



PG3. Manejo de datos en un Entornos Grid

Conceptos de la Computación en Grid y Cloud



Práctica 1: Lanzamiento de Trabajos Grid RSL (GASS Server – standard 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 a 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 fichero 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 (Utiliza otro terminal). 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 máquina 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 del GASS, 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.ccgc.mastercpd.upv.es -f ejercicio1.rsl
```

Paso 6. Comprueba que se te ha generado el Fichero miStdOut en la localización que has indicado.

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

Evidencias – Ejercicio 1



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

ejercicio1.rsl → Trabajo RSL del ejercicio 1.
miStdIn → Fichero de Entrada Estándar del ejercicio 1.
miStdOut → Fichero de Salida Estándar del ejercicio 1.
miStdErr → Fichero de Salida Error del ejercicio 1.

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

- Para esta práctica se deben de organizar Grupos de tres alumnos.
- En este ejercicio se quiere lanzar un trabajo desde vuestro equipo local (alumno 1) a una máquina remota (Equipo Remoto 1 – alumno 2) donde el fichero de entrada estándar se ubique en una segunda máquina remota (Equipo Remoto 2 – alumno 3) y que los resultados se guarden en vuestra máquina local (alumno 1).
- 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.
- Para realizar el ejercicio se debe de lanzar el GASS como no seguro en la máquina que proporciona los datos.

Nota: en este escenario utiliza la opción `-i` cuando arranques el GASS

- 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 Máquina Remota 2 en el directorio `$HOME/evidencias/Grid/03_Manejo_Datos/P1`. (reutiliza el fichero `miStdIn` del ejercicio 1)
- El trabajo RSL que definas, 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 – ejercicio 2



- En la carpeta `$HOME/evidencias/Grid/03_Manejo_Datos/P1/ejercicio2` debe de aparecer los ficheros:

`ejercicio2.rsl` → Trabajo rsl del ejercicio 2.
`miStdIn2` → Fichero de Entrada Estándar del ejercicio 2.
`miStdOut2` → Fichero de Salida Estándar del ejercicio 2.
`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 (Utiliza otro terminal). 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 máquina 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/P2/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 2: Lanzamiento de Trabajos Grid RSL (GASS Server) Evidencias – ejercicio 1



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

`ejercicio1.rsl` → Trabajo rsl del ejercicio 1.
`Entrada1.txt` → Fichero de texto con nombres.
`Entrada2.txt` → Fichero de texto con nombres.
`Salida.txt` → Fichero resultado del ejercicio1.
`miStdOut` → Fichero de Salida Estándar del ejercicio 1.
`miStdErr` → Fichero de Salida Error del ejercicio 1.

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

- Para esta práctica se deben de organizar Grupos de tres alumnos.
- En este ejercicio se quiere lanzar un trabajo desde vuestro equipo local (alumno 1) a una máquina remota (Equipo Remoto 1 – alumno 2) donde el fichero de entrada estándar se ubique en una segunda máquina remota (Equipo Remoto 2 – alumno 3) y que los resultados se guarden en vuestra máquina local (alumno 1).
- Para realizar el ejercicio se debe de lanzar el GASS como no seguro en la máquina que proporciona los datos.
- 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 Máquina 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/ejercicio2` debe de aparecer los ficheros:

ejercicio2.rsl → Trabajo rsl del ejercicio 2.
Entrada1.txt → Fichero de texto con nombres.
Entrada2.txt → Fichero de texto con nombres.
Salida2.txt → Fichero resultado del ejercicio 2.
miStdOut2 → Fichero de Salida Estándar del ejercicio 2.
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:
 - Compila de forma Remota. Utiliza `cpi.c` de PoliformaT.
 - Pon permisos de ejecución al ejecutable.
 - Retorna el resultado en un fichero llamado `miStdOut`, `miStdError` donde se redirija la salida Estándar y de Error.

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

Evidencias



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

ejercicio.rsl	→ Trabajo rsl del ejercicio 1.
Compila_y_ejecuta.sh	→ script que compila y ejecuta cpi
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)

- 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

Nota: para transferir los ficheros deben de estar activado el servicio globus-gridftp-server como root.

sudo /etc/init.d/globus-gridftp-server start

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 4: 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.