



Análisis Comparativo GT 2.4 - GT 4.0

Juan Eduardo Murrieta León

eml@super.unam.mx





Temario

- Escenario
- Globus y el Globus Toolkit.
- Globus Toolkit 2.4
 - Los cuatro componentes: GSI, GRAM, GridFTP, MDS
- Evaluación del GT 2.4
- OGSA, Web Services y el GT 4.0
- Globus Toolkit 4.0
 - Pre-WS
 - GT-WebServices: GSI, WS-GRAM, GridFTP, MDS4
- Conclusiones





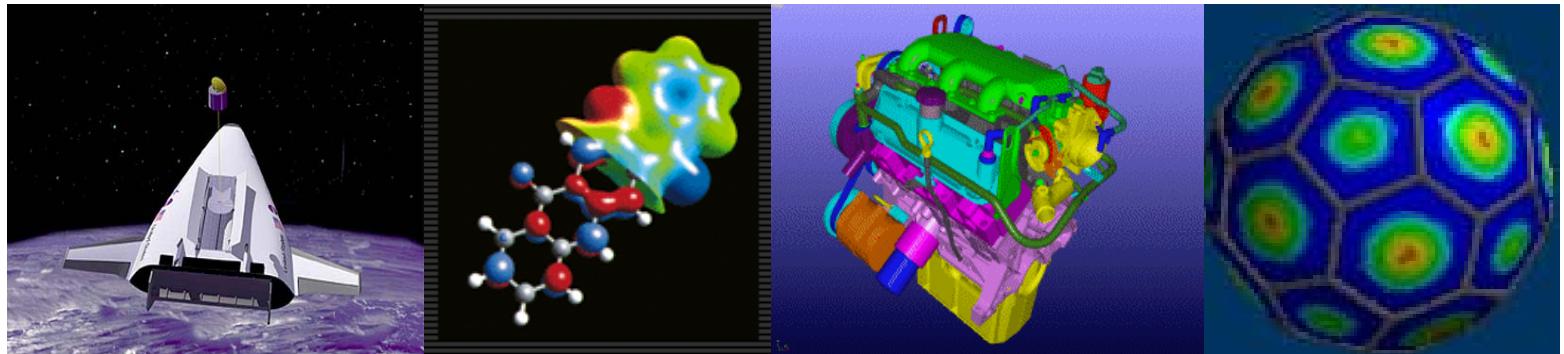
Escenario





Escenario

- El *cómputo científico* es una herramienta tan poderosa y necesaria como lo es el trabajo teórico y experimental en la investigación científica actual.





Escenario

- Los centros de investigación cuentan con una gran diversidad de recursos técnicos y computacionales accesibles vía red:
 - Computadoras
 - Bases de datos
 - Instrumentos
 - Personal
 - etc.



The Earth Simulator Center



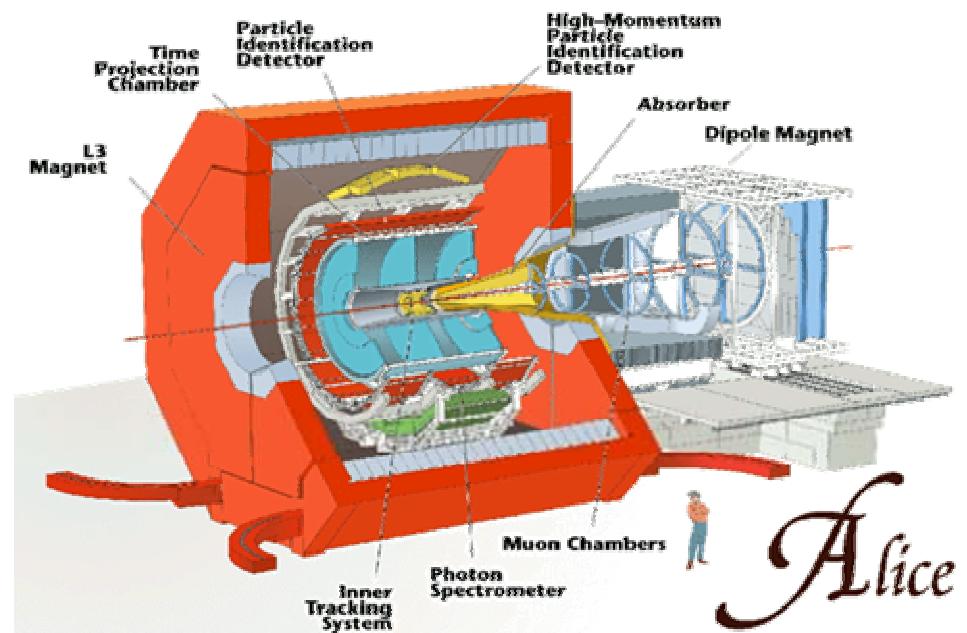


Escenario

- Las necesidades de cómputo son extremas, mientras que los recursos son limitados.

LHC y *Alice*

10 Petabytes de datos al año.





Escenario

- Una **GRID** es un sistema que:
 1. Cordina recursos que no son susceptibles de un control centralizado
 2. Utiliza protocolos e interfaces estandarizadas, abiertas y de propósito general
 3. Proporciona calidades de servicio no triviales

Ian Foster, *What is the Grid? A three point checklist.*
GRIDToday, Julio 20, 2002 .



Grids en el Mundo

L C G (LHC Computing Grid)



NPACI Grid



EU-DataGrid Project (EDG)

DataTAG (Trans Atlantic Grid)



INFN Grid



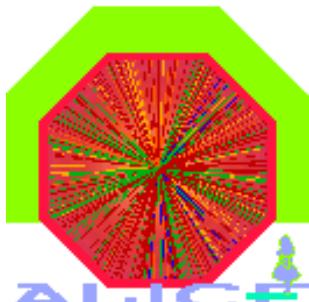
GriPhyN (Grid Physics Network)

PPDG (Particle Physics Data Grid)



International Virtual Data Grid Laboratory (iVDgL)

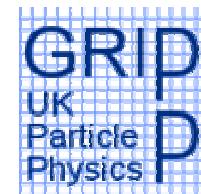
Grids en el Mundo



ApGrid (Asia Pacific Grid)



(Canada)



PRAGMA (Pacific Rim Applications and Grid Middleware Assembly)



GRAMA
Grid Académica Mexicana

www.grama.org.mx



Globus y el Globus Toolkit





Globus y el Globus Toolkit

- Se han explorado diversos mecanismos para crear la infraestructura de software de la GRID.
- **Globus** es un proyecto muti-institucional de *The Globus Alliance* para proveer la infraestructura necesaria para la construcción de grids.
 - ANL, Universidad de Chicago, EPCC, NCSA, USC, Instituto Real de Tecnología de Suecia.





