

GT

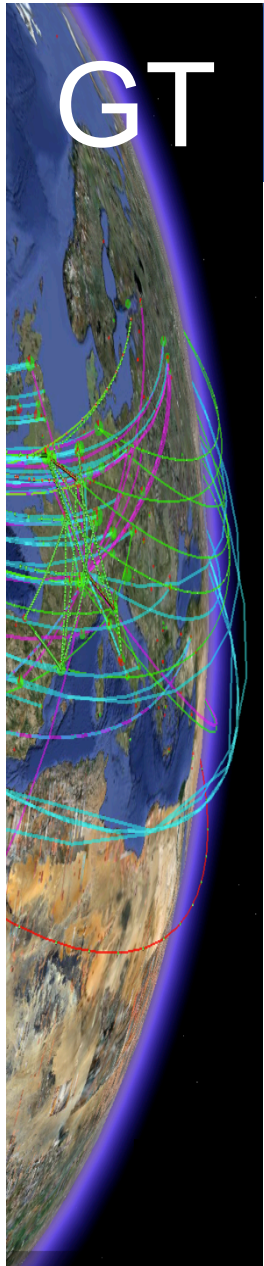
Grid Technology

CERN IT
Department

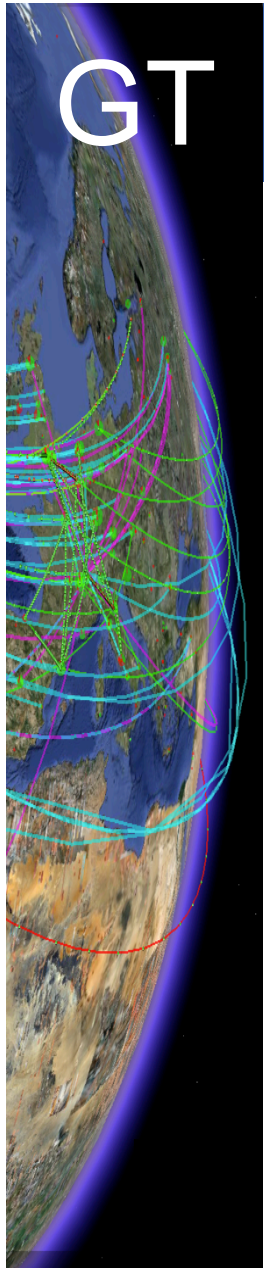
dpm-xrootd v3

Creating Federated Data Stores for the LHC
David Smith, on behalf of IT-GT, CERN

14 Sep 2012

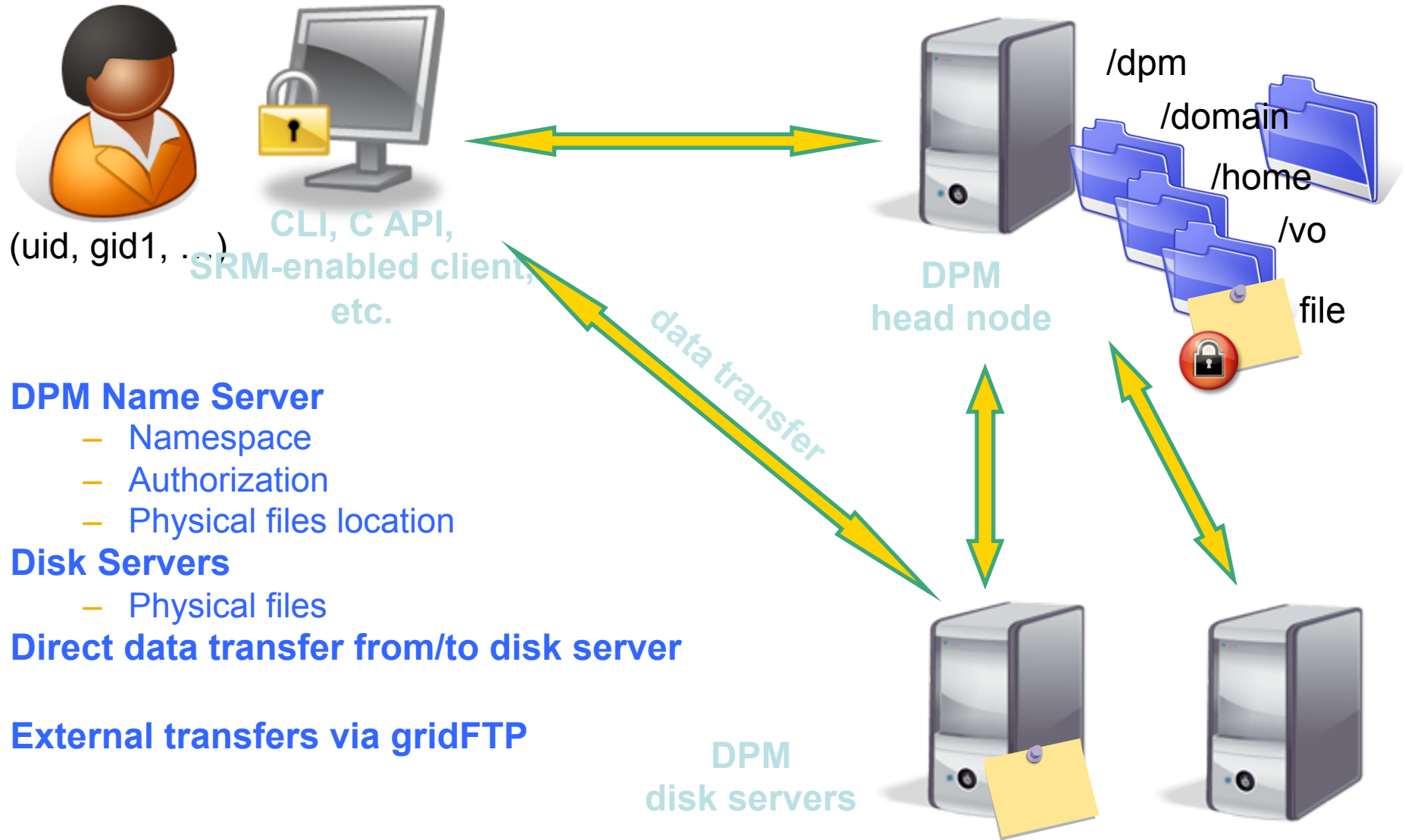


- An existing dpm-xrootd written 2006
 - A pair of plugins for OFS, XMI
 - Only ALICE token based access
 - Some performance issues with dispatch of requests to dpm



Aim to:

- Provide xroot file I/O for DPM for all VOs
- Allow participation in xrootd federations
- Allow ALICE token based access
- Other VOs to use GSI for authentication
 - Support VOMS extensions
 - Authorization is done by DPM system use of an identity (user, [/vo,/vo/group,...]).



DPM Name Server

- Namespace
- Authorization
- Physical files location

Disk Servers

- Physical files

Direct data transfer from/to disk server

External transfers via gridFTP

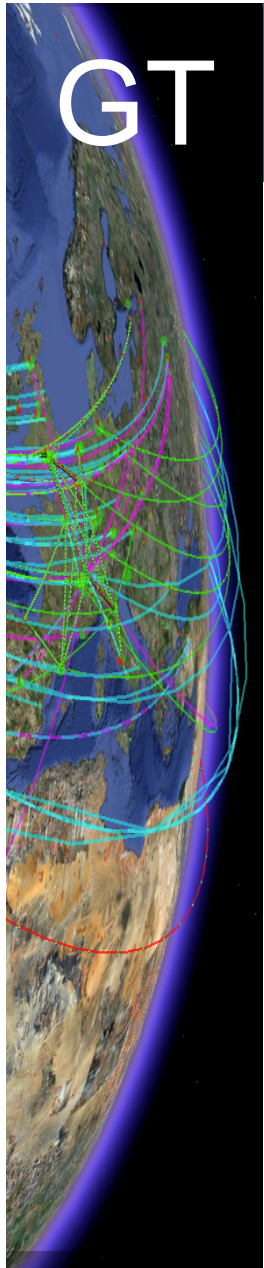
- DPM has a central service
 - A **get** request for a SURL (/dpm/example.com/home/dteam/file) gives an SFN (disk001.example.com:/data1/dteam/2012-09-14/file.1353.0) to read
 - A **put** request for a SURL returns an SFN to write to
 - Put concluded with **put_done**
 - Authorization check during get or put
 - Files can not be modified once written
 - DPM instance is often run for multiple VOs
- Disk servers providing I/O
 - Clients have connectivity to disk servers

