The Cambridge CFD Grid Portal for Large-Scale Distributed CFD Applications

Xiaobo Yang[1], Mark Hayes[1], Karl Jenkins[2], and Stewart Cant[2]

[1] Cambridge eScience Centre, University of Cambridge
Wilberforce Road, Cambridge CB3 0WA, United Kingdom
`xy216,mah1002@cam.ac.uk`
[2] Department of Engineering, University of Cambridge
Trumpington Street, Cambridge CB2 1PZ, United Kingdom
`kwj20,rsc10@eng.cam.ac.uk`

Abstract. The Cambridge CFD (computational fluid dynamics) Web Portal (CamCFDWP) has been set up in the Cambridge eScience Centre to provide transparent integration of CFD applications to non-computer scientist end users who have access to the Cambridge CFD Grid. Besides the basic services provided as other web portals such as authentication, job submission and file transfer through a web browser, the CamCFDWP makes use of the XML (extensible markup language) techniques which make it possible to easily share datasets between different groups of users.

1 Introduction

CFD is now widely used in aerodynamics, automotive industry, etc. In order to satisfy the increased demands of understanding complex flows, increased computing power becomes more and more important for large-scale CFD applications. With the emerging Grid technique [1], the integration of resources belonging to different organisations is now practical. The Cambridge CFD Grid, a distributed problem solving environment between the Cambridge eScience Centre and the CFD Lab at the Cambridge University Engineering Department has been set up as a testbed for such large-scale distributed CFD applications. At the same time, the Cambridge CFD Web Portal (CamCFDWP) [2] has been developed in the Cambridge eScience Centre to provide end users transparently access to the power of computing resources contributed to the Cambridge CFD Grid through a web browser.

In this paper, we first briefly describe the Cambridge CFD Grid. Then the CamCFDWP with application of the XML techniques is depicted in detail. Finally our conclusions are presented.

2 Cambridge CFD Grid

As mentioned above, the Cambridge CFD Grid is a distributed problem solving environment. The Globus group [3] defined the Grid as "an infrastructure

that enables the integrated, collaborative use of high-end computers, networks, databases, and scientific instruments owned and managed by multiple organisations." Detailed information on the Grid technique was given by Foster *et al.* [1,4,5] in their publications. An introduction to the Grid technique in CFD is reported by Yang *et al.* [6]

Currently, the Cambridge CFD Grid comprises two dedicated linux clusters, a web server, database and dedicated data storage machines. The network link between the two sites is currently investigated by considering a virtual private network (VPN) for security although this has not been fully tested yet. Once setup, the VPN will provide a route around the departmental firewalls. The clusters run the Globus Toolkit [7] and Condor [8] for remote job submission, file transfer and batch queue management.

SENGA, a parallel combustion DNS (direct numerical simulation) code developed by Jenkins *et al.* [9] at the Cambridge CFD Lab, has been tested in the Cambridge CFD Grid. The CFD code is used to study the effects of a turbulent flame kernel, in which there exists a strong coupling between turbulence, chemical kinetics and heat release.

3   Cambridge CFD Web Portal

The Globus Toolkit v2.4.3 used in the Cambridge CFD Grid provides a set of command line tools to manage remote computing resources. This means extra work for end users to get accustomed to these commands. In order to provide transparent access to remote resources including computing resources, large datasets, etc., many web portals such as the ASC Portal [10], the Telescience Project [11] and PACI HotPage [12] have been set up. Basically these portals enable end users to run large-scale simulations through web interfaces. The aim of the CamCFDWP is also to hide command line tools of the Globus Toolkit and resources behind a simple but user friendly interface, i.e., web interface. The CamCFDWP provides the ability to guide users through running the SENGA CFD code inside the Cambridge CFD Grid.

The current version of the CamCFDWP was developed based on the Grid Portal Toolkit (Gridport) 2.2 [13] with the following capabilities through a web browser: 1) login/logout through MyProxy [14] delegation, 2) remote job submission either interactively or in batch mode, 3) a batch job manager, 4) file transfer including upload, download and third party transfer, and 5) a database (Xindice [15]) manager. Fig. 1 shows the architecture of the CamCFDWP. The portal web server plays a key role. Whatever an end user wants to do on remote computing resources, he or she only needs to contact the portal web server through a web browser, from which he/she can execute his/her job.