Globus y el Globus Toolkit

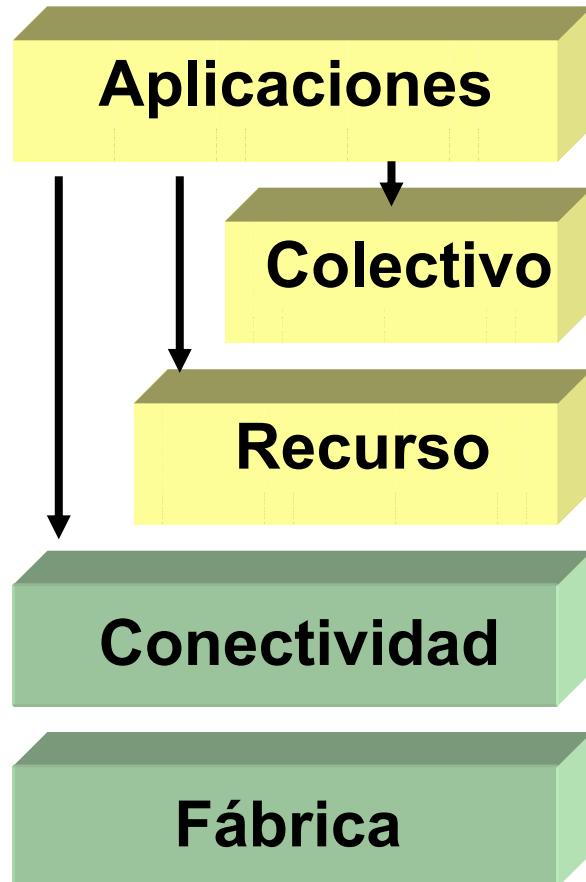
- El elemento central del proyecto Globus es el **Globus Toolkit (GT)**
 - Basado en un **modelo de paquete de servicios**.
- El **GT** es considerado como uno de los elementos clave en desarrollo de las grids computacionales:
 - *“The Globus Toolkit ... is one of the Ten Technologies That Will Change the World”*

MIT Technology Review





Arquitectura por capas del GT



**Administración colectiva
de recursos “lógicos”**

**Administración de recursos
individuales**

Comunicación y seguridad

Recursos “físicos”





Globus Toolkit 2.4



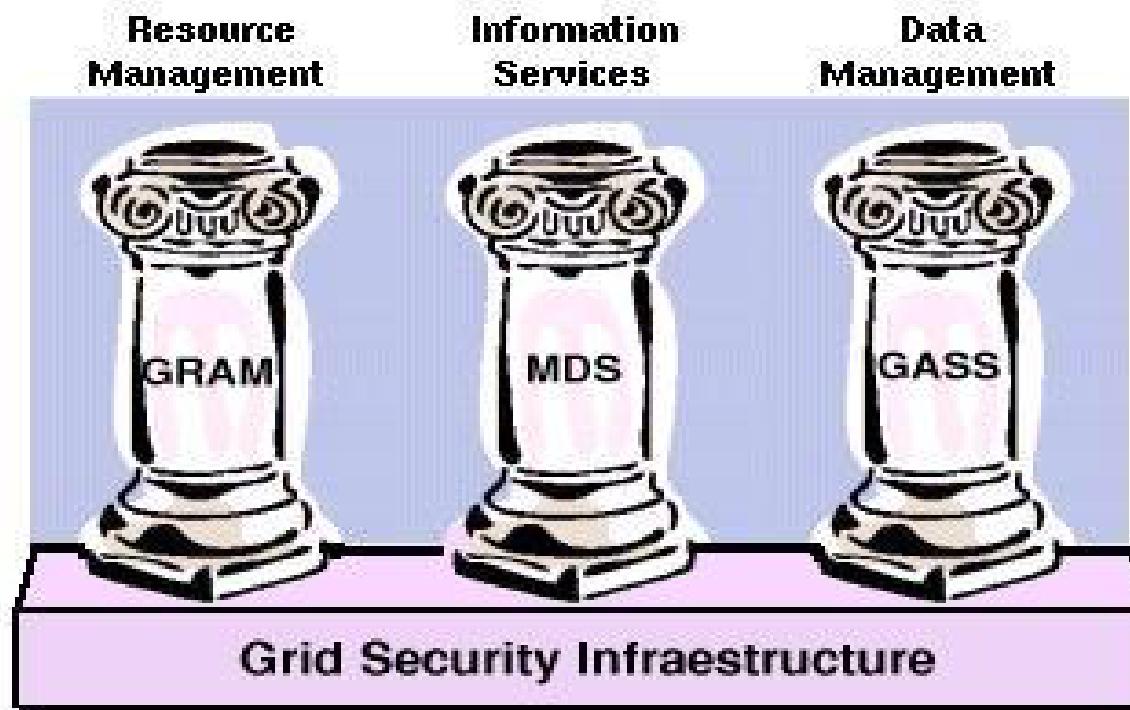


Globus Toolkit 2.4

- El Globus Toolkit se ha convertido en el estándar para la implementación de servicios de GRID.
 - La versión más utilizada y estable en GRIDs de producción en el mundo es la versión 2.4.
- La experiencia generada con esta versión sirvió para diseñar una serie estándares para servicios de Grid.
 - Open Grid Services Architecture (OGSA) 2002.



Globus Toolkit 2.4



The **Globus Toolkit**



Componentes del Globus Toolkit 2.4

G S I

Seguridad





Globus Toolkit 2.4 - GSI

- El elemento básico del GT2.4 es la infraestructura **GSI** (GRID Security Infrastructure) que proporciona servicios de autenticación y comunicación segura.
- Su diseño se basó en las siguientes premisas:
 - Comunicación segura (**autenticación y confidencialidad**) entre los elementos de la GRID.
 - Mecanismo de seguridad **no centralizado**.
 - Soporte para registro único (**single sign-on**), y delegación de credenciales para procesos que involucren múltiples recursos.





Globus Toolkit 2.4 - GSI

- Basado en **PKI**, certificados **X.509** y el protocolo **TLS**.
- La implementación GSI se apega al API: Servicio de Seguridad Genérico (**GSS-API**), propuesto por el IETF para sistemas de seguridad.
- Cada usuario y servicio en la GRID es identificado mediante un certificado.





Componentes del Globus Toolkit 2.4

G R A M

Acceso a Recursos





Globus Toolkit 2.4 - GRAM

- El Gerente de Asignación de Recursos de Grid (**GRAM**) proporciona los mecanismos necesarios para ejecutar y monitorear trabajos en la Grid.
- Una grid construida con Globus, contiene varios GRAMs, cada uno responsable de un recurso físico local.





