DEISA : Technologies and design principles

Distributed European Infrastructure for Supercomputing Applications

Andrea Vanni, CINECA

IBERGRID
May 14-16 2007
Santiago de Compostela (Spain)

Outline

Distributed European Infrastructure for Supercomputing Applications

- *Project overview*
  - *Objectives*
  - *Participating sites and resources*
  - *Activities*
  - *Benefits*
- *Project tecnologies*
  - *WAN global shared file system*
  - *Metascheduler layer*

Outline

Distributed European Infrastructure for Supercomputing Applications

- *Project overview*
  - *Objectives*
  - *Participating sites and resources*
  - *Activities*
  - *Benefits*
- *Project tecnologies*
  - *WAN global shared file system*
  - *Metascheduler layer*

DEISA objectives

Distributed European Infrastructure for Supercomputing Applications

- *Enabling Europe's terascale science by the integration of Europe's most powerful supercomputing systems.*
- *Enabling scientific discovery across a broad spectrum of science and technology is the only criterion for success*
- **DEISA is an European Supercomputing Service built on top of existing national services. This service is based on the deployment and operation of a persistent, production quality, distributed supercomputing environment with continental scope.**

The integration of national facilities and services, together with innovative operational models, is expected to add substantial value to existing infrastructures.

- **Main focus is High Performance Computing (HPC).**
- *RAS + not to be intrusive in yet deployed national services*
- *Better exploitation of the resources both at site level and European level + openness and usage of standards*

Participating Sites

Distributed European Infrastructure for Supercomputing Applications

1	BSC	*Barcelona Supercomputing Centre*	Spain
2	CINECA	*Consorzio Interuniversitario per il Calcolo Automatico*	Italy
3	CSC	*Finnish Information Technology Centre for Science*	Finland
4	EPCC/HPCx	*University of Edinburgh and CCLRC*	UK
5	ECMWF	*European Centre for Medium-Range Weather Forecast*	UK (int)
6	FZJ	*Research Centre Juelich*	Germany
7	HLRS	*High Performance Computing Centre Stuttgart*	Germany
8	IDRIS	*Institut du Développement et des Ressources en Informatique Scientifique – CNRS*	France
9	LRZ	*Leibniz Rechenzentrum Munich*	Germany
10	RZG	*Rechenzentrum Garching of the Max Planck Society*	Germany
11	SARA	*Dutch National High Performance Computing and Networking centre*	The Netherlands

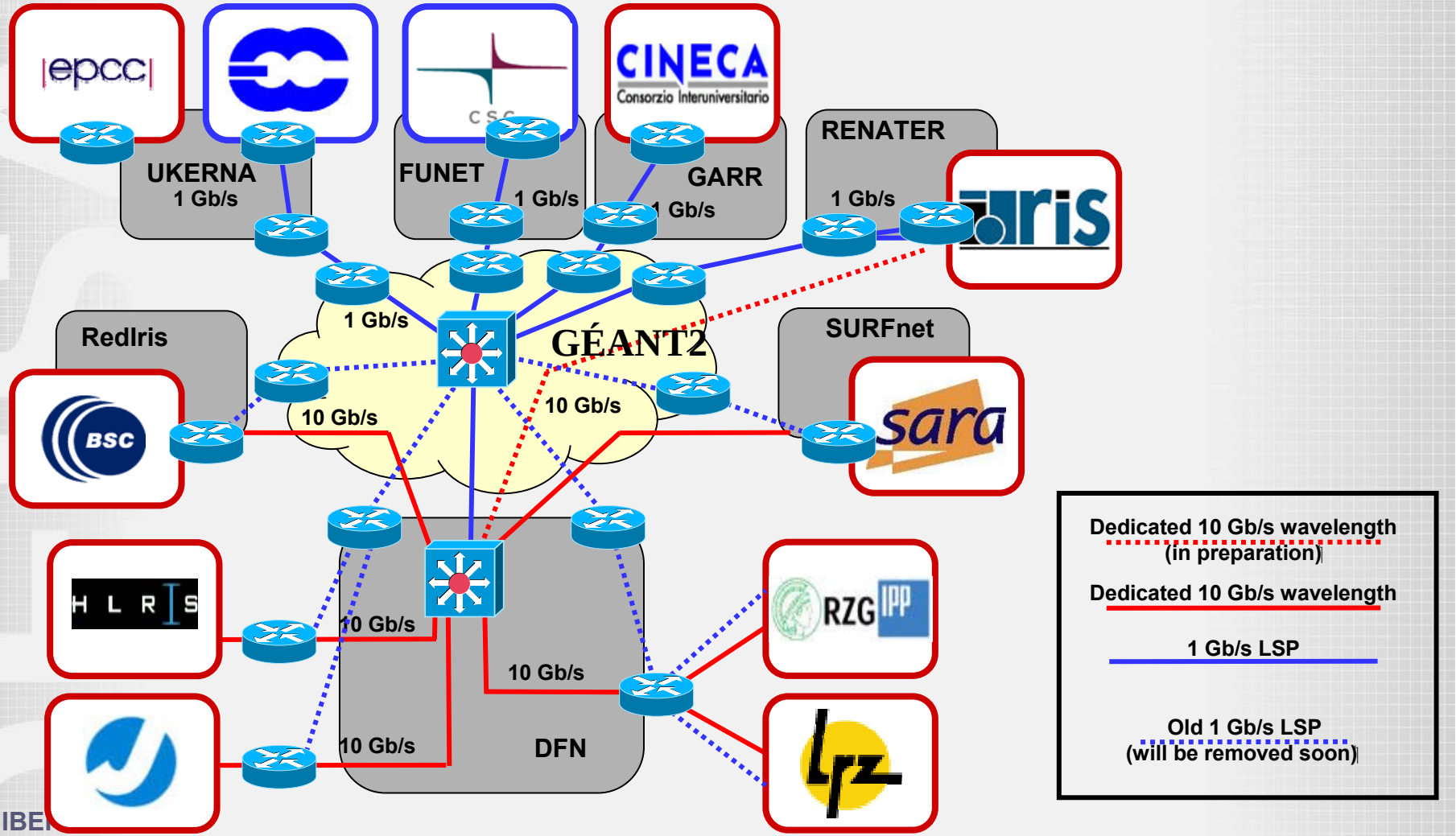
The DEISA supercomputing environment

Distributed European Infrastructure for Supercomputing Applications

IBM AIX Super-cluster			
	FZJ	1312	8.90 TFlop/s
	RZG	812	4.20 TFlop/s
	IDRIS	1024	6.55 TFlop/s
	CINECA	512	3.89 TFlop/s
	CSC	256	1.10 TFlop/s
	ECMWF	5280	40.00 TFlop/s
	HPCx	2560	12.90 TFlop/s
IBM PowerPC Linux system			
	BSC	10240	94.21 TFlop/s
SGI Altix Linux system			
	SARA	416	2.20 TFlop/s
	LRZ	9728	62.30 TFlop/s
NEC SX8 vector ystem			
	HLRS	576	12.70 TFlop/s
TOTAL		**32716**	**248.95 TFlop/s**

DEISA network environment

Distributed European Infrastructure for Supercomputing Applications



The DEISA network environment

Distributed European Infrastructure for Supercomputing Applications

SITE	Provider	Connection Gbis/s
BSC	DFN	10
CINECA	GEANT2	1 (10 July)
CSC	GEANT2	1
EPCC/HPCx	GEANT2	1
ECMWF	GEANT2	1
FZJ	DFN	10
HLRS	DFN	10
IDRIS	GEANT2	10
LRZ	DFN	10
RZG	DFN	10
SARA	DFN	10

