

# Silvia Delgado Olabarriaga

# Introduction to WS-PGRADE

## GUSE GRID AND CLOUD SCIENCE GATEWAY FRAMEWORK



Prepared by Silvia Delgado Olabarriaga

Academic Medical Center of the University of Amsterdam, NL



**SURF SARA**

# Examples of executables (scripts)

- ❑ No ports
- ❑ Command-line arg
- ❑ Input and output files
- ❑ Fixed names: **inputFile.txt**, **outputFile.txt**

hello-script.sh

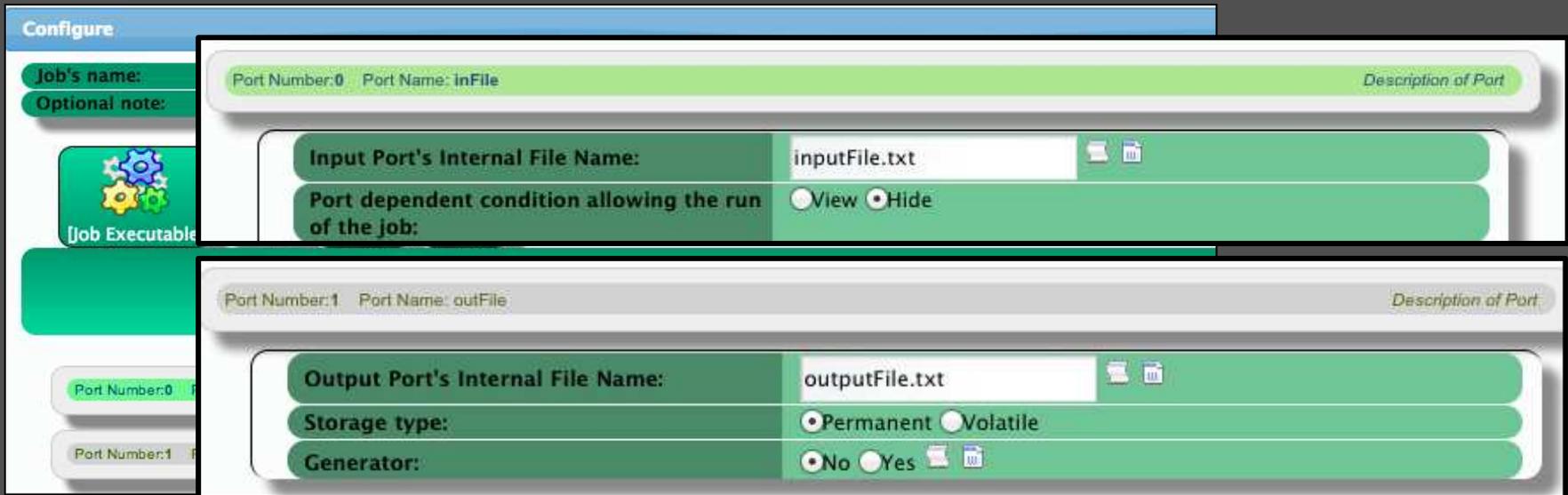
```
#!/bin/bash  
  
echo Hello $1  
echo Today is `date`  
echo This computer is `hostname`
```

copy-script-ports.sh

```
#!/bin/bash  
  
IFile="inputFile.txt"  
OFile="outputFile.txt"  
  
echo -e "start copy (`date`, on `hostname`):\n-----\n" >  
$OFile  
cat $IFile >> $OFile  
echo -e "\n-----\nend $1 (`date`)" >> $OFile
```

# Configure concrete workflow: Job ports internal names

- ❑ Click on **Job I/O**
- ❑ Expand ports
- ❑ Set port names as in the script
- ❑ 



The screenshot shows a 'Configure' window with a sidebar on the left containing 'Job's name:', 'Optional note:', and 'Job Executable' (with a gear icon). The main area displays two port configuration panels. The top panel is for 'Port Number:0' with 'Port Name: inFile'. It includes a text input for 'Input Port's Internal File Name' containing 'inputFile.txt' and radio buttons for 'Port dependent condition allowing the run of the job' set to 'View'. The bottom panel is for 'Port Number:1' with 'Port Name: outFile'. It includes a text input for 'Output Port's Internal File Name' containing 'outputFile.txt', radio buttons for 'Storage type' set to 'Permanent', and radio buttons for 'Generator' set to 'No'.

