

# Globus Middleware

Daniel Cano  
Instituto de Física de Cantabria  
IFCA/CSIC

**GRIDS & e-CIENCIA**

Curso 2005 - IFIC Valencia - SPAIN

14 - 16 de junio de 2005

<http://ific.uv.es/grid/cursogrid/>

## ■ Objetivos

- Conocer globus
- Comprender la cadena básica de envío de trabajos
- Conocimiento inicial de herramientas globus. Quedan como “deberes” conocimientos más avanzados.

## ■ Asumiré

- Conocimientos de GNU/Linux
- Conocimientos de lenguaje de shell script (bash, sh, tcsh, csh ...)

# Proyecto Globus

- Comunidad de individuos y organizaciones que desarrollan tecnologías GRID fundamentales .
- Permite compartir recursos, capacidad de cálculo, datos, etc. sin sacrificar la autonomía.
- Globus Toolkit™: base Open Source sobre la que se construyen la infraestructura y las aplicaciones GRID.
- Desarrollo y promoción de protocolos GRID para permitir la interoperabilidad y compartir las ifaestructuras.
- Desarrollo y promoción de API's para permitir la portabilidad y la compartición de código.

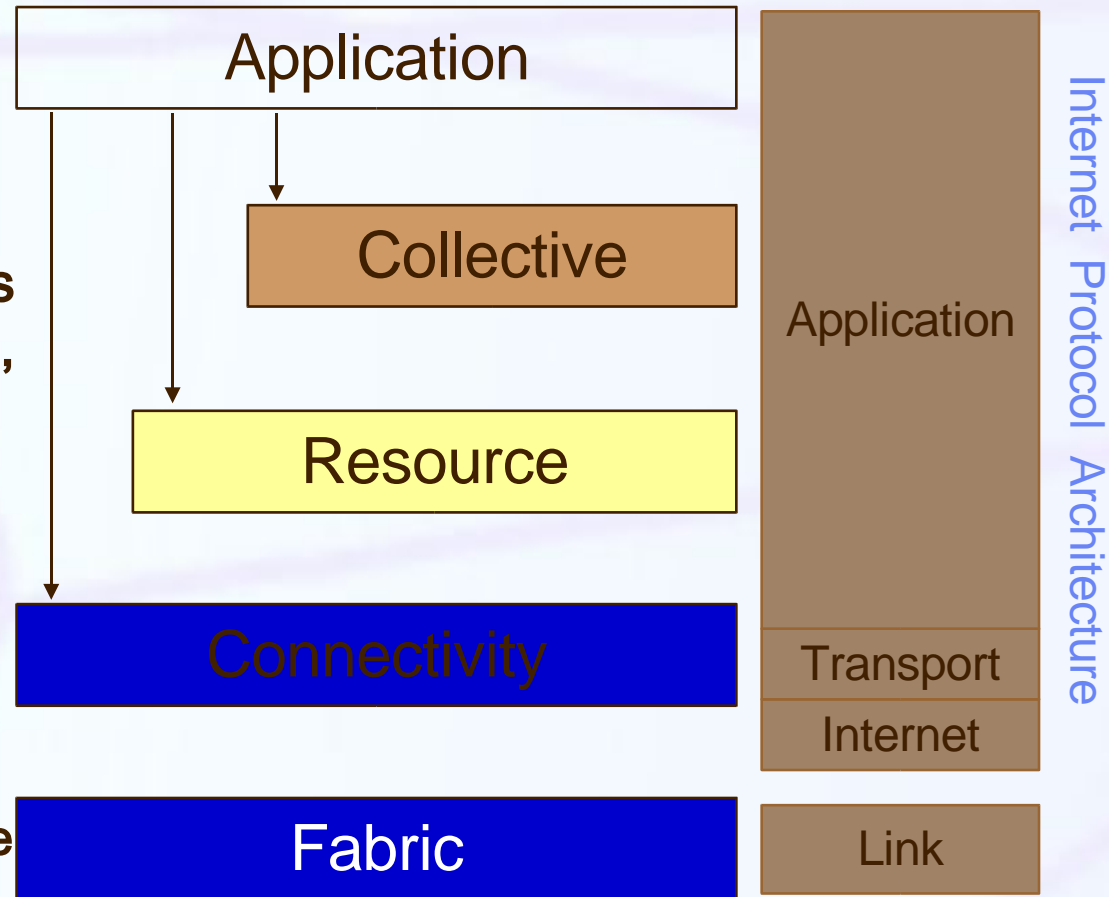
- Conjunto de protocolos para la administración de sistemas distribuidos.
  - Autenticación, delegación, descubrimiento de recursos, monitorización, ejecución, manejo de datos ...
- Implementación Open Source.
  - Gran comunidad internacional de usuarios.
- Permite la utilización de servicios y aplicaciones de más alto nivel.