DEISA activities

Distributed European Infrastructure for Supercomputing Applications

- **Service provision model is a trans-national extension of national supercomputing centres operational model**
- **Service Activities based on trans-national working teams with distributed leadership:**
  - **Network Deployment and Operation**
  - **Infrastructure Operation**
  - **Global File Systems**
  - **Resource management & middleware**
  - **User Support**
  - **Applications Enabling**
  - **Security/AAA**
  - **Middleware R&D (JRA)**

DEISA activities

Distributed European Infrastructure for Supercomputing Applications

- **Scientific Joint Research Activities**
  1. **Materials Sciences**
  2. **Plasma Physics**
  3. **Life Sciences**
  4. **CFD**
  5. **Cosmology**
  6. **Coupled Applications**
- **Development Joint Research Activities**
  7. **Access to Resources in Heterogeneous Environments**

DECI initiative: enabling Science

Distributed European Infrastructure for Supercomputing Applications

- Identification, deployment and operation of a number of « flagship » applications requiring the infrastructure services, in selected areas of science and technology.
- European Call for proposals in May - June every year. Applications are selected on the basis of scientific excellence, innovation potential and relevance criteria, with the collaboration of the HPC national evaluation committees.
- DECI users are supported by the Applications Task Force (ATASKF), whose objective is to enable and deploy the Extreme Computing applications. The activities of the ATASKF are focused on:
  - Hyperscaling of huge parallel applications, data oriented applications
  - Workflows and coupled applications
  - Production of an European Benchmark Suite for HPC systems

23 scientific projects in operation 2006 – 2007
29 scientific projects operated from 2005 to 2006

How is DEISA enhancing HPC services in Europe?

Distributed European Infrastructure for Supercomputing Applications

- **Providing a global data management service whose primordial objectives are:**
  - **Integrating distributed data with distributed computing platforms**
  - **Enabling efficient, high performance access to remote datasets (with Global File Systems and stripped GridFTP).**
  - **Integrating hierarchical storage management and databases in the supercomputing Grid.**

- **DEISA training and DEISA symposium days**

How is DEISA enhancing HPC services in Europe?

Distributed European Infrastructure for Supercomputing Applications

- **Running larger parallel applications** in individual sites, by a cooperative reorganization of the global computational workload on the whole infrastructure, or by the operation of the **job migration service** inside the AIX super-cluster.
- Enabling **workflow applications** with UNICORE (complex applications that are pipelined over several computing platforms)
- Enabling **coupled** multiphysics Grid applications (when it makes sense)
- **Deploying portals** as a way to hide complex environments to new users communities, and to interoperate with another existing grid infrastructures.
- **Deploying and monitoring a Common Production Environment.** Operational over the whole infrastructure.

Outline

Distributed European Infrastructure for Supercomputing Applications

- *Project overview*
  - *Objectives*
  - *Participating sites and resources*
  - *Activities*
  - *Benefits*
- *Project technologies and production environment*
  - *WAN global shared file system*
  - *Metascheduler layer*