e.g.: for [copy-script-ports.sh](#)

# Configure concrete workflow: Job input port value

- ❑ Click on **Job I/O**
- ❑ Select: **Source of input directed to this port**



Input Port's Internal File Name:	inFile.txt
Port dependent condition allowing the run of the job:	<input type="radio"/> View <input checked="" type="radio"/> Hide
Source of input directed to this port:	 Recently defined External File Name: N/A (Standard name paramInputs.zip – to be regard as PS container.) <input type="button" value="Browse..."/> No file selected.
Parametric Input details:	<input type="radio"/> View <input checked="" type="radio"/> Hide

Upload file

Input Port's Internal File Name:	inFile.txt
Port dependent condition allowing the run of the job:	<input type="radio"/> View <input checked="" type="radio"/> Hide
Source of input directed to this port:	 <input type="text"/>
Parametric Input details:	<input type="radio"/> View <input checked="" type="radio"/> Hide

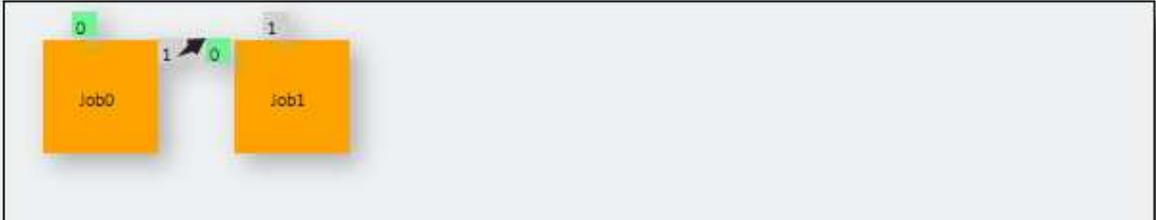
Provide value

# Save concrete workflow configuration

Concrete

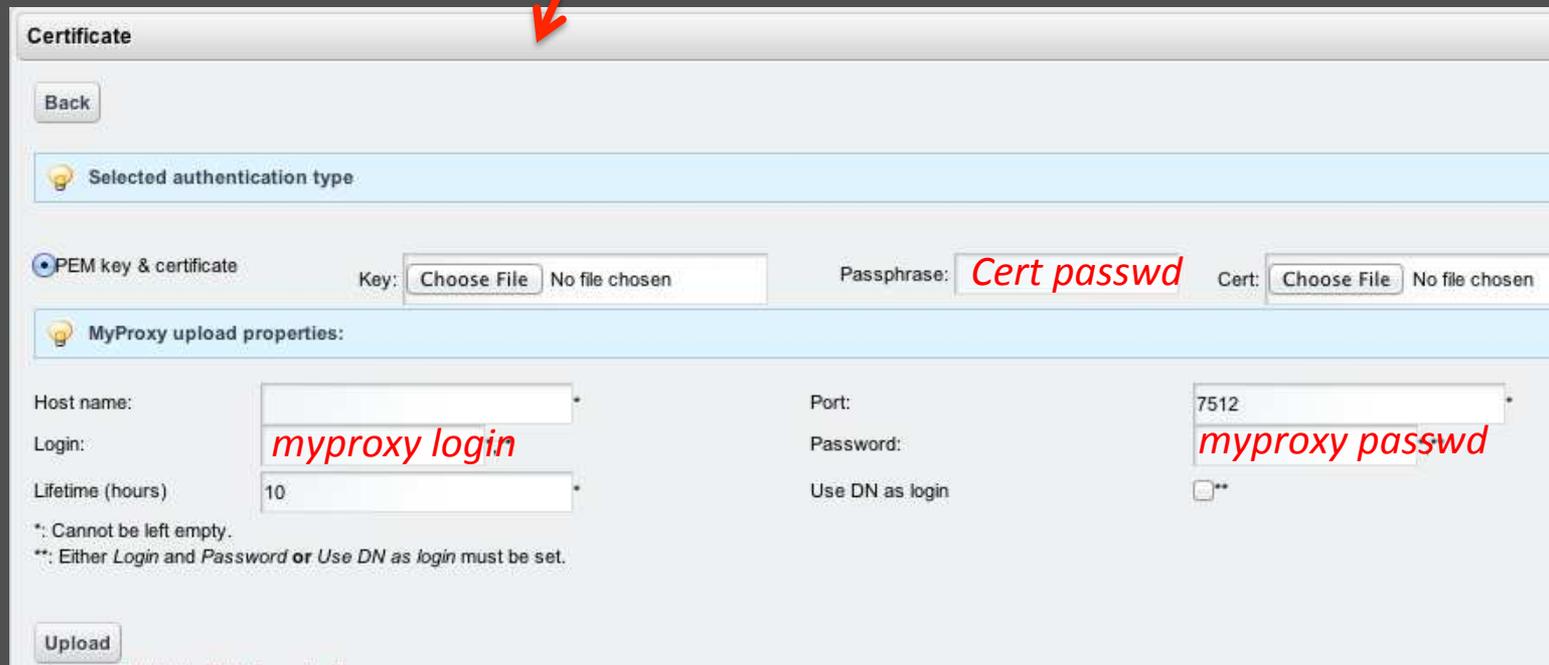
  Delete old instances   Do not delete old instances

