DEISA - Easy Integration of HPC-Systems with UNICORE

Distributed European Infrastructure for Supercomputing Applications

Andrea Righi
a.righi@cineca.it

ETSI Headquarters, Sophia Antipolis, France, October 11-12, 2005

1st UNICORE Summit

Outline

- DEISA Introduction
- GRID: logical layout and UNICORE
- Account management
- UNICORE configuration
- Job submission
- UNICORE availability

Outline

- DEISA Introduction
- GRID: logical layout and UNICORE
- Account management
- UNICORE configuration
- Job submission
- UNICORE availability

DEISA

Distributed European Infrastructure for Supercomputing Applications

It is a consortium

- of leading national supercomputing centers in Europe

Focuses

- deploying an innovative GRID empowered infrastructure to enhance and reinforce High Performance Computing (HPC) in Europe
- set up a stable production infrastructure distributed across Europe

Its goal

- jointly build and operate a distributed terascale supercomputing facility
- deep integration -*using GRID technologies*- of high end national HPC infrastructures that are currently being deployed

Distributed European Infrastructure for Supercomputing Application

Joint Research Activities

Distributed European Infrastructure for Supercomputing Applications

JRA2 – Cosmology. (EPCC)
Migration and deployment on the inner infrastructure and its outer extension, of leading applications carried out by the European project VIRGO.

JRA1 – Material Sciences. (RZG) Migration and deployment of leading applications in this scientific area.

JRA3 – Plasma Physics. (RZG)
Migration and deployment of leading applications in this scientific area.

Joint Research Activities

The first six JRAs applications JRAs addressing modern and challenging application software engineering issues, paving the way in some cases for the heterogeneous extension of the facility. They also provide strong support to early adopters of the DEISA platform, in leading scientific areas vigorously supported by the national research organizations.

The last JRA is the only Grid R&D activity supported by the project.

JRA4 – Life Sciences. (IDRIS)
Migration and deployment of leading applications in Genomics and Health Sciences.

JRA5 – Industry. (CINECA)
Migration and deployment of leading applications in Industrial Computational Fluid dynamics.

JRA7 – Access to Resources in Heterogeneous Environments. (EPCC)
Development of a Heterogeneous Service Management software infrastructure based on OGSA standards.

JRA6 – Coupled Applications. (IDRIS)
Development, migration and deployment of leading multi-physics, multi-scale coupled applications in several areas of science and technology.

Service Activities in DEISA

Distributed European Infrastructure for Supercomputing Applications

- **SA1 – Network Operation and Support.** (FZJ leader)
  - Deployment and operation of a gigabit per second network infrastructure for an European distributed supercomputing platform. Network operation and optimization during project activity.
- **SA2 – Data Management with Global file systems.** (RZG leader)
  - Deployment and operation of global distributed file systems, as basic building blocks of the “inner” super-cluster, and as a way of implementing global data mlanagement in a heterogeneous Grid.
- **SA3 – Resource Management.** (CINECA leader)
  - Deployment and operation of global scheduling services for the European supercluster, as well as for its heterogeneous Grid extension.
- **SA4 – Applications and User Support.** (IDRIS leader)
  - Enabling the adoption by the scientific community of the distributed supercomputing infrastructure, as an efficient instrument for the production of leading computational science.
- **SA5 – Security.** (SARA leader)
  - Providing administration, authorization and authentication for a heterogeneous cluster of HPC systems, with special emphasis on single sign-on.

Service Activities in DEISA

Distributed European Infrastructure for Supercomputing Applications

- **SA1 – Network Operation and Support.** (FZJ leader)
  - Deployment and operation of a gigabit per second network infrastructure for an European distributed supercomputing platform. Network operation and optimization during project activity.
- **SA2 – Data Management with Global file systems.** (RZG leader)
  - Deployment and operation of global distributed file systems, as basic building blocks of the “inner” super-cluster, and as a way of implementing global data mlanagement in a heterogeneous Grid.
- **SA3 – Resource Management.** (CINECA leader)
  - Deployment and operation of global scheduling services for the European supercluster, as well as for its heterogeneous Grid extension.
- **SA4 – Applications and User Support.** (IDRIS leader)
  - Enabling the adoption by the scientific community of the distributed supercomputing infrastructure, as an efficient instrument for the production of leading computational science.
- **SA5 – Security.** (SARA leader)
  - Providing administration, authorization and authentication for a heterogeneous cluster of HPC systems, with special emphasis on single sign-on.

SA3: a three layer architecture

Distributed European Infrastructure for Supercomputing Applications

- **Basic services**: are those located closest to the operating system of the computing platforms and enable the operation of a single or a multiple cluster through local or extended batch schedulers and other cluster-like features.
- **Intermediate services**: are the first-level Grid services that allow access to an enlarged GRID empowered infrastructure, dealing with resource, network monitoring and information systems.
- **Advanced services**: use the previous layers to implement the global management of the distributed resources of the infrastructure.

SA3: a three layer architecture

Distributed European Infrastructure for Supercomputing Applications

- **Basic services**: are those located closest to the operating system of the computing platforms and enable the operation of a single or a multiple cluster through local or extended batch schedulers and other cluster-like features.
- **Intermediate services**: are the first-level Grid services that allow access to an enlarged GRID empowered infrastructure, dealing with resource, network monitoring and information systems.
- **Advanced services**: use the previous layers to implement the global management of the distributed resources of the infrastructure.

