PR Testing and the “Terrible Diagram”

Presented by

William McLendon
A Quick History the Trilinos Pull Request (PR) Testing

The Framework team enabled automatic PR testing in 2018

The “autotester” was developed in-house in Python by a member of department 1424.
- Trilinos PR testing uses the autotester, but it’s not the autotester.
- PR testing is a whole framework of infrastructure due to the complexities of the environment we work in and the complexity of Trilinos itself.

Testing Trilinos is tricky
- There are many packages and options – We really can’t permute them all.
- Trilinos is a large and complex library that takes hours to compile & test.
- Trilinos is hosted and developed on Github
- We must test on a variety of non-standard platforms such as testbeds, GPU systems, etc.
- The early roll out of automated PR testing had the usual expected issues.

The Origins of the Terrible Diagram
- I put the “Terrible Diagram” together in 2018 as a thought exercise to try and capture the pain-points associated with network connections required by the system.
- Jim Willenbring coined the ‘Terrible Diagram’ term.
20+ fixed comm events per PR.

Also polling & monitoring comm events:
- GH for new commits.
- Jenkins for failures.

Hundreds of comm events for a typical PR that can cause ‘random’ failures.
Some Job Stats for Trilinos PR Testing Q4-2020 to Q4-2021

Stats are for both Autotester itself and individual builds.

Stats collected via script that scans Jenkins console logs.

- ~61,200 Jenkins jobs run.
- ~51,700 ‘SUCCESS’ jobs.
- ~7,000 SNL proxy failures.
- ~450 ‘User’ errors
- ~200 SNL Gitlab failures.
- ~1,800 ‘Other’

Job failures tend to occur in clusters.

PR’s are all-or-nothing so 5x jobs pass and 1x fails is still a failed run.

Total Wall Time: ~130,000,000 s.
- Jenkins WALL time not CPU time.
Spotting Framework Errors

A common indicator of a possible ‘terrible diagram’ issue is when a build fails but CDash shows all green:

All 6 jobs shown on CDash passed...
Spotting Framework Errors

A common indicator of a possible ‘terrible diagram’ issue is when a build fails but CDash shows all green:

<table>
<thead>
<tr>
<th>PR</th>
<th>Test Result</th>
<th>Status</th>
<th>Passes</th>
<th>Fails</th>
<th>CI Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR-9929-test-Trilinos_pullrequest_gcc_7.2.0_serial-3536</td>
<td>![Status Icon]</td>
<td>![Pass Icon]</td>
<td>6</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>PR-9929-test-Trilinos_pullrequest_gcc_8.3.0-5993</td>
<td>![Status Icon]</td>
<td>![Pass Icon]</td>
<td>4</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>PR-9929-test-Trilinos_pullrequest_gcc_7.2.0_debug-4054</td>
<td>![Status Icon]</td>
<td>![Pass Icon]</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PR-9929-test-Trilinos_pullrequest_intel_17.0.1-11204</td>
<td>![Status Icon]</td>
<td>![Pass Icon]</td>
<td>4</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>PR-9929-test-Trilinos_pullrequest_clang_10.0.0-4002</td>
<td>![Status Icon]</td>
<td>![Pass Icon]</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

PR-9929-test-ref7_sans-gnu-7.2.0-anaconda3-serial_debug_shared_no-kokkos-arch_no-asan_no-complex_no-fpic_no-mpi_no-pb_no-dtc_pr-framework-638

But the GitHub commit message shows 8 jobs launched:

All 6 jobs shown on CDash passed...
A common indicator of a possible ‘terrible diagram’ issue is when a build fails but CDash shows all green:

```
PR-9929-test-Trilinos_pullrequest_gcc_7.2.0_serial-3536 0 0 0 25 0 0 773
PR-9929-test-Trilinos_pullrequest_gcc_8.3.0-5993 0 0 4 0 50 0 0 1347
PR-9929-test-Trilinos_pullrequest_gcc_7.2.0_debug-4004 0 0 4 0 0 0 0 1327
PR-9929-test-Trilinos_pullrequest_intel_17.0.1-11204 0 0 4 0 50 0 0 1443
PR-9929-test-Trilinos_pullrequest_clang_10.0.0-4902 0 0 4 0 0 0 0 1337
PR-9929-test-riem_7.3.0-anaconda3-serial_debug_shared_no-tokkno-arch_no-sssan_no-complex_no-fpic_no-nmpi_no-rt_no-dcc_pr-framework-638 0 0 0 0 0 0 0 12
```

All 6 jobs shown on CDash passed...

But the GitHub commit message shows 8 jobs launched:

In this situation, what happened is the autotester launched 8 jobs but only 6 ever made it far enough in the framework to report to CDash. Expand console output shows the issue.
Questions?