

Status & Plan of the Xrootd Federation

Wei Yang

Again, why do we do this

A data federation allows accessing data at other locations

- Opens many possibilities and use cases (need to understand/exploit pros and cons)
- See Rob Gardner's slides at SW&C meeting last week:

<https://indico.cern.ch/getFile.py/access?contribId=40&sessionId=9&resId=0&materialId=slides&confId=169695>

Where are we in R&D & Deployment

- Identified site architectures
- Deployed a prototype federation with basic monitoring in the US
- Global name space issue addressed
- X509 implementing, deployed at SLAC
- Studying WAN reading performance using prun and TTreeCache
- Git repository to share configurations, knowledge
- Bi-weekly working group meetings

Site Architectures

Six possible site architectures

<https://www.usatlas.bnl.gov/twiki/pub/Admins/MinutesFedXrootdFeb8/FAX-site-diagram.pptx>

Work with backend storage systems: dCache, Xrootd, Posix

Xrootd proxy based architecture works for all sites

- Deployed at AGLT2, BNL, UTA, OU, SLAC

Overlapping xrootd on dCache

- Deployed at UC

Xrootd with libXrdDcap.so (deprecated)

- Deployed at UIUC

Working on direct access to dCache xrootd door

- Need N2N via authorization plugin

Deployment, Dashboard

<http://uct3-xrdp.uchicago.edu:8080/rsv/>



USATLAS Federated Xrootd Status - 2012-03-18 22:59:43

Frequently Asked Questions

Host: atl-prod09.slac.stanford.edu (atl-prod09.slac.stanford.edu)

Metric	Last Executed	Enabled?	Next Run Time	Status
org.usatlas.xrootd.grid-xrdcp-compare	2012-03-18 22:50:02 CDT	YES	2012-03-18 23:05:00 CDT	OK
org.usatlas.xrootd.grid-xrdcp-direct	2012-03-18 22:50:03 CDT	YES	2012-03-18 23:05:00 CDT	OK
org.usatlas.xrootd.grid-xrdcp-fax	2012-03-18 22:50:02 CDT	YES	2012-03-18 23:05:00 CDT	OK
org.usatlas.xrootd.ping	2012-03-18 22:50:01 CDT	YES	2012-03-18 23:05:00 CDT	OK

Host: atlas29.hep.anl.gov (atlas29.hep.anl.gov)

Metric	Last Executed	Enabled?	Next Run Time	Status
org.usatlas.xrootd.ping	2012-03-18 22:50:02 CDT	YES	2012-03-18 23:05:00 CDT	OK
org.usatlas.xrootd.xrdcp-compare	2012-03-18 22:50:01 CDT	YES	2012-03-18 23:05:00 CDT	CRITICAL
org.usatlas.xrootd.xrdcp-direct	2012-03-18 22:50:02 CDT	YES	2012-03-18 23:05:00 CDT	CRITICAL
org.usatlas.xrootd.xrdcp-fax	2012-03-18 22:50:03 CDT	YES	2012-03-18 23:05:00 CDT	CRITICAL

Host: atlgridftp01.phy.duke.edu (atlgridftp01.phy.duke.edu)

Metric	Last Executed	Enabled?	Next Run Time	Status
org.usatlas.xrootd.ping	2012-03-18 22:50:02 CDT	YES	2012-03-18 23:05:00 CDT	OK
org.usatlas.xrootd.xrdcp-compare	2012-03-18 22:50:02 CDT	YES	2012-03-18 23:05:00 CDT	OK
org.usatlas.xrootd.xrdcp-direct	2012-03-18 22:50:01 CDT	YES	2012-03-18 23:05:00 CDT	OK
org.usatlas.xrootd.xrdcp-fax	2012-03-18 22:50:02 CDT	YES	2012-03-18 23:05:00 CDT	OK

Host: dcdoor09.usatlas.bnl.gov (dcdoor09.usatlas.bnl.gov)

Metric	Last Executed	Enabled?	Next Run Time	Status
org.usatlas.xrootd.ping	2012-03-18 22:50:01 CDT	YES	2012-03-18 23:05:00 CDT	OK
org.usatlas.xrootd.xrdcp-compare	2012-03-18 22:50:03 CDT	YES	2012-03-18 23:05:00 CDT	OK
org.usatlas.xrootd.xrdcp-direct	2012-03-18 22:50:01 CDT	YES	2012-03-18 23:05:00 CDT	OK
org.usatlas.xrootd.xrdcp-fax	2012-03-18 22:50:02 CDT	YES	2012-03-18 23:05:00 CDT	OK