As XML is fast becoming an industry standard because of its intrinsic merit for data exchange, we adopted XML techniques in order to store information about each job. Without too much modification of the legacy FORTRAN CFD code (SENGA, mainly to read in new parameters), a user inputs parameters through a web form in the CamCFDWP. These parameters will first be saved as an XML file, which will then be validated against a schema [16] designed for
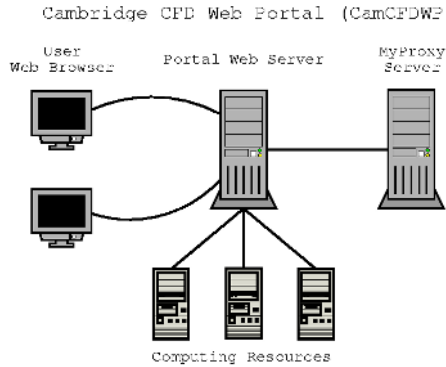
Fig. 1. Architecture of the Cambridge CFD Web Portal (CamCFDWP)

SENGA. Inside the schema, all the input parameters are described as precisely as possible so that they can be set up correctly for SENGA. Xerces-C++ [17] is used to validate the XML file against the schema. If the validation is successful, a plain text file with all input parameters will then be created and transferred to SENGA. Inside the XML file, extra information such as the creator and date are also saved.

When the numerical simulation has finished (on remote machines), all output data are transferred to a file server. During this stage, the location of these data will be recorded in the same XML file mentioned above. Thus for each calculation, the input parameters, location of output data, creator, date, etc. are all recorded in one XML file. Apache Xindice [15], a native XML database has been adopted to manage these small XML files (each job has one XML file accordingly). According to our tests, it has the ability to query an element in an XML database and return elements only or whole XML files. For example, a user may be interested in querying all data created by user "xyang", or all simulations done with the "*Reynolds number*" equals "30.0".

While developing the CamCFDWP, we have also developed a similar web portal for the Cambridge EM (electromagnetic scattering from aircraft) Grid. Basically, we simply modified a configuration file of the CamCFDWP. Although it is really easy to do such work, we realise that for centres like the Cambridge eScience Centre with many projects hosted it is not a good idea to develop one web portal for each project with similar interface. Thus, we are now developing some portlets. These portlets are divided into two classes. First, general portlets for authentication, file transfer, etc. These portlets should be available to all grid users. Second, particular portlets for particular projects. For instance, a RunSENGA portlet which should only be available to CFD people. With the help of Jetspeed [18], a portlet container, each user can customise his/her own web interface, he/she should have permission to run all general portlets and any special portlet. But he/she will not have permission to run portlets for other projects.

4   Conclusions

This paper describes the Cambridge CFD Web Portal for the Cambridge CFD Grid. Through a web browser, the CamCFDWP provides a user friendly interface, which makes the Grid transparent to end users. Besides the basic services of authentication, job submission and file transfer, XML techniques have been introduced to the project. At the current stage, XML brings us two benefits. First, an XML schema has been developed which makes it easy to validate user input parameters through the CamCFDWP. Second, Xindice, a native XML database has been set up which manages all the necessary information on each numerical simulation including all input parameters, user name, date and data location for possible future datasets sharing with other groups of users.

References

1. Foster, I. and Kesselman, C., "The Grid: Blueprint for a New Computing Infrastructure", Morgan Kaufman, San Francisco, Calif, 1999.
2. https://www.escience.cam.ac.uk/portals/CamCFDWP/.
3. http://www.globus.org/.
4. Foster, I. and Kesselman, C., "Globus: A Metacomputing Infrastructure Toolkit", *Int. J. Supercomputer Applications*, 11(2):115-128, 1997.
5. Foster, I., Kesselman, C. and Tuecke, S., "The Anatomy of the Grid: Enabling Scalable Virtual Organizations", *Int. J. Supercomputer Applications*, 15(3), 2001.
6. Yang, X. and Hayes, M., "Application of Grid Technique in the CFD Field", *Integrating CFD and Experiments in Aerodynamics*, Glasgow, UK, 8-9 September 2003.
7. http://www-unix.globus.org/toolkit/.
8. http://www.cs.wisc.edu/condor/.
9. Jenkins, K. and Cant, R.S., "Direct Numerical Simulation of Turbulent Flame Kernels", *Recent Advances in DNS and LES, eds. Knight, D. and Sakell, L.*, pp. 191-202, Kluwer Academic Publishers, New York, 1999.
10. Russel, M., Allen, G., Foster, I., Seidel, E., Novotny, J., Shalf, J., von Laszewski, G. and Daues, G., "The Astrophysics Simulation Collboratroy: A Science Portal Enabling Community Software Development", *Proceedings of High-Performance Distributed Computing 10 (HPDC-10)*, pp. 207-215, San Francisco, CA, 7-9 August 2001.
11. https://telescience.ucsd.edu/.
12. https://hotpage.npaci.edu/.
13. https://gridport.npaci.edu/.
14. http://grid.ncsa.uiuc.edu/myproxy/.
15. http://xml.apache.org/xindice/.
16. http://www.escience.cam.ac.uk/projects/cfd/senga.xsd.
17. http://xml.apache.org/xerces-c/index.html.
18. http://jakarta.apache.org/jetspeed/.