

JORNADAS REDIMADRID
Investigación en Red de Frontera
Universidad Rey Juan Carlos
24 de Abril de 2008

Tendencias en GRID para Potenciar la Investigación Colaborativa

dsa-research.org

Ignacio Martín Llorente

Distributed Systems Architecture Research Group
Universidad Complutense de Madrid



Motivación

- La existencia de **consorcios de investigación** compuestos por grupos complementarios con diferentes especialidades es indispensable en el avance de la investigación
- Las **Tecnologías de la Información** son un soporte a los procesos de negocio e investigación
- Las tecnologías que proporcionan **soporte a la investigación colaborativa** son clave para el desarrollo de una actividad de investigación puntera a nivel mundial.
- El **Grid** es la tecnología que permite que grupos de diferentes organismos compartan recursos como soporte a proyectos de investigación colaborativos



Objetivos

dsa-research.org

- Analizar los objetivos y características de una infraestructura Grid computacional
- Identificar los componentes principales de una arquitectura Grid computacional
- Describir ejemplos de infraestructuras Grid como soporte a la investigación colaborativa
- Caso de Estudio: **GRIDIMadrid**



Contenidos

1. Recursos Computacionales
2. Middleware Grid
3. Ejemplos de Infraestructuras
4. GRIDIMadrid

dsa-research.org

1. Recursos Computacionales

1.1. Computación Paralela y Distribuida

Objetivo de la Computación Paralela y Distribuida

- Ejecución **eficiente** de aplicaciones intensivas en cálculo o datos

Paradigmas de Computación Paralela y Distribuida

High Performance Computing (HPC)

- Reducción del tiempo de ejecución de una aplicación paralela (memoria compartida o distribuida)
- Rendimiento del sistema en FLOPS
- Ejemplos: CFD, modelos meteorológicos

High Throughput Computing (HTC)

- Aumentar el número de ejecuciones por unidad de tiempo
- Rendimiento del sistema en trabajos por segundo, o FLOPY
- Ejemplos: Física de Altas Energías, Bioinformática, Modelos Financieros

1. Recursos Computacionales

1.2. Tipos de Plataformas

**Centralizada
Acoplada**

- Red de Interconexión
- Administración
- Homogeneidad

**Descentralizada
Desacoplada**

**SMP (Symmetric
Multi-processors)**



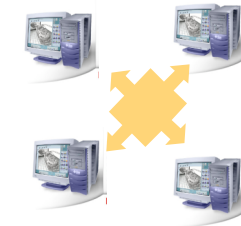
**MPP (Massive
Parallel Processors)**



Clusters



**Network Systems
Intranet/Internet**



High Performance Computing

High Throughput Computing

1. Recursos Computacionales

1.3. Gestores Locales de Recursos

Sistemas Locales para la Gestión de Recursos (LRMS)

- Ofrecen una visión uniforme y única del sistema
- Gestión de los trabajos, optimiza el uso (carga...)
- Habitualmente asumen acceso exclusivo a los recursos
 - ① Sistemas de colas *batch* para servidores HPC
 - ② Gestores de recursos en clusters dedicados
 - ③ Gestores de carga para sistemas en red

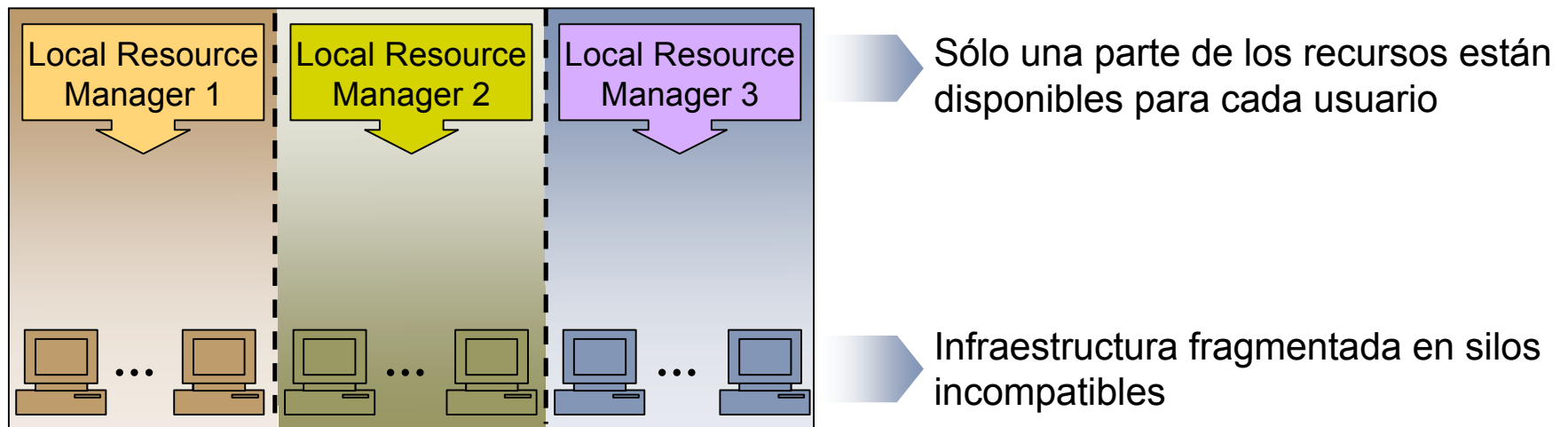
<i>Independent Suppliers</i>	<i>Open Source</i>	<i>OEM Proprietary</i>
② <i>Platform Computing</i> ③ LSF	② <i>Altair</i> Open PBS	① <i>IBM</i> Load Leveler
② <i>Altair</i> PBS Pro	② <i>University of Wisconsin</i> ③ Condor	① <i>Cray</i> NQE
	② <i>Sun Microsystems</i> ③ SGE	

