A Survey on Training and Education Needs for Petascale Computing[1]

Tim Stitt[2] and Tim Robinson[3]
ETH Zurich-CSCS, Switzerland
[2]tstitt@cscs.ch, [3]robinson@cscs.ch

Abstract

In this report we present the outcome of a comprehensive survey of HPC training and educational needs amongst the top-tier users across all PRACE partner sites. We describe the design and implementation of this survey which examines user opinion on their current satisfaction and competence with existing training and educational material, as well as solicit their requirements for new and improved training material aimed at high-performance scalable computing towards the Petascale regime. In addition to surveying technical training requirements, the survey also assesses user satisfaction with existing training delivery methods, and the potential of novel presentation approaches. The results of the survey indicate that there is a significant need for improved training in nearly all areas of HPC education. In particular the survey identifies a definite need for improved training in mixed-mode MPI/OpenMP programming, multi-core programming, optimization techniques, parallel I/O and visualization. Furthermore, it is revealed that users would significantly benefit from a centralised European repository of training material and HPC knowledge dissemination. It is hoped that the recommendations presented in this report describe a roadmap of how user requirements and training deficiencies can be satisfied in conjunction with a sustained, comprehensive PRACE educational programme encompassing summer schools, winter schools and training workshops.

[1] This work was supported by the European Commission under the Seventh Framework Programme (FP7), grant agreement number RI-211528

Table of Contents

A Survey on Training and Education Needs for Petascale Computing i

Abstract i

Table of Contents ii

List of Tables ii

Executive Summary 1

1 Introduction 2

2 Design and Deployment of the HPC Training and Education Needs Survey 3

2.1 Survey Design 4

2.1.1 *Choice of Online Survey Tool* *5*

2.1.2 *HPC Survey Themes* *5*

2.2 Survey Deployment 6

2.1.3 *Submission Prize* *6*

2.1.4 *Survey Completion* *6*

3 Survey Results 7

3.1 HPC Background (Demographics) 7

3.2 Programming Languages 9

3.3 Parallel Programming Methodologies 16

3.4 Programming Tools and Libraries 22

3.5 Debugging, Profiling and Optimisation Tools 27

3.6 Parallel I/O and Fault-Tolerance 33

3.7 Third-Party Scientific Applications 35

3.8 Basic Linux/Linux Tools and Techniques for HPC 37

3.9 Distributed Systems Access and Grid Middleware 38

3.10 Scientific Visualisation 41

3.11 Training Requirements 45

4 Conclusions and Recommendations 50

4.1 Immediate Requirements 50

4.2 Short-Term Requirements 50

4.3 Long-Term Requirements 51

4.4 General Requirements 51

References and Applicable Documents 53

List of Tables

Table 1: The eleven HPC themes assessed in the survey 5

Table 2: Views of some respondents on HPC architecture optimization 18

Executive Summary

This report describes the preliminary steps in developing and implementing a comprehensive and focused training and educational programme for tier-0 users, aimed at exploiting high-performance scalable computing on leadership-class systems, towards the petascale regime.

In this document we describe the design, implementation and results of a comprehensive survey which was conducted online to assess user competence and satisfaction with existing training and educational material, as well as solicit requirements for new and improved training material and delivery methods aimed at high-performance scalable computing. To motivate completion of the survey, a reward was offered to a randomly selected user who successfully submitted a response by the survey expiration deadline.

The survey was undertaken by 119 participants, featuring the top users (in terms of system usage) at each of the PRACE partner sites, between the 15th April 2008 and 17th May 2008. By analysing the data obtained from the survey responses we present recommendations which will be used to steer future PRACE training and education activities. In summary, we offer four (4) classifications of recommendations:

1. Immediate requirements
2. Short-term requirements
3. Long-term requirements
4. General requirements.

The immediate requirements that were identified include a significant need for improved mixed-mode (hybrid) MPI/OpenMP and visualization training materials (both in quality and content). Furthermore, PRACE should immediately begin to investigate the introduction of a centralised European repository for training material and HPC knowledge dissemination. Short-term requirements include the development of world-class training materials covering multi-core programming techniques, code and compiler optimizations, parallel I/O, parallel debugging and code testing strategies. Longer-term goals should be to continue to observe and adapt to the impact of technologies such as next-generation HPC programming languages, scripting languages and Grid/Distributed computing tools within the HPC community. Finally, our general requirements recommend good practice for the delivery of future training and education programmes based upon users' prior training experiences and satisfaction.

