



Enabling Grids for  
E-science in Europe

*Nicosia, April 23<sup>rd</sup>, 2007*

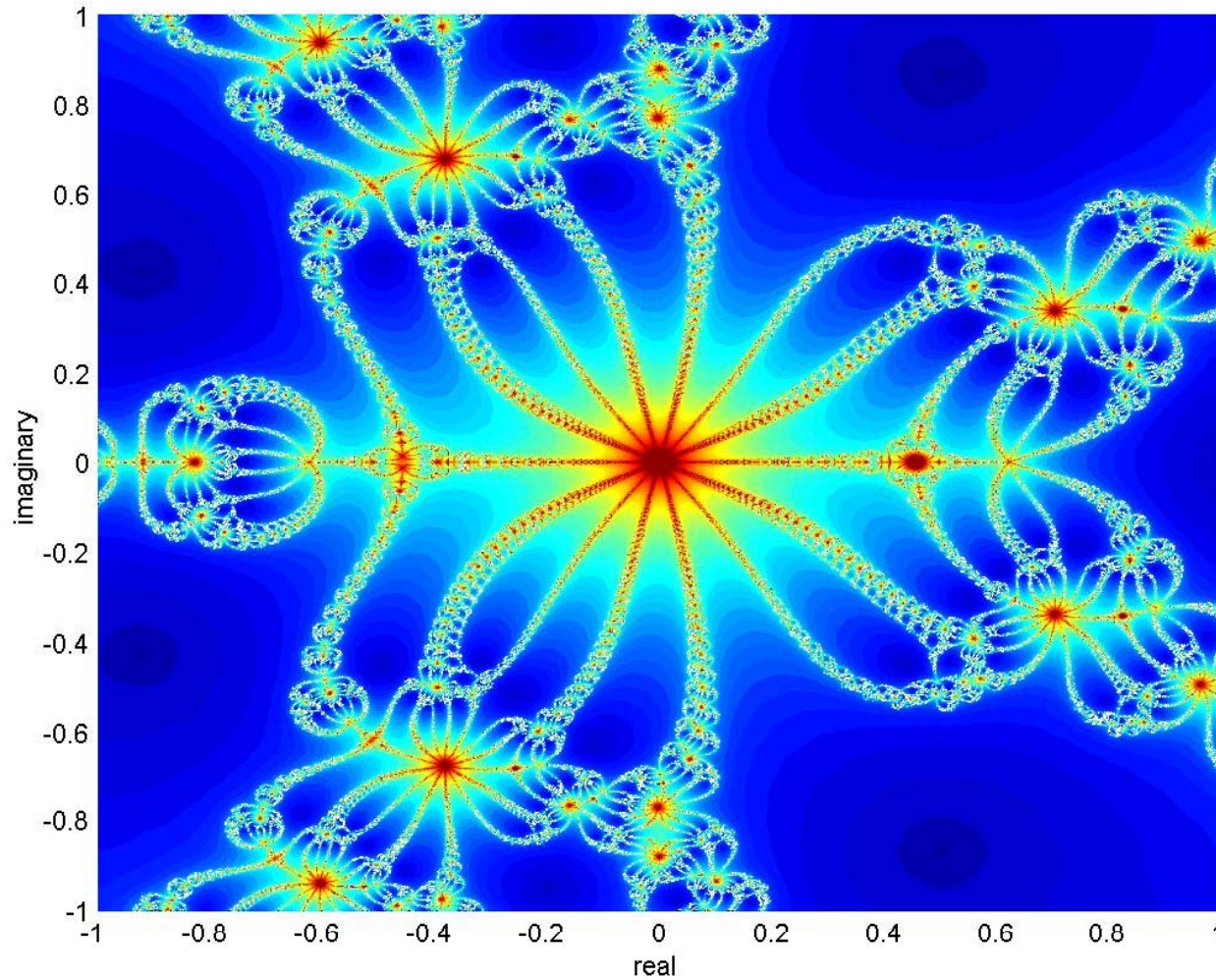
# Hands-on: Lifecycle of a Grid job

Fotis Georgatos <fotis@mail.cern.ch>  
