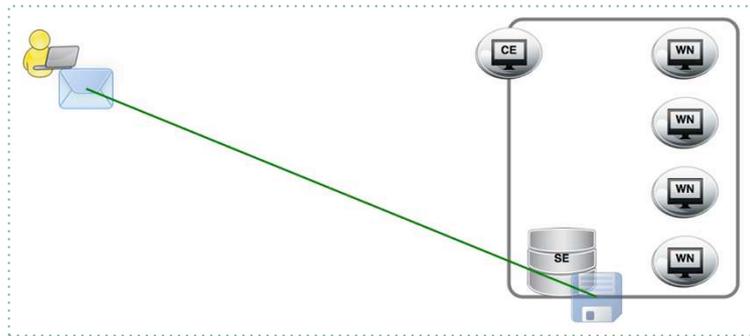


Data Management

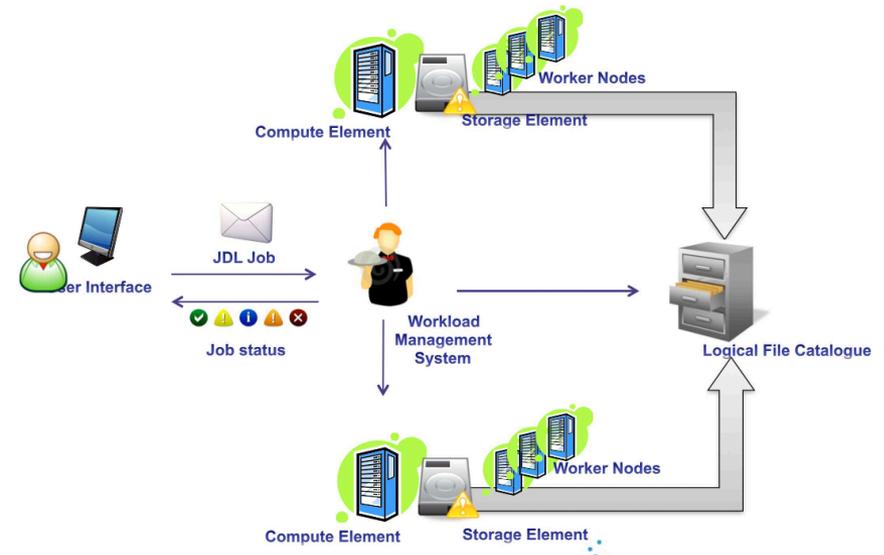
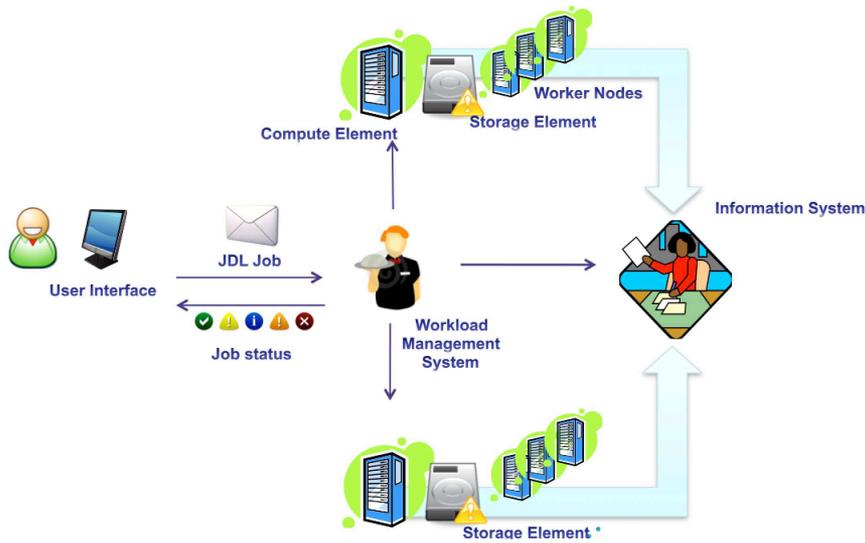


Outline

- ❑ Storage Information System
- ❑ Grid Identifiers
- ❑ Data Requirements attribute
- ❑ Large files on the Grid



