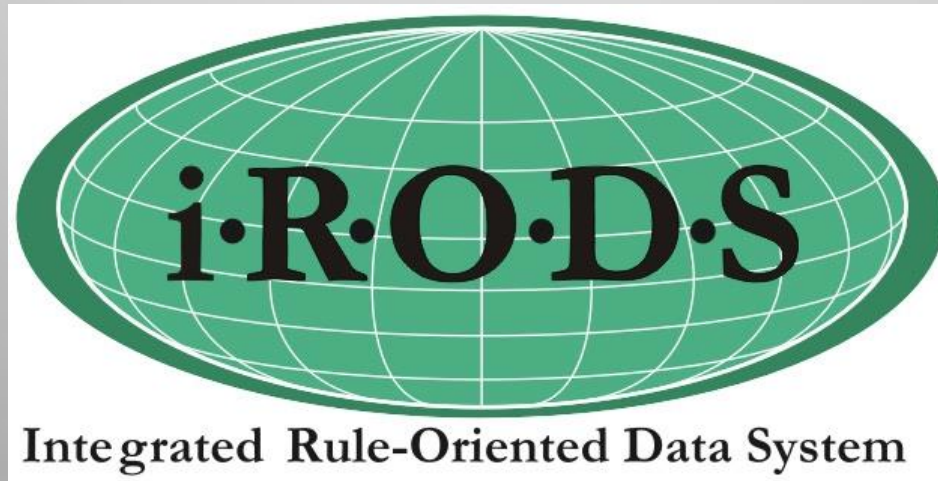


Solving IT Security Problems with iRODS

Alan Hall – NOAA's National Climatic Data Center

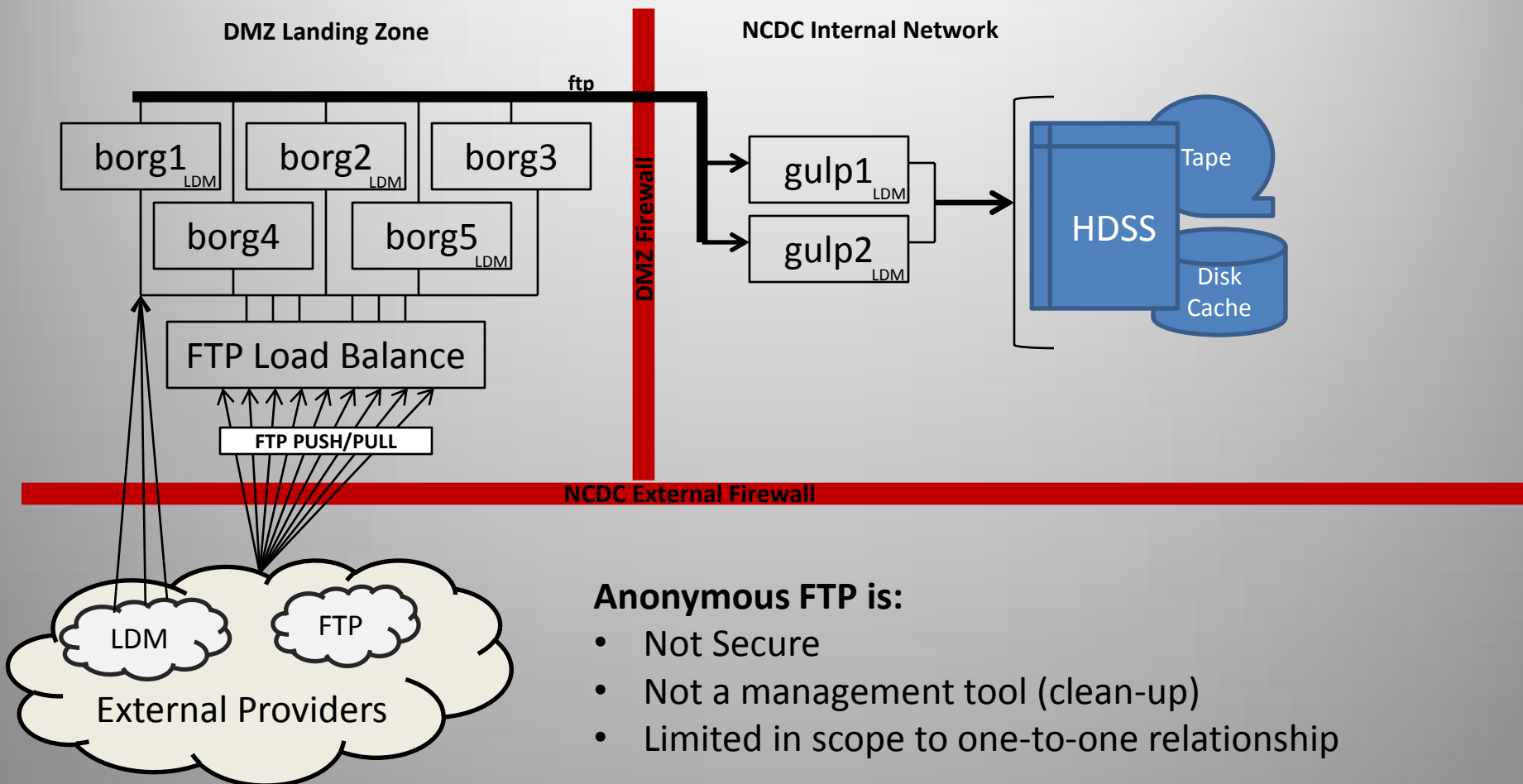


iRODS User Group

Garching, Germany

February 28 – March 1, 2013

Ingest & Archive: FTP



Anonymous FTP is:

- Not Secure
- Not a management tool (clean-up)
- Limited in scope to one-to-one relationship

