The GRID

E Wuensch
JUNE 2008

as I know it

Since 2004 DESY operates the DESY Production Grid
which is an official Grid site in the LHC Computing Grid (LCG).

The Grid activities are carried out in the context of
- the EU-project Enabling Grids for E-siencE (EGEE),
- the project Physics at the Terascale
- and the national project D-GRID.

DESY provides Grid resources as Tier-2 centre
for the LHC experiments
ATLAS, CMS, and LHCb. DESY supports the EGEE VO 'biomed'.
The HERA experiments and the International Linear Collider Community
(ILC) use the DESY Grid

Part I

- WN Worker Node
- Software on WN or my application on the WN
- CE Computer Element
- SE Storage Element
- LCG Middleware - I can access the GRID files
- gLite Middleware - I can submit jobs

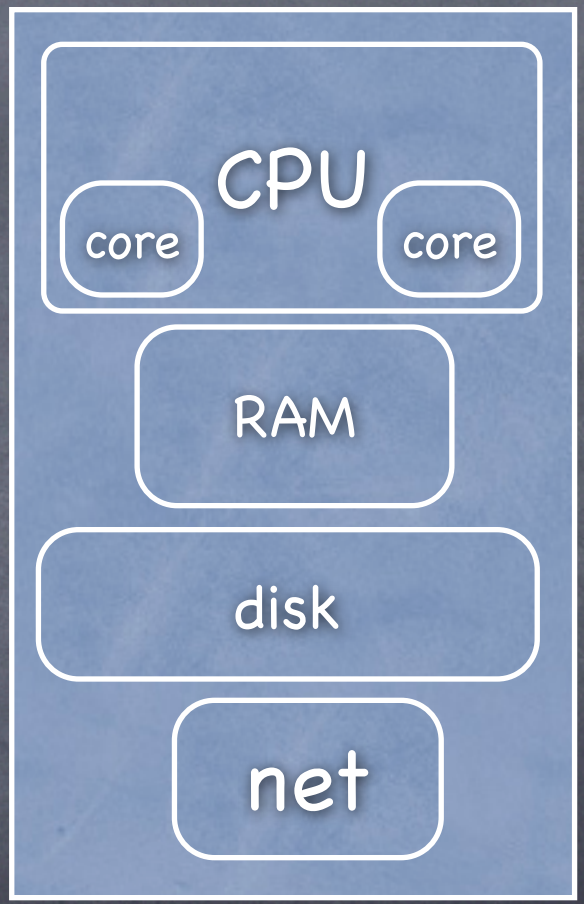
Part II

- UI my userinterface - scripts ..
- submit a request
- TOP Topologies
- APP other applications
- ERRORs - Nasty Business
- END - What is our request doing ?

WN Worker Node

- A WN comprises a classical computer without keyboard, monitor and other devices

- shared RAM bigger than 2 GB
- shared DISK, more than 6 GB
- GB Ethernet



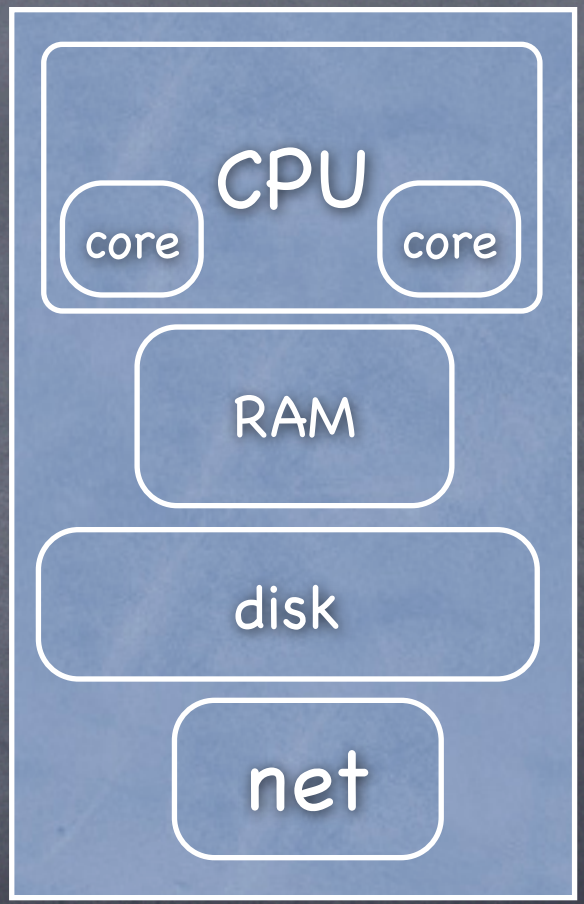
Software on WN

- Linux Operating System, delivered, customized by site, mostly SL4
- selection of rpm, configured, tuned
- selection of libs, applic, tools, configured, tuned
- network configuration

