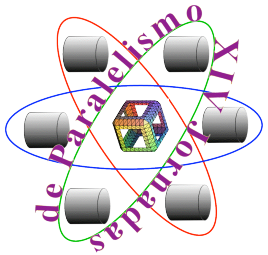
16 de Septiembre de 2003

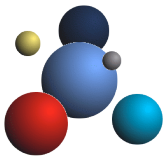
XIVJORNADAS DE PARALELISMO



Estado de la Tecnología Grid y la Iniciativa IrisGrid

Ignacio Martín Llorente
www.dacya.ucm.es/nacho

UCM



Grupo de Arquitectura de Sistemas Distribuidos y Seguridad
Departamento de Arquitectura de Computadores y Automática
Universidad Complutense de Madrid

Laboratorio de Computación Avanzada y Simulación
Centro de Astrobiología CSIC/INTA
Asociado al NASA Astrobiology Institute

¿Qué es la computación en red?

¿Qué alternativas existen?

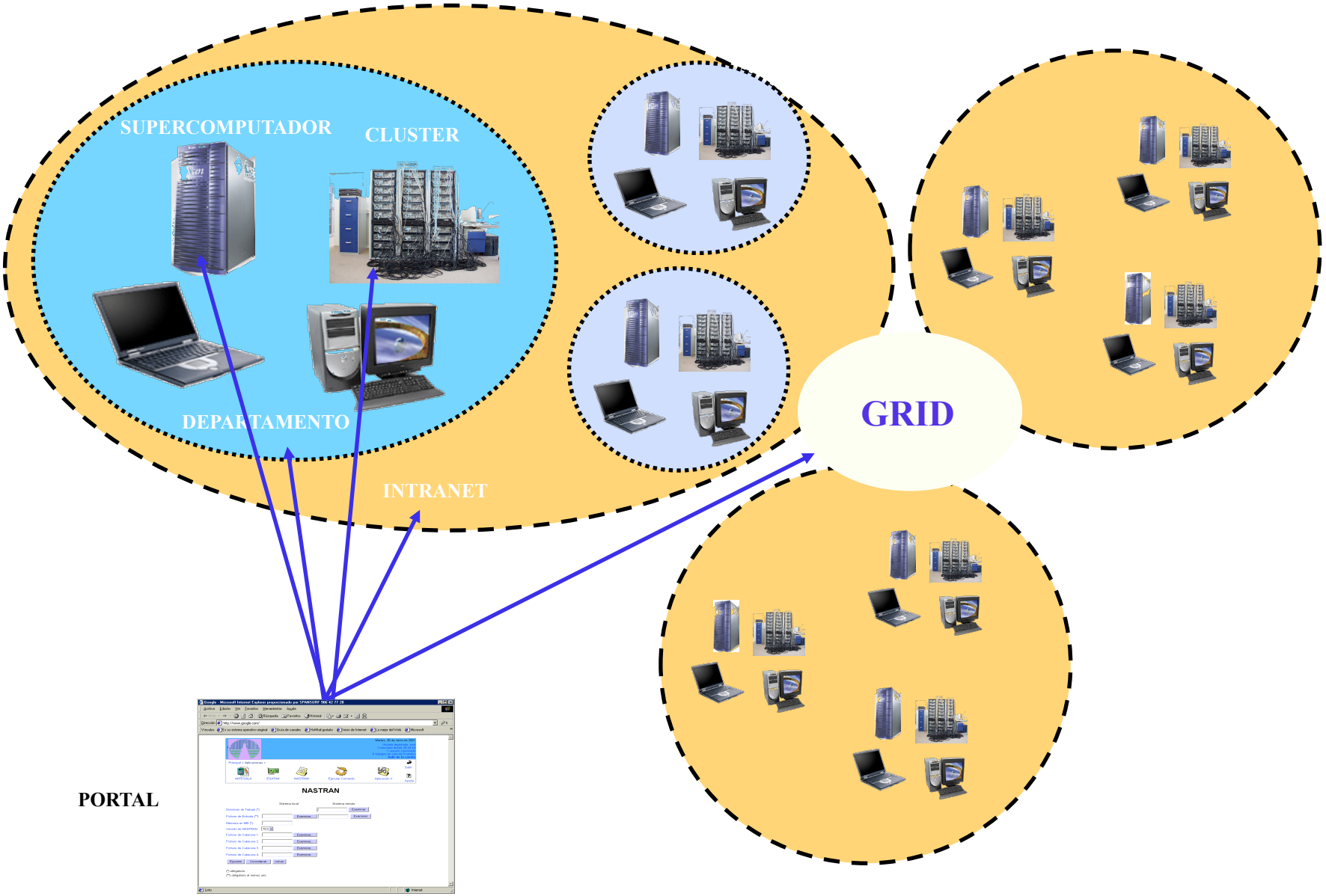
¿Qué es Grid Computing?

¿Cuál es el estado de la tecnología?

La iniciativa IRISGRID

El banco de pruebas IRISGRID

La Tecnología Grid constituirá la próxima generación de Internet


SUPERCOMPUTADOR
CLUSTER
DEPARTAMENTO
INTRANET
GRID
PORTAL
NASTRAN

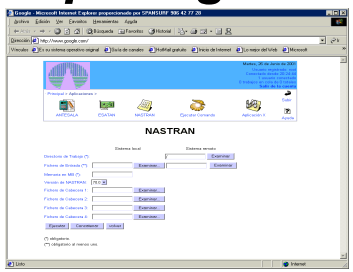
Introducción

Cluster Computing

Intranet Computing

Grid Computing

Computing Portals



Internet Computing

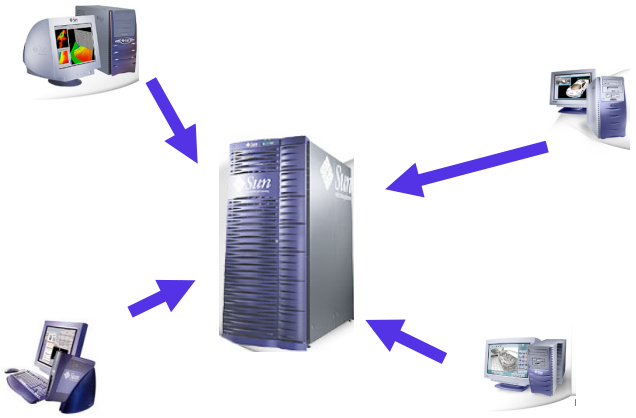
¿Cuál es la visión clásica de la supercomputación?

Tipos de aplicaciones que requieren potencia computacional

- Ejecución de una aplicación en menos tiempo (*Alto Rendimiento*)
  - ✓ Paralelización previa de la aplicación
- Ejecución de un número mucho mayor de aplicaciones (*Alta Productividad*)

Solución clásica

- Computación centralizada basada en servidor



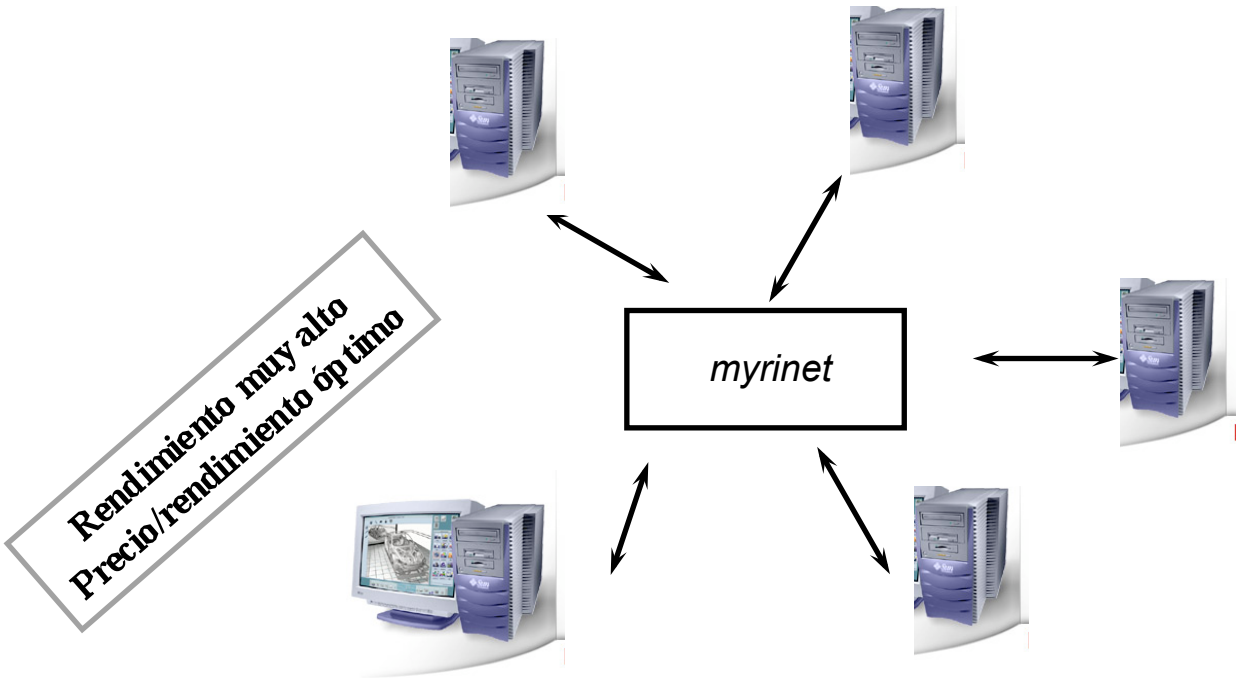
Problemas de la supercomputación basada en servidor

- Falta de escalabilidad
- Equipos muy caros
- Mantenimiento muy caro
- Las demandas de cálculo podrían ser puntuales
- Posibles problemas de fiabilidad

¿Cuál es la diferencia entre una sistema multiprocesador y un conjunto de equipos interconectados por red?