DEISA supercomputers

Distributed European Infrastructure for Supercomputing Applications

IDRIS IBM P4

FZJ IBM P4

HLRS NEC SX8

ECMWF IBM P4

HPCX IBM P5

CSC IBM P4

LRZ SGI ALTIX

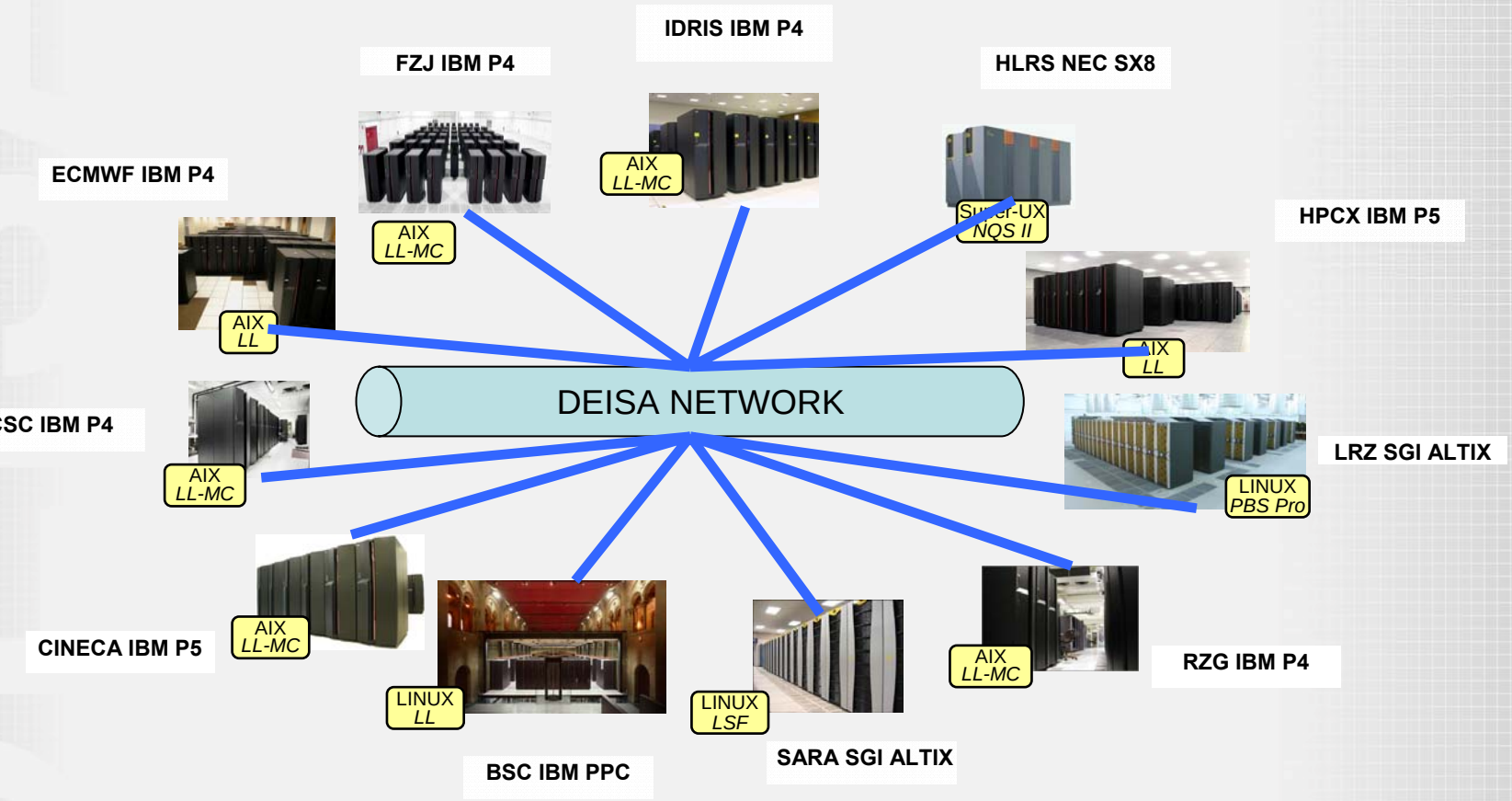
CINECA IBM P5

BSC IBM PPC

SARA SGI ALTIX

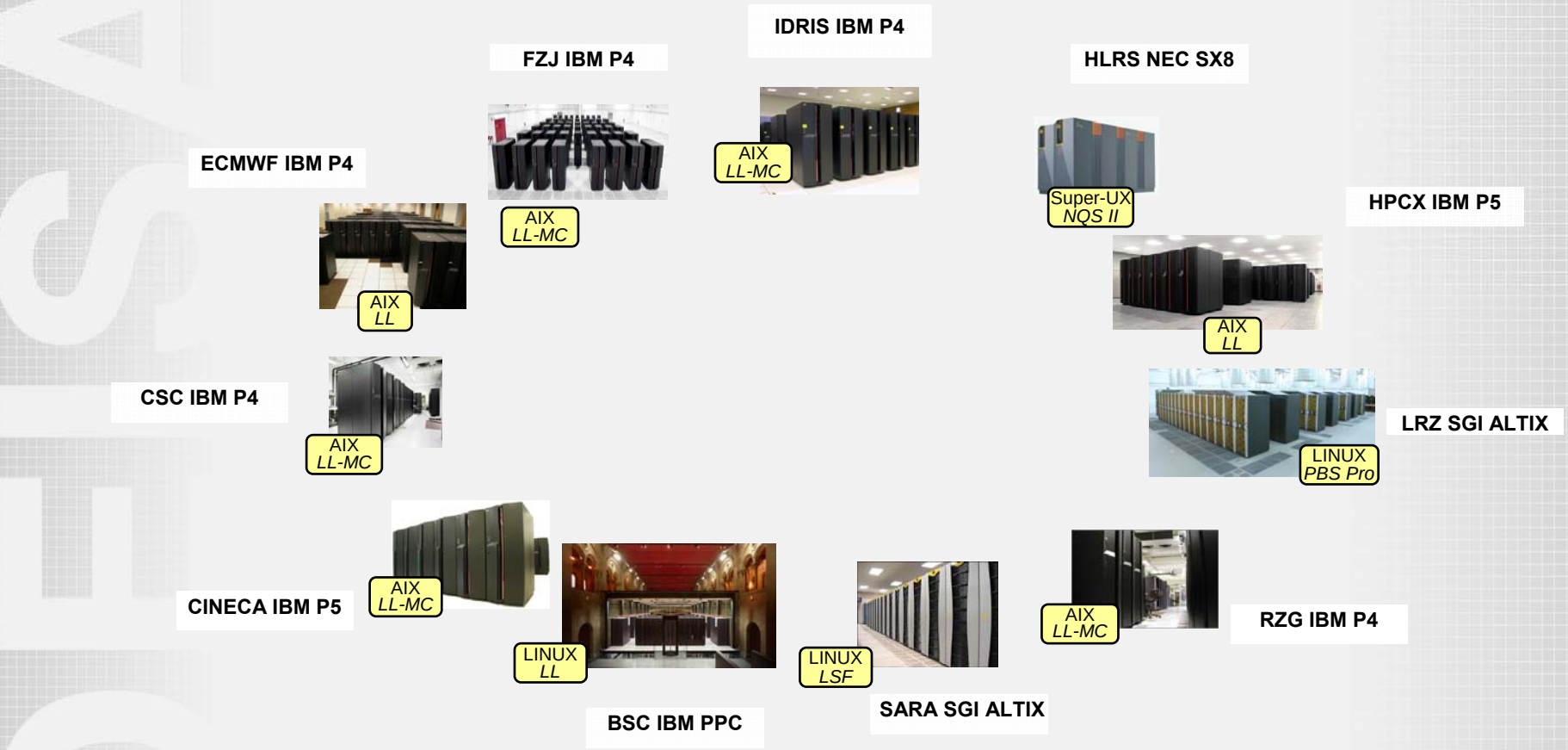
RZG IBM P4

DEISA GPFS

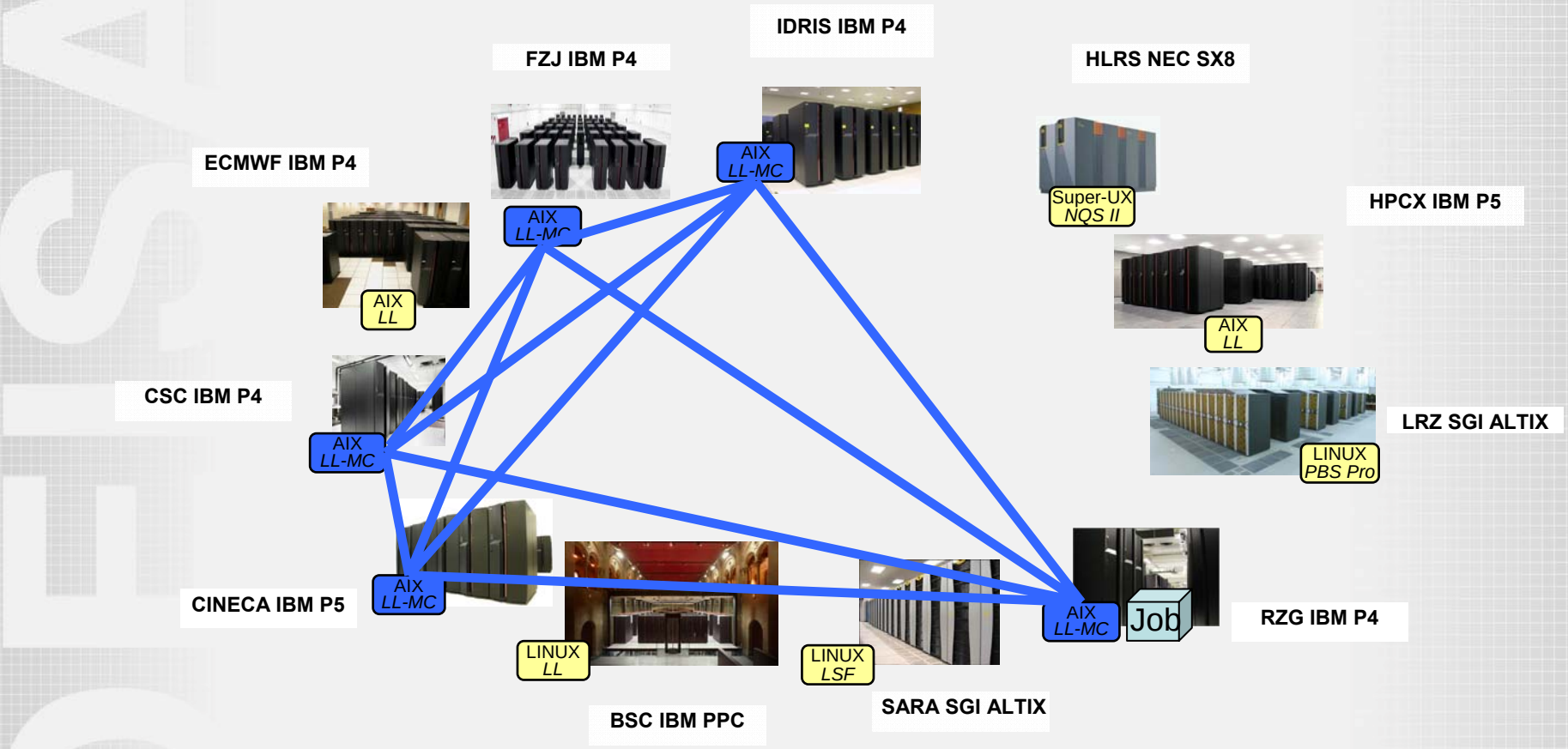

IDRIS IBM P4
FZJ IBM P4
HLRS NEC SX8
ECMWF IBM P4
AIX LL-MC
AIX LL-MC
Super-UX NQS II
HPCX IBM P5
AIX LL
AIX LL
DEISA NETWORK
CSC IBM P4
AIX LL-MC
LINUX PBS Pro
LRZ SGI ALTIX
AIX LL-MC
CINECA IBM P5
LINUX LL
LINUX LSF
AIX LL-MC
RZG IBM P4
BSC IBM PPC
SARA SGI ALTIX