Now:

- dpm-xrootd v3 is a set of plugins to xrootd server, using plugin interfaces (as for previous version)
- But different interfaces:
 - Use XrdOss, XrdCmsClient and XrdAccAuthorize plugins
- VOMS extraction disabled for now

GT

Example: file access



GT

dpm-xrootd basic file access

CERN IT
Department



xrootd:1095
(disk server)



xrootd:1095
(disk server)



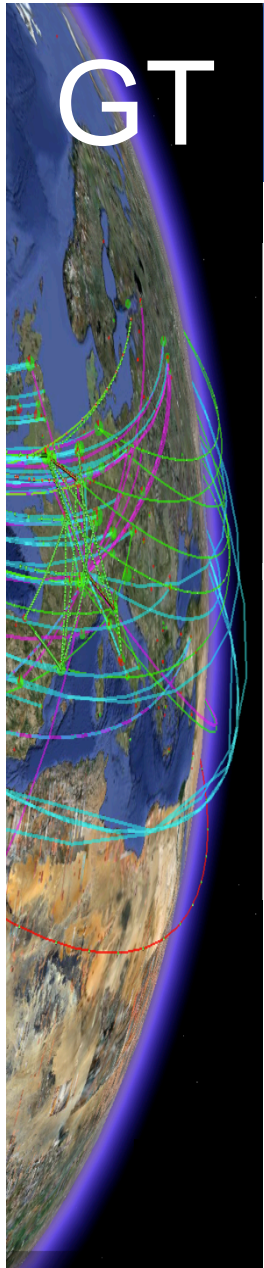
xrootd
(redirector)

DPM service

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

dpm-xrootd - 8





xrootd:1095
(disk server)



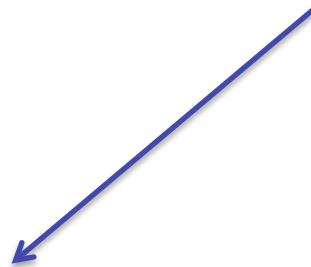
xrootd:1095
(disk server)

xroot client: open

<xroot://dpm.example.com/dpm/example.com/dteam/file>



xrootd
(redirector)

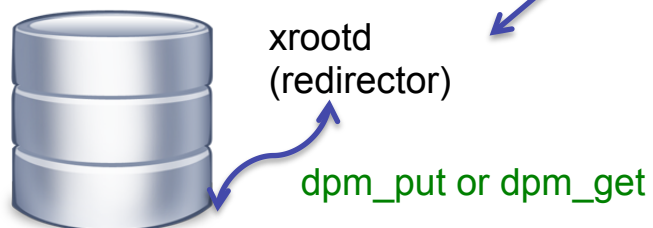


DPM service

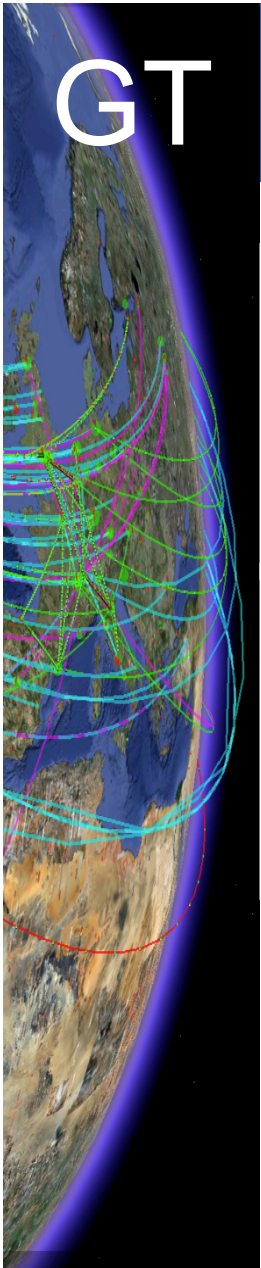


xroot client: open

<xroot://dpm.example.com/dpm/example.com/dteam/file>



DPM service



xrootd:1095
(disk server)



xrootd:1095
(disk server)



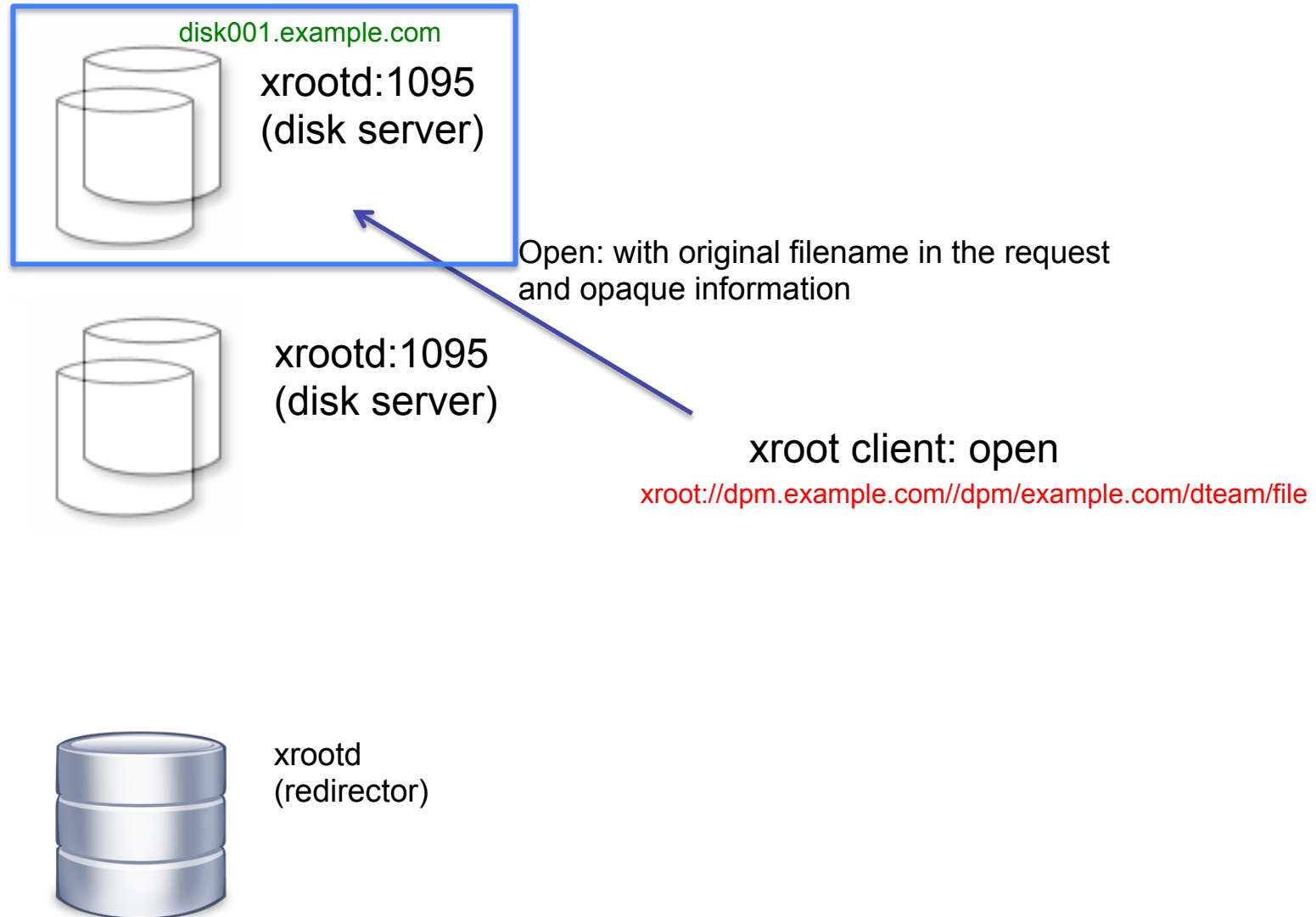
xrootd
(redirector)

