Data management on the Grid
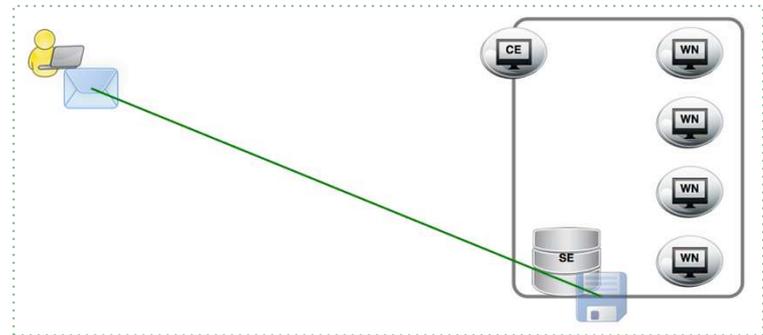
Outline

□ **Data Management overview**

- Sandbox

- Logical File Catalog (LFC)

- Srm
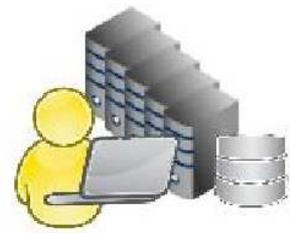
- Gridftp protocol

□ **Grid storage tools:**

- Mapping lfn and physical files

- Glite

- Lcg

- Lfc

- Srm

Data Management (1)

- **Sandbox**
  - Easy to use, but
  - Works for small files
  - Network connection limitations

- **Logical File Catalog**
  - Virtual interface to Grid Storage
  - Transparent file replication to multiple sites

- **Storage Resource Manager (srm)**
  - Interface for the disk/tape storage
  - Staging

- **Gridftp protocol**

Data Management (2)

- ❑ **Storage Element**
  - ▪ $ lcg-infosites --vo tutor se

- ❑ **Replication**
  - ▪ Jobs run close to the data
  - ▪ Reduces network traffic

- ❑ **Storage URL or SURL**
  - ▪ SE's identify files by SURL
  - ▪ SURL's are not very descriptive,eg.

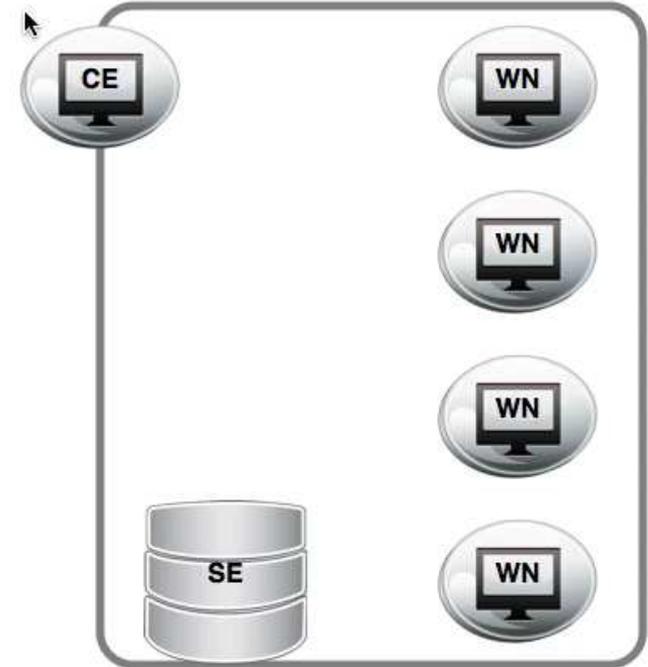   < srm://gb-se-ams.els.sara.nl/dpm/els.sara.nl/home/tutor/generated/2013-10-28/file2868595b-ae21 >

- ❑ **Logical File Catalog or LFC**
  - ▪ It is a database, *not* a storage system
  - ▪ Descriptive names to files

Data transfer

- **http://mooc-inst.sara.cloudlet.sara.nl/mooc/data_transfer.html**

Data transfer with central storage

lcg vs lfc

❑ **lcg-...**

  ▪ Operate on the data itself

  ▪ lcg-cr :

    ✓ uploads a file to an SE *and* registers this file in the LFC under a certain name

  ▪ Lcg command line utilities

    ✓ https://grid-deployment.web.cern.ch/grid-deployment/documentation/LFC_DPM/lcg_util/

❑ **lfc-...**

  ▪ Operate on the LFC, and don't manipulate data

  ▪ lfc-mkdir :

    ✓ creates a new directory in the LFC