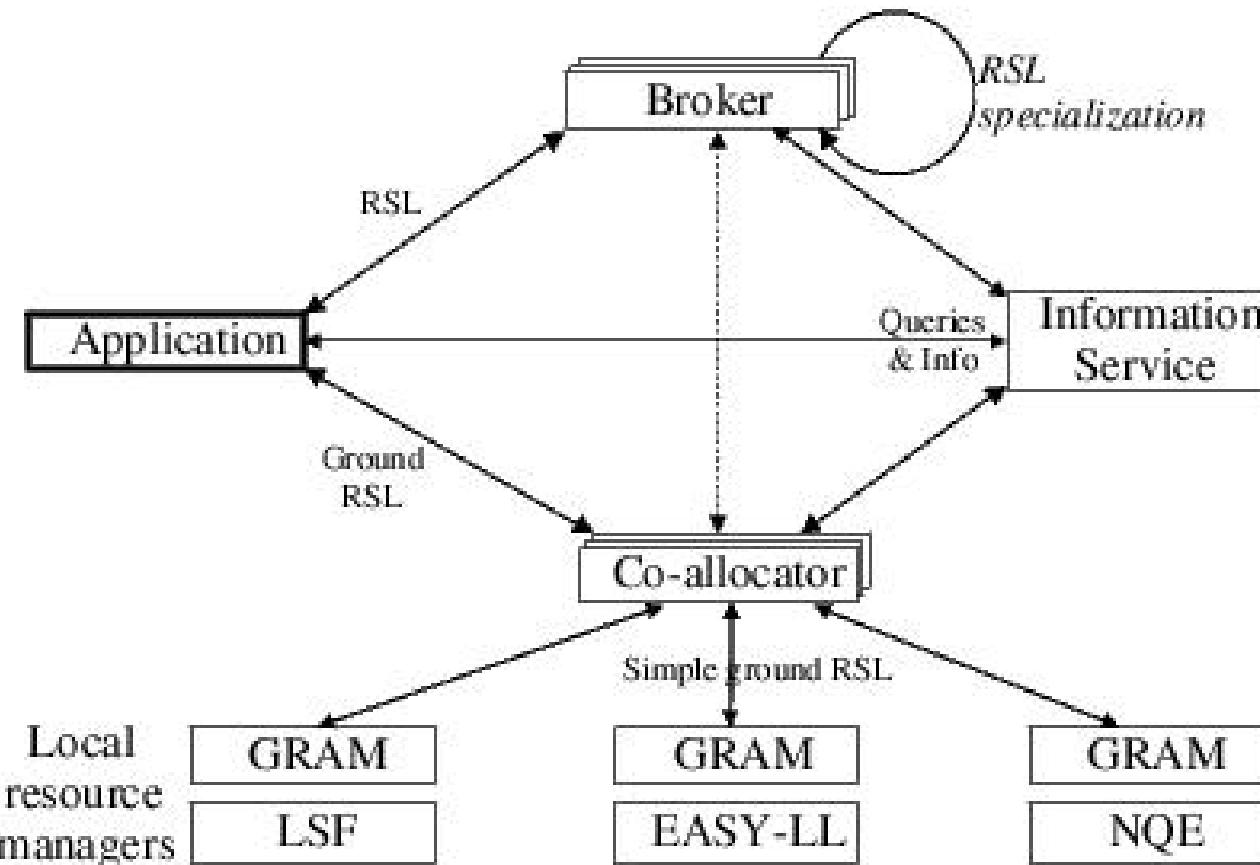
Globus Toolkit 2.4 - GRAM

- Las solicitudes de recursos son expresadas en términos de un Lenguaje de Especificación de Recursos (**RSL**).
- Interfaz única para todos los mecanismos locales de envío de trabajos.
 - LSF, NQE, fork, etc.
- No es necesario recordar/conocer la sintaxis de cada sistema.





Globus Toolkit 2.4 - GRAM



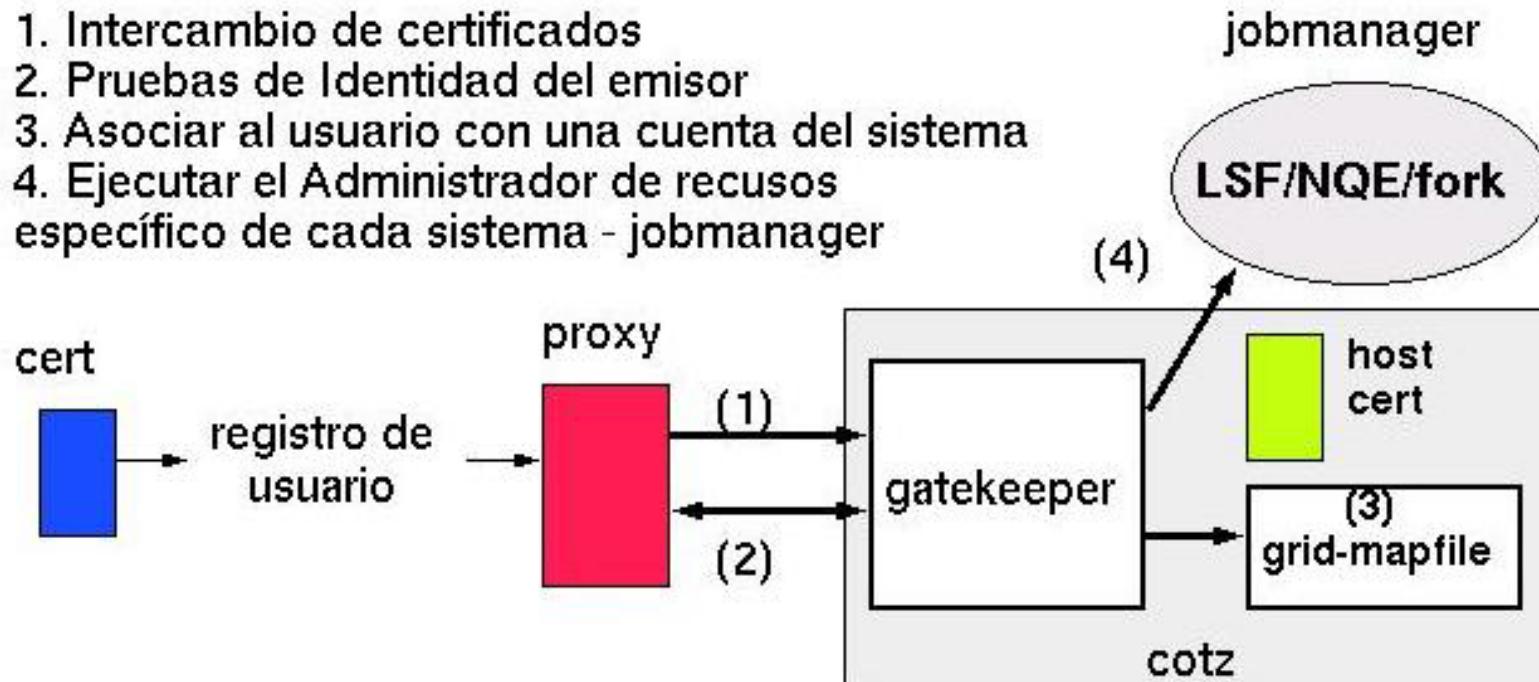


Globus Toolkit 2.4 - GRAM

Ejecución de procesos:

- Para ejecutar el comando "date" en el host "cotz"
% globus-job-run cotz /bin/date

1. Intercambio de certificados
2. Pruebas de Identidad del emisor
3. Asociar al usuario con una cuenta del sistema
4. Ejecutar el Administrador de recusos específico de cada sistema - jobmanager





Componentes del Globus Toolkit 2.4

M D S

Servicios de Información





Globus Toolkit 2.4 - MDS

- El Servicio de Monitoreo y Descubrimiento (**MDS**) está diseñado para proporcionar información sobre los recursos de la Grid, adaptándose a los cambios y estado del sistema.
- MDS utiliza el protocolo LDAP como una interfaz de acceso a la información.
- MDS 2.2 cuenta con:
 - Un proveedor de información local: **GRIS** (Grid Resource Information Service).
 - Un servicio de directorio : **GIIS** (Grid Index Information Service).