Ingest & Archive: iRODS

DMZ Landing Zone

**iRODS DMZ
Grid**

/DMZ

/Archive

/NR2

/NR3

NCDC Internal Network

**iRODS NCDC
Grid**

/NCDC

/gulp1

/Ingest

/NR2

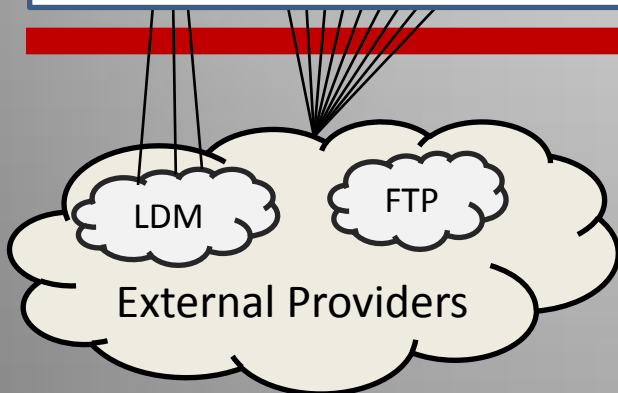
/gulp2

/Ingest

/NR3

DMZ Firewall

NCDC External Firewall

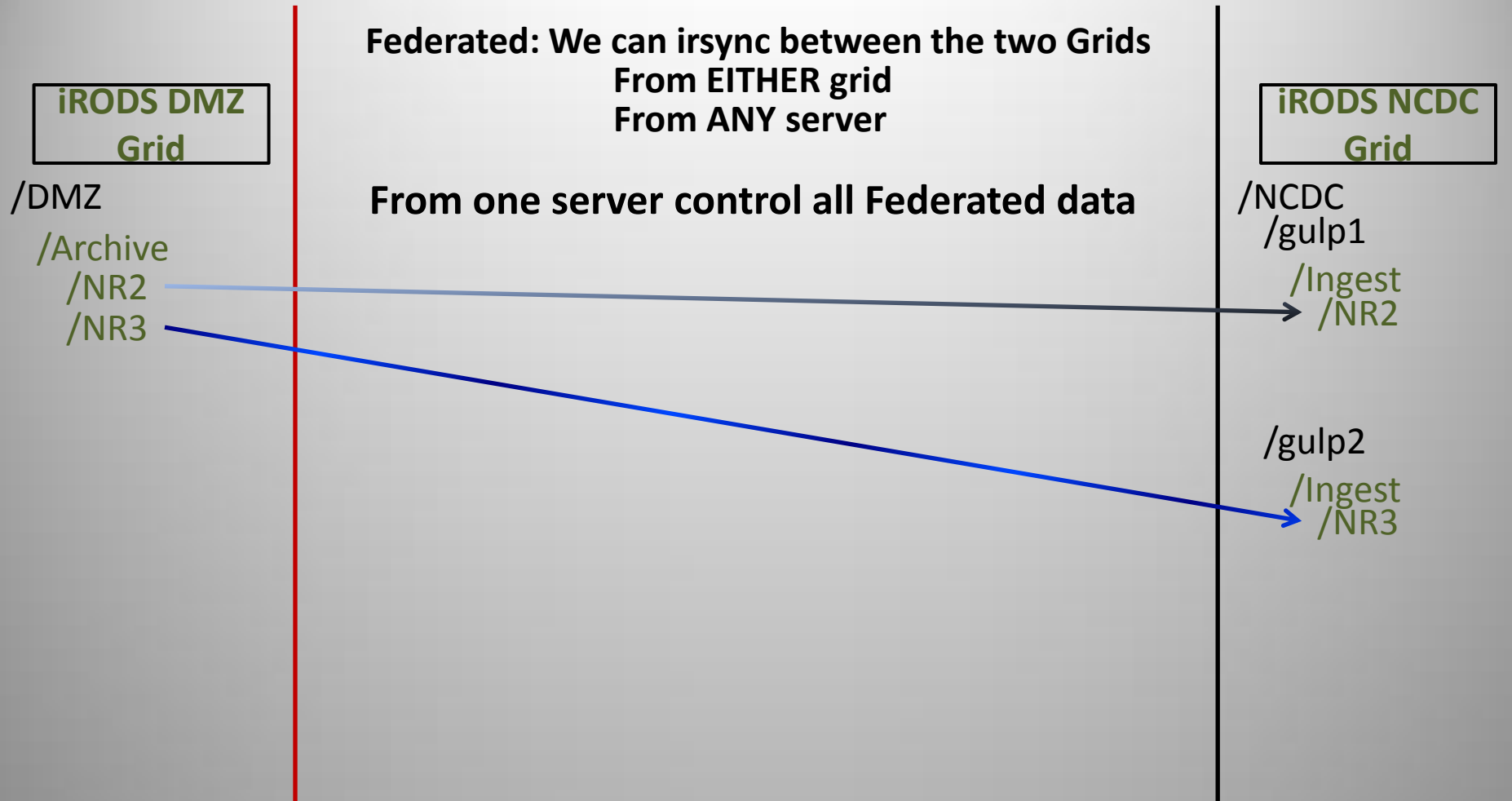


iRODS is:

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



Future NCDC iRODS Implementation



irsync i:/DMZ/Archive/NR2 i:/NCDC/gulp1/Ingest/NR2

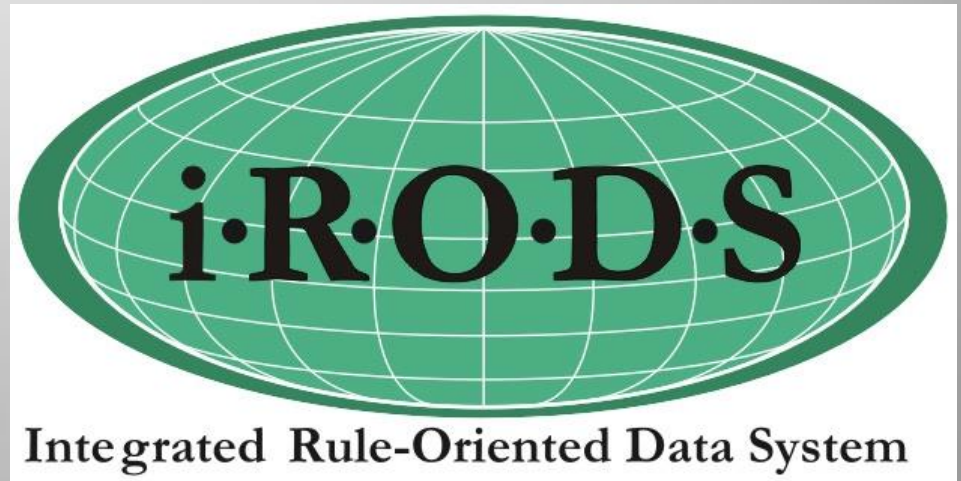
irsync i:/DMZ/Archive/NR3 i:/NCDC/gulp2/Ingest/NR3