- SANDBOX: container for input & output files

- USER/Appication Environment: your 1st scripts, applications, libs, db, input files

- INIT: executing 1st scripts, adjusting limits & watchdogs, download applications, libs, db, input files



Software on WN - running phase

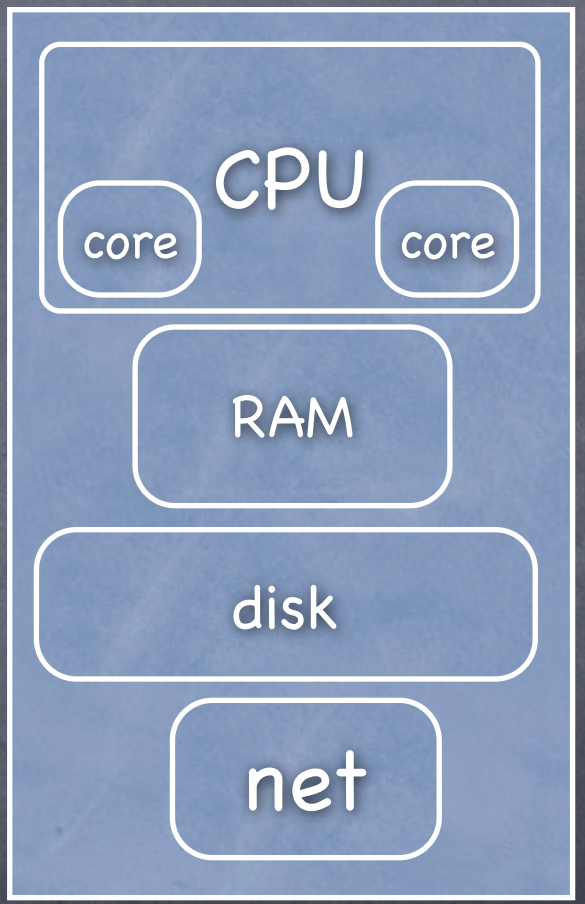
- As a minimum only one application is running and producing some output

- SANDBOX: container for input & output files
The script/application fetches the input from the input-sandbox; after processing a selection of output-files/logs may be stored into the output-sandbox.

- H1 mc: we watch the groth of the output-file and logfiles.

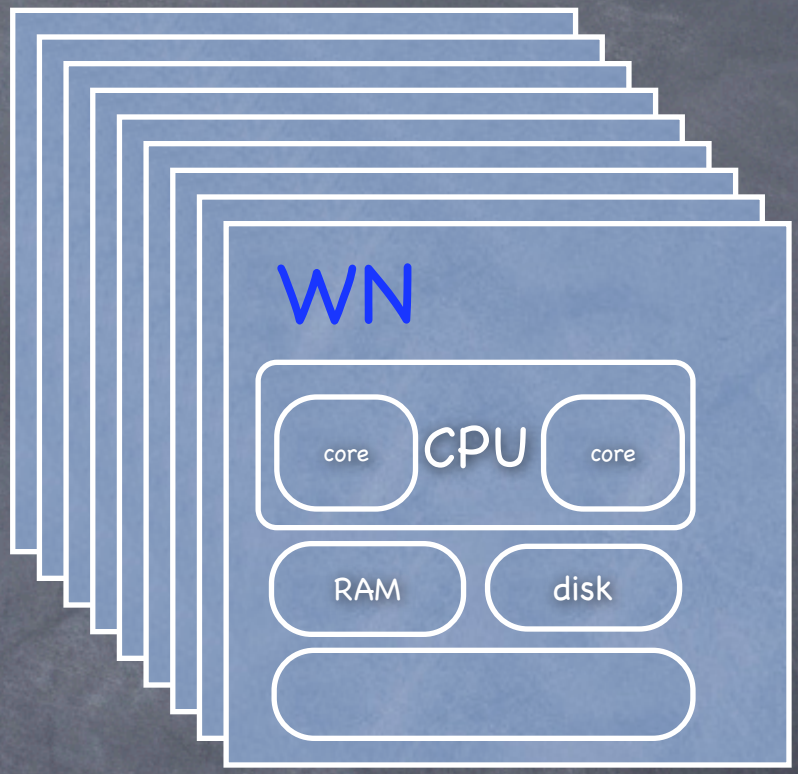
- OUTPUT: resulting in large filesizes will be stored directly on a SE