DEISA Batch systems

Distributed European Infrastructure for Supercomputing Applications

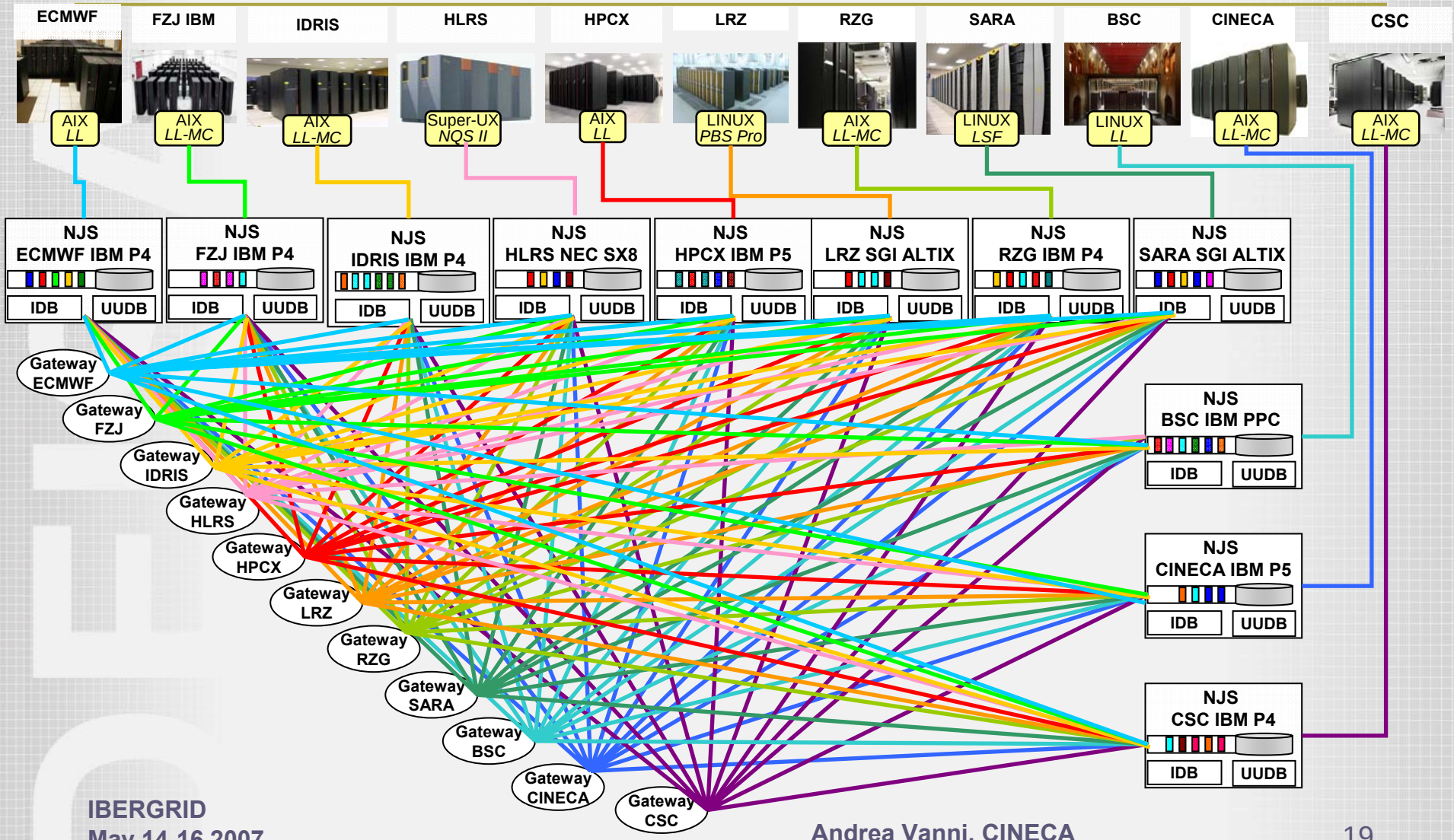


DEISA LL-MC & re-routing



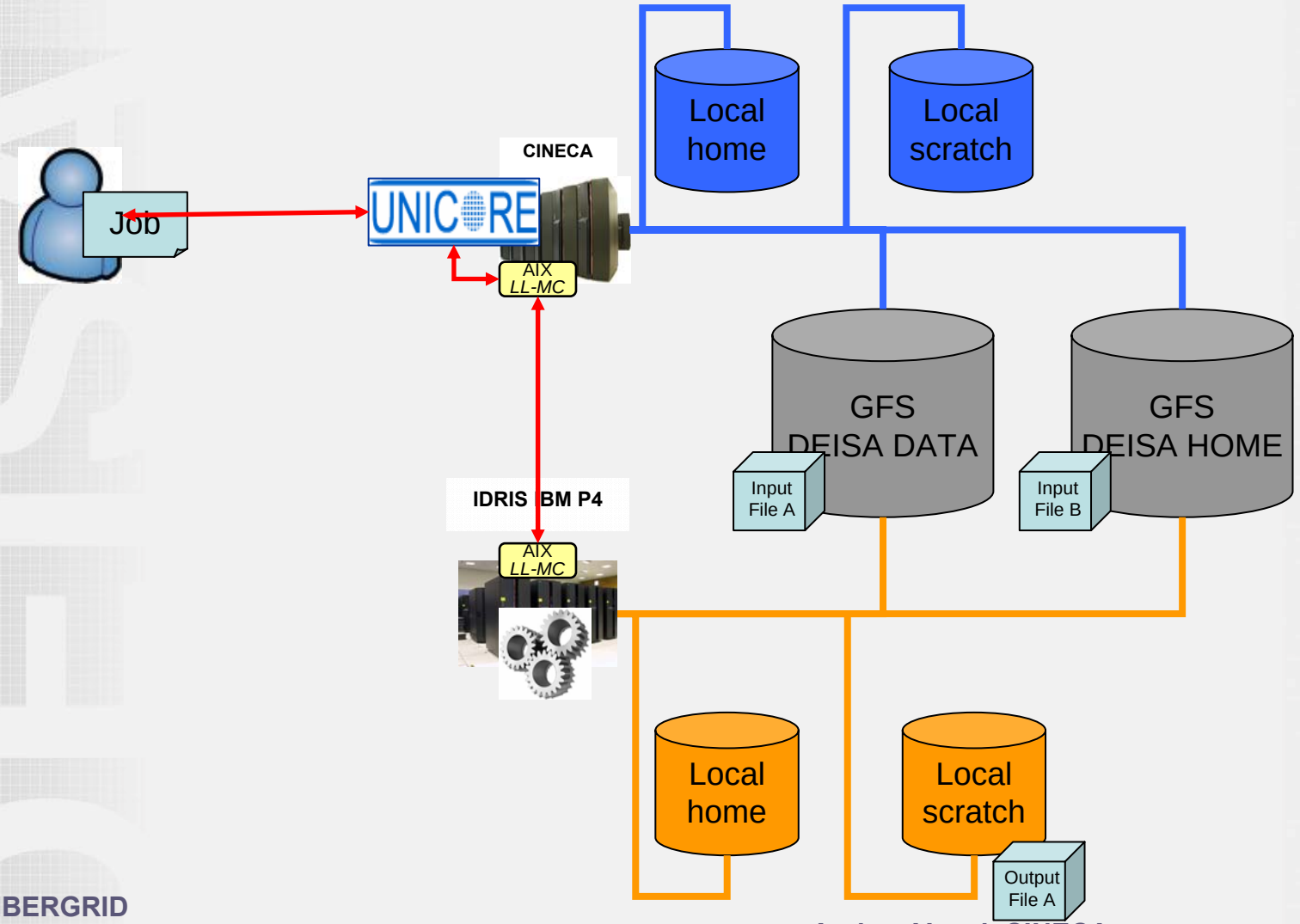
DEISA UNICORE

Distributed European Infrastructure for Supercomputing Applications



UNICORE + Job re-routing + GPFS

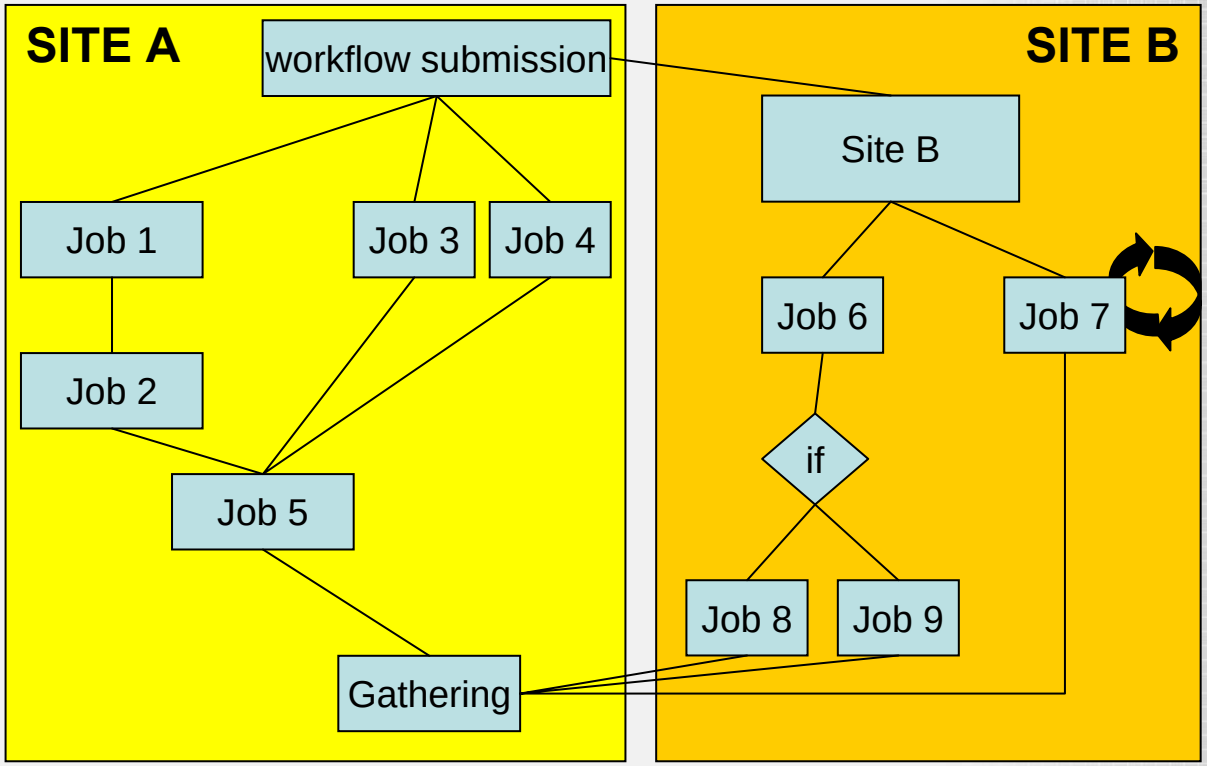
Distributed European Infrastructure for Supercomputing Applications



DEISA Workflow

DEISA is a workflow enabled infrastructure via UNICORE

- Simultaneous job submission intra/extra site
- Contiditional execution
- Cyclic execution



DESHL

Distributed European Infrastructure for Supercomputing Applications

- DESHL provides OGSA standards-based access for users and their applications to manage jobs and transfer files in the DEISA heterogeneous supercomputing infrastructure. This is provided by a command line client application which supports SAGA standards for remote job and file management

- DESHL comprises:
  - command line tool based on the Open Group Batch Environment Services specification;
  - client library exposing an API based on the SAGA standard currently being developed by a OGF Research Group;