Alternativa económica a la adquisición de un sistema multiprocesador

- ⇨ Cluster de estaciones o computadores personales homogéneo dedicado a computación paralela
- ⇨ Suele estar dotada de una red avanzada basada en *router* con *Fast Ethernet* (LAN) o *Myrinet* (SAN)

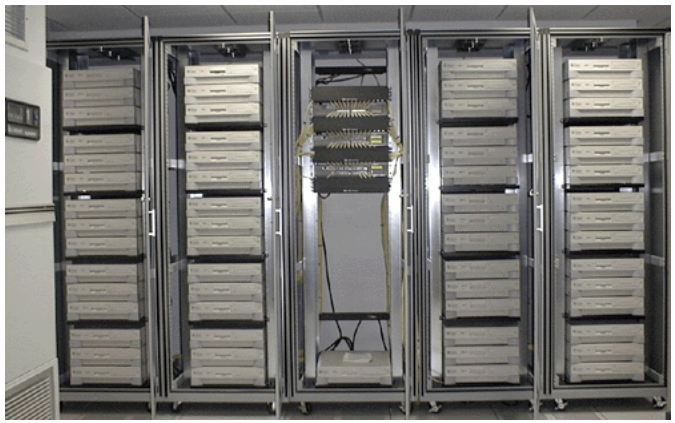


Ventajas:

- Mejor relación coste/rendimiento (3-10 veces)

Inconvenientes:

- Coste de las comunicaciones muy alto
  - Bus lento
  - Acceso secuencial al bus
  - Sobrecarga TCP/IP
- Mantenimiento
- Modelo de programación



Conclusión:

- Buena solución para aplicaciones con grano medio o HTC (*Computación de Alta Productividad*)

Front-end: Alpha DS20 a 500 Mhz

Nodos de trabajo: 30 nodos Alpha DS10 a 466 Mhz

Memoria RAM: 8 Gb

Disco: 300 Gb.

Rendimiento pico: 30 Gflops

Red
- **Servicios:** Fast-Ethernet (100Mbits/seg)
- **Comunicación:** ServerNet II (1Gbit/seg)

Sistema Operativo: Alpha 7.0 de Suse

Software de gestión de carga:PBS

Librerías de paralelización: MPI

Gestión de usuarios: NFS y NIS

Babieca

http://dalbe.inta.es/~LCASAT/trab/o_babieca.htm

¿Cómo implementar una máquina paralela de bajo coste?

HispaCluster	`www.hispacluster.org`
IEEE TFCC (*reports*)	`www.ieeetfcc.org`
Sun BluePrints	`www.sun.com/blueprints`

Ejemplos

Beowulf Project en CESDIS	`www.beowulf.org`
Avalon	`cnls.lanl.gov/Internal/Computing/Avalon/`
Sandia Labs Computational Plant	`www.cs.sandia.gov`
Coral en ICASE	`www.icase.edu`
Babieca en el CAB	`dalbe.inta.es/~LCASAT/trab/o_babieca.htm`

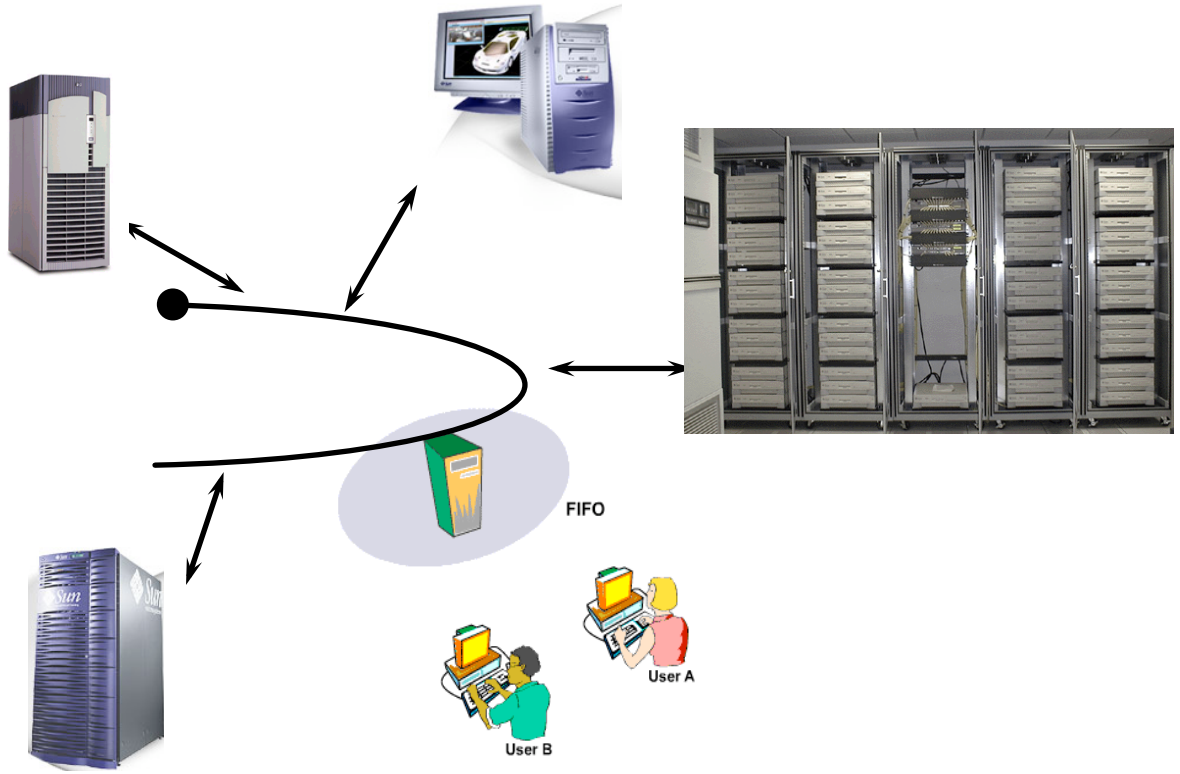
Software

- Herramientas de distribución de carga, NFS, NIS...
- Interesante el software Mosix de la Universidad de Israel (`www.mosix.cs.huji.ac.il`)
- Herramientas de gestión de colas (PBS, LSF, SGE...)

¿Cuánto tiempo al día permanecen desaprovechados sus recursos ?

¿Si tengo un pico de demanda de CPU puedo utilizar mis equipos distribuidos?

- Utilización de los equipos de una red departamental para ejecutar trabajos secuenciales o paralelos por medio de una herramienta de gestión de carga
- Explotación de potencia computacional distribuida



Ventajas

- Aumentar el aprovechamiento de los recursos informáticos
- Ciclos de CPU a coste bajo
- Mejora de la escalabilidad
- Mejora de fiabilidad
- Facilidad de administración
- Facilidad de sustitución de equipos obsoletos

Software disponible

Sun Grid Engine de Sun Microsystems	`www.sun.com/gridware`
LSF de Platform Computing	`www.platform.com`
Condor de la Universidad de Wisconsin	`www.cs.wisc.edu/condor`

Específico para ejecución paramétrica

Ejemplos

InnerGrid de Grid Systems	`www.gridsystems.com`
appLES de la Universidad de California	`apples.ucsd.edu`
Nimrod de la Universidad de Monash	`www.csse.monash.edu.au/~rajkumar/ecogrid/`

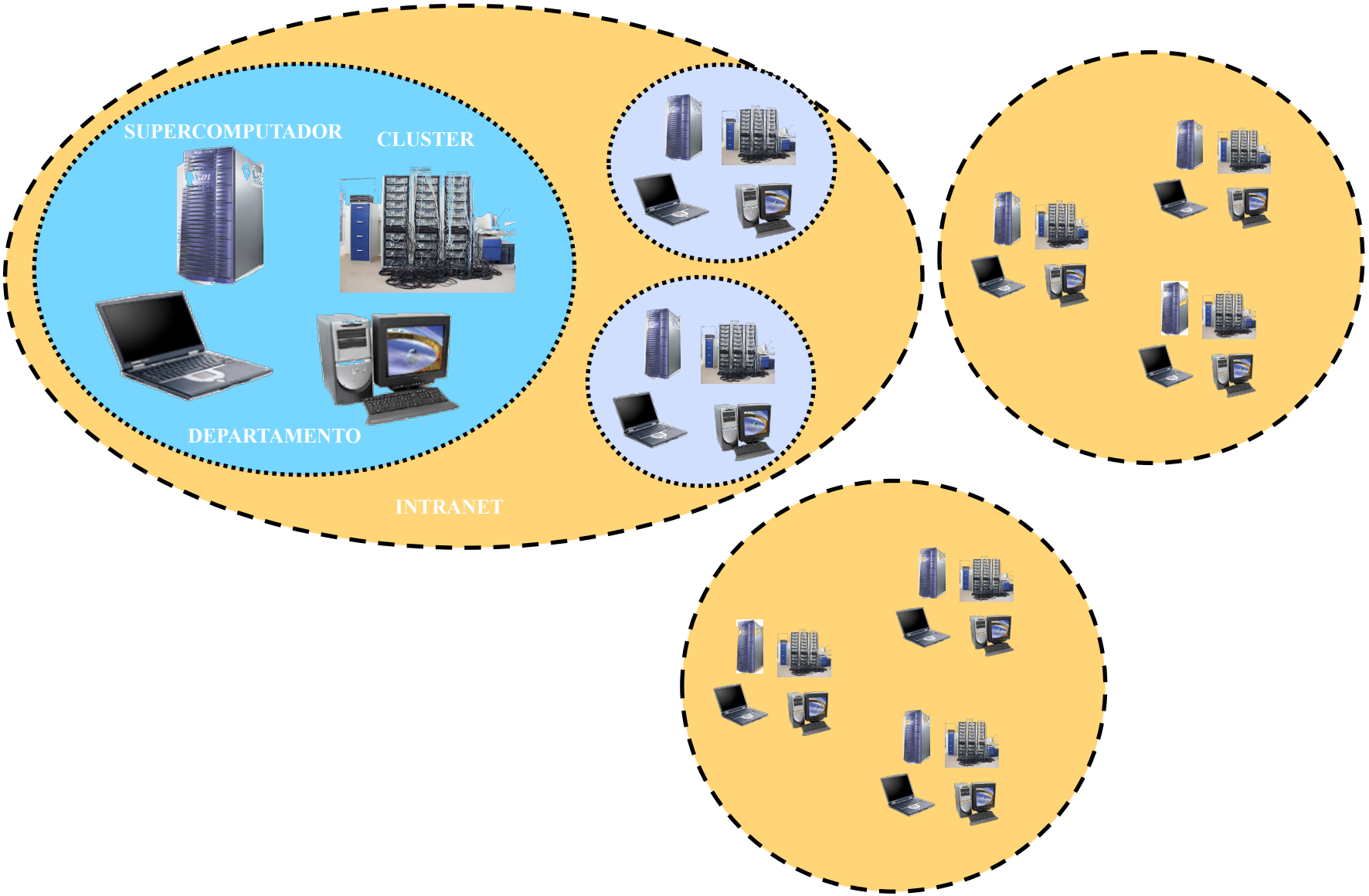
Otras herramientas de empresas como Avaki, United Devices, Entropia, Parabon...