- ERROR: no output, to big files, overtime conditions will abort the job



CE Computer Element

- eg DESY builds an CE

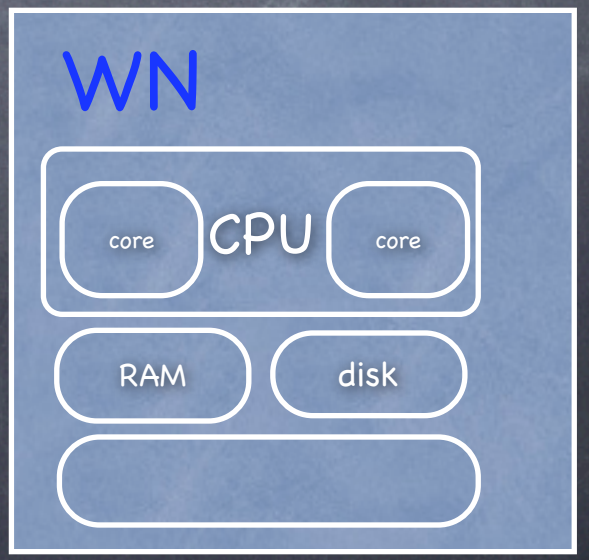


- CE: initializes, controls and monitors the WN via a batch-queue system, transfers sandboxes

- H1 mc: we watch the groth of the output-file and logfiles.

