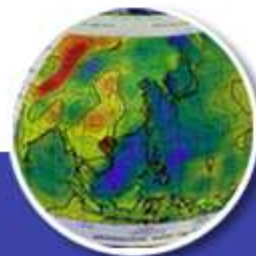




# Architecture of the gLite Data Management System

Jhen-Wei Huang, ASGC  
EAGSS09  
July 27-31 2009, UPM Malaysia



# Outline



- **General Concept**
- **Storage Elements and SRM**
- **File Catalog and DM tools**



## Data Management System (DMS)

- **DMS enables the location, access and transfer of data in grid**
  - User do not need to know data location, just the logical name
  - Data is accessed through standard interfaces
  - Data can be replicated or transferred to several locations as needed
  - Data is shared within a VO

# Scope of data services



- **Simply, DMS provides all operation that all of us are used to performing**
  - **Uploading /downloading files**
  - **Creating file /directories**
  - **Renaming file /directories**
  - **Deleting file /directories**
  - **Moving file /directories**
  - **Listing directories**
  - **Creating symbolic links**
- **Note: Files are write-once, read-many**
  - Files cannot be changed unless remove or replaced

# Data services in gLite



- **3 types of services for DM:**
  - **Storage** (SE's): where files are “physically” located
    - Storage URL or **SURL**:
      - **srm://castor.sc.grid.sinica.edu.tw/data/dteam/mytest.dat**
  - **Catalogs**: High level hierarchical namespace, maps the “physical” files to a virtual “logical” filename
    - Logical File Name or **LFN**:
      - **lfn:/grid/dteam/mytest.dat**
  - **Movement**: put/get files into grid SE's, move/replicate files between SE's.
    - File Transfer Service or FTS (*Not covered in this tutorial*)
    - Transport URL or **TURL**:
      - **gsiftp://sc003.grid.sinica.edu.tw:2811/data/dteam/mytest.dat**

# Grid Data Management Challenge



- **Resource centers need meet growing demand for storage**
  - DPM
- **Data is stored on different storage systems technologies**
  - SRM
- **Data is stored at different locations with separate namespace**
  - LFC
- **Large amount of data need be reliably distributed**
  - FTS

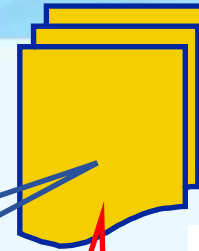
# Data management example



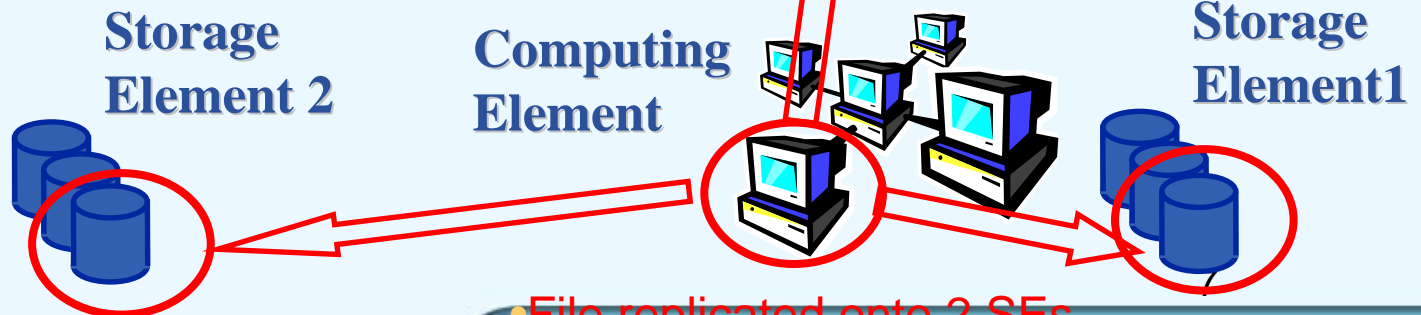
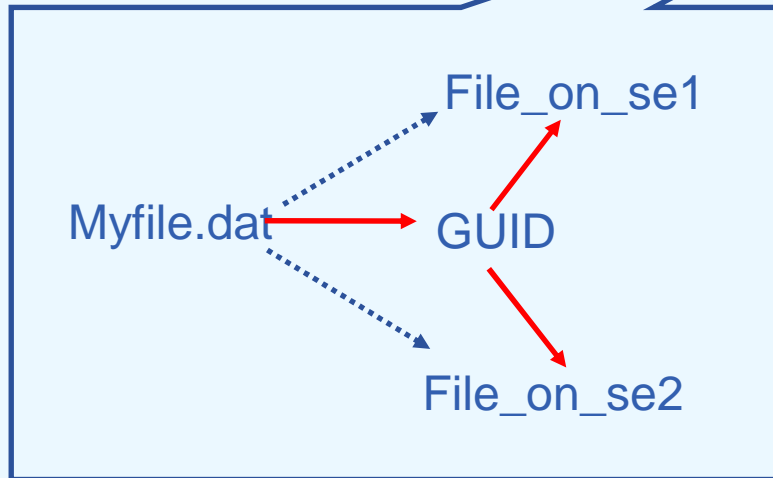
“User interface”