Outline

- DEISA Introduction
- GRID: logical layout and UNICORE
- Account management
- UNICORE configuration
- Job submission
- UNICORE availability

The GRID logical layout

Distributed European Infrastructure for Supercomputing Applications

Access				
Workflow management				
Job rerouting	Data staging	Brokering	Co-allocation	Multiple accounting
Policies implementation through the scheduler (workload, advance reservation, accounting)				
Resource manager				
OS and communication				
Hardware				

The GRID logical layout

Distributed European Infrastructure for Supercomputing Applications

Access				
Workflow management				
Job rerouting	Data staging	Brokering	Co-allocation	Multiple accounting
Policies implementation through the scheduler (workload, advance reservation, accounting)				
Resource manager				
OS and communication				
Hardware				

Middleware choice (1/3)

Distributed European Infrastructure for Supercomputing Applications

- Given the previous logical layout the DEISA partners have:
  1. Gather the project requirements
  2. Evaluated GTK and UNICORE
  3. Chosen UNICORE because:
     - Respect of requirements and usefulness
     - Project partner experience with the product
     - Product maturity

Middleware choice (2/3)

Distributed European Infrastructure for Supercomputing Applications

- UNICORE as access layer:
  - UNICORE adds a powerful complex workflow manager able to handle articulated operative chains spanning over **several heterogeneous machines**
  - Authentication and authorization
    - Single Sign-on with x509 certificates and username mapping (UUDB)

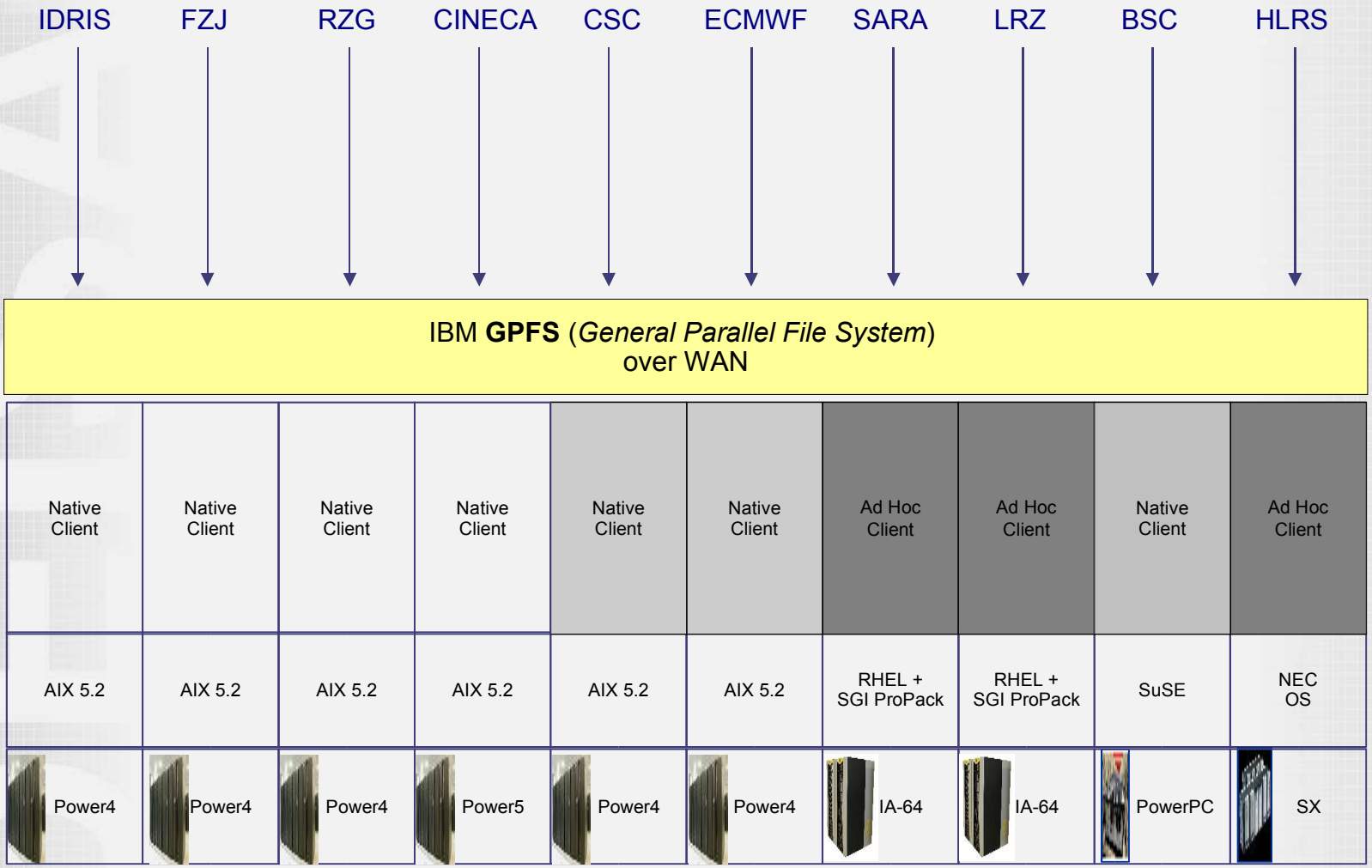
Middleware choice (3/3)