# Arquitectura Grid

Capacidades de alto nivel  
Globusrun, MPICH-G2

Servicios Grid. Integran los  
componentes de Fabric. GRAM,  
MDS, GSI.

Máquinas, SO, redes, sistemas de  
almacenamiento



# Arquitectura Grid

## Applications



## Application Toolkits



## Grid Services

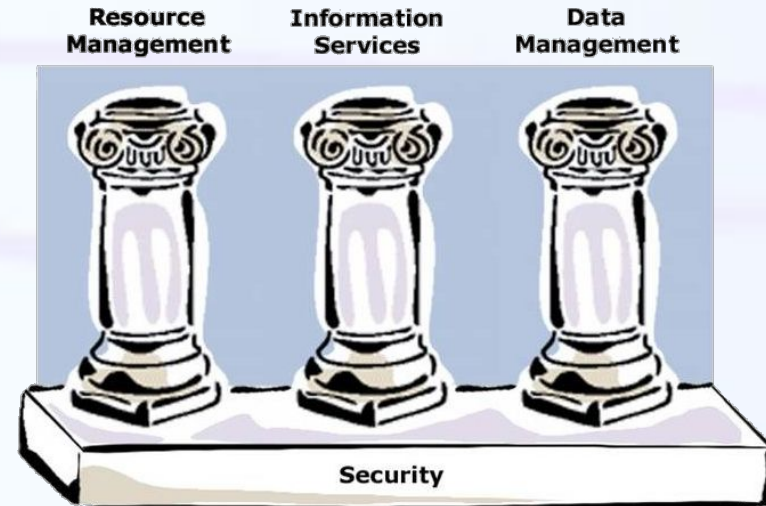
Protocols, authentication, policy, instrumentation, resource management, discovery, events, etc.

## Grid Fabric

Storage, networks, computers, display devices, etc. and their associated local services

# Protocolos Clave

- Globus Toolkit™ v2 (GT2) se centra en 4 protocolos clave.
  - Seguridad. Grid Security Infrastructure (GSI).
  - Recursos.
    - Manejo de recursos. Grid Resource Allocation Management (GRAM).
    - Información. Grid Resource Information Protocol (GRIP/LDAP).
    - Transferencia de datos. Grid File Transfer Protocol (GridFTP).
- También se apoya en una serie de protocolos colectivos
  - Monitorización y descubrimiento, replicación ...

