

## Grid Computing

Un nuevo paradigma de  
computación distribuida



*Borja Sotomayor*

*9 de abril de 2003*

Grid Computing



## Grid Computing



Duración aproximada: 1h

Grid Computing



## Índice

- ▶ ¿Qué es la Computación Grid?
- ▶ Aplicaciones Prácticas
- ▶ Arquitectura de la Grid
- ▶ Globus Toolkit 3
- ▶ Del "Web Service" al "Grid Service"
- ▶ Futuro de la Computación Grid

Grid Computing



## Índice

- ▶ ¿Qué es la Computación Grid?
- ▶ Aplicaciones Prácticas
- ▶ Arquitectura de la Grid
- ▶ Globus Toolkit 3
- ▶ Del "Web Service" al "Grid Service"
- ▶ Futuro de la Computación Grid

Grid Computing



## ¿Qué es la Computación Grid? (1)



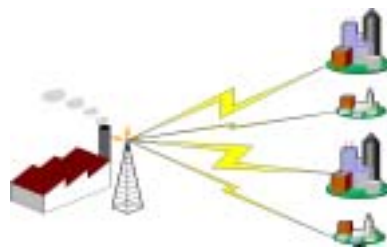
- ▶ A comienzos del siglo XX, si querías tener electricidad, hacía falta estar situado cerca de un generador electrico.

Grid Computing

• • • • • • • •



## ¿Qué es la Computación Grid? (2)



- ▶ Ahora, no tenemos generadores individuales, sino unos *pocos* 'supergeneradores' que abastecen a *muchos* clientes.
- ▶ Red Electrica (Electric Grid)

Grid Computing

• • • • • • • •



## ¿Qué es la Computación Grid? (3)

- ▶ Algo similar pasa con la informática ahora. Si en mi organización quiero disponer de cierto recurso computacional, no tengo más remedio que comprarlo.
  - ▶ Potencia computacional
  - ▶ Almacenamiento
  - ▶ Equipos para aplicaciones específicas



## ¿Qué es la Computación Grid? (4)

- ▶ Actualmente, puedo subcontratar el acceso a recursos computacionales ajenos, pero cada compañía lo realiza de manera *ad hoc*.
- ▶ Además, el acceso a esos recursos dista mucho de ser óptimo.



## ¿Qué es la Computación Grid? (5)

### ► Ejemplo

- Supongamos que cuatro veces al año el centro de calculo de nuestra empresa genera varios *terabytes* ( $10^6$  MB) de información que deben ser analizados.
- Puesto que no nos merece la pena tener un supercomputador, subcontratamos el proceso de los datos a otra empresa.

Grid Computing



## ¿Qué es la Computación Grid? (6)



- Se generan los datos, que son volcados a cintas magneticas.

Grid Computing



## ¿Qué es la Computación Grid? (7)

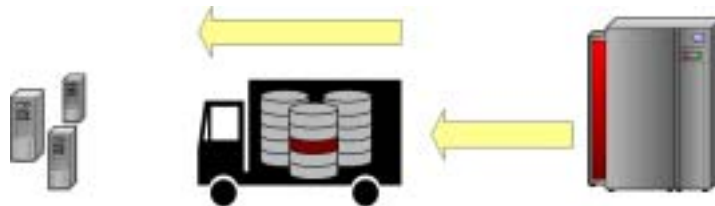


- ▶ Las cintas son enviadas a la empresa que realiza el proceso de datos.

Grid Computing



## ¿Qué es la Computación Grid? (8)



- ▶ La empresa tarda 5 días en realizar el proceso de datos. Los resultados son volcados a cintas magneticas.

Grid Computing



## ¿Qué es la Computación Grid? (9)



- ▶ Recibimos el resultado. Descubrimos que hay un fallo en los datos. ¡Vuelta a empezar!

Grid Computing

• • • • •



## ¿Qué es la Computación Grid? (10)

- ▶ ¿Sería posible crear algo similar a la red eléctrica pero en el ámbito computacional?
  - ▶ Acceder a recursos computacionales *geográficamente dispersos* (como las centrales eléctricas) con la misma facilidad con la que enchufo un electrodoméstico a la red eléctrica.
- ▶ Solución propuesta: Grid Computing

Grid Computing

• • • • •



## ¿Qué es la Computación Grid? (11)

### ► Grid Computing

- Un nuevo paradigma de computación distribuida propuesto por Ian Foster y Carl Kesselman a mediados de los 90.
- Está enfocado al acceso remoto a recursos computacionales.
- Es un paradigma. No describe ni favorece a ninguna tecnología concreta. Sienta las bases teóricas (arquitectura, requisitos previos, ...)
  - Tecnología estandar *de facto*: Globus Toolkit

