



Tema 3. Introducción a las Tecnologías Grid

Tema 3.4. Manejo de datos en un Entornos Grid

Conceptos de la Computación en Grid y Cloud



- Describir la Forma en la que se Manejan los Datos en un Entorno Grid.
- Describir las Formas Más Eficientes de Disponer la Información para su Uso en un Entorno Grid.
- Describir los Protocolos que se Usan para Transferir Ficheros en un Entorno Grid.
- Aplicar los Mecanismos de Globus Toolkit para Acceder y Transmitir Datos.

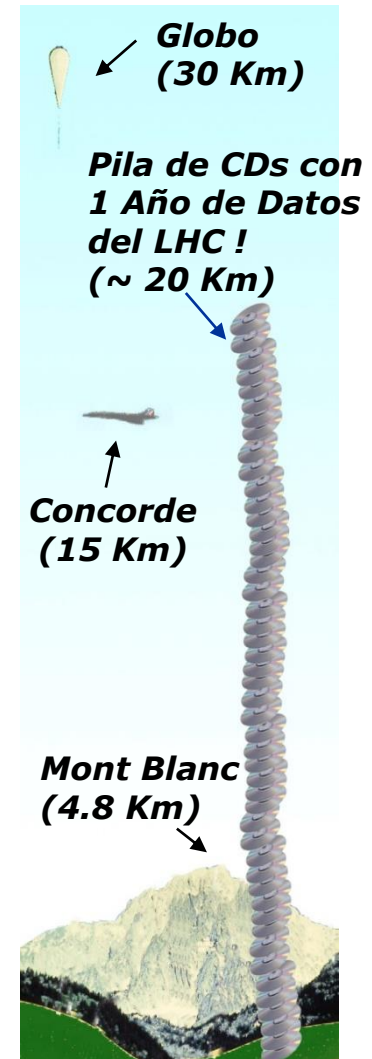
- Modelos de Acceso a Datos en Grids
 - Necesidades (Proyectos)
 - Conceptos Generales
 - Alternativas
- Soluciones propuestas por Globus Toolkit
 - Global Access to Secondary Storage (GASS)
 - Integración GASS-GRAM
 - GridFTP

- **Práctica 1:** Lanzamiento de Trabajos Grid RSL (GASS Server – Input/Output).
- **Práctica 1:** Lanzamiento de Trabajos Grid RSL (GASS Server - Staging).
- **Práctica 3:** Cálculo del Número Pi Mediante una Aproximación por Rectángulos (RSL).
- **Práctica 4:** Lanzamiento de Trabajos Grid (GridFTP).

- **Modelos de Acceso a Datos en Grids**
 - Necesidades (Proyectos)
 - Conceptos Generales
 - Alternativas
- **Soluciones propuestas por Globus Toolkit**
 - Global Access to Secondary Storage (GASS)
 - Ejemplos
 - Integración GASS-GRAM
 - GridFTP

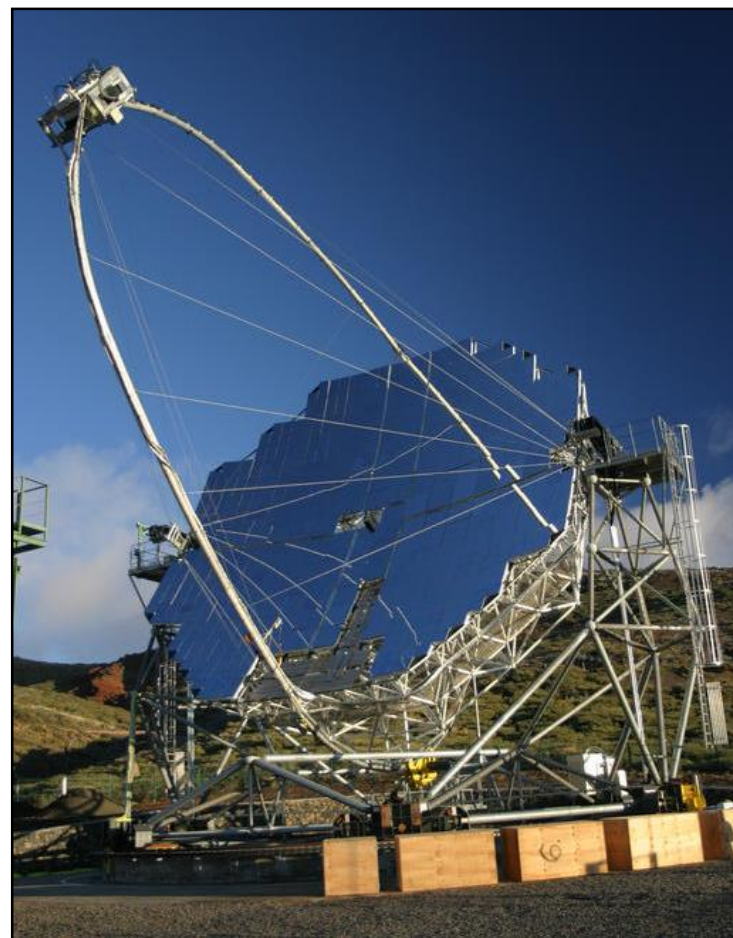
- Almacenamiento, Preproceso y Análisis de los Datos Obtenidos por el Large Hadron Collider (LHC)