Deployment, CMS-like monitoring

- Courtesy Matevz Tadel UCSD
- Detailed xrootd monitoring information sent to collector
- Tracks files (global names) in use, when opened, server, client, MB read
- Provides IO visibility in federation

OpenAgo	ServerDomain	ClientDomain	User	Read [MB]	UpdateAgo
02:38:26	slac.stanford.edu	23.40.189		347.835	00:00:05
00:04:05	slac.stanford.edu	uchicago.edu		11.003	00:02:54
00:04:05	slac.stanford.edu	uchicago.edu		11.008	00:02:53
00:02:53	slac.stanford.edu	uchicago.edu		11.126	00:01:48
00:02:53	slac.stanford.edu	uchicago.edu		11.020	00:01:51
00:01:49	slac.stanford.edu	uchicago.edu		10.982	00:00:46
00:01:49	slac.stanford.edu	uchicago.edu		11.125	00:00:48
00:01:49	slac.stanford.edu	uchicago.edu		11.125	00:00:49
00:01:49	slac.stanford.edu	uchicago.edu		11.125	00:00:48
					00:00:6
					00:00:7
					00:00:7
					00:00:4

```
File
/atlas/xrootd/atlasdatadisk/data11_7TeV/AOD/r260_p659_tid493619_00/AOD.493619_000001.pool.root.1
/atlas/dq2/mc10_7TeV/NTUP_SMWZ/e773_s933_5_WW2lep.merge.NTUP_SMWZ.e773_s933_s946_r2302_r2300_p591_tid408566_00/NTUP_SMWZ.408566_000001.root.1
/atlas/dq2/mc10_7TeV/NTUP_SMWZ/e773_s933_5_WW2lep.merge.NTUP_SMWZ.e773_s933_s946_r2302_r2300_p591_tid408566_00/NTUP_SMWZ.408566_000001.root.1
/atlas/dq2/mc10_7TeV/NTUP_SMWZ/e773_s933_5_WW2lep.merge.NTUP_SMWZ.e773_s933_s946_r2302_r2300_p591_tid408566_00/NTUP_SMWZ.408566_000002.root.1
/atlas/dq2/mc10_7TeV/NTUP_SMWZ/e773_s933_5_WW2lep.merge.NTUP_SMWZ.e773_s933_s946_r2302_r2300_p591_tid408566_00/NTUP_SMWZ.408566_000002.root.1
/atlas/dq2/mc10_7TeV/NTUP_SMWZ/e773_s933_5_WW2lep.merge.NTUP_SMWZ.e773_s933_s946_r2302_r2300_p591_tid408566_00/NTUP_SMWZ.408566_000003.root.1
/atlas/dq2/mc10_7TeV/NTUP_SMWZ/e773_s933_5_WW2lep.merge.NTUP_SMWZ.e773_s933_s946_r2302_r2300_p591_tid408566_00/NTUP_SMWZ.408566_000003.root.1
/atlas/dq2/mc10_7TeV/NTUP_SMWZ/e773_s933_5_WW2lep.merge.NTUP_SMWZ.e773_s933_s946_r2302_r2300_p591_tid408566_00/NTUP_SMWZ.408566_000003.root.1
/atlas/dq2/mc10_7TeV/NTUP_SMWZ/e773_s933_5_WW2lep.merge.NTUP_SMWZ.e773_s933_s946_r2302_r2300_p591_tid408566_00/NTUP_SMWZ.408566_000004.root.1
/atlas/dq2/mc10_7TeV/NTUP_SMWZ/e773_s933_5_WW2lep.merge.NTUP_SMWZ.e773_s933_s946_r2302_r2300_p591_tid408566_00/NTUP_SMWZ.408566_000004.root.1
/atlas/dq2/mc10_7TeV/NTUP_SMWZ/e773_s933_5_WW2lep.merge.NTUP_SMWZ.e773_s933_s946_r2302_r2300_p591_tid408566_00/NTUP_SMWZ.408566_000004.root.1
/atlas/dq2/mc10_7TeV/NTUP_SMWZ/e773_s933_5_WW2lep.merge.NTUP_SMWZ.e773_s933_s946_r2302_r2300_p591_tid408566_00/NTUP_SMWZ.408566_000004.root.1
```