“Myfile.dat”



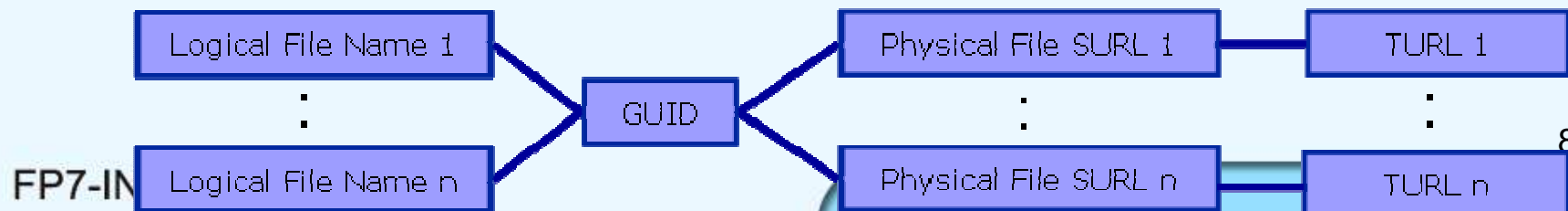
LCG File Catalogue (LFC)



# Name conventions



- **Logical File Name (LFN)**
  - An alias created by a user to refer to some item of data,  
e.g. “lfn:/grid/cms/20030203/run2/track1”
- **Globally Unique Identifier (GUID)**
  - A non-human-readable unique identifier for an item of data,  
e.g. “guid:f81d4fae-7dec-11d0-a765-00a0c91e6bf6”
- **Storage URL (SURL) or Physical File Name (PFN)**
  - The location of an actual piece of data on a storage system,  
e.g. <sfm | srm>://<SE\_hoatname><SE\_Accesspoint><VO\_Path><filename>  
“srm://dpm.grid.sinica.edu.tw/grid/home/dteam/mytest.dat ”
- **Transport URL (TURL)**
  - Temporary locator of a replica + access protocol: understood by a SE,  
e.g. <protocol>://<SE\_hoatname>:<port>/<path>  
“gridftp://dpm.grid.sinica.edu.tw:2811/grid/home/dteam/mytest.dat ”