Subsequent tasks within the project will focus on implementing the recommendations presented in this report as part of a sustained, comprehensive training and education programme including world-class summer schools, winter schools and workshops.

1 Introduction

The Partnership for Advanced Computing in Europe (PRACE) has the overall objective to prepare for the formation of a persistent pan-European HPC service. PRACE is a synergy of eight work packages with the objective to:

1. Create and implement by 2009/2010, a persistent, sustainable pan-European HPC service with three to five HPC leadership systems of Petaflop/s performance;
2. Define and establish a legal and organizational structure involving HPC Centers, national funding agencies, and scientific user communities;
3. Prepare for the deployment of Petaflop/s systems in 2009/2010 under the responsibility of European supercomputing centers having the expertise, competency, and required infrastructure to provide a comprehensive service to industry and academia user groups;
4. Collect requirements and demands from the user community about future challenging applications.

One work package (WP3) coordinates the dissemination of PRACE activities and implements an education and training programme for computational science aiming at scalable computing.

While the remaining seven work packages focus primarily on preparing the hardware, software and legal infrastructure for high-performance scalable computing in Europe, it is critical that the potential users of these leadership-class architectures are adequately trained and prepared to fully exploit these systems as soon as they become available. Otherwise, the satisfaction in achieving this infrastructure will be short-lived as users fail to maximize its potential.

This document describes the initial steps in assessing the current and future training and education needs of the potential top-tier users within the PRACE HPC infrastructure. It is anticipated that the recommendations produced from this report will steer subsequent implementing a world-class education and training service for the European HPC community.

The report is structured as follows:

- Section 2 describes the design and implementation of a survey to assess the current and future HPC training and education needs of top-tier users across PRACE partner sites
- Section 3 presents the results from this survey
- Section 4 provides conclusions and recommendations to direct subsequent PRACE education and training activities.

2 Design and Deployment of the HPC Training and Education Needs Survey

A sustained, high-quality training and education programme is paramount in guaranteeing that top-tier users of the PRACE HPC Research Infrastructure will:

- *remain productive;*
- *evolve with innovative hardware and software environments;*
- *cultivate new communities of users who can respond to, and disseminate information regarding new developments and computational techniques;*
- *be in a position to fully exploit the strengths of existing and future hardware and software resources.*

Tomorrow's HPC users will be required to harness the huge parallelism embodied in petascale computing architectures containing hundreds of thousands of cores. This activity will be tremendously non-trivial but necessary, if researchers wish to remain competitive and continue to advance knowledge in their field.

> *"Application developers today write programs that are as complex as describing where every single bit must move between the 6,000 transistors of the 808a microprocessor"* [4]

In addition to commanding thousands of cores in a scalable manner, users will also most likely need to embrace the following expertises:

- *develop scalable parallel codes across heterogeneous architectures;*
- *hybrid OpenMP/MPI programming techniques and/or partitioned global address space languages (PGAS) approaches such as Co-Array Fortran, UPC and Titanium;*
- *novel HPC programming languages such as Chapel and X10;*
- *mixed-language code development;*
- *fault-tolerant programming techniques;*
- *many-core debugging and testing techniques;*
- *many-core load-balancing and latency mitigation;*
- *high-performance parallel I/O techniques for large (possibly petabyte) file production.*

Unfortunately, improving existing training materials will not be enough to keep users at the leading edge of simulation science. There will need to be a concerted effort to develop new material which exploits tried-and-tested education methodologies, as well as new modes of training delivery.

