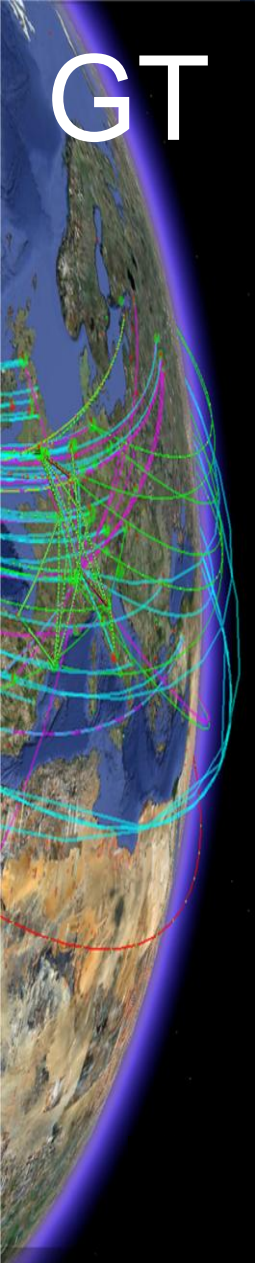


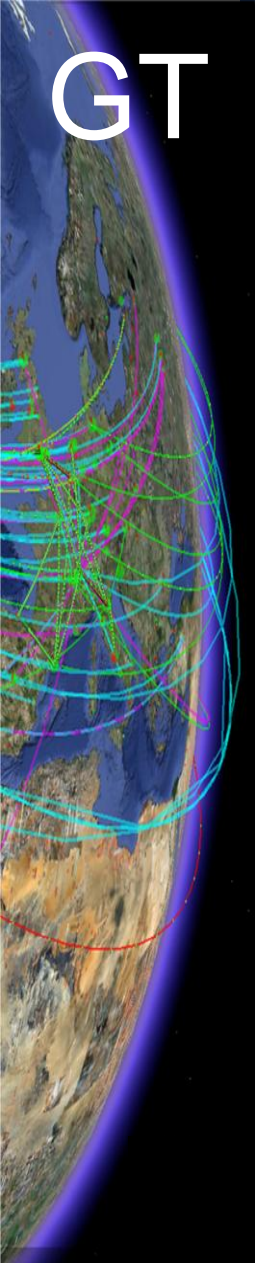
Federation Concepts and DPM's experience

Oliver Keeble

On behalf of IT-GT-DMS, CERN

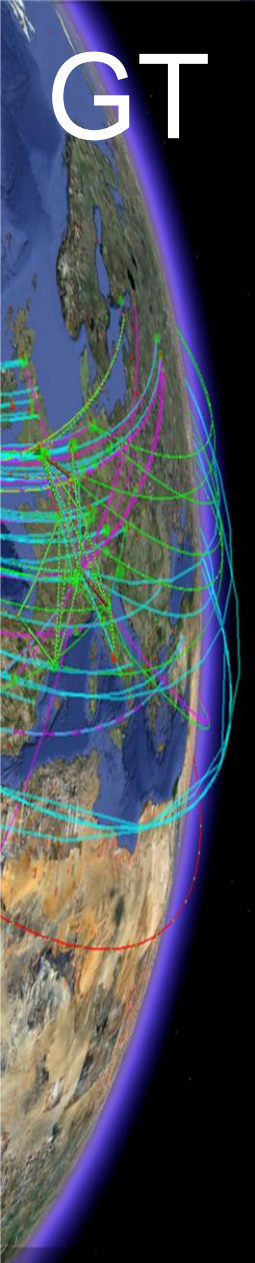


- IT-GT-DMS and their quest for federation
- What flavour of federation?
 - Which are the important federation advantages?
 - Which are the important concepts?
 - Which of the building blocks are needed?
 - What risks are associated with federation?
 - What balance between control & commoditisation?
- Some federation examples