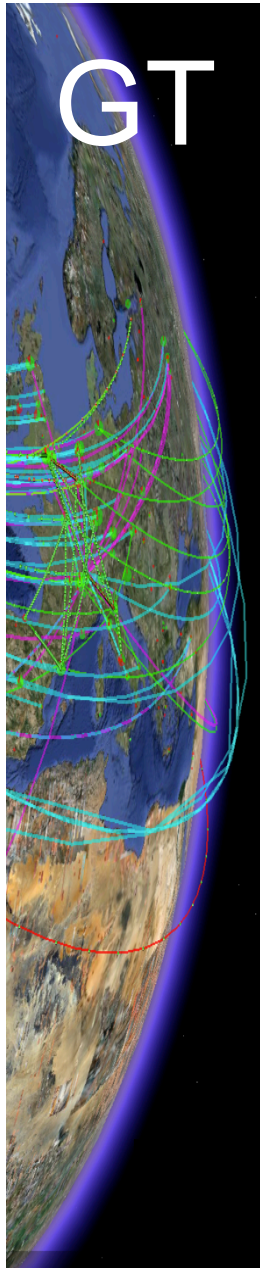
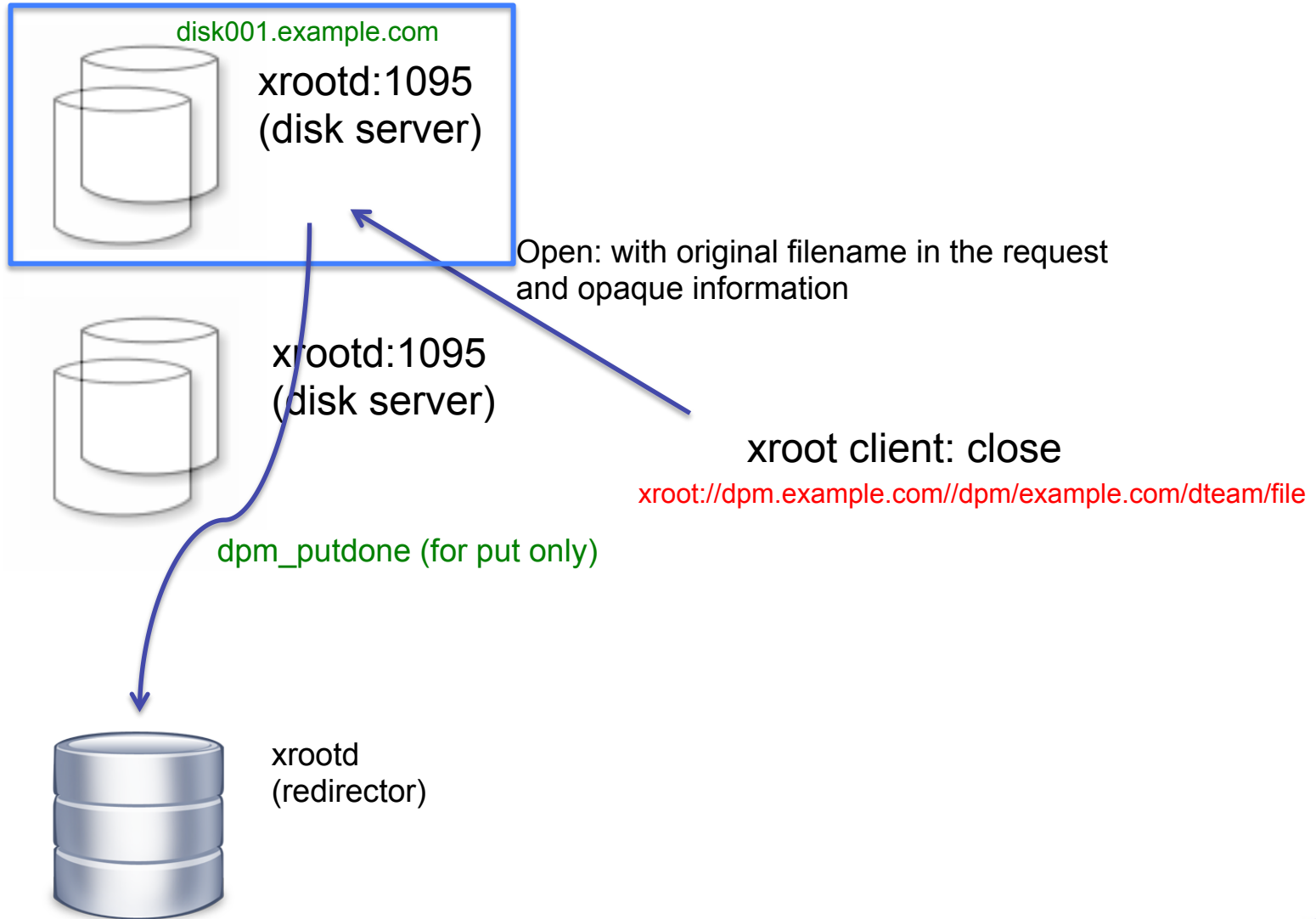
xroot client: open

`xroot://dpm.example.com/dpm/example.com/dteam/file`

Redirect: with host/port and opaque info
`&dpm.sfn=/data1/dteam/2012-09-14/file.
1353.0`

DPM service





- Redirection xrootd interacts with DPM
 - Configured to offer XrdSecunix and XrdSecgsi to the xroot client
 - If a valid ALICE token is present in the opaque data from the client the identity is a preset. (Check if **authz** or **signature** in opaque data)
 - Otherwise an identity must be derivable from the XrdSecEntity authentication data (at disk too)
 - Goes through **get** or **put** sequence with DPM system (currently using classic dpns/dpm api)
 - May need to return a wait time to the client if the **get** or **put** sequence takes more than ~1 second

- If get or put is successful, redirect is returned directing to the target disk server.
- Opaque data is added:

Name	Included when	Encoded (&, ?, space)	Description
dpm.time	Always	No	UNIX time and optional validity duration
dpm.dn	Only for preset identity	Yes	DPM Username in case this data is not derived from XrdSecEntity
dpm.voms	Only for preset identity if group data available	Yes	DPM Group data if available and is not derived from XrdSecEntity
dpm.dhost	Always	No	Target disk host name
dpm.surl	For put* if different from xrd filename	Yes	DPM file name
dpm.sfn	For get or put*	Yes	File name on disk server
dpm.tk	For put*	No	A DPM identifier for the request
dpm.nonce	If no dpm.tk	No	Equivalent of 48 random bits
dpm.put	For put	No	Indicates a put request
dpm.hv1	Always	No	Keyed hash of various parameters (base64)

- Disk servers require keyed hash (with limited duration) in opaque data
- Use HMAC-SHA256-128 for keyed hash
- Key stored in disk and memory of xrootd processes (required to be 32 to 64 bytes)

Hash over string (and terminating nul) Notes

File name in xrd request

dpm.surl

Decoded form

dpm.dhost

dpm.sfn

Decoded form

dpm.tk

Decimal number 'flags'

Bitmapped flags indicating identity source and put/get

DN

From dpm.dn or XrdSecEntity

Group info

From dpm.voms or XrdSecEntity

dpm.time

Binary representation is converted to string

dpm.nonce

- Hash and `dpm.dhost` checked
- `dpm.time` must be within validity window (300 seconds default) of current time
- Access mode of request must be consistent with `dpm.put`
- Then disk I/O via native XrdOss
- `ofs.persist auto hold 0` is set, to remove files not closed at the end of writing

- For Put
 - On success call `dpm_putdone` with dpm token and surl.
 - On fail (no close) call `dpm_abortfiles` with dpm token and surl.
- For Get
 - No interactions with central DPM

FEDERATION of XROOT access

- With the basic xroot file access vo/users could devise federations.. but
- Want to provide integrated method (as for a native xroot site)
- Next: a couple of example sequences



xrootd:1095
(disk server)

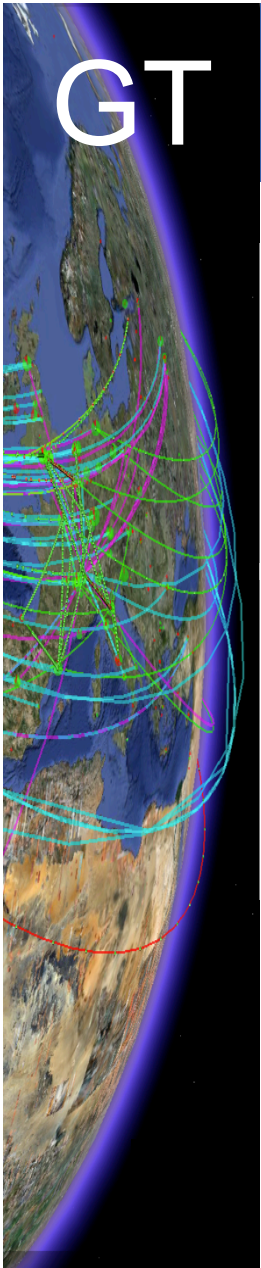


xrootd:1095
(disk server)