- El LHC Genera 40 Millones de Colisiones por Segundo.
- Una Vez Clasificados, Se Analizarán de Forma Intensiva Para Verificar Teorías Correspondientes a la Estructura Básica de la Materia, Como el Boson de Higgs.
- Tras el Filtrado Quedan 100 Colisiones Interesantes por Segundo.
- Cada Colisión Genera un Megabyte de Datos, lo que Implica un Ratio de Registro de 0.1 Gigabytes/seg.
- En Total ~ 10 Petabytes/por Año, lo que Corresponde a 20 millones de CDs por Año!
- Su Proceso Requeriría $\sim 100,000$ PCs Actuales.



Proyectos - MAGIC (Major Atmospheric Gamma-ray Imaging telesCope)

- Telescopio de 17 metros (Uno de los mas grandes de su Tecnología)
- Ubicado en la Isla de La Palma a 2200 m. de Altitud.
- El Objetivo es Detectar Fuentes de Rayos Gamma en un Rango Inexplorado.
- Requerimientos
 - Cada Evento Implica ~ 20 kBytes.
 - Se Adquieren 500 Eventos por Seg.
 ~ 10 MByte/seg.
 - Dado que Hay unas 1000 h de Observación Anuales, Esto Implica ~ 36 TByte/año.



Modelos de Acceso a Datos en Grids

Conceptos Generales

- Las Aplicaciones Científicas Distribuidas Requieren a Menudo el Acceso a Grandes Cantidades de Datos (Terabytes, Petabytes).
- En estos Casos, el Acceso a Datos Distribuidos es Tan Importante como el Acceso a Recursos de Proceso.
- Por Tanto, Un Objetivo Prioritario es: “Permitir a una Gran Comunidad Geográficamente Distribuida Coordinar sus Recursos para la Realización de Análisis Sofisticados y Computacionalmente Intensivos de Petabytes de Datos”.
- En Cualquier Caso y **de forma General, el Acceso a Datos Implica Simplemente la Transmisión de Datos y Resultados De y Desde los Computadores de la Infraestructura Grid Implicados en un Proceso.**

- Generalmente los Grids de Datos Actúan como Grandes Sistemas de **Ficheros Distribuidos y Compartidos**.
 - La Organización de los Datos Es Más Cercana a una Estructura Jerárquica de Directorios que a un Entorno Relacional.
 - No Obstante, Es Posible Definir en Muchos Casos Metadatos que Enriquezcan la Información.
- Los Datos en los Grids Son **Generalmente de Sólo Lectura**.
 - La Modificación Generalmente Implica la Creación de una Nueva Entrada de Datos.

- La **Replicación de Datos** es un Factor Importante
 - **Fiabilidad y Robustez.** Los Datos Deben Estar Accesibles el Mayor Tiempo Posible.
 - **Eficiencia.** Puede Haber Disponibles Diferentes Copias en Diferentes Ubicaciones Para Repartir Ancho de Banda y Seleccionar los Almacenamientos Más Próximos.
 - **Integridad.** La Condición de Sólo Lectura Simplifica la Coherencia de Diferentes Réplicas.

- Existen Numerosos Sistemas de Ficheros Distribuidos
 - Distributed Parallel Storage Systems (DPSS).
 - High Performance Storage Systems (HPSS)
 - Utiliza Técnicas de Altas Prestaciones Para la Transferencia de Datos.
 - Distributed File Systems (DFS)
 - Gran Volumen, Replicación de Datos, Tamponado (cache) Local.
 - Storage Resource Broker (SRB)
 - Conecta Colecciones de Datos Heterogéneos desde un Interfaz Uniforme y Permitiendo Búsquedas sobre el Metadata.
- Problemas
 - Los Protocolos son Incompatibles, Propietarios y a Veces No son Públicos
 - Cada Cliente Debe Incorporar el Soporte a los Diferentes Protocolos.
 - No hay Interoperabilidad.
 - Cada Protocolo Tiene Una Parte de la Funcionalidad Deseada.
 - Ninguno Se Usa Para la Transmisión de Datos sobre Internet.

- Modelos de Acceso a Datos en Grids
 - Necesidades (Proyectos)
 - Conceptos Generales
 - Alternativas
- **Soluciones propuestas por Globus Toolkit**
 - Introducción
 - Global Access to Secondary Storage (GASS)
 - Integración GASS-GRAM
 - GridFTP

- Acceso a Datos y Transferencia
 - **GASS:** Proporciona Servicios Principalmente Orientados para el GRAM (Redirección de la E/S, Volcado de Ficheros).
 - **GridFTP:** Proporciona Altas Prestaciones, Transferencia Fiable sobre Redes WAN.