1. Recursos Computacionales

1.3. Gestores Locales de Recursos

Desventajas de los Gestores Locales

- No disponen de interfaces ni infraestructuras de seguridad comunes
- Basados en protocolos propietarios
- **Silos computacionales incompatibles en una misma organización**
 - Habilidades de administración específicas
 - Aumenta el coste operacional
 - Sobre-aprovisionamiento y desequilibrio global de la carga
- Imposibilidad de construir Infraestructuras computacionales que abarcan varios dominios de administración (políticas de seguridad y LRMS)





2. Middleware Grid

1. Recursos Computacionales
2. Middleware Grid
3. Ejemplos de Infraestructuras Grid
4. GRIDIMadrid

2. Middleware Grid

2.1. Integración de Diferentes Dominios de Administración

Un Nuevo Problema ...

“The real and specific problem that underlies the Grid concept is coordinated resource **sharing** and problem solving in dynamic, multi-institutional **virtual organizations**.”

*Ian Foster, Carl Kesselman, Steven Tuecke
The Anatomy of the Grid. Enabling Scalable Virtual Organizations (2001)*

sharing ...

- de **recursos** de naturaleza dispar (software, hardware, datos...)
- entre entidades arbitrarias y dinámicamente (**interoperabilidad**)
- sobre **protocolos** comunes, un Grid es una arquitectura de protocolos
- conjunto de individuos y/o instituciones que colaboran compartiendo los recursos para alcanzar objetivos comunes (VO)

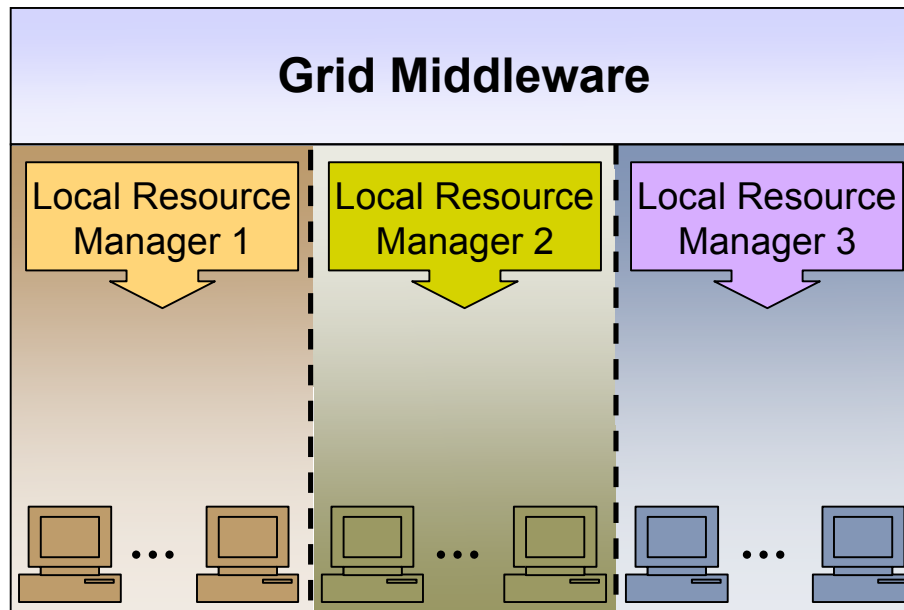
2. Middleware Grid

2.1. Integración de Diferentes Dominios de Administración

"Any problem in computer science can be solved with another layer of indirection... But that usually will create another problem." David Wheeler

... Un Nuevo Nivel de Abstracción

“Un grid (*computacional*) ofrece una capa de abstracción (*middleware*) para integrar diferentes dominios de administración (plataformas y políticas heterogéneas) “



Interfaces Comunes: Los usuarios pueden acceder un amplio conjunto (número y tipo) de recursos

Infraestructura: Recursos computacionales y de almacenamiento, red y LRMS

2. Middleware Grid

2.1. Integración de Diferentes Dominios de Administración

Características de un Grid

“Un grid es un sistema que ...