Distributed European Infrastructure for Supercomputing Applications

- Other UNICORE advantages
  - Scalability
    - Easy to add new sites
    - Separate Gateways and NJSs help to balance the workload of the metascheduling operations
  - Server logging and troubleshooting
    - Typically NJS log contains all the necessary informations to discover the problems
    - NJS log can be seen in the UNICORE client
  - Setting up UNICORE server mainly requires property file editing

The GRID physical layout: data management

Distributed European Infrastructure for Supercomputing Applications

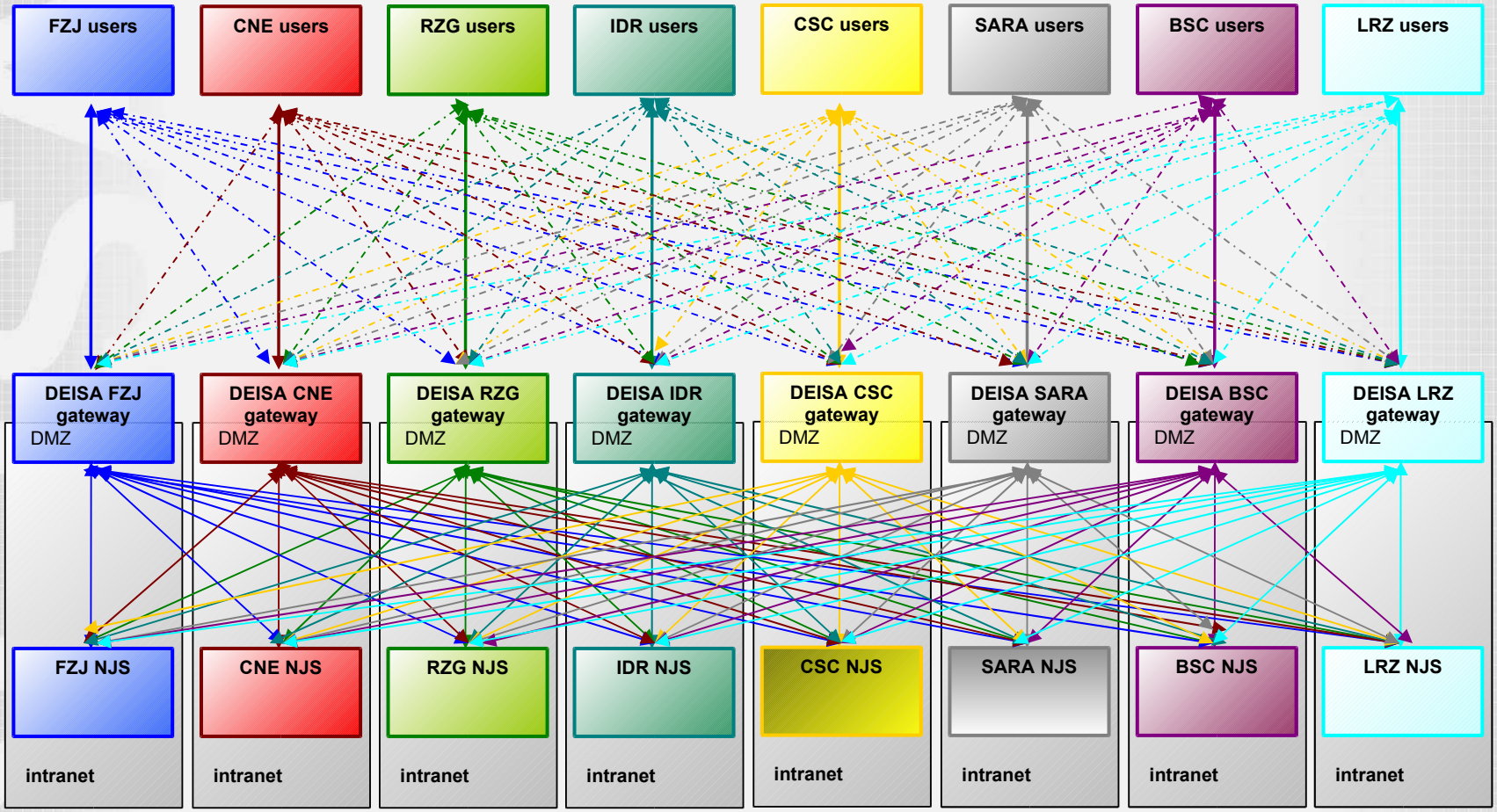


The GRID physical layout: resource management

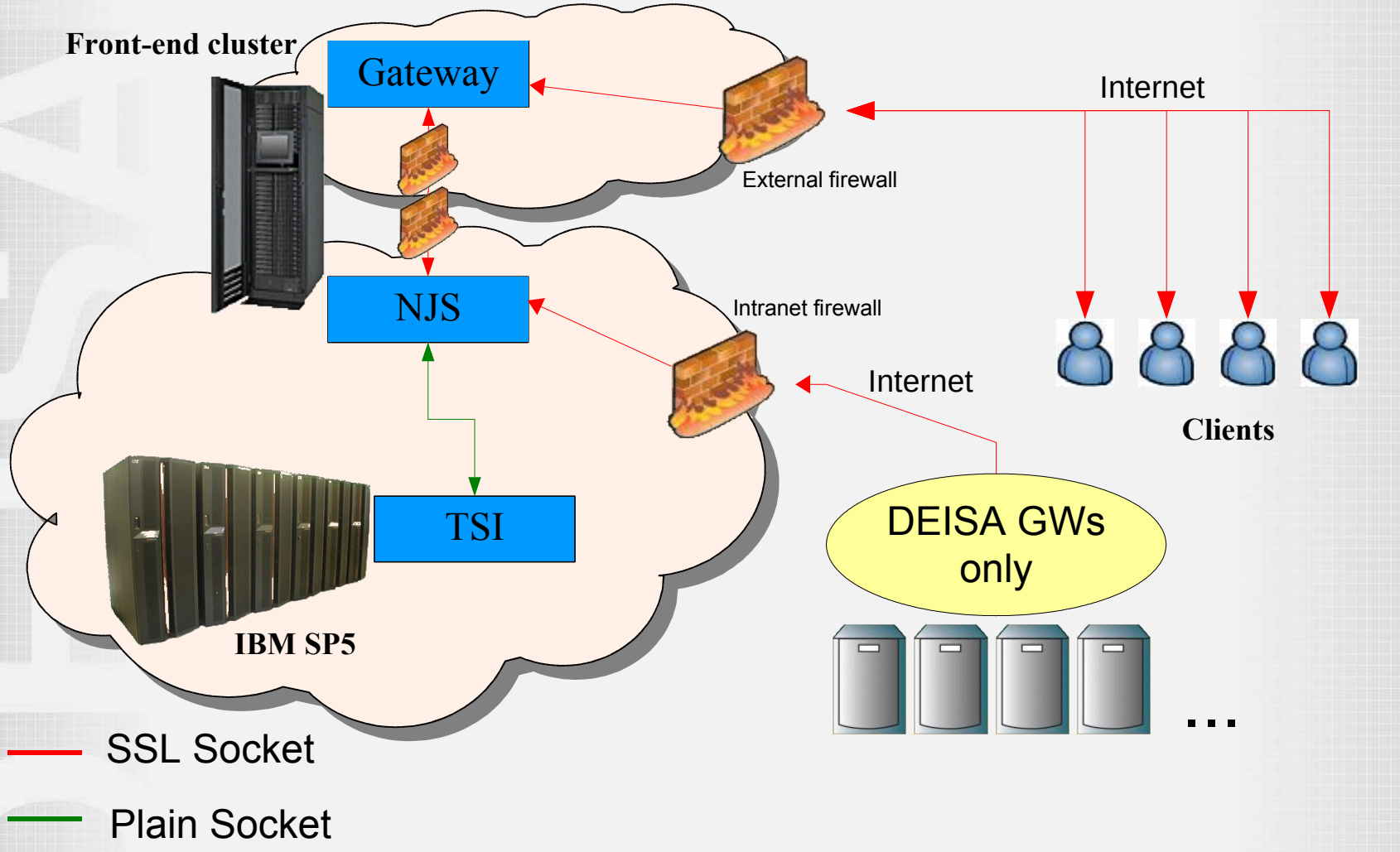
Distributed European Infrastructure for Supercomputing Applications

DEISA Unicore Infrastructure (architecture)

