Technology Update

Terrell Russell, Ph.D.
@terrellrussell
Chief Technologist, iRODS Consortium

June 5-7, 2018
iRODS User Group Meeting 2018
Durham, NC
In The Last Year

<table>
<thead>
<tr>
<th>iRODS Release</th>
<th>Issues Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1.11</td>
<td>92</td>
</tr>
<tr>
<td>4.2.2</td>
<td>39</td>
</tr>
<tr>
<td>4.2.3</td>
<td>76 + 18</td>
</tr>
</tbody>
</table>

Plugins

- Python Rule Engine Plugin
- Storage Tiering Rule Engine Plugin
- Auditing (AMQP) Rule Engine Plugin
- S3 Resource Plugin
- GSI Authentication Plugin
- Kerberos Authentication Plugin
- Curl Microservice Plugin

Clients

- Python iRODS Client
- Automated Ingest Framework
- Cloud Browser
In The Last Year

~/irods $ git shortlog --summary --numbered 4.1.10..4-1-stable
  31  Ben Keller
  20  Alan King
  20  Terrell Russell
   7  Hao Xu
   7  Rick Skarbez
   7  jkgill
   6  Zoey Greer
   5  Jason Coposky
   2  Matthew Vernon
   1  Brian N. Makin

~/irods $ git shortlog --summary --numbered 4.2.1..4-2-stable
  51  Alan King
  27  Jason Coposky
  22  Hao Xu
  18  Terrell Russell
  17  Ben Keller
  17  Zoey Greer
  15  Kory Draughn
   9  jkgill
   8  Justin James
   5  Andrew Kelly
   5  Kyle Ferriter
   2  d-w-moore
   2  xu-hao
   1  Felix A. Croes
   1  Marco Grossi
   1  Sam Morris
Ongoing Development Work

- iRODS 4.1.12
- iRODS 4.2.4
- iRODS 4.3.0
- Python iRODS Client (PRC)
- Indexing Framework
- Lustre Integration
- Metalnx
- CockroachDB Database Plugin
- Ceph RADOS Resource Plugin
- QueryArrow Database Plugin
- R Client Library (R-iRODS)
- Multipart Transfer, v5 API
- NFS4J Integration
- Metadata Templates
- Testing Infrastructure
4.1.12 Update

Replicas in a complex tree (across 2 zones)...

Not enough replicas! #3746
Problem: 1 copy == BAD
Cause(s): Network, storage, OS...
Solution: Try harder
Replication retry (tested with mungefs)

Too many replicas! #3665
Problem: Filling up storage
Cause: Surprise concurrency? Repl resc?
Solution: ???
"Enhanced logging" with PID/user info

To be continued...
4.3.0 Update

- irodsMonitor process
- Checksums moving down into resource plugins
- Containerize installation
- Logging overhaul
4.3.0 Logging Update

Today

- Inconsistent
- Incomplete (syslog support)
- Not very helpful in tracking what caused an error
- Not very helpful when multiple servers are involved

Design Goals

- Leverage existing logging library (spdlog, etc.)
- Enable admins to easily capture, process, and analyze activity
- Consistent
- Easily track errors across multiple servers (hostname, timestamp, PID, plugin, etc.)
- Tie in existing infrastructure
- Provide more options for controlling output

Join Us!

- What would you like to see accomplished with the logging refactor?
- RFC: https://github.com/irods/irods_rfcs/blob/master/0001_logging.md
- Have questions/suggestions? Let's talk!
iRODS Build and Test - Yesterday

July 2011
  • Python → Node.js → RabbitMQ → Celery → Eucalyptus

October 2012
  • Python → Node.js → ssh → OpenStack

January 2013
  • Hudson → Python → OpenStack

October 2013
  • Hudson → Python → vSphere long-running VMs

Spring 2015
  • Jenkins → Python → Ansible → vSphere dynamic VMs
iRODS Build and Test - Today

Spring 2015 - onwards

- Jenkins → Python → Ansible → zone_bundles → vSphere dynamic VMs

Changes Since 2017

- Ended support for CentOS 6 and Ubuntu 12
- Moved iRODS build/test logic from Ansible to per-repository python modules
- Unified Jenkins workflow to test all supported plugins
- Consolidated to two parameterized Jenkins jobs
  - run-script-on-vms
  - run-script-on-irods-zone
## iRODS Build and Test - Today