While existing HPC training and education programmes routinely exploit traditional learning techniques such as:

- *classroom style teaching;*
- *online tutorials/courseware;*
- *documentation;*
- *workshops, seminars and conferences;*

the time may be opportune to investigate the introduction of novel learning methods in HPC education such as:

- *screencasting (narrated software movies)*
- *flash documentation*
- *wiki and knowledge repositories*
- *multicast remote training (Access Grid)*

Determining the current and future training requirements of PRACE HPC Infrastructure users, in preparedness for a comprehensive education programme, requires great responsibility. Assessing the educational needs of users not only requires a thorough understanding of their existing skills and competences in traditional HPC techniques but also needs to predict the utility and relevance of forthcoming technologies that may shape the HPC landscape in both the short and long term. Obtaining this complete information is critical in directing the development of any prospective educational programme.

A number of conventional approaches that can be employed to carry out this user assessment are:

- *Face-to-face Meetings*
- *Focus Groups*
- *Surveys*
- *User Forums*

Due to the time and human resource constraints (as defined in the PRACE Description of Work document [2]) it was decided that a survey would be the most effective means of assessing the educational and training requirements of PRACE HPC Infrastructure users. A summary of arguments supplementing the decision to employ a survey are:

1. Many people can be polled in a short time
2. Closed-form questions can be quick to tabulate
3. Different question formats to extract relevant information
4. Open-ended questions can be used to garner new ideas

2.1 Survey Design

To facilitate the design of survey questions and the collection of survey responses it was decided that an online survey tool would be a valuable instrument in making the process as effortless and adaptable as possible. In addition to the providing a powerful environment for designing and deploying surveys, surveys conducted online also tend to have the following benefits:

- *are easier and faster to complete*
- *have a better response percentage*
- *prevent email and attachment difficulties*
- *no specific software requirements (incompatibilities). . . just a browser*
- *no illegible/spoiled answers, missing/skipped pages etc.*
- *reduced printing and mailing costs*
- *reduction in compilation time and errors*

- *automatic tabulation of closed-form answers*

2.1.1 *Choice of Online Survey Tool*

Prior to embarking on the design of the training and education needs survey, a review of existing online survey tools was undertaken. The outcome of this review indicated that the online *SurveyMonkey* website (www.surveymonkey.com) provided the most appropriate tool in terms of features, functionality and price that satisfied our requirements. Along with a powerful and flexible question builder interface (that provides over twenty question types ranging from multiple choice to rating scales to open text) it provides automatic collector and analyser modules in a complete system.

> *The TeraGrid and CyberInfrastructure projects in the US recently used the Survey Monkey online survey tools to assess the ongoing needs of their HPC users. These users span some of the most important HPC centres in the US including SDSC, NCSA, PSC, TACC, Purdue, Indiana University, University California, ANL, ORNL, NCAR* [5]

2.1.2 *HPC Survey Themes*

Both authors of this report have extensive experience in developing and delivering HPC training courses to beginners and advanced students alike. Where the authors felt they had insufficient knowledge to satisfactorily survey a specialist topic in HPC, they recruited the help of domain-specific experts to develop relevant survey questions.

After an intensive design process a final draft of the survey was made available for review in March 2008 to members of both WP3[2] and WP6[3]. The survey comprised of 95 questions spanning the eleven (11) topical HPC areas given in Table 1.

1. *HPC background (demographics)*
2. *Programming languages*
3. *Parallel programming methodologies*
4. *Programming tools and libraries*
5. *Debugging, profiling and optimisation*
6. *Parallel I/O and Fault-tolerance*
7. *Third-party scientific applications*
8. *Unix/Linux tools and techniques*
9. *Distributed systems access and grid middleware*
10. *Visualisation*
11. *Training methodologies and preferences*

Table 1: The eleven HPC themes assessed in the survey

Due to the diverse backgrounds of the surveyed users, it was necessary to implement two tracks through the survey. Many HPC users routinely exploit third-party scientific applications without performing any in-house development or requiring highly-technical HPC development skills. To prevent this type of user from being forced to answer detailed questions on HPC development topics that are irrelevant to their needs, a skip-logic question was implemented near the outset of the survey; the answer to this question determined if the

[2] WP3 members primarily provided feedback on the survey design and presentation

[3] WP6 members were to provide feedback on the technical content of the survey (unfortunately at the time of deployment only one WP6 member responded with comments)

respondent continued to complete the full survey or was directed to a condensed version eliminating highly technical and detailed HPC development questions. This form of conditional question logic and redirection is easily implementable within the *SurveyMonkey* tool.

