

Porting application to Grid

Viet D. Tran

Institute of Informatics, Slovakia

1. Preparation

- Preparation of application
- Register to Grid

2. Porting application

- First job
- Full computation

3. Advanced Grid application

- Portal
- Data management

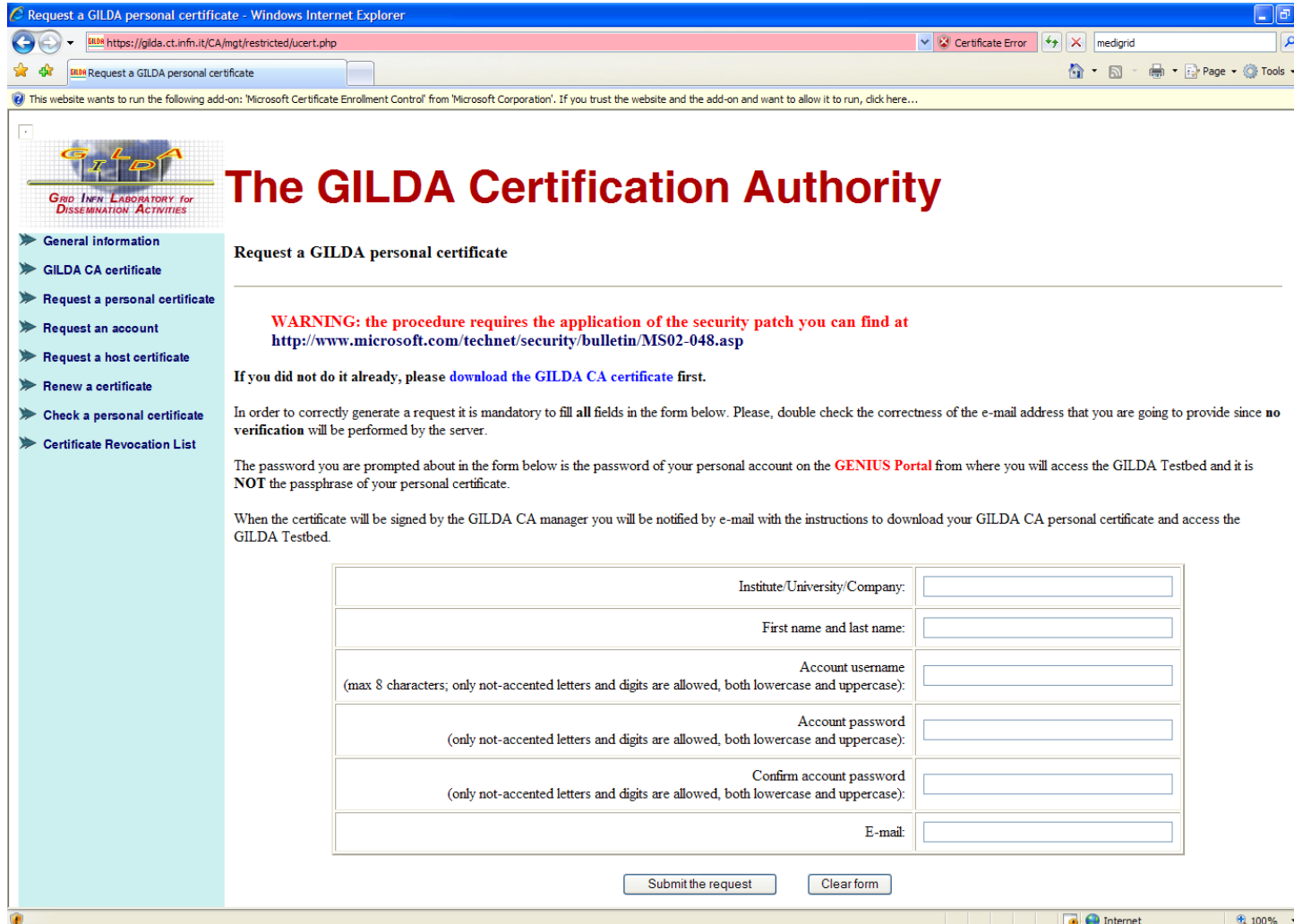
4. Conclusion remarks

Preparation

- **Application should run in batch mode:**
 - Separate computations from graphical interfaces
 - Input and output data from/to files
- **Choosing methods of parallelization**
 - Multiple independent executions of a sequential program with different input data
 - Typical for Monte-Carlo algorithms, parametric study, searching, data conversion
 - Most suitable for Grid computing
 - MPI parallel program with communication during execution
 - Typical for linear algebra, finite elements/differences, genetic algorithms
 - Combination of both: multiple independent executions of parallel programs
- **Porting application to Linux**
 - Compile, link and test the program on Linux machines
 - For MPI parallel program, testing on Linux cluster
 - Non-standard libraries should be linked statically

- **Choosing the infrastructure for porting**
 - GILDA testing infrastructure
 - Simple registration, available for every ones
 - Suitable for testing, learning and demonstration
 - Should not be used for stress testing (too long or too many jobs), especially during some courses or demonstrations
 - EGEE production infrastructure
 - Suitable for running application in production