1. coordina recursos que no están sujetos a un control centralizado ...
2. usando protocolos e interfaces estándar, abiertos y de propósito general ...
3. para proporcionar calidades de servicio no triviales.”

Ian Foster

What is the Grid? A Three Point Checklist (2002)

Un arquitectura Grid puede definirse de forma funcional mediante el conjunto de servicios que soportan las actividades de una VO (**OGSA**)

Ian Foster, Carl Kesselman, Jeffrey M. Nick y Steven Tuecke

“The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration”

2. Middleware Grid

2.1. Integración de Diferentes Dominios de Administración Grid Middleware (una visión computacional)

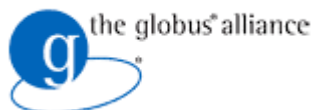
- **Servicios básicos**

- Seguridad
- Información y monitorización
- Gestión de datos
- Ejecución
- **Meta-planificación**

- **Distribuciones “Open-Source” de Middleware Grid**



- **Comunidades “Open-Source”**



The Globus Alliance (dev.globus.org)

2. Middleware Grid

2.2. Globus Toolkit

¿Por qué Globus?...

- **Proyecto “Open Community”** basado en el modelo Apache Jakarta:
 - El control de cada proyecto está en manos de los *comitters*
 - Infraestructura pública de desarrollo para cada proyecto: CVS, bugzilla, listas de correo, y Wiki
 - Cada proyecto sigue un proceso de *incubación* antes de ser un proyecto Globus
- **Globus Toolkit (GT)** integra una selección de proyectos Globus
- GT ofrece un conjunto de servicios para acceder de forma segura múltiples recursos en diferentes dominios de administración, cada uno con sus políticas y gestores locales



2. Middleware Grid

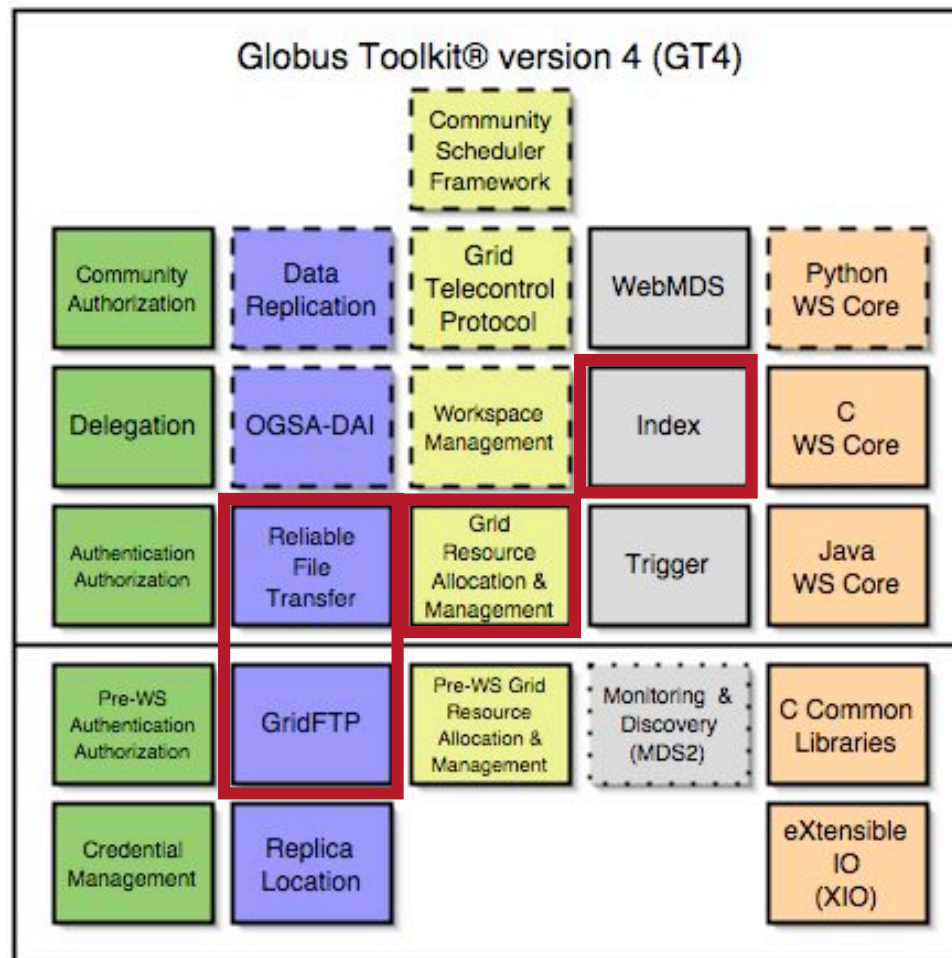
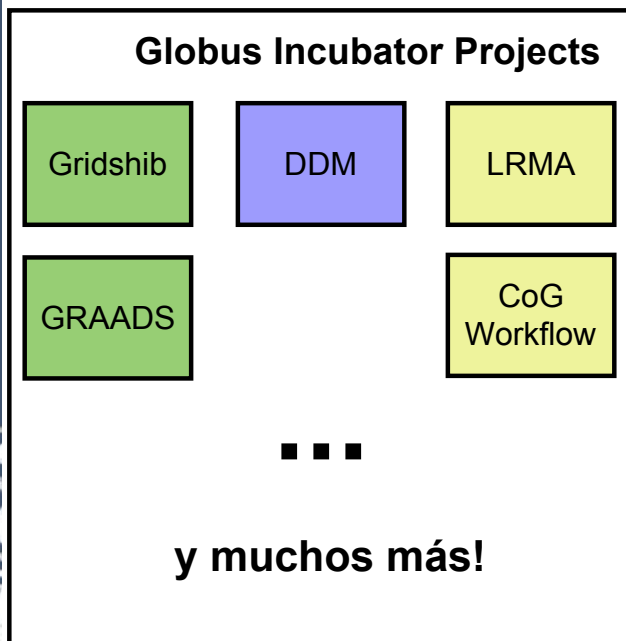
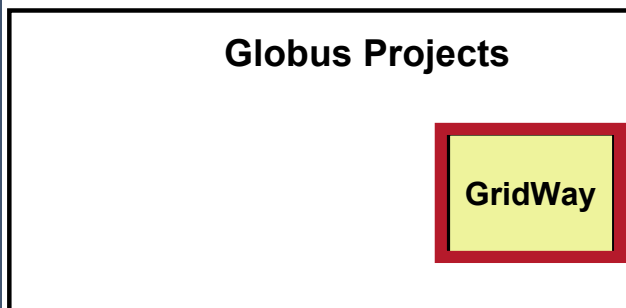
2.2. The Globus Toolkit

