I Upgraded iRODS And I Still Have All My Hair

John Constable

john.constable@sanger.ac.uk
Background

Wellcome Trust Sanger Institute has had iRODS since 2011

“Implementing a genomic data management system using iRODS in the Wellcome Trust Sanger Institute”

by Gen-Tao Chiang, .. Clapham, Coates DOI: 10.1186/1471-2105-12-361

~14PB, in five federated Zones, plus another.

Most iCAT’s in modest VMWare (a couple high traffic ones are physical servers) instances

Sanger DC - DDN SFA10k’s x4 and 2U iRES via fibrechannel, some HP SL4540 (inbuilt 60 disks)
(about 36 in total)

Janet Shared Data Centre - HP SL4540 (inbuilt 60 disks) - about 30

Configuration Management using CFEngine 3.1
Background (2)

All running iRODS 3.3.1 on Ubuntu 12.04, with Oracle RAC database backend.

Resource Groups to organise resources at sanger and resources at JSDC

Bash script to move resources where filesystem is > 98% full to a ‘full’ group.

Rules;

  Force uploads to go into certain groups (the ‘not full’ ones)
  Replicate data that goes into one group is copied to the other

Other Rules;

  Disable trash (we want to be able to use the filesystem to directly retrieve files in extremis)
  Force checksum generation on ingest

On one zone, don’t replicate files if they are put into a particular collection
Reason to move

- iRODS 3.3.1 end-of-life, not getting security fixes etc
- SSL end-to-end encryption
- Composable Resource Tree’s potential to simplify rules and scripting needed
- Gen3 Query aka external metadata search
- The 2000’s called and wanted their software back.
Fail to Plan, Plan to Fail

We were going to do this *right*!

Also, 14PB of 10 years of scientific research, often where you couldn't get another sample.

Scientific colleagues looking on with some trepidation.

We had a placement student (Hi Andy!).

We shared our plans with RENCI and Had A Lot Of Conference Calls

(are you a consortium member? Perhaps consider joining, this kind of thing was very helpful. I have not been paid by RENCI to write this)
### Plan to Fail Test Plan

<table>
<thead>
<tr>
<th>Raiser</th>
<th>Test Prior</th>
<th>Reference Number</th>
<th>User Story</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISG</td>
<td>MUST</td>
<td>1.1.01</td>
<td>As a SysAdmin I can upgrade an existing 3.3.1 iRODS install on Ubuntu 12.04 to iRODS 4.1.7 and have a client <code>iget</code> and <code>imeta</code> commands return the same result</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.1.02</td>
<td>As a SysAdmin I can upgrade an existing 3.3.1 iRODS install on RedHat 7.1 to iRODS 4.1.6 and have a client <code>iget</code> and <code>imeta</code> commands return the same result</td>
</tr>
<tr>
<td>ISG</td>
<td>MUST</td>
<td>1.1.03</td>
<td>As a user I should be able to authenticate using Kerberos before and after an iRODS upgrade</td>
</tr>
<tr>
<td>ISG</td>
<td>MUST</td>
<td>1.1.04</td>
<td>As a SysAdmin I should be able to follow the RENCI upgrade procedure to upgrade from iRODS 3.3.1 to 4.1.x</td>
</tr>
</tbody>
</table>

For more details, refer to [Update iRODS project](https://docs.google.com/spreadsheets/d/1m5I4ZQfCryz-).
Basic plan

Spin up 3.3.1 in a virtual environment.
Run BATS test framework to setup basic configuration & baseline features
Upgrade to 4.1.X
Run bats framework again to verify retained functionality
Rinse, repeat as bugs found
Test Plan

Coverage

- **Must**
  - Functionality we must have
  - Cannot upgrade without this working
  - Issues we have had fixed before
- **Should**
  - Would make operating the Zones easier
  - Better for the future
  - Functionality we’ve not used but plan to
- **Could**
  - Nice to have
  - Functionality we don’t plan to use but good to have ‘in the back pocket’
Unit tests!

- Surprisingly, not a pun on the iRODS name
- Simple to write
- Everyone knows bash
- No learning curve e.g. server spec
- Some use of shunit previously, so some history
- Helpful guide when getting started:
  https://blog.engineyard.com/2014/bats-test-command-line-tools
The BATS-mobile in action

./scripts/v3/icat/setup
./bats_scripts/icommands.bats
✓ Check the output of ils
✓ Check the output of ipwd
✓ make a collection
✓ remove a collection
✓ Check that iput stores a txt document correctly
✓ Check that iget can retrieve the txt document correctly
✓ Add Metadata
✓ List Metadata
✓ remove temporary file using irm
✓ clean up

10 tests, 0 failures
#!/usr/bin/env bats

@test "iput a file" {
    iput -K -f $INSERT_FILE
    run ils
    for i in $lines[@]; do
        if [ i = $INSERT_FILE ]; then
            [ true ]
        fi
    done
    [ false ]
Unit tests!

Did our BAT fly?

What we found:

- Unit tests fine, but making a functional test out of them was painful (we’d use server spec for that next time)
- Writing tests while doing testing lead to confusion over passing tests
- We could have done with using branches and tags, but we were new at this.
- Idempotence hard to implement
- Learned as much *writing* the tests as running them
- Orchestration across servers hard, ended up being manual (ansible? Fabric? serverspec?)
Virtualised Environments!