  - Grid Access Library for interacting with a UNICORE Grid.

Advanced resource management capabilities: co-allocation and co-reservation

Distributed European Infrastructure for Supercomputing Applications

LSF based systems

- LSF MC tests
  - internal to CINECA
    - LSF MC (master) on a small Intel xeon based cluster
    - LSF HPC on SGI Altix (64 IA64)
    - LSF HPC on HP XC2 (128 IA64)
  - between CINECA and SARA
    - LSF MC (master) on a small Intel Xeon based cluster (CIN)
    - LSF HPC on SGI Altix (SARA)
  - two use cases for co-allocation

Advanced resource management capabilities: co-allocation and co-reservation

Distributed European Infrastructure for Supercomputing Applications

Heterogeneous batch system via Universus and LSF

- Universus is translator for batch system submission and reservation requests.
- LSF uses Universus for allocate resources on systems with different batch system and reservation policies.

Accessing remote data: high performance remote I/O and file transfer

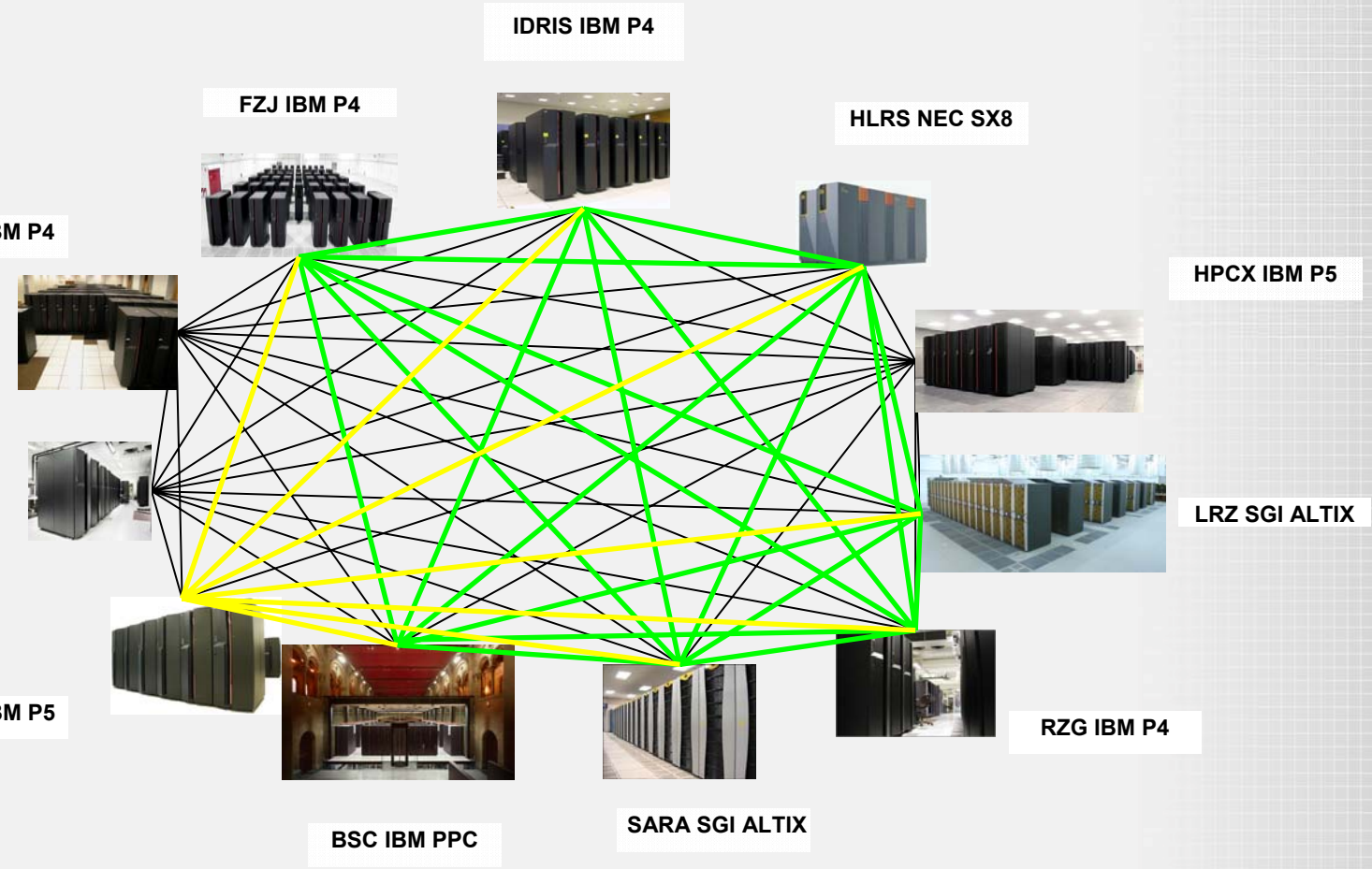
Distributed European Infrastructure for Supercomputing Applications