A final design requirement was that the survey be completed anonymously. In conjunction with questioning on training preferences and requirements, respondents were also asked to describe their proficiency in various technical HPC areas. The authors agreed that respondents were more likely to be honest in their responses, regarding their own levels of proficiency, if their identities were not associated with their submission.

The final version of the survey can be previewed at the following URL:

http://www.surveymonkey.com/s.aspx?sm=dyRFpC_2bpdwB_2fnmY_2fR2yvFA_3d_3d

2.2 Survey Deployment

On April 15th 2008, the Training and Education Needs survey was made available online to a select list of users within the PRACE partner community, via the notification facilities provided by the SurveyMonkey tool. The list of eligible users was determined from usage statistics provided by each partner site. It was agreed at the WP3 level that the Top 10 users at each site (it is anticipated that these users should in general have an extensive background in HPC and be potential tier-0 users of the PRACE HPC Research Infrastructure) should be invited to participate in completing the survey.

To accommodate for the situation where a Top 10 user may not necessarily be the code developer (but be a project PI), it was stressed in the survey invitation that users were obliged to redistribute the invitation to other group members who were primarily involved in the technical development of the user code. It was hoped that this fan-out would result in a representative cross-section of highly-skilled developers and users. In total, the invitational email was sent to 177 users throughout PRACE partner sites (except Portugal who failed to provide contact addresses for their Top 10 users).

2.1.3 *Submission Prize*

To encourage respondents to successfully complete the online survey before the expiry deadline, a prize of an iPod-nano was put on offer to one user, who would be selected at random from the set of completed surveys. On the 29th May 2008, a random draw was made at CSCS, Switzerland and the iPod-nano will be awarded to the winning user.

2.1.4 *Survey Completion*

The survey was made available online between April 15th 2008 and 17th May 2008. There were 93 fully completed surveys and 26 partially completed surveys[4] returned, representing all PRACE partner countries except Portugal.

[4] The answers from partially completed surveys are also included in the final survey data analysis. Only those respondents who submitted fully completed surveys were entitled to enter the draw for the iPod-nano.

3 Survey Results

In this section we introduce the results of the HPC Training and Education Needs survey that was described in Section 2. In presentation of the results we have made use of the reporting features provided by the SurveyMonkey online survey tool. For each question in the survey we visually present the raw data collected online and provide an explanatory paragraph analyzing the results and identifying significant trends where possible. Each subsection in Section 3 corresponds to one of the eleven HPC topic areas presented in Table 2.

3.1 HPC Background (Demographics)

1. What is your academic status?

	Response Percent	Response Count
Graduate Student	18.5%	22
Postdoctoral Fellow	19.3%	23
Researcher	**32.8%**	39
Professor	26.9%	32
Other (please specify)	2.5%	3
	answered question	119
	skipped question	0

2. How many years experience do you have in HPC?

	Response Percent	Response Count
1-2 years	20.2%	24
3-5 years	19.3%	23
6-10 years	25.2%	30
>10 years	**35.3%**	42
	answered question	119
	skipped question	0

3. In which scientific domains do you utilise HPC (please select all that apply)?	Response Percent	Response Count
Astrophysics	18.5%	22
Accelerator Physics	0.8%	1
Biophysics and Bioinformatics	7.6%	9
Chemistry	**26.9%**	32
Climate Modelling	9.2%	11
Computational Fluid Dynamics	16.0%	19
Combustion	2.5%	3
Engineering	2.5%	3
Financial Modelling	0.0%	0
Fusion	3.4%	4
Geophysics	10.9%	13
High Energy Physics	11.8%	14
Material Science	25.2%	30
Nano-science	11.8%	14
Nuclear Energy	0.0%	0
Nuclear Physics	3.4%	4
Signal and Image Processing	1.7%	2
Other (please specify)	9.2%	11
	answered question	119
	skipped question	0

The first three questions in this section determine the respondent's scientific and HPC background i.e. their professional status (e.g. student, professor), their years of experience in the HPC field, and the scientific domains in which they carry out their computational research.