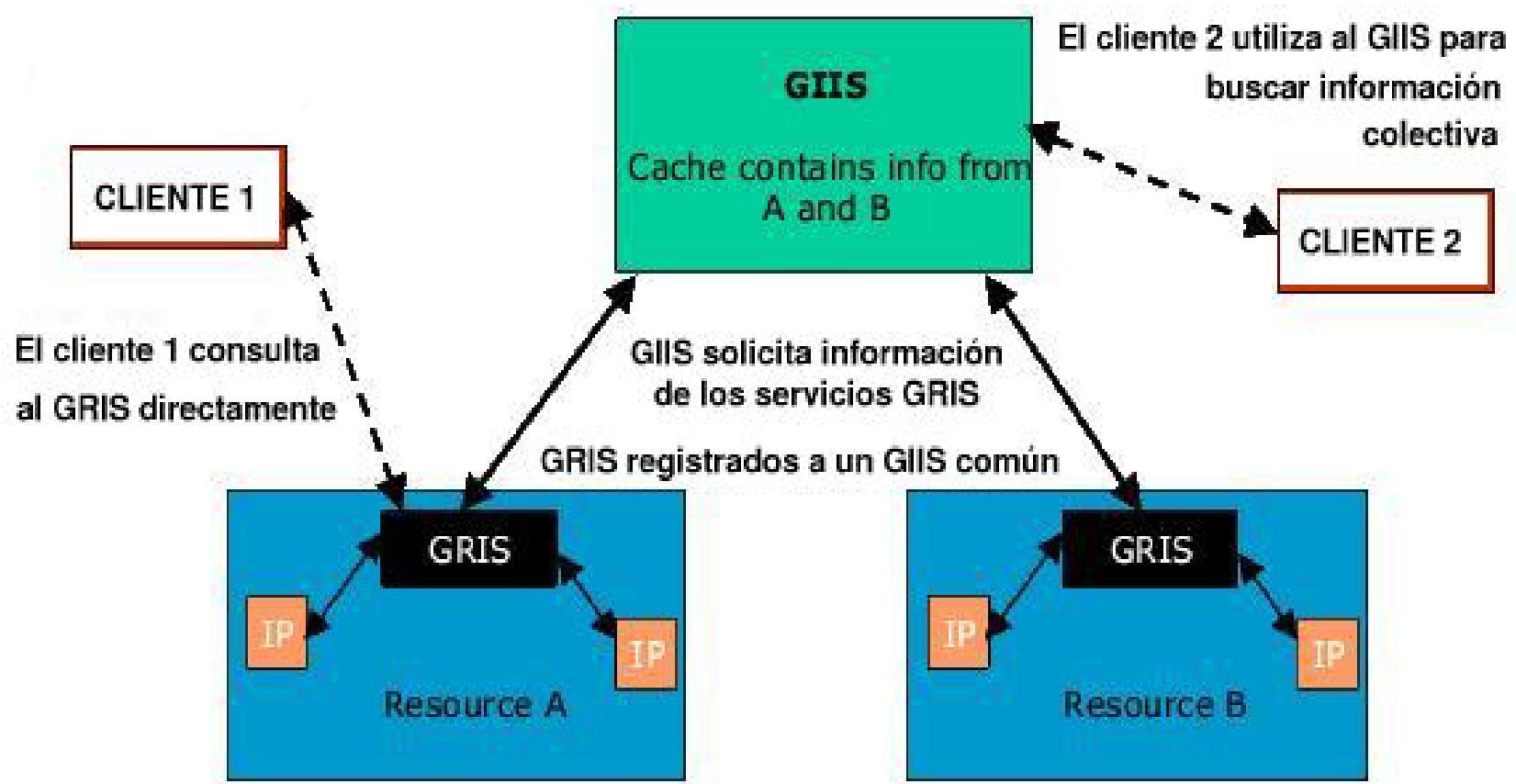
Globus Toolkit 2.4 - MDS

- MDS 2.2 permite obtener información sobre los recursos y el estado operacional del sistema.
 - **Plataforma y arquitectura.**
 - **Sistema operativo** (del host): nombre y versión.
 - **CPU**: tipo, cantidad, versión, velocidad, caché, etc.
 - **Memoria** (física y virtual): tamaño, espacio libre, etc.
 - **Interfaz de red**: nombre de máquina y dirección.
 - **Sistema de archivos**: tamaño, espacio libre, etc.
 - Otros





Globus Toolkit 2.4 - MDS





Componentes del Globus Toolkit 2.4

GridFTP, GASS, GRC, GRM
Control de Datos





Globus Toolkit 2.4 – Datos

- El proyecto Globus se interesa en resolver los retos existentes para implementar una “**GRID de Datos**”.
- Proporciona una infraestructura en la forma de:
 - **Servicios de transferencia de datos.**
 - **Bibliotecas genéricas para el control de datos.**
- Estos bloques fundamentales pueden usarse para construir sistemas y aplicaciones destinadas a usuarios finales.





Globus Toolkit 2.4 – Datos

- ***Global Access to Secondary Storage (GASS)***
 - biblioteca que permite acceder a datos remotos, vía interfaces estándar de entrada y salida.
- ***GridFTP***
 - mecanismo de transferencia de datos seguro, robusto y de alto desempeño.
- ***Globus Replica Catalog (GRC)***
 - un mecanismo para mantener un catálogo de réplicas de conjuntos de datos.
- ***Globus Replica Management (GRM)***
 - mecanismo que conjunta el GRC y el GridFTP para permitir a las aplicaciones, crear y administrar réplicas de datos.





Evaluación del Globus Toolkit 2.4





Evaluación del GT 2.4

Evaluación (+) del Globus Toolkit

- **Buena solución técnica para problemas clave:**
 - Autenticación y autorización.
 - Descubrimiento de recursos y monitoreo.
 - Petición de recursos remotos confiable.
 - Acceso de alto desempeño a datos remotos.
- **Este diseño ha permitido el progreso de:**
 - Esquemas de calidad, soporte de múltiples interfaces y sistemas, una gran base de usuarios y soporte comercial.





