Solving IT Security Problems with iRODS

Alan Hall – NOAA’s National Climatic Data Center

iRODS User Group
Garching, Germany
February 28 – March 1, 2013
Anonymous FTP is:
- Not Secure
- Not a management tool (clean-up)
- Limited in scope to one-to-one relationship
Ingest & Archive: iRODS

iRODS is:
- Secure authentication
- Security via Obscurity (one to bind them)
- A virtual management tool (clean-up)
- Scope is entire grid
Future NCDC iRODS Implementation

Federated: We can irsync between the two Grids
From EITHER grid
From ANY server

From one server control all Federated data

irsync i:/DMZ/Archive/NR2 i:/NCDC/gulp1/Ingest/NR2
irsync i:/DMZ/Archive/NR3 i:/NCDC/gulp2/Ingest/NR3
S3 Cloud Storage Using iRODS: A Possible Access Paradigm

Alan Hall – NOAA’s National Climatic Data Center
Cloud Pilot Flow

NWS NEXRAD Level-2

CLASS
- Ingest
- FTP

NCDC

hook

AWS S3

SIPGenSys
iRODS Compound Resource
iRODS

AIPs are removed from borg5 when they reach the archive by iRODS
Cloud Pilot Access

- AWS S3
- iRODS
- NCDC
- RENCI
- HydroNEXRAD
- Other VA Access
- Future Possibilities
CLASS Cloud Pilot: NCDC

SIPGenSys
iRODS Compound Resource
iRODS

iRODS Workflow

HydroNEXRAD

AWS S3

Other VA Access

CLASS

Ingest

FTP

NCDC

borg5

hook

NCDC HDSS

SIPGenSys
iRODS Compound Resource
iRODS

iRODS Workflow

HydroNEXRAD

AWS S3

Other VA Access
iRODS S3 Connection is from borg5 (DMZ) and the local cache for S3 is on borg5.

Any transfers to S3 must be staged on borg5 first, then transferred to S3.

If sync is not necessary, then no staging.
S3 Upload/Download Transfer Rates Using iRODS

<table>
<thead>
<tr>
<th>Number of Streams</th>
<th>Upload Rate (MB/s)</th>
<th>Download Rate (MB/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>8,646,086</td>
<td>15,077,304</td>
</tr>
<tr>
<td>16</td>
<td>20,507,405</td>
<td>26,308,937</td>
</tr>
<tr>
<td>24</td>
<td>26,837,760</td>
<td>26,308,937</td>
</tr>
</tbody>
</table>

- Upload Rate
  - 8 streams: 8,646,086 MB/s (740GB/day)
  - 16 streams: 20,507,405 MB/s (1.8TB/day)
  - 24 streams: 26,837,760 MB/s (2.3TB/day)

- Download Rate
  - 8 streams: 15,077,304 MB/s (1.3TB/day)
  - 16 streams: 26,308,937 MB/s (2.2TB/day)
  - 24 streams: 26,308,937 MB/s (2.2TB/day)
### Borg5 Network Traffic

