eGee

Enabling Grids for E-sciencE

Data Management System (DMS)

HungChe Jen

EGEE Tutorial, June 5, 2007

Academia Sinica Grid Computing

www.eu-egee.org

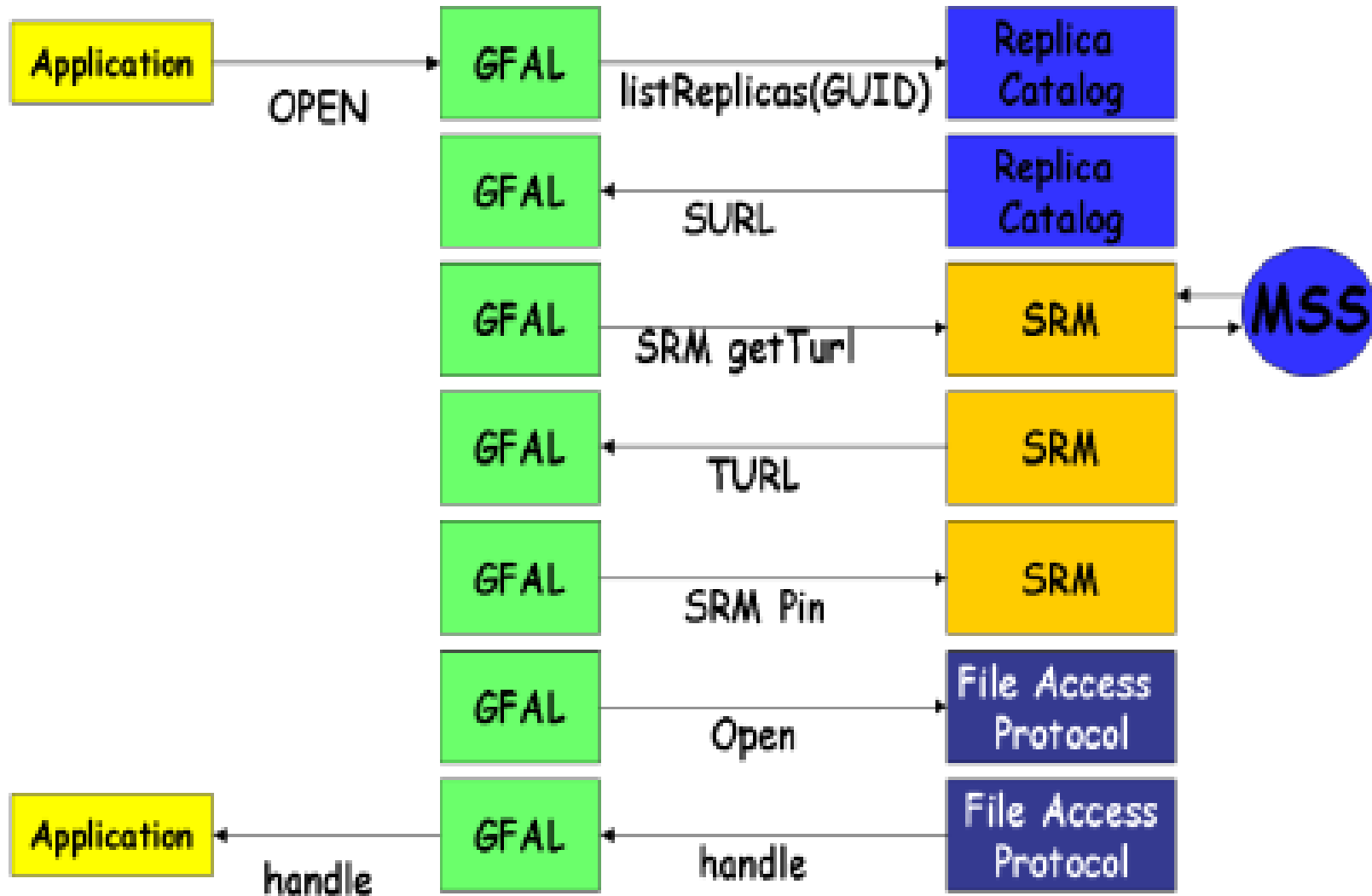Information Society

Enabling Grids for E-sciencE

- **Data Management**
  - Introduction
  - Examples
  - Name Convention
  - Storage Elements
  - LCG File Catalog