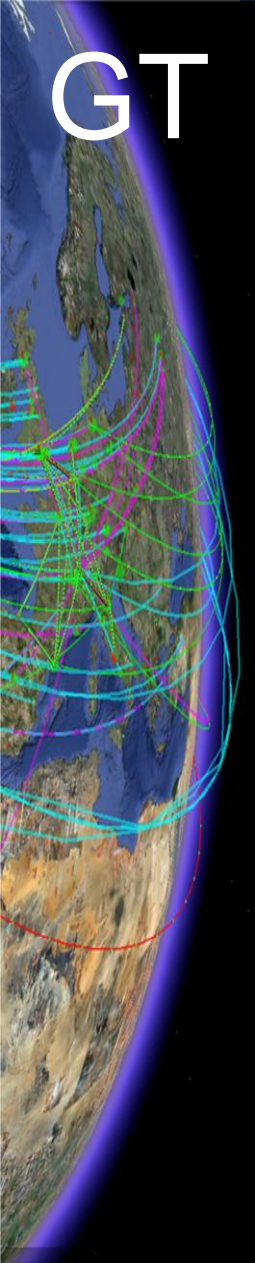


- DPM has the following relevant features
 - Xrootd plugin
 - HTTP/WebDAV interface (beta)
 - NFSv4.1 (pNFS) interface (beta)
 - Refactored nameserver/storage library (dev)
 - Now modular, allowing easy incorporation of multiple data sources
- LFC is the same codebase
 - Already a global namespace
- gfal/lcg_util can tie all this (and more) together
- FTS schedules file transfers
- Semsg – messaging interface to DPM/LFC
- This offers numerous federation possibilities
- We are here to understand what is required
 - And which directions should be explored

- Simplicity (users and sites)
- Robustness
 - Failover
- Efficiency
 - Fewer unnecessary replicas
 - Cost of network usage and chaotic access patterns
- Performance
 - Use of multiple replicas
- Elasticity
 - eg easy provisioning via cloud providers
- Data placement
 - Push/pull, controlled/on-demand
- In isolation, most of the above can be addressed without invoking federation



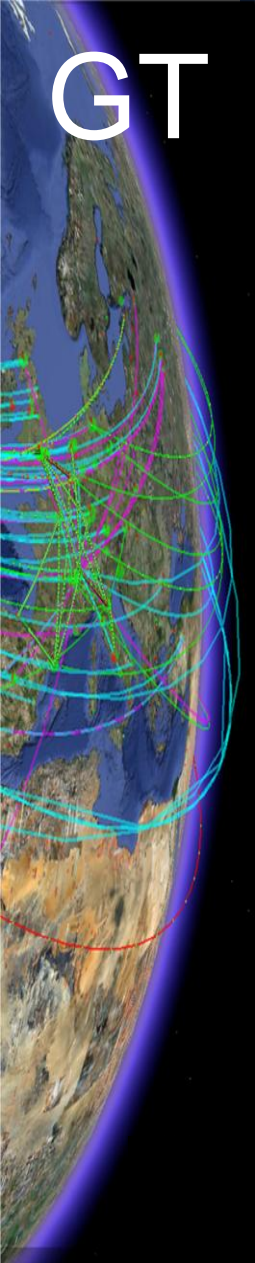
- Integrated namespace
 - Xrootd, NFS, LFC, Fedfs, AFS, DFS
- Replica management
 - Xrootd, NFS, LFC, AFS
- File access and redirection
 - Xrootd, HTTP, NFS, AFS
- File transfer
 - Xrootd, HTTP, FTS, gridFTP
- Data Caching
 - Xrootd, SQUID, NFS, AFS, HTCP
- Metadata Caching
 - Xrootd
- Writing
 - Just say no