Distributed European Infrastructure for Supercomputing Applications



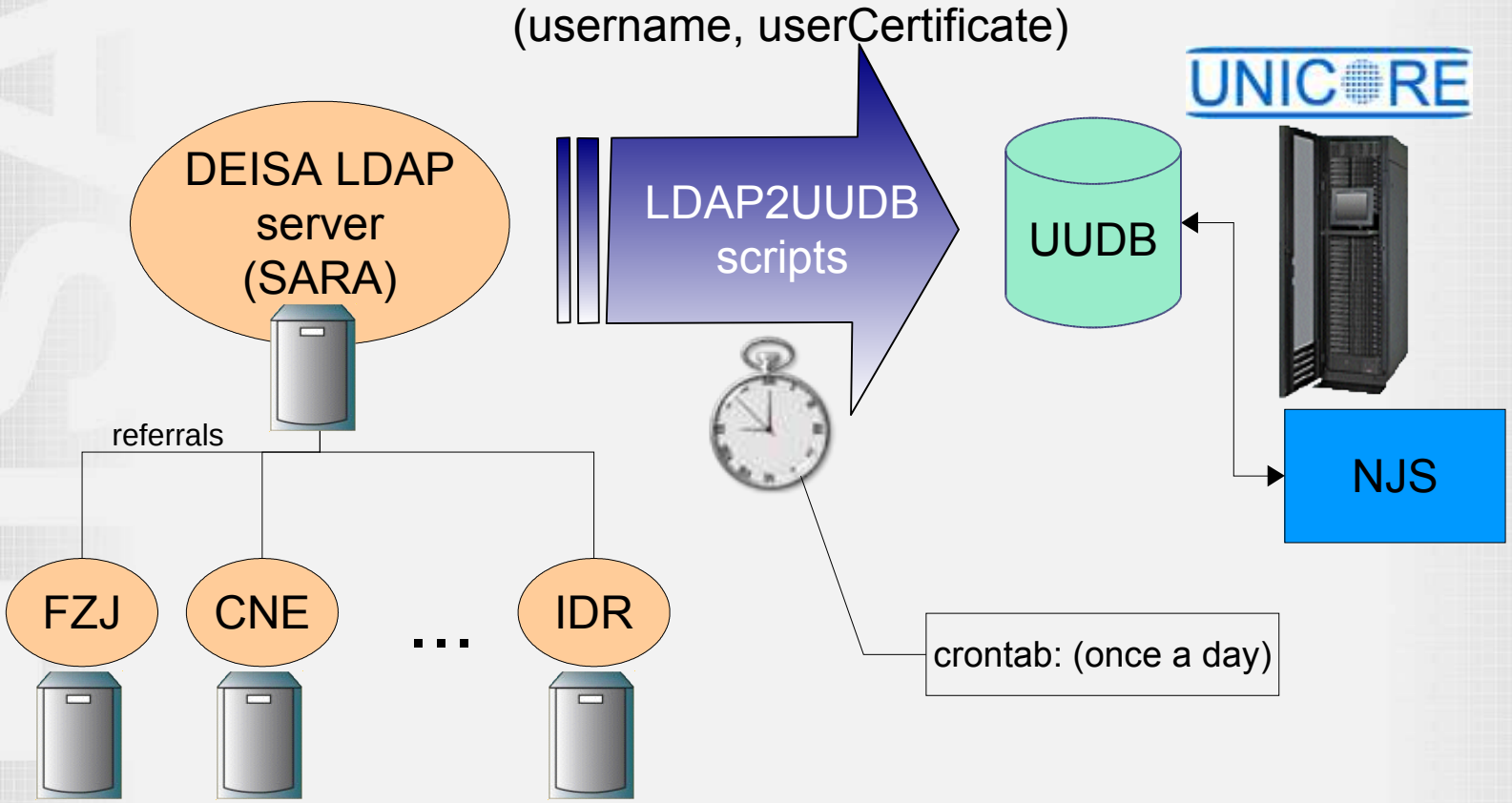
CINECA implementation



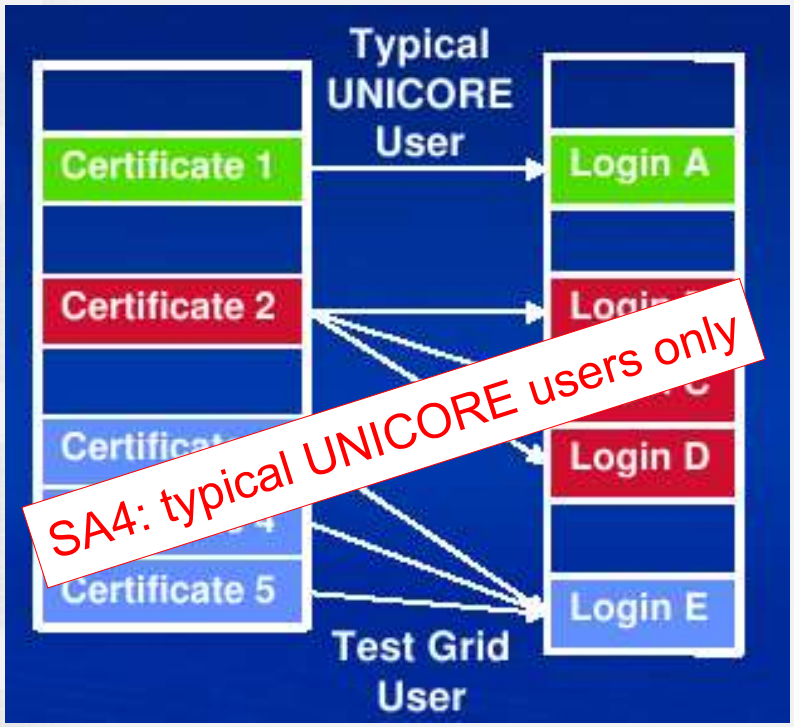
Outline

- DEISA Introduction
- GRID: logical layout and UNICORE
- Account management
- UNICORE configuration
- Job submission
- UNICORE availability

Account management: LDAP and UUDB



Multiple accounting



- User mapping with x509 certificate into more Xlogins
  - Useful for portals:
    - authentication and authorization is managed inside the portal
    - only a single entity (the portal) has to be authenticated in UNICORE
- More x509 certificates to a single Xlogin
  - Account sharing

Outline

- DEISA Introduction
- GRID: logical layout and UNICORE
- Account management
- UNICORE configuration
- Job submission
- UNICORE availability

DEISA Unicore configuration

Distributed European Infrastructure for Supercomputing Applications

- Public HTML page (Internet) to configure the server-side components:
  - Packages version,
  - Gateways and NJSs ports,
  - Certificates of the trusted NJSs,
  - Trusted CAs (CAcerts file),
  - Gateways XML file (to import into the UNICORE client),
  - PDF server configuration tutorial,
  - ... and contacts! (sa3-wg-unicore@deisa.org)