Evaluación del GT 2.4

Evaluación (-) del Globus Toolkit

- **Deficiencias en protocolos**
 - Base heterogénea: HTTP, LDAP, FTP.
 - Mecanismos no estándar de invocación, notificación, propagación de error, autorización, terminación, etc.
- **Funcionalidad deficiente**
 - Bases de datos, sensores, instrumentos, etc.
 - Dependencia del sistema.
 - Falta de calidad de servicio de punto-a-punto.
 - Virtualización del sistema final.





OGSA, Web Services y el Globus Toolkit 4





OGSA, Web Services y el GT4

- Open Grid Services Architecture (**OGSA**)
 - Desarrollada por *The Global Grid Forum*
 - Define una arquitectura abierta y estándar para el desarrollo de aplicaciones de grid
 - Especifica las interfaces estándar para cualquier servicio que pueda estar disponible en una grid.
 - Basado en estándares ya existentes.
 - Implementado sobre Servicios Web





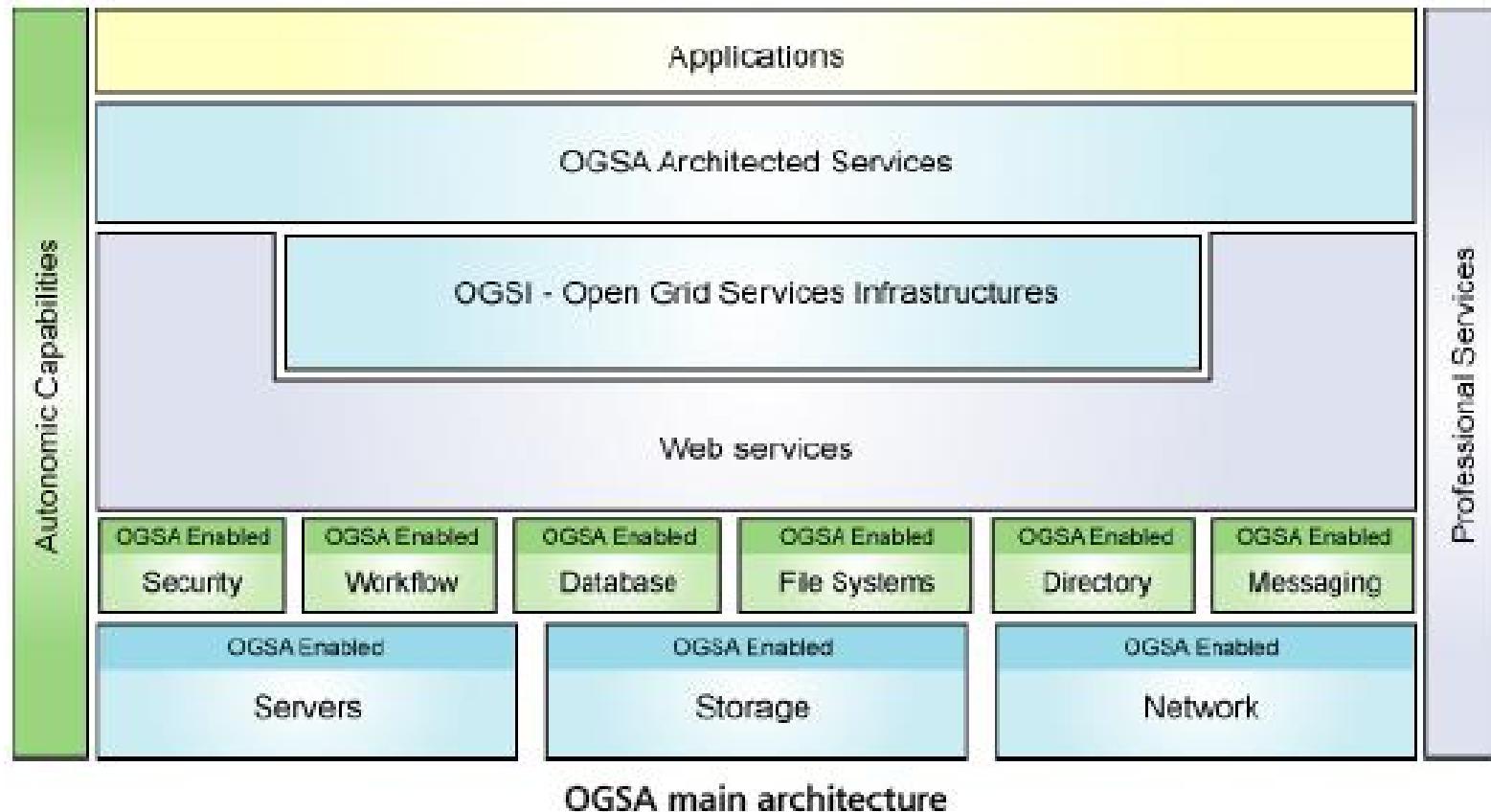
OGSA, Web Services y el GT4

- Técnicamente, OGSA permite
 - Reconstrucción de los protocolos: GRAM, MDS2, etc., conservando todas las características y conceptos del GT 2.4.
 - Un conjunto amplio y estándar de servicios.
 - Integración de la industria en el desarrollo.
 - Valores agregados de tipo comercial.
 - El Globus Toolkit 3.0 fué la primera implementación del estándar OGSA.