Sin embargo:

- No pueden gestionar recursos fuera del dominio de administración
  -Algunas herramientas (Condor, LSF o SGE) permiten colaboración inter-departamental asumiendo la misma estructura administrativa
- No respetan las políticas de seguridad y de gestión de recursos de las organizaciones
- Protocolos e interfaces básicos no basados en estándares abiertos
- El único recurso que gestionan es la CPU
  -¿Qué ocurre con los datos compartidos entre organizaciones?

Por tanto:

- Escalabilidad limitada a la organización en picos de demanda
- No puedo amortizar mis recursos cuando están desaprovechados
- No puedo compartir recursos con otras organizaciones


SUPERCOMPUTADOR
CLUSTER
DEPARTAMENTO
INTRANET

¿Cómo puedo salir de los límites de mi centro o laboratorio?

¿Puedo ceder mis recursos a otras organizaciones?

¿Puedo compartir recursos con otras organizaciones?

“It’s hard to make predictions, especially about the future”

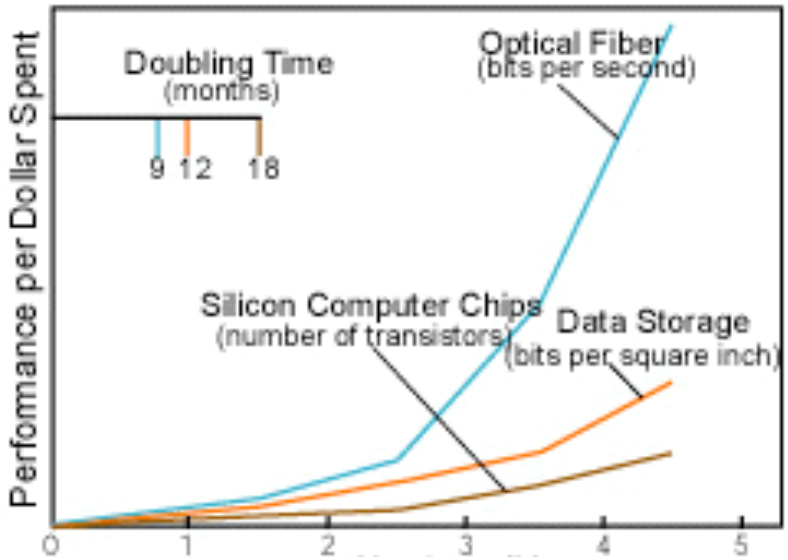
Yogi Berra

1. Necesito más potencia de cálculo

- La potencia de los supercomputadores solo crece linealmente (dos órdenes de magnitud cada 10 años) (`www.top500.org`)

La capacidad de almacenamiento se dobla cada 12 meses
El ancho de banda de red se dobla cada 9 meses
El rendimiento de un procesador se dobla cada 18 meses

- 1986 to 2000
  - Computers: x 500
  - Networks: x 340,000
- 2001 to 2010
  - Computers: x 60
  - Networks: x 4000



Moore's Law vs. storage improvements vs. optical improvements. Graph from **Scientific American** (Jan-2001) by Cleo Vilett, source Vined Khoslan, Kleiner, Caufield and Perkins.

Conclusiones:

Un único sistema no será capaz de analizar los datos que almacenen sus discos
Un único centro no podrá analizar el volumen de información generado
La red permitirá de forma eficiente usar recursos distribuidos
(1 orden de magnitud de diferencia entre procesamiento y red)

2. Los equipos son caros y difíciles de mantener

- ¿Cómo puedo amortizar la inversión realizada?

3. ¿Qué hago si tengo un pico de demanda de cálculo y no tengo presupuesto para adquirir un supercomputador?

4. Desequilibro de carga en el tiempo

- ¿Cómo puedo conseguir una carga más homogénea?

“Cuando Internet sea tan rápido como los buses internos de un computador, éste se desintegrará en la red en un conjunto de recursos de propósito específico”

Gilder Technology Report, Junio 2000

⇨ Nueva tecnología cuyo objetivo es la **compartición de recursos en Internet** de forma uniforme, transparente, segura, eficiente y fiable

⇨ Análoga a las **redes de suministro eléctrico**:

- Ofrecen un único punto de acceso a un conjunto de recursos distribuidos geográficamente **en diferentes dominios de administración** (supercomputadores, clusters, almacenamiento, fuentes de información, instrumentos, personal, bases de datos...)

⇨ **La tecnología Grid es complementaria a las anteriores**

- Permite interconectar recursos en diferentes dominios de administración respetando sus políticas internas de seguridad y su software de gestión de recursos en la Intranet

A three point checklist

A Grid is a system that...

1) ...coordinates resources that are not subject to a centralized control...

2) ...using standard, open, general-purpose protocols and interfaces...

3) ...to deliver nontrivial qualities of services.

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

Beneficios

- ⇨Alquiler de recursos
- ⇨Amortización de recursos propios
- ⇨Gran potencia de cálculo a precio bajo sin necesidad de adquirir equipamiento
- ⇨Mayor colaboración y compartición de recursos entre varios centros
- ⇨Creación de organizaciones virtuales
- ⇨Negocios basados en proveer recursos

Desafíos técnicos

- Recursos heterogéneos
- Descubrimiento, selección, reserva, asignación, gestión y monitorización de recursos
- Desarrollo de aplicaciones
- Desarrollo de modelos eficientes de uso
- Comunicación lenta y no uniforme

Desafíos socioeconómicos

- Organizativos: Dominios de administración, modelo de explotación y costes, política de seguridad...
- Económicos: Precio de los recursos, oferta/demanda...

Infraestructura Grid: Servicios y protocolos básicos para interconectar recursos

- **Globus: www.globus.org**
- Legion: www.cs.virginia.edu/~legion/
- Polder: www.science.uva.nl/research/scs/PSCS4.html
- MOL: www.uni-paderborn.de/pc2/projects/mol/

Toolkits de aplicación: Módulos para construir aplicaciones Grid específicas

- Nimrod/G:www.csse.monash.edu.au/~rajkumar/ecogrid/
- Condor: www.cs.wisc.edu/condor/condorg
- Data Grid: www.eu-datagrid.org
- Portal: dast.nlanr.net/Projects/GridPortal
- MPI/G: www.globus.org/mpi
- GridWay: www.dacya.ucm.es/asds

Aplicaciones Grid

- CACTUS (www.cactuscode.org)
- Virtual Laboratory (www.csse.monash.edu.au/~rajkumar/vlab/)

Bancos de prueba Grid: Sistemas Grid para prototipos y producción

- *NASA's Information Power Grid:* www.nas.nasa.gov/About/IPG/ipg.html
- *European Data Grid:* www.eu-datagrid.org
- *Grid Physics Network:* www.griphyn.org
- *Teragrid:* www.teragrid.org
- *Tidewater Research Grid:* www.tidewaterrgp.org

Globus Toolkit

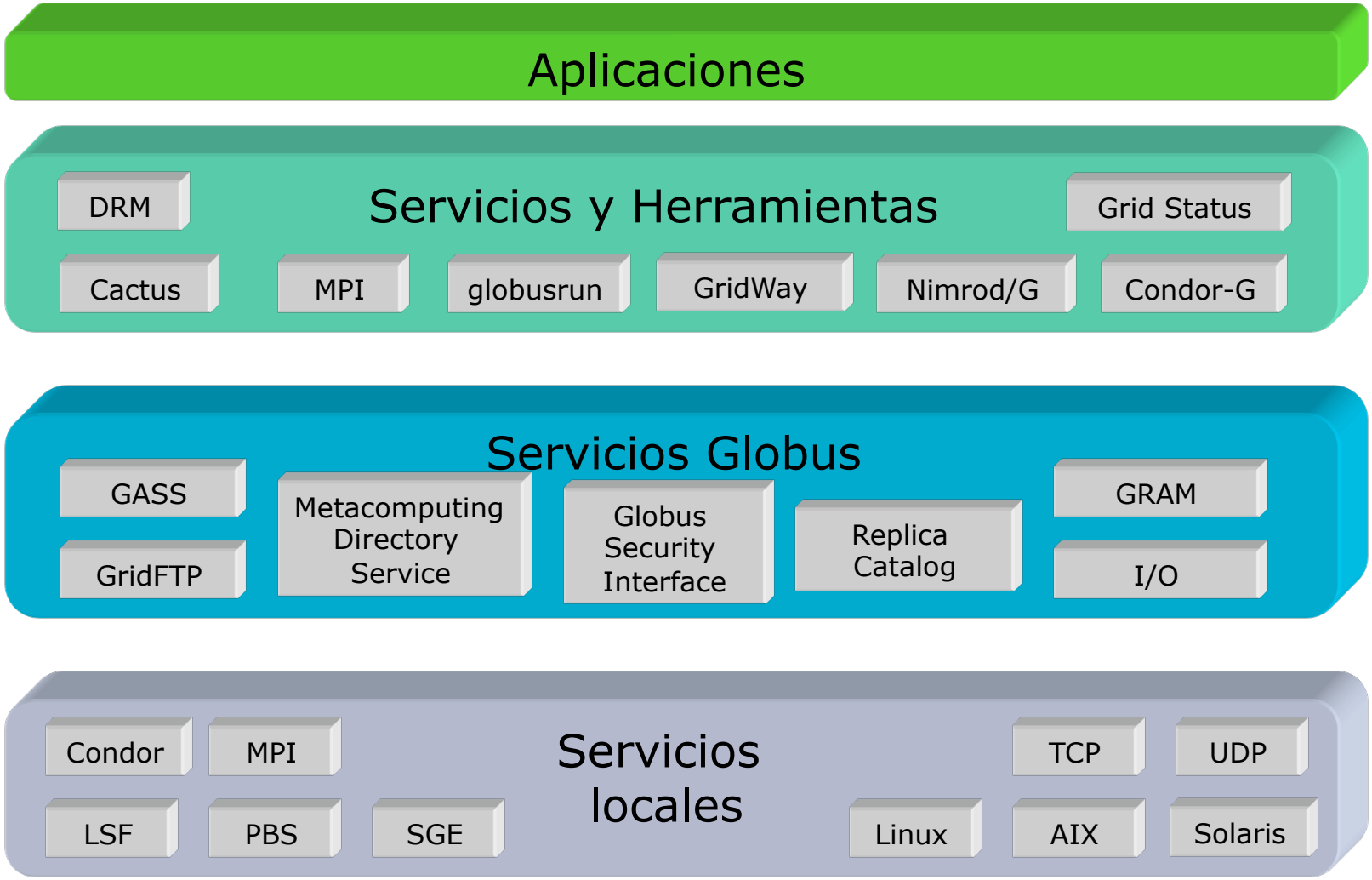
Permite compartir recursos localizados en diferentes dominios de administración, con diferentes políticas de seguridad y gestión de recursos.