R&D, Technical Stuffs

Global Name Space is addressed

- Global File Name to Storage File Name (N2N) translation ready
- N2N via a local LFC introduces small latency (3ms?)
- dq2-client 1.0 supports GFN

X509 work in progress

- Implemented and working. Fixing a security hole

Prun jobs run successfully against the federation

- Using prun --pfnList to supply a list of input files in GFN

Git repository at BNL

Next Steps

Continue Deployment

- More sites with Xrootd 3.1.1, X509, etc.
- Can we do this in a month or two?

Latency study and TTreeCache

- Probably need to get a sense of worst cases first
- Talk to experts

Panda Analysis, *for example:*

- Have pilot replacing local missing files with GFN
- Take advantage of regional redirectors
- Local caching
- **Or can we help ourselves?** Note:

```
lsm-get [-t token] [--size N] [--checksum csum] [--guid guid] source destination
```

Next Steps, cont'd

Global redirector evaluation

- Reliability and performance
- explore the concept of regional redirector

Monitoring

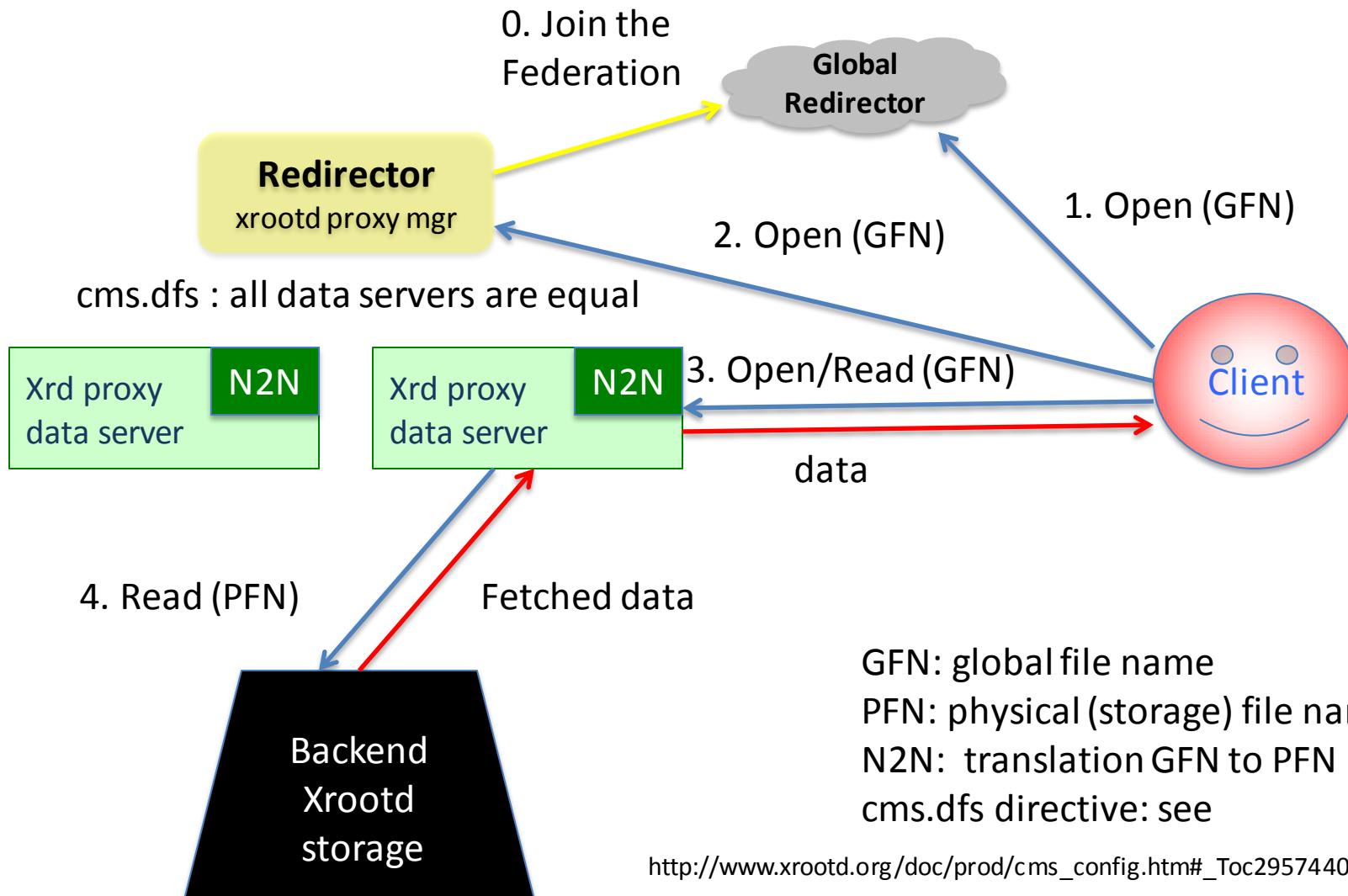
- Leaning toward DB-less detailed stream monitoring
- CMS-like monitoring: Know what is happening, needs aggregation and history
- T3mon