S3 Cloud Storage Using iRODS: A Possible Access Paradigm

Alan Hall – NOAA's National Climatic Data Center



Amazon S3

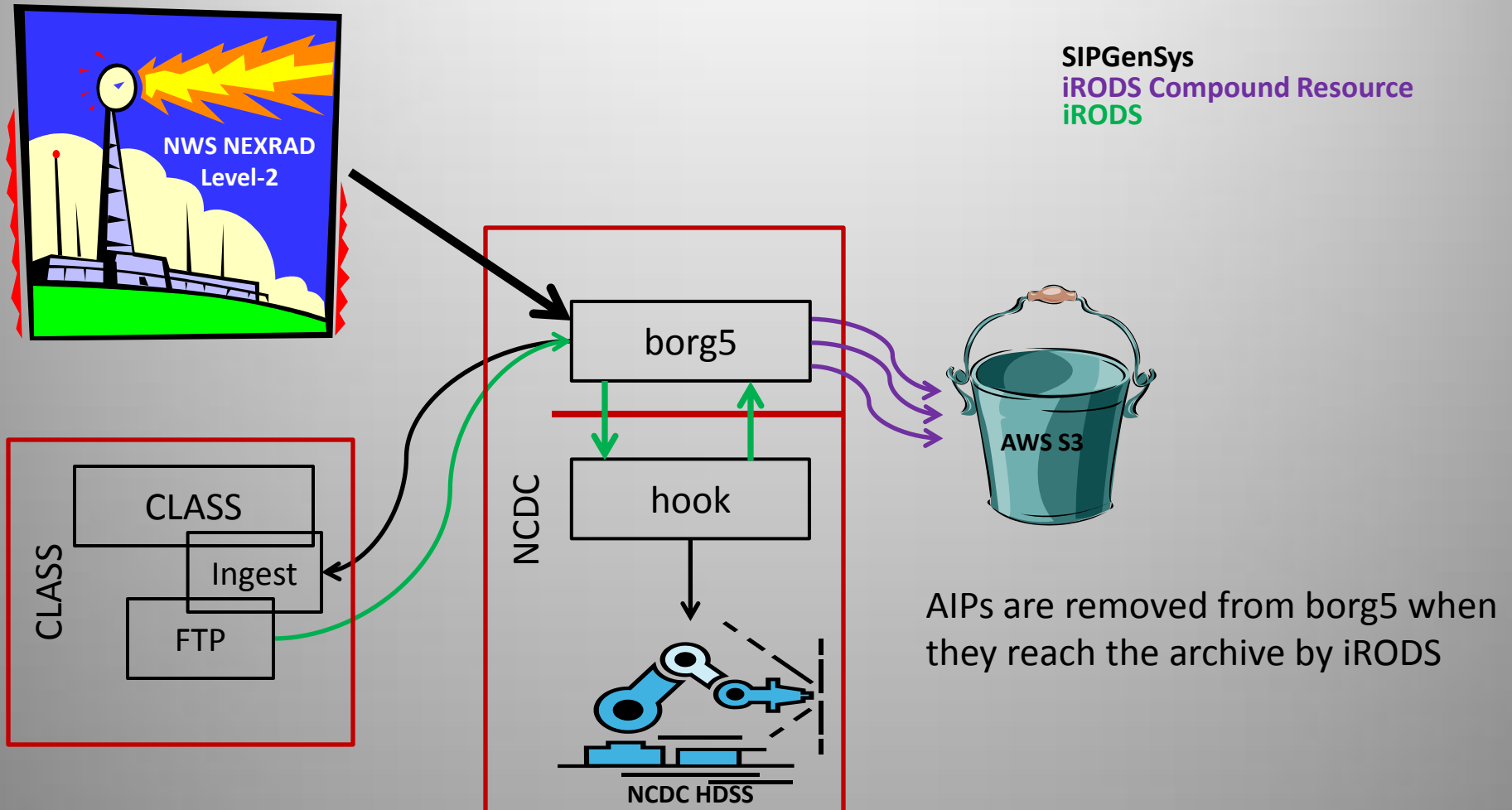


iRODS User Group

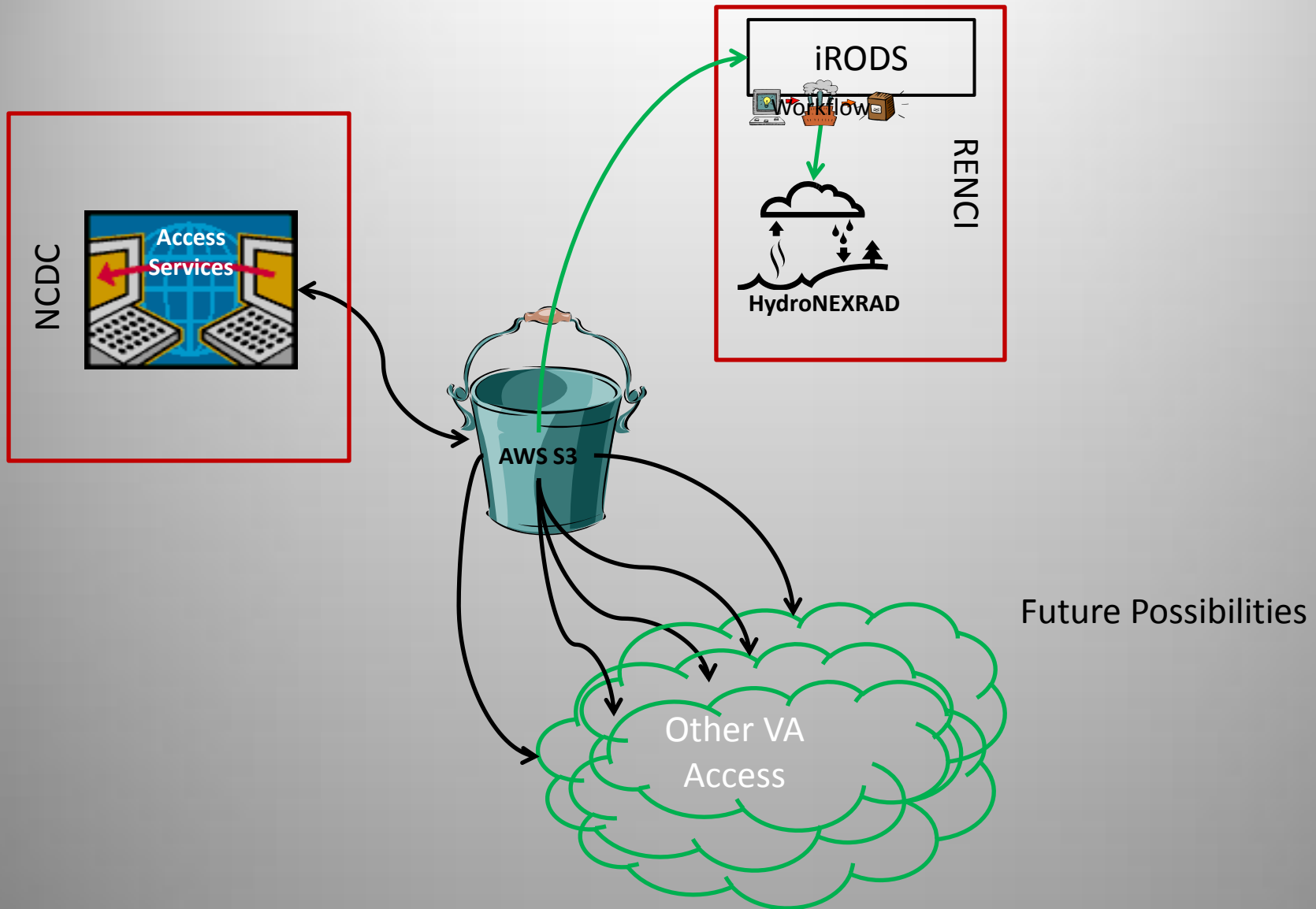
Garching, Germany

February 28 – March 1, 2013

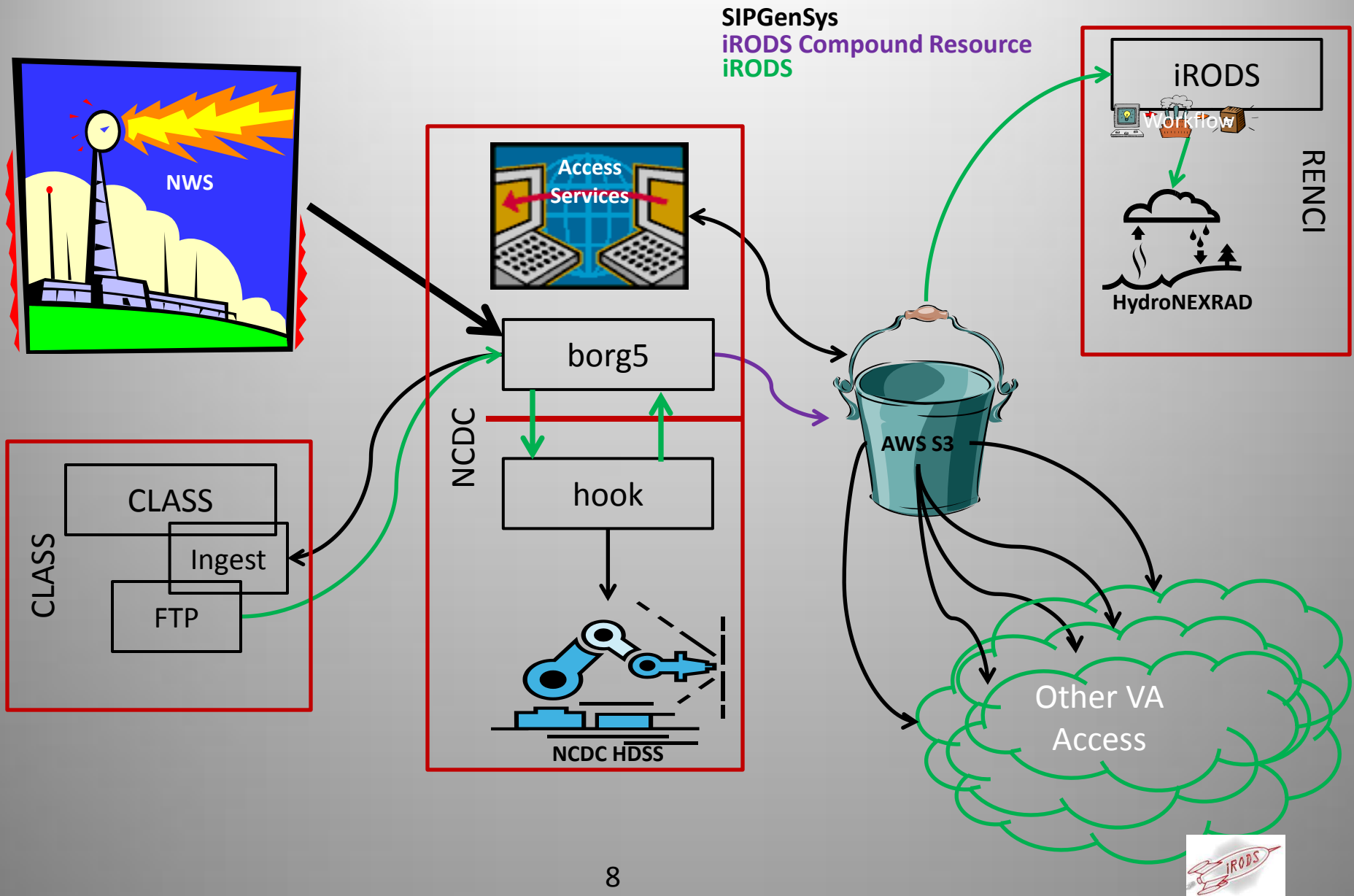
Cloud Pilot Flow



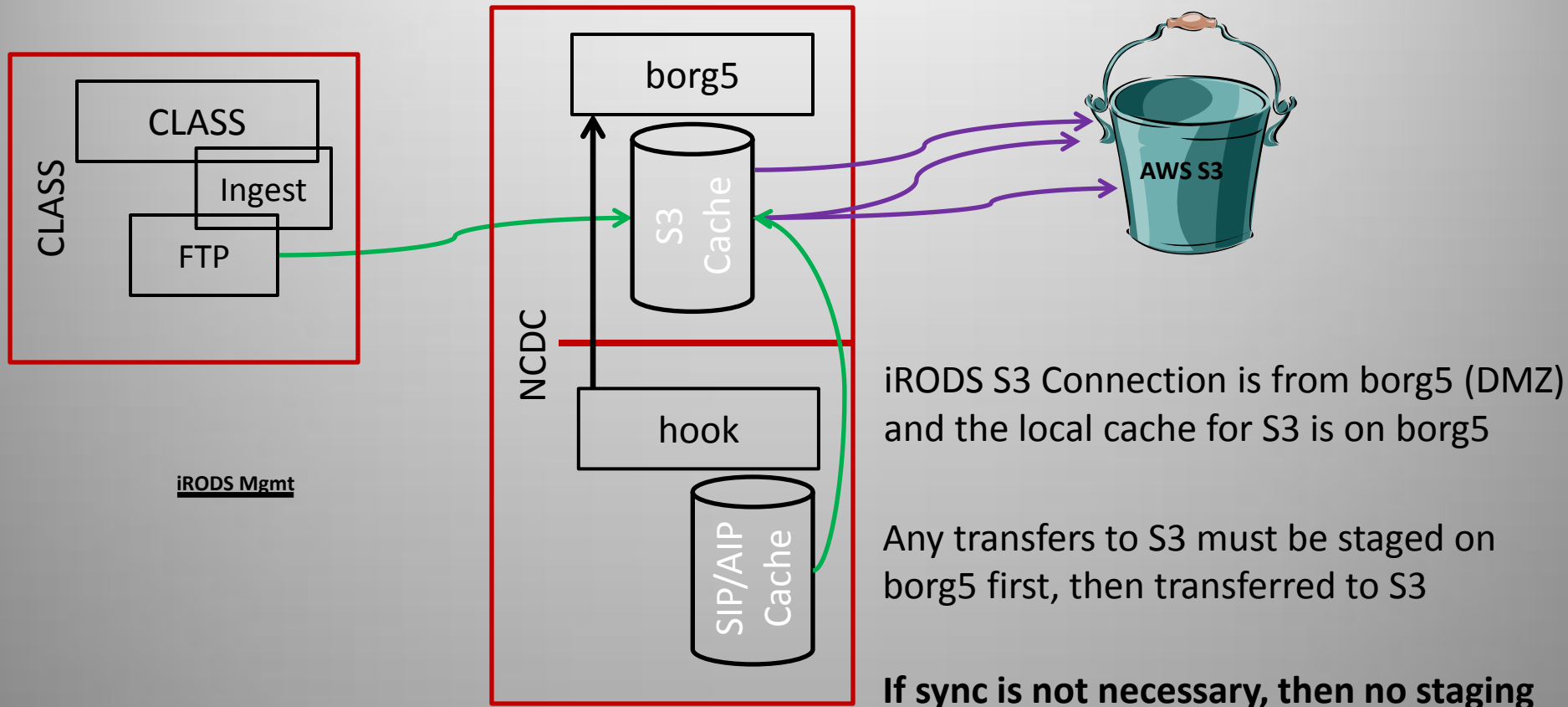
Cloud Pilot Access



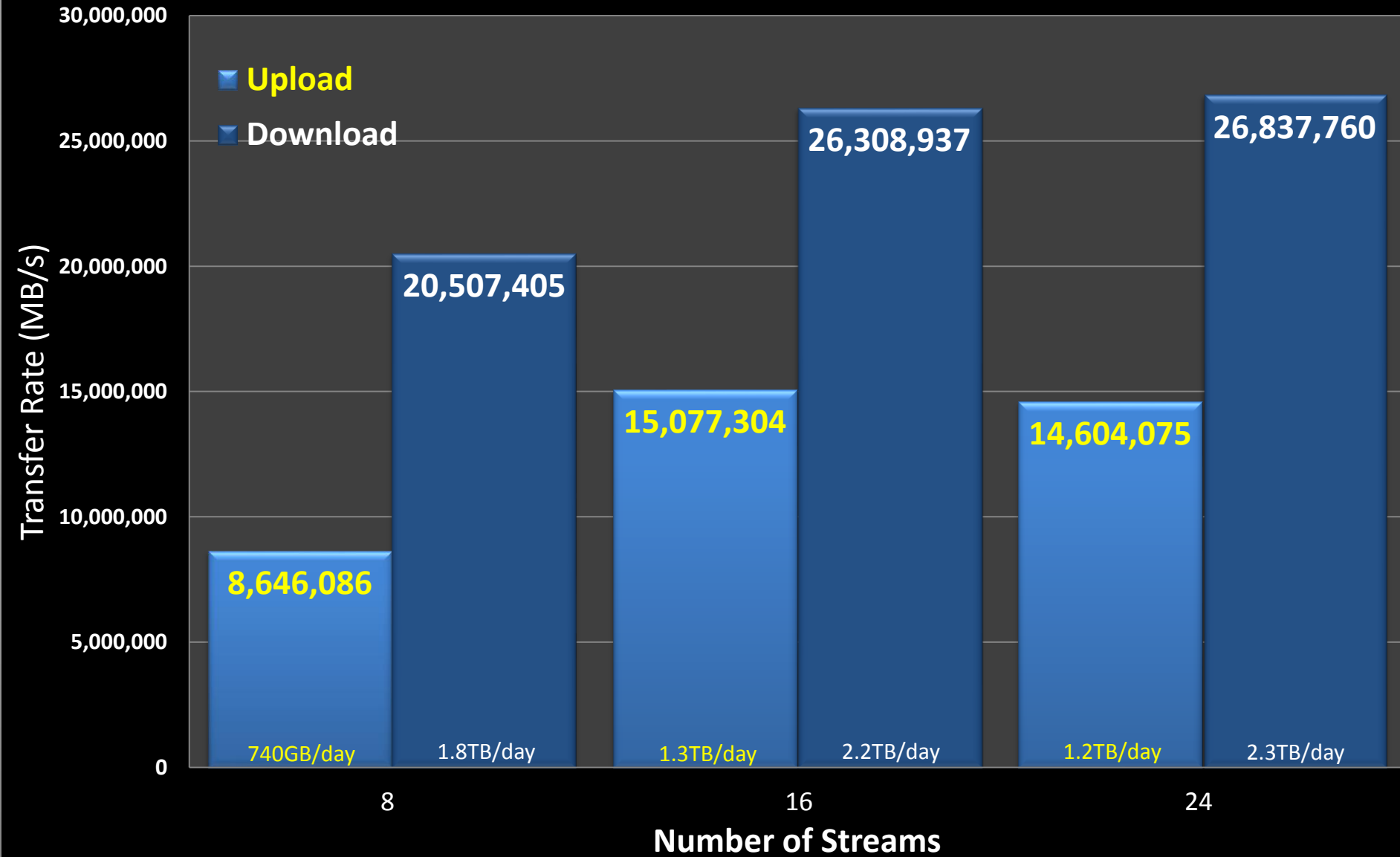
CLASS Cloud Pilot: NCDC



iRODS Compound Resource



S3 Upload/Download Transfer Rates Using iRODS



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

ConversationOUTReport

Src IP	Dst IP	Application	Port	Protocol	DSCP	Traffic
borg5.ncdc.noaa.gov	igor.class.ncdc.noaa.gov	TCP_App	*	TCP	000010	12.21 GB
borg5.ncdc.noaa.gov	s3-1.amazonaws.com	https	443	TCP	Default	2.36 GB