Globus es…

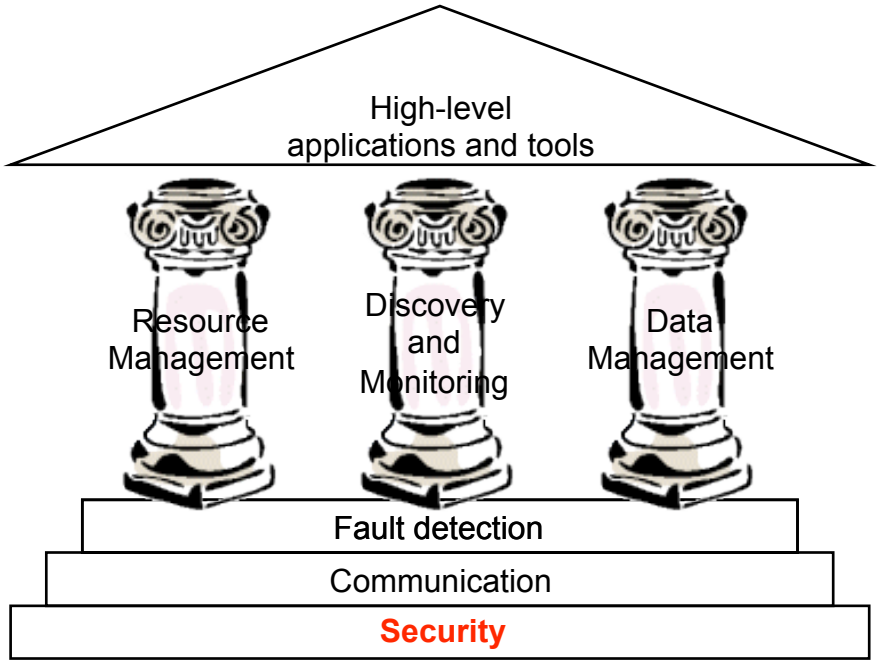
- Un middleware software
- Un conjunto de librerías, servicios y APIs

Globus no es...

- Una herramienta de usuario o planificador
- Una aplicación


Aplicaciones
DRM
Servicios y Herramientas
Grid Status
Cactus
MPI
globusrun
GridWay
Nimrod/G
Condor-G
Servicios Globus
GASS
GridFTP
Metacomputing Directory Service
Globus Security Interface
Replica Catalog
GRAM
I/O
Condor
MPI
LSF
PBS
SGE
Servicios locales
TCP
UDP
Linux
AIX
Solaris

Seguridad



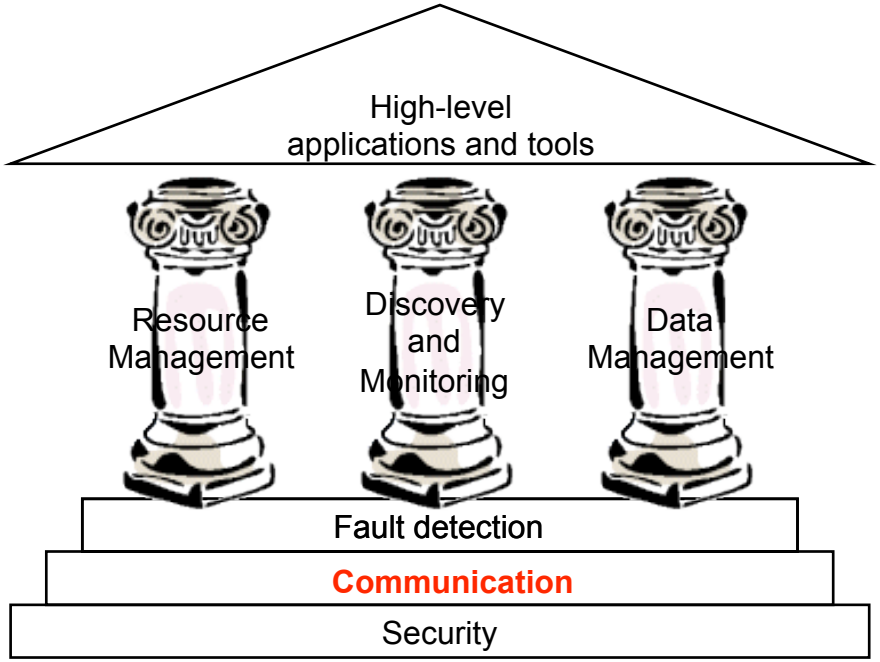
Globus Security Infrastructure (GSI)

- Transport Layer Security (TLS), a.k.a. Secure Socket Layer (SSL), basado en OpenSSL
- Public Key Infrastructure (PKI) con certificados X.509
- Generic Security Services API (GSS-API)

Community Authorization Service (CAS)

- GSI

Comunicación

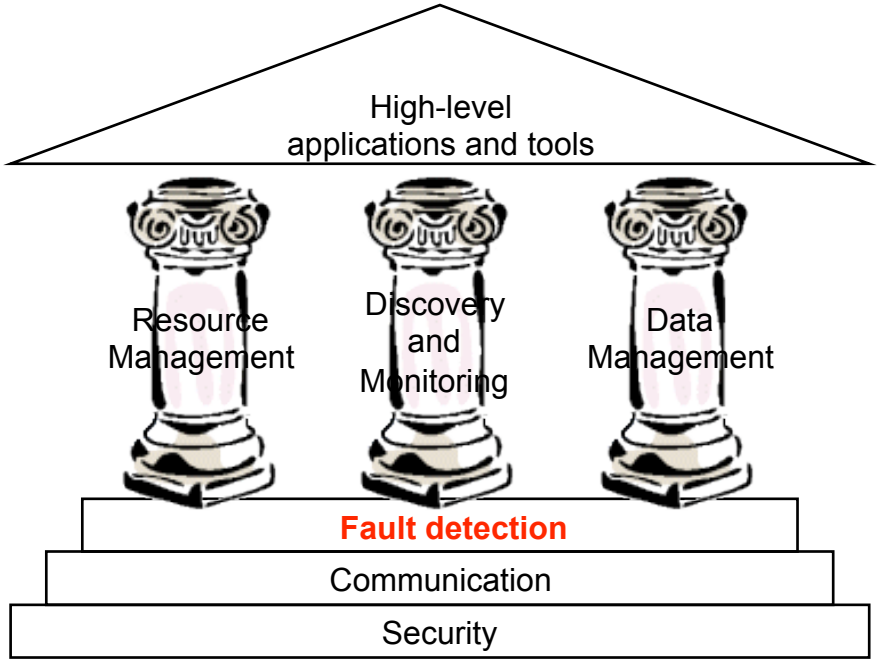


Nexus (en desuso)

Globus I/O

- Transport Control Protocol (TCP)
- User Datagram Protocol (UDP)
- File I/O
- GSI

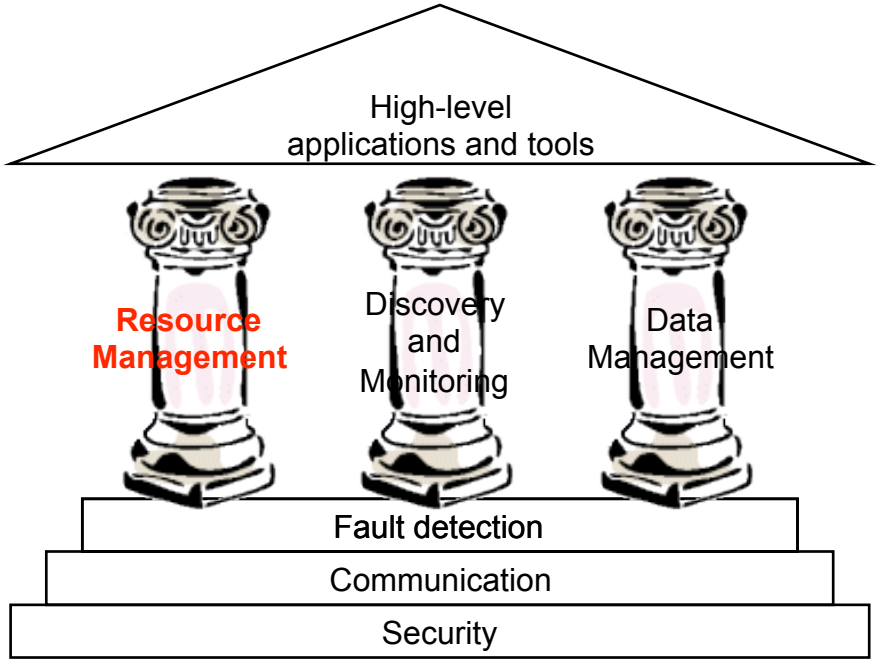
Detección de fallos



Heart Beat Monitor (HBM)

- Nexus
- Globus I/O

Gestión de recursos



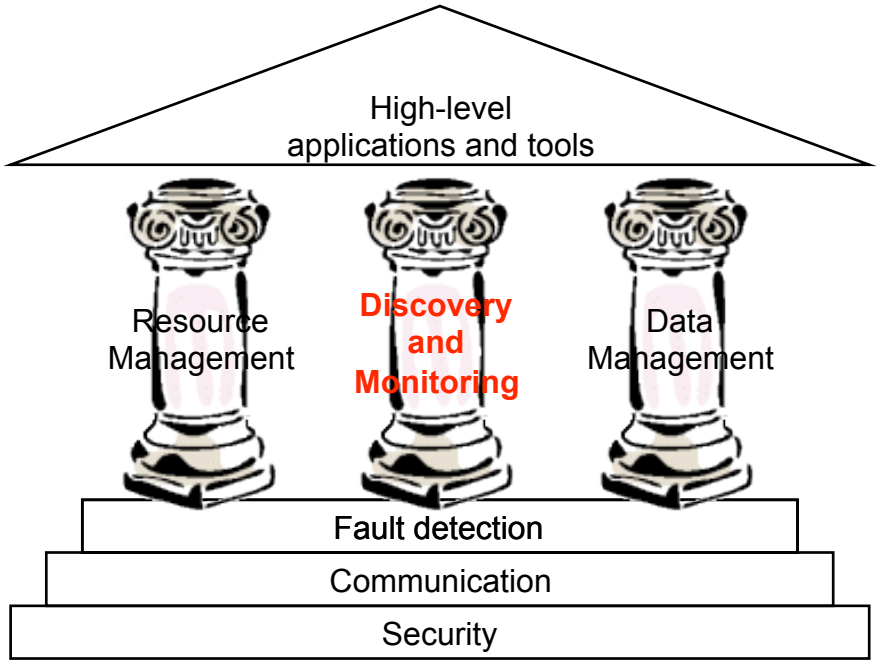
Globus Resource Allocation Manager (GRAM)

