

Arvados - Bug #15579

Staging a large number of files with "loadListing: no_listing" still takes more than 30 mins

08/21/2019 04:03 PM - Jiayong Li

Status:	New	Start date:	
Priority:	Normal	Due date:	
Assigned To:		% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:	To Be Groomed		

Description

I have a workflow (https://workbench.su92l.arvadosapi.com/container_requests/su92l-xvhdp-isnujmvprksam2q) that scatters the step "gvcf2fastj", each of them creates a directory structure as follows

```
sample_name
|
---- 0000.fj.gz
|
---- 0001.fj.gz
|
----
```

The gathering step stages all of those directories in a single directory. There are 228 samples, and each directory has ~1700 files. The gathering step takes more than 30 mins to finish, even if I used "loadListing: no_listing" as Peter suggested.

Supporting information: the gathering javascript `array-to-dir.cwl`

```
$namespaces:
  arv: "http://arvados.org/cwl#"
  cwltool: "http://commonwl.org/cwltool#"
class: ExpressionTool
cwlVersion: v1.0
hints:
  cwltool:LoadListingRequirement:
    loadListing: no_listing
inputs:
  arr:
    type:
      type: array
      items: [File, Directory]
  dirname:
    type: string
outputs:
  dir: Directory
requirements:
  InlineJavascriptRequirement: {}
expression: |
  ${
    var dir = {"class": "Directory",
              "basename": inputs.dirname,
              "listing": inputs.arr};
    return {"dir": dir};
  }
```

Log for failed run due to javascript timeout (my eval-timeout is turned to 2000).

```
2019-08-21T08:01:58.136588884Z cwltool WARNING: Failed to evaluate expression:
2019-08-21T08:01:58.136588884Z Expression evaluation error:
2019-08-21T08:01:58.136588884Z Long-running script killed after 2000.0 seconds: Javascript expression was: {
```

```
2019-08-21T08:01:58.136588884Z   var dir = {"class": "Directory",
2019-08-21T08:01:58.136588884Z           "basename": inputs.dirname,
2019-08-21T08:01:58.136588884Z           "listing": inputs.arr};
2019-08-21T08:01:58.136588884Z   return {"dir": dir};
2019-08-21T08:01:58.136588884Z }
2019-08-21T08:01:58.136588884Z stdout was: {"dir":{"class":"Directory","basename":"fjdir","listing":
": [{"basename":"A-UPN-UP000009-BL-UPN-3714","nameext":"","nameroot":"A-UPN-UP000009-BL-UPN-3714","
location":"keep:38b03b88c07b0fee01871a0e4f748829+51042/stage/A-UPN-UP000009-BL-UPN-3714","listing"
: [{"basename":"00ce.fj.gz","nameroot":"00ce.fj","nameext:".gz","location":"keep:38b03b88c07b0fee0
1871a0e4f748829+51042/stage/A-UPN-UP000009-BL-UPN-3714/00ce.fj.gz","class":"File","size":1237495},
{"basename":"001e.fj.gz","nameroot":"001e.fj","nameext:".gz","location":"keep:38b03b88c07b0fee018
71a0e4f748829+51042/stage/A-UPN-UP000009-BL-UPN-3714/001e.fj.gz","class":"File","size":1244790}, {"
basename":"00c9.fj.gz","nameroot":"00c9.fj","nameext:".gz","location":"keep:38b03b88c07b0fee01871
a0e4f748829+51042/stage/A-UPN-UP000009-BL-UPN-3714/00c9.fj.gz","class":"File","size":5199364},

...

{"basename":"01f8.fj.gz.gzi","nameroot":"01f8.fj.gz","nameext:".gzi","location":"keep:dbca4359019
69db7a1b2f38fb60ec240+51091/stage/A-CUHS-CU000305-BL-COL-60321BL1/01f8.fj.gz.gzi","class":"File","
size":7144}, {"basename":"0089.fj.gz.gzi","nameroot":"0089.fj.gz","nameext:".gzi","location":"keep
:dbca435901969db7a1b2f38fb60ec240+51091/stage/A-CUHS-CU000305-BL-COL-60321BL1/0089.fj.gz.gzi","cla
ss":"File","size":5576}, {"basename":"01f7.fj.gz","nameroot":"01f7.fj","nameext:".gz","location":"
keep:dbca435901969db7a1b2f38fb60ec240+51091/stage/A-CUHS-CU000305-BL-COL-
2019-08-21T08:01:58.375853555Z stderr was:
2019-08-21T08:02:15.664004152Z cwltool ERROR: [step handle-fjdirs] Output is missing expected fiel
d file:///var/lib/cwl/workflow.json#main/handle-fjdirs/dir
2019-08-21T08:02:58.086903477Z cwltool WARNING: [step handle-fjdirs] completed permanentFail
```

History

#1 - 09/05/2019 03:46 PM - Jiayong Li

Still having this problem after setting the overall workflow to have "no_listing"

https://workbench.su92l arvadosapi.com/container_requests/su92l-xvhdp-jfqj82mdufr0740

Tried debugging with Bryan, still have the same problem. The output listing is what's causing the problem.