The Cambridge CFD Grid for Large-Scale Distributed CFD Applications

X. Yang [a], M. Hayes [a,*], K. Jenkins [b], S. Cant [c]

[a] *Cambridge eScience Centre, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK*

[b] *Department of Aerospace Sciences, School of Engineering, Cranfield University, Cranfield, Bedfordshire MK43 0AL, UK*

[c] *CFD Laboratory, Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK*

Abstract

The Cambridge CFD (computational fluid dynamics) Grid is a distributed problem solving environment for large-scale CFD applications set up between the Cambridge eScience Centre and the CFD Lab in the Engineering Department at the University of Cambridge. A Web portal, the Cambridge CFD Web Portal (CamCFDWP) has been developed to provide transparent integration of CFD applications to non-computer scientist end users. In addition to the basic services provided of authentication, job submission and file transfer, the CamCFDWP makes use of XML (extensible markup language) techniques which make it possible to easily share datasets between different groups of users. A Web service interface has recently been implemented for a CFD database which could be integrated in the CamCFDWP in the near future. We also review how this Web service can be made secure using SSL, XML signatures and XML encryption.

Key words: Grid, Portal, Web services, XML, Security

* Corresponding author.
Email address: mah1002@cam.ac.uk (M. Hayes).
URL: http://www.escience.cam.ac.uk/ (M. Hayes).

1 Introduction

CFD is now widely used in industry and academia for the development and modelling of complex problems such as the aerodynamic aspects of an aircraft. 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 large-scale distributed CFD applications. 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.

At the same time, the Cambridge CFD Web Portal (CamCFDWP) [2] has been developed in the Cambridge eScience Centre to provide end users with transparent access to the power of computing resources contributed to the Cambridge CFD Grid through a Web browser. 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. These XML descriptions are stored in a native XML database. We have developed a Web service interface to this database which allows us, with little further effort, to automatically annotate, store and query CFD simulation result sets from within the portal.

In this paper, we first describe the Cambridge CFD Grid and the Web portal–CamCFDWP. Then application of the XML techniques in the CamCFDWP is depicted in detail followed by the development of a Web service interface for a database of CFD simulation results. Finally our conclusions are presented.

2 Cambridge CFD Grid

As mentioned above, the Cambridge CFD Grid is a distributed problem solving environment. The "Grid" concept was developed during the mid-1990s based on decades of previous work on distributed computing. 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.

3 Cambridge CFD Web portal (CamCFDWP)

3.1 Architecture of the CamCFDWP

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 NPACI 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.

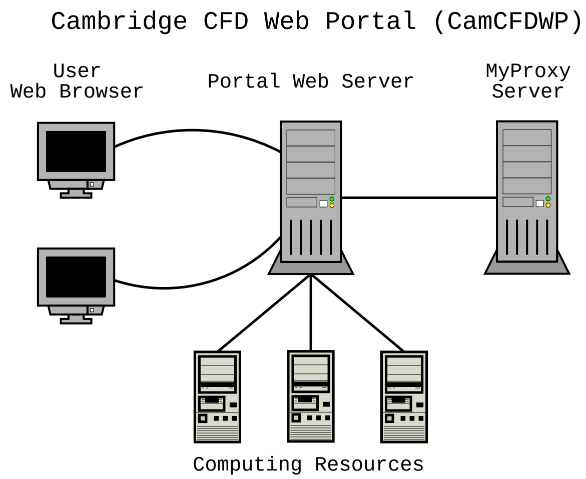

Fig. 1. Architecture of the Cambridge CFD Web Portal (CamCFDWP)

3.2 XML in CamCFDWP

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 (Fig. 2) in the CamCFDWP. These parameters will first be saved as an XML file, which will then be validated against a schema [16] designed for 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.