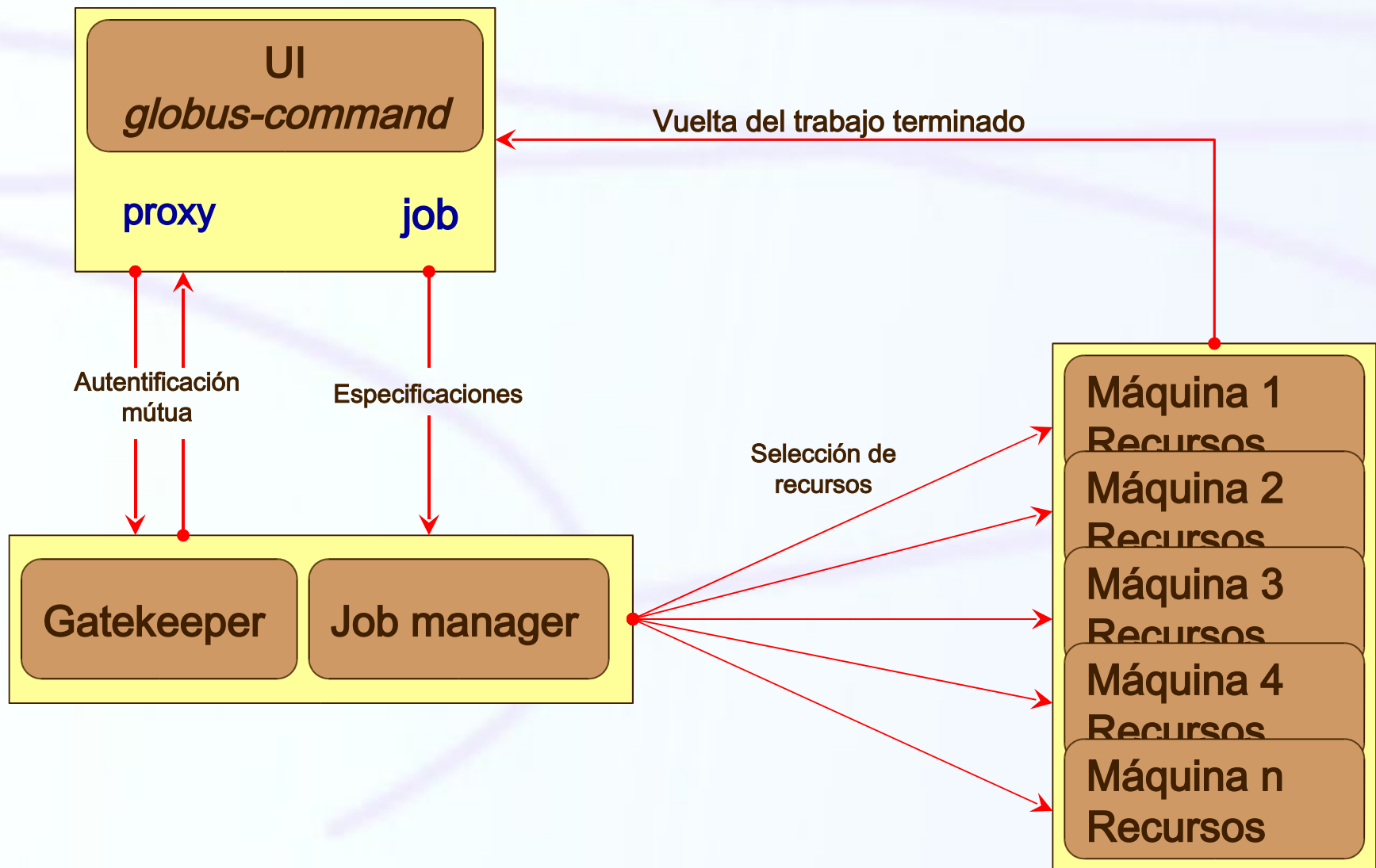


- GRAM: Globus Resource Allocation Manager.
  - Librería básica
  - Envío remoto de trabajos
  - Unificador de máquinas grid
    - User Interface
  - Ubicuo

- MDS: Monitoring and Discovering Service.
  - También GIS: Grid Information Service
    - Número de CPUs
    - Memoria
    - Ancho de banda
    - Almacenamiento
  - LDAP: Lightweight Directory Access Protocol
    - Información del middleware en una interfaz común
    - Equipos dispares → imagen común