borg5.ncdc.noaa.gov	s3-1.amazonaws.com	https	443	TCP	Default	1.95 GB
borg5.ncdc.noaa.gov	s3-1.amazonaws.com	https	443	TCP	Default	1.83 GB
borg5.ncdc.noaa.gov	s3-1.amazonaws.com	https	443	TCP	Default	1.38 GB
borg5.ncdc.noaa.gov	s3-1.amazonaws.com	https	443	TCP	Default	1.06 GB
borg5.ncdc.noaa.gov	s3-1.amazonaws.com	https	443	TCP	Default	916.65 MB
borg5.ncdc.noaa.gov	s3-1.amazonaws.com	https	443	TCP	Default	113.92 MB
borg5.ncdc.noaa.gov	s3-1.amazonaws.com	https	443	TCP	Default	45.61 MB
borg5.ncdc.noaa.gov	ftp.class.ncdc.noaa.gov	ftp	21	TCP	Default	18.19 MB
borg5.ncdc.noaa.gov	ftp.class.ncdc.noaa.gov	TCP_App	*	TCP	Default	6.3 MB
borg5.ncdc.noaa.gov	igor.class.ncdc.noaa.gov	ftp	21	TCP	Default	2.05 MB
borg5.ncdc.noaa.gov	igor.class.ncdc.noaa.gov	TCP_App	*	TCP	Default	269.42 KB
borg5.ncdc.noaa.gov	s9a1.psmtp.com	smtp	25	TCP	Default	22.62 KB