Trainer, University of Cyprus

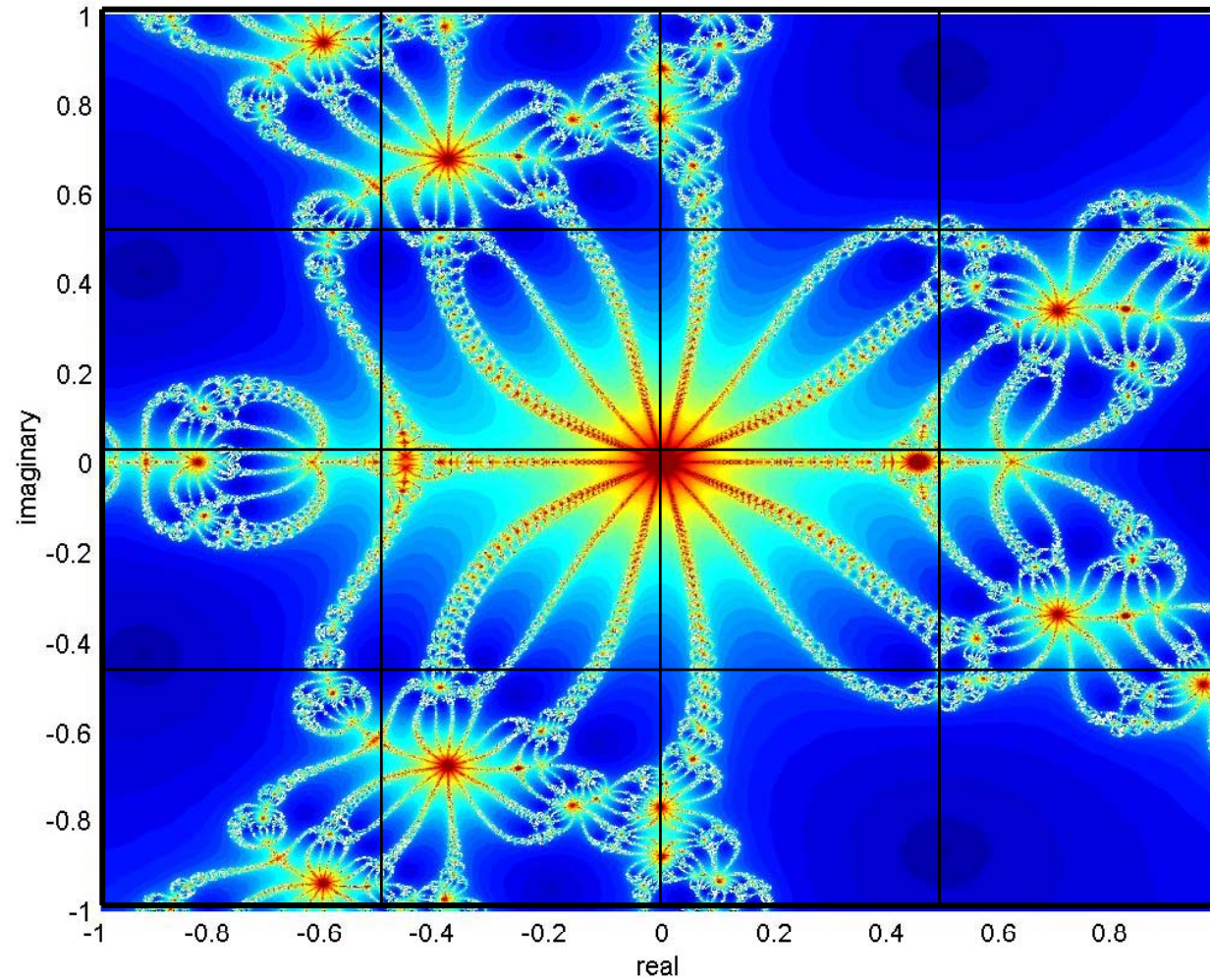


EGEE is a project funded by the European Union

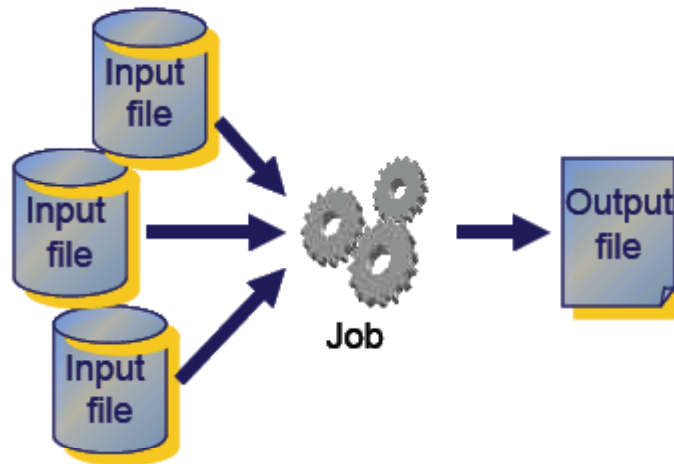
# An application at a single computer