- **GILDA main web page:** <https://gilda.ct.infn.it/> with many tutorials, documentations and training materials
- **Register and get a GILDA certificate**
<https://gilda.ct.infn.it/CA/>
- **Get an account on User Interface (UI) machines**
 - GILDA automatically create an account on `glite*.gilda.infn.it` during registration but the UI machines have firewalls for SSH
 - Contact with us viet.ui@savba.sk to have an account on our UI `dgt03.ui.sav.sk`.
- **Start to learn grid commands and test them**



Request a GILDA personal certificate - Windows Internet Explorer

Request a GILDA personal certificate

This website wants to run the following add-on: 'Microsoft Certificate Enrollment Control' from 'Microsoft Corporation'. If you trust the website and the add-on and want to allow it to run, click here...

The GILDA Certification Authority

Request a GILDA personal certificate

WARNING: the procedure requires the application of the security patch you can find at <http://www.microsoft.com/technet/security/bulletin/MS02-048.asp>

If you did not do it already, please [download the GILDA CA certificate](#) first.

In order to correctly generate a request it is mandatory to fill **all** fields in the form below. Please, double check the correctness of the e-mail address that you are going to provide since **no verification** will be performed by the server.

The password you are prompted about in the form below is the password of your personal account on the **GENIUS Portal** from where you will access the GILDA Testbed and it is **NOT** the passphrase of your personal certificate.

When the certificate will be signed by the GILDA CA manager you will be notified by e-mail with the instructions to download your GILDA CA personal certificate and access the GILDA Testbed.

Institute/University/Company:	<input type="text"/>
First name and last name:	<input type="text"/>
Account username (max 8 characters; only not-accented letters and digits are allowed, both lowercase and uppercase):	<input type="text"/>
Account password (only not-accented letters and digits are allowed, both lowercase and uppercase):	<input type="password"/>
Confirm account password (only not-accented letters and digits are allowed, both lowercase and uppercase):	<input type="password"/>
E-mail:	<input type="text"/>

Submit the request Clear form

- **Register and get an EGEE certificate**
 - Contact with Miroslav Dobrucky (dobrucky.ui@savba.sk)
- **Join to a Virtual Organization (VO)**
 - If your application belong to areas of some existing VO (e.g. biomedicine, chemistry, earth science, ...) you can contact with the VO manager and ask them to add you to his VO
 - Send us short description of your application, we can give contact of the corresponding VO manager
 - If your application does not belong to any VO, you can still join the generic VOCE VO (Central European VO)
<https://voce-register.farm.particle.cz/voce>

Porting application to Grid

- **Copy your executable program (myprogram.bin) and input data (myinput.dat) to User Interface (UI) machine**
- **Log in to the UI (ssh dgt03.ui.sav.sk)**
- **Create a simple JDL file (myfirstjob.jdl)**

```
Executable= "myprogram.bin";  
Arguments= "myinput.dat";  
InputSandbox = {"myprogram.bin", "myinput.dat"};  
StdOutput = "std.out";  
StdError = "std.err";  
OutputSandbox = {"myoutput.dat", "std.out", "std.err"};
```
- **Create proxy certificate:** voms-proxy-init
- **And submit the program:** glite-job-submit myfirstjob.jdl
- **That is applicable only for smaller data and binary (< 2MB)**

- If you have large data and binary, first upload them to storage elements
- Then create a simple script to download the binary, input data and upload the output data (myscript.sh)

```
globus-url-copy gsiftp://storage.ui.sav.sk/data/myprogram.bin \  
                file://$PWD/myprogram.bin
```

```
chmod a+x myprogram.bin
```

```
globus-url-copy gsiftp://storage.ui.sav.sk/data/myinput.dat \  
                file://$PWD/myinput.dat
```

```
./myprogram.bin myinput.dat
```

```
globus-url-copy file://$PWD/myoutput.dat \  
                gsiftp://storage.ui.sav.skdata/myoutput.dat
```

- Create a JDL to submit a job with the script

```
Executable= "myscript.sh";
```

```
InputSandbox = {"myscript.sh"};
```

- **For parametric study, first create the JDL (param.jdl):**

```
JobType = "Parametric";
```

```
Executable = "myprogram.bin";
```

```
Arguments = "myinput_PARAM_.dat";
```

```
StdOutput = "std_PARAM_.out";
```

```
StdError = "std_PARAM_.err";
```

```
Parameters = 100;
```

```
ParameterStart = 1;
```

```
ParameterStep = 1;
```

```
InputSandbox = {"myprogram.bin", "myinput_PARAM_.dat";
```

```
OutputSandbox = {"myoutput_PARAM.dat", std_PARAM_.out",  
"std_PARAM_.err"};
```

- **Then submit the JDL to Grid:** `glite-job-submit param.jdl`

- **For MPI jobs, just add to the JDL**

```
JobType = "MPICH";
```

```
NodeNumber = 16;
```

```
Executable= "myprogram.bin";
```

```
Arguments= "myinput.dat";
```

```
InputSandbox = {"myprogram.bin", "myinput.dat"};
```

```
StdOutput = "std.out";
```

```
StdError = "std.err";
```

```
OutputSandbox = {"myoutput.dat", "std.out", "std.err"};
```