- Gestores locales: PBS, LSF, NQE, LoadLeveler, UNIX fork...
- Resource Specification Language (RSL)
- Globus I/O
- GSI

Dynamic User Runtime On-line Co-allocator (DUROC)

- GRAM
- Globus I/O

Descubrimiento y monitorización de recursos

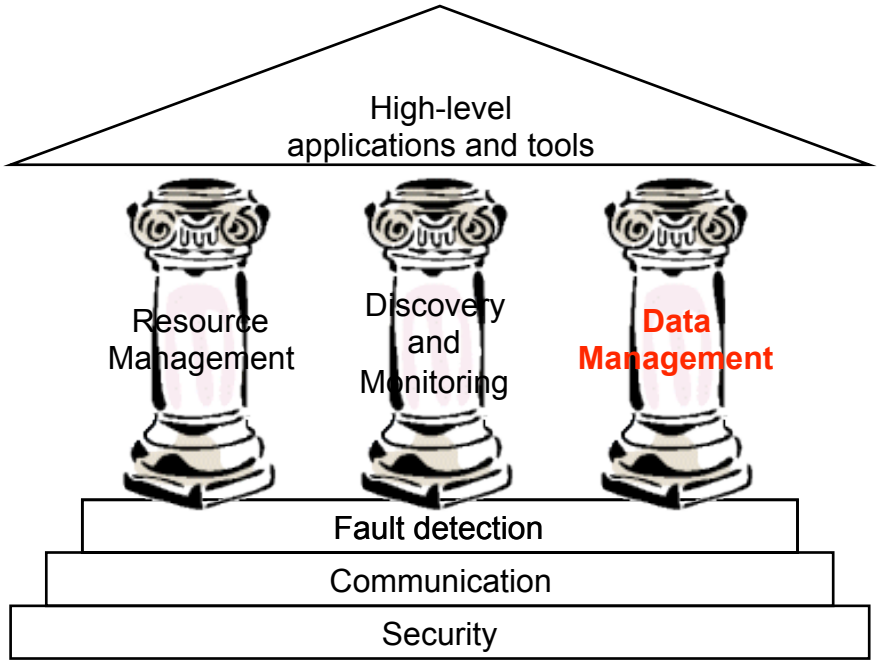


Monitoring and Discovery Service (MDS)
- GIIS
- GRIS

Globus Index Information Service (GIIS)
- GRIS
- Lightweight Directory Access Protocol (LDAP), basado en OpenLDAP
- GSI

Globus Resource Information Service (GRIS)
- GRAM
- LDAP
- GSI

Gestión de datos



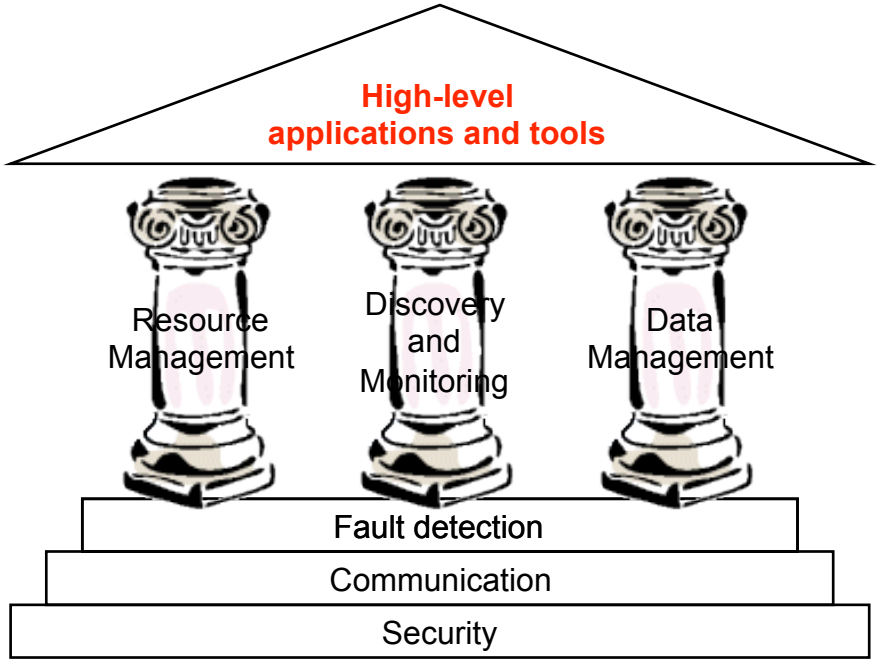
Global Access to Secondary Storage (GASS)
- Globus I/O
- GridFTP
- GSI

Grid File Transfer Protocol (GridFTP)
- FTP
- GSI

Globus Replica Catalog
- LDAP
- GSI

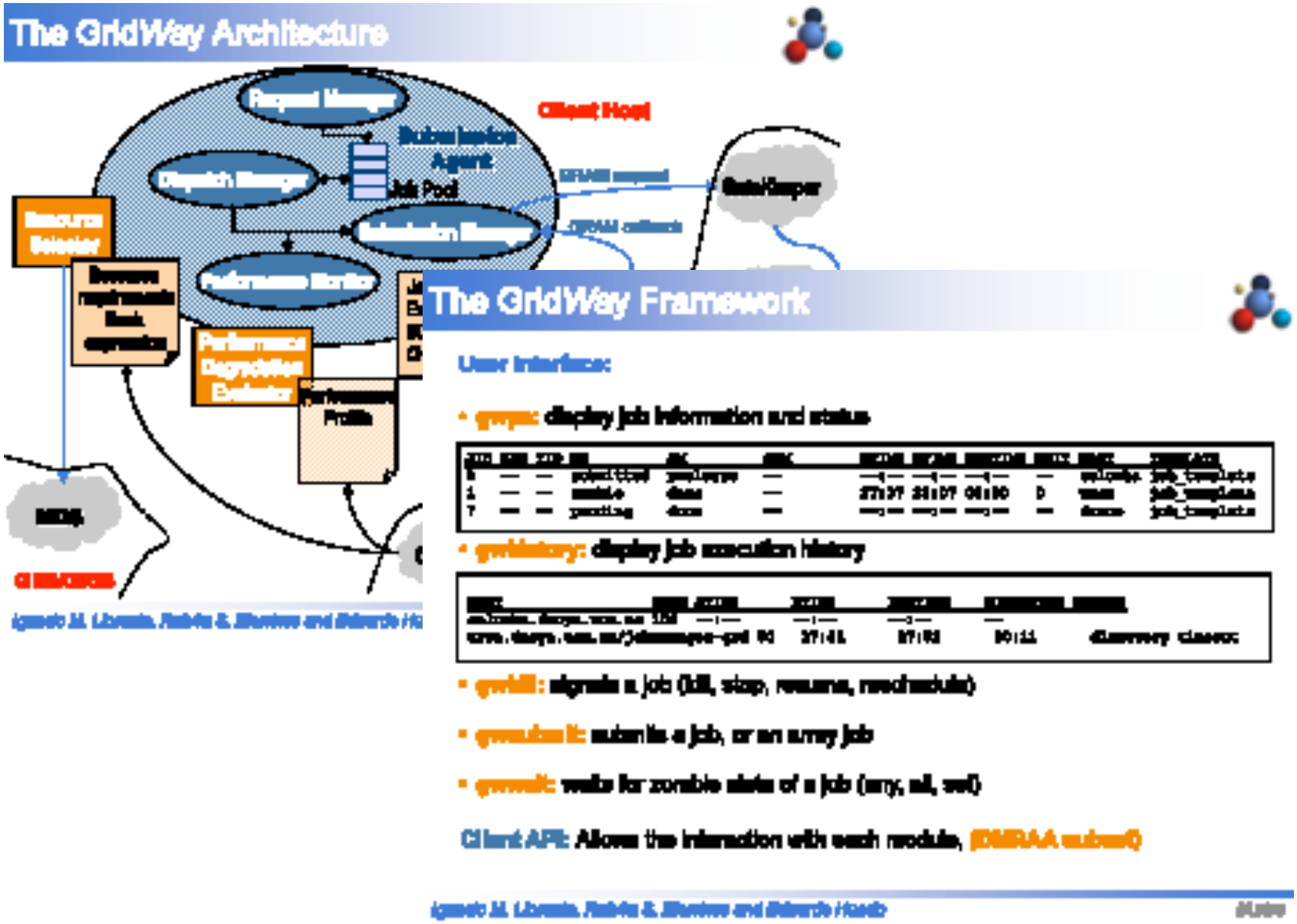
Globus Replica Management
- Globus Replica Catalog
- GridFTP
- GSI