### Build History

<table>
<thead>
<tr>
<th>S</th>
<th>W</th>
<th>Name</th>
<th>Last Success</th>
<th>Last Failure</th>
<th>Last Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>✔</td>
<td>✔</td>
<td>build_and_test_all_plugins_workflow</td>
<td>6 hr 25 min - #16</td>
<td>4 days 16 hr - #15</td>
<td>6 hr 25 min</td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>irods-build-and-test-workflow</td>
<td>9 hr 21 min - #1035</td>
<td>4 days 16 hr - #1028</td>
<td>7 hr 23 min</td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>plugin-auth-gsi-test-workflow</td>
<td>23 min - #23</td>
<td>1 hr 5 min - #22</td>
<td>9 min 2 sec</td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>plugin-auth-kerberos-test-workflow</td>
<td>1 hr 5 min - #20</td>
<td>10 days - #16</td>
<td>28 min</td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>plugin-microservice-curl-build-test-workflow</td>
<td>1 hr 5 min - #16</td>
<td>4 days 16 hr - #14</td>
<td>8 min 26 sec</td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>plugin-resource-s3-test-workflow</td>
<td>13 days - #10</td>
<td>1 hr 5 min - #26</td>
<td>2 hr 28 min</td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>python-irodsclient-coverage-and-test-results</td>
<td>1 day 18 hr - #2973</td>
<td>N/A</td>
<td>0.19 sec</td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>python-irodsclient-github-tracking-master</td>
<td>1 day 18 hr - #65</td>
<td>3 days 2 hr - #94</td>
<td>17 min</td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>python-irodsclient-test-workflow</td>
<td>1 day 18 hr - #303</td>
<td>3 days 2 hr - #302</td>
<td>17 min</td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>rule-engine-plugin-audit-amqp-build-test-workflow</td>
<td>1 hr 15 min - #68</td>
<td>1 hr 5 min - #67</td>
<td>26 min</td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>rule-engine-plugin-audit-amqp-github-tracking-master</td>
<td>1 hr 15 min - #23</td>
<td>9 days 22 hr - #22</td>
<td>26 min</td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>rule-engine-plugin-python-build-test-workflow</td>
<td>4 days 16 hr - #55</td>
<td>5 days 16 hr - #51</td>
<td>6 hr 25 min</td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>rule-engine-plugin-python-github-tracking-master</td>
<td>1 day 0 hr - #16</td>
<td>11 days - #15</td>
<td>5 hr 28 min</td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>rule-engine-plugin-storage-tiering-build-test-workflow</td>
<td>1 hr 5 min - #180</td>
<td>11 days - #171</td>
<td>32 min</td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>rule-engine-plugin-storage-tiering-github-tracking-master</td>
<td>1 hr 36 min - #73</td>
<td>11 days - #71</td>
<td>32 min</td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>run-script-on-irods-zone</td>
<td>19 min - #5207</td>
<td>1 hr 1 min - #5193</td>
<td>4 min 56 sec</td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>run-script-on-irods-zone</td>
<td>4 min 3 sec - #2137</td>
<td>4 days 16 hr - #2109</td>
<td>3 min 42 sec</td>
</tr>
</tbody>
</table>
iRODS Build and Test - Future

- Increase coverage  
  (more plugins in CI)
- Move pipeline scripts to GitHub  
  (no logic in Jenkins)
- Address inconsistency  
  (false reds / pyvmomi errors)
- Containerize Jenkins  
  (easier to test / update / redeploy)
- Possibly move from VMs to containers  
  (speed / fewer moving parts)

Testing a distributed system is hard.

Input and other ideas are welcome.
Big Picture

4.2 is about sending things out
- Dynamic PEPs
- Pluggable Rule Engine
- Enterprise Messaging

4.3 (and friends) will be about bringing them back
- Facilitated by GQv2
Big Picture - GenQuery

Current GQ

Future GQv2

iRODS

GQ

SQL

iRODS

GQv2

SQL

GraphDB

MongoDB

...
Big Picture

Enterprise Messaging - Scatter

iRODS 4.2

Rule Engine Plugin Interface
Audit Plugin
Compute Plugin
Index Plugin
Triple Store Plugin

Audit Service

Enterprise Message Bus

Compute
Index
Triple Store
Big Picture

Metadata Virtualization - Gather

iRODS 4.3
GQv2 - Suite of Plugins Virtualizing Metadata
Open Source Community Engagement

- Technology Working Group
- Metadata Templates Working Group
- GitHub Issues
- Pull Requests
- Chat List
- Consortium Membership