Storage Information System



Grid Identifiers

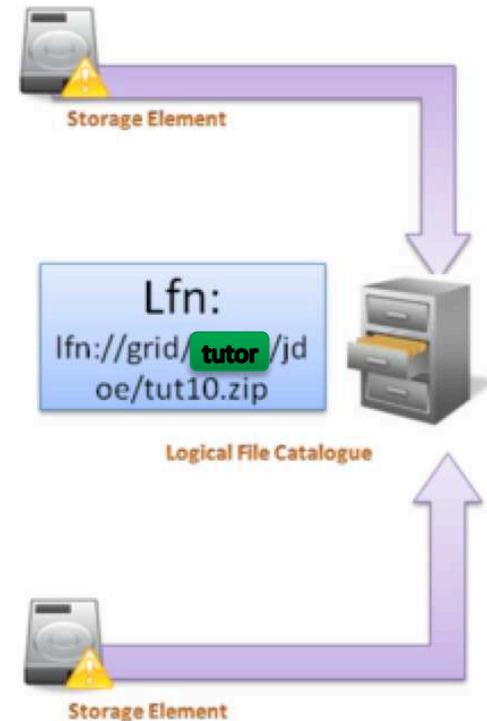
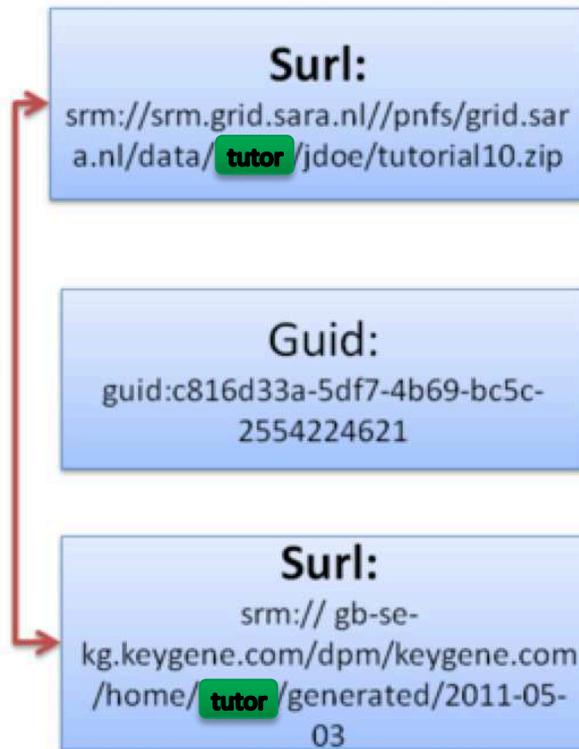
File identifiers

- GUID
- Lfn
- Surl
- Turl

1 GUID per file

1 SURL per replica

Replica



Grid Identifiers

❑ GUID - Grid Unique Identifier

- LFC assigns this unique name to files on the grid and maps LFNs toSURLs.

✓ `guid:<36_bytes_unique_string>`

eg. `guid:38ed3f60-c402-11d7-a6b0-f53ee5a37e1d`

❑ LFN - Logical File Name

- A user friendly (and user defined) name of a grid file

✓ `lfn:/grid/<vo>/<directory>/<file>`

eg. `lfn:/grid/tutor/mooc_test/filefromVM`

❑ SURL - Storage URL

- The physical stored version of a file on a grid storage element

✓ `srm://<SE_hostname>:<port>/<some_string>`

eg. `srm://srm.grid.sara.nl:8443/pnfs/grid.sara.nl/data/tutor/mooc_test/filefromVM`

❑ TURL - Transport URL

- A URL that a storage element issues for a SURL which can be used to retrieve or store data.

✓ `<protocol>://<SE_hostname>:<port>/<path>`

eg. `gsiftp://gridftp.grid.sara.nl:2811/pnfs/grid.sara.nl/data/tutor/mooc_test/filefromVM`