Workflow name:	TwoJobsCopyJava
Note:	2013-10-11
Workflow Graph:	TwoJobsCopy <input type="button" value="Edit"/> <input data-bbox="807 806 1284 842" type="button" value="Fit copied workflow to a new graph"/>
Workflow Template:	--



# Grid Security: Upload to myproxy server

- ❑ Security -> Certificate
- ❑ Upload
- ❑ Select PEM key & certificate
- ❑ Fill fields
- ❑ Upload



Certificate

Back

Selected authentication type

• PEM key & certificate

Key:  No file chosen

Passphrase:

Cert:  No file chosen

MyProxy upload properties:

Host name:

Login:

Lifetime (hours):

Port:

Password:

Use DN as login:

\*: Cannot be left empty.  
\*\*: Either Login and Password or Use DN as login must be set.

Upload

# Grid security: Download proxy

❑ Security->Certificate

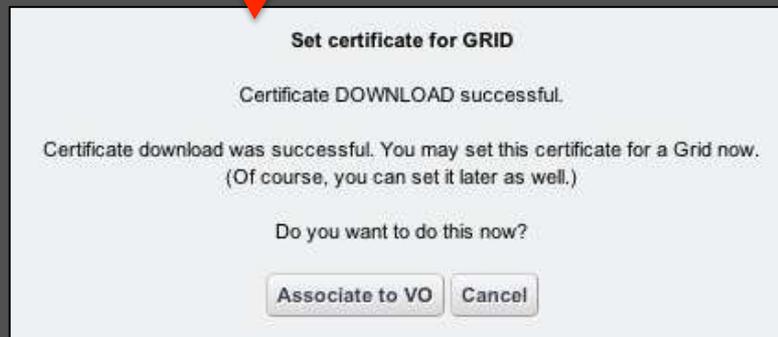
❑ Download

❑ Fill fields

❑ Download

❑ Associate VO

❑ Choose VO



# Submit workflow

- ❑ Workflow->Concrete
- ❑ Submit

Concrete						
						
<input type="button" value="Refresh"/>						
Workflow name	Running	Finished	Error	Suspended	Actions	
<b>TestSQL</b> 2013-10-14	0	1	0	0	<input type="button" value="Configure"/>	<input type="button" value="Info"/> <input type="button" value="Details"/> <input type="button" value="Submit"/> <input type="button" value="Delete"/> <input type="button" value="Export"/>
<b>TwoJobsCopyJava</b> 2013-10-11	0	0	0	0	<input type="button" value="Configure"/>	<input type="button" value="Info"/> <input type="button" value="Details"/> <input type="button" value="Submit"/> <input type="button" value="Delete"/> <input type="button" value="Export"/>
<b>concrete-cesup</b> 2013-3-1	0	4	1	0	<input type="button" value="Configure"/>	<input type="button" value="Info"/> <input type="button" value="Details"/> <input type="button" value="Submit"/> <input type="button" value="Delete"/> <input type="button" value="Export"/>
<b>hello-deaf-mute</b> 2013-1-9	0	7	6	0	<input type="button" value="Configure"/>	<input type="button" value="Info"/> <input type="button" value="Details"/> <input type="button" value="Submit"/> <input type="button" value="Delete"/> <input type="button" value="Export"/>