Servicios de alto nivel y herramientas

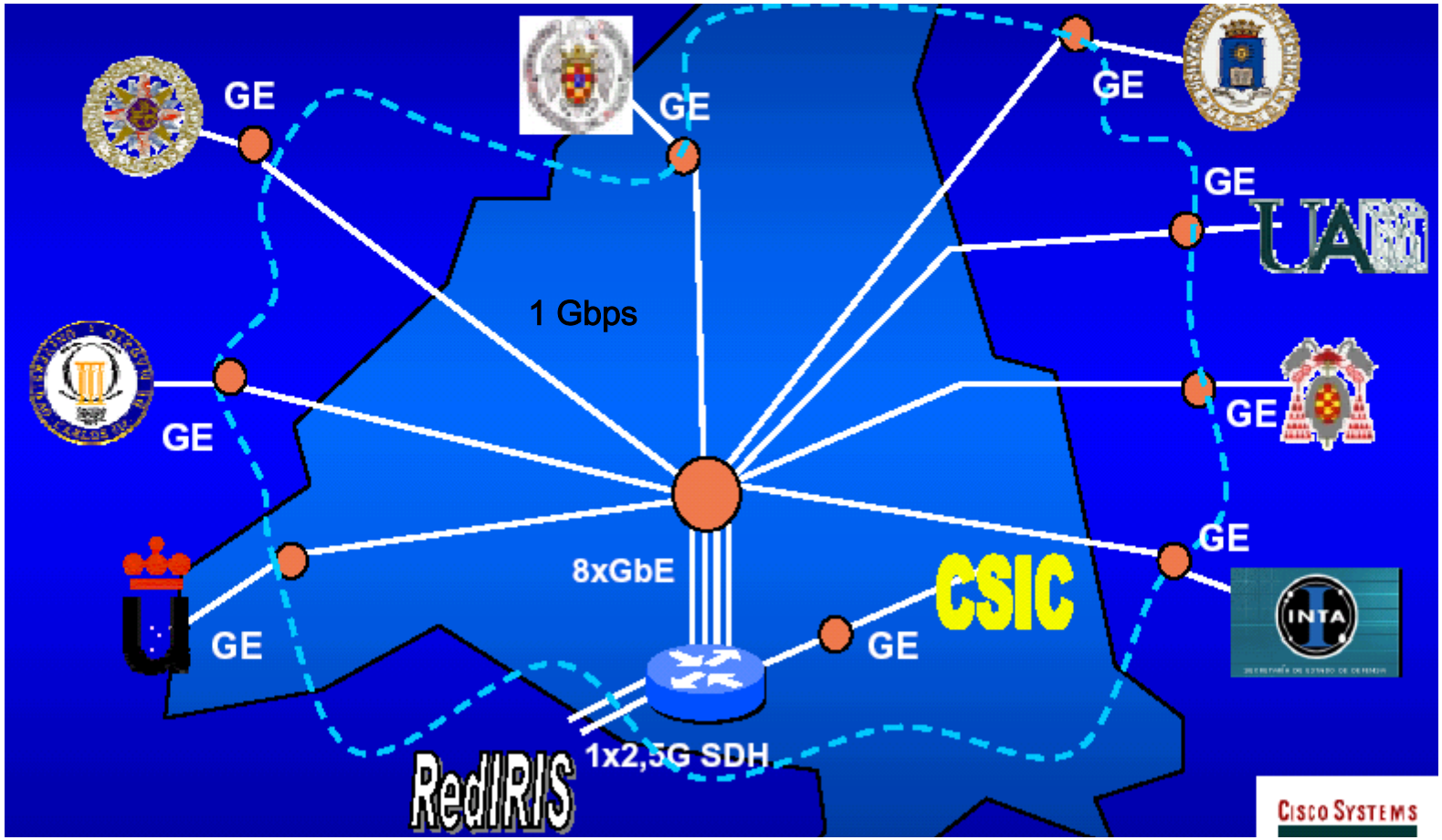


Herramientas de alto nivel

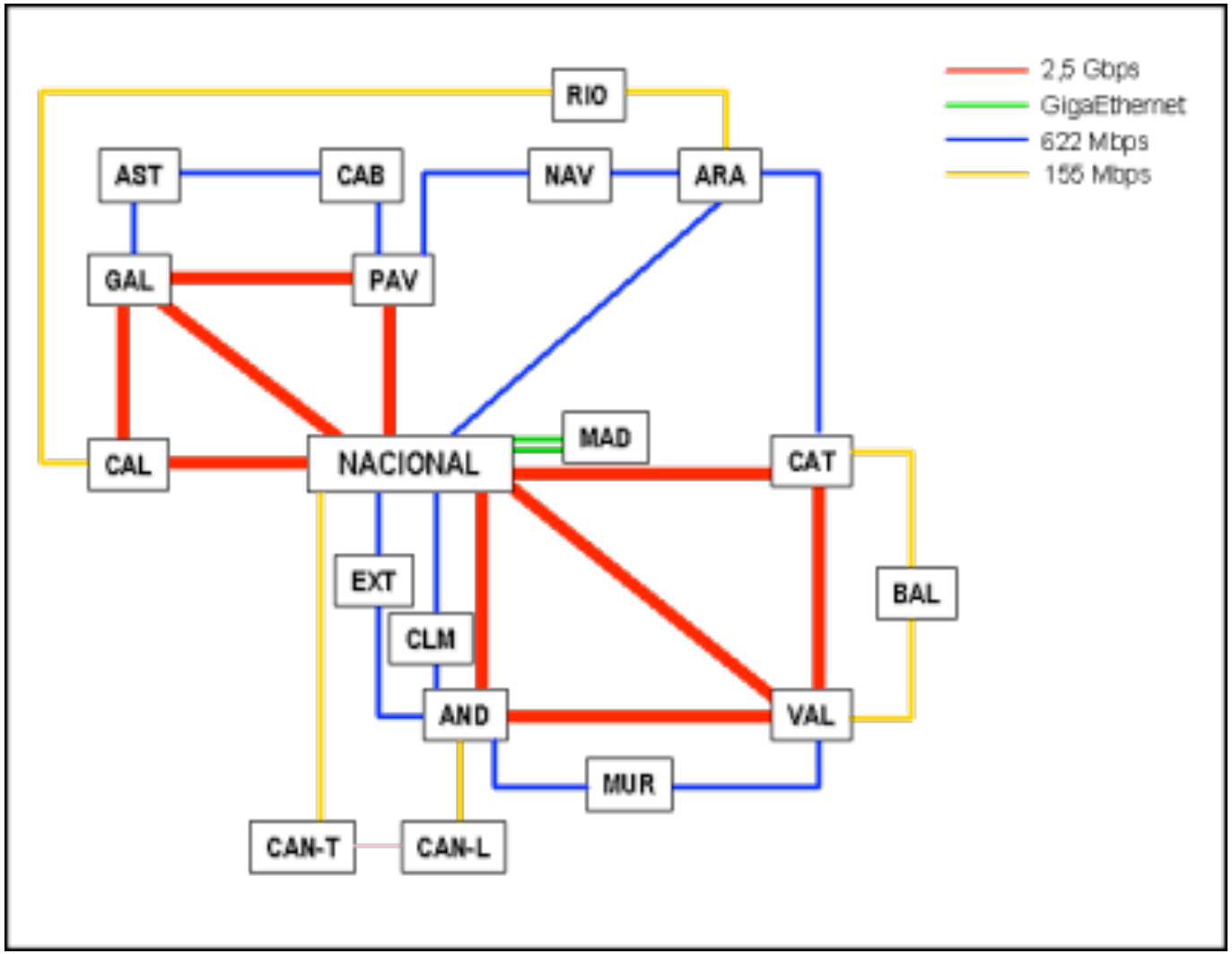
- Órdenes Globus
- MPICH-G2
- Condor-G
- Nimrod/G
- **GridWay**


The GridWay Architecture
Client Host
Job Pool
The GridWay Framework
User Interface:
• gwps: display job information and status
• gwhistory: display job execution history
• gwkill: signals a job (kill, stop, resume, reschedule)
• gwsubmit: submits a job, or an array job
• gwwait: waits for zombie state of a job (any, all, set)
Client API: Allows the interaction with each module, DRMAA submit()

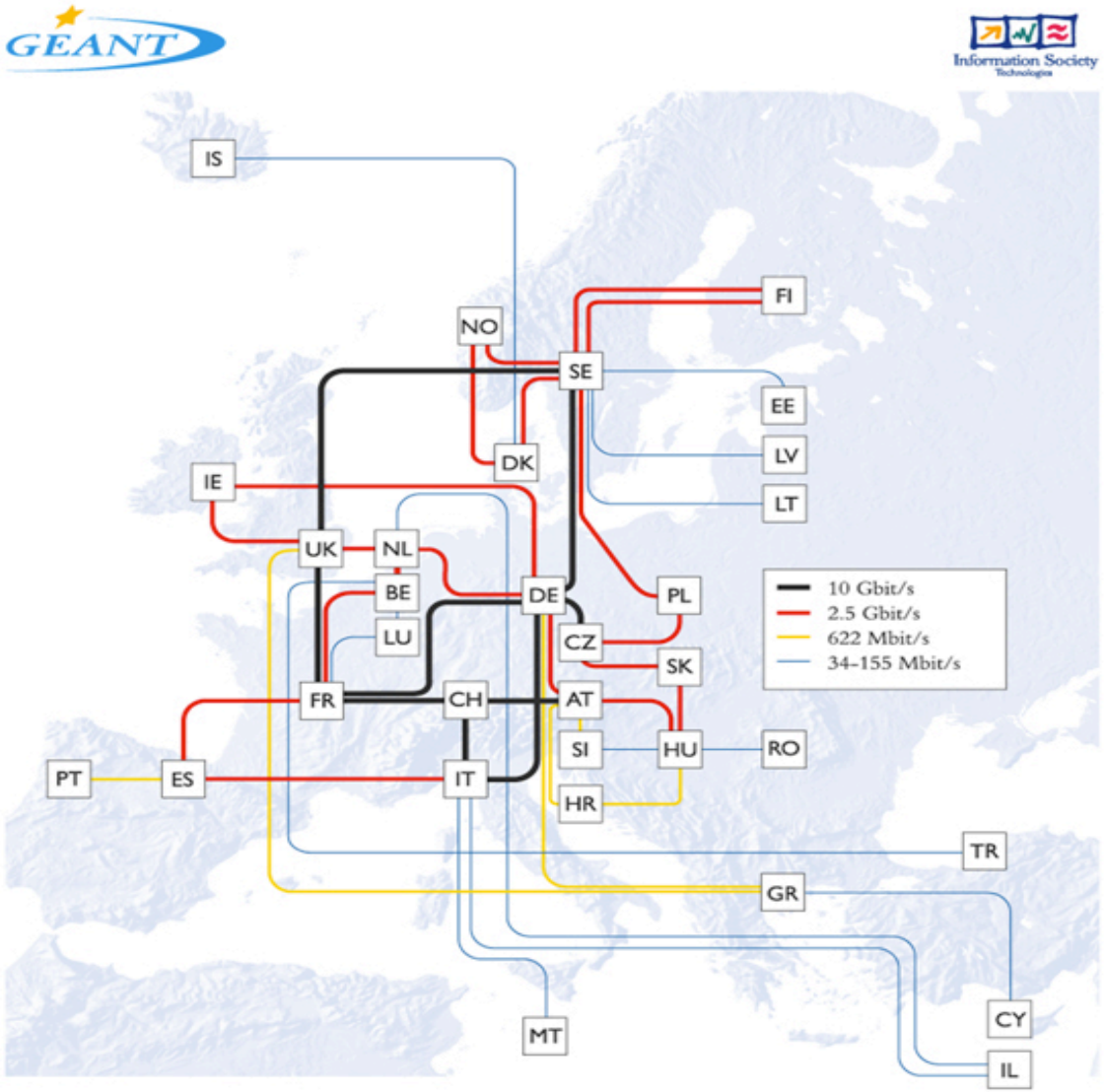
Conexión en la Comunidad de Madrid por medio de la nueva
Red Telemática de Investigación



