Arvados - Bug #9865
[CWL] Fix undefined behavior after ignoring an unhandled exception
08/26/2016 06:12 PM - Tom Clegg

<table>
<thead>
<tr>
<th>Status</th>
<th>Resolved</th>
<th>Start date:</th>
<th>02/25/2019</th>
</tr>
</thead>
<tbody>
<tr>
<td>Priority</td>
<td>Normal</td>
<td>Due date:</td>
<td></td>
</tr>
<tr>
<td>Assigned To</td>
<td>Eric Biagiotti</td>
<td>% Done:</td>
<td>100%</td>
</tr>
<tr>
<td>Category</td>
<td>SDKs</td>
<td>Estimated time:</td>
<td>0.00 hour</td>
</tr>
<tr>
<td>Target version</td>
<td>2019-03-13 Sprint</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Description

Example from the wild

2016-08-26 13:28:37 arvados.cwl-runner[20416] ERROR: While getting final output object: global name 'adjustFiles' is not defined

There's a section here where any exception will be reported but then leave "outputs" in some half-baked state.

https://github.com/curoverse/arvados/blob/master/sdk/cwl/arvados_cwl/runner.py#L131

```python
try:
    outc = arvados.collection.Collection(record["output"])  # c8d9a898cde654b53200bda0b0ef8b406dd71739
    with outc.open("cwl.output.json") as f:
        outputs = json.load(f)
    def keepify(fileobj):
        path = fileobj["location"]
        if not path.startswith("keep:"):
            fileobj["location"] = "keep:%s/%s" % (record["output"], path)
        adjustFileObjs(outputs, keepify)
        adjustDirObjs(outputs, keepify)
    except Exception as e:
        logger.error("While getting final output object: %s", e)
```

This code should either:

- Reset outputs to None in the "except" block; or
- Make the "try" scope smaller, so once "outputs" isn't None, unexpected exceptions get propagated up.

It should also log the full backtrace for the caught exception.

The "try" block looks like it started accidentally including too much code in c8d9a898cde654b53200bda0b0ef8b406dd71739

Another example from the wild

2019-01-30T04:31:12.159992495Z Exception RuntimeError: 'maximum recursion depth exceeded while calling a Python object' in <bound method _fileobject.__del__ of <socket._fileobject object at 0x7f10930130f50>> ignored

After this, arvados-cwl-runner stopped producing logs every 3 seconds, and appeared to be deadlocked.

Another example the following day reported a "maximum recursion depth" error, but kept logging "cwltool DEBUG: [workflow workflow.json#main] job step [...] not ready".

Subtasks:

- Task # 14840: Review 9865-cwl-fix-ignored-exceptions Resolved

Associated revisions

Revision e231ef3 - 02/28/2019 08:33 PM - Eric Biagiotti
Merge branch '9865-cwl-fix-ignored-exceptions'

refs #9865

Arvados-DCO-1.1-Signed-off-by: Eric Biagiotti <ebiagiotti@veritasgenetics.com>

History

#1 - 01/30/2019 06:06 PM - Tom Clegg
- Description updated

#2 - 01/31/2019 10:03 PM - Tom Clegg
- Description updated

#3 - 01/31/2019 10:23 PM - Tom Clegg
The first example was at least partly addressed in b9f7a8693579045f05d142db8ff42c5660dtfc (it now fails instead of proceeding with undefined results). I'm not sure whether logger here is configured such that logger.exception() prints a stack trace.

Grep reveals some other dubious uses of "except Exception", like this

```python
except Exception as e:
    logger.info("Setting container output: %s", e)
```

#4 - 02/13/2019 04:18 PM - Tom Morris
- Target version set to 2019-02-27 Sprint

#5 - 02/13/2019 04:33 PM - Tom Morris
- Assigned To set to Eric Biagiotti

#6 - 02/20/2019 02:52 PM - Eric Biagiotti
- Status changed from New to In Progress

#7 - 02/25/2019 08:09 PM - Eric Biagiotti
9865-cwl-fix-ignored-exceptions at 6ab33cb91de748f48ec8e12377f0b0650804db1

Added stack trace info by using logger.exception where appropriate, cleaned up formatting, added comments clarifying exclusion of stack traces in some instances, and fixed some exception handling. More details in the commit logs.

Unit tests: https://ci.curoverse.com/view/Developer/job/developer-run-tests/1085

#8 - 02/25/2019 10:58 PM - Peter Amstutz
Reviewing 9865-cwl-fix-ignored-exceptions @ 6ab33cb91de748f48ec8e12377f0b0650804db1

I think you missed this one in executor.py:364

```python
except Exception as e:
    logger.warning("Error checking states on API server: %s", e)
    remain_wait = self.poll_interval
    continue
```

#9 - 02/26/2019 02:30 PM - Eric Biagiotti

Peter Amstutz wrote:

Reviewing 9865-cwl-fix-ignored-exceptions @ 6ab33cb91de748f48ec8e12377f0b0650804db1

I think you missed this one in executor.py:364

```python
except Exception as e:
    logger.warning("Error checking states on API server: %s", e)
    remain_wait = self.poll_interval
    continue
```

#10 - 02/26/2019 03:08 PM - Peter Amstutz

Eric Biagiotti wrote:

I left line 364 deliberately. It seemed like a stack trace would be more clutter than helpful here, but I'll defer to you on this if you think it would be helpful.
Peter Amstutz wrote:

Reviewing 9865-cwl-fix ignored exceptions @ 6ab33cb91de748f4f8ec8e12377f0b0650804db1

I think you missed this one in executor.py:364

[...]

I left line 364 deliberately. It seemed like a stack trace would be more clutter than helpful here, but I'll defer to you on this if you think it would be helpful.

The entire motivation for this ticket to be put on the sprint was because there's a bug that happens in that specific try/catch block and we were hoping to get a stack trace for it.

#12 - 02/26/2019 06:34 PM - Eric Biagiotti

Peter Amstutz wrote:

Eric Biagiotti wrote:

Peter Amstutz wrote:

Reviewing 9865-cwl-fix ignored exceptions @ 6ab33cb91de748f4f8ec8e12377f0b0650804db1

I think you missed this one in executor.py:364

[...]

I left line 364 deliberately. It seemed like a stack trace would be more clutter than helpful here, but I'll defer to you on this if you think it would be helpful.

The entire motivation for this ticket to be put on the sprint was because there's a bug that happens in that specific try/catch block and we were hoping to get a stack trace for it.

Thanks for the clarification. Got caught up in the first example and missed the second one in the description. Fixed in 50ecc8767d980670db0f8906da4da57e56779e54

Unit tests: https://ci.curoverse.com/view/Developer/job/developer-run-tests/1087/

Finally got workbench integration tests to pass: https://ci.curoverse.com/job/developer-run-tests-apps-workbench-integration/1127/

#13 - 02/27/2019 01:57 PM - Eric Biagiotti

- Target version changed from 2019-02-27 Sprint to 2019-03-13 Sprint

#14 - 02/28/2019 06:50 PM - Peter Amstutz

Unfortunately error handling is hard to test, and the unit tests don't cover most of these exception cases. I've had exception branches that themselves have bugs that go undetected for months until the one time the error gets thrown, and then instead of logging the error we get a new one instead.

However as a second pair of eyes, I don't see any obvious mistakes, so I think this LGTM.

#15 - 02/28/2019 08:36 PM - Eric Biagiotti

- Status changed from In Progress to Resolved

#16 - 03/01/2019 07:33 PM - Tom Morris

- Release set to 15