# Monitor workflow execution

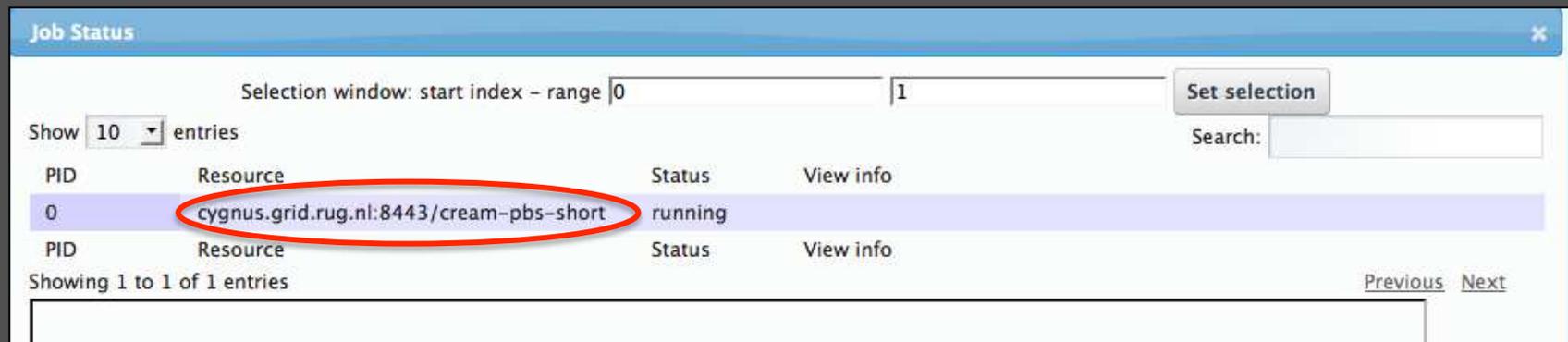
- ❑ Workflow->Concrete
- ❑ Details
- ❑ Refresh, Details (again)

The screenshot shows the Concrete workflow monitoring interface. At the top, there is a 'Concrete' header and a help icon. Below the header, there are 'Back' and 'Refresh' buttons. The 'Refresh' button is circled in red. The main content area displays workflow details for 'TestCopyFile.JavaGrid' with a note of '2013-10-16' and a workflow graph of 'one-component-input-outputFile'. Below this, there is a table of workflow instances. The first instance, dated '2013-10-16 11:16', is in a 'running' state and has a 'Details' button circled in red, along with 'Suspend' and 'Cost of Instance' buttons. The second instance, dated '2013-10-16 11:8', is also in a 'running' state and has 'Details', 'Suspend', and 'Cost of Instance' buttons. Below the table, there is a section for 'Selected WF Instance:' with the date '2013-10-16 11:16'. At the bottom, there is a table with columns 'Job', 'Status', 'Instances', and '[ Actions ]'. The 'Job0' row shows a 'running' status, '1' instance, and a 'View running' button.

Job	Status	Instances	[ Actions ]
Job0	running	1	View running

# Monitor workflow execution (Running)

- ❑ Workflow->Concrete
- ❑ Details
- ❑ Details (again)
- ❑ View Running



The screenshot shows a 'Job Status' window with a table of job entries. The first entry is highlighted in blue and circled in red. The table has columns for PID, Resource, Status, and View info. The selected entry has PID 0, Resource 'cygnus.grid.rug.nl:8443/cream-pbs-short', and Status 'running'. The window also includes a selection window (start index - range 0 - 1), a 'Set selection' button, a search field, and navigation links for 'Previous' and 'Next'.