Advanced Grid applications

- **Command line interface** (glite-job-* commands) **is powerful but not user-friendly**
- **Portal is useful because:**
 - Accessible from any where
 - Hide complexity: the actual users do not have to learn about gLite or similar things
 - More secure
 - Useful for demonstration, can be combined with visualization
- **Recommended Grid portal framework: Gridsphere**
 - <http://www.gridisphere.org/>
 - Supporting portlet (JSR168)
 - Easy to use and extend
- **Ready to use portlets:**
 - Certificate management
 - Job submission
 - Data browser

New Credential

This credential can be retrieved from myproxy.gridlab.org.

User Name: (Your credential repository username)

Credential Name: (Name you assigned to credential in repository)

Credential Label: (Label to display for credential in portlet)

Passphrase: (Your credential repository password)

Use Portal Credential: (Leave checked to use the portal credential)

Single Sign-On: (Leave checked to sign-on to the portal)

File Browser Portlet

Physical Files Logical Files

File Browser 1

<Select Resource>

Portal
Payote
Helix
Venus
Skirit
FSO
NO
Rage1
Litchi
HitCross

Path: /mnt/shared/people/glab047

File Browser 2

<Select Resource>

Portal
Payote
Helix
Venus
Skirit
FSO
NO
Rage1
Litchi
HitCross

Path: /home/glab047

Job creation

Select template:

Job name:

Factory URI:

Job configuration

daveFOutputFile:

visualizationOutputFile:

User job list

Job: **DaveF 2D visualization** State: CREATED Creation date: 2006-08-23 15:43:35.472

Job configuration

daveFOutputFile: davef-out-2005-06-24

visualizationOutputFile: davef-2d-visualization-2005-06-24

Job: **HSPF experiment 1** State: CREATED Creation date: 2006-08-23 15:43:07.27

Job: **DaveF 2D visualization** State: CREATED Creation date: 2006-08-23 15:42:56.399

Welcome Grid Job services **Data Services** Visualization

Metadata Services VFS Browser Data Resource Explorer

Data Resource Explorer

Target host:

Connected: https://gaia.ui.savba.sk:8443/wsrff/services/medigrid/dataservices/DataResourcesManager

Host browser: All resources #fixed_data#

Content	Attributes
[..]	
[test_data]	
ahoj.meta	View Attributes
ahoj	View Attributes
metatest.txt.meta	View Attributes
ahoj.meta~	View Attributes
metatest.txt	View Attributes
rftInx.tmp	View Attributes
rftInx.tmp.meta	View Attributes
testdata1	View Attributes

Attributes:

Owners	Users	Hosts
/O=MediGrid/O=IISAS/CN=ROOT		
	/O=MediGrid/O=ALGO/CN=Miltos Kokkosoulis RW	
	/C=SK/O=SlovakGrid/O=IISAS/CN=Martin Maliska RW	
	/C=SK/O=SlovakGrid/O=IISAS/CN=Viet Tran R	
	/O=MediGrid/O=IISAS/CN=Branislav Simo RW	
	/O=MediGrid/O=IISAS/CN=Martin Maliska RW	
	/O=MediGrid/O=IISAS/CN=Marek Ciglan RW	

Upload file to current directory

Add new attribute

New identified as:

Read Write

- **Big applications usually also have big data sets**
- **Some data management tools in Grid**
 - Replica management
 - Each piece of data may have multiple copies (files) to improve the data access
 - Can provide some optimization
 - Metadata catalogs
 - Describe the data in Grid
 - Searching, browsing
 - Access to databases (OGSA-DAI)
 - Access to databases (e.g. Oracle, MySQL) from your program running on the Grid

Conclusion remarks

- **How to exploit the full power of the Grid?**
- **Example:**
 - A scientist runs his application on the Grid and gets the speed-up of $N!$
 - Very nice results, but the Grid power is not fully exploited
 - A group of scientists run the application, sometimes simultaneously, and all are satisfied with the speed
 - The computational power of Grid is exploited
 - A group of research institutes run the application, collaborate with each other, coordinate the computations, sharing their data and knowledge, for the common research aim
 - The full power of the Grid technologies (including computation, data management, security, portal, ...) is exploited
 - Example: high energy physics, biomedicine
- **Think about larger scale of your applications:**
 - Larger computation
 - Larger data sets
 - Larger user community

- **Contact to us:**

- Miroslav Dobrucky: dobrucky.ui@savba.sk
- Viera Sipkova: sipkova.ui@savba.sk
- Viet Tran: viet.ui@savba.sk

- **Useful links**

- EGEE project: <http://www.eu-egee.org/>
<http://www.ui.sav.sk/egee/>
- gLite manual:
<http://glite.web.cern.ch/glite/documentation/userguide.asp>
- GILDA testbed: <https://gilda.ct.infn.it>
- VVT information portal: <http://www.ui.sav.sk/vvt/>