Componentes Globus

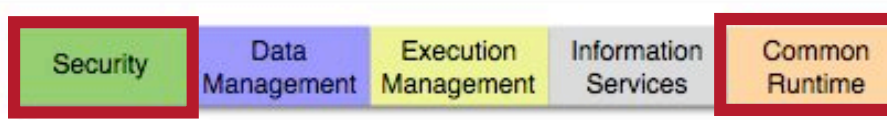
BioGridNet



dsa-research.org



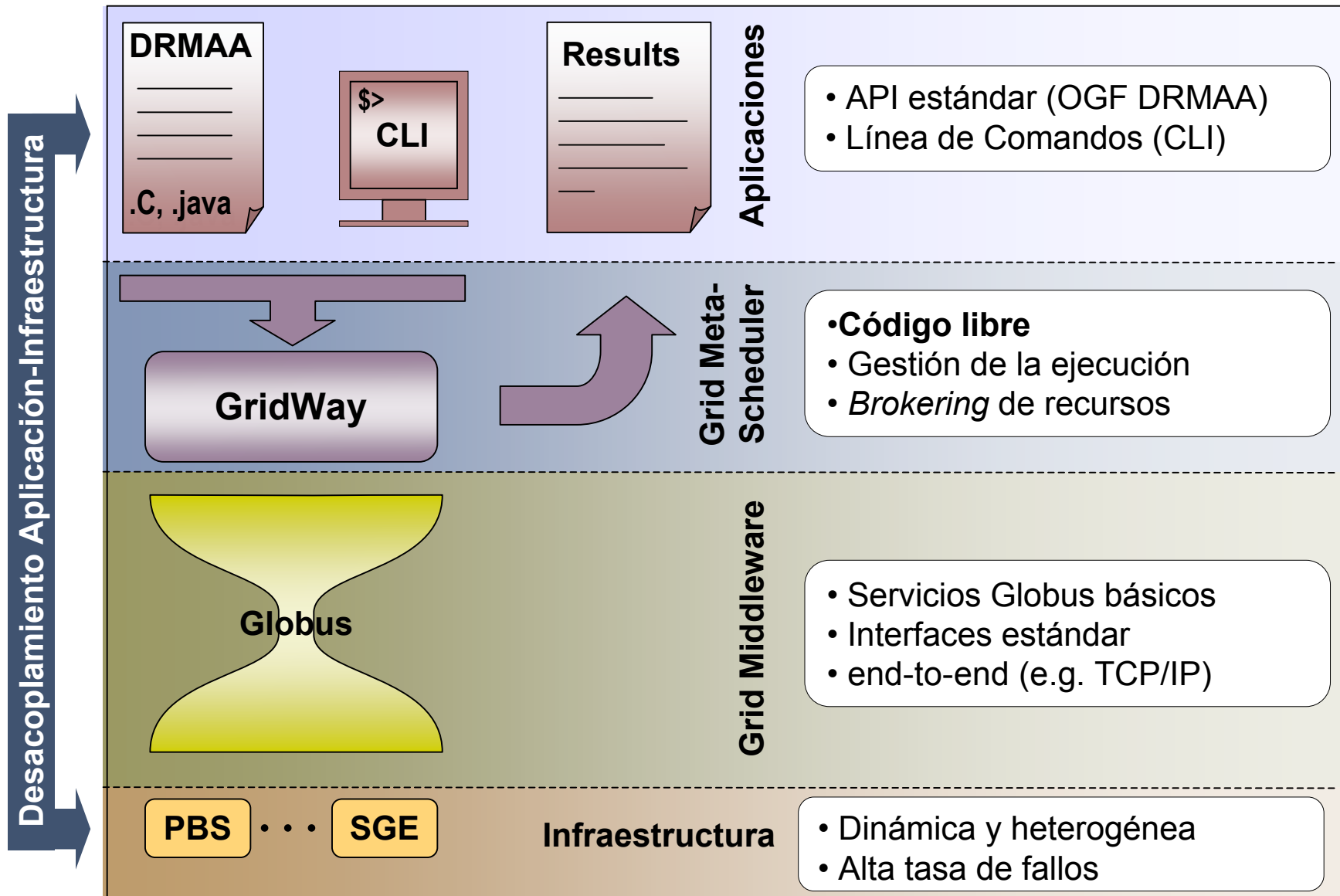
Componentes mínimos para un Grid computacional



