

## Arvados - Bug #12606

### Symlink in output points to invalid location -- no such file or directory

11/17/2017 02:46 PM - Brad Chapman

<b>Status:</b>	Resolved	<b>Start date:</b>	11/17/2017
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assigned To:</b>	Peter Amstutz	<b>% Done:</b>	50%
<b>Category:</b>		<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>			

#### Description

Hi all;  
I'm running into an issue on runs with symlinked outputs, specifically for strelka2 variant calling. The run finishes cleanly and then fails when uploading results:

```
017-11-17T11:31:37.852090800Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-ploidy.vcf.gz (403 bytes) 2017-11-17T11:31:37.854757100Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-ploidy.vcf.gz.tbi (385 bytes) 2017-11-17T11:31:37.854791000Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-regions-merged.bed (285572 bytes) 2017-11-17T11:31:37.854875100Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-regions-merged.bed.gz (94913 bytes) 2017-11-17T11:31:37.854917900Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-regions-merged.bed.gz.tbi (13964 bytes) 2017-11-17T11:31:37.854941700Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-regions-merged.bed (285572 bytes) 2017-11-17T11:31:37.855083700Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/results/runStats.tsv (170 bytes) 2017-11-17T11:31:37.856529200Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/results/stats/runStats.xml (451 bytes) 2017-11-17T11:31:37.856914000Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/results/variants/genome.S1.vcf.gz (24603795 bytes) 2017-11-17T11:31:37.868911500Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/results/variants/genome.S1.vcf.gz.tbi (34744 bytes) 2017-11-17T11:31:37.868975600Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/results/variants/genome.vcf.gz (24603795 bytes) 2017-11-17T11:31:37.878724000Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/results/variants/genome.vcf.gz.tbi (34744 bytes) 2017-11-17T11:31:37.878796700Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/results/variants/variants.vcf.gz (4136633 bytes) 2017-11-17T11:31:37.880521500Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/results/variants/variants.vcf.gz.tbi (59036 bytes) 2017-11-17T11:31:37.880587900Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/runWorkflow.py (7795 bytes) 2017-11-17T11:31:37.882003000Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/runWorkflow.py.config.pickle (4785 bytes) 2017-11-17T11:31:37.882031100Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/workflow.error.log.txt (0 bytes) 2017-11-17T11:31:37.882054800Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/workflow.exitcode.txt (2 bytes) 2017-11-17T11:31:37.882076800Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/workflow.warning.log.txt (0 bytes) 2017-11-17T11:31:37.882133000Z While uploading output files: Symlink in output "/strelka2/chr19/NA24385-chr19_0_16089283-block.vcf.gz" points to invalid location "genome.S1.vcf.gz": lstat /tmp/603101599/strelka2/chr19/genome.S1.vcf.gz: no such file or directory 2017-11-17T11:31:37.882144900Z Cancelled
```

This is an example project and run that show the problem:

[https://cloud.curoverse.com/container\\_requests/qr1hi-xvhdp-j10b0yribnprohf](https://cloud.curoverse.com/container_requests/qr1hi-xvhdp-j10b0yribnprohf)  
[https://cloud.curoverse.com/container\\_requests/qr1hi-xvhdp-ewo8ck7owxbuyud#Log](https://cloud.curoverse.com/container_requests/qr1hi-xvhdp-ewo8ck7owxbuyud#Log)

It appears to move files, then fail when a symlink that points to those files is uploaded later. Here is what the directory structure for one of those strelka2 runs looks like:

```
├── Test1-chrM_0_1000-block-ploidy.vcf.gz
├── Test1-chrM_0_1000-block-ploidy.vcf.gz.tbi
├── Test1-chrM_0_1000-block-regions.bed
├── Test1-chrM_0_1000-block-regions-merged.bed
├── Test1-chrM_0_1000-block-regions-merged.bed.gz
├── Test1-chrM_0_1000-block-regions-merged.bed.gz.tbi
├── Test1-chrM_0_1000-block.vcf.gz -> Test1-chrM_0_1000-block-work/results/variants/genome.vcf.gz
├── Test1-chrM_0_1000-block.vcf.gz.tbi -> Test1-chrM_0_1000-block-work/results/variants/genome.vcf.gz.tbi
```

```

└─ Test1-chrM_0_1000-block-work
  └─ results
    ├── stats
    │   ├── runStats.tsv
    │   └─ runStats.xml
    └─ variants
        ├── genome.S1.vcf.gz
        ├── genome.S1.vcf.gz.tbi
        ├── genome.vcf.gz -> genome.S1.vcf.gz
        ├── genome.vcf.gz.tbi -> genome.S1.vcf.gz.tbi
        ├── variants.vcf.gz
        └─ variants.vcf.gz.tbi
  └─ runWorkflow.py
  └─ runWorkflow.py.config.pickle
  └─ workflow.error.log.txt
  └─ workflow.exitcode.txt
  └─ workflow.warning.log.txt

```

Thanks for any tips or clues on how to work around or fix.