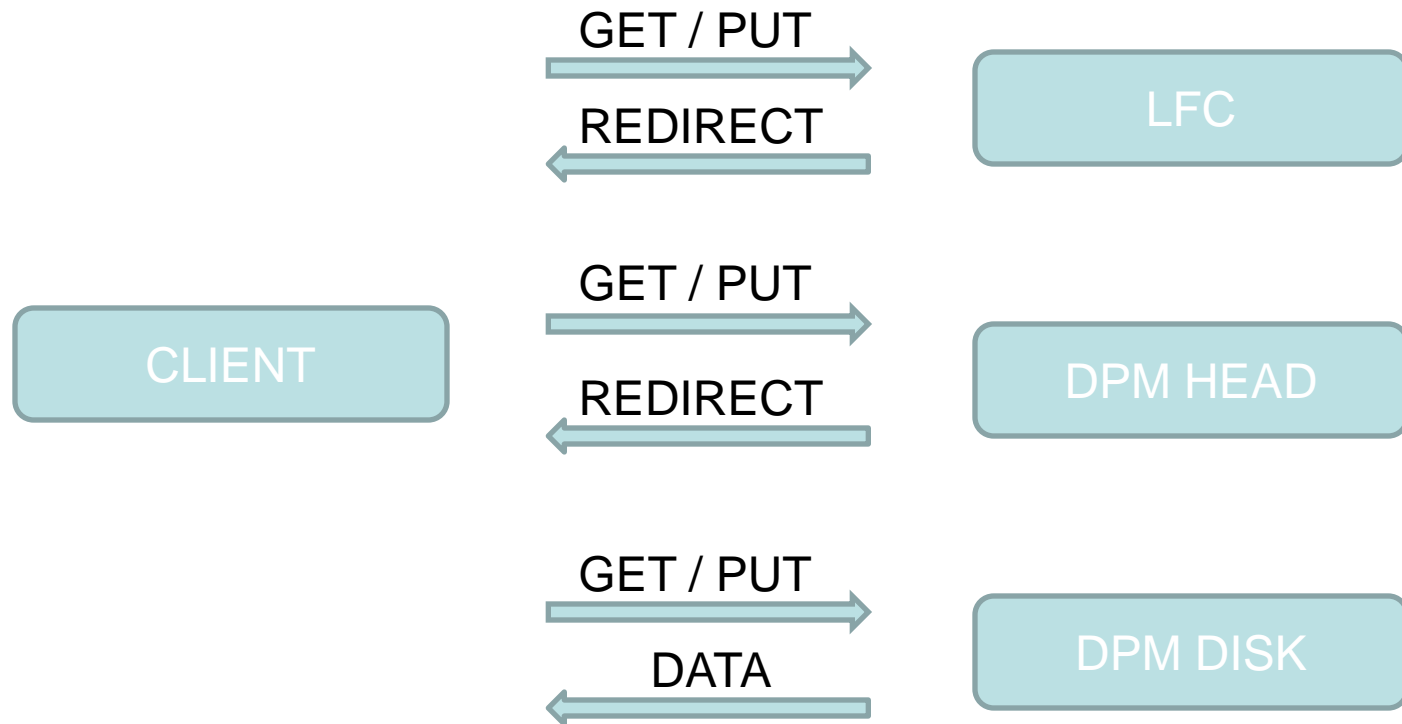
- Integrated Namespaces & replica location
 - Federation varieties
 - /site1/a/b.dat, /site2/c/d.dat – concatenated, no replica awareness
 - /store/group/run/dir/data.dat – redirects transparently to the above replicas
 - Can one federation solution fit everybody?
 - Frameworks and policies differ
 - Role of ‘namespace congruence’
 - guides in SEs
 - Metadata catalogues will always be needed
 - What replica location information is in the metadata catalogue?
 - What metadata is present in the federation?
 - Properties – ACLs, checksums...
 - Replica selection information (server load, network location...)

- File access
 - What is required for federation?
 - Redirection (I/O, but also AA)
 - WAN friendly protocol
 - Supported by numerous protocols (some as old as AFS)
 - So on what basis is a decision on preference to be made?
 - What's the impact on a site of remote access of its files?
 - What will the network impact be of transparent failover?
- File transfer
 - Can the traffic be controlled?
 - Role for FTS?
 - Integration with lcg_util
 - Scheduling / throttling
 - dynamic replication, migration

- Caching
 - Many (non exclusive) strategies and varieties exist
 - OS, application, site, file, block...
 - Is cache querying envisaged?
 - Does a remotely fetched and locally cached file join the federation?
 - Role of remote access
- Writing
 - The federations considered are read-only
 - How is data imported/registered to the federation?
 - What relation to experiment DM frameworks



- pNFS
 - Decoupled namespace / WAN friendly data access
 - Redundancy / parallelism
 - Kernel block cache
 - Global namespace
 - Conceptually good, implementations unverified and immature
- DPM/xrootd
 - X509, performance enhancements, federation support can all be added to the DPM xrootd plugin
- A DPM/dCache federation based on HTTP access and redirection
 - Both SEs now have HTTP interfaces
- Federated LFCs – a possible option for Atlas to manage CERN/BNL
 - Same nameserver extension above could manage federation between LFCs



- This has been demonstrated using the HTTP interface to LFC & DPM
- Can even incorporate dCache



How do we get there?

- Transition
 - Building small islands
 - What else is needed for truly global federation?
 - Should existing data be imported or transparently available?
 - Are experiments contemplating policy changes which will make federation easier?
 - Solution diversity and existing installations
 - Coexistence
- Costs
 - Sustainability
 - Site and software maintenance
 - HTTP implementation for DPM was easy to do using apache
 - HEP / NFS+AFS / Web dev effort
 - few, some, army



- CERN data management tools offer a rich set of potential directions for federation
- Various federation solutions exist already and many more are possible
- The key to identifying how to proceed lies in
 - Identifying which constituent concepts are the important ones
 - Understanding the motivations for federation
 - Understanding the costs
 - Deciding how we get there



Thanks

- Thanks to the following for their contributions
 - Markus Schulz
 - Dirk Duellmann
 - Fabrizio Furano
 - Ricardo Rocha
 - IT-GT-DMS team