Grid Computing



## ¿Qué es la Computación Grid? (12)

- "The Grid": *pool* de recursos computacionales *geográficamente dispersos* al que tenemos acceso. "Red eléctrica de recursos computacionales"
- Sin embargo, todavía no existe "*the Grid*"
  - Hay muchos 'mini-Grids' a lo largo de todo el mundo (testbeds) que se están utilizando para poner a prueba tecnologías Grid.
  - Futuro: Un Grid a nivel mundial al que todos podremos tener acceso.

Grid Computing





## ¿Qué es la Computación Grid? (13)

- ▶ ¿Por qué un nuevo paradigma y tecnologías?
- ▶ Los paradigmas (Cliente/Servidor 'tradicional') y tecnologías (CORBA, EJB, .Net, ...) actuales son insuficientes para la Computación Grid.
  - ▶ Orientados a la *información*, no a los *recursos computacionales*.
  - ▶ Sistemas fuertemente acoplados.
  - ▶ Seguridad en segundo plano.
  - ▶ Orientados a latencias bajas y a comunicación síncrona.
  - ▶ No suelen contemplar QoS.

Grid Computing



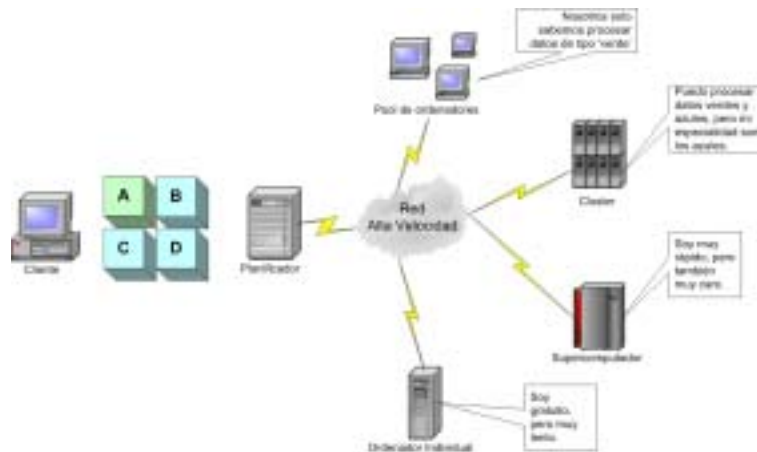
## ¿Qué es la Computación Grid? (14)

- ▶ Ejemplo
  - ▶ ¿Cómo se beneficiaría la empresa anterior de un enfoque Grid?
  - ▶ En vez de enviar los datos a una empresa subcontratada, 'lanzaría' el trabajo a la Grid.

Grid Computing



## ¿Qué es la Computación Grid? (15)



Grid Computing



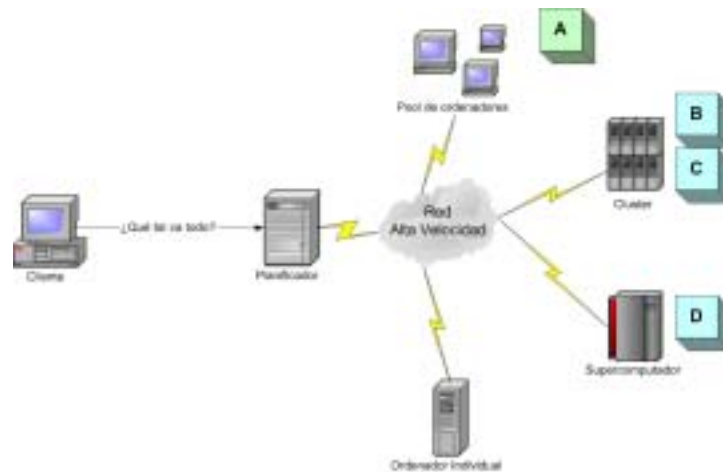
## ¿Qué es la Computación Grid? (16)

- Los datos (en general: el problema a resolver) se dividen en particiones manejables. Cada parte es enviada a un nodo de la Grid a través de una red de alta velocidad. El destino lo decide un 'planificador' en función de las características de cada nodo y de cada partición.

Grid Computing



## ¿Qué es la Computación Grid? (17)



Grid Computing



## ¿Qué es la Computación Grid? (18)

- Cada nodo procesa la parte que le toca. Periodicamente podemos contactar con cada nodo, para asegurarnos de que todo va bien, e incluso realizar correcciones sobre la marcha.