- **Data Management Practical**

- **Provides file manipulation services for users and other Grid services.**

- **DMS enables the location, access and transfer of data**
    - 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
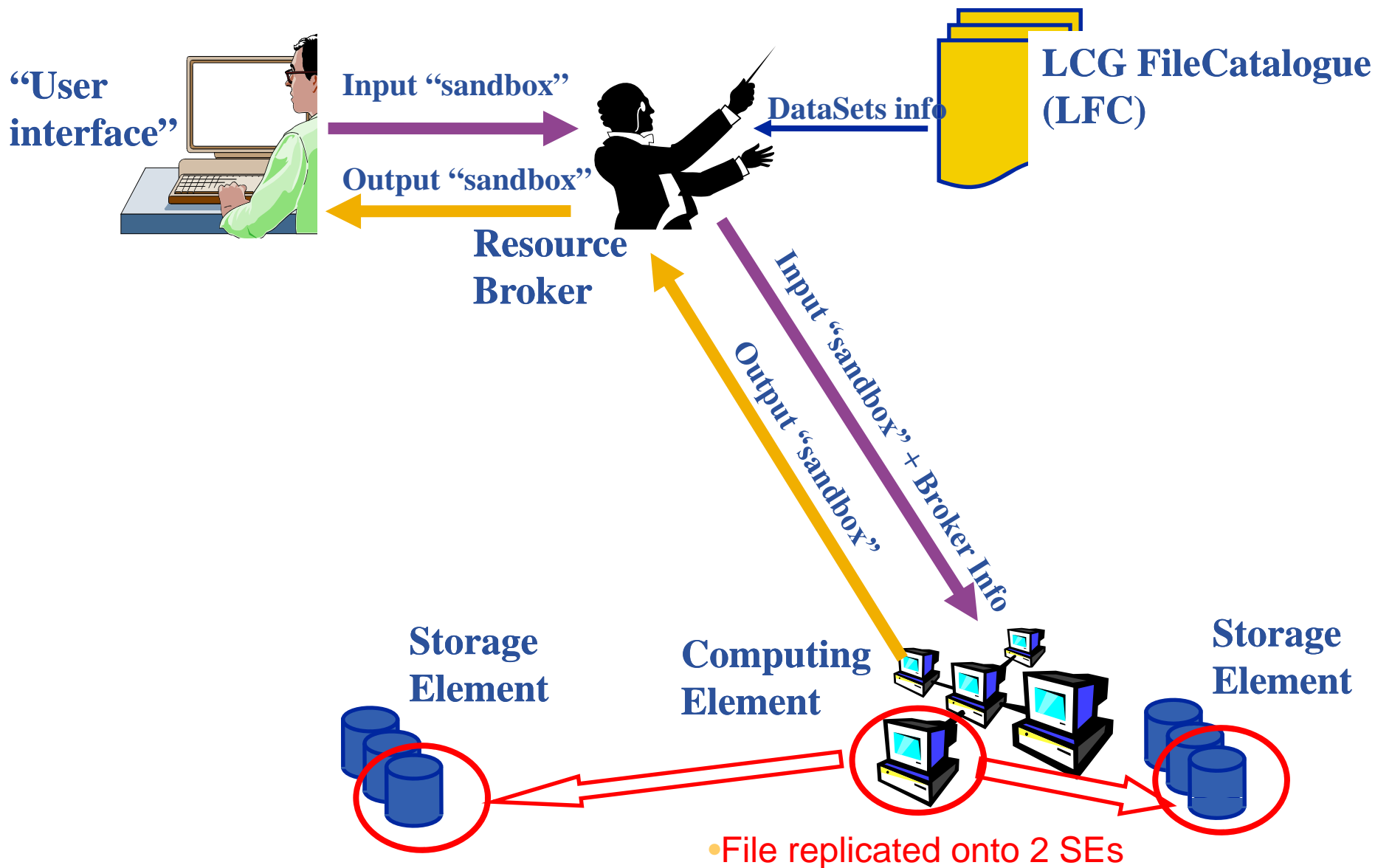
    - Data is shared within a VO

eGee

Enabling Grids for E-sciencE

- **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
  - – No intention of providing a global file management system