PID	Resource	Status	View info
0	cygnus.grid.rug.nl:8443/cream-pbs-short	running	
PID	Resource	Status	View info

Showing 1 to 1 of 1 entries

# Monitor workflow execution

- ❑ Workflow->Concrete
- ❑ Details
- ❑ Details (again)
- ❑ View Finished

The screenshot displays a web interface for monitoring workflow execution. At the top, there are 'Back' and 'Refresh' buttons. Below them, the following information is shown:

- Workflow name: TestCopyFileJavaCluster
- Note: 2013-10-16
- Workflow Graph: one-component-input-outputFile
- Workflow Template:

A table lists workflow instances:

Time	Status	Actions
2013-10-16 10:56	finished	Details, Delete, Cost of Instance
2013-10-16 11:1	finished	Details, Delete, Cost of Instance

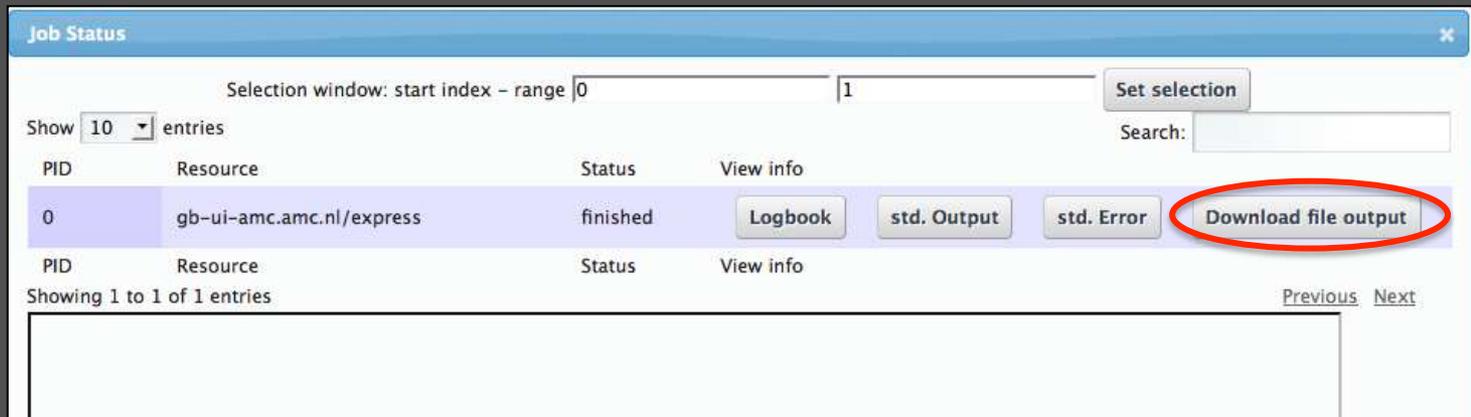
The 'Details' button for the 11:1 instance is circled in red. Below the table, the 'Selected WF Instance:' is identified as '2013-10-16 11:1'.

A summary table at the bottom shows the overall status:

Job	Status	Instances	[ Actions ]
Job0	finished	1	View finished

# Get workflow results

- ❑ Workflow->Concrete
- ❑ Details
- ❑ Details (again)
- ❑ View Finished
  - Logbook, std.Output, std.Error
- ❑ Download file output



The screenshot shows a 'Job Status' window with a table of workflow entries. The first entry is highlighted in blue and has several action buttons next to it. The 'Download file output' button is circled in red.

PID	Resource	Status	View info
0	gb-ui-amc.amc.nl/express	finished	<a href="#">Logbook</a> <a href="#">std. Output</a> <a href="#">std. Error</a> <a href="#">Download file output</a>

Showing 1 to 1 of 1 entries

# Beyond the basics

- ❑ Running jobs on the grid
- ❑ Connecting ports of different jobs
- ❑ Handling big data
- ❑ Parameter sweeps

# Configure concrete workflow: Job executable (on grid)

❑ Workflow->Concrete

❑ Configure

❑ Click on Job

❑ Type: glite

❑ Grid: VO

Configure

Job's name: Job0

Optional note: Description of Job

[Job Executable] [Job I/O] [JDL/RSL] [History]

Workflow Service Binary

Type: glite

Grid: vlemed

Replicate settings in all Jobs:

Kind of binary:  Sequential  Java  MPI

MPI Node Number:

Executable code of binary: Recently stored: HelloCopyFile.jar  
Browse... No file selected.