- OUTPUT: resulting in large filesizes will be stored directly on a SE

- ERROR: no output, to big files, overtime conditions will abort the job

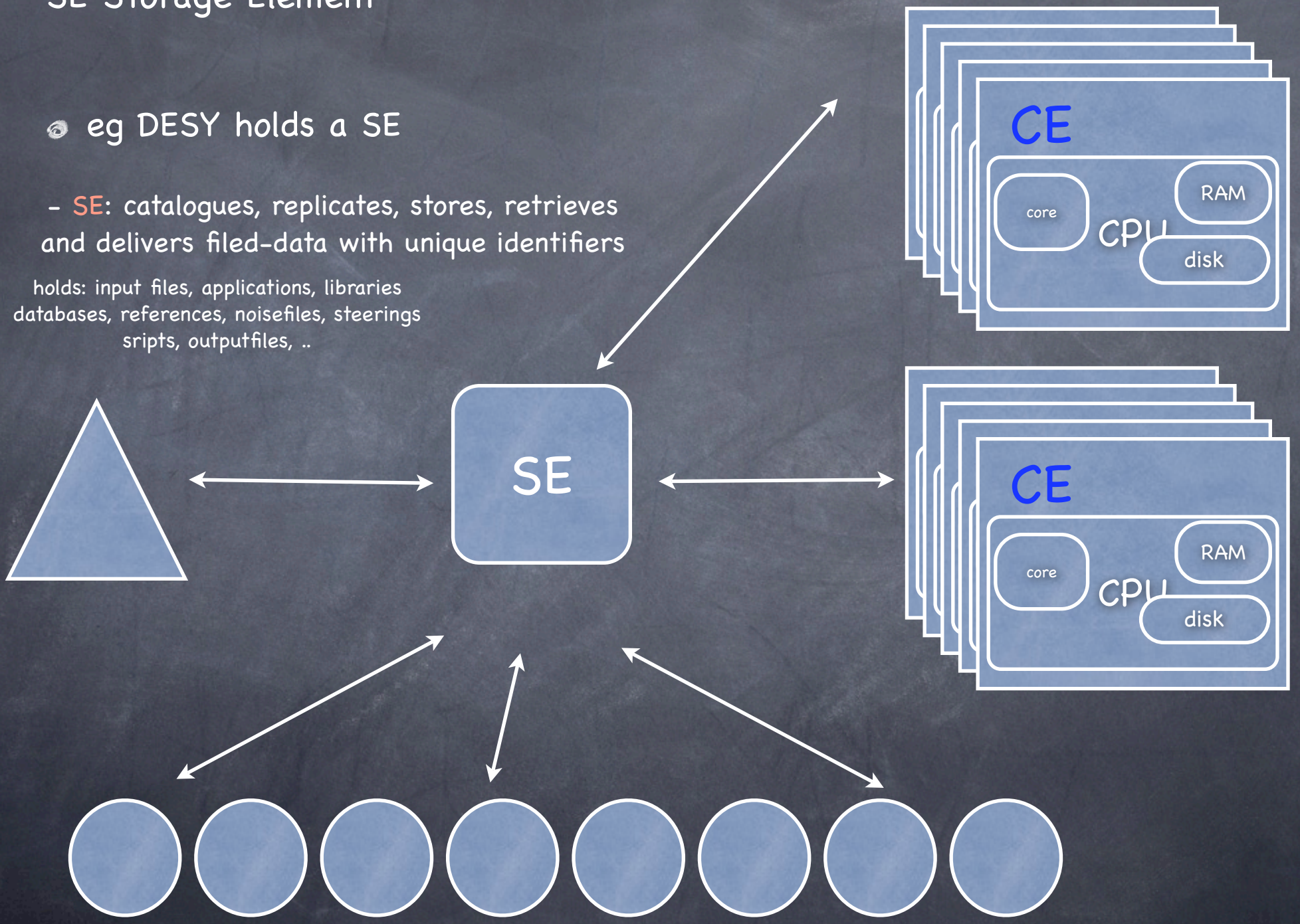


SE Storage Element

- eg DESY holds a SE

- SE: catalogues, replicates, stores, retrieves and delivers filed-data with unique identifiers

holds: input files, applications, libraries databases, references, noisefiles, steerings sripts, outputfiles, ..



- huge disk space: ~100 PetaBytes

UI my userinterface - scripts ..

- UNIX account - token only 24 h
- GRID CERTificate - valid 1 y, initialise PROXI, valid 39 d
- Membership in Virtual-Organisation VO (e.g. hone for H1)
- scripts to submit, monitor, control jobs and collect data
- configurations: which CE support our VO hone

  lcg-infosites --vo hone -f ce
- steerings - how to run h1simrec

Proxies are certificates signed by the user, or by another proxy, that do not require a password to submit a job. They are intended for short-term use, when the user is submitting many jobs and cannot be troubled to repeat his password for every job.

LCG Middleware - I can access the GRID files

LCG is based on gLite middleware

```
lfc-ls -l /grid/hone/h1mc/input/5525/
-rw-rw-r--   1 44205    2049            30040721 Jun 21 17:30
RAPGAP.NC.ELEC920.CTEQ5L.Q2G100.W150.norm.Z0398.tar.gz
-rw-rw-r--   1 44205    2049            36491901 Jun 21 17:30
RAPGAP.NC.ELEC920.CTEQ5L.Q2G100.W150.norm.Z0399.tar.gz
-rw-rw-r--   1 44205    2049            31107935 Jun 21 17:30
RAPGAP.NC.ELEC920.CTEQ5L.Q2G100.W150.norm.Z0400.tar.gz
[h1grid01] /x01/usr/wuensch/h1mc.grid_production.glite-wms/h1mcDaemon $
```