Grid Computing



## ¿Qué es la Computación Grid? (19)



- ▶ Cada nodo devuelve el resultado de la computación realizada.

Grid Computing



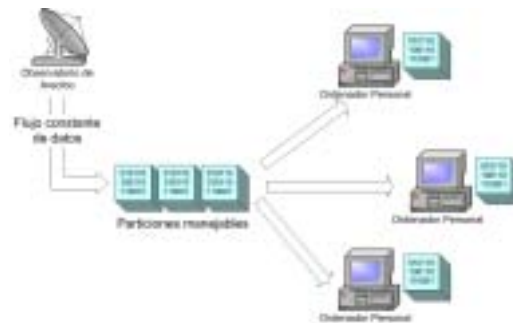
## ¿Qué es la Computación Grid? (20)

- ▶ Esto suena muy bien pero...¿es real?
  - ▶ Sí. La Computación Grid es una de las áreas más activamente investigadas actualmente.
  - ▶ Ya hay toolkits de desarrollo que nos permiten realizar programación Grid.
  - ▶ Existen muchos proyectos a nivel mundial y europeo: Crossgrid, Eurogrid, Datagrid, ...
  - ▶ Algunas grandes empresas ya apuestan por las tecnologías Grid (p.ej. IBM, Oracle, Intel)

Grid Computing



## ¿Qué es la Computación Grid? (21)



### ► Ejemplo Real #1

- SETI@Home: Los datos recogidos en el observatorio de Arecibo (Puerto Rico) son procesados por ordenadores de todo el mundo.

Grid Computing



## ¿Qué es la Computación Grid? (22)

### ► Ejemplo Real #2

- Analisis de Mamografías en Grid
- Se reduce el tiempo de 1h a 40 segundos.
- Diagnóstico más completo
- 100 nodos de computacion en Grid
- Acceso a una BD distribuida con imagenes de mamografías
- IBM, Miranda Solutions, Gobierno UK

Grid Computing



## ¿Qué es la Computación Grid? (23)

### ► Ejemplo Real #3

- Sistema de Apoyo a Decisiones para inundaciones.
- Procesamiento en tiempo real de datos obtenidos en el lugar de la catastrofe, o en lugares susceptibles a grandes inundaciones.
- Monitorización, Predicción, Simulación
- Proyecto Crossgrid (Europa)



## Índice

- ¿Qué es la Computación Grid?
- **Aplicaciones Prácticas**
- Arquitectura de la Grid
- Globus Toolkit 3
- Del "Web Service" al "Grid Service"
- Futuro de la Computación Grid



## Aplicaciones Prácticas (1)

- ▶ Actualmente, hay 5 aplicaciones generales para la Computación Grid:
  - ▶ Supercomputación distribuida
  - ▶ Sistemas distribuidos en tiempo real
  - ▶ Servicios puntuales
  - ▶ Proceso intensivo de datos
  - ▶ Entornos virtuales de colaboración (Teleinmersión)

Grid Computing



## Aplicaciones Prácticas (2)

- ▶ Supercomputación distribuida
  - ▶ Aplicaciones cuyas necesidades es *imposible* satisfacer en un único nodo.
  - ▶ Estas aplicaciones se distinguen porque satisfacen necesidades *puntuales* e *intensivas* de computación.
  - ▶ Ejemplos: Simulaciones, cálculos numéricos, Data Mining, análisis de grandes volúmenes de datos.

Grid Computing



## Aplicaciones Prácticas (3)

- ▶ **Sistemas distribuidos en tiempo real**
  - ▶ Sistemas que generan un flujo de datos a alta velocidad que debe ser analizado en tiempo real.
  - ▶ Ejemplos: e-Medicine, experimentos de física de alta energía, control remoto de un recurso no-trivial (un microscopio, equipo médico, ...)



## Aplicaciones Prácticas (4)

- ▶ **Servicios puntuales**
  - ▶ Acceso puntual a un recurso que no nos merece la pena tener en nuestra organización.
  - ▶ Similar a las dos aplicaciones anteriores.
    - ▶ Diferencias: no nos referimos a 'potencia computacional', y no tiene que ser en tiempo real.
  - ▶ Ejemplos: acceso a hardware específico para ciertos tipos de análisis (químico, biológico, ...)