# An application at the Grid

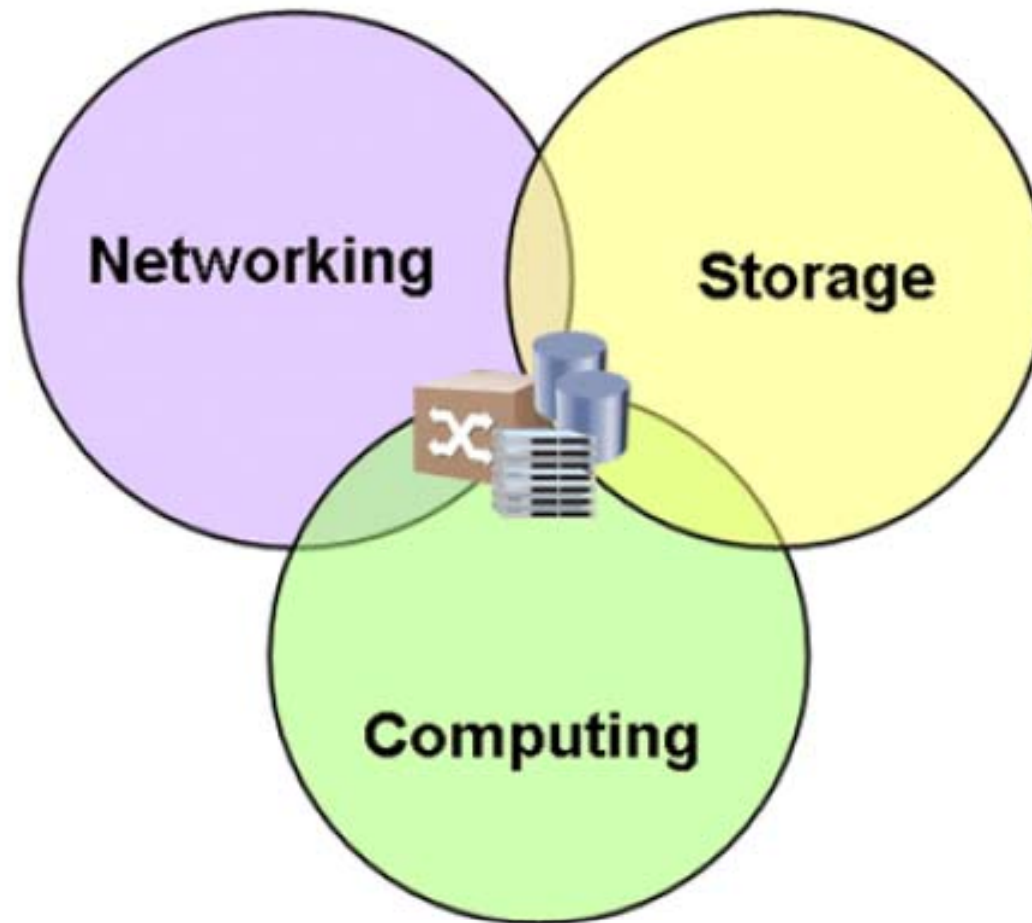


# Structure of a Grid application

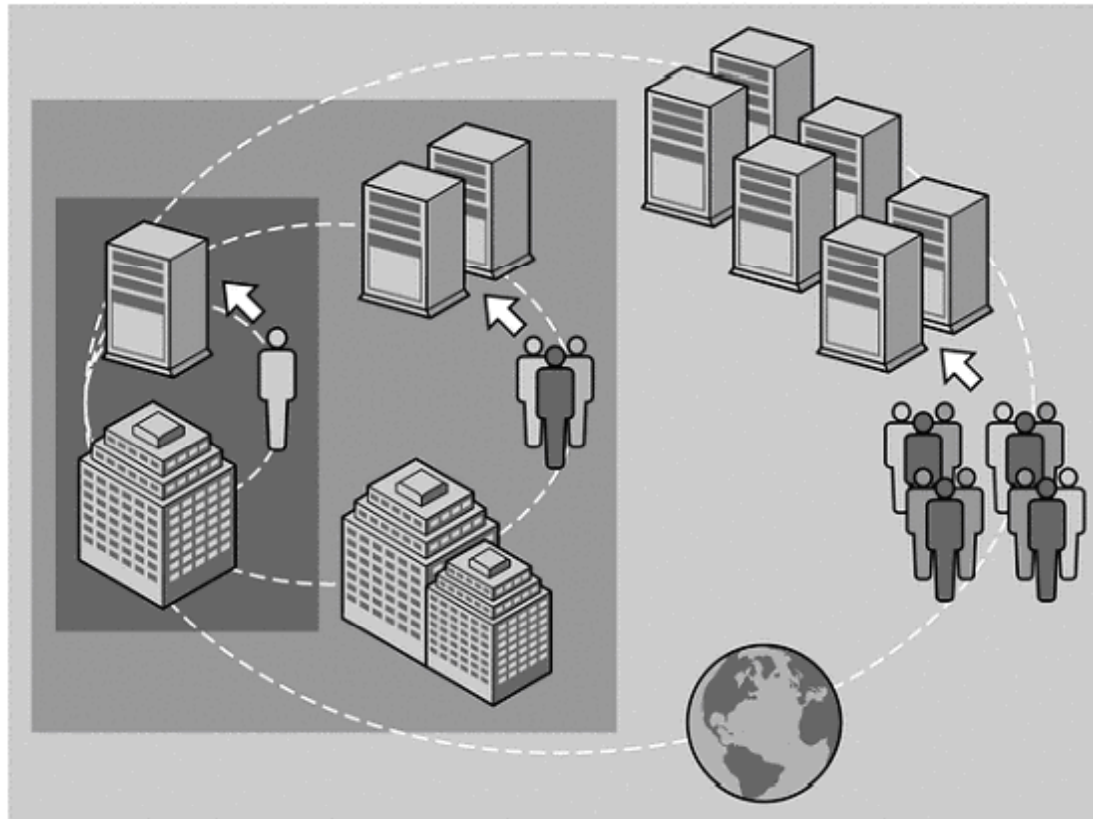


The user asks for the execution of an application at a remote system. The application handles input data (that exist on the Grid) and produces some output data, which are also stored on the Grid, quite likely on the Storage Element or the UI itself.