Two ways

- Implicit via GPFS
- Explicit
  - via GridFTP
  - via Unicore

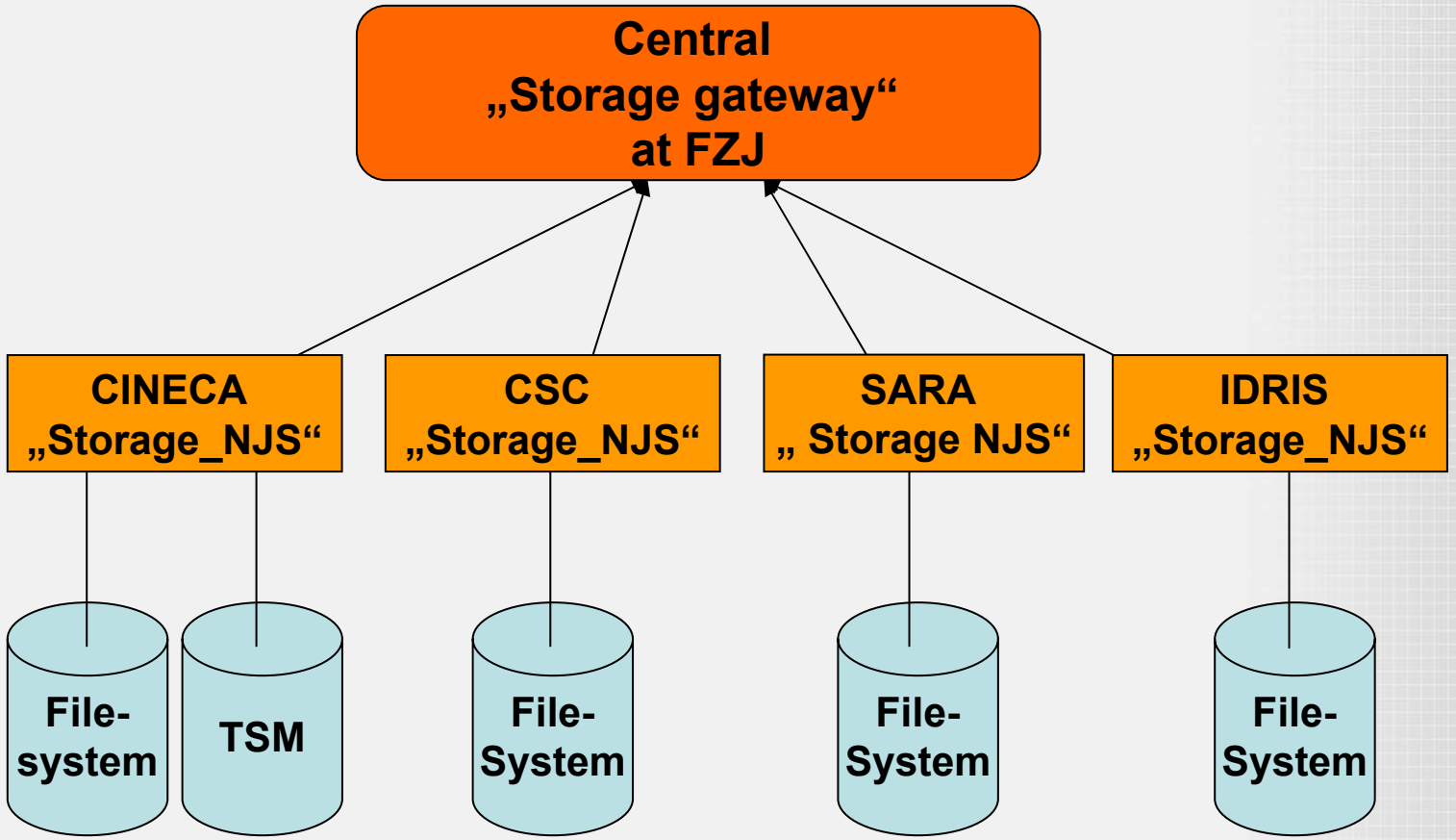
GridFTP – 10 GBit/s

Distributed European Infrastructure for Supercomputing Applications



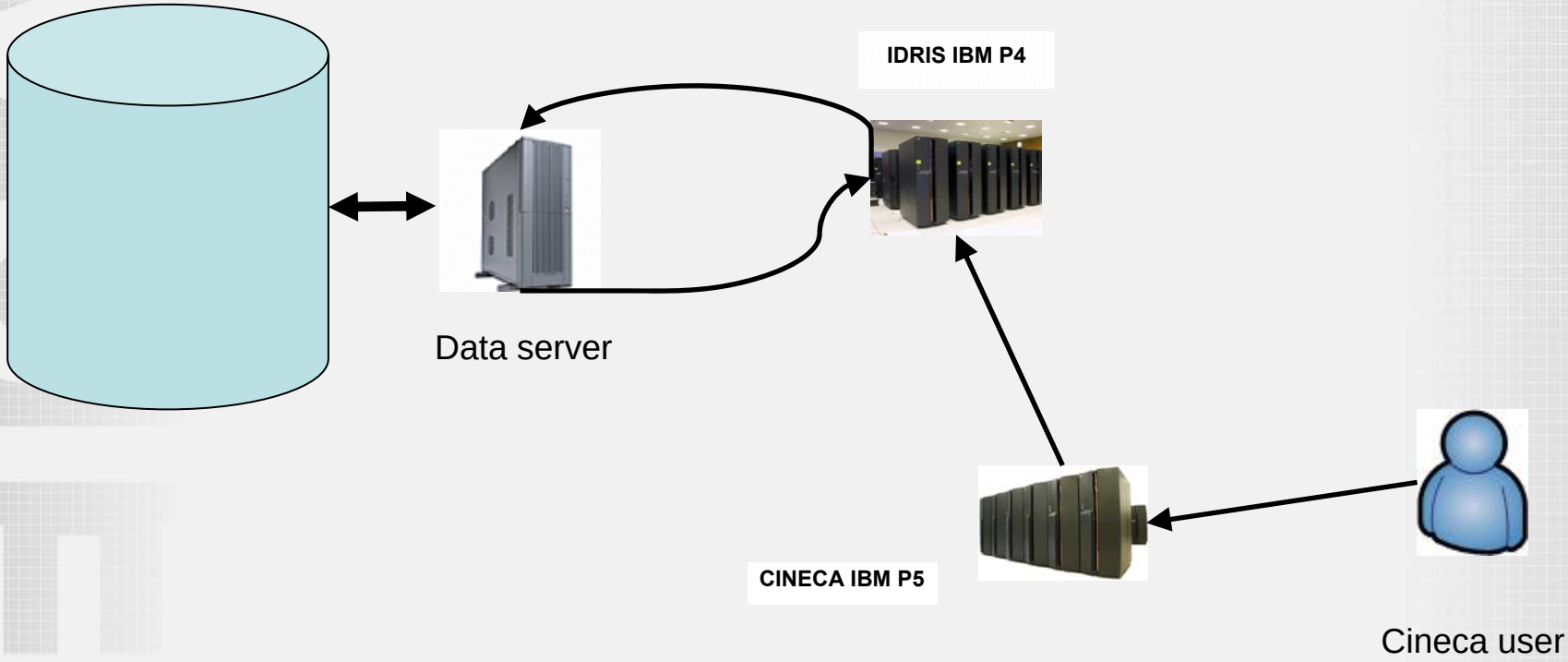
Data federation via UNICORE

Distributed European Infrastructure for Supercomputing Applications

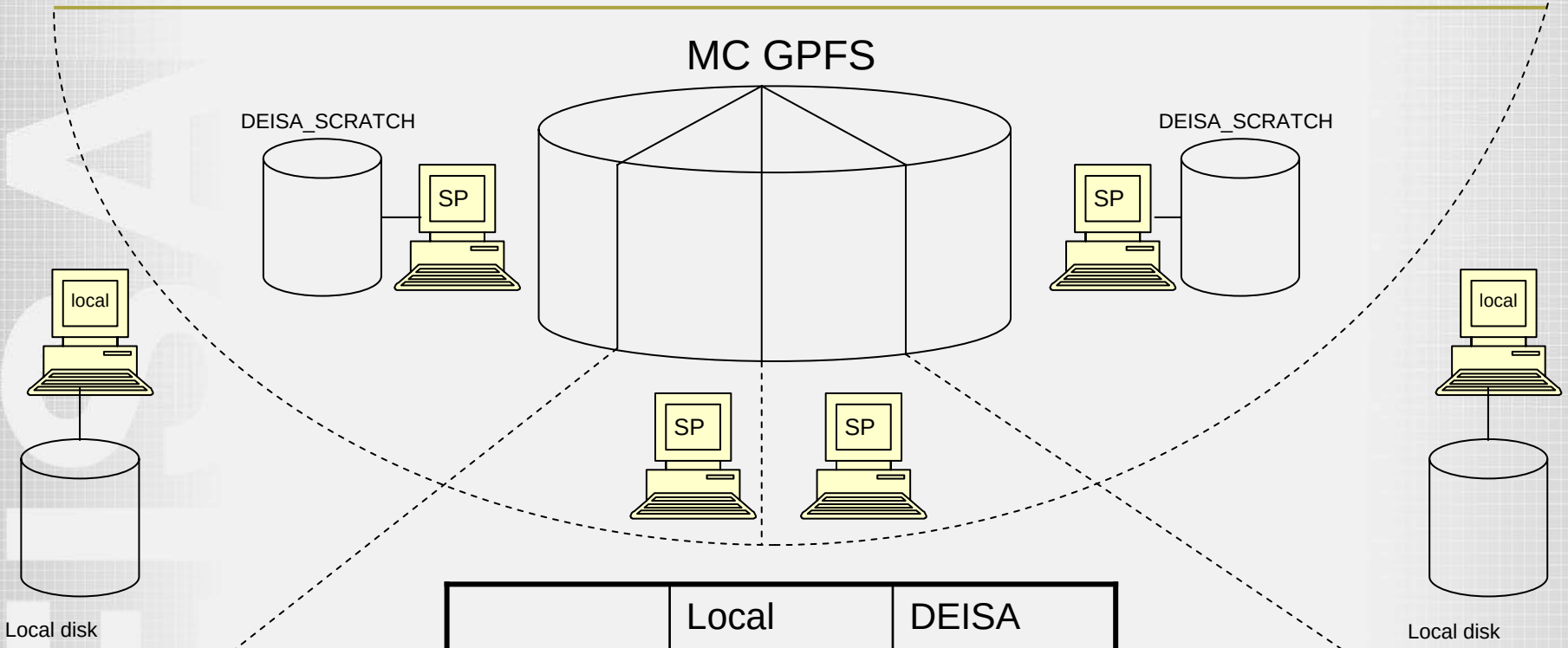


Data federation via DF DEISA tool

Distributed European Infrastructure for Supercomputing Applications

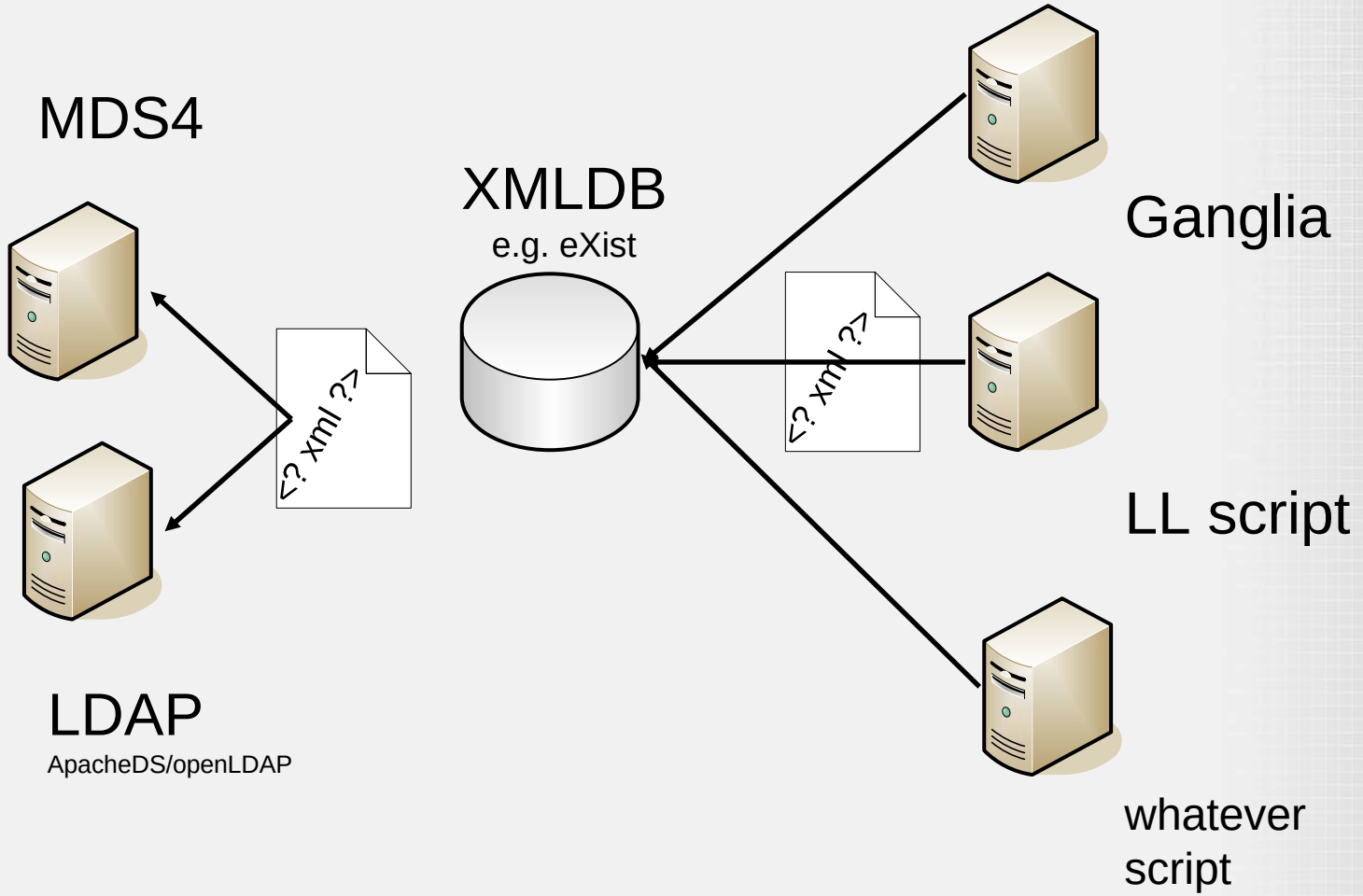


Data management integrated view

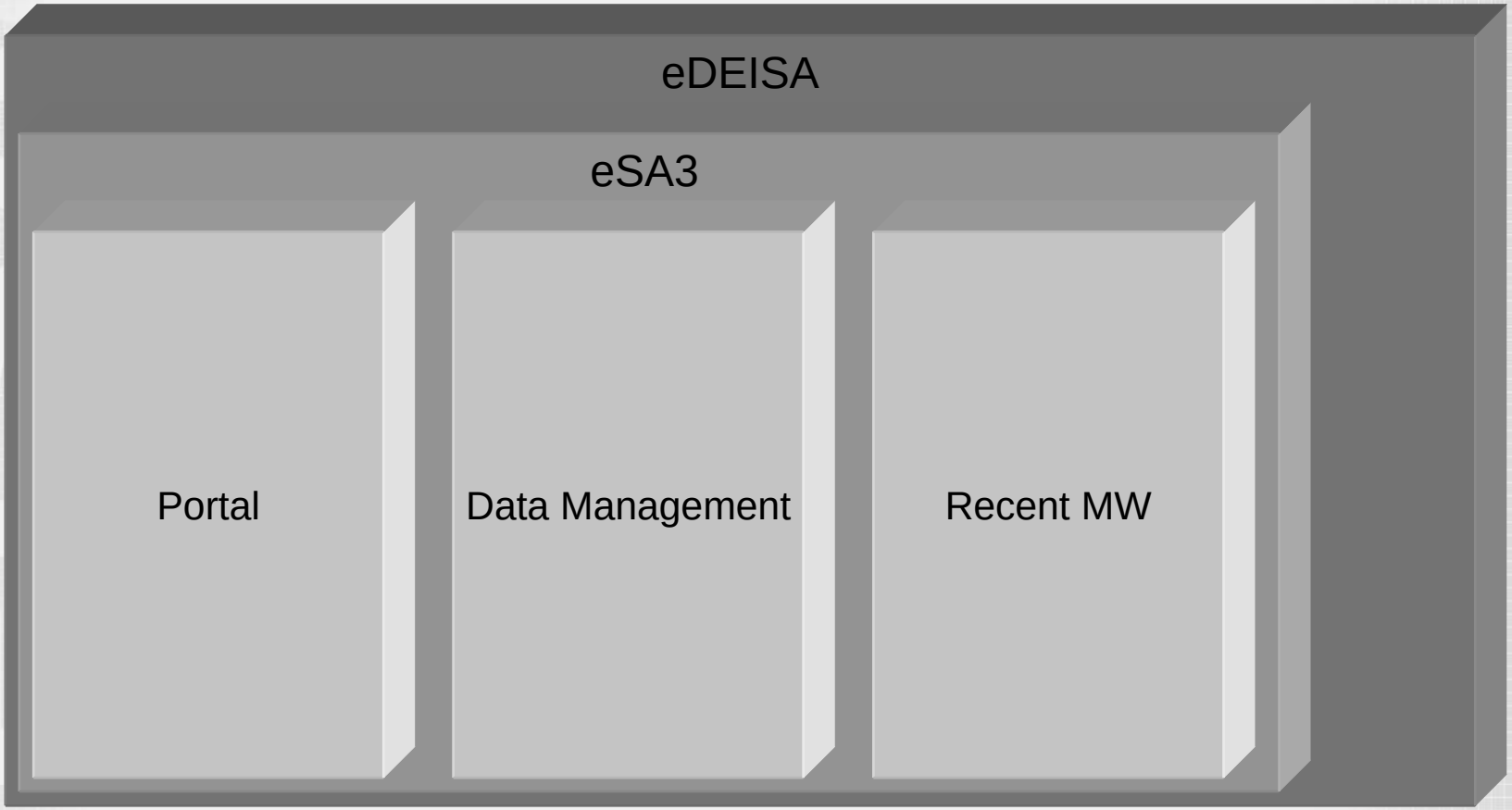


	Local	DEISA
To Local	GridFTP	Stage Out
To DEISA	Stage in	GPFS (GridFTP)

Resource management information system