## Aplicaciones Prácticas (5)

- ▶ Proceso intensivo de datos
  - ▶ Aplicaciones que trabajan con grandes volúmenes de datos, y que es imposible *almacenar* en un único nodo.
  - ▶ Los datos se distribuyen a lo largo de la Grid.



## Aplicaciones Prácticas (6)

- ▶ Teleinmersión
  - ▶ Utilizar la potencia computacional y la naturaleza distribuida de la Grid para crear entornos virtuales 3D distribuidos.



## Aplicaciones Prácticas (7)

- ▶ Actualmente, las aplicaciones del Grid Computing están muy orientadas al mundo científico.
- ▶ Las empresas empiezan a interesarse por estas tecnologías.
- ▶ Futuro (10 años): Un Grid Público a nivel mundial, en el que cualquier persona puede disponer, si lo necesita, del poder computacional de un supercomputador. Además, todos podemos ser también proveedores de ciclos de CPU.

Grid Computing



## Índice

- ▶ ¿Qué es la Computación Grid?
- ▶ Aplicaciones Prácticas
- ▶ **Arquitectura de la Grid**
- ▶ Globus Toolkit 3
- ▶ Del "Web Service" al "Grid Service"
- ▶ Futuro de la Computación Grid

Grid Computing



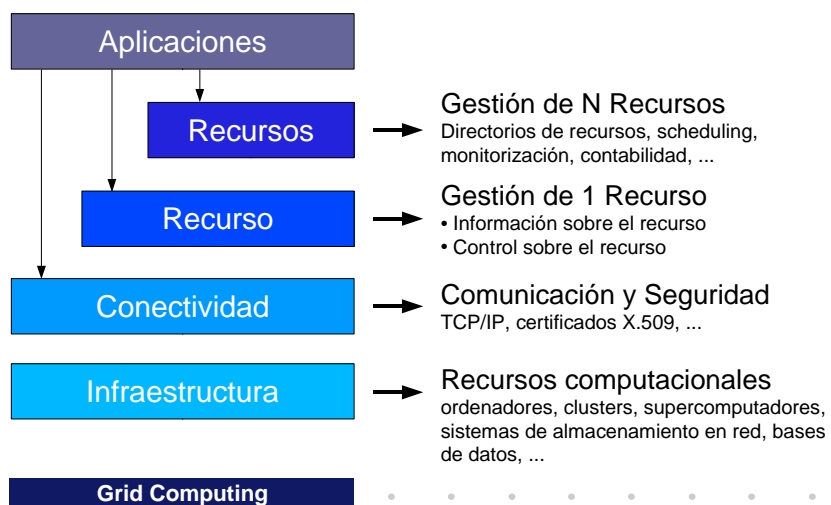
## Arquitectura de la Grid (1)

- ▶ En el artículo 'clasico' *Anatomy of the Grid: Enabling Scalable Virtual Organizations* (2000), Ian Foster, Carl Kesselman, y Steven Tuecke proponen una arquitectura general que deberían cumplir aplicaciones, toolkits, APIs, SDKs, etc. relacionadas con la Computación Grid.