**Subtasks:**

Task # 12637: Review	<b>Closed</b>
Task # 12609: Diagnose	<b>New</b>

**Related issues:**

Related to Arvados - Bug #12183: [crunch-run] Handle symlinks with absolute p...	<b>Resolved</b>	<b>09/28/2017</b>
Related to Arvados - Bug #13100: [crunch-run] Replace custom manifest-writing...	<b>Resolved</b>	<b>03/15/2018</b>

**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](mailto:tclegg@veritasgenetics.com)>

**History**

**#1 - 11/17/2017 02:48 PM - Brad Chapman**

Sorry the error message got formatted all on one line for some reason. Here it is in a more readable format:

```

2017-11-17T11:31:37.852090800Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-ploidy.vcf.gz (403 bytes)
2017-11-17T11:31:37.854757100Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-ploidy.vcf.gz.tbi (385 bytes)
2017-11-17T11:31:37.854791000Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-regions-merged.bed (285572 bytes)
2017-11-17T11:31:37.854875100Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-regions-merged.bed.gz (94913 bytes)
2017-11-17T11:31:37.854917900Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-regions-merged.bed.gz.tbi (13964 bytes)
2017-11-17T11:31:37.854941700Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-regions.bed (285572 bytes)
2017-11-17T11:31:37.855083700Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/results/stats/runStats.tsv (170 bytes)
2017-11-17T11:31:37.856529200Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/results/stats/runStats.xml (451 bytes)
2017-11-17T11:31:37.856914000Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/results/variants/genome.S1.vcf.gz (24603795 bytes)
2017-11-17T11:31:37.868911500Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/results/variants/genome.S1.vcf.gz.tbi (34744 bytes)
2017-11-17T11:31:37.868975600Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/results/variants/genome.vcf.gz (24603795 bytes)
2017-11-17T11:31:37.878724000Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/results/variants/genome.vcf.gz.tbi (34744 bytes)
2017-11-17T11:31:37.878796700Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/results/variants/v

```

```
ariants.vcf.gz (4136633 bytes)
2017-11-17T11:31:37.880521500Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/results/variants/v
ariants.vcf.gz.tbi (59036 bytes)
2017-11-17T11:31:37.880587900Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/runWorkflow.py (77
95 bytes)
2017-11-17T11:31:37.882003000Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/runWorkflow.py.con
fig.pickle (4785 bytes)
2017-11-17T11:31:37.882031100Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/workflow.error.log
.txt (0 bytes)
2017-11-17T11:31:37.882054800Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/workflow.exitcode.
txt (2 bytes)
2017-11-17T11:31:37.882076800Z Uploading strelka2/chr19/NA24385-chr19_0_16089283-block-work/workflow.warning.l
og.txt (0 bytes)
2017-11-17T11:31:37.882133000Z While uploading output files: Symlink in output "/strelka2/chr19/NA24385-chr19_
0_16089283-block.vcf.gz" points to invalid location "genome.S1.vcf.gz": lstat /tmp/603101599/strelka2/chr19/ge
nome.S1.vcf.gz: no such file or directory
2017-11-17T11:31:37.882144900Z Cancelled
```

## #2 - 11/17/2017 04:01 PM - Peter Amstutz

I think what it is trying to tell you that the output file `strelka2/chr19/NA24385-chr19_0_16089283-block.vcf.gz` is a symlink to `genome.S1.vcf.gz`, but `genome.S1.vcf.gz` doesn't exist.

It looks like it might be resolving the relative symlink incorrectly, which would be a bug. I'll have to investigate.

## #3 - 11/17/2017 07:09 PM - Brad Chapman

Thanks Peter. I'll work around this by avoiding the symlinks and copying the files out of the `strelka2` specific directory into the `out` directory. If you end up digging into it, I'd be interested to know if it's the relative symlinks or the nested double symlink that causes issues, in case I do any symlinking elsewhere. I'll report back after testing with the copies output.

## #4 - 11/17/2017 07:28 PM - Peter Amstutz

- Assigned To set to Peter Amstutz

- Target version set to 2017-11-22 Sprint

## #5 - 11/22/2017 07:23 PM - Peter Amstutz

- Target version changed from 2017-11-22 Sprint to 2017-12-06 Sprint

## #6 - 12/06/2017 07:33 PM - Peter Amstutz

- Target version changed from 2017-12-06 Sprint to 2017-12-20 Sprint

## #7 - 12/06/2017 08:30 PM - Peter Amstutz

- Target version changed from 2017-12-20 Sprint to Arvados Future Sprints

## #8 - 04/09/2018 03:33 PM - Tom Clegg

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

## #9 - 04/09/2018 07:46 PM - Peter Amstutz

Here's a minimal CWL test case:

```
class: CommandLineTool
cwlVersion: v1.0
inputs: []
outputs:
  out:
    type: File
    outputBinding:
      glob: Test1-chrM_0_1000-block.vcf.gz
requirements:
  ShellCommandRequirement: {}
arguments:
  - shellQuote: false
    valueFrom: |
      mkdir -p Test1-chrM_0_1000-block-work/results/variants
      echo genome | gzip -c > Test1-chrM_0_1000-block-work/results/variants/genome.S1.vcf.gz
      cd Test1-chrM_0_1000-block-work/results/variants
      ln -s genome.S1.vcf.gz genome.vcf.gz
      cd ../../..
      ln -s Test1-chrM_0_1000-block-work/results/variants/genome.vcf.gz Test1-chrM_0_1000-block.vcf.gz
```

This fails on current master but passes on [#13100](#) branch.

**#10 - 04/10/2018 01:35 PM - Tom Clegg**

- *Status changed from New to Resolved*

- *% Done changed from 0 to 100*

Applied in changeset [arvados!95be914af0ab0a82c4fa92b3f9c29ebec88e8595](#).

**#11 - 09/12/2018 04:55 PM - Tom Morris**

- *Target version deleted (Arvados Future Sprints)*