Data Issues and Grid Solutions

- **Issue: Resource centers need meet growing demand for storage**
- **Storage Element capable to manage multiple disk pools**
  - Disk Pool Manager (DPM), dCache, CASTOR

- **Issue: Data is stored on different storage systems technologies**
- **Common interface required to hide underlying complexity**
  - Storage Resource Manager (SRM) – storage management protocol
  - GridFTP – secure file transfer

- **Issue: Data is stored at different locations with separate namespace**
- **File catalogue to provide uniform view of Grid data**
  - LCG File Catalog (LFC)

- **Issue: Applications need to access Grid data management services**
- **Data management API**
  - GFAL

"User interface"

Input "sandbox"

Output "sandbox"

Resource Broker

DataSets info

LCG FileCatalogue (LFC)

Input "sandbox" + Broker Info

Output "sandbox"

Storage Element

Computing Element

Storage Element

•File replicated onto 2 SEs

"User interface"

"Myfile.dat"

LCG FileCatalogue (LFC)

File_on_se1

Myfile.dat → guid

File_on_se2

Storage Element 2

Storage Element1

•File replicated onto 2 SEs

"User interface"

"Myfile.dat"

LCG FileCatalogue (LFC)

File_on_se1 ("SURL": site URL)

Myfile.dat
"Logical filename"

"GUID" Global Unique Identifier

File_on_se2 ("SURL": site URL)