Conexión nacional por medio de RedIris-2

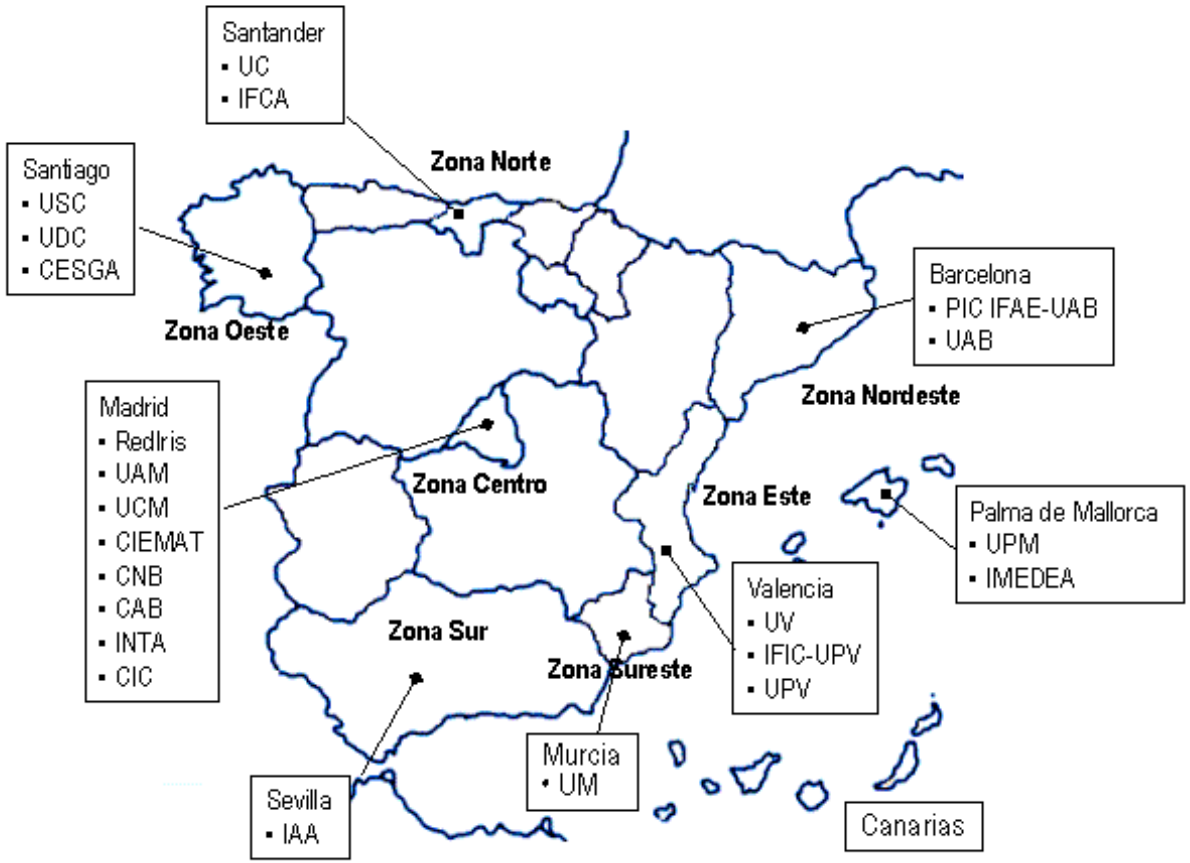


Conexión por medio de la nueva red paneuropea Geant



www.cab.inta.es/~**CABGrid**/

www.rediris.es/**irisgrid**/

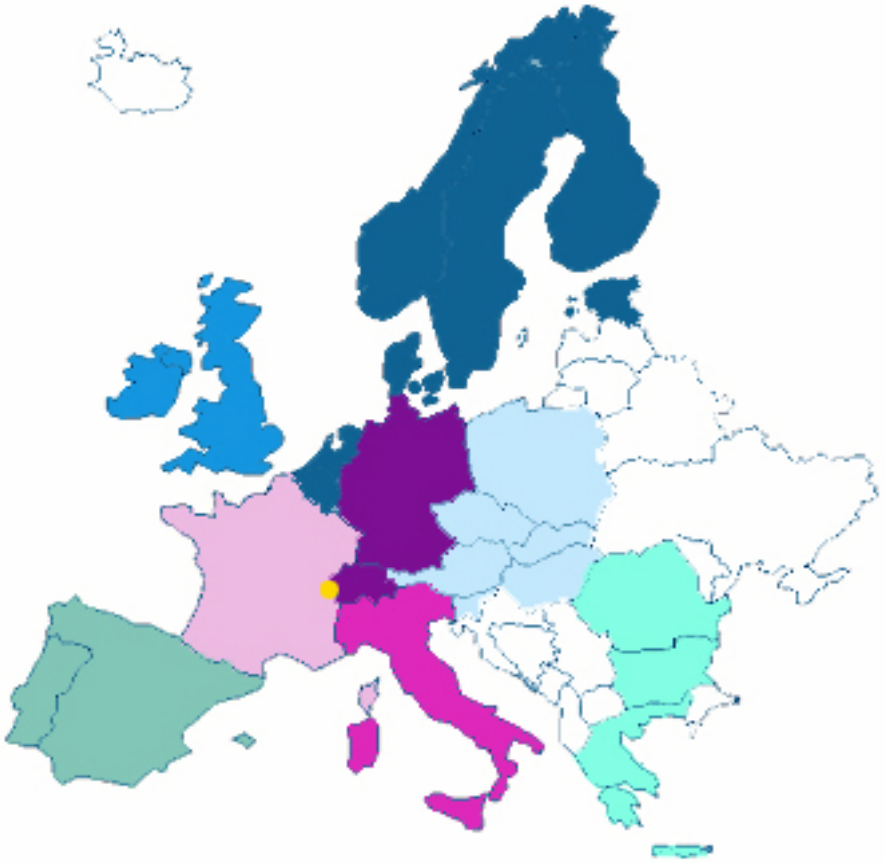


www.rediris.es/**irisgrid**/

Nombre	V.O.	Modelo	CPUs	Rendimiento	S.O.	GRAM
bw	CESGA	Intel PIII	16	16 GFLOPS	Linux 2.4	PBS
svg	CESGA	Intel PIII	16	8.8 GFLOPS	Linux 2.4	PBS
Kadesh	CEPBA	Power3	128	192 GFLOPS	AIX5.1	LoadLeveler
grid-w2	PIC-IFAE	Intel P4	4	8 GFLOPS	Linux 2.2	PBS
ramses	DSIC-UPV	Intel PIII	20	17 GFLOPS	Linux 2.4	PBS
sherlock	DIF-UM	Intel PIII	1	550 MFLOPS	Linux 2.4	FORK
grid00x	IFCA	Intel PIII	10	12.6 GFLOPS	Linux 2.4	FORK
babieca	CAB	Compaq DS10	28	30 GFLOPS	Linux 2.2	PBS
ursa	DACYA-UCM	Sun Blade 100	1	1 GFLOPS	Solaris 8	FORK
draco	DACYA-UCM	Sun Ultra I	1	334 MFLOPS	Solaris 8	FORK
pc-ruben	DACYA-UCM	Intel P4	1	2.4 GFLOPS	Linux 2.4	FORK
solea	QUIM-UCM	Sun Entr. 250	2	1.2 GFLOPS	Solaris 8	FORK
		TOTAL	**228**	**290 GFLOPS**		

www.cern.ch/**egee**

Enabling Grids for E-Science and industry in Europe



- CERN
- Central Europe Consortium (Austria, Czech Republic, Hungary, Poland, Slovakia)
- France
- Germany and Switzerland
- Italy
- Northern Europe Consortium (Belgium, Denmark, Finland, Estonia, The netherlands, Norway, Sweden)
- South-East Europe Consortium (Bulgaria, Cyprus, Greece, Israel, Romania)
- South-West Europe Consortium (Spain, Portugal)
- UK and Ireland

CrossGrid

http://www.crossgrid.org

EU-DataGrid

http://www.eu-datagrid.org

FlowGrid

http://www.unizar.es/flowgrid

Damien

http://www.hlrs.de/organization/pds/projects/damien/

iAstro: Cost Action

http://main.cs.qub.ac.uk/~fmurtagh/iastro/

CESGA-CESCA Grid

http://www.cesga.es/Novas/defaultL.html?2003/2003_07_28.html&2/

www.rediris.es/**irisgrid**/

La iniciativa IRISGRID, evolución de la Red Temática sobre Grid lanzada en el 2002, surge con el objetivo de ***coordinar a nivel académico y científico a los grupos de investigación interesados en esta tecnología****, tanto en su desarrollo, implantación y aplicaciones.*

Además de esta coordinación, IRISGRID tiene como objetivo ***crear la infraestrutura GRID nacional*** *que permita el uso de esta tecnología tanto a nivel de aplicabilidad en diferentes ámbitos, como a nivel de desarrollo e innovación en este campo.*

Grupos:

- RedIRIS (CSIC-MCYT)
- Instituto de Física de Cantabria (IFCA, CSIC)
- ATC. Grupo de Arquitectura y Tecnología de Computadores. Universidad de Cantabria (unican)
- Instituto de Física Corpuscular de Valencia (IFIC, CSIC).
- Universidad de Valencia (UV)
- Centro de Astrobiología (CAB, CSIC-INTA)
- Grupo de Arquitectura de Sistemas Distribuidos y Seguridad (UCM). Universidad Complutense de Madrid (UCM)
- Grupo de Arquitectura y Tecnología de Computadores. Universidad Complutense de Madrid (UCM)
- Centro de Supercomputación de Galicia (CESGA)
- Port de Informació Cientifica de Barcelona (PIC)
- Institut de Fisica d'Atles Energies (IFAE)
- Unidad de Arquitectura de Ordenadores y Sistemas Operativos (AOSO). Universidad Autónoma de Barcelona (UAB)
- Grupo de Redes y Computación de Altas Prestaciones (GRyCAP), Universidad Politécnica de Valencia (UPV)
- Unidad de Biocomputación.Centro Nacional de Biotecnología (CNB, CSIC)
- Grupo CIRI (CEPBA-IBM Research Institute, Universidad Politécnica de Cataluña, Barcelona)
- Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT)
- Instituto Nacional de Técnica Aeroespacial (INTA)
- Arquitectura de Computadores, Comunicaciones y Sistemas (ARCOS), Universidad Carlos III de Madrid (UCIIIM)
- Laboratorio de Mecánica de Fluidos Computacional, Universidad Politécnica de Madrid (UPM)
- Grupo de Circuitos y Sistemas para Procesamiento de la Información, Universidad de Granada (UGR)
- Instituto Mediterráneo de Estudios Avanzados (IMEDEA, CSIC) & Universidad Islas Baleares (UIB)
- Instituto de Astrofísica de Andalucía (IAA, CSIC)
- Centro de Investigación del Cáncer (CIC, CSIC) & Universidad de Salamanca
- Grupo Arquitectura de Computadores (GAC).Universidad de A Coruña (UDC)
- Grupo Arquitectura de Computadores (GAC), Universidad de Santiago (USC)
- Grupo de Física Experimental de Altas Energías, Universidad Autonoma de Madrid (UAM)
- Grupo de Supercomputación: Algoritmos, Universidad de Almeria
- Instituto de Astrofisica de Canarias (IAC)
- Universidad de Murcia (UM)
- Universidad de Oviedo (UNIOVI)
- Grupo experimental de Fisica de Altas Energias. Universidad de Santiago de Compostela

Última Reunión en Junio de 2003

- **Demo Grid**
  - Equipos de 9 centros
  - 228 CPUs (290 GFlops), código CFD
  - Globus 2, EDG-CA española, GridWay RB

- **Definición de intereses de investigación en diferentes áreas**
  - Middleware
  - Astrophysics
  - Health Area
  - Bio-computing
  - High Energy Physics
  - Computational Chemistry
  - Complex Systems
  - Environmental Research

Próxima Reunión en Septiembre/Octubre

www.rediris.es/**irisgrid**/testbed

IRISGRID pretende

- **Aportar los estándares, protocolos, procedimientos y guías de "buenas prácticas"** necesarios para construir dentro de España un Grid de investigación coordinando a los diferentes Grupos (Organizaciones Virtuales) interesados en investigación sobre Grids.
- Esta iniciativa pretende **unir recursos distribuidos geográficamente** para que los Grupos involucrados tengan un banco de pruebas donde realizar la investigación en cualquiera de las Areas Grid.