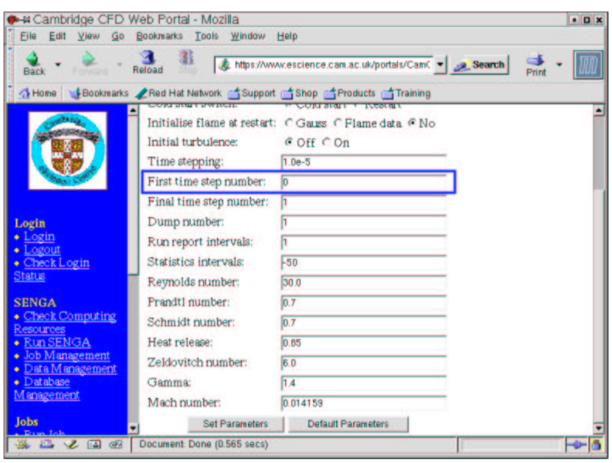

Fig. 2. Set parameters for CFD code SENGA within the Cambridge CFD Web Portal (CamCFDWP)

As shown in Fig. 2, the input parameter *"First time step number"* is set to "0" by a user. In fact, as you can see from the schema [16] below, this parameter is defined as type "*positive_integer*", i.e. 1, 2, 3... and "0" in this case will therefore not pass validation. The XML file saved by the CamCFDWP and the validation result are shown in Fig. 3 and Fig. 4 accordingly.

$$< xsd : element\ name = "firstTimeStepNumber"$$

$$type = "positive\_integer" / >$$

$$< xsd : simpleType\ name = "positive\_integer" >$$

$$< xsd : restriction\ base = "xsd : integer" >$$

$$< xsd : minInclusive\ value = "1" / >$$

$$< /xsd : restriction >$$

$$< /xsd : simpleType >$$

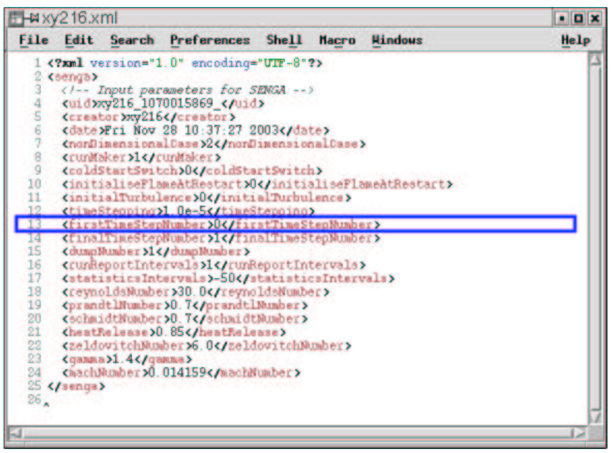

Fig. 3. User input parameters for SENGA are saved as an XML file

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". Fig. 5 shows a result obtained by querying the current database for simulations with the "*Reynolds number*" less then "40.0".

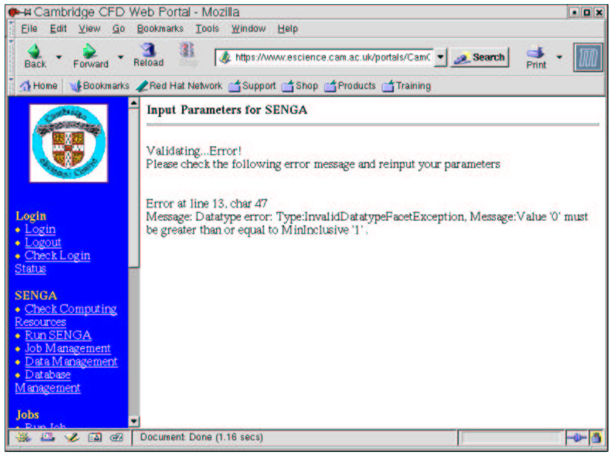

Fig. 4. User input parameters for SENGA are validated against the schema

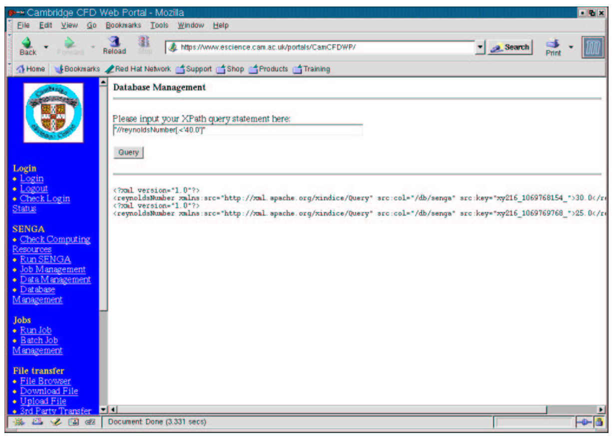

Fig. 5. Querying the Xindice database

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 have developed 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.