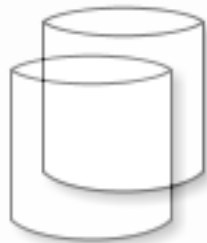


xrootd
(redirector)

DPM service



xrootd:1095
(disk server)



xrootd:1095
(disk server)

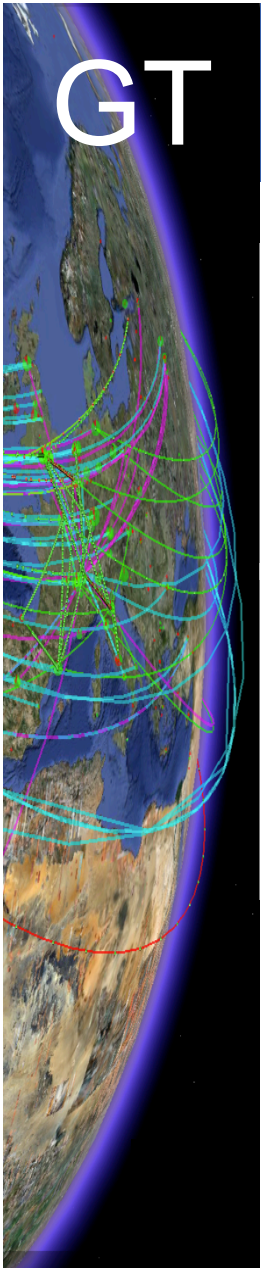


xrootd
(redirector)

xrootd:11000
(fedredir_vo)

cmsd
(fedredir_vo)

DPM service



xrootd:1095
(disk server)



xrootd:1095
(disk server)



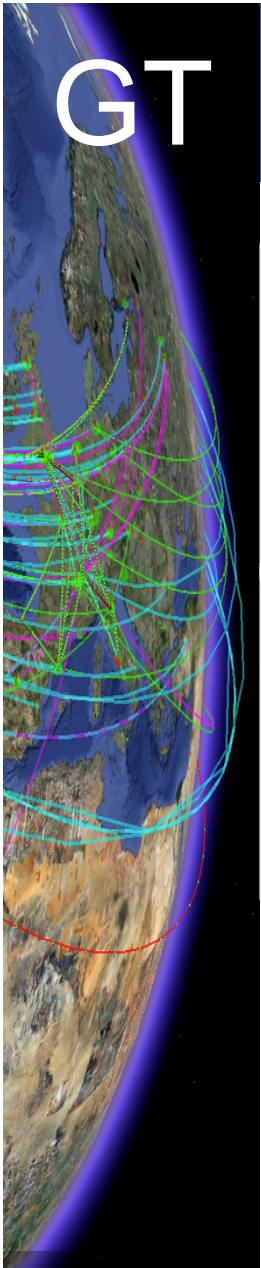
DPM service

xrootd
(redirector)

xrootd:11000
(fedredir_vo)



cmsd
(fedredir_vo)



xrootd:1095
(disk server)



xrootd:1095
(disk server)



DPM service

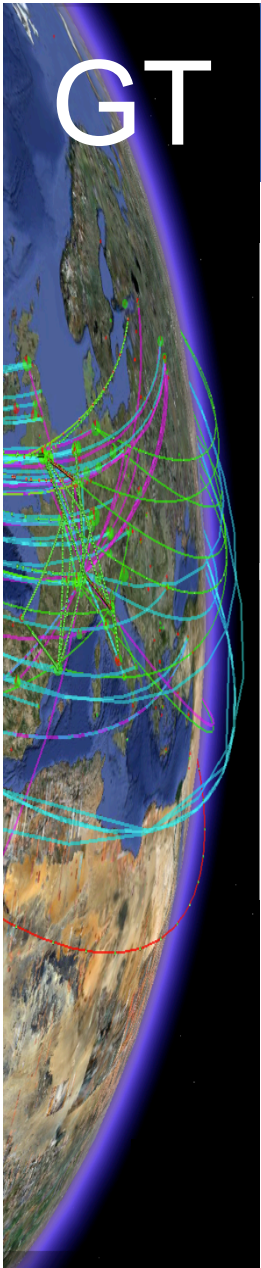
xrootd
(redirector)

xrootd:11000
(fedredir_vo)

cmsd
(fedredir_vo)



regional
redirector



xrootd:1095
(disk server)



xrootd:1095
(disk server)



xrootd
(redirector)

xrootd:11000
(fedredir_vo)

cmsd
(fedredir_vo)



regional
redirector

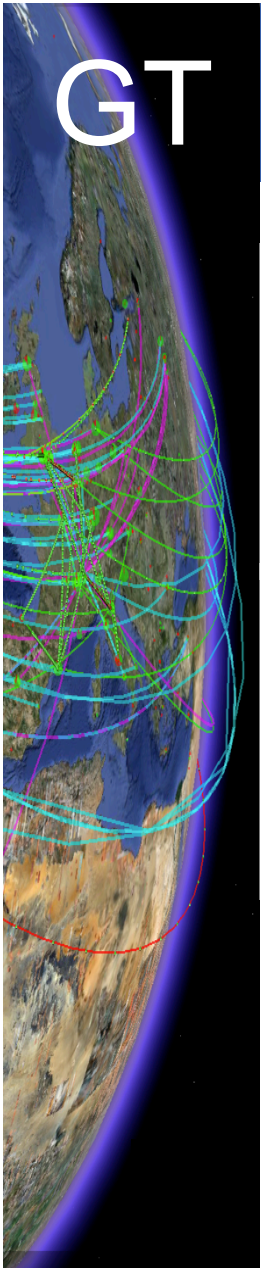
xroot client: open

`xroot://regional.example.org/vo/example.dat`

`/vo/example.dat ?`

DPM service





xrootd:1095
(disk server)



xrootd:1095
(disk server)



DPM service

xrootd (redirector) xrootd:11000 (fedredir_vo)

cmsd (fedredir_vo)

XrdPss: stat /vo/example.dat
(specially trusted at xrootd)

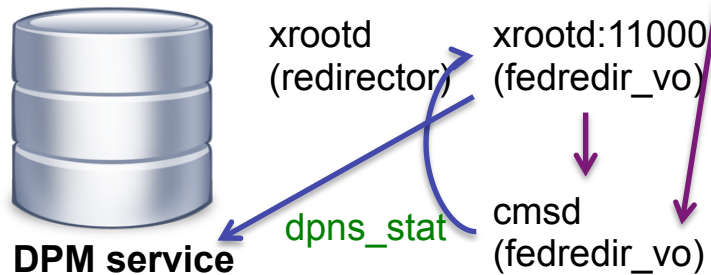
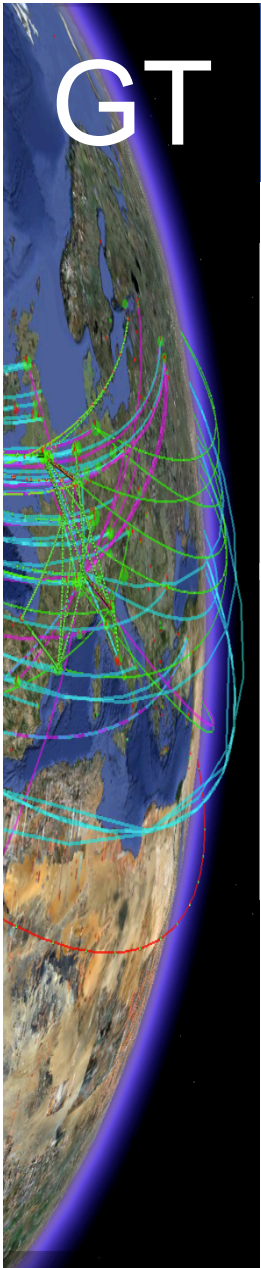


regional
redirector

xroot client: open

xroot://regional.example.org/vo/example.dat

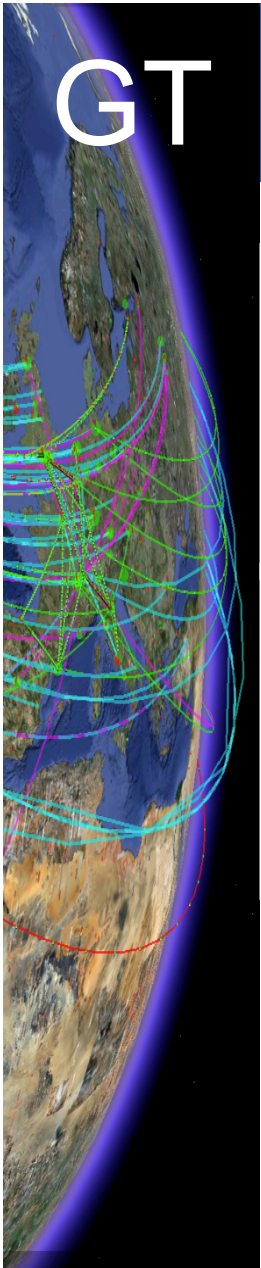
/vo/example.dat ?



