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

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
```
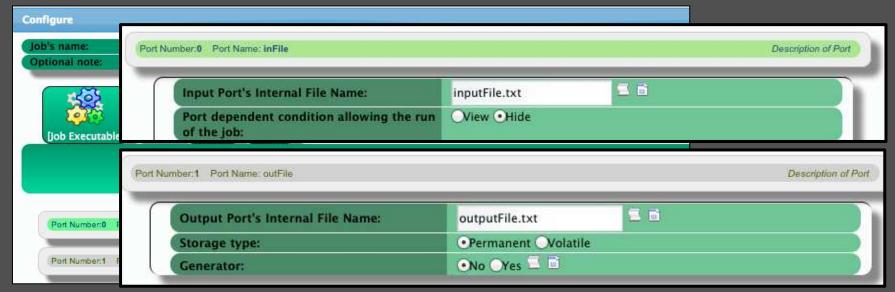
Configure concrete workflow: Job ports internal names

❑ **Click on Job I/O**

❑ **Expand ports**

❑ **Set port names as in the script**

❑ ✅

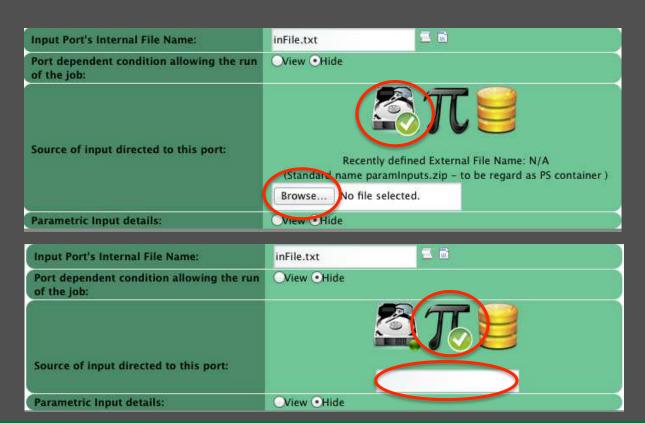e.g.: for copy-script-ports.sh

Configure concrete workflow:
Job input port value

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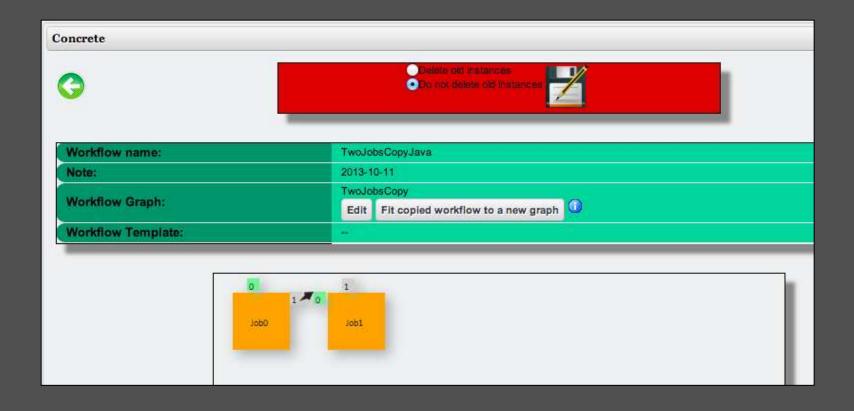
- **Select: Source of input directed to this port**

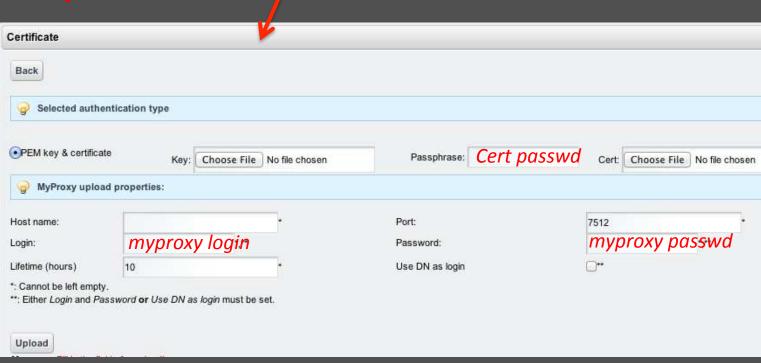
Upload file

Provide value

Save concrete workflow configuration

Grid Security: Upload to myproxy server

❑ **Security -> Certificate**

❑ **Upload**

❑ **Select PEM key & certificate**

❑ **Fill fields**

❑ **Upload**

Certificate

There is no downloaded certificate.

Download — (Dowload certificate from MyProxy server.)
Upload — (Upload authentication data to MyProxy server.)