Fig. 6 demonstrates a Web portal based on portlets. One of the portlets, xportlets:ProxyManager is developed by the Alliance Portal team [19]. The Alliance Portal has a collection of several portlets inlcuding LDAP Browser

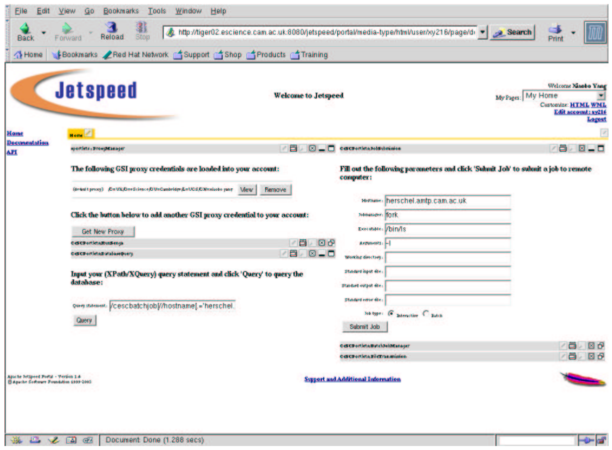

Fig. 6. Web portal based on portlets

and GridFTP Client. These general purpose protlets can be easily re-used as the Proxy Manager portlet in our portal. But currently there is still need to recompile the code locally. With the help of Web service, the next generation of portals should be able to publish local portlets as remote portlet Web services to make them available to other portals [20].

4 Developing Web service interfaces for a CFD database

After the short-lived Open Grid Services Infrastructure (OGSI) [21], WS-RF [22–24] is now under development by the Globus Alliance and IBM. Both OGSI and WS-RF are based on Web services architecture but with extensions to address problems such as the state of services with lifetime management. As Grid services are moving towards Web services, we continue to make use of the Globus Toolkit v2.4.3 but with investigation in developing Web services in a computational grid environment [25]. In this section, we are going to describe some ongoing work at the Cambridge eScience Centre in developing Web service interfaces for a CFD database.

As described in Section 3.2, the XML database records the meta-data, e.g., creator and simulation parameters, rather then the raw CFD data. In this test, we adopt eXist [26], another open source native XML database, to store the meta-data. Two Web services have been developed to manage the CFD database, database query service and database add service. Both services are exposed with Apache Axis [27] deployed in Tomcat [28].

The database query service accepts the XQuery statement as its input parameter. A Java servlet has been developed to act as the service client. For instance, a query condition is set in Fig. 7 as querying the "Peak wave number" with the condition "< 20.0". The query statement will then be constructed in the servlet as "$//camcfddataset[peakWaveNumber < 20.0]$" and transferred to the service for query. The query service will then return the result in XML format. With the help with XSLT, the result can be transformed to other formats like HTML or plain text.

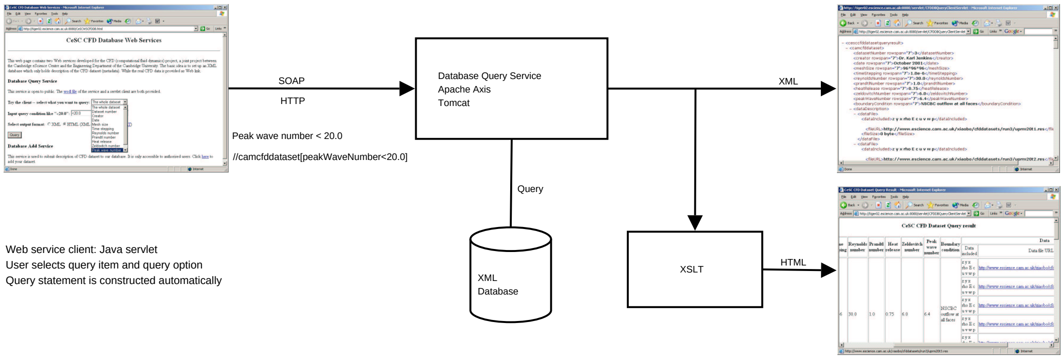

Fig. 7. Database query service

The database add service shown in Fig. 8 is different from the database query service, which is available to public. This service is currently protected using basic HTTP authentication over SSL. Although not implemented in this service, different security techniques have been tested successfully. XML signature has been tested to sign the meta-data that data providers want to add to the database to keep their integrity. With the help of Java CoG [29] package, Globus proxy has been transferred by a Web service client to the service for authentication purpose successfully.