Bug fixes

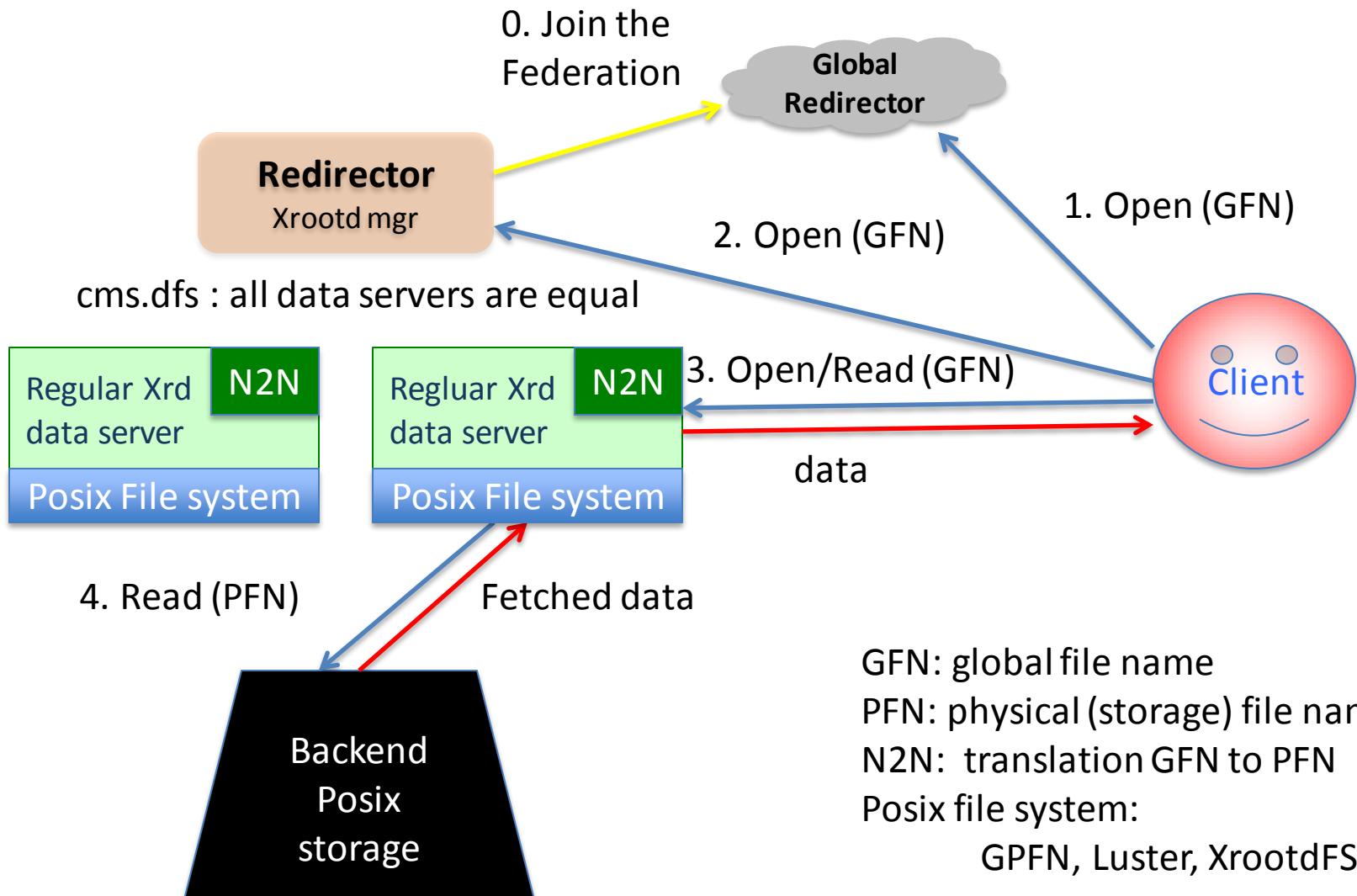
- Extreme copy mode in Xrdcp
- Patch X509 security issue