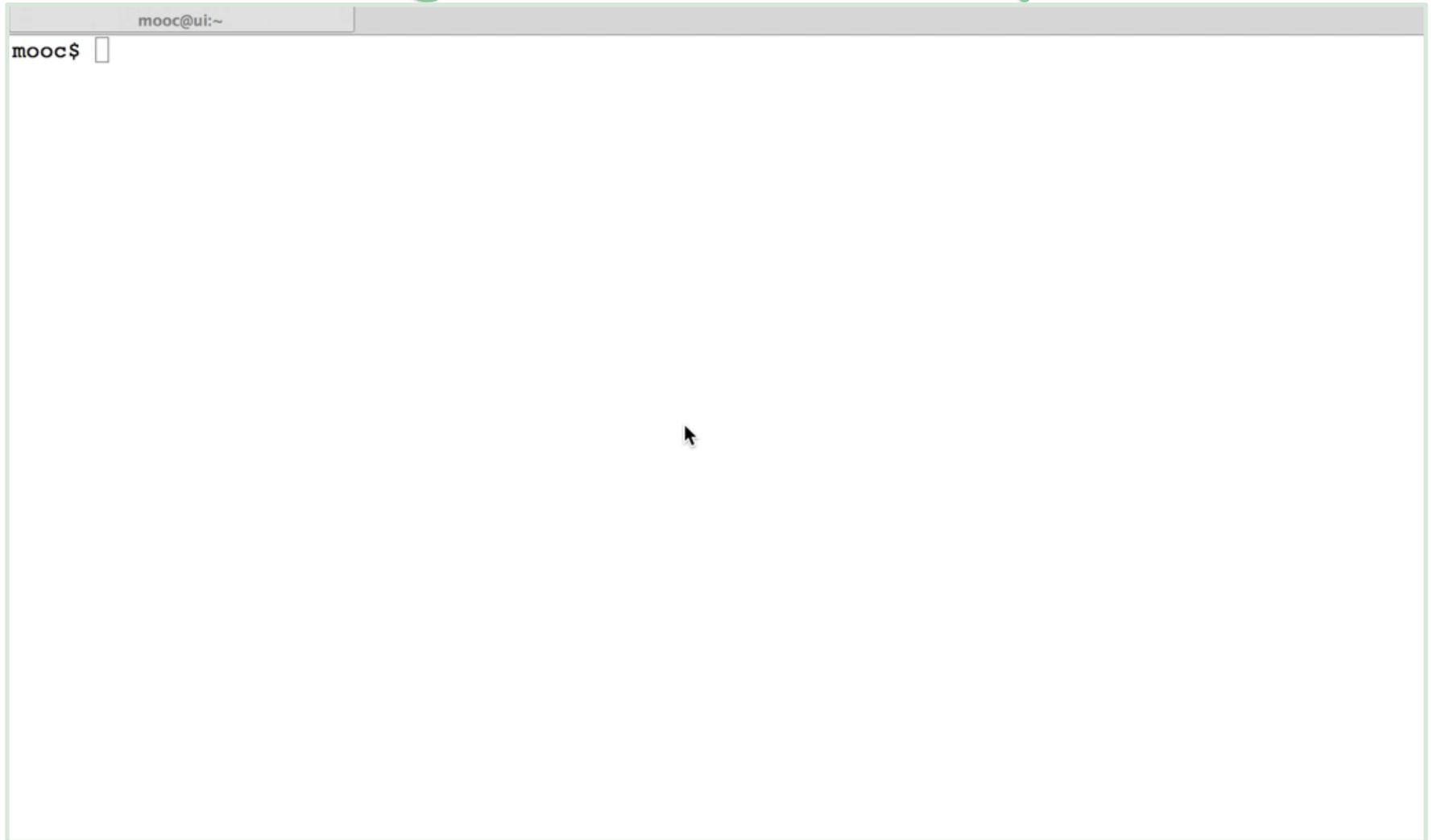
lfn- create first file (Optional)

```
mooc@ui:~  
mooc$
```

Data Requirements Attribute in jdl

```
DataRequirements = {  
    [  
        InputData = {"lfn:/grid/tutor/mooc/test.txt"};  
        DataCatalogType = "DLI";  
        DataCatalog = "http://lfc.grid.sara.nl:8085";  
    ]  
};  
DataAccessProtocol = {"gsiftp"};
```