xroot client: open

xroot://regional.example.org/vo/example.dat

/vo/example.dat ?



xrootd:1095
(disk server)



xrootd:1095
(disk server)



DPM service

xrootd
(redirector)

xrootd:11000
(fedredir_vo)

cmsd
(fedredir_vo)



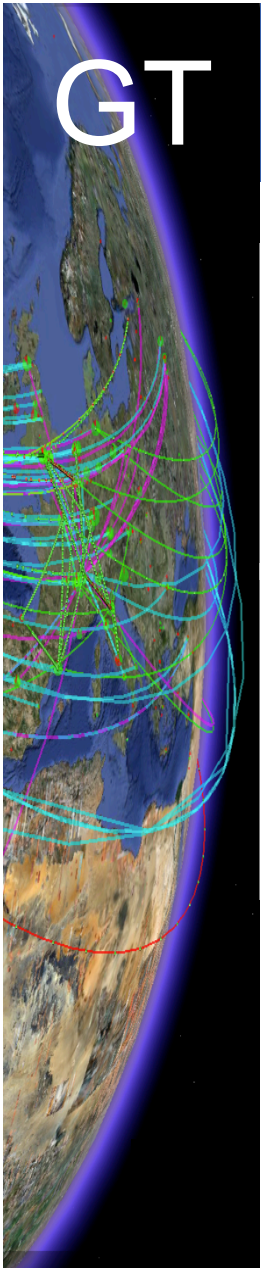
regional
redirector

...if site has example.dat
client may be directed there

Redirect to dpm.example.com:11000

xroot client: open

`xroot://regional.example.org/vo/example.dat`



xrootd:1095
(disk server)



xrootd:1095
(disk server)



xrootd
(redirector)

xrootd:11000
(fedredir_vo)

cmsd
(fedredir_vo)

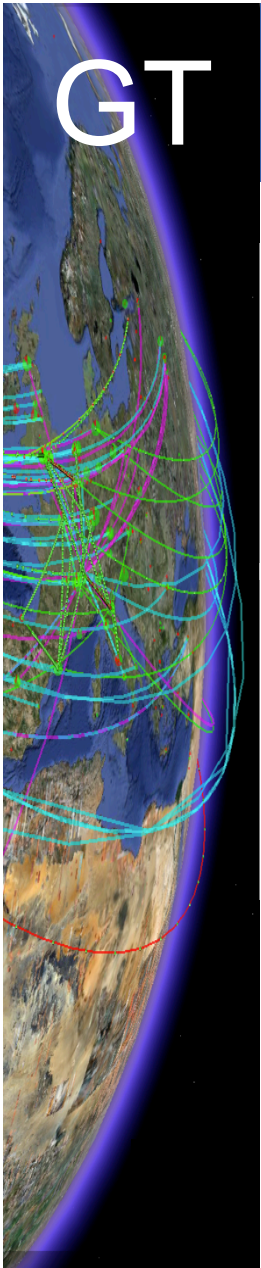


regional
redirector

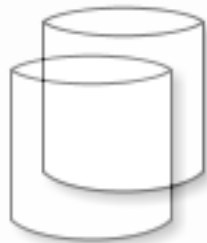
xroot client: open

`xroot://regional.example.org/vo/example.dat`

DPM service



xrootd:1095
(disk server)



xrootd:1095
(disk server)



DPM service

Not Involved

~~xrootd
(redirector)~~

xrootd:11000
(fedredir_vo)

cmsd
(fedredir_vo)



regional
redirector

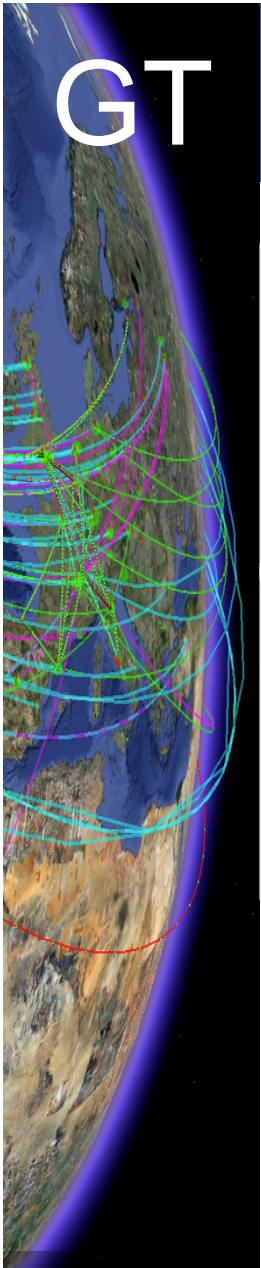
xroot client: open

`xroot://regional.example.org/vo/example.dat`

GT

Client example 2





xrootd:1095
(disk server)



xrootd:1095
(disk server)



DPM service

xrootd
(redirector)

xrootd:11000
(fedredir_vo)

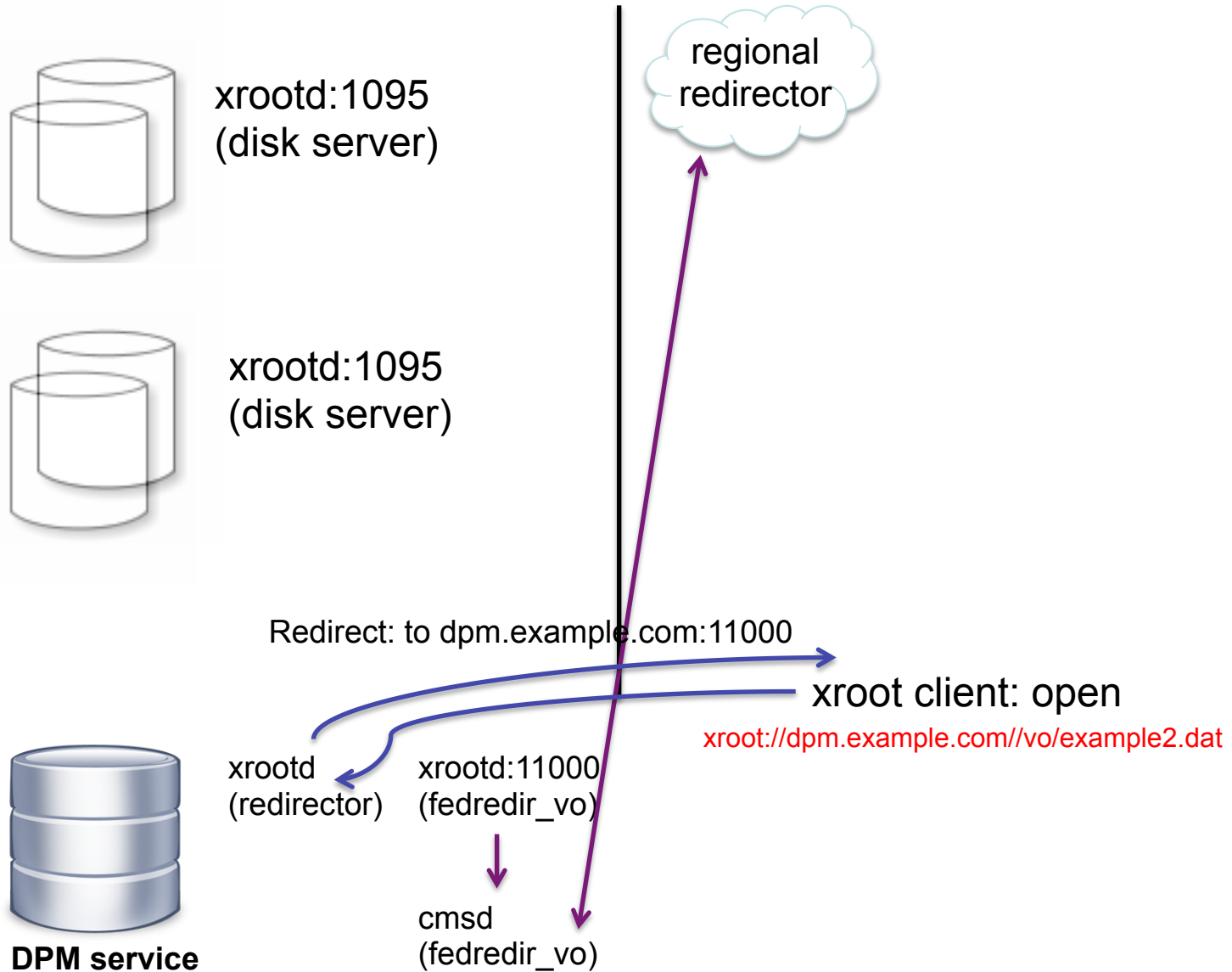
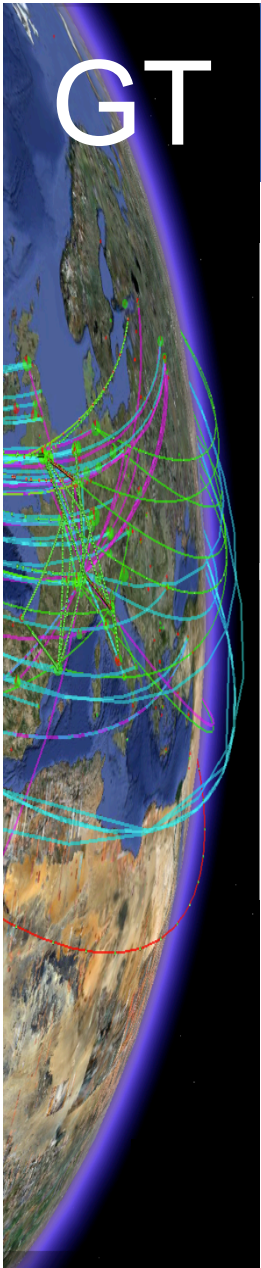
cmsd
(fedredir_vo)