borg5.ncdc.noaa.gov	time-c.timefreq.bldrdoc.gov	ntp	123	UDP	Default	1.29 KB
borg5.ncdc.noaa.gov	62-25-50-200-static.centennialpr.net	ntp	123	UDP	Default	1.29 KB
borg5.ncdc.noaa.gov	131.107.13.100	ntp	123	UDP	Default	1.29 KB
borg5.ncdc.noaa.gov	113.64.79.216.68.in-addr.arpa	ntp	123	UDP	Default	1.29 KB
borg5.ncdc.noaa.gov	128.138.188.172	ntp	123	UDP	Default	1.21 KB
borg5.ncdc.noaa.gov	time-b.nist.gov	ntp	123	UDP	Default	1.21 KB

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:
 - Must be linked to a local filesystem cache
- Security via Obscurity
 - iRODS Commands are run from hook, not borg5
 - When transferring to S3 from hook, borg5 makes the connection
- HTTPS Transfers
- iRODS Restful Download supported multi-part download (threads)
 - Not on upload because iRODS is not installed on the S3 Resource
- Our processes do NOT have to be tightly coupled
 - 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:

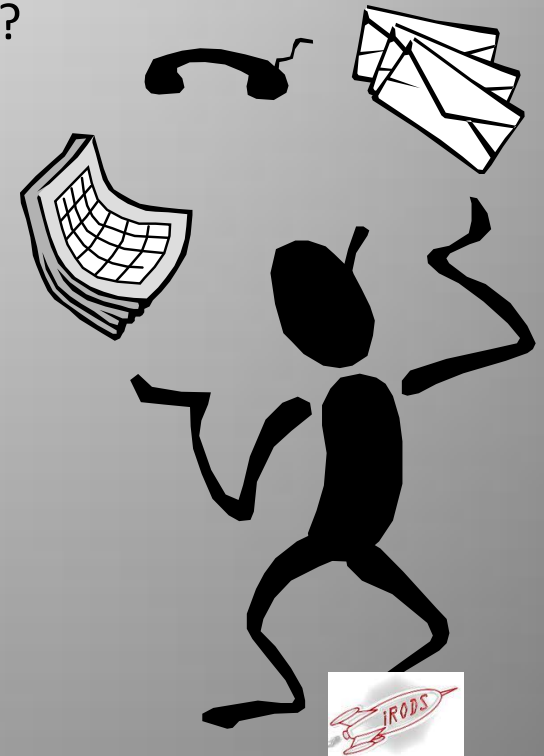
- Restful APIs did not implement basic list function
 - Need listings to know what is actually on the resource
- No way to force Reduced Redundancy on upload thru iRODS
- iRODS registration of files on compound resource difficult

iRODS Developers are working on these



Challenges

- To utilize Cloud Storage
 - 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)?
 - 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/
22:25:08: SysCmd: Cmd: 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:
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: ArcSB Remove: rm /raid/gtsnp/tmp/iRODSirsyncTh_MultiStream.2242
22:25:08: ----- iRODSirsyncTh hook MultiStream END --- ET: 00:05 -----
22:25:08:
22:25:03:
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/
22:25:09: SysCmd: Cmd: 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:
22:25:09: iRODSirsyncTh: hook From: i:/DMZ/Archive/NR2/KGSP To: i:/DMZ/NWS-WxRadar-2/KGSP/ ET: 6 Err: none ET: 00:06
22:25:09: ArcSB Remove: rm /raid/gtsnp/tmp/iRODSirsyncTh_MultiStream.2093
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: irsync -v -r -R s3Resc i:/DMZ/Archive/NR2/KGSP i:/DMZ/NWS-WxRadar-2/KGSP/
00:25:42: SysCmd: Cmd: irsync -v -r -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:
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: irsync -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:
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 -----
```