In general there was a reasonable balance in responses from graduate students, postdoctoral fellows, researchers and professors (the definition of which may vary between European countries). The majority of respondents (60%) have 6 or more years experience in HPC and over one third of all respondents over 10 years experience. As described earlier in this report, one of the main aims of this survey was to determine the training needs of potential tier-0 users within the PRACE HPC Infrastructure, so the prevalence of responses from those with significant levels of experience is a pleasing result, and suggests that the target audience was appropriate.

Question 3 was included to determine those subject areas which dominate usage on national resources, and could indicate a core computing domain of initial tier-0 use[5]. The most dominant areas were Chemistry, Materials Science, Astrophysics and CFD, in decreasing order.

4. What types of codes do you run on HPC systems? (please select the first option if both apply)		
	Response Percent	Response Count
Codes developed by me and/or my research project	86.6%	103
Third-party scientific applications only	13.4%	16
	answered question	119
	skipped question	0

The final question in this section implemented the skip logic (as described in Section 2.1.2) to determine the track taken by the respondent through the survey, based on their HPC experience. The 13.4% of respondents who reported that they only use third-party codes i.e. "black-box users" were subsequently steered to section 7, while all remaining respondents continued onto section 2.

It is interesting to note that the overwhelming majority of respondents were involved in the development of HPC codes and applications. This indicates that a sustained PRACE HPC training and education infrastructure is relevant and necessary if these users wish to continue to remain competitive and knowledgeable of HPC skills and expertise both now and in the future.

3.2 Programming Languages

This section was designed to assess the respondent's competence and/or training satisfaction in traditional as well as novel HPC programming languages and related techniques.

5. Are the majority of your current programming language skills self-taught?		
	Response Percent	Response Count
Yes	93.8%	91
No	6.2%	6
	answered question	97
	skipped question	22

[5] This task is taken up in detail in WP6

6. Please indicate the importance of the following programming languages for your HPC code development:

	Not important	Somewhat important	Very important	Rating Average	Response Count
FORTRAN77	24.7% (22)	**38.2% (34)**	37.1% (33)	1.12	89
Fortran 95	12.9% (12)	20.4% (19)	**66.7% (62)**	1.54	93
Fortran 2003	**50.0% (42)**	32.1% (27)	17.9% (15)	0.68	84
C	27.0% (24)	36.0% (32)	**37.1% (33)**	1.10	89
C++	**40.7% (35)**	30.2% (26)	29.1% (25)	0.88	86
Java	**87.3% (69)**	11.4% (9)	1.3% (1)	0.14	79
Scripting Languages (Python, PERL, Ruby etc.)	27.3% (24)	**45.5% (40)**	27.3% (24)	1.00	88
If you use another language, please specify and indicate its relative importance					12
				answered question	**97**
				skipped question	**22**

7. Do you believe you would benefit from formal training in a given programming language?

		Response Percent	Response Count
Yes		**68.0%**	66
No		32.0%	31
		answered question	**97**
		skipped question	**22**

8. Please indicate your requirements for comprehensive formal training in the following programming languages:					
	Not important	Somewhat important	Very important	Rating Average	Response Count
FORTRAN77	**75.9% (60)**	16.5% (13)	7.6% (6)	0.32	79
Fortran 95	31.8% (27)	**41.2% (35)**	27.1% (23)	0.95	85
Fortran 2003	34.9% (29)	**36.1% (30)**	28.9% (24)	0.94	83
C	**37.5% (33)**	35.2% (31)	27.3% (24)	0.90	88
C++	**38.9% (35)**	33.3% (30)	27.8% (25)	0.89	90
Java	**78.1% (57)**	17.8% (13)	4.1% (3)	0.26	73
Scripting Languages (Python, PERL, Ruby etc.)	39.1% (34)	**41.4% (36)**	19.5% (17)	0.80	87
If you use another language, please indicate the relative importance of training in it:					7
	answered question				**97**
	skipped question				**22**