OGSA, Web Services y el GT4





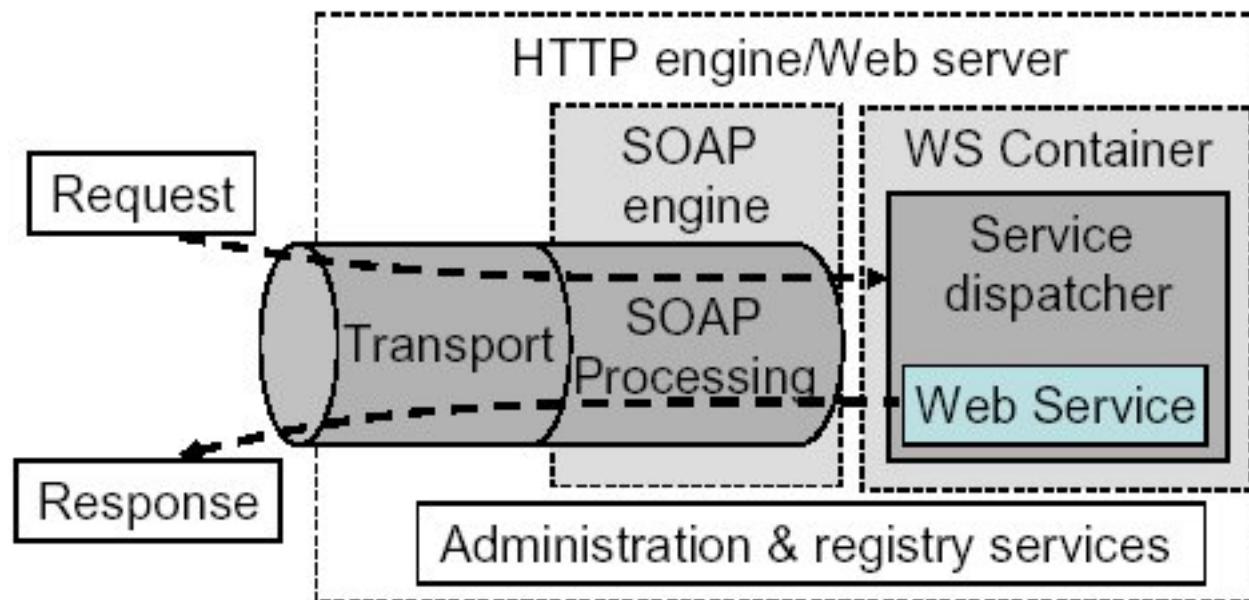
OGSA, Web Services y el GT4

- Servicios Web
 - Creciente popularidad en la estandarización de marcos de trabajo para acceder a aplicaciones de red.
 - Permiten crear aplicaciones cliente/servidor distribuidas y poco acopladas.
 - Independientes de plataforma y lenguaje:
 - XML
 - WSDL: Lenguaje de descripción de servicios Web.
 - SOAP: Protocolo de Acceso a Objeto Simple.
 - Desventajas: sobrecarga, falta de madurez.





Servicios Web





OGSA, Web Services y el GT4

- Globus Toolkit 4.0
 - Liberado en abril del 2005
 - Implementa los principales servicios definidos por OGSA
 - Monitoreo de recursos
 - Servicios de descubrimiento
 - Ejecución de trabajos
 - Seguridad
 - Administración de datos
 - Implementado sobre Web Services Resource Framework (**WSRF**), una extensión a los Web Services.

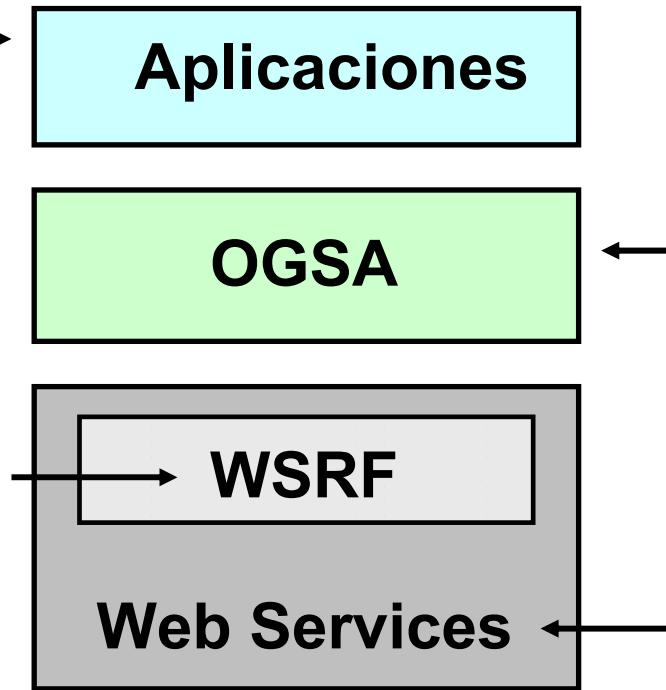




OGSA, Web Services y el GT4

Las aplicaciones se →
construyen sobre los
servicios definidos
por OGSA

Estandarizado (OASIS)
implementado (GT4)



Estándares definidos
por GGF: seguridad,
manejo de recursos,
manejo de trabajos,
servicios de datos,
etc.

GT4 incorpora muchos
de estos servicios.

Estandarizado (W3C,
OASIS, IETF, etc.)
e implementado
(Apache axis)





Globus Toolkit 4.0



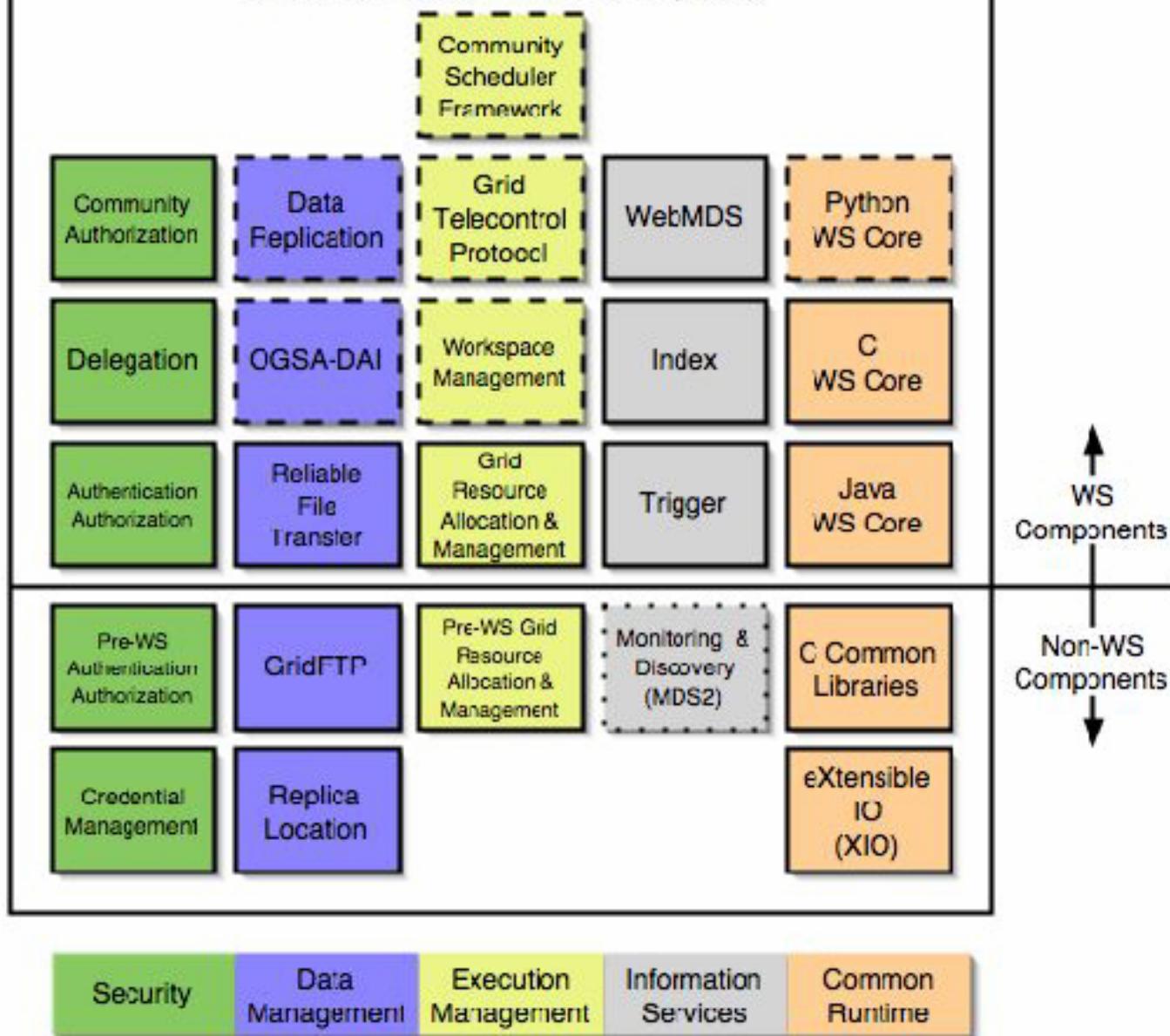


Globus Toolkit 4.0

- En producción, liberado en abril de 2005
- Enfocado a la calidad, robustez, facilidad de uso y documentación!
- Basado en estándares: OGSA, WSRF, XML
- Herramientas de usuario para el desarrollo de interfaces
- Compatible GT3 y con soporte para GT2.4
- Soporte comercial de *UNIVA Corporation*