- Global Access to Secondary Storage (GASS)
Es un Servicio que Permite Manejar Datos Almacenados de Forma Remota.
- GASS Permite Definir Identificadores que Identifican el Recurso, la Ubicación y el Nombre de los Archivos Accesibles por el Sistema.
- GASS Se Integra con los Sistemas de Seguridad, Lanzamiento de Tareas y Comunicaciones de globus.

- **Los Servicios del GASS Permiten:**

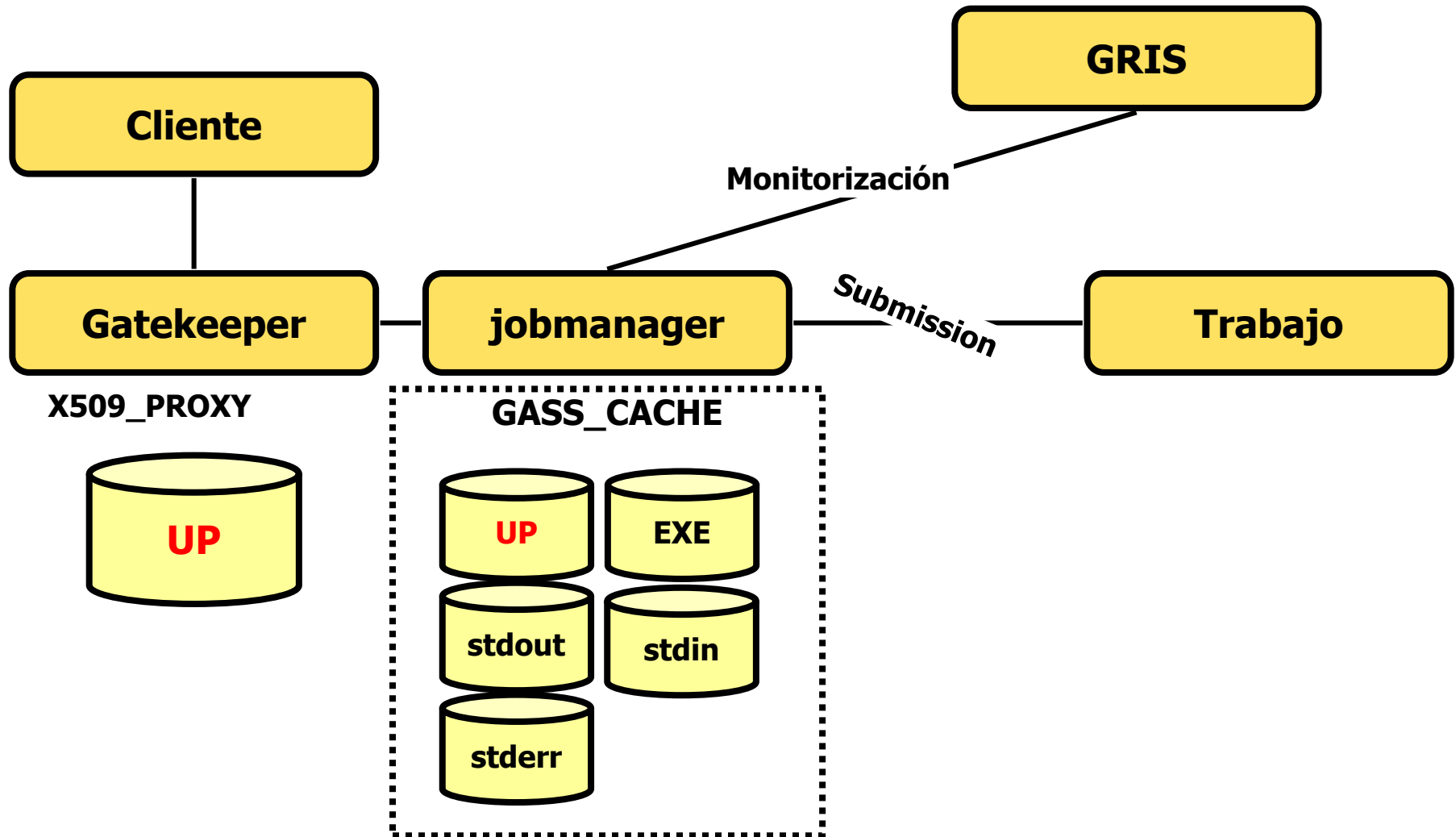
- Indicar al Servicio GRAM la Ubicación de los Ejecutables Cuando Éstos no se Encuentran en la Máquina Destino.
- Acceder a Ficheros de Datos que se Encuentran en un Recurso Diferente del Elegido para la Ejecución.
- Gestionar la Entrada (stdin), Salida (stdout) y Error (stderr) Estándar Desde una Ubicación Remota.