2. Middleware Grid

2.3 Plataformas de Computación

Arquitectura de un Grid Computacional



2. Middleware Grid

2.3 Plataformas de Computación

**Centralizada
Acoplada**

- Red de Interconexión
- Administración
- Homogeneidad

**Descentralizada
Desacoplada**

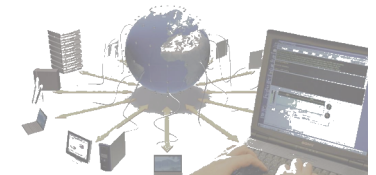
**SMP (Symmetric
Multi-processors)**

**MPP (Massive
Parallel Processors)**

Clusters

**Network Systems
Intranet/Internet**

**Grid
Infrastructures**



High Performance Computing

High Throughput Computing



3. Ejemplos de Infraestructuras Grid

1. Recursos Computacionales
2. Middleware Grid
3. Ejemplos de Infraestructuras Grid
4. GRIDIMadrid

3. Ejemplos de Infraestructuras Grid

3.1 Uso de GridWay en el Mundo



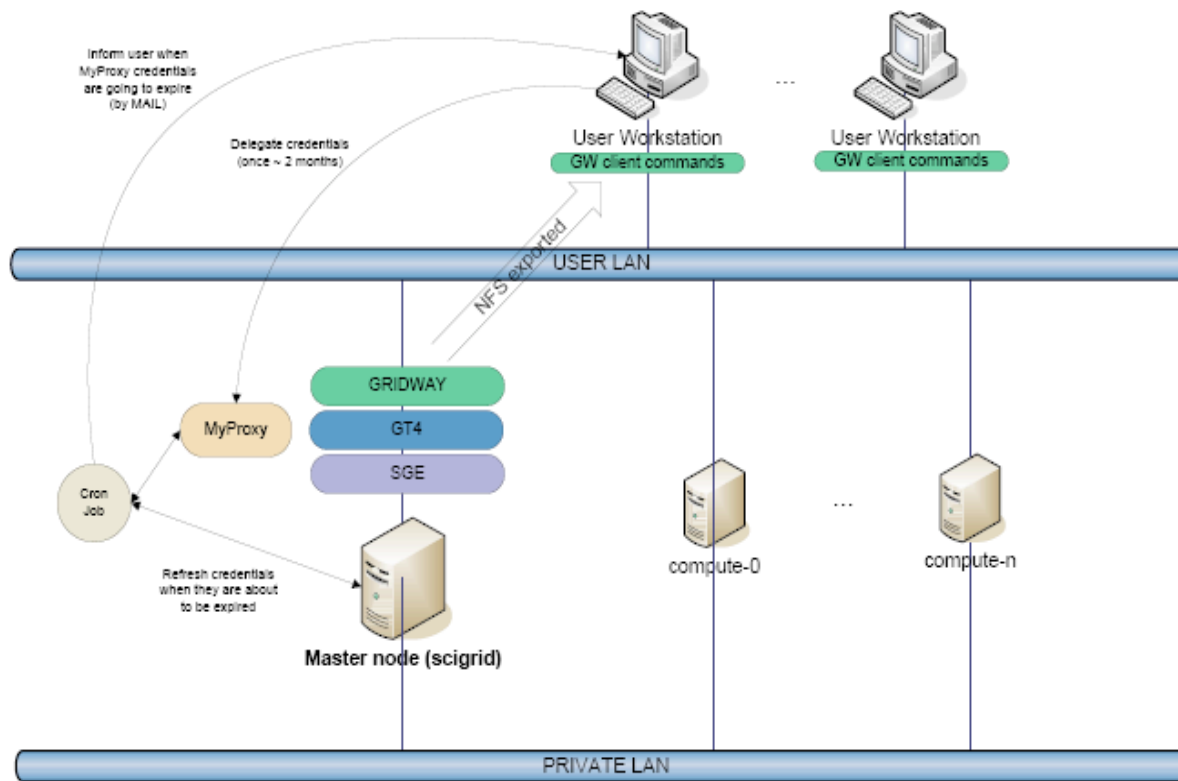
3. Ejemplos de Infraestructuras Grid

3.2. Ejemplos

European Space Agency



- Meta-planificador para el sitio
- Un cluster 20 CPUs, 60 GB memoria principal
- Análisis de datos de las misiones espaciales (DRMAA)

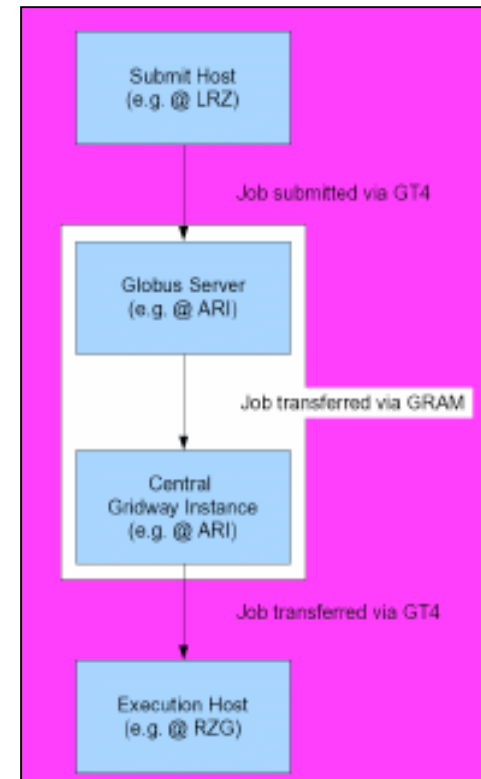
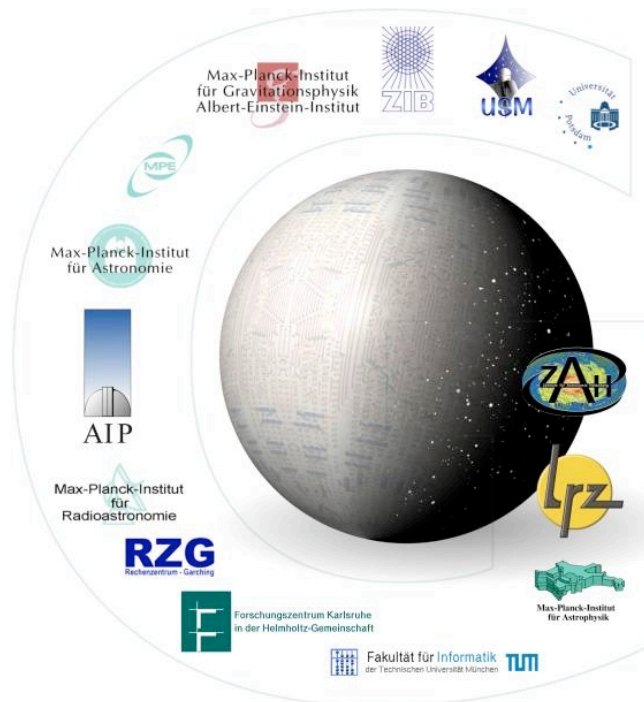


3. Ejemplos de Infraestructuras Grid

3.2. Ejemplos

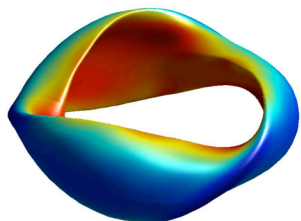
AstroGrid-D, German Astronomy Community Grid