Grid Computing

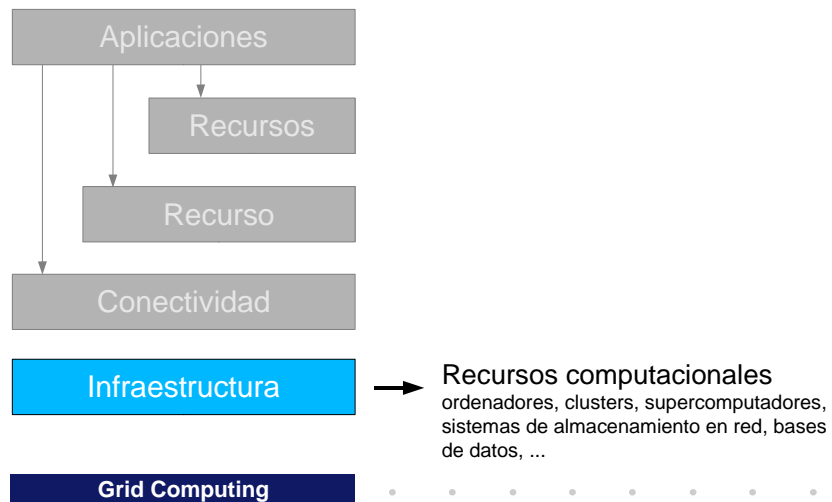


## Arquitectura de la Grid (2)





## Arquitectura de la Grid (3)

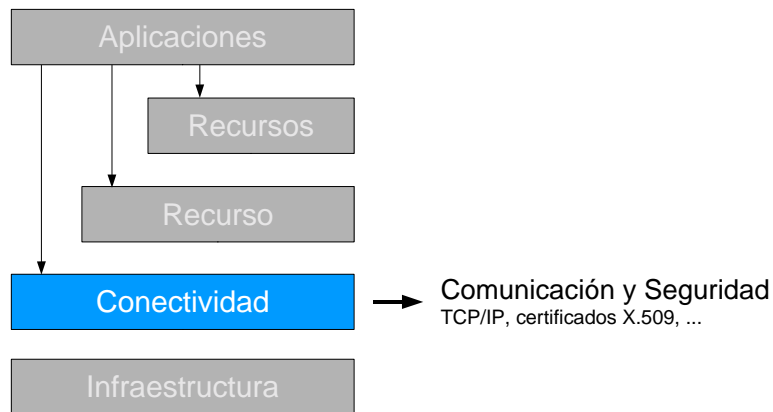


## Arquitectura de la Grid (4)

- ▶ La infraestructura de la Grid está compuesta por los recursos computacionales que queremos compartir.
  - ▶ Ordenadores individuales
  - ▶ Pool de ordenadores
  - ▶ Clusters
  - ▶ Supercomputadores
  - ▶ Sistemas de almacenamiento
- ▶ También incluye la infraestructura de red y sus mecanismos de gestión y control.



## Arquitectura de la Grid (5)



Grid Computing



## Arquitectura de la Grid (6)

- ▶ La capa de conectividad incluye los protocolos de comunicación y de seguridad que permiten que los recursos computacionales se comuniquen.
  - ▶ Pila de protocolos TCP/IP
  - ▶ Se están estudiando otros protocolos que proporcionen mejor rendimiento en redes de alta velocidad.
  - ▶ SSL, Certificados X.509

Grid Computing

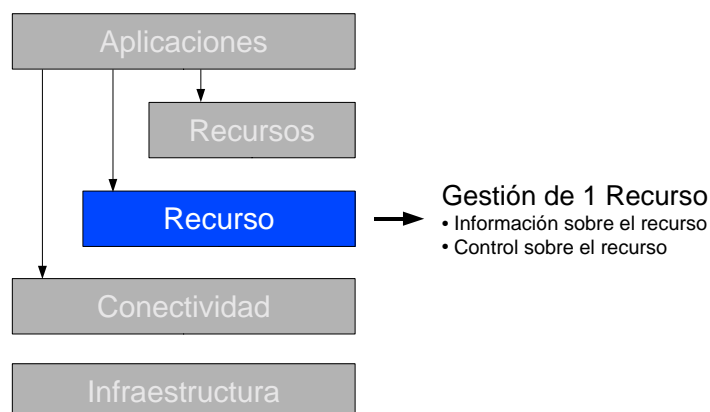


## Arquitectura de la Grid (7)

- ▶ En la Computación Grid es especialmente importante la seguridad.
  - ▶ Compartimos recursos entre distintas organizaciones (con distintas políticas de seguridad, etc.)
  - ▶ Importante --> Delegación de autenticación y autorización.



## Arquitectura de la Grid (8)





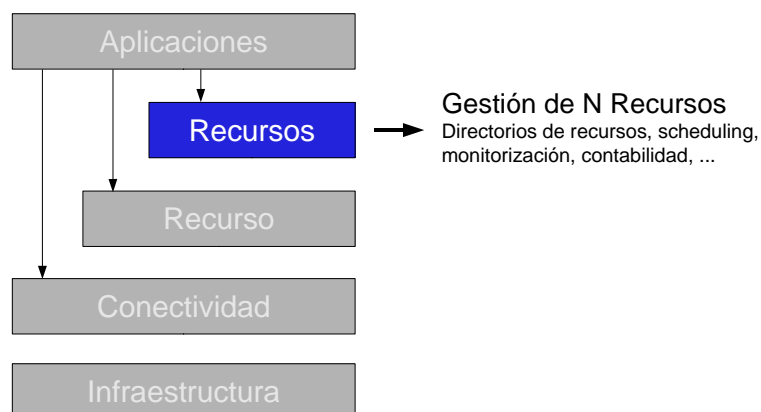
## Arquitectura de la Grid (9)

- ▶ Esta capa incluye protocolos para el control y gestión de recursos individuales.
  - ▶ **Información:** Debe ser posible obtener información sobre un recurso (características técnicas, carga actual, precio, ...)
  - ▶ **Control:** Acceso, arranque, gestión, parada, monitorización, contabilidad, auditoría del recurso.

