID/datapoints endpoint

This is the first commit to try to implement Historical Forecasting
https://dev.arvados.org/projects/arvados/wiki/API_HistoricalForecasting_data_for_CR

This will enable to create Gantt diagrams that will help visualize the times and dependencies of a run.

refs #16462

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

Revision c0916b95 - 12/14/2020 09:40 PM - Nico Cesar

Added /a/v1/<cr_uuid>/datapoints endpoint to controller

This builds on top of the work we've done in 17014 this is the inclusion of the forecast datapoints and the relevant tests

refs #16462

refs #17014

Arvados-DCO-1.1-Signed-off-by: Nico Cesar <nico@curii.com>

History

#1 - 06/03/2020 03:44 PM - Peter Amstutz

- Target version changed from 2020-06-03 Sprint to 2020-06-17 Sprint

**

#2 - 06/03/2020 11:02 PM - Nico César

- File *timeline.html* added

- File *report.html* added

As a reference this is what nextflow has as a report

<https://www.nextflow.io/docs/latest/tracing.html>

#3 - 06/06/2020 01:46 PM - Nico César

- File *su92l-xvhdp-3gri0mi1vtakaf4.png* added

- File *su92l-xvhdp-3gri0mi1vtakaf4_just_circles.dot* added

- File *su92l-xvhdp-3gri0mi1vtakaf4_intermediate_graph.pdf* added

In order to determine dependencies I added a "<workflow>-start" and "<workflow>-end" in the intermediate graph you can see this in: *su92l-xvhdp-3gri0mi1vtakaf4_intermediate_graph.pdf*

Output.Source and Input.Source have been a source of trouble to connect the dots to make dependencies work.

#4 - 06/08/2020 03:21 PM - Nico César

Longest running haplotypcaller: 1d21h49m:

https://workbench.su92l.arvadosapi.com/container_requests/su92l-xvhdp-igq6soi7sz4runm (Notice that the command doesn't have the -L)

While most of the other ones are 2h to 4h range, as an example haplotypcaller_9:

https://workbench.su92l.arvadosapi.com/container_requests/su92l-xvhdp-cpoakbi81qgsbke

As we discussed before, we can't take just the name of the step to do the bookkeeping of the metrics (running time in this case, but more to come), Currently we use: "duration:<checkpoint>#<containerUUID>" as the key where checkpoint is "haplotypcaller" and anything after "#" is ignored on summarizing results.

Here the strategies are several:

1. get the parent container request that has a workflow.json to compare to "the family of workflows" and use that as part of the key when storing the
2. make some kind of signature of the command used for that Container, this will take into account the command line, but also get a little more creative and have input parsing and having some extra metrics comparing input sizes for example.
3. a mix of 1 and 2, starting with one and try to get a sense of the clustering happens for all data

#5 - 06/08/2020 08:40 PM - Nico César

added branch 16462-go-sdk-patch [3e8a3101cb9191813f3c8ed557d6f189d3e42063](https://github.com/Arvados/arvados/commit/3e8a3101cb9191813f3c8ed557d6f189d3e42063)

#6 - 06/08/2020 08:45 PM - Tom Clegg

[3e8a3101cb9191813f3c8ed557d6f189d3e42063](https://github.com/Arvados/arvados/commit/3e8a3101cb9191813f3c8ed557d6f189d3e42063) LGTM, thanks

#7 - 06/10/2020 04:24 PM - Nico César

- Related to Story #16517: Workflow runtime/cost visibility and forecasting added

#8 - 06/17/2020 03:31 PM - Nico César

- Target version changed from 2020-06-17 Sprint to 2020-07-01 Sprint

After talking to Lucas and Tom this is the WIP: https://dev.arvados.org/projects/arvados/wiki/API_HistoricalForecasting_data_for_CR

(for later review)

#9 - 06/17/2020 03:52 PM - Tom Clegg

- Related to Story #16534: [controller] facilitate database access by localdb API methods added

