

The GridFactory computing system

Introduction and quick-start

Frederik Orellana

Niels Bohr Institute

Copenhagen University

December 2010

1 Introduction

This document will show you how to use GridFactory services for running a few example jobs on a GridFactory cluster. It will also show you how you can contribute your own computers to extend such a cluster.

2 Prerequisites

The setup we'll be using for the following examples is the simplest possible. For more advanced configurations, involving your own credentials, GridFactory server and virtual organizations, see the [documentation section on the GridFactory web site](#).

- We'll assume that have **NO existing grid/x.509 credentials**, i.e. that you don't have a directory called “.globus” in your home directory¹. If you do have such a directory, you're recommended to move it out of the way before running the examples.
- If you want to run your own GridWorker:
 - you need an Internet connection fast enough that you can download a disk image of ~1 GB without losing patience
 - VirtualBox or Qemu should be installed on your machine²
 - you need a reasonably powerful machine with ≥ 2 GB of RAM and ≥ 4 GB of free disk space

3 Choosing servers and services to use

Before running any jobs, you should first decide where you want to run them. If you want to run any serious work, you'll probably need to set up your own GridFactory server. How to do this is described on <http://www.gridfactory.org/grid/>. For the examples in this document, you can use the central GridFactory server www.gridfactory.org.

4 Running a few simple jobs

To GridFactory, a job is just a shell or bat script. It can be as simple as a file, say “my_script.sh” containing:

```
#!/bin/bash
echo "Hello world"
```

This can be run by

```
> psub -b www.gridfactory.org my_script.sh
Job submitted. Identifier is
https://www.gridfactory.org/gridfactory/jobs/716c1e9d-0c21-11e0-ac32-
3181dc42bf59
```

Instructions to GridFactory on what input to download, where to upload output files, what software to provision, how much RAM is needed etc. are given in the form of in-line comments – just like on other batch systems. E.g. to calculate the MD5 sum of a file “my_file.txt” on your own hard disk, you could use the following script:

```
#!/bin/bash
md5sum my_file.txt
#GRIDFACTORY -i my_file.txt
```

¹On Windows, typically something like “C:\Documents and Settings\Frederik Orellana\globus”.

²This is not strictly necessary, but recommended in order to take full advantage of the software provisioning capabilities of GridFactory.

As another example, say you want to convert an image from JPG to PNG format. This could be done by:

```
#!/bin/bash
convert my_image.jpg my_image.png
#GRIDFACTORY -i my_image.jpg
#GRIDFACTORY -o my_image.png
```

Running this job will cause the file “my_image.png” to be available for download on the GridFactory server. After the job has finished, you can download it by executing

```
> pget job_identifier
```

where `job_identifier` is the string returned by `psub`.

You can get a list of all your jobs and see the status of several or all of them with the command `pstat`. Other commands of interest are: `pcat` and `pkill`. All commands output a short description if you give them the argument `-h`.

5 Running GridWorkers

If your jobs don't start fast enough or don't start at all, it could well be because the server doesn't have any running worker nodes attached. You can fix this by contributing your own machine as a worker node. This amounts to downloading and running a standard installer and starting up GridWorker as described on <http://www.gridfactory.org/share/>. Notice that GridWorker does not have configuration file; instead you need to start GridWorker from the command line and use the available arguments. E.g.

```
gridworker.sh -j 2 -u 4 -y 2 https://gridfactory.nbi.dk/db/
```

will cause GridWorker to pick up jobs from `gridfactory.nbi.dk` and run maximally two jobs with maximally two other jobs queueing (downloading input files) and each job running in its own virtual machine.