Storage Element2

Storage Element1

eGee

Enabling Grids for E-sciencE
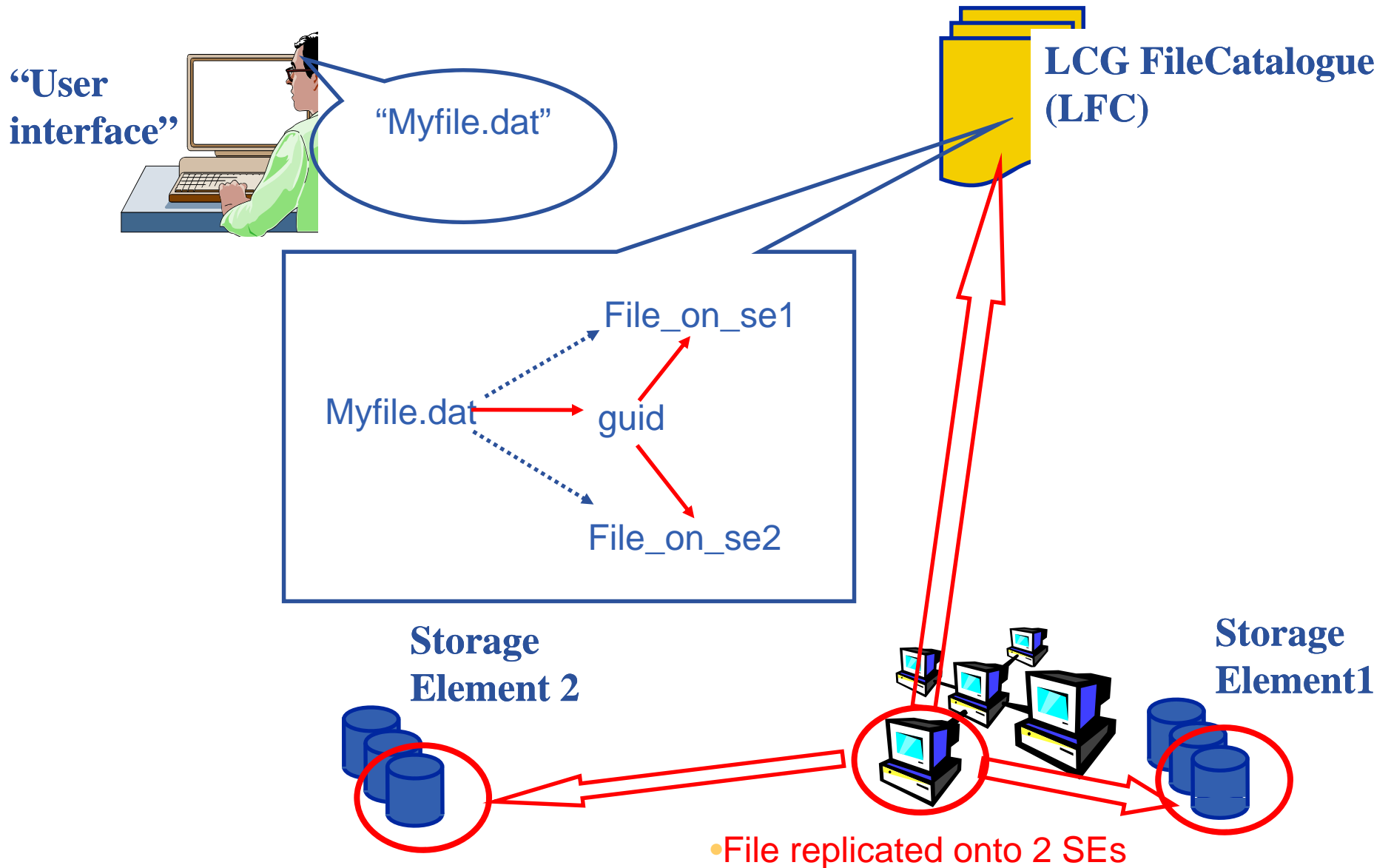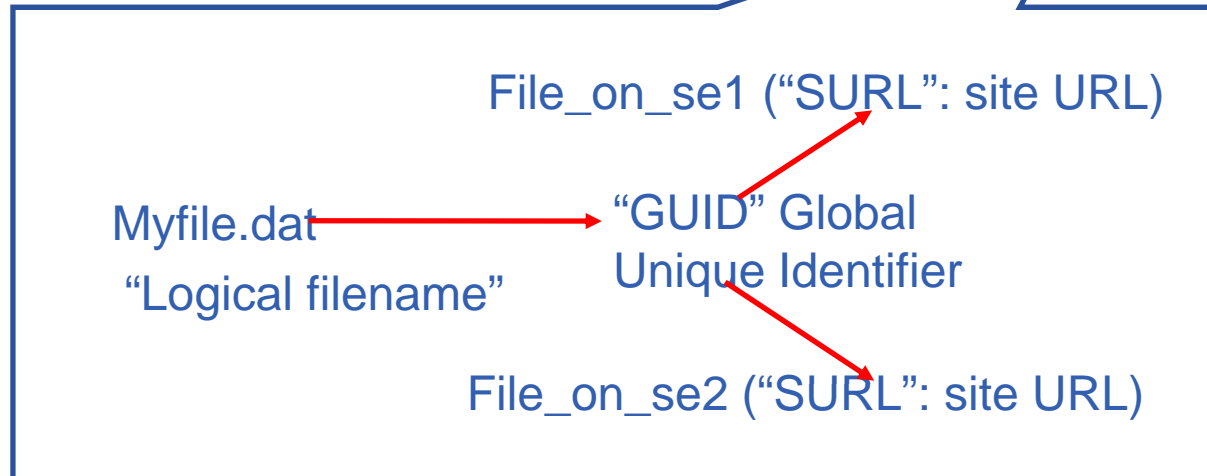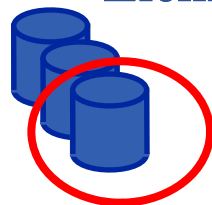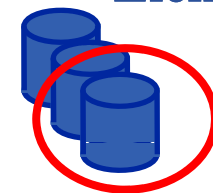
- Logical File Name (LFN)
  - An alias created by a user to refer to some item of data, e.g.
    "lfn:/grid/cms/20030203/run2/track1"

- 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.
    srm:<SE_hoatname>/<path>
    "srm://castor.grid.sinica.edu.tw/data/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://castor.grid.sinica.edu.tw:2811/data/dteam/mytest.dat"