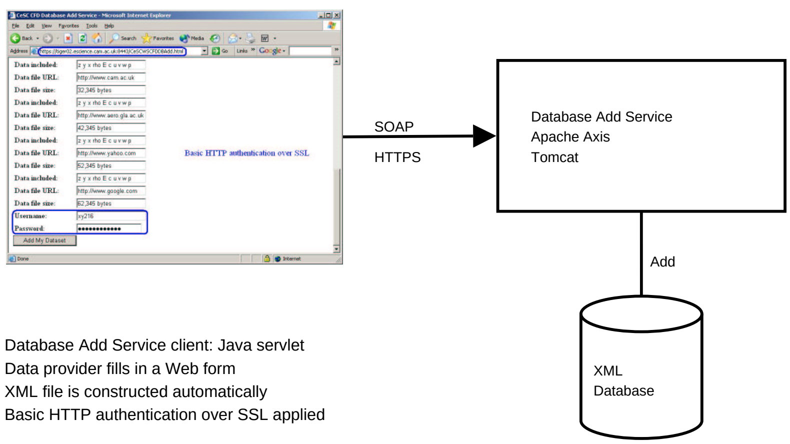

Fig. 8. Database add service

5 Conclusions and future work

This paper describes a distributed problem solving environment, the Cambridge CFD Grid, and the Web Portal – CamCFDWP. 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, native XML database like Xindice and eXist could be used to manage meta-data on each numerical simulation including all input parameters, user name, date and data location for possible future datasets sharing with other groups of users.

Web service interfaces have been set up successfully for the CFD database. The big issue we find is how to keep the service secure. Although more proposals are emerging, such as XML digital signature, SAML and XKMS, there is now no widely accepted standard. As far as wee see, XML digital signature and XML encryption are two footstones to solve this security issue on message level for Web services. As a mature technique, SSL over HTTP could be heavily used in the near future. Further investigation may be carried out to develop GSI-enabled Web services [30].

The integration of the Web service interfaces into the CamCFDWP would be our future work together with the cooperation with the Cambridge-Cranfield High Performance Computing Facility.

Acknowledgements

This work was undertaken at the Cambridge eScience Centre supported by EPSRC and the DTI under the UK eScience Programme.

References

[1] I. Foster and C. Kesselman, The Grid: Blueprint for a New Computing Infrastructure (Morgan Kaufman, San Francisco, Calif, 1999).

[2] CamCFDWP. https://www.escience.cam.ac.uk/portals/CamCFDWP/.

[3] Globus group. http://www.globus.org/.

[4] I. Foster and C. Kesselman, Globus: A Metacomputing Infrastructure Toolkit, Int. J. Supercomputer Applications, 11(2) (1997) 115-128.

[5] I. Foster, C. Kesselman, S. Tuecke, The Anatomy of the Grid: Enabling Scalable Virtual Organizations, Int. J. Supercomputer Applications, 15(3) (2001).

[6] X. Yang and M. Hayes, Application of Grid Technique in the CFD Field, in: Integrating CFD and Experiments in Aerodynamics (Glasgow, UK, 8-9 September 2003).

[7] Globus Toolkit. http://www-unix.globus.org/toolkit/.

[8] Condor. http://www.cs.wisc.edu/condor/.

[9] K. Jenkins and R.S. Cant, Direct Numerical Simulation of Turbulent Flame Kernels, in: Recent Advances in DNS and LES, eds. Knight, D. and Sakell, L. (Kluwer Academic Publishers, New York, 1999), 191-202.

[10] M. Russel, G. Allen, I. Foster, E. Seidel, J. Novotny, J. Shalf, G. von Laszewski, G. Daues, The Astrophysics Simulation Collboratroy: A Science Portal Enabling Community Software Development, in: Proc. HPDC-10 (San Francisco, USA, 7-9 August 2001), 207-215.

