

# Grid Computing Programming Assignment

## Grid Computing Module

### October 2004

#### Assignment

This assignment will explore various performance aspects of Grid technologies. The basis of the investigation is to:

1. Write a Java based application that will perform two main operations:
  - search a text file (and retrieve the number of times a certain term occurs);
  - sort the text file so that the terms and number of times they occur is given.

Students may implement whatever search/sorting algorithm they wish. The text file to be used is located in `/home/gc5/assignment/inputFile/shakespeareWorks.txt`

2. Perform benchmarking on the speed of the application on a single PC for searching and sorting the text file. (Benchmarking should include several timed runs of the search and sort algorithms on a single PC followed by statistical averaging.)

3. Extend and parallelise the application to make use of the training lab Condor pool.

4. Perform benchmarking on the speed of the application across the Condor pool. What are the performance overheads in using Condor? How is the performance affected by the number of nodes used and the distribution of the data across these nodes?

5. Wrap the *parallelised* application as a GT3.3 Grid service called "*searchSortGridService*" with method names "*searchMethod*" and "*sortMethod*" and develop a client to test them.

6. Perform benchmarking on the speed of the GT3.3 service. What are the performance overheads in using GT3.3?

7. Extend the GT3.3 based parallelised application so that it supports secure communications using GSI between the Grid service client and the Grid service (not between the Grid service and the Condor nodes). This security should support server side security so that the service is only accessible by **yourself** and the **lecturing staff** (Dr's Perkins, Sinnott and Watt). It is not necessary to support message level security (encryption) also. The DN of the lecturing staff are C=GB, O=University of Glasgow, OU=Lecturers, CN=John Watt/Richard Sinnott/Colin Perkins.

8. Perform benchmarking on the speed of the GSI secure GT3.3 version of the application. What are the performance overheads in using the GSI security infrastructure?

9. Take the application in step 5 and using the XML policy developed in Problem Set 2, ensure that method *sortMethod* can **only** be invoked by members of your student group and the lecturing staff, and that method *searchMethod* can be invoked by everyone.

Deploy the service developed in part 5 and ensure that it is configured to use the PERMIS authorisation infrastructure with the XML policy that was created in Problem Set 2.

To allow you to test that the policy is being adhered to by the service, the following students are assigned to these groups:

#### Student Group 1:

Alexandros Koliouris  
Sakshi Anand  
Christopher Bayliss

#### Student Group 2:

Kenchangouda Patil  
Yong Que  
Arun Kumar Sathyanarayan

Gary Fleming  
Jiyu Jiang  
Omar Kooheji  
Ross McIlroy  
Derek Murray

Vikas Shas  
Stuart MacDonald  
Krunal Thakkar  
Kashif Saleem  
Vinit Shah

10. Perform benchmarking on the speed of the PERMIS secured GT3.3 version of your application? What are the performance overheads in using a role based authorisation infrastructure?

### **Marking**

This assignment is due at 5:00PM on Friday, 29<sup>th</sup> November 2004. The marks will be assigned 50% for the design and implementation of the various implementations of the search/sort algorithm, 30% for the way in which performance aspects are addressed, and 20% for the write up and discussion (including the issues in performance benchmarking on a shared set of resources and the pro's and con's of GSI and PERMIS). You should submit the source code for the Grid services and any design notes you have made.