IRISGRID no pretende

- Dar servicio técnico.

Requisitos

- **Requisitos para Añadir Recursos**
  - Mínimos para garantizar la disponibilidad del servicio
  - Proveer recursos a IRISGRID conllevará la responsabilidad de proporcionar el nombre de una persona responsable de la administración local de los sistemas que esté localizable durante un horario bien definido.

- **Requisitos para Usar Recursos**
  - El único requisito que se exige es la pertenencia a alguna organización pública o privada.

Middleware

- Globus Toolkit 2.4 (compatibilidad con 2.0 a nivel cliente)
  - GRAM 1.6
  - MDS 2.4
  - GridFtp 1.5
  - GSI

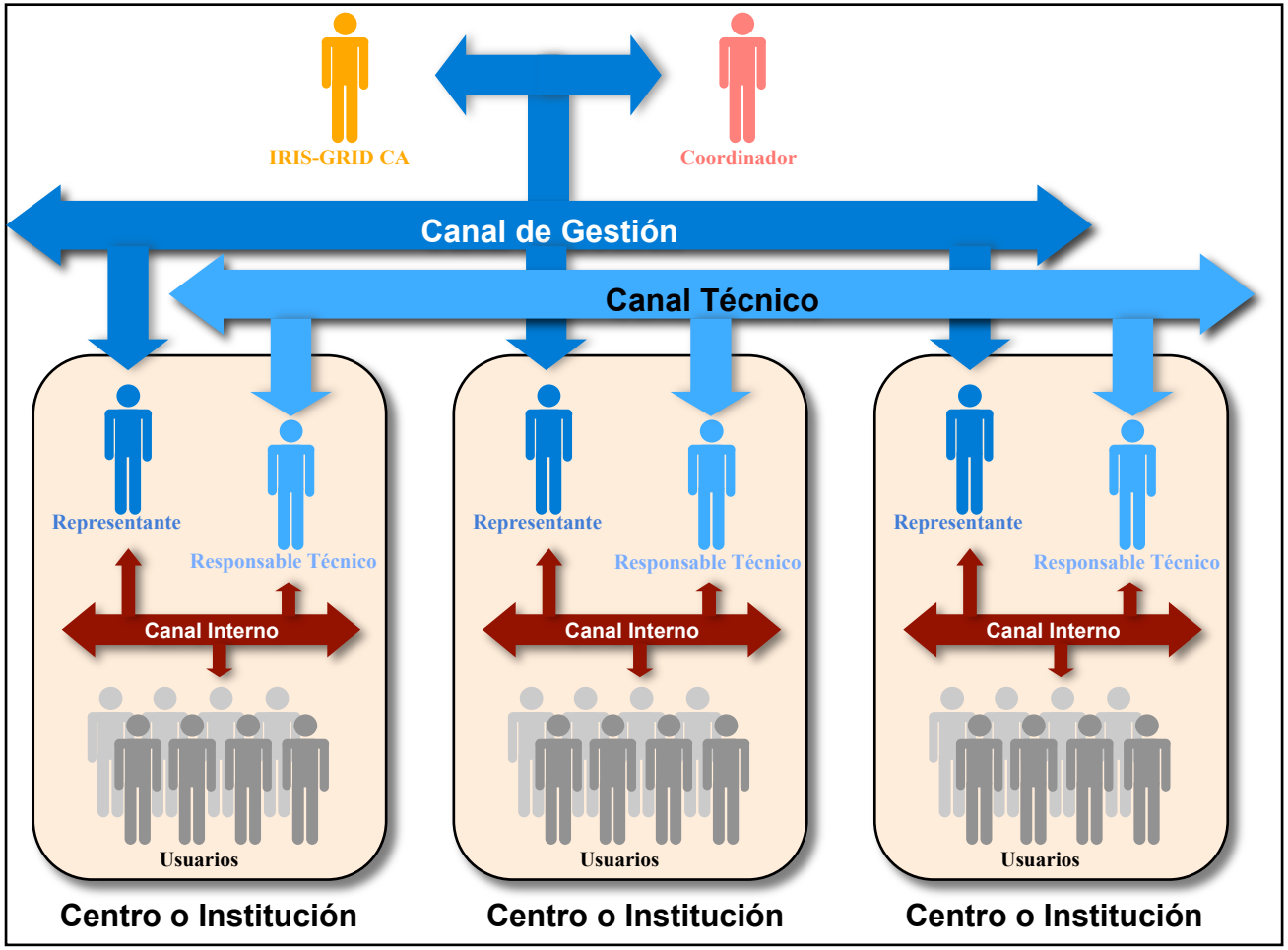
Procedimientos

- Solicitar el Alta de Instituciones o Centros
- Solicitar el Alta de Nuevos Recursos
- Solicitar el Alta de Nuevos Usuarios
- Solicitar la Autorización para el uso de Recursos
- Solicitar la Baja un Usuario

Autenticación

- **Certificación:** X509 de Globus
- **Autoridad de certificación:** IRISGRID CA
- **Canales de Comunicación:**
  - Nivel 1: Autenticación emisor e integridad
  - Nivel 2: Confidencialidad
  - Implementación: Por medio de importación de certificados a las herramientas de correo más comunes

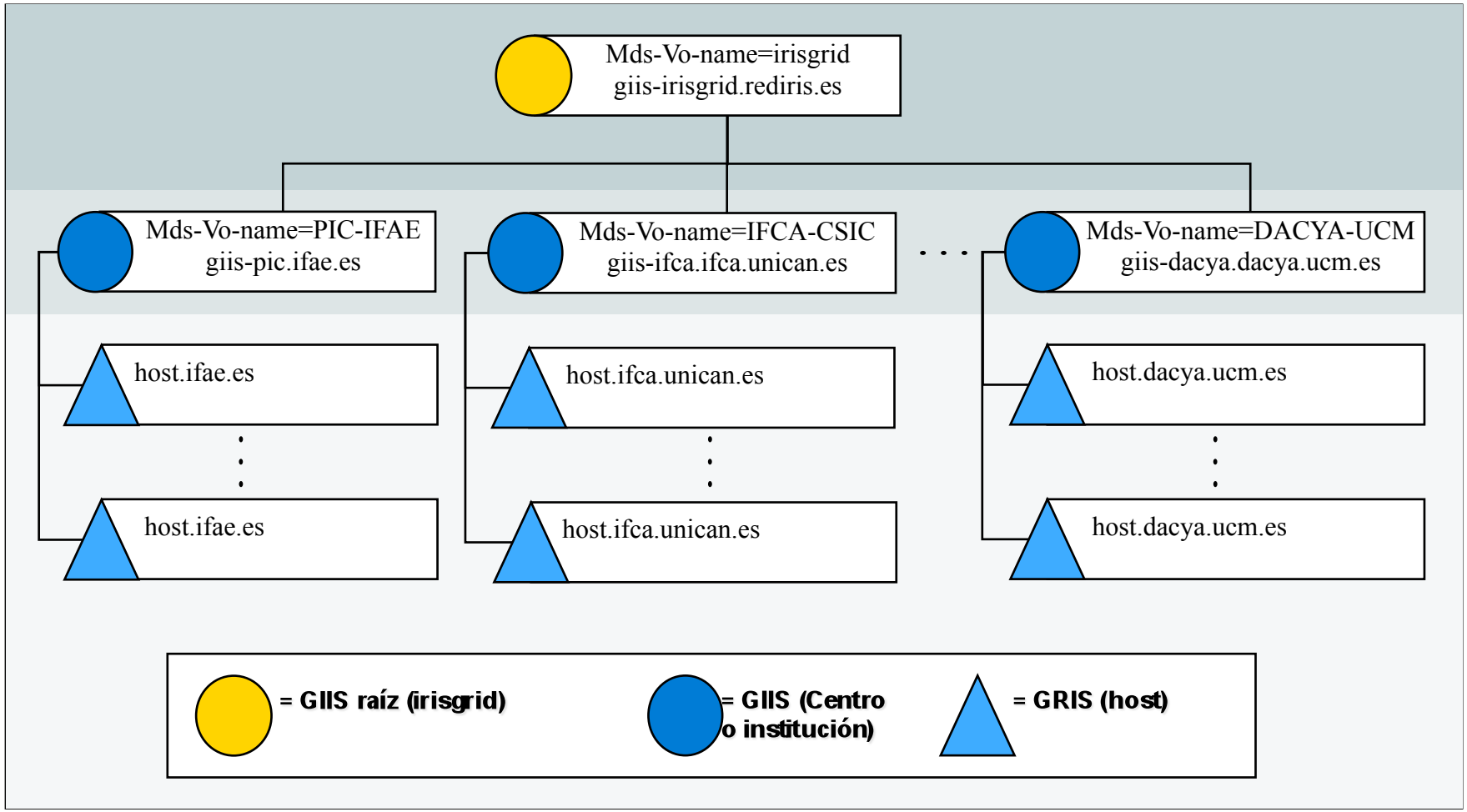
Canales de Comunicación



Autorización

- La decisión final de que usuarios podrán usar los recursos será, por supuesto, de los propietarios de los mismos siguiendo sus normas locales.

Sistemas de Información



"Technology does not drive change at all. Technology merely enables change. It's our collective cultural response to the options and opportunities presented by technology that drives change"

Paul Saffo