lfc-ls -l /grid/hone/h1mc/input/5520

gLite Middleware - I can submit jobs

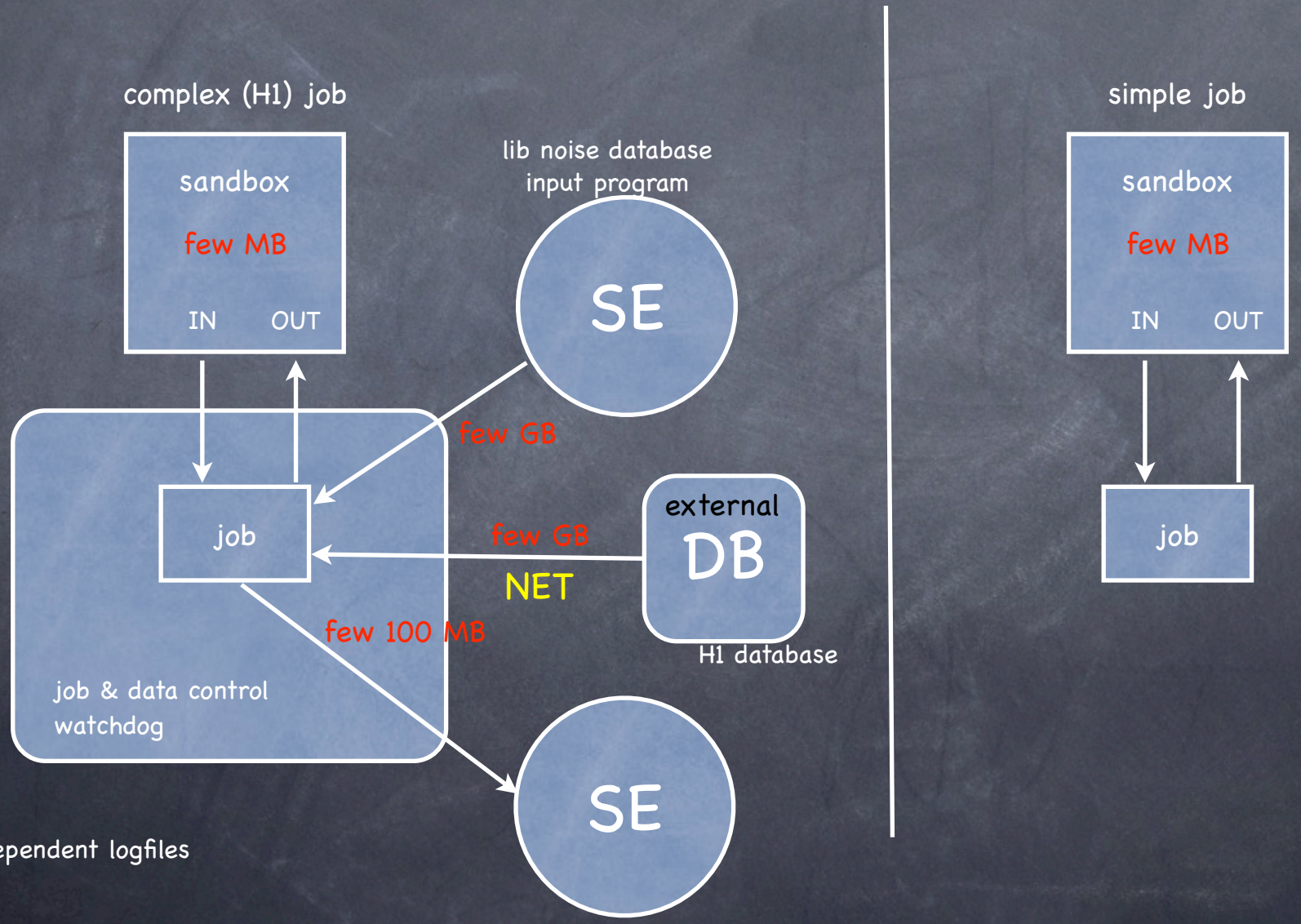
- glite-job-submit <jdl_file: h1mcjob.jdl >

```
h1mcjob.jdl
VirtualOrganisation = "hone";
Executable = "h1mcLauncher_perl.sh";
StdOutput = "std.out";
StdError = "std.err";

InputSandbox = {
"/x01/usr/wuensch/h1mc.grid_production.glite-wms/h1mcJobwrapper/h1mcLauncher_perl.sh",
"/x01/usr/wuensch/h1mc.grid_production.glite-wms/h1mcJobwrapper/download_perl_hone.pl",
"/x01/usr/wuensch/h1mc.grid_production.glite-wms/h1mcJobwrapper/h1mcJobwrapper.tar.gz",
"/x01/usr/wuensch/h1mc.grid_production.glite-wms/h1mcProduction/mcreq5520/checksums.txt",
"/x01/usr/wuensch/h1mc.grid_production.glite-wms/h1mcProduction/mcreq5520/jobs/0001/wrapper.conf",
"/x01/usr/wuensch/h1mc.grid_production.glite-wms/h1mcProduction/mcreq5520/jobs/0001/steering_mcrequest_5520"};

OutputSandbox = {
"std.out","std.err","h1mc.log","h1simrec.out",
"statistics.log","cmds.log","files.lfn","files.guid","logs.tar.gz","rescue.tar.gz"};
```

TOPOLOGY

Program & Data

IN: simple scripts, configs, steerings, H1-PERL



OUT: run dependent logfiles

ERRORs - Nasty Business

- no free CE sites available
- H1 share limit
- transfer too slow -> REPEAT
- checksum error -> automatic REPEAT
- waiting too long in queues
- memory getting low
- disk write error -> abort repeat
- SE write error - check SE catalog, download manually, manipulate catalog
- finally: data on ACS -> cleanUP GRID and UI spaces
- site problems -- communicate to GRID admins
- adapt software to changing GRID conditions
- open questions: requirement-specification, efficiency improvement

OUTLOOK

- The future in computing will be the GRID
- LHC -> data samples too big & numerious to transfer to local UI space
- GRID hold the data
- GRID holds the applications, libraries ..
- HARDWARE: 4 cores/cpu --> 256 cores/cpu
memory - diskspace -- shared, virtual, terrabytes
Gigabit and higher network connection
- universal and global FILESYSTEMS
- communicate to GRID via scripts & logfiles & presentation files
and probably webservices

scripts take a key role for reproduction and validation

http://grid.desy.de/