9. If your primary programming language is Fortran, please select which (if any) of the following constructs you routinely use in your codes:			
		Response Percent	Response Count
Modules		**92.4%**	61
Generic Procedures		56.1%	37
Operator Overloading		10.6%	7
Derived Types		42.4%	28
Object-Oriented Support		10.6%	7
	answered question		**66**
	skipped question		**53**

According to the survey results, over **90% of respondents claimed their programming language skills are self-taught**; 68% also believe they would benefit from formal training in a given programming language. This significant trend could well be indicating an immediate need for user's to obtain up-to-date training on a given programming language so they are maximising the language features (both fundamental and advanced) as efficiently and effectively as possible.

Fortran 95 was rated the most important programming language followed by FORTRAN77 and C. Interestingly, scripting languages, such as Python, Perl, and Ruby were rated next most important, followed by C++ and then Fortran 2003. Java was rated least important programming language, with 87% of respondents considering it '*not important*'. In addition to the suggested languages, assembly programming was considered by two respondents as being very important, while shell scripting, Matlab, Maple and IDL were also mentioned by at least one respondent.

According to the respondents, **the need for comprehensive formal training is strongest for Fortran 95, Fortran 2003, C and C++ languages**. In each case over 25% of respondents considered their requirements as '*very important*'. Not surprisingly, there is very little desire for formal training in Java and FORTRAN77, however some **60% of respondents considered training in scripting languages as '*somewhat important*' or '*very important*'**. In addition to the options given, a need for formal training was expressed for Matlab, compiler design, IDL and development tools e.g. Autotools, XML, CVS/Subversion[6], by a small number of respondents.

Fortran programmers were asked whether or not they routinely use some of the modern software engineering language constructs available in Fortran95 and 2003 within their codes e.g. *modules*, *generic procedures*, *operator overloading*, *derived types*, and *object-oriented support*. **Only *modules* and *generic procedures* were used by more than 50% of Fortran programmers, whilst object oriented support and operator overloading was selected by only 10%**. These results are probably indicative of the fact that 75% of all survey respondents considered the FORTRAN77 language at least somewhat important in Question 6.

From the responses to this set of questions, the authors believe there is meaningful need for formal training in modern Fortran 95 and/or Fortran 2003 programming skills. It is also interesting to note the increasing importance of scripting languages such as Python.

10. Are you familiar with Partitioned Global Adress Space (PGAS) languages? (PGAS languages provide a shared-memory style programming environment in the context of distributed array structures)

	Response Percent	Response Count
Yes	7.2%	7
No	92.8%	90
answered question		97
skipped question		22

11. Describe your level of experience with the following PGAS languages:

	None	Basic	Advanced	Rating Average	Response Count
Co-Array Fortran	93.8% (91)	6.2% (6)	0.0% (0)	0.06	97
Titanium	99.0% (96)	1.0% (1)	0.0% (0)	0.01	97
UPC	97.9% (95)	1.0% (1)	1.0% (1)	0.03	97
Global Array Toolkit	94.7% (90)	4.2% (4)	1.1% (1)	0.06	95
Other (please specify and describe level of experience)					1
answered question					97
skipped question					22

[6] these topics are independently surveyed in Section 3.5

12. Do you believe your code development and productivity could be enhanced by exploiting a PGAS language?		Response Percent	Response Count
Yes		17.5%	17
No		6.2%	6
Don't know		76.3%	74
	Please justify your response (if possible):		19
	answered question		97
	skipped question		22

Questions 10-12 addressed the state of knowledge with respect to Partitioned Global Address Space (PGAS) languages[7]. **Only 7 of 97 respondents were familiar with PGAS languages in general**. For each specific language assessed – Co-array Fortran, Titanium, UPC, and Global Array Toolkit – over 90% of respondents described their level of experience as '*none*'. Not surprisingly, when asked if they believed that exploiting a PGAS language could enhance their code development and productivity, **76% replied that they did not know**. This result, and the fact that 18% of respondents answered '*yes*', indicates that it is perhaps a lack of knowledge, rather than lack of interest, which is limiting a more widespread adoption of PGAS languages.