Large files submit – lfn input



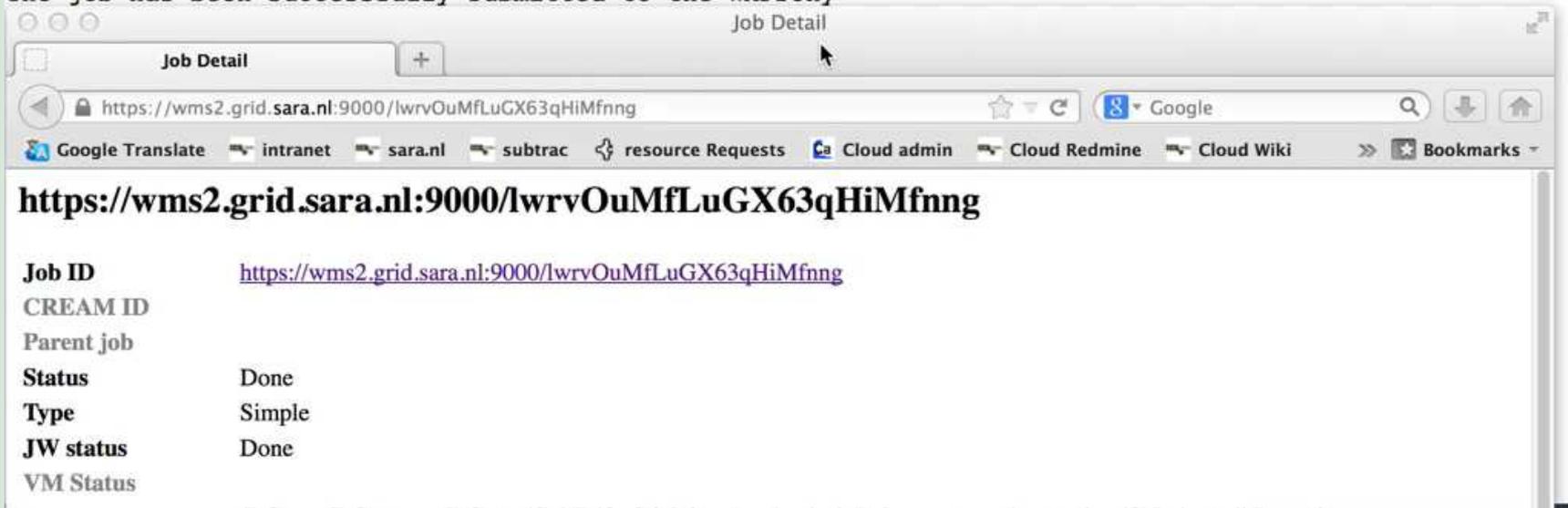
Large files lfn - Output

```
moooc@ul:~/moo...ples/4_largefiles
moooc$ cd mooocexamples/4_largefiles/
moooc$ ls
inputdata.jdl  scriptInputData.sh
moooc$ vim scriptInputData.sh
moooc$ vim inputdata.jdl
moooc$
moooc$ glite-wms-job-submit -d $USER -o /home/moooc/mooocexamples/jobIDs inputdata.jdl

Connecting to the service https://wms2.grid.sara.nl:7443/glite_wms_wmproxy_server

===== glite-wms-job-submit Success =====

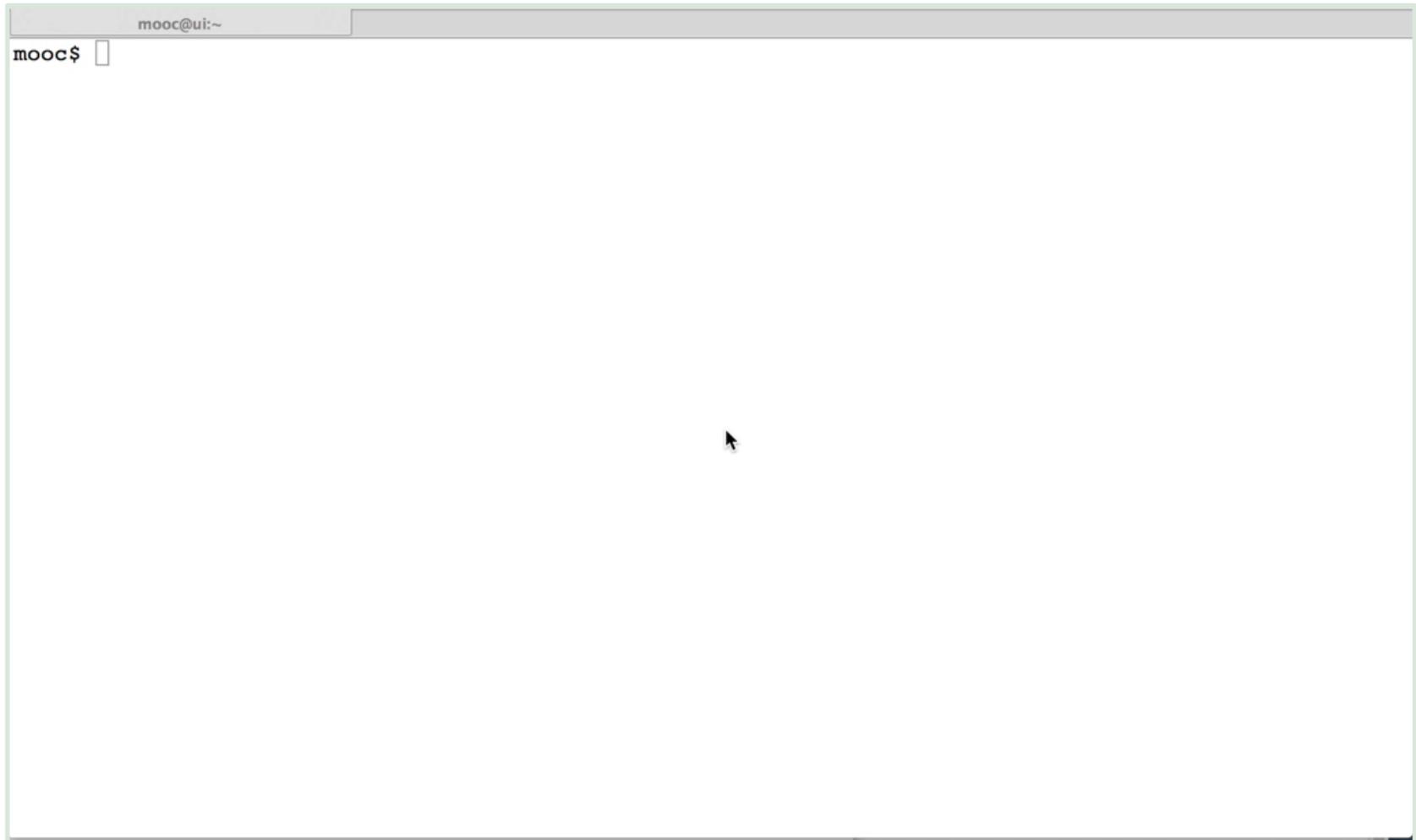
The job has been successfully submitted to the WMPProxy
```