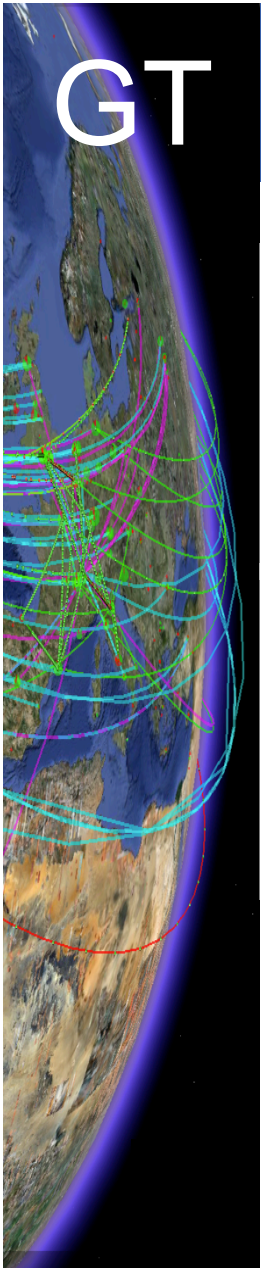


regional
redirector

xroot client: open

`xroot://dpm.example.com//vo/example2.dat`





xrootd:1095
(disk server)



xrootd:1095
(disk server)



DPM service

xrootd
(redirector)

xrootd:11000
(fedredir_vo)

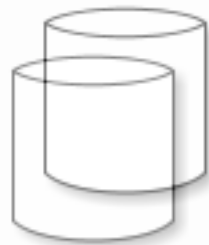
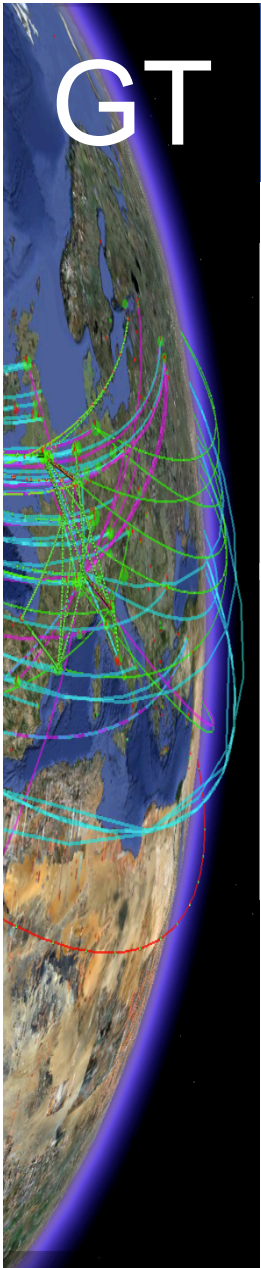
cmsd
(fedredir_vo)



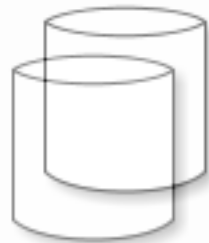
regional
redirector

xroot client: open

`xroot://dpm.example.com//vo/example2.dat`



xrootd:1095
(disk server)



xrootd:1095
(disk server)



DPM service

xrootd
(redirector)

xrootd:11000
(fedredir_vo)

cmsd
(fedredir_vo)



regional
redirector

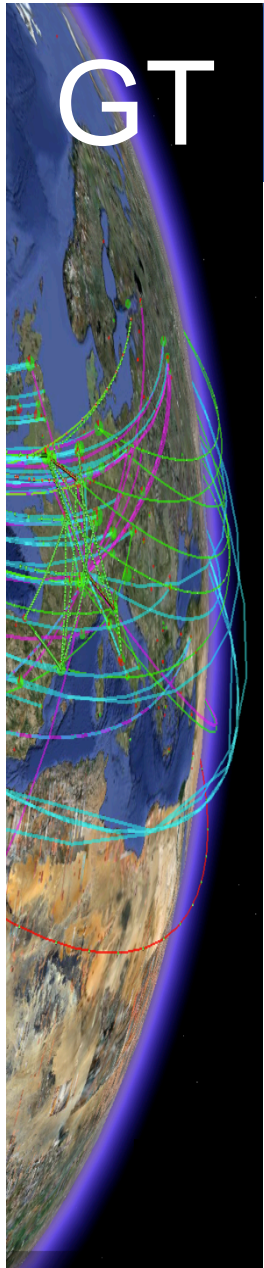
xroot client: open

`xroot://dpm.example.com//vo/example2.dat`

In case of no file:
Redirect to regional.example.org

dpm_get

Few other details





xrootd:1095
(disk server)



- May use a standard name2name lib
- But loaded by our cmslib not OSS
- Calls lfn2pfn method, e.g. for global namespace to surl



DPM service

xrootd
(redirector)

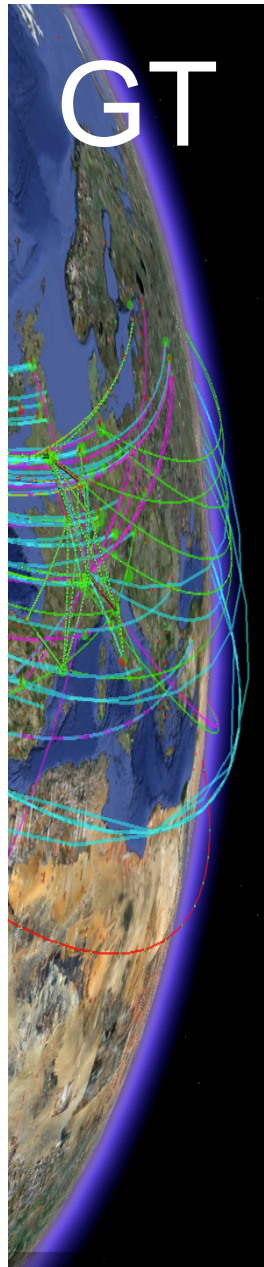
xrootd:11000
(fedredir_vo)

xrootd:11001
(fedredir_vo2)