Grid Computing



## Arquitectura de la Grid (10)



Grid Computing



## Arquitectura de la Grid (11)

- ▶ Esta capa engloba todos los servicios que nos permiten gestionar varios recursos.
  - ▶ Servicios de directorio: Localización de los recursos que nos interesan.
  - ▶ Schedulers distribuidos: Asignación de tareas a cada recurso.
  - ▶ Monitorización y diagnostico de la ejecución de las distintas tareas de una aplicación.
  - ▶ Contabilidad: Calculo del coste de la utilización de varios recursos heterogeneos.
  - ▶ Acceso a datos distribuidos. Gestión de replicación.



## Arquitectura de la Grid (12)

- ▶ Los schedulers distribuidos suelen ser una de las partes más complicadas de una aplicación grid, porque se 'pelean' tres schedulers distintos:
  - ▶ Planificador de jobs: Su objetivo es maximizar el *throughput* de jobs (jobs/unidad de tiempo).
  - ▶ Planificador de recursos: Su objetivo es maximizar el uso de los recursos.
  - ▶ Planificador de la aplicación: Divide la aplicación en tareas, las asigna a recursos concretos, vigila la ejecución, ...
- ▶ 1 y 2 priman la eficiencia del sistema. 3 prima la eficiencia de la aplicación.



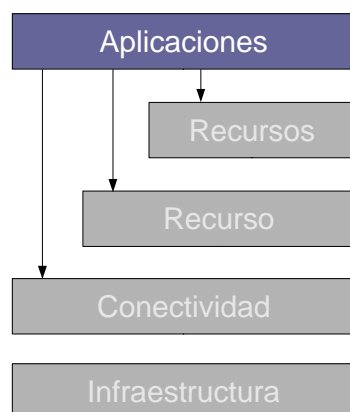


## Arquitectura de la Grid (13)

- ▶ *Virtual Organization (VO)*
  - ▶ Concepto fundamental en Computación Grid.
  - ▶ Agrupación de recursos de varios individuos y/o organizaciones distintas que colaboran para alcanzar una meta común.
  - ▶ La pertenencia a una VO no es permanente. Puede cambiar según las necesidades.
  - ▶ Agrupación en VOs --> Facilita gestión de recursos y seguridad.



## Arquitectura de la Grid (14)



### Aplicaciones Grid

Acceden a la infraestructura de la Grid a través de las distintas capas. Según las exigencias de la aplicación, puede ser necesario pasar por todas las capas, o conectarse directamente a la infraestructura.



## Índice

- ▶ ¿Qué es la Computación Grid?
- ▶ Aplicaciones Prácticas
- ▶ Arquitectura de la Grid
- ▶ **Globus Toolkit 3**
- ▶ Del "Web Service" al "Grid Service"
- ▶ Futuro de la Computación Grid

Grid Computing



## Globus Toolkit 3 (1)

- ▶ El Globus Toolkit es el estandar *de facto* para la implementación de aplicaciones Grid.
- ▶ La versión 2 goza de una gran aceptación, pero tiene algunas deficiencias.
- ▶ La versión 3 está en fase alfa, y está basada (casi) completamente en tecnologías estándar como XML, SOAP, WSDL, Servicios Web.
  - ▶ Implementado integralmente en Java.

Grid Computing



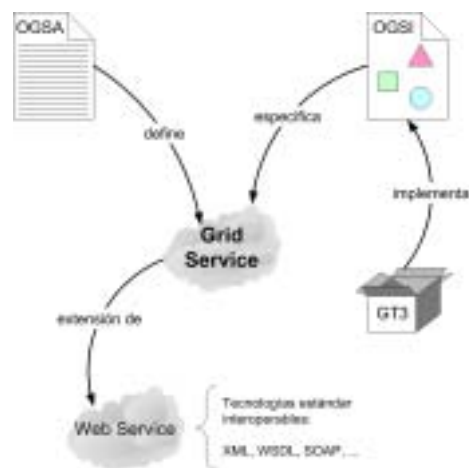
## Globus Toolkit 3 (2)

- ▶ El Globus Toolkit 3 (GT3) introduce el concepto de *Grid Service*, que es una ampliación considerable a los Web Services.
- ▶ La arquitectura Grid Service está especificada por el Global Grid Forum
  - ▶ OGSA: Open Grid Services Architecture
  - ▶ OGSi: Open Grid Services Infrastructure
- ▶ Los Grid Services son la base de la arquitectura GT3.