- GSI: librería gss-api para añadir autenticación
  - grid-proxy-init
  - Independiente de las políticas de seguridad de cada centro.
  - Independiente del hardware
  - Simplificación de múltiples accesos remotos

# Cadena de un trabajo



# En una página

- User Interface
- Cuenta, certificado y proxy
- `globusrun -a -r <hostname>`
- `globus-job-run <hostname> /bin/echo "Hello World"`
- `<globus command> -help`

# Requisitos para la utilización del Testbed

- Cuenta en una máquina UI
- Certificado válido y no caducado
  - .globus:
    - *usercert.pem: Permiso de lectura para todos los usuarios (644)*
    - *userkey.pem: Permiso de lectura para el usuario (400)*
- Organización Virtual (VO)
  - CMS, ATLAS, DTEAM, BIOMED, LHCB, .....

# Globus Proxy

- El acceso al Testbed se controla via proxy.
- proxy: Credencial de tiempo limitado firmado por la clave privada del usuario.
  - Comandos básicos:
    - **grid-proxy-init** : obtención del proxy, se requiere la clave.
      - ✓ -valid <hh:mm>: cambia la duración del proxy.
      - ✓ -cert <\${PATH\_TO\_USERCERT}>: en caso de que usercert.pem no se encuentre en el lugar por defecto .
      - ✓ -key <\${PATH\_TO\_USERKEY}>: en caso de que userkey.pem no se encuentre en el lugar por defecto.
    - **grid-proxy-info**: da toda la información relevante del proxy existente
    - **grid-proxy-destroy**: borra el proxy de esa sesión
    - **grid-proxy-xxx -help**: muestra como usar el comando *grid-proxy-xxx*

# Globus Proxy

## Ejemplos:

- grid-proxy-init

```
[cano@egeeui cano]$ grid-proxy-init -valid 24:00
Your identity: /C=ES/O=DATAGRID-ES/O=IFCA/CN=Daniel Cano Fernandez
Enter GRID pass phrase for this identity:
Creating proxy ..... Done
Your proxy is valid until: Mon Jun 6 19:09:01 2005
```

- grid-proxy-info

```
[cano@egeeui cano]$ grid-proxy-info
subject : /C=ES/O=DATAGRID-ES/O=IFCA/CN=Daniel Cano Fernandez/CN=proxy
issuer  : /C=ES/O=DATAGRID-ES/O=IFCA/CN=Daniel Cano Fernandez
identity : /C=ES/O=DATAGRID-ES/O=IFCA/CN=Daniel Cano Fernandez
type    : full legacy globus proxy
strength : 512 bits
path    : /tmp/x509up_u501
timeleft : 23:57:07
```

- grid-proxy-destroy

```
[cano@egeeui cano]$ grid-proxy-destroy
```

# Obteniendo información

- De momento:

```
lcg-infosites -vo <vo_name> [options]
```

- **Varios modos**
  - globus-job-run
  - globus-job-submit
  - globusrun
  - mpirun
  - globus-sh-exec

# globus-job-run

- Forma más básica de enviar trabajos
- Corre en foreground y la salida es enviada al terminal.

`globus-job-run <hostname> </path/executable> <arguments>`

- Ejercicio 1. Enviar un “Hello World”

# globus-job-run con ficheros

- Ejercicio 2. Enviar un trabajo con dos número como argumentos de entrada y que devuelva el nombre de la máquina, los argumentos y las suma de ambos.

ej2\_file1.sh

- ¿Funciona?

# globus-job-run con ficheros

- No funciona, ¿por qué?