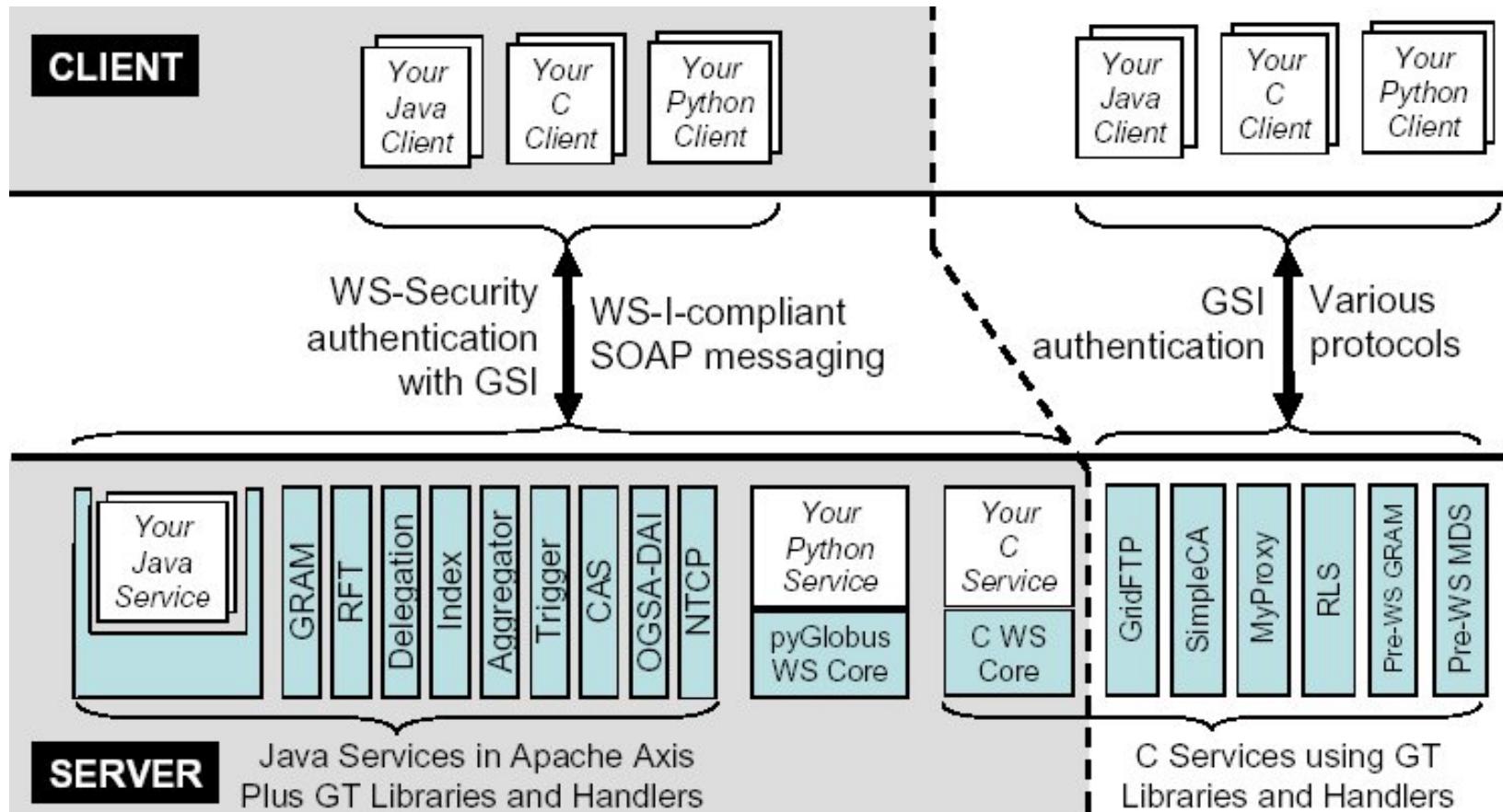


Globus Toolkit® version 4 (GT4)





Componentes del GT4.0





Pre-WebServices

- Técnicamente los mismos servicios del GT2.4
 - GSI, GRAM, MDS, GridFTP
- Permiten la operación de clientes y servicios del GT2.4 con los del GT4.0
 - Servidor GridFTP del GT4 compatible con versiones anteriores
 - Pre WS MDS está descontinuado y desaparecerá en el GT4.4
 - Pre WS GRAM continuará operando, se espera una migración de usuarios hacia WS GRAM





WS Seguridad

- GSI realiza cuatro funciones básicas:
 - Protección de mensajes, autenticación, delegación y autorización.
- Se basa en cuatro estándares
 - TLS (a nivel de transporte) o **WS-Security** y **WS-SecureConversation** (a nivel de mensaje)
 - **X.509 End Entity Certificates** o **Usuario-Contraseña** para autenticación
 - **X.509 Proxy Certificates** y **WS-Trust** para delegación
 - **SAML** para autorización





WS Seguridad

- ***Autenticación y Autorización***: controla el acceso a los servicios y recursos, permitiendo el uso de métodos implementados por el usuario.
- ***Delegación***: servicio que delega credenciales a un contenedor.
- ***Autorización comunitaria***: el Servicio de Autorización Comunitaria (CAS) permite administrar políticas en Organizaciones Virtuales sobre sus recursos.
- ***Manejo de Credenciales***: simpleCA y MyProxy
- ***GSI-OpenSSH***: parche del OpenSSH para aceptar autenticación con certificados.





WS Seguridad

	Message-level Security w/X.509 Credentials	Message-level Security w/Usernames and Passwords	Transport-level Security w/X.509 Credentials
Authorization	SAML and grid-mapfile	grid-mapfile	SAML and grid-mapfile
Delegation	X.509 Proxy Certificates/ WS-Trust		X.509 Proxy Certificates/ WS-Trust
Authentication	X.509 End Entity Certificates	Username/ Password	X.509 End Entity Certificates
Message Protection	WS-Security WS-SecureConversation	WS-Security	TLS
Message format	SOAP	SOAP	SOAP





WS Manejo de datos

- Servicios de descubrimiento, transferencia y acceso a grandes bancos de datos.
 - ***GridFTP***: cliente y servidor FTP optimizado para transferencias de grandes cantidades de datos.
 - ***Reliable File Transfer (RFT)***: servicio de transferencia confiable, a prueba de interrupciones.
 - ***Replica Location (RLS)***: servicio de ubicación de réplicas
 - ***Data Replication (DRS)***: Servicio de replicación de datos
 - ***OGSA-DAI***: acceso e integración de datos, permite integrar conjuntos de datos en formatos diferentes.