Backup Slides

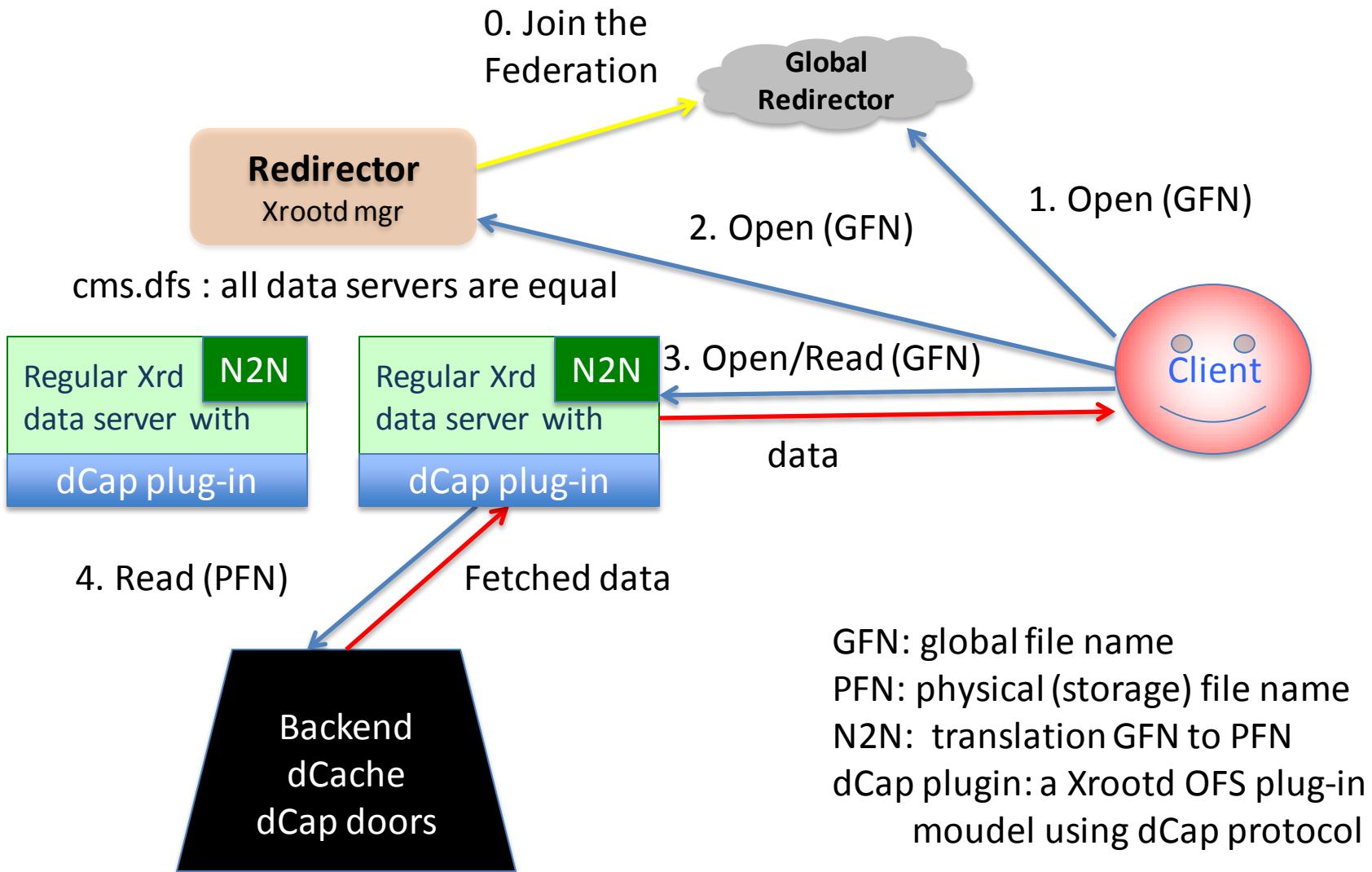
Export Xrootd storage via Xrootd Proxy cluster