Pista: ejecutar

*globus-job-run máquina /bin/pwd*

# globus-job-run con ficheros

- El script no está en la máquina donde se va a ejecutar.
- Solución: añadir la opción `-stage`
- Modificaremos el script para ver que hace esta opción

`ej2_file2.sh`

- Otra solución: *globus-url-copy*

# globus-url-copy

- Permite copiar hacia/desde máquinas remotas
- Sintaxis básica

```
globus-url-copy file:$PATH_TO_FILE/filename gsiftp://hostname/$PATH_TO_FILE/filename
```

O bien

```
globus-url-copy gsiftp://hostname/$PATH_TO_FILE/filename file:$PATH_TO_FILE/filename
```

- Ejercicio 3:
  - Correr el script con `globus-job-run` sin usar `-stage`

# globus-job-submit

- Utilizar colas remotas (PBS, LSF ...).
- Corre en background.
- la salida se escribe en la máquina en la que se ejecutó.
- globus-job-status devuelve el estado del trabajo.
- globus-job-get-output recoge la salida.
- globus-job-clean limpia la salida.
- globus-job-cancel cancela el trabajo

# globusrun

- “Ejecutables” escritos en RSL, Globus Resource Specification Language.
- Trabajos en foreground o background.
- globus-job-run y globus-job-submit son scripts que usan globusrun.

# Algunas relaciones RSL comunes

Relación	Ejemplo	Significado
(executable=string)	(executable="/bin/echo")	Programa ejecutable
(arguments=list)	(arguments= Hola)	Argumentos para el programa
(directory=string)	(directory="/tmp")	Directorio activo
(environment=list)	(environment= (PATH "/opt/bin")	Variables de entorno para el job
(stdin=string) (stdout=string) (stderr=string)	(stdin=/tmp/cmds)	Donde deben ir las variables stdin, stdout y stderr
(count=integer)	(count=4)	Número de procesos, por defecto 1

# Algunas relaciones RSL comunes

Relación	Ejemplo	Significado
(hostCount=integer)	(hostCount=6)	Número de CPU's en las que correr el proceso
(maxCpuTime=integer)	(maxCpuTime=10)	Máximo tiempo de CPU en minutos
(maxMemory=integer)	(maxMemory=100)	Máxima memoria por proceso en MB
(minMemory=integer)	(minMemory=15)	Mínima memoria por proceso en MB
(jobType=string)	(jobType=condor)	Como iniciar el proceso