- Meta-planificador para el Grid (interfaz GRAM)
- 22 resources @ 5 sites, 800 CPUs
- Gestión y uso conjunto de recursos computacionales y específicos de astronomía

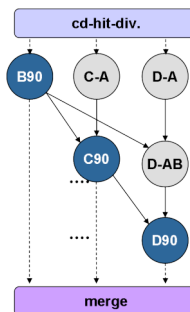
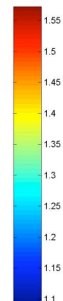


3. Ejemplos de Infraestructuras Grid

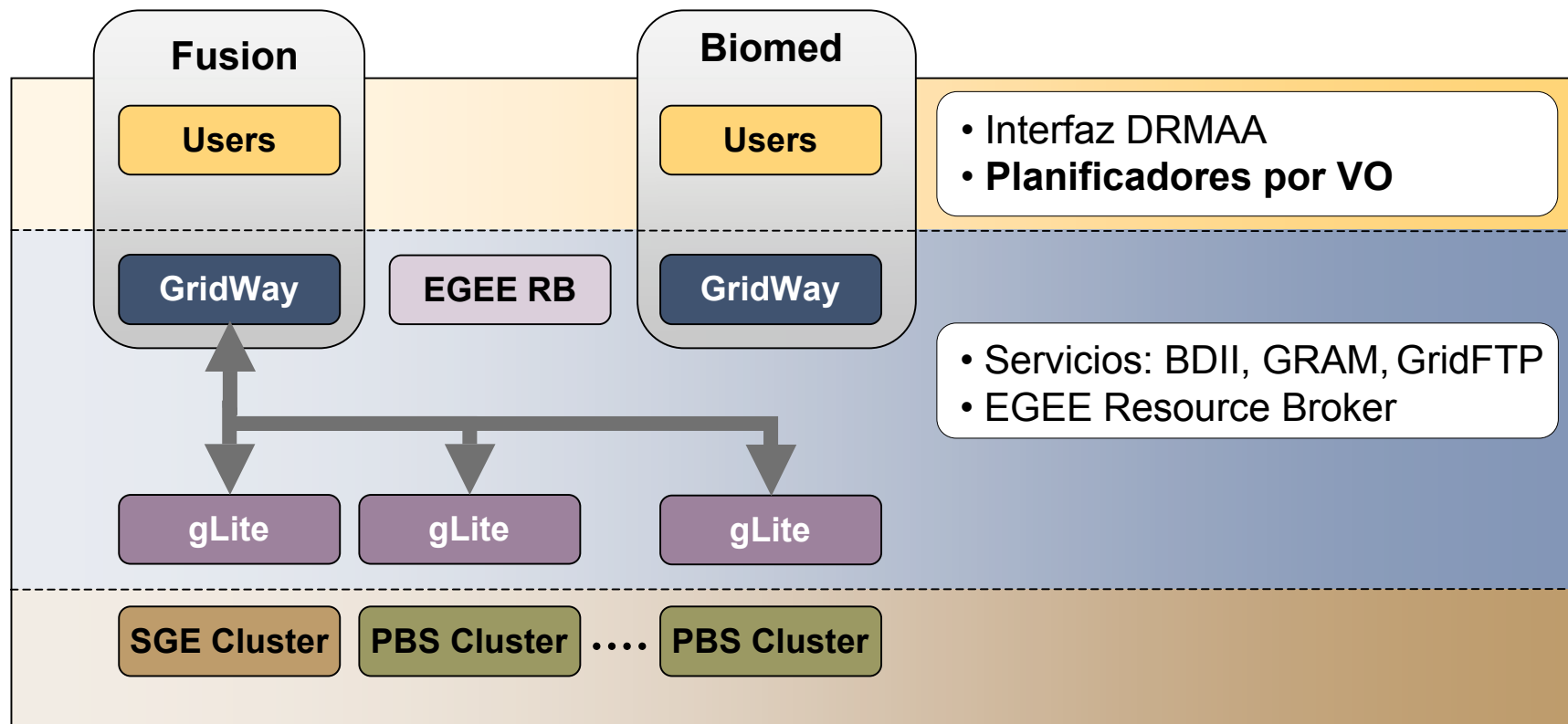
3.2. Ejemplos



Massive Ray Tracing



CD-HIT workflow





4. GRIDIMadrid



1. Recursos Computacionales
2. Middleware Grid
3. Arquitectura de Infraestructuras Grid Computacionales
4. GRIDIMadrid



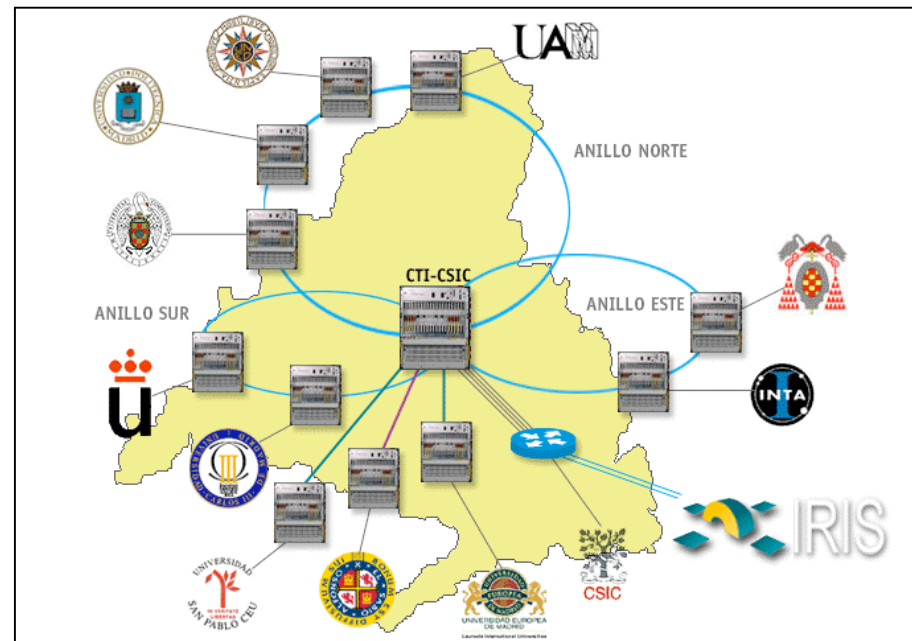
4. GRIDIMadrid



dsa-research.org

Objetivos

- Establecer una infraestructura Grid de investigación en el ámbito de la Comunidad de Madrid, que a su vez proporcione tránsito hacia otras infraestructuras Grid nacionales e internacionales.
- Fomentar la colaboración entre las instituciones, los proyectos y las redes temáticas relacionadas con la investigación en tecnología Grid y migración de aplicaciones.





4. GRIDIMadrid



Estructura de Gestión

- **Coordinador**
- **Coordinador de Operaciones**
- **Representante** de un Centro
- **Responsable de Operaciones** de un Centro
- **Comité Ejecutivo**, compuesto por el Coordinador, el Coordinador de Operaciones y unos Representantes seleccionados

4. GRIDIMadrid

Middleware básico

- Globus Toolkit versión 4.
- Componentes basados en WSRF:
 - Infraestructura de seguridad (GSI)
 - Gestión de recursos (GRAM)
 - Gestión de datos (RFT y GridFTP)
 - Sistema de información (MDS)
- Excepcionalmente, se instalará la versión pre-WS de GRAM.