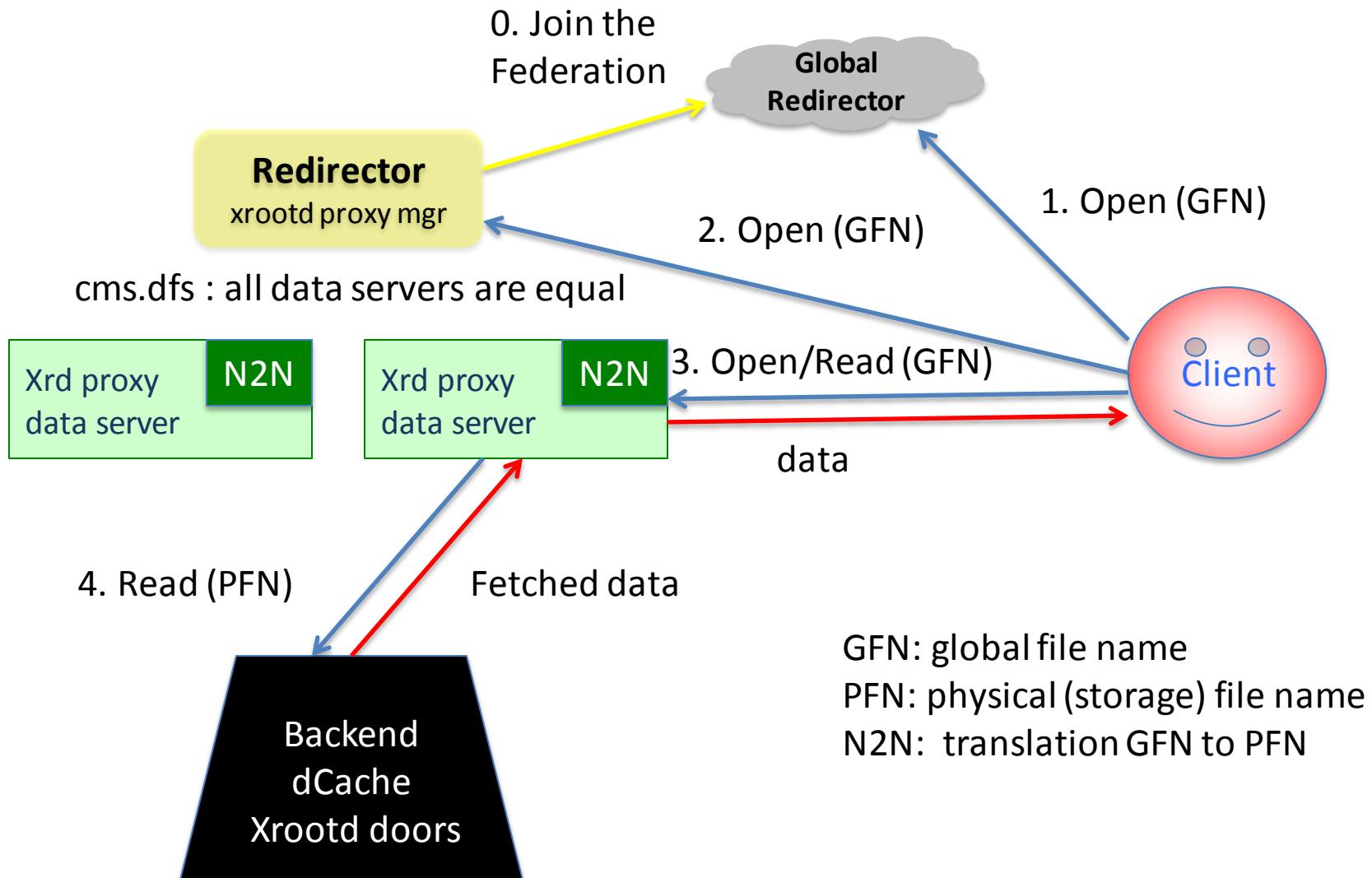
Export Posix storage via regular Xrootd cluster