# Resources available on the Grid



# Outlook of the Grid

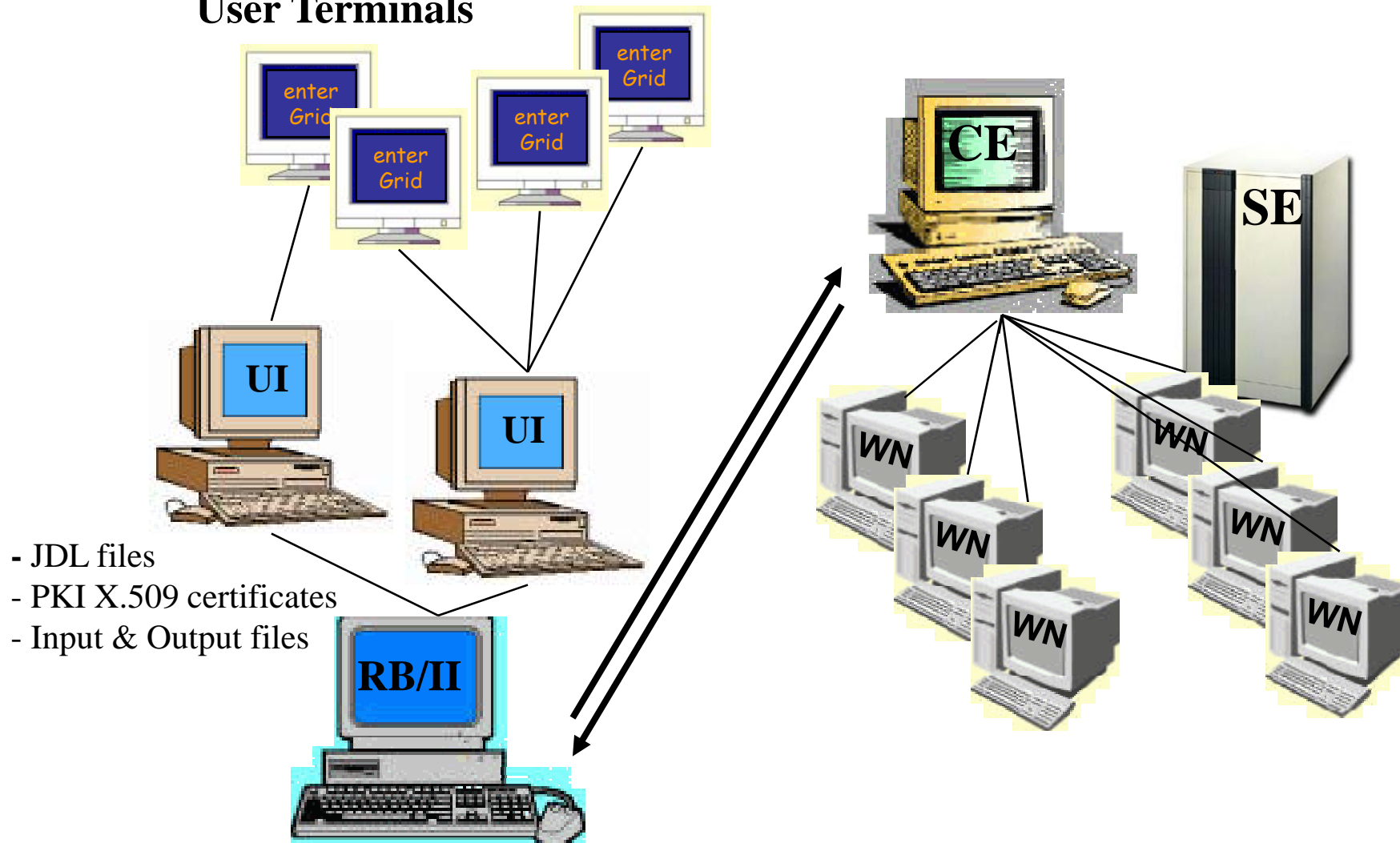


- The Grid (LCG in particular) tries to coordinate resources that are far greater than a local or campus-wide cluster.
- The expanded facilities are provided, but this implies a certain increased complexity in the middleware.
- It is required for optimal data and task management to follow a certain command set, which is specific to the LCG Grid.

