



Enabling Grids for
E-science in Europe

Nicosia, Feb 16, 2009

the gEclipse platform: Application Porting and Deployment

Asterios Katsifodimos (asteriosk_AT_cs.ucy.ac.cy)

Nicholas Loulloudes (loulloudes.n_AT_cs.ucy.ac.cy)

Trainers, University of Cyprus



Session Requirements

- Successful completion of the following:
 - VO Setup
 - CA Certificates Setup
 - Grid Project Creation
 - Authentication Token Creation
 - Job Submission and Job Status Monitoring
 - Data Management

Session Aim and Scenario

Aim

- The aim of this session is to demonstrate to the trainee how an application can be ported in such a way that it can be successfully executed on the EGEE infrastructure.

Scenario

- How to search a huge dataset in a distributed manner.
- Search dataset for a specified word.
- Dataset: 5000 e-books (ASCII text) from the Gutenberg project.
(<http://www.gutenberg.org>)

Session Aim and Scenario

- The dataset will be downloaded from the Gutenberg website to a Grid Storage Element.
- N Jobs will be submitted. Each job will search into a portion of the dataset for the specific word.
- Each job output will be stored to an output folder to a Grid Storage Element.
- Create a Bash shell script that does the above and use g-Eclipse to deploy the script to the Grid (job submission)
- Output can be retrieved by the user at any point later in time.

Exemplary Bash Script

```
01. #!/bin/bash -x
02. input="urls.txt"
03. output="texts.tar.gz"
04. job_count=5
05.
06. word="condition"
07. id=$2
08. num_of_urls=`wc -l $input | cut -d ' ' -f 7`
09.
10. let "my_urls=$num_of_urls/$job_count"
11. let "start_line=$id*$my_urls"
12. let "end_line=$((id+1)*$my_urls) - 1"
13. let "current_line=0"
14.
15. mkdir -p output/
16.
17. for url in `cat urls.txt`
18. do
19.     if [ $current_line -ge $start_line ] && [
20.         ↪ $current_line -le $end_line ]
21.     then
22.         curl $url | grep -i $word >> output.temp
23.         if [ ! -s output.temp ]
24.         then
25.             cd output;
26.             curl -O $url ;
27.             cd ../
28.         fi
29.         rm output.temp;
30.     fi
31.     let "current_line=$current_line + 1"
32. done
33.
34. tar -zcf $output output/
35. rm -rf output/
```

URL Reference File

<http://www.gutenberg.org/files/10940/10940.txt>
<http://www.gutenberg.org/files/10941/10941.txt>
<http://www.gutenberg.org/files/10942/10942.txt>
<http://www.gutenberg.org/files/10943/10943.txt>
<http://www.gutenberg.org/files/10944/10944.txt>
<http://www.gutenberg.org/files/10945/10945.txt>
<http://www.gutenberg.org/files/10946/10946.txt>
<http://www.gutenberg.org/files/10947/10947.txt>