- **GASS**

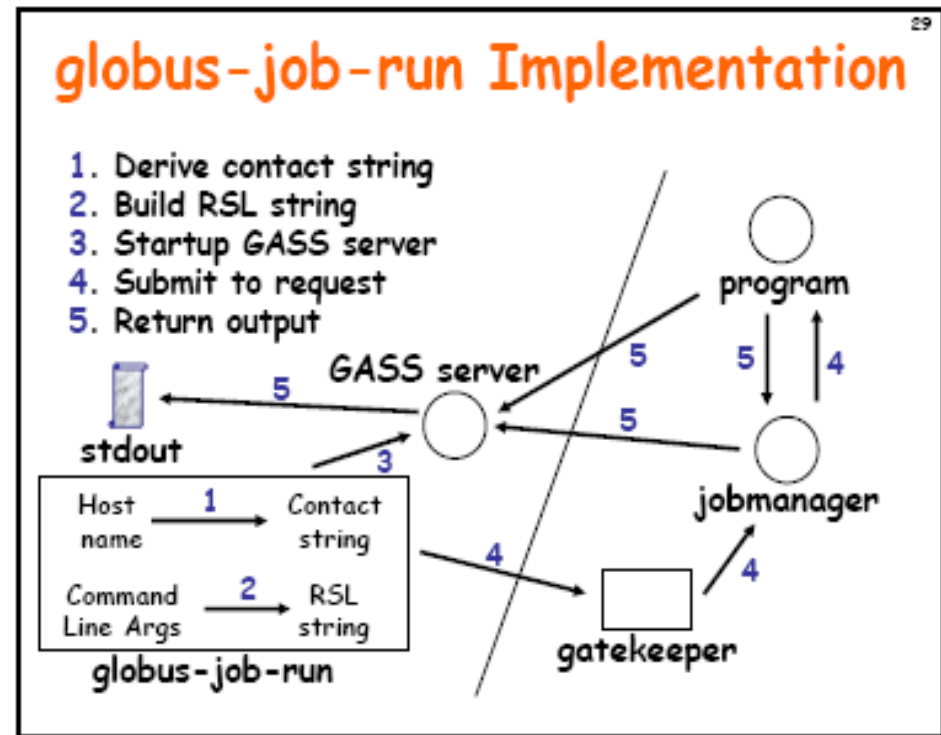
- Utiliza URLs para Identificar Ejecutables, Ficheros de Datos Entrada, Salida y Error Estándar, en una Sintaxis Integrada en el Resource Specification Language (RSL).
- Permite Sustituir las Llamadas a Funciones de E/S (open/close) por Funciones Equivalentes (globus_gass_open/close, globus_gass_fopen/fclose).

Soluciones Propuestas por Globus Toolkit

GASS – Integración con GRAM



- En la Implementación del Comando globus-job-run se Incluye Directamente el Uso del GASS.
 - A Partir del Nombre de la Máquina y de la Línea de Argumentos se Construyen las Direcciones Remotas para los Datos y Ejecutables (1, 2).
 - Se Activa un Servidor GASS (3).
 - Se Ejecutan Los Trabajos a Partir de globusrun (4).
 - La Salida se Recupera a Través del GASS en el Computador Local (5).



- ## Ventajas

- Carga Transparente y Bajo Demanda de Ficheros de Datos.
- Tamponado (caché) de Ficheros Pequeños.
- Volcado Automático de Código y Datos en Recursos Remotos.
- Volcado Automático y en Tiempo Real de la Salida Por Pantalla.

- ## Limitaciones

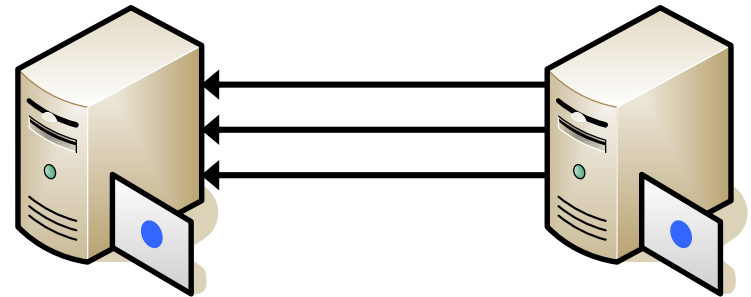
- No es Adecuado para la Transferencia de Grandes Volúmenes de Datos en Entornos de Altas Prestaciones.

P1, P2 y P3

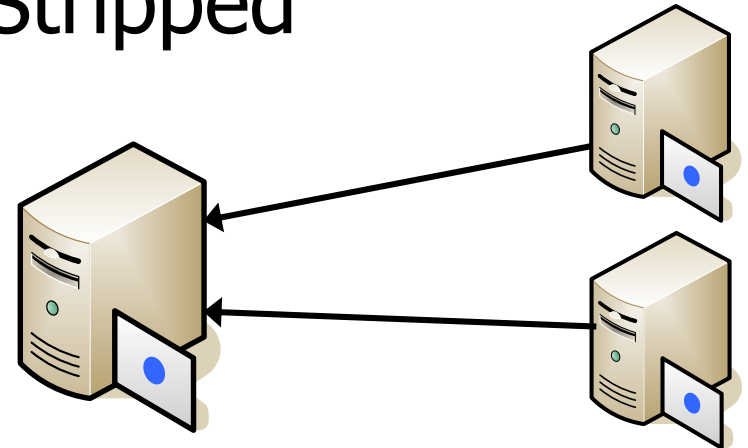
GridFTP

- GridFTP Esconde las Particularidades de los Sistemas Proporcionando un Interfaz Único
 - Similar al Papel que Juega GRAM en el Lanzamiento de Trabajos.
- GridFTP está Basado en FTP, Incorporando una Serie de Características Adecuadas para el Grid.
 - GridFTP > FTP
 - FTP es el Estándar “De-facto” para la Transferencia de Ficheros.
 - FTP es Interoperable con Muchas Herramientas Propietarias.
- GridFTP Proporciona un Interfaz a Muchos Sistemas de Almacenamiento
 - HPSS, DPSS, Sistemas de Fichero.