■ Cluster Grid    ■ Campus Grid    ■ Global Grid

# Node cooperation on the Grid

## User Terminals

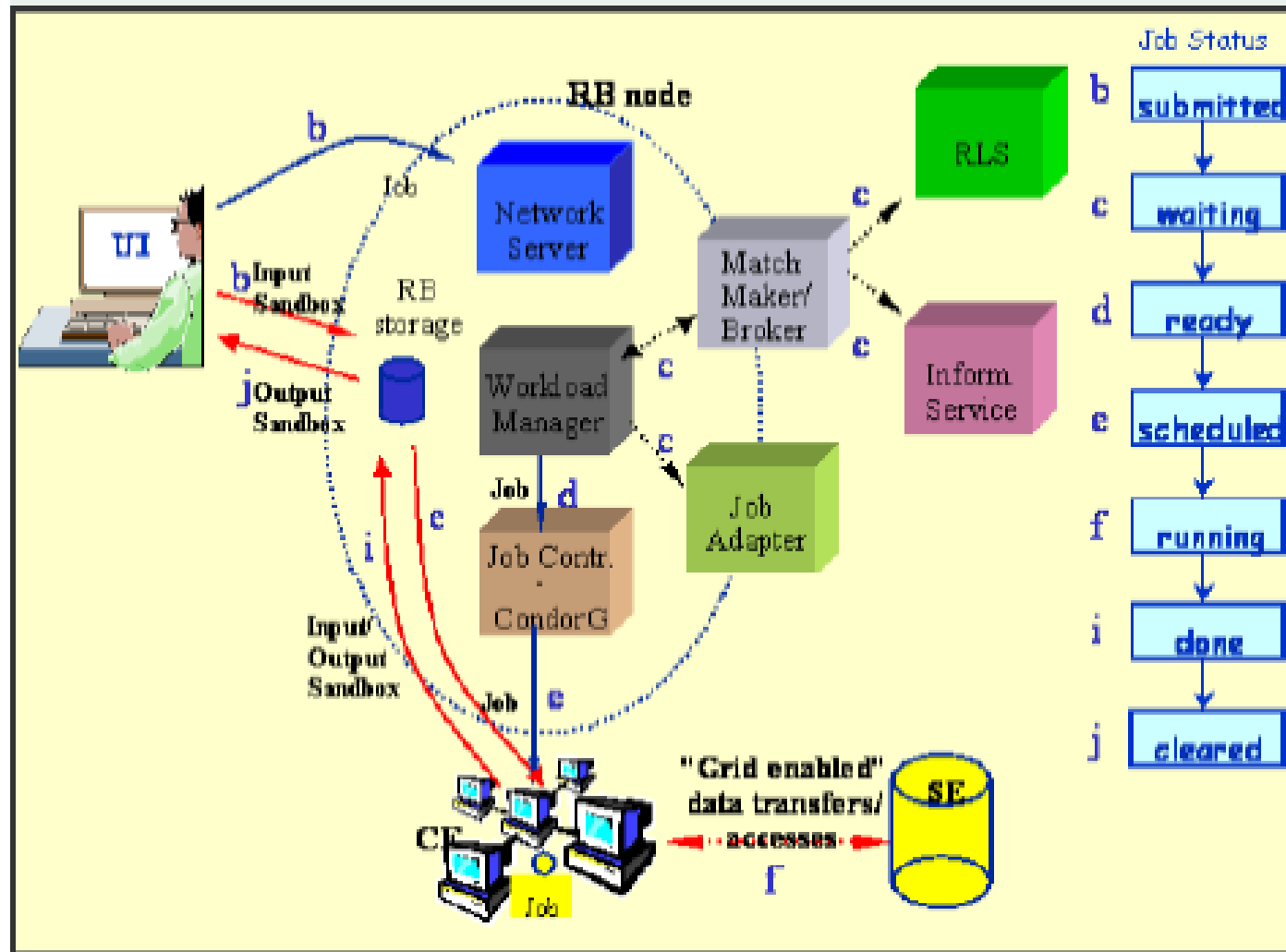


# The major node roles with LCG

- **UI: User Interface**
  - This is the system through which Users submit their jobs
  - It might be any Linux system, including a portable notebook.
- **RB: Resource Broker (or WMS)**
  - This is the node where User's jobs arrive, before being sent to a CE
- **CE: Computing Element**
  - This is the front-end node that provides access to a set of Worker Nodes
- **WN: Worker Node**
  - These are the nodes that provide the computational services to the Grid
- **SE: Storage Element**
  - These nodes provide access to disk and tape subsystems
- **BDII: Berkeley Database Information Index**
  - It provides the information «Where/who/what is the Grid»