New design



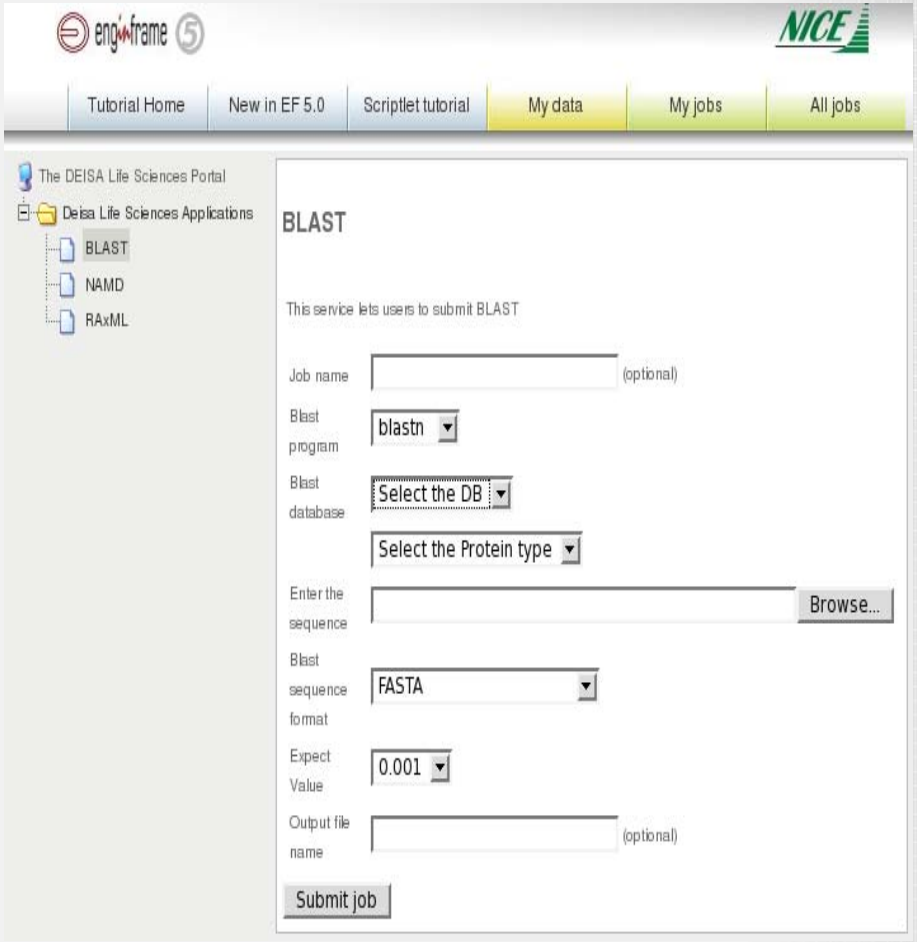
eDEISA – eSA3



eDEISA – eSA3 - portal

Distributed European Infrastructure for Supercomputing Applications

- Portal
  - RAxML
  - BLAST
  - NAMD



- Data Management
  - GridFTP
  - OGSA-DAI
- RRM
  - GRAM
  - NAREGI
  - GSI-SSH

DEMOs

Distributed European Infrastructure for Supercomputing Applications

- Let's start …

Conclusions

Distributed European Infrastructure for Supercomputing Applications

- DEISA aims at deploying a **sustained and persistent**, basic European infrastructure for general purpose high performance computing.

- **We expect services and existing synergies to be persistent**. We do not claim persistency of the current organizational model. The DEISA Consortium is ready to adapt to the new FP7 strategies and establish a roadmap incorporating cooperation or merging with new HPC initiatives.

Conclusions

Distributed European Infrastructure for Supercomputing Applications

- Our next challenge is establishing an efficient organization embracing all relevant HPC organizations in Europe.
- Interfaced with the other grid-enabled complementary infrastructures, DEISA expects to continue to contribute to a global European eInfrastructure for science and technology
- *Integrating leading supercomputing platforms with Grid technologies and reinforcing capability with shared petascale systems is needed to open the way to new research dimensions in Europe.*