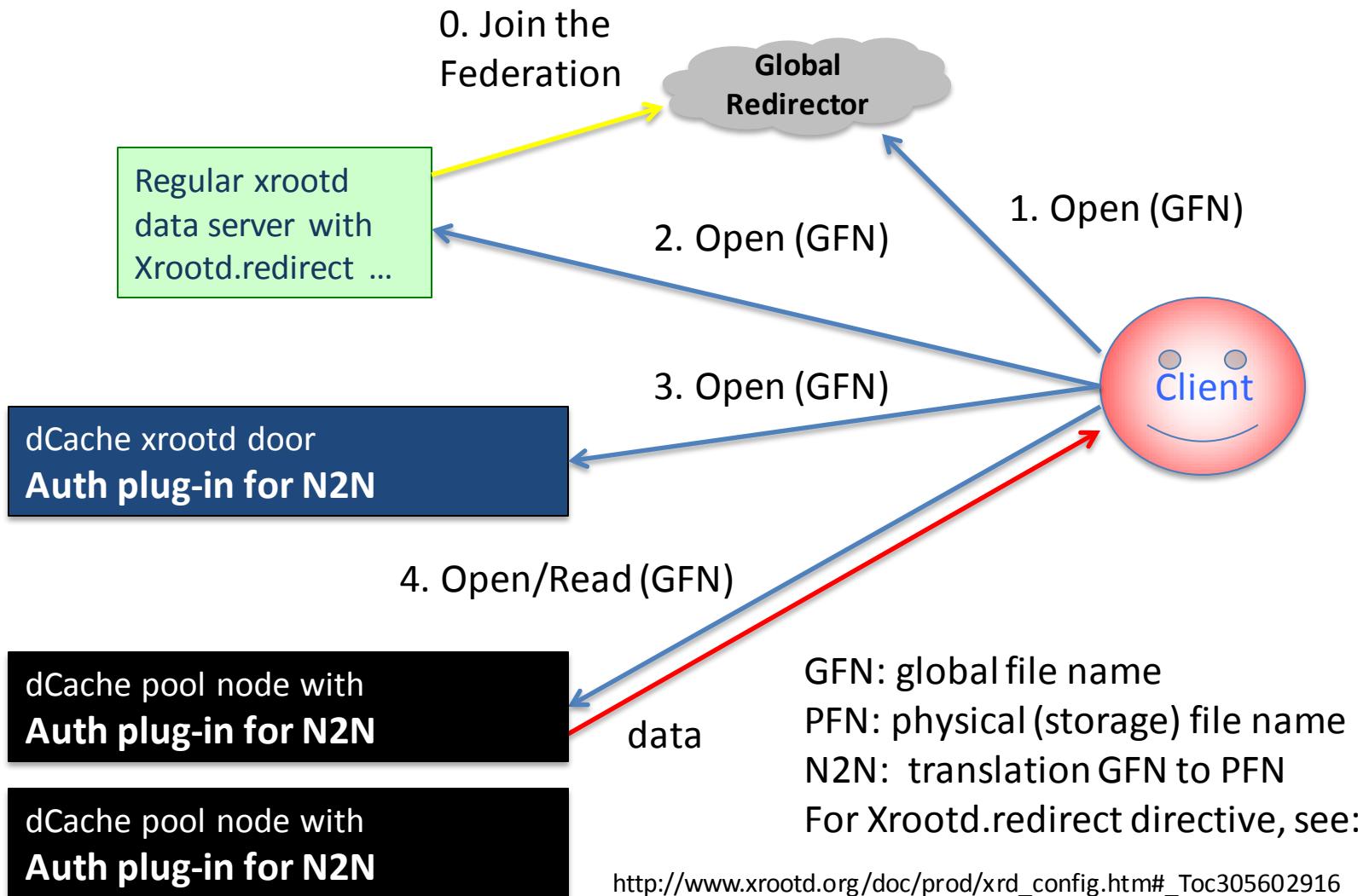
Export dCache dCap doors via regular Xrootd cluster (deprecated)



Export dCache Xrootd doors via Xrootd Proxy cluster, 1



Export dCache Xrootd doors via Xrootd Proxy cluster, 2



Overlapping Xrootd cluster on top of dCache