The screenshot shows a web browser window titled "Job Detail" with the URL <https://wms2.grid.sara.nl:9000/lwrvOuMfLuGX63qHiMfnng>. The browser's address bar and bookmarks are visible. The page content displays the following job details:

Job ID	https://wms2.grid.sara.nl:9000/lwrvOuMfLuGX63qHiMfnng
CREAM ID	
Parent job	
Status	Done
Type	Simple
JW status	Done
VM Status	

Large files submit – srm input



Large files srm - output

```
moo@ui:~/mooexamples/6_WNtoSE
=====

moo$ glite-wms-job-output --dir . -i /home/moo/mooexamples/jobIDs
-----
1 : https://wms2.grid.sara.nl:9000/yMB2XrBtulqFp74r0NP5EA
2 : https://wms2.grid.sara.nl:9000/qBqkHRsTYdY3_nSkJfuObA
3 : https://wms2.grid.sara.nl:9000/h-29Nk9Zrb6spPEe2J--LA
4 : https://wms2.grid.sara.nl:9000/lwrvOuMfLuGX63qHiMfnng
5 : https://wms2.grid.sara.nl:9000/cavYvt3TdIX03xYKK2CVdg
6 : https://wms2.grid.sara.nl:9000/32ZYV9xEaJpmtXfvvPhPA
7 : https://wms2.grid.sara.nl:9000/TY6zFPpbLFXy8q2nK2wKuw
a : all
q : quit
-----
Choose one or more jobId(s) in the list - [1-7]all (use , as separator or - for a range): 7

Connecting to the service https://wms2.grid.sara.nl:7443/glite_wms_wmproxy_server

=====

JOB GET OUTPUT OUTCOME

Output sandbox files for the job:
https://wms2.grid.sara.nl:9000/TY6zFPpbLFXy8q2nK2wKuw
have been successfully retrieved and stored in the directory:
/home/moo/mooexamples/6_WNtoSE/moo_TY6zFPpbLFXy8q2nK2wKuw

=====
```

Summary

❑ Interaction with the WMS

- The InputSandbox/OutputSandbox should only be used for small amounts of data.
- Large files should be on SEs.

❑ jdl file: `InputData = "lfn:/grid/tutor/MyFile";`

- The lfn's needed by the job as an input to the process
- Tells RB to schedule job on CE close to SE holding the file

❑ `SE=srm.grid.sara.nl;`

- location of a SE where the output data will be stored

❑ `DataAccessProtocol="gsiftp";`

- The list of protocols that the application is able to “speak” for accessing files listed in the InputData