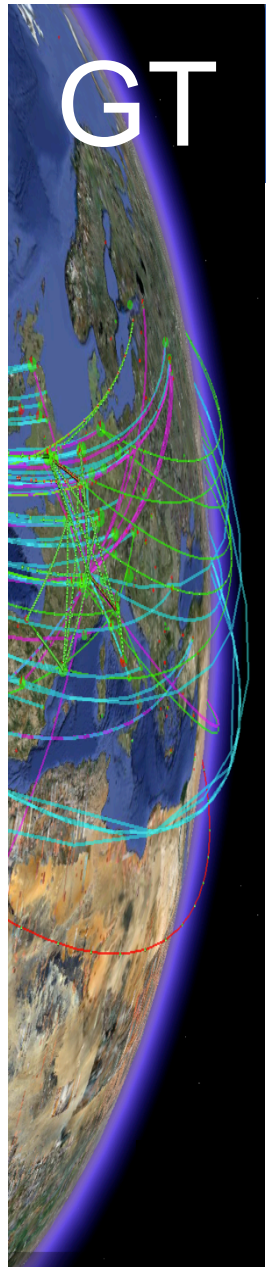
cmsd
(fedredir_vo)

cmsd
(fedredir_vo2)

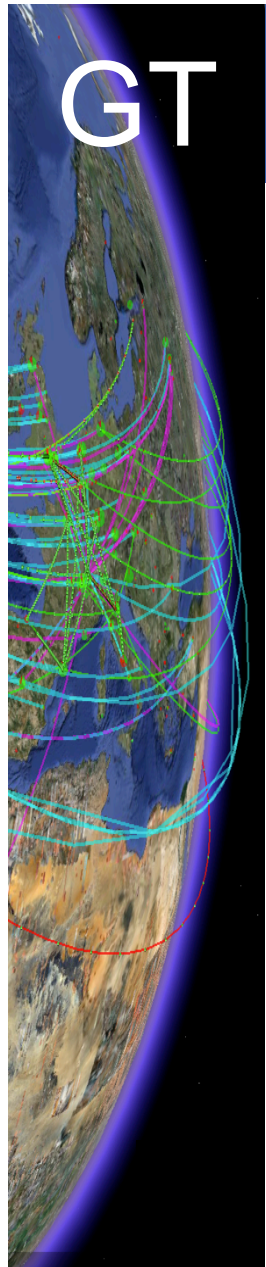
- xrootd **redirector**
 - libXrdDPMFinder (XrdCmsClient)
 - libXrdDPMOss (XrdOss)
 - libXrdDPMRedirAcc (XrdAccAuthorize)
- xrootd **disk**
 - libXrdDPMDiskAcc (XrdAccAuthorize)
 - libXrdDPMOss



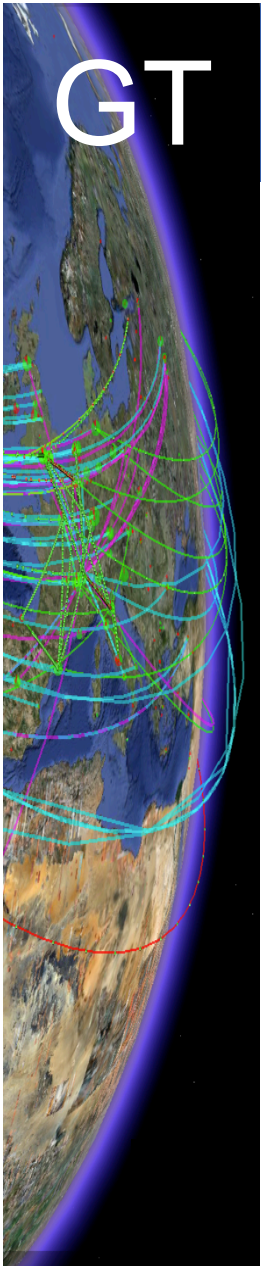
Option	Where	Description
dpm.putlifetime	xroot (XrdDPMFinder) config	Default parameter for put requests
dpm.putfiletype	xroot (XrdDPMFinder) config	Default parameter for put requests
dpm.putstoken	xroot (XrdDPMFinder) config	Default parameter for put requests
dpm.putsize	xroot (XrdDPMFinder) config	Default parameter for put requests
dpm.getlifetime	xroot (XrdDPMFinder) config	Default parameter for get requests
dpm.getfiletype	xroot (XrdDPMFinder) config	Default parameter for get requests
dpm.getstoken	xroot (XrdDPMFinder) config	Default parameter for get requests
dpm.principal	xroot (XrdDPMFinder) config	Fixed ID: username
dpm.fqan	xroot (XrdDPMFinder) config	Fixed ID: /vo/group
dpm.fixedidrestrict	xroot (XrdDPMFinder) config	Requests using the preset ID must start with this path
dpm.namelib	xroot (XrdDPMFinder) config	name2name library for filename translation
dpm.localroot	xroot (XrdDPMFinder) config	lroot argument for namelib
dpm.namecheck	xroot (XrdDPMFinder) config	Require result from namelib to begin with this prefix
dpm.enablecmsclient	xroot (XrdDPMFinder) config	Start a XrdCmsFinder target client
dpm.defaultprefix	xroot (XrdDPMFinder) config	If no namelib, fixup filename to begin with this prefix
dpm.replacementprefix	xroot (XrdDPMFinder) config	In no namelib, do string replace on filename prefix
dpm.allowvo	xroot (XrdDPMFinder) config	Restrict access to VO (requires VOMS)
dpm.xrdserverport	xroot (XrdDPMFinder) config	Port number for disk server xrootd in redirects
dpm.mkpath	xroot (XrdDPMFinder) config	Make leading paths in create requests
dpm.mmreqhost	xroot (XrdDPMFinder) config	Allowed hosts for anonymous stat requests
dpm.gracetime	xroot (XrdDPMFinder) config	Validity time to write into redirect opaque data
dpm.reqsize, oss.asize	Request opaque data	Per request size for put
dpm.ftype	Request opaque data	Per request file type
dpm.stoken	Request opaque data	Per request space token
dpm.lifetime	Request opaque data	Per request lifetime
	1st argument of XrdDPMDiskAcc	Enforced maximum validity time in hash check
	2nd argument of XrdDPMDiskAcc	HMAC Key (not recommended to set this here)
	1st argument of XrdDPMRedirAcc	XrdAccAuthorize plugin for fixed ID requests



Option	Where	Description
dpm.putlifetime	xroot (XrdDPMFinder) config	Default parameter for put requests
dpm.putfiletype	xroot (XrdDPMFinder) config	Default parameter for put requests
dpm.putstoken	xroot (XrdDPMFinder) config	Default parameter for put requests
dpm.putsize	xroot (XrdDPMFinder) config	Default parameter for put requests
dpm.getlifetime	xroot (XrdDPMFinder) config	Default parameter for get requests
dpm.getfiletype	xroot (XrdDPMFinder) config	Default parameter for get requests
dpm.getstoken	xroot (XrdDPMFinder) config	Default parameter for get requests
dpm.principal	xroot (XrdDPMFinder) config	Fixed ID: username
dpm.fqan	xroot (XrdDPMFinder) config	Fixed ID: /vo/group
dpm.fixedidrestrict	xroot (XrdDPMFinder) config	Requests using the preset ID must start with this path
dpm.namelib	xroot (XrdDPMFinder) config	name2name library for filename translation
dpm.localroot	xroot (XrdDPMFinder) config	lroot argument for namelib
dpm.namecheck	xroot (XrdDPMFinder) config	Require result from namelib to begin with this prefix
dpm.enablecmsclient	xroot (XrdDPMFinder) config	Start a XrdCmsFinder target client
dpm.defaultprefix	xroot (XrdDPMFinder) config	If no namelib, fixup filename to begin with this prefix
dpm.replacementprefix	xroot (XrdDPMFinder) config	In no namelib, do string replace on filename prefix
dpm.allowvo	xroot (XrdDPMFinder) config	Restrict access to VO (requires VOMS)
dpm.xrdserverport	xroot (XrdDPMFinder) config	Port number for disk server xrootd in redirects
dpm.mkpath	xroot (XrdDPMFinder) config	Make leading paths in create requests
dpm.mmreqhost	xroot (XrdDPMFinder) config	Allowed hosts for anonymous stat requests
dpm.gracetime	xroot (XrdDPMFinder) config	Validity time to write into redirect opaque data
dpm.reqsize, oss.asize	Request opaque data	Per request size for put
dpm.ftype	Request opaque data	Per request file type
dpm.stoken	Request opaque data	Per request space token
dpm.lifetime	Request opaque data	Per request lifetime
	1st argument of XrdDPMDiskAcc	Enforced maximum validity time in hash check
	2nd argument of XrdDPMDiskAcc	HMAC Key (not recommended to set this here)
	1st argument of XrdDPMRedirAcc	XrdAccAuthorize plugin for fixed ID requests