#10 - 07/01/2020 03:33 PM - Peter Amstutz

- Target version changed from 2020-07-01 Sprint to 2020-07-15

#11 - 07/15/2020 03:52 PM - Nico César

- Target version changed from 2020-07-15 to 2020-08-12 Sprint

#12 - 08/05/2020 03:22 PM - Nico César

- Status changed from New to In Progress

#13 - 08/12/2020 03:31 PM - Nico César

- Target version changed from 2020-08-12 Sprint to 2020-08-26 Sprint

#14 - 08/26/2020 03:33 PM - Nico César

- Target version changed from 2020-08-26 Sprint to 2020-09-09 Sprint

#15 - 09/09/2020 03:12 PM - Nico César

- Target version changed from 2020-09-09 Sprint to 2020-09-23 Sprint

#16 - 09/23/2020 03:55 PM - Nico César

- Target version changed from 2020-09-23 Sprint to 2020-10-07 Sprint

#17 - 09/23/2020 03:57 PM - Nico César

[9cce9619f3438011a42e5d10ca6c8584f0788381](https://dev.arvados.org/projects/arvados/wiki/API_HistoricalForecasting_data_for_CR) (branch 16462-forecast-wip) has the initial version of database tests

#18 - 09/28/2020 03:07 PM - Nico César

Note to future self/reviewer : make sure GET https://dev.arvados.org/v1/container_requests/ works as expected

#19 - 10/07/2020 02:11 AM - Peter Amstutz

- Release set to 25

#20 - 10/07/2020 01:56 PM - Peter Amstutz

- Release deleted (25)

#21 - 10/07/2020 01:59 PM - Peter Amstutz

- Target version changed from 2020-10-07 Sprint to 2020-10-21 Sprint

#22 - 10/15/2020 08:19 PM - Nico César

16462-forecast-wip2 ([f6ccc08c3f6b1ad42f2c827b19df0300f2c3c3db](https://dev.arvados.org/projects/arvados/wiki/API_HistoricalForecasting_data_for_CR)) can be reviewed now, This is unfinished work but might be useful to know if I'm heading in the right direction.

I see that some of my changes broke federation:

<https://ci.arvados.org/job/developer-run-tests-remainder/2227/console>

#23 - 10/21/2020 03:55 PM - Nico César

- Blocked by Feature #17014: Add Container Requests endpoint to controller added

#24 - 10/21/2020 03:55 PM - Nico César

- Target version changed from 2020-10-21 Sprint to 2020-11-18

#25 - 10/29/2020 02:33 PM - Nico César

- Related to Feature #17054: Custom naming for scatter steps added

#26 - 11/04/2020 05:05 PM - Nico César

- Target version changed from 2020-11-18 to 2020-12-02 Sprint

#27 - 12/02/2020 04:41 PM - Peter Amstutz

- Target version changed from 2020-12-02 Sprint to 2020-12-16 Sprint

#28 - 12/02/2020 04:41 PM - Peter Amstutz

- Target version changed from 2020-12-16 Sprint to 2021-01-06 Sprint

#29 - 12/14/2020 09:44 PM - Nico César

[c0916b956054b2e16ff61bd33b11bfe07e81787d](https://github.com/Arvados/c0916b956054b2e16ff61bd33b11bfe07e81787d) branch 16462-CR-datapoints

<https://ci.arvados.org/view/Developer/job/developer-run-tests/2227/>

#30 - 12/15/2020 06:13 PM - Peter Amstutz

- Target version changed from 2021-01-06 Sprint to 2021-01-20 Sprint

#31 - 01/20/2021 03:15 PM - Nico César

- Target version changed from 2021-01-20 Sprint to To Be Groomed

#32 - 03/31/2021 02:22 PM - Nico César

- Status changed from In Progress to New

Files

timeline.html	8.34 KB	06/03/2020	Nico César
report.html	2.81 MB	06/03/2020	Nico César
su92l-xvhdp-3gri0mi1vtakaf4_intermediate_graph.pdf	26.7 KB	06/06/2020	Nico César
su92l-xvhdp-3gri0mi1vtakaf4_just_circles.dot	8.1 KB	06/06/2020	Nico César
su92l-xvhdp-3gri0mi1vtakaf4.png	810 KB	06/06/2020	Nico César