DEISA Unicore configuration

Distributed European Infrastructure for Supercomputing Applications

Centre	GW adress	NJS adress	CN of GW certificate	CN of NJS certificate	CA certificate	NJS certificate	Client-GW port	NJS-GW port
	zam177.zam.kfa-juelich.de 134.94.168.56	zam178.zam.kfa-juelich.de 134.94.168.57	DEISA Demo Gateway	DEISA Demo NJS	projects-ca-fz-juelich.pem	fzj-njs.pem	XXXX	XXXX
CSC	uni.csc.fi 193.166.7.126	hiekka.csc.fi 193.166.7.127	uni.csc.fi	hiekka.csc.fi	NorduGrid_Certification_Authority.pem	csc-njs.pem	XXXX	XXXX
sara	uni-gw1.sara.nl 145.100.29.233	uni-njs1.sara.nl 145.100.29.234	uni-gw1.sara.nl	uni-njs1.sara.nl	nikhef_ca_sara.pem	sara-njs.pem	XXXX	XXXX
RZG IPP	unigate.rzg.mpg.de 130.183.3.42	unicorn.rzg.mpg.de 130.183.8.180	unigate.rzg.mpg.de	unicorn.rzg.mpg.de	toplevel DFN-CA.pem gwdg-ca-ebene3-generic-ca.pem	njs-rzg.pem	XXXX	XXXX
CINECA Consorzio Interuniversitario	reunion.cineca.it 130.186.1.43	reunion.cineca.it 130.186.1.43	reunion.cineca.it	login011.sp4.cineca.it	INFN-CA.pem	njs.sp.sp4.cineca.it.pem	XXXX	XXXX
IDRIS CNRS	sirius.idris.fr 130.84.37.20	canopus.idris.fr 192.54.160.20		njszahir	Grid-FR.pem CNRS.pem	idris-njs.pem	XXXX	XXXX
BSC Barcelona Supercomputing Center	opsuni01.bsc.es	opsuni02.bsc.es	opsuni01.bsc.es	unicore02.bsc.es	projects-ca-fz-juelich.pem	bsc-njs.pem	XXXX	XXXX
lrz	unicore.lrz-muenchen.de 129.187.254.70	lxsrv0.lrz-muenchen.de 129.187.20.237		lxsrv0.lrz-muenchen.de	dfn-grid-pca.pem dfn-grid-server-ca.pem	lrz-njs.pem	XXXX	XXXX

Outline

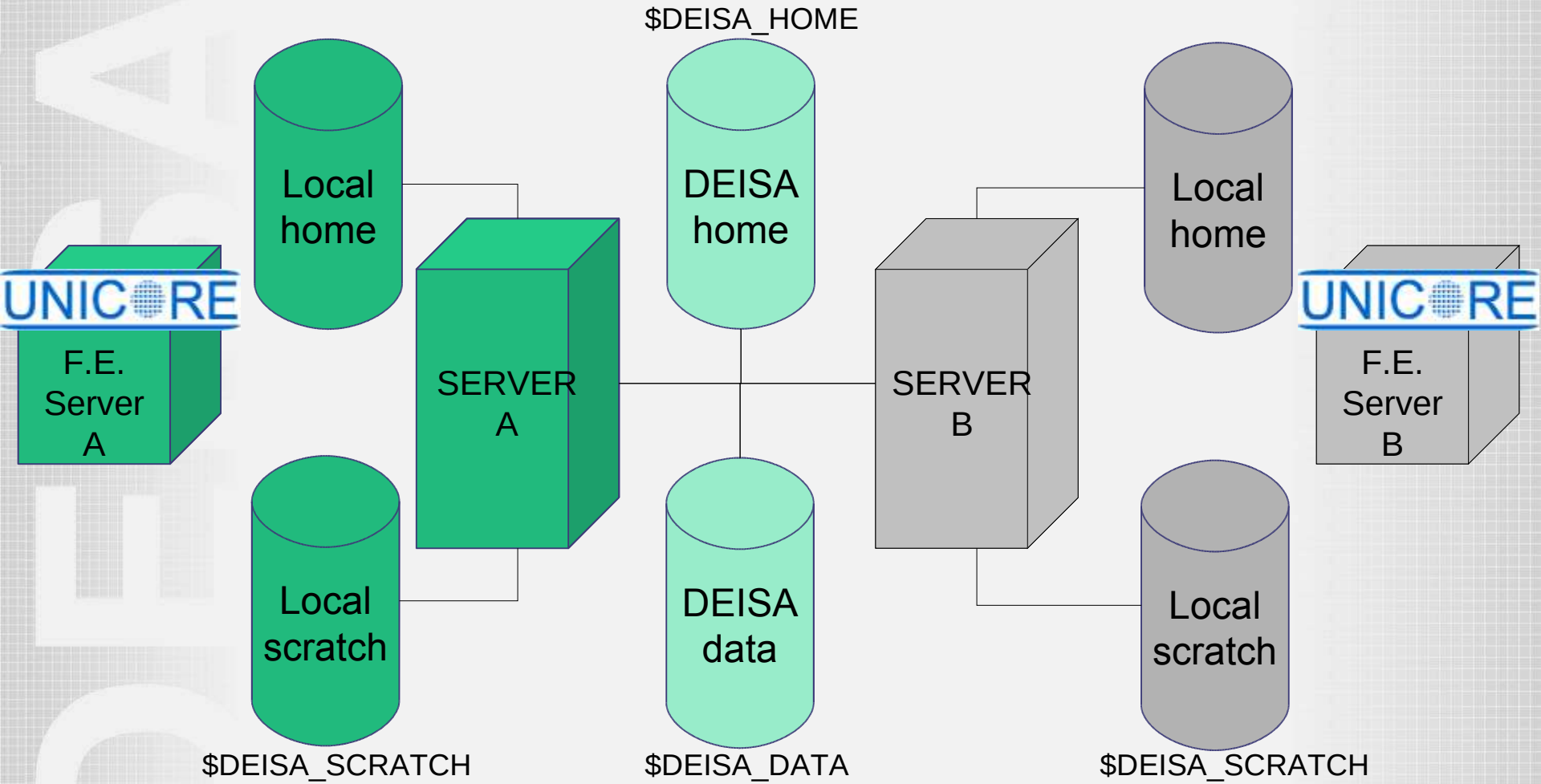
- DEISA Introduction
- GRID: logical layout and UNICORE
- Account management
- UNICORE configuration
- Job submission
- UNICORE availability