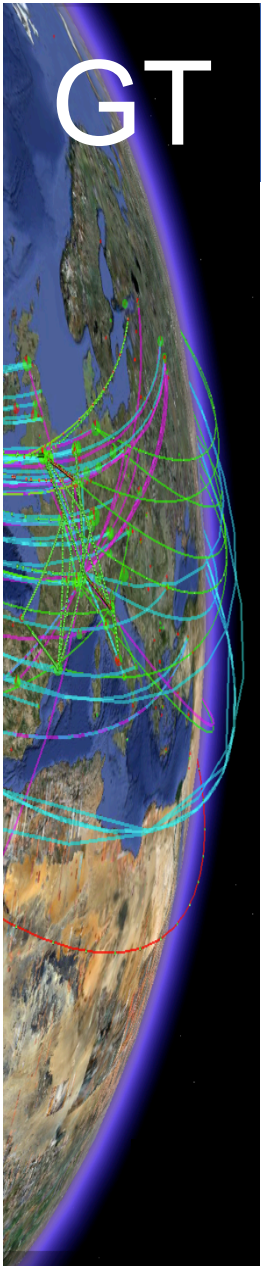


Option	Where	Description
dpm.putlifetime	xroot (XrdDPMFinder) config	Default parameter for put requests
dpm.putfiletype	xroot (XrdDPMFinder) config	Default parameter for put requests
dpm.putstoken	xroot (XrdDPMFinder) config	Default parameter for put requests
dpm.putsize	xroot (XrdDPMFinder) config	Default parameter for put requests
dpm.getlifetime	xroot (XrdDPMFinder) config	Default parameter for get requests
dpm.getfiletype	xroot (XrdDPMFinder) config	Default parameter for get requests
dpm.getstoken	xroot (XrdDPMFinder) config	Default parameter for get requests
dpm.principal	xroot (XrdDPMFinder) config	Fixed ID: username
dpm.fqan	xroot (XrdDPMFinder) config	Fixed ID: /vo/group
dpm.fixedidrestrict	xroot (XrdDPMFinder) config	Requests using the preset ID must start with this path
dpm.namelib	xroot (XrdDPMFinder) config	name2name library for filename translation
dpm.localroot	xroot (XrdDPMFinder) config	lroot argument for namelib
dpm.namecheck	xroot (XrdDPMFinder) config	Require result from namelib to begin with this prefix
dpm.enablecmsclient	xroot (XrdDPMFinder) config	Start a XrdCmsFinder target client
dpm.defaultprefix	xroot (XrdDPMFinder) config	If no namelib, fixup filename to begin with this prefix
dpm.replacementprefix	xroot (XrdDPMFinder) config	In no namelib, do string replace on filename prefix
dpm.allowvo	xroot (XrdDPMFinder) config	Restrict access to VO (requires VOMS)
dpm.xrdserverport	xroot (XrdDPMFinder) config	Port number for disk server xrootd in redirects
dpm.mkpath	xroot (XrdDPMFinder) config	Make leading paths in create requests
dpm.mmreqhost	xroot (XrdDPMFinder) config	Allowed hosts for anonymous stat requests
dpm.gracetime	xroot (XrdDPMFinder) config	Validity time to write into redirect opaque data
dpm.reqsize, oss.asize	Request opaque data	Per request size for put
dpm.ftype	Request opaque data	Per request file type
dpm.stoken	Request opaque data	Per request space token
dpm.lifetime	Request opaque data	Per request lifetime
	1st argument of XrdDPMDiskAcc	Enforced maximum validity time in hash check
	2nd argument of XrdDPMDiskAcc	HMAC Key (not recommended to set this here)
	1st argument of XrdDPMRedirAcc	XrdAccAuthorize plugin for fixed ID requests

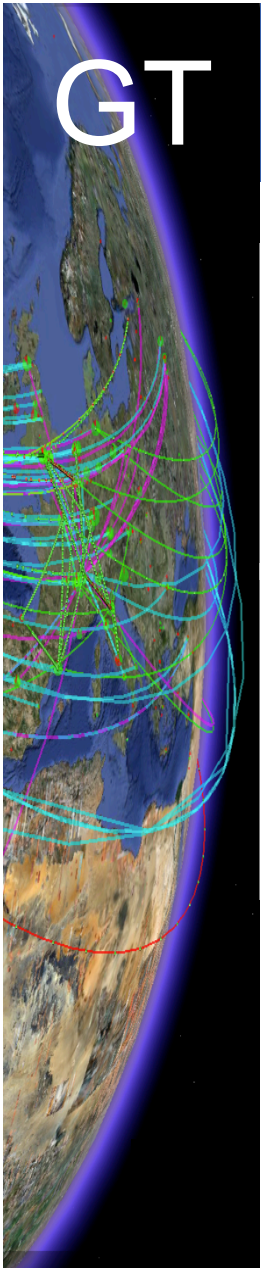


Option	Where	Description
dpm.putlifetime	xroot (XrdDPMFinder) config	Default parameter for put requests
dpm.putfiletype	xroot (XrdDPMFinder) config	Default parameter for put requests
dpm.putstoken	xroot (XrdDPMFinder) config	Default parameter for put requests
dpm.putsize	xroot (XrdDPMFinder) config	Default parameter for put requests
dpm.getlifetime	xroot (XrdDPMFinder) config	Default parameter for get requests
dpm.getfiletype	xroot (XrdDPMFinder) config	Default parameter for get requests
dpm.getstoken	xroot (XrdDPMFinder) config	Default parameter for get requests
dpm.principal	xroot (XrdDPMFinder) config	Fixed ID: username
dpm.fqan	xroot (XrdDPMFinder) config	Fixed ID: /vo/group
dpm.fixedidrestrict	xroot (XrdDPMFinder) config	Requests using the preset ID must start with this path
dpm.namelib	xroot (XrdDPMFinder) config	name2name library for filename translation
dpm.localroot	xroot (XrdDPMFinder) config	lroot argument for namelib
dpm.namecheck	xroot (XrdDPMFinder) config	Require result from namelib to begin with this prefix
dpm.enablecmsclient	xroot (XrdDPMFinder) config	Start a XrdCmsFinder target client
dpm.defaultprefix	xroot (XrdDPMFinder) config	If no namelib, fixup filename to begin with this prefix
dpm.replacementprefix	xroot (XrdDPMFinder) config	In no namelib, do string replace on filename prefix
dpm.allowvo	xroot (XrdDPMFinder) config	Restrict access to VO (requires VOMS)
dpm.xrdserverport	xroot (XrdDPMFinder) config	Port number for disk server xrootd in redirects
dpm.mkpath	xroot (XrdDPMFinder) config	Make leading paths in create requests
dpm.mmreqhost	xroot (XrdDPMFinder) config	Allowed hosts for anonymous stat requests
dpm.gracetime	xroot (XrdDPMFinder) config	Validity time to write into redirect opaque data
dpm.reqsize, oss.asize	Request opaque data	Per request size for put
dpm.ftype	Request opaque data	Per request file type
dpm.stoken	Request opaque data	Per request space token
dpm.lifetime	Request opaque data	Per request lifetime
	1st argument of XrdDPMDiskAcc	Enforced maximum validity time in hash check
	2nd argument of XrdDPMDiskAcc	HMAC Key (not recommended to set this here)
	1st argument of XrdDPMRedirAcc	XrdAccAuthorize plugin for fixed ID requests

- Available via dedicated repository
 - including xrootd packages (EPEL packaging)
- Expect xrootd 3.2.x in the EMI third party repo first
 - Eventually recent version of xrootd to go into EPEL
- dpm-xrootd setup instructions on the web
 - YAIM module written for yaim generation of the config files
- Deployments at four sites
 - May be others



- Eventually use dmlite rather than classic dpm/dpns api for dpm service access.
- VOMS
- dpm-xrootd uses few xrootd interfaces for which ABI compatibility across minor releases isn't guaranteed. Remove when possible.
- **Monitoring**
 - dpm site probably multi-vo



xrootd:1095
(disk server)



xrootd:1095
(disk server)



DPM service

xrootd
(redirector)

xrootd:11000
(fedredir_vo)

xrootd:11001
(fedredir_vo2)

cmsd
(fedredir_vo)

cmsd
(fedredir_vo2)

- DPM Support: hep-service-dpm@cern.ch
- dpm-xrootd setup wiki:
 - <https://svnweb.cern.ch/trac/lcgdm/wiki/Dpm/Xroot/Setup>