Enabling Grids for E-sciencE
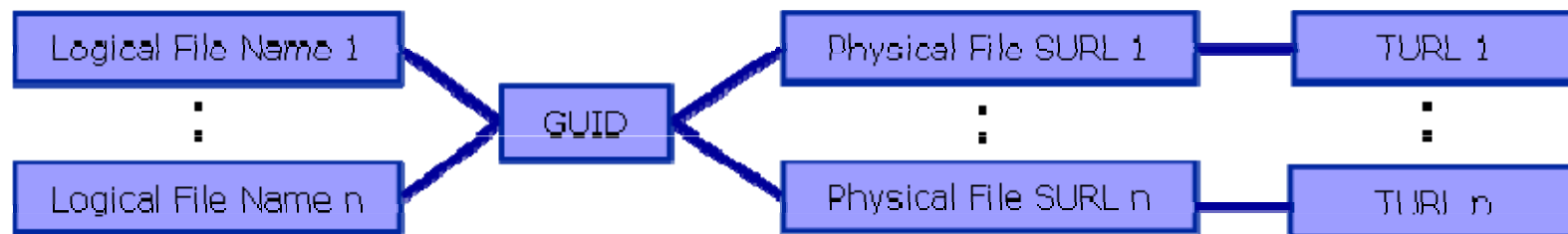
- **Provides**
  - Storage space for files
  - SRM Interface
  - Transfer protocol (gsiFTP) ~ GSI based FTP server
  - POSIX-like file access
    - Accessed via Grid File Access Layer (**GFAL**)
      - *API interface*
      - *To read parts of files too big to copy*

- **Example is Disk Pool Manager (DPM)**
  - Scalable management for independent disk pools for sites
  - Easy to install, configure and manage
  - Secure remote and local transfer protocols
    - GridFTP, secure RFIO

Enabling Grids for E-sciencE

- **LFC = LCG File Catalogue**
  - LCG = LHC Compute Grid
  - LHC = Large Hadron Collider

- **Provides**
  - Mapping between LFN, GUID and SURL
  - Transactions, Sessions, Bulk queries
  - Hierarchical namespace, symbolic links
  - System metadata
  - single string user metadata

- **All members of a given VO have read-write permissions in their directory**

- **Commands look like UNIX with "lfc-" in front (often)**

Enabling Grids for E-sciencE

- Users primarily access and manage files through "logical filenames"

LFC has a directory tree structure
/grid/<VO_name>/ <you create it>

LFC Namespace   **Defined by the user**
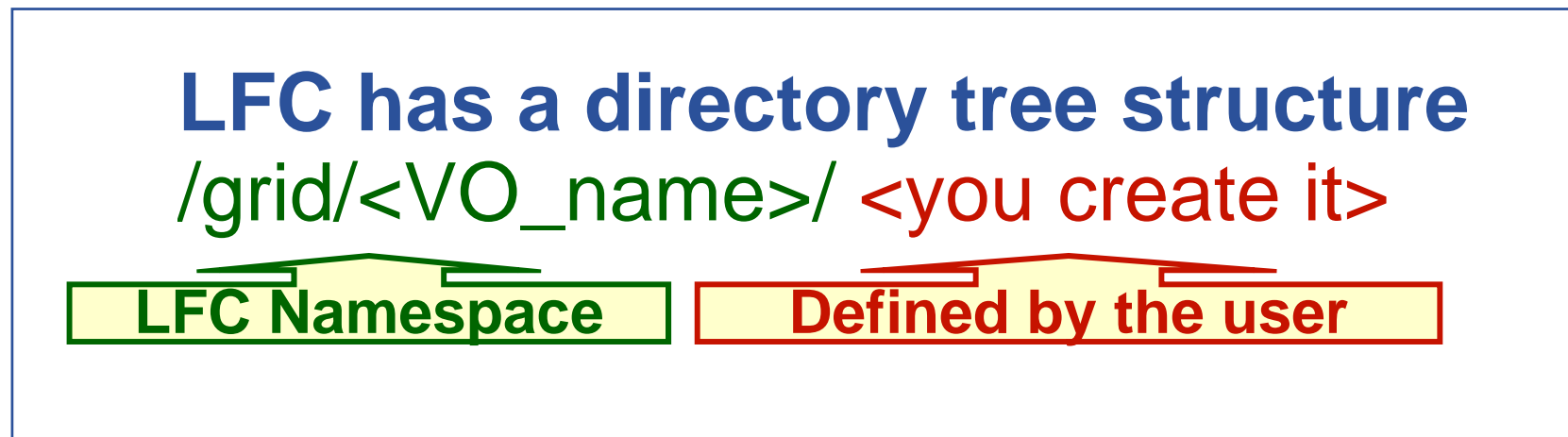
- Example :   /grid/dteam/hungche/testfiles.001