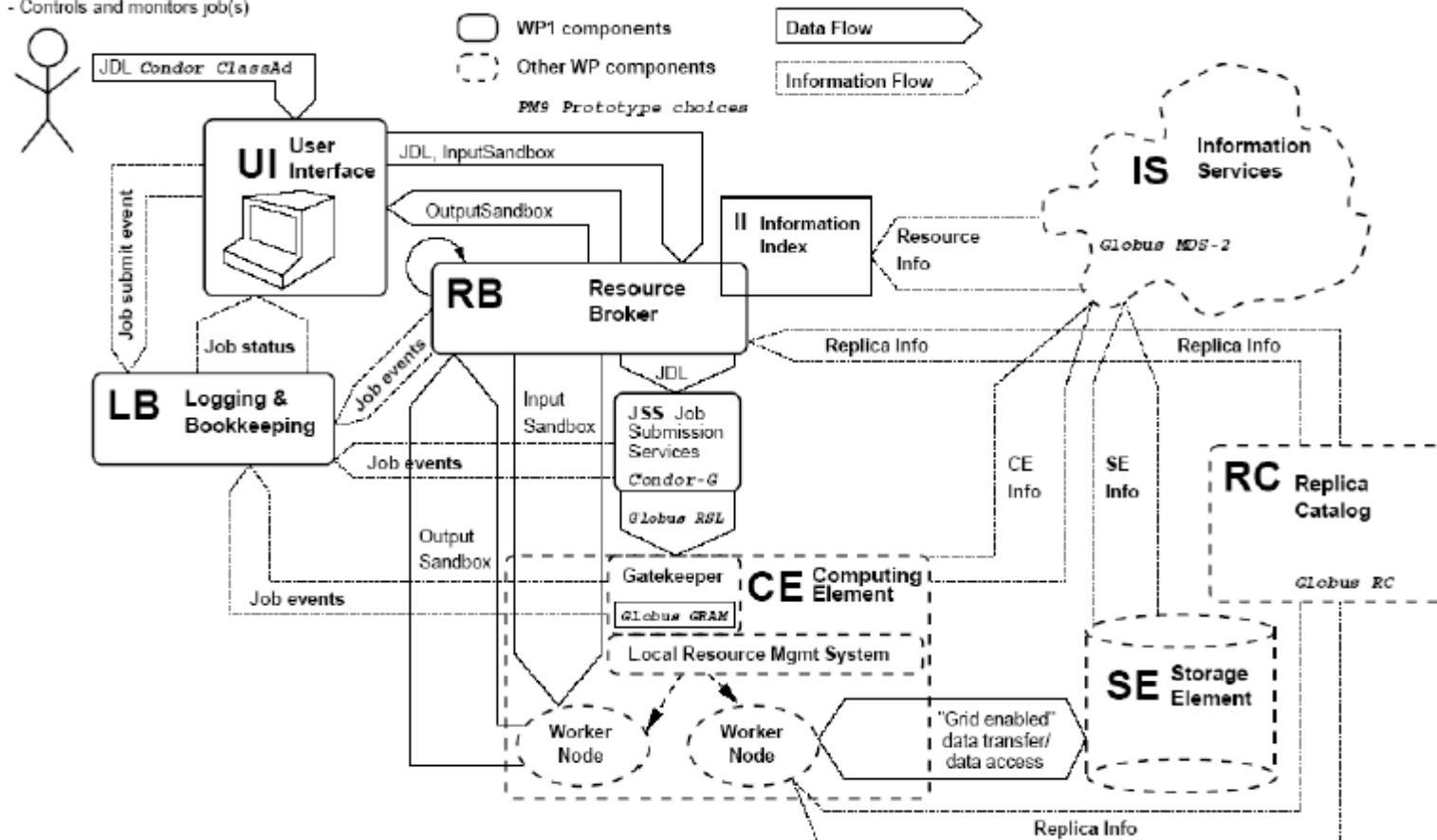
# Job Status of a Grid application



# Flow of a Grid application

## End User

- Specifies job using JDL
- Submits job using UI
- Controls and monitors job(s)



# Execution of a Grid application

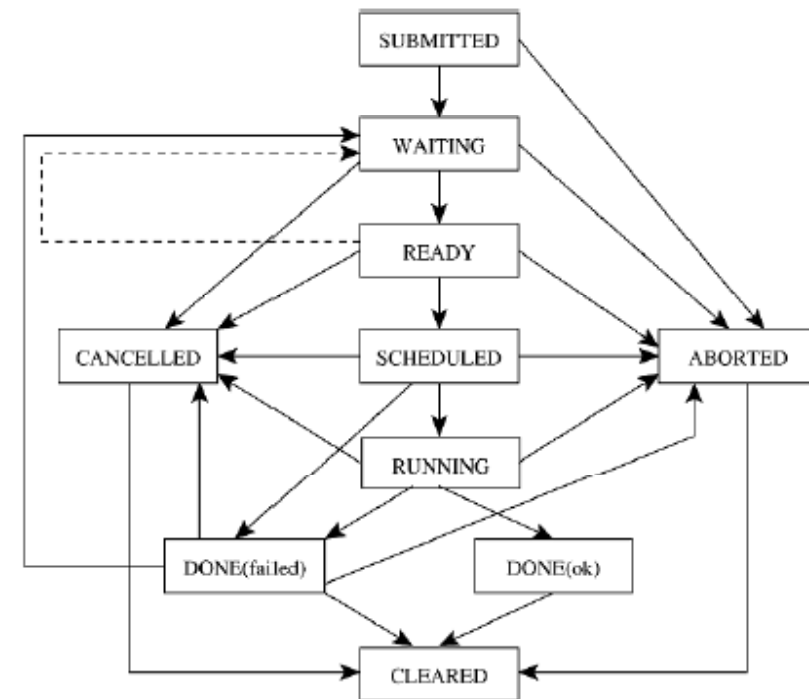
- Use of an ssh client, eg. putty
  - You can easily find it, fi. with google, or going directly to the address:  
<http://www.putty.nl/download.html>
- Access to a Linux system of an LCG User Interface type:
  - `ssh <MyAccount>@<MyUserInterface>`
  - `password: <mypassword>`
  - `ls -als .globus`
- Your certificate lives in the file `~/.globus/usercert.pem`
- Activation of the certificate (by default for 12 hours)
  - `voms-proxy-init --voms gilda`, the reply with the `<passphrase>` (“CYPRUS”)
  - `voms-proxy-info` , to confirm that the certificate is already active