CPE: Common Production Environment

Distributed European Infrastructure for Supercomputing Applications

- Definition:
  - To ensure a high level coherence of the software environment in the DEISA supercomputing infrastructure a *Common Production Environment (CPE)* has been defined and deployed on each computer integrated in the platform.
  - The CPE includes:
    - shells (bash, tcsh)
    - compilers (C, C++, Fortran and Java)
    - libraries (communication, data formatting, numerical analysis, etc.)
    - tools (debuggers, profilers, etc.)

CPE: Common Production Environment



cp $DEISA_HOME/x $DEISA_SCRATCH
cp $DEISA_DATA/y $DEISA_SCRATCH
cp $DEISA_SCRATCH/z $DEISA_HOME

Distributed European Infrastructure for Supercomputing Applications

Submit

Local home

DEISA home

z

Local home

UNICORE

F.E. Server A

SERVER A

transfjob B

SERVER B

JOB

UNICORE

F.E. Server B

Local scratch

DEISA data

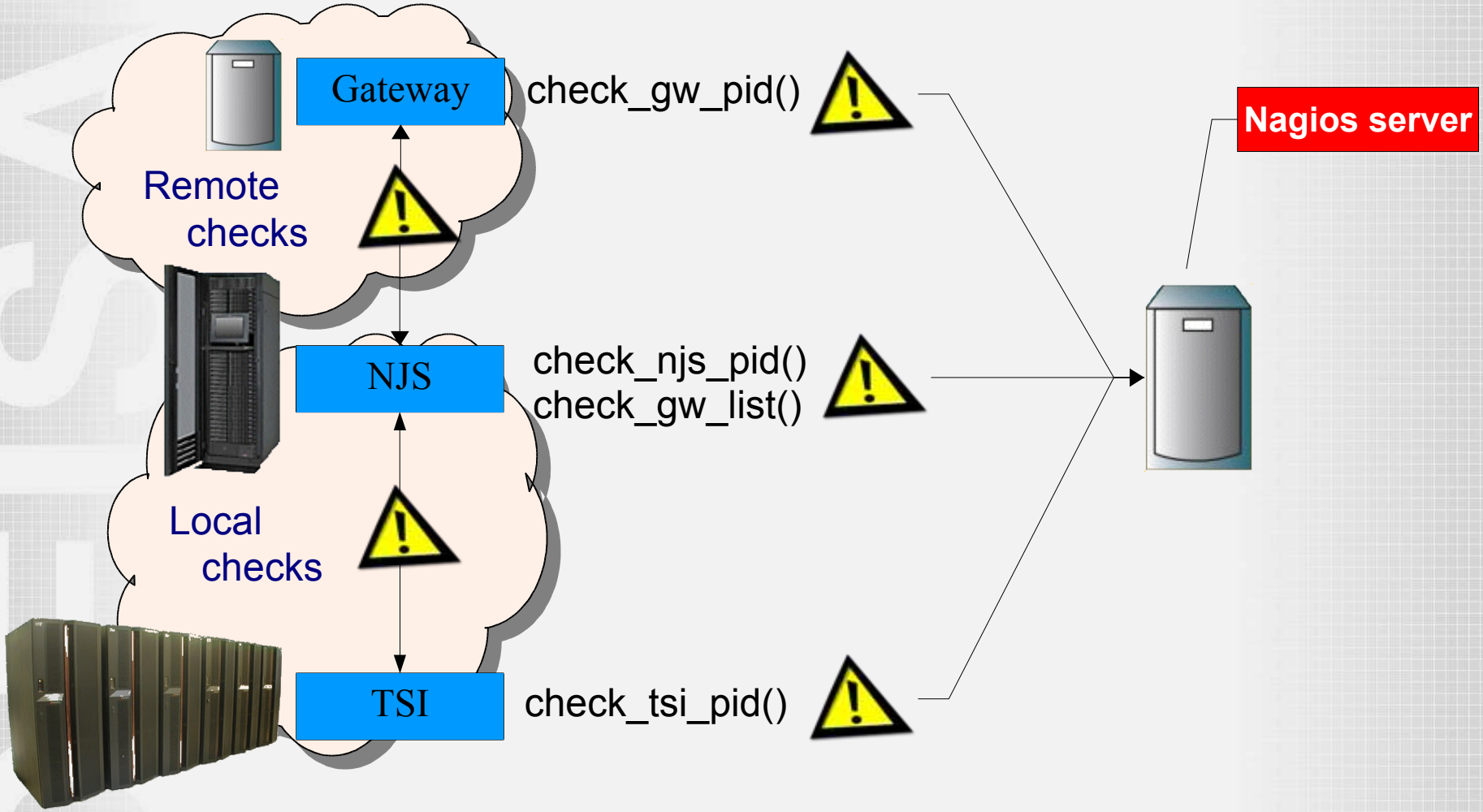
y

Local scratch

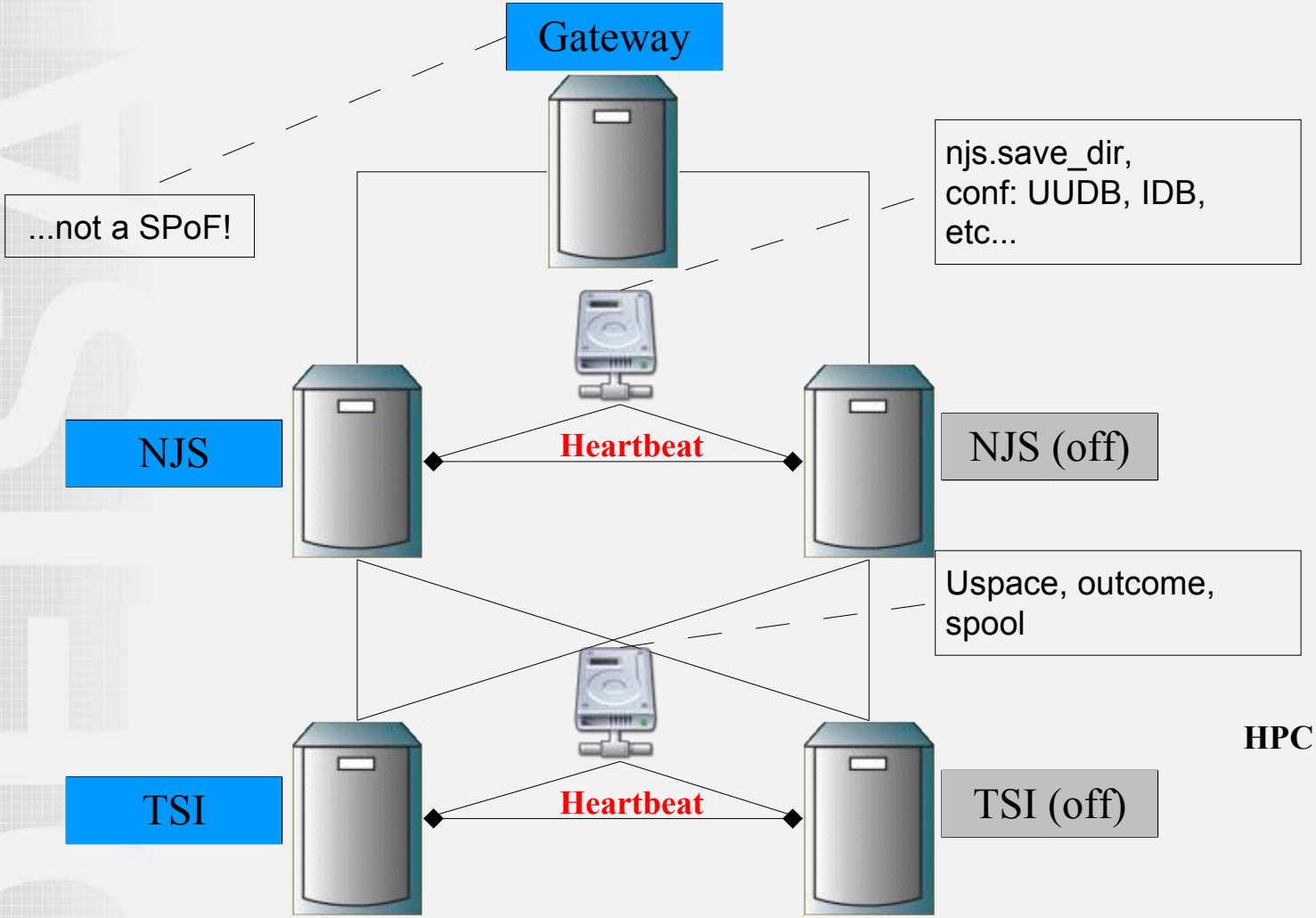
Outline

- DEISA Introduction
- GRID: logical layout and UNICORE
- Account management
- UNICORE configuration
- Job submission
- UNICORE availability

Healt-checking



Unicore and High Availability



Thank you for attending!!!