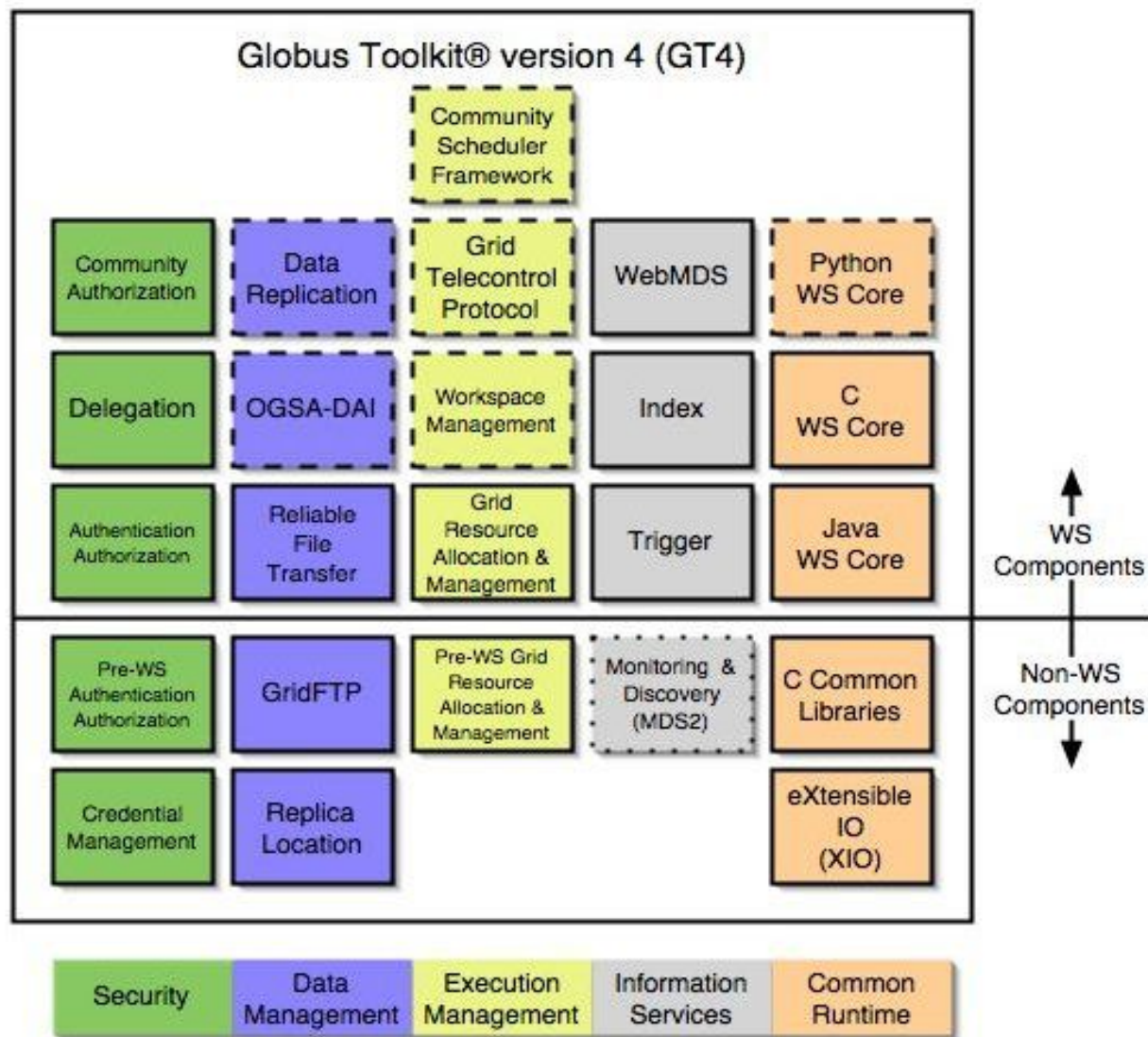
- Ejercicio 4. Usar globusrun para enviar un Hello World

ej4.rsl

- Ejercicio 5. Correr el ejercicio 2, usando globusrun  
ej5.rsl

- MPICH-G2 es una implementación del estándar MPI (Message Parsing Interface), basado en globus. Tiene el comando mpirun que puede usarse para enviar trabajos MPI paralelos a varias máquinas de un grid.
- globus-sh-exec permite correr un script de bash shell en una máquina sin necesidad de preocuparse por paths. Completamente en desuso

# Componentes GT4



# Scripts ejemplo

ej2\_file1.sh

---

```
#!/bin/sh -f
```

```
echo "Hola desde"; $GLOBUS_LOCATION/bin/globus-hostname
```

```
echo arg 1 = $1
```

```
echo arg 2 = $2
```

```
echo -n "la suma es "
```

```
echo "$1+$2" | /usr/bin/bc -l
```

---

[Volver](#)

# Scripts ejemplo

ej2\_file2.sh

---

```
#!/bin/sh -f
```

```
echo "Hola desde"; $GLOBUS_LOCATION/bin/globus-hostname
```

```
echo arg 1 = $1
```

```
echo arg 2 = $2
```

```
echo -n "la suma es "
```

```
echo "$1+$2" | /usr/bin/bc -l
```

```
echo stage path = $0
```

---

[Volver](#)

# Scripts ejemplo

ej4.rsl

---

& (count=1)

(executable=/bin/echo)

(arguments="Hello World")

---

[Volver](#)

# Scripts ejemplo

ej5.rsl

---

```
&(executable=$(GLOBUSRUN_GASS_URL) # "/home/valencia20/Dani/ej2_file1.sh")  
(arguments= "4" "5")
```

---

[Volver](#)