- Integración con GSI
 - Directamente Utiliza las Credenciales del Usuario para la Autenticación.
- Restart Extendido
 - Permite una Recuperación Automática Tras Fallos en las Conexiones.
- Canales de Datos Paralelos
 - Mejora el Ancho de Banda Total Efectivo Mediante la Creación de Múltiples Streams de TCP en Paralelo, Incluso sobre el Mismo Recurso.



- **Transferencia Parcial de Datos**
 - Extensión de los Comandos de FTP para la Transferencia de Partes de un Fichero.
 - No Sólo Implica una Mejora en la Fiabilidad, Sino que Además es Posible Recuperar una Parte Concreta de un Fichero Únicamente.
- **Transferencia de Datos "Stripped"**
 - Múltiples Conexiones desde Servidores Distribuidos Para un Único Fichero.



- **globus-url-copy**
 - Permite la Copia de Ficheros Remotos y Soporta Diversos Protocolos (file://, http://, https://, ftp://, gsiftp://).
 - Necesita Disponer de un Proxy Válido.
- **Sintaxis**
 - globus-url-copy [opciones] URL_Fuente URL_Destino
 - Donde la URL es Cualquier URL Válida Definida en RFC 1738 que Contenga un Protocolo Soportado.
 - Formato URL: **protocol://[host]:[port]/path.**

- Especificando Todos los Parámetros
 - `gsiftp://myhost.mydomain.com:2812/data/foo.dat`
- Sin Indicar el Puerto (Por Defecto, 80 en Este Caso)
 - `http://myhost.mydomain.com/mywebpage/default.html`
- Fichero Local (Protocolo File, No Se Especifica Puerto ni Host).
 - `file:///foo.dat`

● Opciones

- -vb | -verbose Mostrar Información Adicional Durante la Transferencia:
 - Número de Bytes Transferidos, Tasa de Envío Actual y Media, etc.
- -dbg | -debug, Muestra Información de Depuración, Muy Adecuado Cuando Hay Problemas.
- -f <filename> Toma de un Fichero la Lista Pares de URLs Para cada Transferencia. Cada Línea del Fichero Debe Tener el Formato: <sourceURL> <destURL> Indicadas entre Comillas. Comentarios Empezados con "#".
- -r | -recurse, Copia Ficheros en los Directorios.

- **Correcto Uso de los Protocolos**
 - file://
 - Permite Especificar Ficheros en la Máquina Local.
 - http://, https://
 - Permite Definir como Fuente un Fichero Accesible por Web.
 - ftp://
 - Permite Acceder a Un Servidor FTP.
 - Admite Lectura y Escritura, Permitiendo Especificar Usuario y Contraseña (Altamente Inseguro).
 - gsiftp://
 - Protocolo Básico del GridFTP.
 - Admite Lectura y Escritura, Requiere que el Demonio de GridFTP esté en Marcha en la Máquina Destino y que el Directorio Especificado Tenga para el Usuario Local los Permisos Correspondientes.
 - Incorpora la Autenticación por GSI.

GridFTP: Ejemplo 3

• Origen

- Host: Local
- Protocolo: file
- Puerto: Por Defecto
- Directorio: /home/profesor
- Fichero: datos.txt

• Destino

- Host: idunv08.i3m.upv.es (Remoto)
- Protocolo: gsiftp
- Puerto: Por Defecto
- Directorio: /home/globus
- Fichero: inp1.txt

```
globus-url-copy file:///home/profesor/datos.txt  
gsiftp://idunv08.i3m.upv.es/home/profesor/datos.txt
```

GridFTP: Ejemplo 4

• Origen

- Host: idunv08.i3m.upv.es (remoto)
- Protocolo: GridFTP
- Puerto: Por Defecto
- Directorio: /proc
- Fichero: cpuinfo

• Destino

- Host: local
- Protocolo: file
- Puerto: Por Defecto
- Directorio: /home/iblanque
- Fichero: cpu_geb.txt

`globus-url-copy`

`gsiftp://idunv08.i3m.upv.es/home/profesor/datos.txt` `file:///home/profesor/datos2.txt`

Catálogos y Réplicas

- Tareas del Replica Management
 - Ofrecer Nombres Lógicos Para Datos Almacenados en El Grid.
 - Mantener la Correspondencia entre los Nombres Lógicos y las Colecciones de Réplicas, Considerando su Ubicación Física.
 - Integrarse con los Sistemas de Información y Seguridad del Grid.
 - Utilizarlo como un Componente para Otros Servicios y Aplicaciones de Mayor Nivel.
 - Ofrecer la Información de Dónde y Cuándo se Han Creado las Réplicas.
- Limitaciones
 - No Incorpora Coherencia con los Datos, por lo que Éstos Deben ser de Sólo Lectura o la Coherencia la Debe Proporcionar un Nivel Superior.