**Top Conversation OUT Report**

**IP Group Name:** Borg

**Report Start Time:** 2013-01-16 08:25  
**Report End Time:** 2013-01-16 14:25

<table>
<thead>
<tr>
<th>Src IP</th>
<th>Dst IP</th>
<th>Application</th>
<th>Port</th>
<th>Protocol</th>
<th>DSCP</th>
<th>Traffic</th>
</tr>
</thead>
<tbody>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>igor.class.ncdc.noaa.gov</td>
<td>TCP_App</td>
<td>*</td>
<td>TCP</td>
<td>000010</td>
<td>12.21 GB</td>
</tr>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>s3-1.amazonaws.com</td>
<td>https</td>
<td>443</td>
<td>TCP</td>
<td>Default</td>
<td>2.36 GB</td>
</tr>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>s3-1.amazonaws.com</td>
<td>https</td>
<td>443</td>
<td>TCP</td>
<td>Default</td>
<td>1.95 GB</td>
</tr>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>s3-1.amazonaws.com</td>
<td>https</td>
<td>443</td>
<td>TCP</td>
<td>Default</td>
<td>1.83 GB</td>
</tr>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>s3-1.amazonaws.com</td>
<td>https</td>
<td>443</td>
<td>TCP</td>
<td>Default</td>
<td>1.38 GB</td>
</tr>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>s3-1.amazonaws.com</td>
<td>https</td>
<td>443</td>
<td>TCP</td>
<td>Default</td>
<td>1.06 GB</td>
</tr>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>s3-1.amazonaws.com</td>
<td>https</td>
<td>443</td>
<td>TCP</td>
<td>Default</td>
<td>916.65 MB</td>
</tr>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>s3-1.amazonaws.com</td>
<td>https</td>
<td>443</td>
<td>TCP</td>
<td>Default</td>
<td>113.92 MB</td>
</tr>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>s3-1.amazonaws.com</td>
<td>https</td>
<td>443</td>
<td>TCP</td>
<td>Default</td>
<td>45.61 MB</td>
</tr>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>ftp.class.ncdc.noaa.gov</td>
<td>ftp</td>
<td>21</td>
<td>TCP</td>
<td>Default</td>
<td>18.19 MB</td>
</tr>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>ftp.class.ncdc.noaa.gov</td>
<td>TCP_App</td>
<td>*</td>
<td>TCP</td>
<td>Default</td>
<td>6.3 MB</td>
</tr>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>ftp.class.ncdc.noaa.gov</td>
<td>ftp</td>
<td>21</td>
<td>TCP</td>
<td>Default</td>
<td>2.05 MB</td>
</tr>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>igor.class.ncdc.noaa.gov</td>
<td>TCP_App</td>
<td>*</td>
<td>TCP</td>
<td>Default</td>
<td>269.42 KB</td>
</tr>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>s9a1.psmtp.com</td>
<td>smtp</td>
<td>25</td>
<td>TCP</td>
<td>Default</td>
<td>22.62 KB</td>
</tr>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>time-c.timefreq.bldrdoc.gov</td>
<td>ntp</td>
<td>123</td>
<td>UDP</td>
<td>Default</td>
<td>1.29 KB</td>
</tr>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>62-25-50-200-.static.centennialpr.net</td>
<td>ntp</td>
<td>123</td>
<td>UDP</td>
<td>Default</td>
<td>1.29 KB</td>
</tr>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>131.107.13.100</td>
<td>ntp</td>
<td>123</td>
<td>UDP</td>
<td>Default</td>
<td>1.29 KB</td>
</tr>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>113.64.79.216.68.in-addr.arpa</td>
<td>ntp</td>
<td>123</td>
<td>UDP</td>
<td>Default</td>
<td>1.29 KB</td>
</tr>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>128.136.188.172</td>
<td>ntp</td>
<td>123</td>
<td>UDP</td>
<td>Default</td>
<td>1.21 KB</td>
</tr>
<tr>
<td>borg5.ncdc.noaa.gov</td>
<td>time-b.nist.gov</td>
<td>ntp</td>
<td>123</td>
<td>UDP</td>
<td>Default</td>
<td>1.21 KB</td>
</tr>
</tbody>
</table>
Lessons Learned: Cheers

AWS:
• Overall very good experience with setup and basic management

iRODS:
• iRODS S3 setup was amazingly easy
• Basic Restful APIs were easy to use
• The S3 Resource in iRODS is a compound resource:
  o Must be linked to a local filesystem cache
• Security via Obscurity
  o iRODS Commands are run from hook, not borg5
  o When transferring to S3 from hook, borg5 makes the connection
• HTTPS Transfers
• iRODS Restful Download supported multi-part download (threads)
  o Not on upload because iRODS is not installed on the S3 Resource
• Our processes do NOT have to be tightly coupled
  o We did NOT wait for CLASS to provide AIP on ftp subscription
• Almost magic!
Lessons Learned: Jeers

AWS:
• AWS GUI Tools were adequate but hopefully will improve
  IAM for managing users good, but rules could be complex
  S3 tools were pretty basic

iRODS:
• Restful APIs did not implement basic list function
  o Need listings to know what is actually on the resource
• No way to force Reduced Redundancy on upload thru iRODS
• iRODS registration of files on compound resource difficult

iRODS Developers are working on these
Challenges

• To utilize Cloud Storage
  o Much thought has to go into the hierarchy:
    ✓ How much data to post? Rotating Window?
    ✓ How to make relationships with other data easier?
    ✓ Common Formats? Compress? Collections (tar)?
  o Most of the cost are downloads:
    ✓ How to predict?
    ✓ How to limit/control?
    ✓ Pay as you go?
iRODS Demo

[ldm@gulp2 logs]$ icd /DMZ/alan
[ldm@gulp2 logs]$ ils
/DMZ/alan:
   6910_QPE_short_2012120413.tar
   KGSP2012120415.tar
   srb_rel3.1_lw_monthly_200701_200712.tar

[ldm@gulp2 logs]$ ils -l
/DMZ/alan:
   rods  0  DMZBorg2Resc  39383040  2012-12-04.14:26 & 6910_QPE_short_2012120413.tar
   rods  0  DMZBorg1Resc  2826240  2012-12-04.14:11 & KGSP2012120415.tar
   rods  0  DMZBorg3Resc  18769920  2012-12-04.14:20 & srb_rel3.1_lw_monthly_200701_200712.tar
Log: From borg5 to hook

22:25:03: ----- -- iRODSirsyncTh hook MultiStream BEGIN -----------------------------------------------
22:25:03: irsync From: i:/DMZ/Archive/NR2/KLTX To: i:/DMZ/NWS-WxRadar-2/KLTX/
22:25:03: irsync -v -r -R s3Resc i:/DMZ/Archive/NR2/KLTX i:/DMZ/NWS-WxRadar-2/KLTX/
--StdOut: NWS_NEXRAD_NXL2DP_KLTX_20 5.869 MB | 4.476 sec | 0 thr | 1.311 MB/s
--StdOut: NWS_NEXRAD_NXL2DP_KLTX_20 0.000 MB | 0.317 sec | 0 thr | 0.000 MB/s
22:25:08: SysCmd: StdOut END