Grid Computing



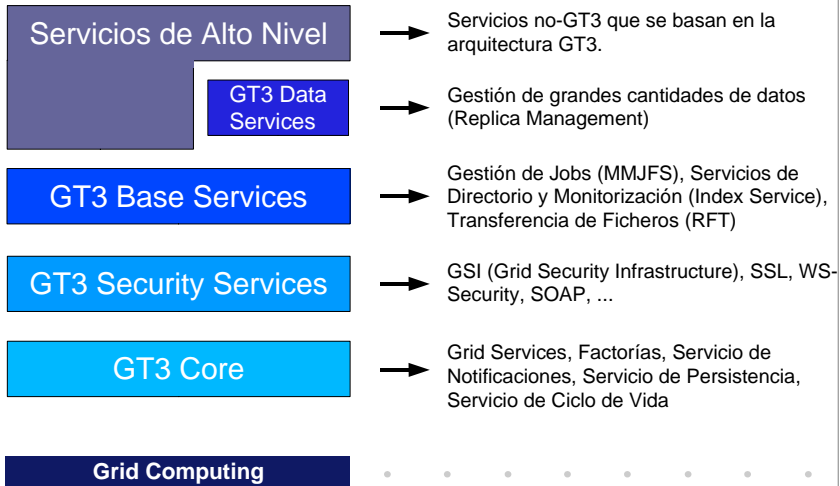
## Globus Toolkit 3 (3)



Grid Computing



## Arquitectura Globus Toolkit 3



## Globus Toolkit 3 (4)



Globus Project: <http://www.globus.org/>

Global Grid Forum: <http://www.ggf.org/>



## Índice

- ▶ ¿Qué es la Computación Grid?
- ▶ Aplicaciones Prácticas
- ▶ Arquitectura de la Grid
- ▶ Globus Toolkit 3
- ▶ Del "Web Service" al "Grid Service"
- ▶ Futuro de la Computación Grid



## Grid Services (1)

- ▶ Al plantear la siguiente generación de tecnologías grid (OGSA y OGSF), el Global Grid Forum buscó una tecnología de objetos distribuidos que se adaptase a las necesidades de una aplicación grid.
- ▶ Los Web Services son la mejor opción. Sin embargo, a pesar de sus ventajas, también tienen importantes limitaciones.



## Grid Services (2)

- ▶ Principales desventajas de los Web Services:
  - ▶ Stateless
  - ▶ No-transientes ('Persistentes')
  - ▶ No tienen 'servicios de apoyo' (notificaciones, servicio de persistencia, gestión del ciclo de vida, etc.)



## Grid Services (3)

- ▶ Los *Grid Services* son Web Services mejorados.
- ▶ Están basados en SOAP y WSDL y, por lo tanto, 100% compatibles con Web Services 'tradicionales'.



## Grid Services (4)

- ▶ Características de los Grid Services
  - ▶ *Stateful*: El estado del Grid Service se mantiene de una invocación a otro.
  - ▶ Puede ser transiente: Podemos crear varias *instancias* de un mismo Grid Service '*on-the-fly*' y destruirlas cuando ya no son necesarias.
  - ▶ Servicios de apoyo

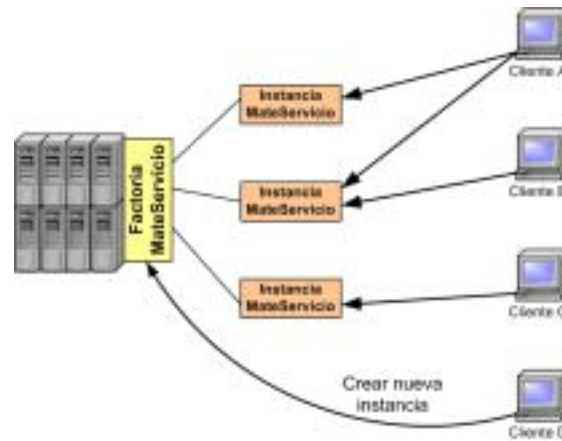


## Grid Services (5)

- ▶ Novedad: Los Grid Services utilizan un enfoque de 'Factorías de objetos'.
- ▶ En lugar de tener un único servicio sin estado, compartido por todos los usuarios (Web Service), tenemos un servicio-factoria para crear instancias individuales del servicio (Grid Service).



## Grid Services (6)



## Grid Services (7)

- ▶ Las instancias se crean a través de la factoría.
- ▶ Cuando queremos invocar una operación del servicio, accedemos a la instancia no a la factoría.
- ▶ Podemos tener una instancia por cliente, varias instancias por cliente, varios clientes por instancia, ...
- ▶ La destrucción de la instancia puede correr a cargo del cliente o de la factoría.