# Outline



- General Concept
- **Storage Elements and SRM**
- File Catalog and DM tools

# Storage Element



- **Provides**

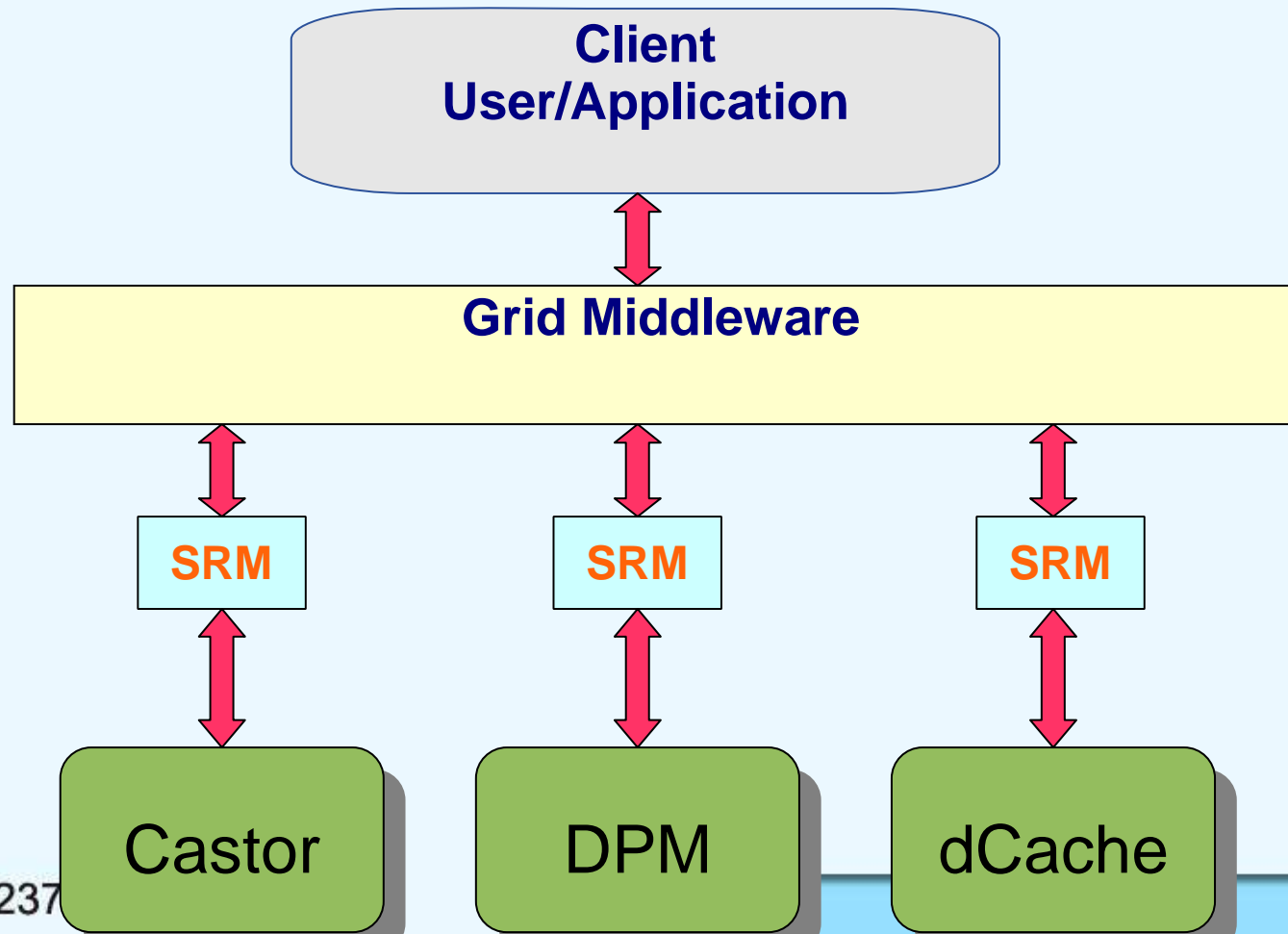
- Most common types of SE's found in the Grid infrastructures:
  - DPM
  - dCache
  - Castor (used at CERN, ASGC...)
  - StoRM , xrootd and so on
- Accessed via Grid File Access Layer (GFAL)
- Storage space for files
  - **Scalable** management for independent disk pools for sites
- Transfer protocol (gsiFTP) ~ GSI based FTP server
- SRM Interface

# Storage Resource Manager



- SRM (Storage Resource Manager)

Provides standardized Uniform Access to Storage and protocol negotiation.