P4

PRÁCTICAS

Práctica 1: Lanzamiento de Trabajos Grid RSL (GASS Server – Input/Output)

Práctica 1: Lanzamiento de Trabajos Grid RSL (GASS Server – Input/Output)

- Sitúate en la carpeta siguiente:

`$HOME/evidencias/Grid/03_Manejo_Datos/P1`

- Todos los ficheros que generes en esta práctica debes de guardarlo en esta carpeta.

Práctica 1: Lanzamiento de Trabajos Grid RSL (GASS Server – Input/Output)

- Para Poder Acceder Ficheros de forma remota es Necesario Crear un Servidor GASS en la Máquina Local (como Usuarios).
- Este Servidor se Pone en Marcha Por el Usuario antes de lanzar el trabajo Grid, Para Permitir el Acceso a Sus Ficheros desde el equipo remoto.
- Utiliza Por Defecto el Protocolo Https.
- La Sintaxis para arrancar el servicio es la siguiente:
 - `globus-gass-server -r -w -t`
 - -r: Permite la Lectura de Ficheros en este Servidor.
 - -w: Permite la Escritura de Ficheros en este Servidor.
 - -i: Pone en Marcha un Servidor No Seguro (http).
 - -t: Extiende el `~` como home en los Paths
 - -p: Puerto
 - -help: Ayuda.

Práctica 1: Lanzamiento de Trabajos Grid RSL (GASS Server – Input/Output) Ejercicio 1

- En este ejercicio se quiere lanzar un “cat” a un equipo remoto (p.e. el de tu compañero) de un richero local.
- Para ello los pasos a seguir son los siguientes:
 - Paso 1.** Crear un fichero Local llamado “miStdIn” y escribe en el tu nombre.
 - Paso 2.** Crea un certificado Proxy con tus credenciales de Usuario.
 - Paso 3.** Activar el GASS en la Máquina Local (Segundo Plano). Utiliza el comando siguiente:

```
[ccgc@XXX P1]$ globus-gass-server -r -w -t  
https://xxx.ccgc.mastercpd.upv.es:50428
```

Verás que te ha devuelto una URL, que será la dirección que habrá que indicar en nuestros trabajos GRID RSL, para que de forma remota acceda a los ficheros de nuestra maquina local a través del servicio GASS.

Nota: Si te aparece el mensaje siguiente “Warning: You do not have valid credentials at this time” es que no has creado las credenciales de usuario. Debes de matar el proceso, crear las credenciales y volverlo a lanzar.

Práctica 1: Lanzamiento de Trabajos Grid RSL (GASS Server – Input/Output) Ejercicio 1

Paso 4. Crea un fichero que describa vuestro trabajo Grid (llámalo “ejercicio1.rsl”), utilizando el lenguaje RSL. Debe de contener el siguiente texto, sustituyendo la URL de vuestro servicio GASS (marcado en rojo).

```
&
(count = 1)
(executable = /bin/cat)
(rsl_substitution = (GASS_URL https://<mi_maquina>.ccgc.mastercpd.upv.es:<puerto>))
(stdin = $(GASS_URL)/home/ccgc/Evidencias/Grid/03_Manejo_Datos/P1/miStdIn)
(stdout = $(GASS_URL)/home/ccgc/Evidencias/Grid/03_Manejo_Datos/P1/miStdOut)
(stderr = $(GASS_URL)/home/ccgc/Evidencias/Grid/03_Manejo_Datos/P1/miStdErr)
```

Paso 5. Lanzar el Trabajo al equipo remoto de tu compañero con globusrun.

```
[ccgc@XXX P1]$ globusrun -r XXX.ccgcc.mastercpd.upv.es -f ejercicio1.rsl
```

Paso 6. Comprueba que se te ha generado el Fichero miStdOut.

Práctica 1: Lanzamiento de Trabajos Grid RSL (GASS Server – Input/Output) Ejercicio 2

- En este ejercicio se quiere lanzar un trabajo desde vuestro equipo local a una maquina remota (Equipo Remoto 1) donde el fichero de entrada estándar se ubique en una segunda maquina remota (Equipo Remoto 2) y que los resultados se guarden en vuestra maquina local.
- Elige dos de tus compañeros, uno para que os ceda su recurso donde ejecutar el trabajo y otro donde se alberguen el fichero que actúe como entrada estándar.
- El trabajo a lanzar es un cat, y pretende sacar por la salida estándar la entrada estándar que corresponderá con el fichero creado por vuestro compañero en la Maquina Remota 2 en el directorio *\$HOME/evidencias/Grid/03_Manejo_Datos/P1*. (reutiliza el fichero *miStdIn* del ejercicio 1)
- El trabajo RSL que defináis, debe llamarse *ejercicio2.rsl* y los ficheros de salida Estándar y Error Estándar deben de llamarse *miStdIn2* y *miStdOut2*.