- Virtualised environment using virtualbox
- Can build irods 4.1.8 or 3.3.1 in a few mins, with oracle database, iCAT and two iRES servers.
<table>
<thead>
<tr>
<th>Status</th>
<th>Pipeline</th>
<th>Commit</th>
<th>Stages</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>passed</td>
<td>#658 by latest</td>
<td>master → 7c62efda</td>
<td>✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔</td>
<td>01:02:22, 2 months ago</td>
</tr>
<tr>
<td>passed</td>
<td>#6327 by latest</td>
<td>master → a86a2acf</td>
<td>✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔</td>
<td>01:02:07, 2 months ago</td>
</tr>
</tbody>
</table>
Design Documents!

ONE DOES NOT SIMPLY
'UPGRADE IRODS'
Waterfall and Gant Charts!

Agile time boxes not waterfall honest guv
Started on 4.1.7
OUR GOAL IS TO WRITE BUG-FREE SOFTWARE. I'LL PAY A TEN-DOLLAR BONUS FOR EVERY BUG YOU FIND AND FIX.

YAHOO! WE'RE RICH! YES!!! YES!!!

I HOPE THIS DRIVES THE RIGHT BEHAVIOR. I'M GONNA WRITE ME A NEW MINIVAN THIS AFTERNOON!
A few initial teething issues

- You can’t use v4 client to talk to v3 servers
- icommands packaged in /opt/renci not /software/irods
- Compile libraries will now be in /usr/include/irods
- Oracle not initially supported in 4.x
- When it became supported, client version 11.1 wasn’t so we’d need to upgrade
- No way to mark resources as full in Composite Resource Tree
- Rule ordering is important in server_config.json
- Because irods 3.3.1's Remote Zone SID uses a "-" to delimit server authentication, RENCI have removed it from the list of allowable characters from the V2/V3 schema. As a result, we have to use a local schema repo
End Result: An upgrade plan

- Upgrade Dev zones first, wait a couple of weeks for people to test
- Update Config management
- Update all 1000 users irods environment files
- Snapshot VM’s & backup /usr/local/iRODS
- Packages put config in separate place from 3.3.1, so easy(?) to revert if needed
- Automate installation of Oracle 11.2 client
- Run unit tests
- Backup database
- Install packages (not done via CFEngine due to phased rollout)
- Run database schema update
- Add specific environment variables for Unicode support for the Oracle plugin, Oracle paths to irods config files and set checksum to MD5 - all using RENCI’s update_json.py script (hidden gem!)
- Manually run setup script (should have automated this w/answers)
- Add Kerberos config to iRODS config files
- Disable irods_client_server_negotiation (until all zones upgraded)
- Check iget files OK

- Run unit tests
- Enable 'high water mark' (with 4.1.10 this would be set minimum_free_space_for_create)
- Modify server_config.json to include federation stanza for federation master and itself
- Modify server_config.json to include rules
- Check you can retrieve files
- Do all of the above for each iRES
Upgrade day!
4.1.8! 4.1.8!

August 2016, nearly a year after we started.

Began with the Federation Master, and icommands, as this would then broker the communication to the other Zones.

We found in short order;

- we couldn’t list across Zones because acACLPolicy wasn’t disabling Strict ACL’s.
- we couldn’t upload files to Federee zones because msiSetDefaultResc wasn’t honoured across zones, and so the composite tree didn’t know where to put the files.
- IIs -A run across zones was astonishingly slow (31 mins for 500 objects)

Paused the upgrade of the other zones until fixed
After upgrading to debug versions we found we couldn’t upload large files (~2G).
Most of our BAM files are at least this big.
Three weeks later and many debug builds and conf calls with RENCI, 4.1.9!
November 2016, we found that Resources whose file systems were full were being written to, despite taking advantage of the ‘High Water Mark’ functionality.

After much debug builds and work with RENCI, the ‘minimum free space for create in bytes’ was introduced in 4.1.10, in conjunction with updating the freespace on a resource (we have a cron script)
Once the dust from all of *that* had settled, we checked that we had enough replicas.

We didn’t, in part due to a congestion issue between our data centres.

So.. 4.1.11 might be coming to you soon, with the fixes from the assorted investigations into replication.

But that’s another talk (it was only 145TB and 70k files...).
Disruption

Total disruption;

- at least a month of unavailability (mostly due to three weeks working out why large files not uploading)
- Instability every couple of weeks as patch s/w installed
Lessons Learned

- Working with your customers scientists is invaluable.
- Check you have enough disk space
- Terrell and Jason are very patient people (also see above)
- Become a consortium member and discuss things on the iRODS-CHAT mailing list
- Test Frameworks such as BATS
- ‘Infrastructure As Code’ (config management and virtualised images in this case)
- CI helps to make the same mistakes again reliable packages if you must do that kind of thing
- naming resources by server name and LUN/disk combination e.g. irods-seq-i23-bc

```python
python /var/lib/irods/packaging/update_json.py /etc/irods/server_config.json string environment_variables,"ORACLE_HOME" "opt/oracle/instantclient_11_2"
```

- Izonereport and jq BFF

```json
izonereport | jq ".["zones"][].icat_server.resources[], .["zones"][].resource_servers[].resources[] | select(.host="irods-gl-dev.internal.sanger.ac.uk") | {name: .name, path: .vault_path, host: .host}
```

```json

```
```
The Future

4.1.11 to 4.2.

Must only take us two weeks to test..

Upgrade has to be smooth

Also, turn on SSL everywhere

GenQuery3

Ceph?
Thanks

- RENCI, especially
  - Terrell
  - Jason
  - Ben
  - Antoine
- NPG, especially
  - Keith James
- ISG
  - Andy Perry
  - Pete Clapham
  - Jon Nicholson