[11] Telescience. https://telescience.ucsd.edu/.

[12] HotPage. https://hotpage.npaci.edu/.

[13] GridPort. https://gridport.npaci.edu/.

[14] MyProxy. http://grid.ncsa.uiuc.edu/myproxy/.

[15] Xindice. http://xml.apache.org/xindice/.

[16] CFD schema. http://www.escience.cam.ac.uk/projects/cfd/senga.xsd.

[17] Xerces. http://xml.apache.org/xerces-c/index.html.

[18] Jetspeed. http://jakarta.apache.org/jetspeed/.

[19] Alliance Portal. http://www.extreme.indiana.edu/xportlets/.

[20] I. Naick and J. Wilson, Developing Remote Portal Web Services, IBM developerWorks (August 2004).

[21] S. Tuecke, K. Czajkowski, I. Foster, J. Frey, S. Graham, C. Kesselman, T. Maguire, T. Sandholm, P. Vanderbilt, D. Snelling, Open Grid Services Infrastructure (OGSI) Version 1.0, Global Grid Forum Draft Recommendation (27 June 2003).

[22] D. Sabbah, Bringing Grid & Web Services Together, presented in: GlobusWORLD 2004 (San Francisco, USA, 20-23 January 2004).

[23] I. Foster, WS-Resource Framework: Globus Alliance Perspectives, presented in: GlobusWORLD 2004 (San Francisco, USA, 20-23 January 2004).

[24] WS-RF whitepaper. http://www-106.ibm.com/developerworks/library/ws-resource/ws-wsrfpaper.html.

[25] X. Yang, M. Hayes, A. Usher, M. Spivack, Developing Web Services in a Computational Grid Environment, in: CG&SC-SCC2004 (Shanghai, China, 15-18 September 2004).

[26] W. Meier, eXist: An Open Source Native XML Database. In: Web, Web-Services, and Database Systems, NODe 2002 Web- and Database-Related Workshops (Erfurt, Germany, October 2002), Springer LNCS Series, 2593.

[27] http://ws.apache.org/axis/.

[28] http://jakarta.apache.org/tomcat/.

[29] http://www-unix.globus.org/cog/java/.

[30] http://www.epcc.ed.ac.uk/ neilc/gsiws/.

Xiaobo Yang is a scientific programmer at the Cambridge eScience Centre, University of Cambridge. He received a BEng from Tsinghua University in 1997, and in 2003 a PhD from the University of Glasgow, where he studied computational fluid dynamics. Dr. Yang currently works on two projects – EMGrid and Cambridge CFD Grid. Both projects aim to develop integrated grid solutions. His research interests are in the application of grid and web services techniques to science and engineering. Dr. Yang is also interested in developing user-friendly web portals for end users.

Mark Hayes is the Technical Director of the Cambridge eScience Centre. He graduated from Kings College, Cambridge with a BA(Hons) in Mathematics. Before joining the eScience Centre in November 2001, Mark was a Senior Technical Officer in the Systems group at the European Bioinformatics Institute, and before that Systems Manager for E*Trade UK. He is interested in all forms of wide area distributed computing and its applications to engineering, particle physics, biology and environmental sciences.

Dr. Karl Jenkins is a Lecturer in the Computational Aerodynamics Design group in the department of Aerospace Sciences within the School of Engineering, Cranfield University. Dr. Jenkins joined Cranfield University in April 2004. After leaving Manchester University with a PhD in numerical and experimental fluid flows, he joined the Computational Fluid Dynamics Laboratory at Cambridge University. He has also worked for Allott and Lomax consulting engineers and Davy Distington Ltd working with various commercial CFD codes such as FLOW3D (Los Alomos). Dr. Jenkins is currently a consultant for Cambridge Flow Solutions Ltd (CFS). He has developed parallel codes for various fluid applications from academic to blue chip industrial companies, such as Rolls Royce plc. Current research interests are in the areas of computational fluid dynamics (CFD), numerical simulation of turbulent combustion and fluid flows, high performance computing (HPC) for fluid dynamics, numerical modelling of industrial problems, eScience grid technology for CFD.