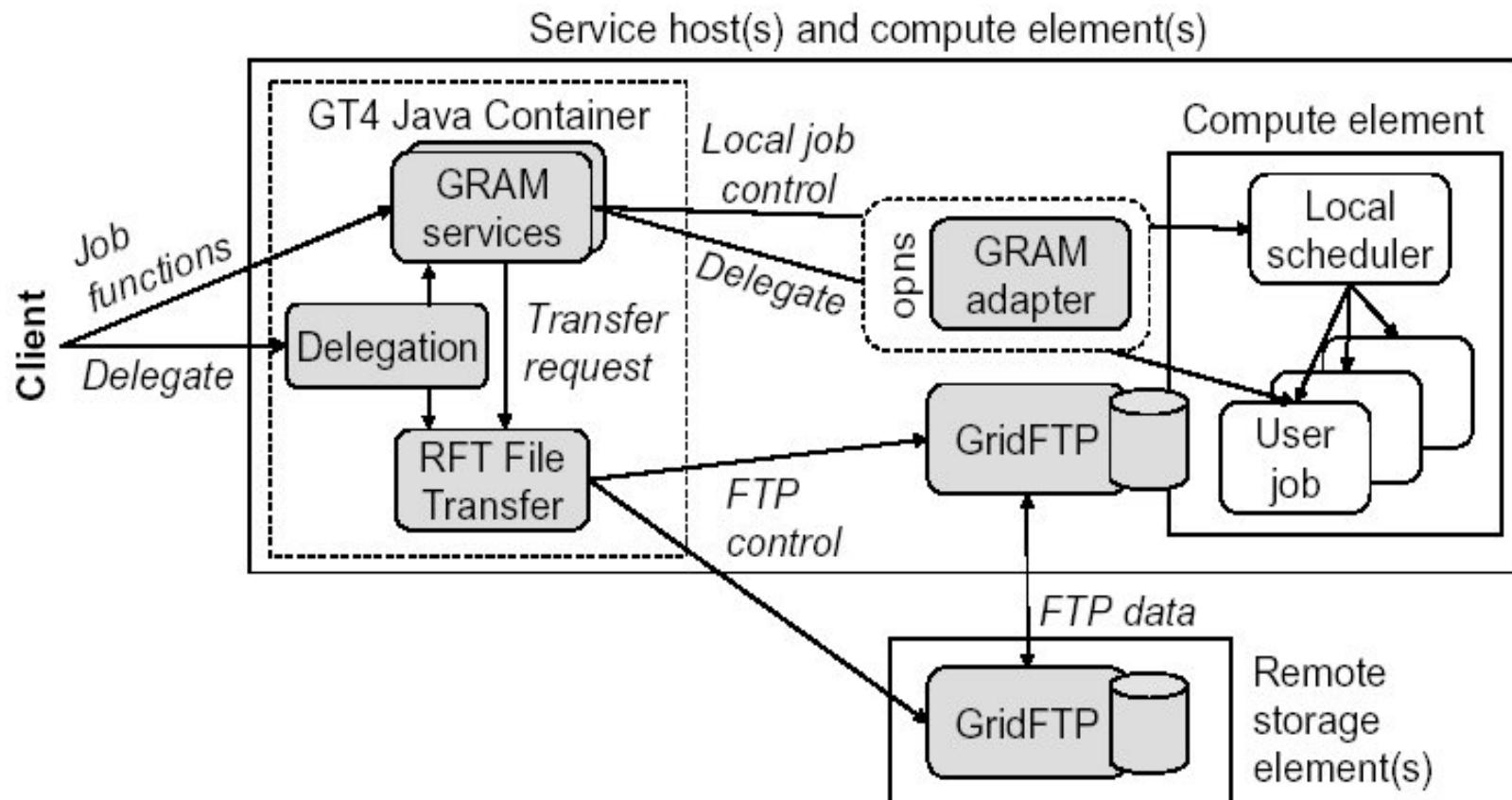
WS Control de ejecución

- Ejecución, calendarización y monitoreo de trabajos.
 - ***WS-GRAM***: ejecuta y monitorea trabajos
 - ***Community Scheduler Framework (CSF)***: interfaz unificada de diferentes calendarizadores (PBS, Condor, LSF, SGE).
 - ***Workspace Management***: permite crear y administrar espacios de trabajo en hosts remotos.
 - ***Grid Telecontrol Protocol***: servicio WSRF para control remoto de instrumentos.





WS-GRAM





WS Servicios de información

- MDS4, permite monitorear y descubrir recursos en la grid.
 - *Servicio de indexado*: permite agregar recursos de interés a la grid.
 - *Trigger Service*: colecta datos de los recursos y realiza acciones basado en esa información.
 - *WebMDS*: proporciona una vista apta para un navegador de los datos colectados por el servicio de indexado.





WS Common Runtime

- Conjunto de bibliotecas y herramientas para desarrollar nuevos servicios.
 - **C**: Incluye herramientas, bibliotecas y hospedaje en WS de programas en C
 - **Python**: Incluye herramientas, bibliotecas y hospedaje en WS de programas en Python
 - **Java**: Incluye herramientas, bibliotecas y servicio de hospedaje para desarrollos en Java.





Conclusiones





Conclusiones

- El GT2.4
 - ha demostrado estabilidad y eficiencia para la implementación de grids en el mundo.
 - Utiliza diversos protocolos, algunos no estándares, que dificultan su uso en aplicaciones más generales o comerciales.
 - Documentación deficiente
 - Sirvió de base para la formulación del estándar OGSA.





Conclusiones

- El GT4.0 (+)
 - Es la primera implementación estable de OGSA
 - Cuenta con una gran diversidad de servicios
 - Protocolos basados en estándares reconocidos por la industria (i.e. soporte comercial)
 - Es compatible en gran medida con el GT2.4
 - Gran mejoría en la documentación
 - Herramientas para desarrollo de aplicaciones





Conclusiones

- El GT4.0 (-)
 - Es más complejo que el GT2.4 y con muchos requerimientos adicionales (Ant, Java, C, Apache, Tomcat, Postgresql).
 - Se requiere el conocimiento de muchas tecnologías no típicas del CAR para comprender su operación.
 - Pese a que el rendimiento de los WebServices ha mejorado sigue siendo una tecnología costosa computacionalmente.





Semana de Cómputo Científico Supercómputo, Visualización y Realidad Virtual

2006

Ciudad de México
del 24 al 28 de abril de 2006
Ciudad Universitaria

<http://dci.dgsca.unam.mx/semana2006>