# Execution of a Grid application

- **Job Description Language**
  - According to the standard of Condor classified advertisements
  - The information hereby included is used for optimal assignment of a job
  - Example of a **hostname.jdl**:
    - Executable = "/bin/hostname";
    - Arguments = "-f";
    - StdOutput = "std.out";
    - StdError = "std.err";
    - OutputSandbox = {"std.out","std.err"};
    - VirtualOrganisation = "gilda";
    - Requirements = other.arch="i386";
    - Rank = other.FreeCPUs;

# Execution of a Grid application

- edg-job-list-match hello.jdl
- lcg-infosites --vo gilda ce
- lcg-infosites --vo gilda se
  
- lcg-infosites --vo see ce
- lcg-infosites --vo see se
- lcg-infosites --vo atlas ce
- lcg-infosites --vo atlas se
- lcg-infosites --vo dteam ce
- lcg-infosites --vo dteam se
- lcg-infosites --vo biomed ce
- lcg-infosites --vo biomed se



# Execution of a Grid application

- `edg-job-submit -o myjobids.txt hello.jdl`
  - It sends the job to the Resource Broker
  - RB finds the most suitable CE and assigns the job to it
  - CE identifies a free WN and assigns the job to it
- `edg-job-status -i myjobids.txt`
  - `submitted`: The job has been submitted to the RB
  - `waiting`: The job awaits processing at the RB
  - `ready`: The job has been assigned to a CE, but is not yet into the LRMS
  - `scheduled`: The job is in the queue of the LRMS at the CE
  - `running`: The job is running within a Worker Node
  - `done`: The job is over at the RB and the collection by an UI is pending
- `edg-job-cancel` , is useful for canceling a job
- `edg-job-get-output -i myjobids.txt --dir .`

# Data Management on the LCG Grid

- Query to find a Storage Element which is accessible:
  - `lcg-infosites --vo gilda se`
- Depositing a file at a Storage Element:
  - `lcg-replica-manager cr -d <SE> -l lfn:<file> --vo=<myvo> file://`pwd`/<file>`
- Withdrawing a file from a Storage Element(s):
  - `lcg-replica-manager cp -l lfn:<file> --vo=<myvo> file://`pwd`/<file>`
- Replicating a file to another Storage Element:
  - `lcg-replica-manager rep --vo <myvo> -d <SE> lfn:<file>`
- Query regarding the replicas of a file (based on lfn):
  - `lcg-replica-manager lr --vo <myvo> lfn:<file>`

# Who can benefit from the Grid

- **Medical/Healthcare**
  - Imaging
  - Diagnosis & Treatment
  - Drug design (d2ol, for SARS, anthrax, embola etc)
- **Bioinformatics**
  - Study of the human and other genomes (genome@home)
  - Protein folding (folding@home, predictor@home)
- **Geological and climate applications**
  - Weather Forecasting
  - Climate Simulation (climate@home)
  - Ocean current analysis
  - Oil and Gas Exploration
  - Seismic Signal Analysis
- **Pharmaceutical, Chemical, Biotechnology**
  - atmospheric chemistry
  - systems biology
  - materials science
    - material interaction simulations
    - catalysis investigations)
  - molecular modeling
  - nanotechnology
- **Mathematics and Basic Research**
  - prime numbers (gimps/mprimes effort)
  - The verification of [Riemann's Hypothesis](#)
- **Business decision support**
  - Financial analysis
  - Portfolio optimization
  - Risk management applications
  - Route Optimization
    - Transportation
    - LAN and WAN Networking
  - Supply Chain and Demand Chain Optimization
  - Search and Retrieval (huge databases, data mining)
- **Electrical, Mechanical and Civil Engineering**
  - Energy production and distribution strategy optimization
  - Engineering and digital design
  - CAD / CAM
  - Construction verification against earthquakes
    - eg. finite elements method
  - Aerodynamic simulation (wind tunnel simulation)
  - Digital Rendering (raytracing, digital video synthesis)
- **Physics & Astrophysics**
  - High Energy Physics simulations and signal analysis
  - N-body problem simulation
  - space probe signal analysis (einstein@home)
  - radio telescope signal analysis (seti@home)
- **Computer Science**
  - Cryptography (distributed.net)
  - Search Engines (grud, a distributed Internet crawler)

**Many! <http://distributedcomputing.info/distrib-2003/distrib-projects.html>**



# Q & A