Práctica 1: Lanzamiento de Trabajos Grid RSL (GASS Server) Evidencias

- En la carpeta *\$HOME/evidencias/Grid/03_Manejo_Datos/P1* debe de aparecer los ficheros:

ejercicio1.rsl	→ Trabajo rsl del ejercicio 1.
ejercicio2.rsl	→ Trabajo rsl del ejercicio 2.
miStdIn	→ Fichero de Entrada Estándar Generado
miStdOut	→ Fichero de Salida Estándar del ejercicio 1
miStdOut2	→ Fichero de Salida Estándar del ejercicio 2
miStdErr	→ Fichero de Salida Error del ejercicio 1
miStdErr2	→ Fichero de Salida Error del ejercicio 2

Práctica 2: Lanzamiento de Trabajos Grid RSL (GASS Server - Staging)

Práctica 2: Lanzamiento de Trabajos Grid RSL (GASS Server – Staging)

- Sitúate en la carpeta siguiente:

`$HOME/evidencias/Grid/03_Manejo_Datos/P2`

- Todos los ficheros que generes en esta práctica debes de guardarlo en esta carpeta.

Práctica 2: Lanzamiento de Trabajos Grid RSL (GASS Server - Staging) Ejercicio 1

- En este ejercicio se quiere lanzar un trabajo cuyos argumentos de entrada son ficheros, por lo que dichos ficheros deben de viajar con el trabajo (staging).
- El trabajo a lanzar es un sort, y pretende ordenar dos ficheros de texto "Entrada1.txt y Entrada2.txt".
- Para ello los pasos a seguir son los siguientes:

Paso 1. Crear dos ficheros Locales llamados "Entrada1.txt" y "entrada2.txt" y escribe dentro de ellos unos cuantas palabras que quieras ordenar.

Paso 2. Crea un certificado Proxy con tus credenciales de Usuario.

Paso 3. Activar el GASS en la Máquina Local (Segundo Plano). Utiliza el comando siguiente:

```
[ccgc@XXX P1]$ globus-gass-server -r -w -t  
https://xxx.ccgcmastercpd.upv.es:50428
```

Verás que te ha devuelto una URL, que será la dirección que habrá que indicar en nuestros trabajos GRID RSL, para que de forma remota acceda a los ficheros de nuestra maquina local a través del servicio GASS.

Nota: Si te aparece el mensaje siguiente "Warning: You do not have valid credentials at this time" es que no has creado las credenciales de usuario. Debes de matar el proceso, crear las credenciales y volverlo a lanzar.

Práctica 2: Lanzamiento de Trabajos Grid RSL (GASS Server – Staging) Ejercicio 1

Paso 4. Crea un fichero que describa vuestro trabajo Grid (llámalo “ejercicio1.rsl”), utilizando el lenguaje RSL. Debe de contener el siguiente texto, sustituyendo la URL de vuestro servicio GASS (marcado en rojo).

```
&
(count = 1)
(executable = /bin/sort)
(arguments = Entrada1.txt Entrada2.txt -o Salida.txt )
(rsl_substitution = (GASS_URL https://<mi\_maquina>.ccgc.mastercpd.upv.es:<puerto>))
(stdout = $(GASS_URL)/home/ccgc/Evidencias/Grid/03_Manejo_Datos/P2/miStdOut)
(stderr = $(GASS_URL)/home/ccgc/Evidencias/Grid/03_Manejo_Datos/P2/miStdErr)
(file_stage_in = ($(GASS_URL)/home/ccgc/Evidencias/Grid/03_Manejo_Datos/P2/Entrada1.txt Entrada1.txt)
  ($(GASS_URL)/home/ccgc/Evidencias/Grid/03_Manejo_Datos/P1/Entrada2.txt Entrada2.txt))
(file_stage_out = (Salida.txt $(GASS_URL)/home/ccgc/Evidencias/Grid/03_Manejo_Datos/P2/Salida.txt))
```

Paso 5. Lanzar el Trabajo al equipo remoto de tu compañero con globusrun.

```
[ccgc@XXX P1]$ globusrun -r XXX.ccgc.mastercpd.upv.es -f ejercicio1.rsl
```

Paso 6. Comprueba que se te ha generado el Fichero miStdOut.