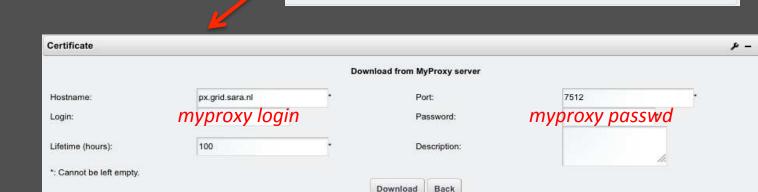Credential Management — (Display information, change MyProxy passphrase, remove a credential from MyProxy server.)
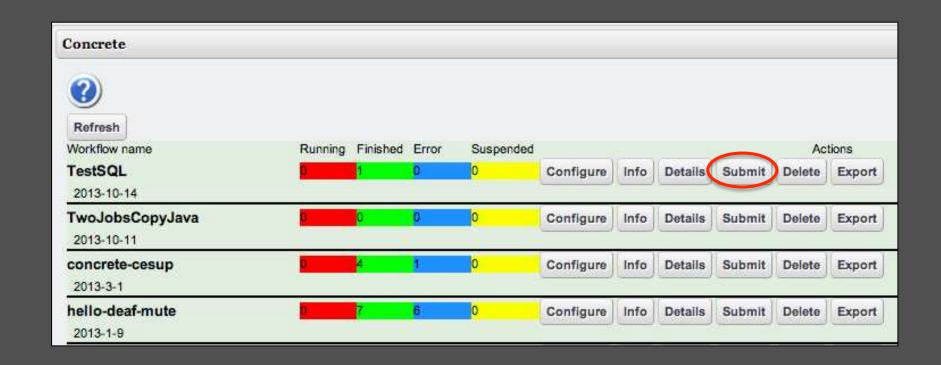
Certificate

Back

💡 Selected authentication type

⦿ PEM key & certificate   Key: [Choose File] No file chosen   Passphrase: *Cert passwd*   Cert: [Choose File] No file chosen

💡 MyProxy upload properties:

Host name:	[ ] *	Port:	7512 *
Login:	*myproxy login*	Password:	*myproxy passwd*
Lifetime (hours):	10 *	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**

Monitor workflow execution

❑ **Workflow->Concrete**

❑ **Details**

❑ **Refresh, Details (again)**

Monitor workflow execution (Running)

- ❑ **Workflow->Concrete**

- ❑ **Details**

- ❑ **Details (again)**

- ❑ **View Running**

Monitor workflow execution

- **Workflow->Concrete**

- **Details**

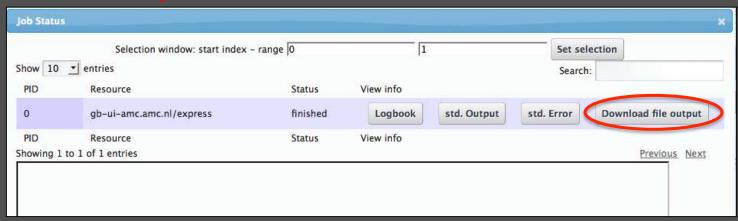- **Details (again)**

- **View Finished**

Get workflow results

- **Workflow->Concrete**

- **Details**

- **Details (again)**

- **View Finished**
  - Logbook, std.Output, std.Error

- **Download file output**

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**

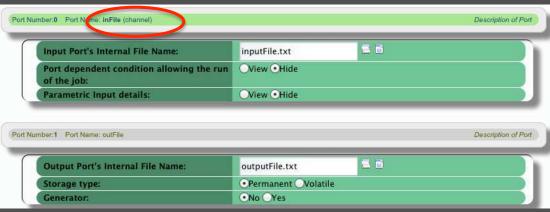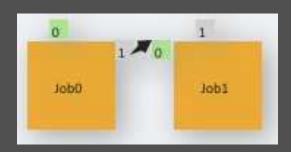- ❑ **Type: glite**

- ❑ **Grid: VO**

Configure 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)

Handling Big Data

❑ **Pass files by reference instead of by value (gLite only)**

❑ **Input files:**

❑ **Output files**

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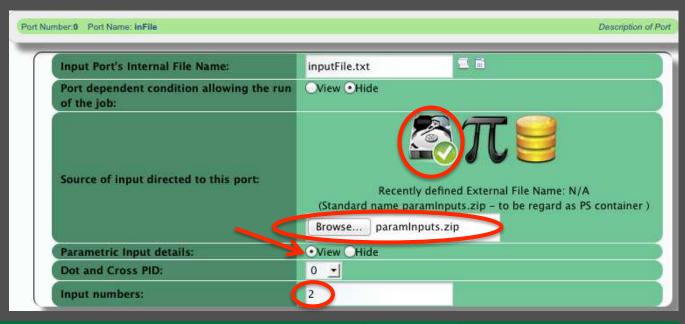
❑ **Indicate the number of files when configuring the job**

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)**

❏

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 interoprable WorkFLOWs and Data
http://www.erflow.eu/

e-infrastructure

SEVENTH FRAMEWORK PROGRAMME

Thanks!