# Outline



- **General Concept**
- **Storage Elements and SRM**
- **File Catalog and DM tools**

# LFC Service



- **LFC = LCG File Catalogue**
  - LCG = LHC Compute Grid
  - LHC = Large Hadron Collider
- **Provides**
  - Mapping between LFN, GUID and SURL
  - Hierarchical namespace, symbolic links
- **All members of a given VO have read-write permissions in their directory**
- **Commands look like UNIX with “lfc-” in front (often)**

# LFC Continued



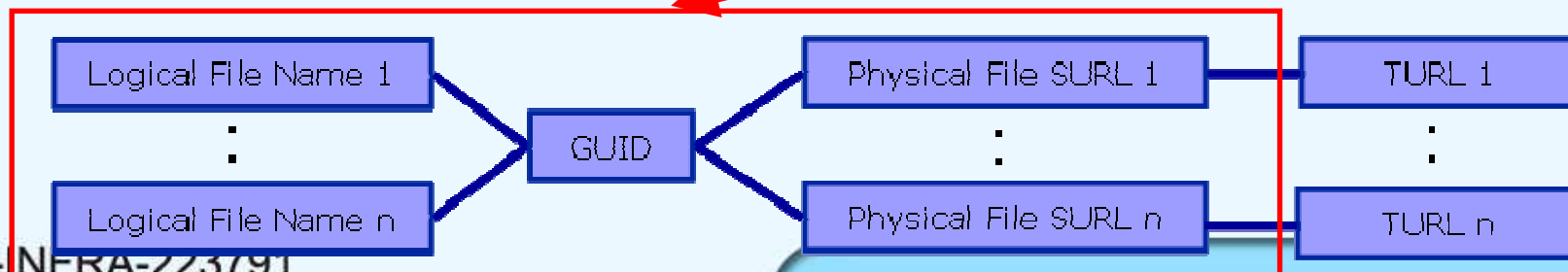
**LFC has a directory tree structure**  
**/grid/<VO\_name>/ <you create it>**

**LFC Namespace**

**Defined by the user**

- Example : /grid/dteam/hungche/testfiles.001

- Mapping by the "LFC" catalogue server



# Two sets of commands



- **lfc commands**
  - Use LFC commands to interact with the catalogue only
    - To create catalogue directory
    - List files
  - Used by you and by lcg-utils
- **lcg-utils**
  - Couples catalogue operations with file management
  - copy files to/from/between SEs
  - Replicated

# LFC Catalog commands



## Summary of the LFC Catalog commands

lfc-chmod	Change access mode of the LFC file/directory
lfc-chown	Change owner and group of the LFC file-directory
lfc-delcomment	Delete the comment associated with the file/directory
lfc-getacl	Get file/directory access control lists
lfc-ln	Make a symbolic link to a file/directory
lfc-ls	List file/directory entries in a directory
lfc-mkdir	Create a directory
lfc-rename	Rename a file/directory
lfc-rm	Remove a file/directory
lfc-setacl	Set file/directory access control lists
lfc-setcomment	Add/replace a comment

# Summary of lcg-utils commands



## Replica Management

lcg-cp	Copies a grid file to a local destination
lcg-cr	Copies a file to a SE and registers the file in the catalog
lcg-del	Delete one file
lcg-rep	Replication between SEs and registration of the replica
lcg-gt	Gets the TURL for a given SURL and transfer protocol
lcg-sd	Sets file status to "Done" for a given SURL in a SRM request

# Questions ?