22:25:08: iRODSirsyncTh: hook From: i:/DMZ/Archive/NR2/KLTX To: i:/DMZ/NWS-WxRadar-2/KLTX/ ET: 5 Err: none ET: 00:05
22:25:08: ----- -- iRODSirsyncTh hook MultiStream END --- ET: 00:05 -----------------------------------------------
22:25:08:-----------------------------------------------

22:25:03: ----- -- iRODSirsyncTh hook MultiStream BEGIN -----------------------------------------------
22:25:03: irsync From: i:/DMZ/Archive/NR2/KGSP To: i:/DMZ/NWS-WxRadar-2/KGSP/
22:25:03: irsync -v -r -R s3Resc i:/DMZ/Archive/NR2/KGSP i:/DMZ/NWS-WxRadar-2/KGSP/
--StdOut: NWS_NEXRAD_NXL2DP_KGSP_20 5.967 MB | 5.496 sec | 0 thr | 1.086 MB/s
--StdOut: NWS_NEXRAD_NXL2DP_KGSP_20 0.000 MB | 0.288 sec | 0 thr | 0.000 MB/s
22:25:09: SysCmd: StdOut END

22:25:09: ----- -- iRODSirsyncTh hook MultiStream END --- ET: 00:06 -----------------------------------------------
Log: From borg5 to S3

00:25:34: ----- -- iRODSirsyncTh hook MultiStream BEGIN ---------------------------------------------------------------------
00:25:34: irsync From: i:/DMZ/Archive/NR2/KGSP To: i:/DMZ/NWS-WxRadar-2/KGSP/
00:25:34: -v -r s3Resc i:/DMZ/Archive/NR2/KGSP i:/DMZ/NWS-WxRadar-2/KGSP/
00:25:42: SysCmd: Cmd: irsync -v -r s3Resc i:/DMZ/Archive/NR2/KGSP i:/DMZ/NWS-WxRadar-2/KGSP/
--StdOut: NWS_NEXRAD_NXL2DP_KGSP_20 7.695 MB --- a match, no sync required
--StdOut: NWS_NEXRAD_NXL2DP_KGSP_20 0.000 MB --- a match, no sync required
--StdOut: NWS_NEXRAD_NXL2DP_KGSP_20 6.348 MB | 4.085 sec | 0 thr | 1.554 MB/s
--StdOut: NWS_NEXRAD_NXL2DP_KGSP_20 0.000 MB | 0.226 sec | 0 thr | 0.000 MB/s
00:25:42: SysCmd: StdOut END
00:25:42: iRODSirsyncTh: hook From: i:/DMZ/Archive/NR2/KGSP To: i:/DMZ/NWS-WxRadar-2/KGSP/ ET: 8 Err: none ET: 00:08
00:25:42: ArcSB Remove: rm /raid/gtsnp/tmp/iRODSirsyncTh_MultiStream.12355
00:25:42: ----- -- iRODSirsyncTh hook MultiStream END --- ET: 00:08 ---------------------------------------------------------------------

01:25:15: ----- -- iRODSirsyncTh hook MultiStream BEGIN ---------------------------------------------------------------------
01:25:15: irsync From: i:/DMZ/Archive/NR2/KLTX To: i:/NCDC/Ingest/NR2/KLTX/
01:25:15: -v -r i:/DMZ/Archive/NR2/KLTX i:/NCDC/Ingest/NR2/KLTX/
01:25:15: SysCmd: Cmd: irsync -v -r i:/DMZ/Archive/NR2/KLTX i:/NCDC/Ingest/NR2/KLTX/
--StdOut: NWS_NEXRAD_NXL2DP_KLTX_20 5.869 MB --- a match, no sync required
--StdOut: NWS_NEXRAD_NXL2DP_KLTX_20 0.000 MB --- a match, no sync required
--StdOut: NWS_NEXRAD_NXL2DP_KLTX_20 3.516 MB | 0.129 sec | 0 thr | 27.211 MB/s
--StdOut: NWS_NEXRAD_NXL2DP_KLTX_20 0.000 MB | 0.033 sec | 0 thr | 0.003 MB/s
01:25:15: SysCmd: StdOut END
01:25:15: iRODSirsyncTh: hook From: i:/DMZ/Archive/NR2/KLTX To: i:/NCDC/Ingest/NR2/KLTX/ ET: 0 Err: none ET: 00:00
01:25:16: ArcSB Remove: rm /raid/gtsnp/tmp/iRODSirsyncTh_MultiStream.632
01:25:16: ----- -- iRODSirsyncTh hook MultiStream END --- ET: 00:00 ---------------------------------------------------------------------