•Mapping by the "LFC" catalogue server

Logical File Name 1			GUID		Physical File SURL 1			TURL 1

Logical File Name 1 :
Logical File Name n

GUID

Physical File SURL 1 :
Physical File SURL n

TURL 1 :
TURL n

Enabling Grids for E-sciencE
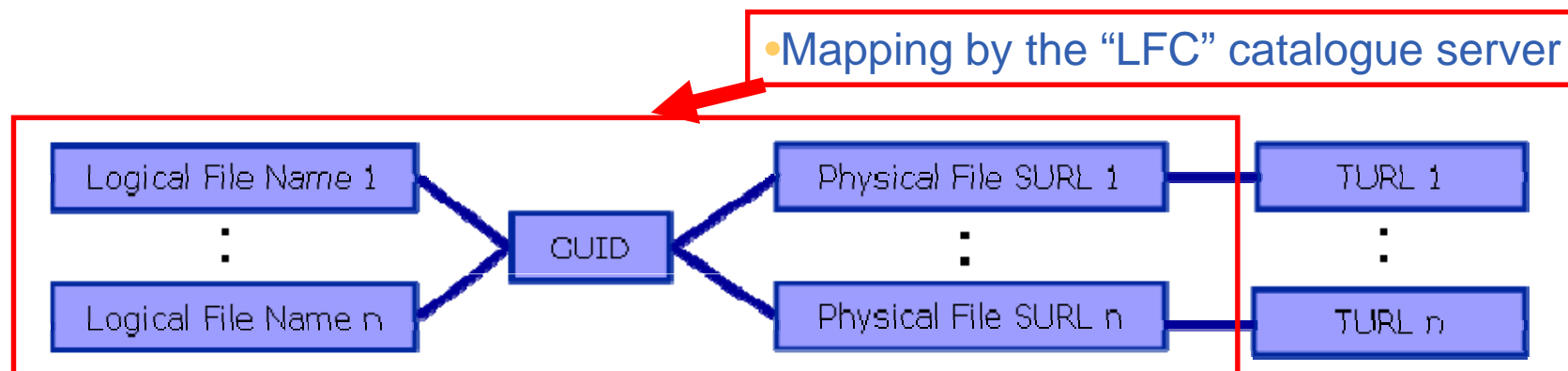
- **lfc commands**
  - Use LFC commands to interact with the catalogue only
    - To create catalogue directory
    - List files
  - Used by you and by lcg-utils

- **lcg-utils**
  - Couples catalogue operations with file management
    - Keeps SEs and catalogue in step!
  - copy files to/from/between SEs
  - Replicated

eGee

Enabling Grids for E-sciencE

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

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

eGee

Enabling Grids for E-sciencE

/grid/dteam/test_SE.002

LFC

c033f475-cc6a-4556-a42c-60fb29274524

sfn://castorsc.grid.sinica.edu.tw/castor/grid.sinica.edu.tw/sc/dteam/generated/2006-03-09/filec025611a-4619-4730-b6e0-5261e5ec095c

gsiftp://castorsc.grid.sinica.edu.tw:2811/castor/grid.sinica.edu.tw/sc/dteam/generated/2006-03-09/filec025611a-4619-4730-b6e0-5261e5ec095c

/castor/grid.sinica.edu.tw/sc/dteam/generated/2006-03-09/filec025611a-4619-4730-b6e0-5261e5ec095c

SE

lcg00116.grid.sinica.edu.tw:/flatfiles/SE00/dt/stage/filec025611a-4619-4730-b6e0-5261e5ec095c.815277

LFN

GUID

SFN

TURL

Logic File Name

Physical File Name

- **List directory**

- **Upload a file to an SE and register a logical name (lfn) in the catalog**

- **Create a duplicate in another SE**

- **List the replicas**

- **Create a second logical file name for a file**

- **Download a file from an SE to the UI**

- **Please go to the web page for this practical**

Practical Page :

- **http://indico.cern.ch/materialDisplay.py?contribId=6&materialId=0 &confId=14306**

Enabling Grids for E-sciencE

- **The following account can be used to log on the GILDA User Interface machines during the SINGAPORE tutorial.**

UI : glite-tutor.ct.infn.it   or glite-tutor2.ct.infn.it (gLite 3.0)

usernames : singapore01~40

passwords : GridSIN01~40

voms-proxy-init -voms gilda

passwords : SINGAPORE

During this practical, you should upload your file under

/grid/gilda/training/singapore/$USER