Parameter: inFile outFile

# Configure concrete workflow: Connected ports

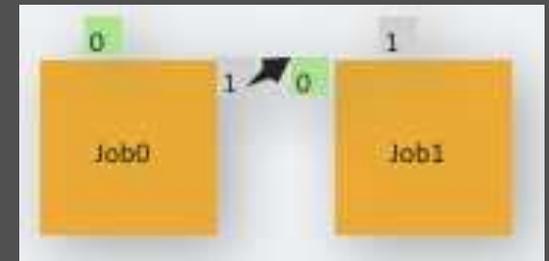
## ❑ Workflow->Concrete, Configure

### ❑ Job0:

- Set port names
- Set input port values

### ❑ Job1:

- Set port names
- (input port value is channeled from Job0)



Port Number:0	Port Name: inFile (channel)	Description of Port
Input Port's Internal File Name: inputFile.txt		
Port dependent condition allowing the run of the job: <input type="radio"/> View <input checked="" type="radio"/> Hide		
Parametric Input details: <input type="radio"/> View <input checked="" type="radio"/> Hide		

Port Number:1	Port Name: outFile	Description of Port
Output Port's Internal File Name: outputFile.txt		
Storage type: <input checked="" type="radio"/> Permanent <input type="radio"/> Volatile		
Generator: <input type="radio"/> No <input checked="" type="radio"/> Yes		

# Handling Big Data

- ❑ Pass files by reference instead of by value (gLite only)
- ❑ Input files:

Input Port's Internal File Name:	inputFile
Port dependent condition allowing the run of the job:	<input type="radio"/> View <input checked="" type="radio"/> Hide
Source of input directed to this port:	 lfn:/grid/vlmed/mahdi/file.txt <input checked="" type="checkbox"/> Copy to WN:
Parametric Input details:	<input type="radio"/> View <input checked="" type="radio"/> Hide

- ❑ Output files

Output Port's Internal File Name:	outputFile
Base of Output Port's Remote File Name:	lfn:/grid/vlmed/ebi
SE, if the definition of the remote file has the prefix 'lfn':	
Storage type:	<input checked="" type="radio"/> Permanent <input type="radio"/> Volatile
Generator:	<input checked="" type="radio"/> No <input type="radio"/> Yes

# Parameter Sweep

## ❑ Run the same job for multiple input values on the same port

- Need to upload various files

## ❑ Make a zip file

- Fixed name

- ✓ paramInputs.zip

- Contents

- ✓ File names inside the zip must be called: 0, 1, 2, ...

- ✓ They will generate jobs and results identified by the same numbers



▼ paramInputs	Today 9:50 AM
0	Mar 27, 2013 5:30 PM
1	Mar 27, 2013 5:30 PM
paramInputs.zip	Mar 27, 2013 5:32 PM

## ❑ Indicate the number of files when configuring the job

# Configure concrete workflow

## Parameter sweep on input port

- ❑ Workflow->Concrete, Configure, Click on Job, select Job I/O
- ❑ Upload zip file
- ❑ View Parametric Input details
- ❑ Set Input numbers (n files in zip)
- ❑ 

paramInputs	Today 9:50 AM
0	Mar 27, 2013 5:30 PM
1	Mar 27, 2013 5:30 PM
paramInputs.zip	Mar 27, 2013 5:32 PM

Port Number:0 Port Name: inFile Description of Port

Input Port's Internal File Name: inputFile.txt

Port dependent condition allowing the run of the job:  View  Hide

Source of input directed to this port:

Recently defined External File Name: N/A  
(Standard name paramInputs.zip – to be regard as PS container )

paramInputs.zip

Parametric Input details:  View  Hide

Dot and Cross PID: 0

Input numbers: 2

# Wrap-up

## ❑ **WS-PGRADE is a workflow management system**

- Wraps executables as jobs to run in various computing resources (local, grid, cluster, ..)
- Transports data to/from the jobs
- Provides facilities to retrieve results, see job logs

## ❑ **It also..**

- Retries failed jobs a number of times
- Allows pausing and resuming (failed) workflows
- Has facilities to split and merge data (generator and collector ports)

# Further information

## □ Documentation map

- <http://www.guse.hu/documentation/how-to>
- <http://www.guse.hu/documentation>

## □ Documentation about WS-PGRADE version used in the course

## □ Publications about WS-PGRADE

- <http://www.guse.hu/about/publications>

# Acknowledgements



SCientific gateway Based User Support  
<http://www.sci-bus.eu/>



Building an European Research community  
through interoperable WorkFLOWS and Data  
<http://www.erflow.eu/>



e-infrastructure



# Thanks!