## Grid Services (8)

- ▶ Servicios de apoyo
  - ▶ Datos del Servicio (Service Data): Podemos asociar información a cada instancia para facilitar la clasificación e indexado de los servicios.
  - ▶ Notificaciones. Modelo Push.
  - ▶ Servicio de Persistencia.
  - ▶ Gestión del Ciclo de vida.



## Grid Services (9)

- ▶ El Globus Toolkit 3 ya incluye una implementación completa de OGSA/OGSI.
- ▶ El Global Grid Forum está en contacto con la W3C para que las mejoras propuestas lleguen a formar parte del estándar oficial de Web Services.
  - ▶ Se espera que, en el futuro, ambas líneas de desarrollo se fusionen en una única.



## Índice

- ▶ ¿Qué es la Computación Grid?
- ▶ Aplicaciones Prácticas
- ▶ Arquitectura de la Grid
- ▶ Globus Toolkit 3
- ▶ Del "Web Service" al "Grid Service"
- ▶ Futuro de la Computación Grid



## Futuro de la Computación Grid (1)

- ▶ La Computación Grid es bastante reciente. Sin embargo, parece que no es simplemente una 'moda', ya que está siendo activamente investigada y desarrollada en todo el mundo.
- ▶ Plantea interesantes desafíos en una gran cantidad de áreas de investigación distintas.



## Futuro de la Computación Grid (2)

- ▶ **Sistemas Distribuidos**
  - ▶ Modelos de componentes para entornos grid.
  - ▶ Grid Services.
- ▶ **Sistemas Operativos Distribuidos y Computación Paralela**
  - ▶ Paralelismo entre recursos heterogeneos con una alta latencia y un ancho de banda limitado y variable.
  - ▶ Scheduling distribuido.
  - ▶ Ejecución ubicua de código.

Grid Computing



## Futuro de la Computación Grid (3)

- ▶ **Redes de Computadores**
  - ▶ Redes de alta velocidad a escala mundial.
  - ▶ Nuevos protocolos con sobrecarga (overhead) mínima.
  - ▶ IPv6
- ▶ **Seguridad**
  - ▶ Nuevo panorama en el que no tenemos los clásicos roles de 'cliente' y 'servidor', sino un montón de nodos geográficamente dispersos colaborando para resolver una tarea común.

Grid Computing



## Futuro de la Computación Grid (4)

- ▶ Bases de datos
  - ▶ Aplicación de tecnologías Grid al Data Mining, Data Warehousing, y a Sistemas de Apoyo a la Decisión.



## Futuro de la Computación Grid (5)

- ▶ I+D liderado principalmente por EEUU
  - ▶ Global Grid Forum. Desarrollo de estándares para la Grid.
  - ▶ The Globus Project
  - ▶ TeraGrid. Redes WAN de alta velocidad (40 Gb/s), y almacenamiento distribuido (1 Petabyte).
  - ▶ Empresas: IBM, Oracle, Intel, HP-Compaq, Entropia, Platform, Avaki, ...



## Futuro de la Computación Grid (6)

### ► Proyectos Europeos

- CrossGrid. Supercomputación y análisis de grandes volúmenes de datos.
- EuroGrid. Aplicación de tecnologías Grid en el ámbito científico e industrial.
- DataGrid. Análisis de grandes volúmenes de datos.
- UK e-Science (Reino Unido). Investigación distribuida.



## Futuro de la Computación Grid (7)

### ► Infraestructura en Europa

- GÉANT: Backbone Gigabit en Europa (10 y 2.5 Gbps)
- REDIRIS2: Backbone Gigabit en España (2.5 Gbps)
- Además...
  - IRIS-GRID: Grupo de trabajo de RED-IRIS sobre tecnologías Grid.



## Futuro de la Computación Grid (8)

### ► ¿El Futuro?

- La Computación Grid es una de las tecnologías más prometedoras actualmente.
- Ahora mismo está presente principalmente en el ámbito científico.
  - Sin embargo, gracias a la Computación Grid ya se está realizando investigación de mayor calidad.
- Todavía falta *mucho* para que llegue al público general --> *Grid Público*.

Grid Computing

## Turno de Preguntas



**Borja Sotomayor**  
Facultad de Ingeniería - ESIDE  
Universidad de Deusto  
[borja@borjanet.com](mailto:borja@borjanet.com)

Grid Computing