Práctica 1: Lanzamiento de Trabajos Grid RSL (GASS Server – Staging) Ejercicio 2

- En este ejercicio se quiere lanzar un trabajo desde vuestro equipo local a una maquina remota (Equipo Remoto 1) donde el fichero de entrada estándar se ubique en una segunda maquina remota (Equipo Remoto 2) y que los resultados se guarden en vuestra maquina local.
- Elige dos de tus compañeros, uno para que os ceda su recurso donde ejecutar el trabajo y otro donde se alberguen el fichero que actúe como entrada estándar.
- El trabajo a lanzar es el mismo que el ejercicio 1, y pretende ordenar dos ficheros ubicados en la Maquina Remota 2 en el directorio *\$HOME/evidencias/Grid/03_Manejo_Datos/P2*.
(reutiliza el fichero "Entrada1.txt y Entrada2.txt" del ejercicio 1) y crear un fichero resultado en local
- El trabajo RSL que defináis, debe llamarse ejercicio2.rsl y el fichero resultado debe de llamarse Salida2.txt

Práctica 1: Lanzamiento de Trabajos Grid RSL (GASS Server) Evidencias

- En la carpeta *\$HOME/evidencias/Grid/03_Manejo_Datos/P2* debe de aparecer los ficheros:

ejercicio1.rsl	→ Trabajo rsl del ejercicio 1.
ejercicio2.rsl	→ Trabajo rsl del ejercicio 2.
Entrada1.txt	→ Fichero de texto con nombres.
Entrada2.txt	→ Fichero de texto con nombres.
Salida.txt	→ Fichero resultado del ejercicio1.
Salida2.txt	→ Fichero resultado del ejercicio2.
miStdOut	→ Fichero de Salida Estándar del ejercicio 1
miStdOut2	→ Fichero de Salida Estándar del ejercicio 2
miStdErr	→ Fichero de Salida Error del ejercicio 1
miStdErr2	→ Fichero de Salida Error del ejercicio 2

Práctica 3: Cálculo del Número Pi Mediante una Aproximación por Rectángulos (RSL)

Práctica 3: Cálculo del Número Pi Mediante una Aproximación de Rectángulos (RSL)

- Sitúate en la carpeta siguiente:

`$HOME/evidencias/Grid/03_Manejo_Datos/P3`

- Todos los ficheros que generes en esta práctica debes de guardarlo en esta carpeta.

Práctica 3: Cálculo del Número Pi Mediante una Aproximación por Rectángulos (RSL)

Ejercicio

- Calcula el número Pi mediante una aproximación por Rectángulos definiendo un trabajo Grid RSL (Ejercicio.rsl) que:
 - Compile de forma Remota
 - Ponga permisos de ejecución al ejecutable
 - Retorne el resultado en un fichero llamado miStdOut, miStdError donde se redirija la salida estándar y de Error.

Práctica 1: Lanzamiento de Trabajos Grid RSL (GASS Server) Evidencias



- En la carpeta *\$HOME/evidencias/Grid/03_Manejo_Datos/P3* debe de aparecer los ficheros:

ejercicio.rsl → Trabajo rsl del ejercicio 1.

miStdOut → Fichero de Salida Estándar del ejercicio 1

Práctica 4: Lanzamiento de Trabajos Grid (GridFTP)

Práctica 4: Lanzamiento de Trabajos Grid (GridFTP)

- Sitúate en la carpeta siguiente:

`$HOME/evidencias/Grid/03_Manejo_Datos/P4`

- Todos los ficheros que generes en esta práctica debes de guardarlo en esta carpeta.

Práctica 4: Lanzamiento de Trabajos Grid

(GridFTP)

Ejercicio 1

- En este ejercicio se quiere lanzar un trabajo cuyos argumentos de entrada son ficheros, en este caso se quiere evitar el uso de staging y utilizar los comandos grid-ftp para la transferencia de ficheros.
- El trabajo a lanzar es sort, y pretende ordenar dos ficheros "Entrada1.txt y Entrada2.txt" y retornarlo en otro fichero llamado "Salida.txt".
- Para ello los pasos a seguir son los siguientes:
 - Paso 1.** Crear dos ficheros Locales llamados "Entrada1.txt" y "entrada2.txt" y escribe dentro de ellos unos cuantas palabras que quieras ordenar.
 - Paso 2.** Crea un certificado Proxy con tus credenciales de Usuario.
 - Paso 3.** Transferir los ficheros al recurso remoto a través de los comandos globus-url-copy.


```
globus-url-copy file:///home/ccgc/Evidencias/Grid/03_Manejo_Datos/P4/Entrada1.txt gsiftp://XXX.ccgc.mastercpd.upv.es/tmp/Entrada1.txt
```

```
globus-url-copy file:///home/ccgc/Evidencias/Grid/03_Manejo_Datos/P4/Entrada2.txt gsiftp://XXX.ccgc.mastercpd.upv.es/tmp/Entrada2.txt
```
 - Paso 4.** Lanzar el Trabajo al equipo remoto de tu compañero con globusrun definiendo un trabajo RSL llamado "ejercicio.rsl".
- Copia en un fichero de texto llamado "comandos.txt" los comandos utilizados en el paso 3 y 4.

Práctica 1: Lanzamiento de Trabajos Grid (GridFTP) Evidencias



- En la carpeta *\$HOME/evidencias/Grid/03_Manejo_Datos/P2* debe de aparecer los ficheros:

ejercicio.rsl → Trabajo rsl del ejercicio.
Entrada1.txt → Fichero de texto con nombres.
Entrada2.txt → Fichero de texto con nombres.
Salida.txt → Fichero resultado del ejercicio.
Comandos.txt → Los comandos utilizados.