- Procedimiento para la instalación de otros servicios, previa aprobación del Comité Ejecutivo.

4. GRIDIMadrid

Autorización y Autenticación

- Autenticación con pkIRISGrid: (<http://www.irisgrid.es/pki>).
- Inicialmente, autorización de usuarios centralizada y en un solo nivel.
- Archivo grid-mapfile público (<http://www.gridimadrid.org/grid-mapfile>).
- Asignación cíclica de los DNS de los certificados a las cuentas grid00, grid01, grid02, grid03.
- En último término, un Centro puede denegar de forma autónoma, temporal o permanentemente, el uso de sus recursos a cualquier usuario.
- Procedimiento para reservar la infraestructura, completa o en parte, para la realización de pruebas.

4. GRIDIMadrid

Miembros Actuales

- Universidad Complutense de Madrid
- Universidad Carlos III de Madrid
- Universidad Rey Juan Carlos
- Universidad Autónoma de Madrid
- RedIRIS/Red.es
- CIEMAT

Requisitos de Adhesión

- Centro público de la Comunidad de Madrid.
- Posibilidad de solicitar certificados a la Autoridad de Certificación de IRISGrid.
- Establecer un Representante del Centro con competencia sobre la infraestructura hardware del mismo.
- Establecer un Responsable de Operaciones del Centro.

4. GRIDIMadrid

Más Información

- Página web de GRIDIMadrid

`www.gridimadrid.org`

- Weblog de GRIDIMadrid en Madri+D:

`weblogs.madrimasd.org/gridimadrid`





- Analizar los objetivos y características de una infraestructura Grid computacional
- Identificar los componentes principales de una arquitectura Grid computacional
- Describir ejemplos de infraestructuras Grid como soporte a la investigación colaborativa
- Caso de Estudio: **GRIDIMadrid**