13. Are you familiar with the next-generation HPC programming languages currently being developed? (eg. Chapel, X10 or Fortress)?		Response Percent	Response Count
Yes		4.1%	4
No		95.9%	93
	answered question		97
	skipped question		22

[7] PGAS languages are gaining widespread attention as a novel paradigm for developing codes on Terascale and petascale systems

14. Describe your level of experience with the following next generation HPC languages:					
	None	Basic	Advanced	Rating Average	Response Count
Chapel	**100.0% (97)**	0.0% (0)	0.0% (0)	0.00	97
X10	**100.0% (96)**	0.0% (0)	0.0% (0)	0.00	96
Fortress	**96.8% (92)**	3.2% (3)	0.0% (0)	0.03	95
		Other (please specify and describe level of experience)			1
				answered question	**97**
				skipped question	**22**

Questions 13 and 14 assessed familiarity with the next generation HPC programming languages Chapel, X10, and Fortress. **Just four (4) respondents claimed to be familiar with these languages in general**. With regards to specific languages, three considered that they had a basic knowledge of Fortress; no respondent professed to have any knowledge of X10 or Chapel.

This result is not particularly surprising considering that the development of these languages is still in the early stages. It is encouraging to note that some users though are keeping abreast of the cutting-edge in HPC programming techniques and languages. It could also be argued that Questions 10-14 reveal that many technical users are inadequately informed or are unaware of the information channels that deliver news from within the HPC community in relation to the latest developments in HPC programming and techniques[8].

15. Do you implement one-sided communication in your codes to improve performance?			
		Response Percent	Response Count
Yes		**58.8%**	57
No		41.2%	40
		answered question	**97**
		skipped question	**22**

[8] see Question 22 for further details

16. If you do implement one-sided communications, please describe your proficiency with the following low-level communication libraries:					
	None	Basic	Advanced	Rating Average	Response Count
ARMCI	**98.2% (56)**	1.8% (1)	0.0% (0)	0.02	57
SHMEM	**65.1% (41)**	23.8% (15)	11.1% (7)	0.46	63
MPI	12.0% (9)	**57.3% (43)**	30.7% (23)	1.19	75
Other (please specify and decribe level of experience)					2
				answered question	**75**
				skipped question	**44**

One-sided communication for message passing was investigated in Questions 15 and 16. It is encouraging to note that nearly 60% of respondents stated that they use one-sided communication in their codes[9]. Of these the majority claimed at least basic (30% advanced) knowledge of the MPI library. One-third of respondents claimed at least basic (11%, advanced) proficiency with the SHMEM library, whereas the vast majority (98%) of respondents claimed no proficiency with the ARMCI[10] library.

As codes continue to scale to higher core counts it will be become necessary to reduce process synchronisation bottlenecks that are inherent in standard message-passing libraries. The results of Questions 15 and 16 indicate that over half of the respondents are practising this technique. This is an encouraging result and suggests that there isn't an immediate need for the promotion and education of this very important scaling technique.

17. Have you received any formal training in multi-core programming?		
	Response Percent	Response Count
Yes	20.6%	20
No	**79.4%**	77
If possible, please provide specific details (courses, locations, URLs) of training/training materials you found particularly useful:		9
	answered question	**97**
	skipped question	**22**

[9] One-sided communication can be used to optimize code performance by allowing the overlapping of computation and communication and the reduction of process synchronisation

[10] The ARMCI library was developed at Pacific Northwest Lab (PNL) to address the limitations in the SHMEM and MPI-2 one-sided communication libraries

18. If you have undertaken multi-core training, please describe your proficiency in the following techniques:					
	None	Basic	Advanced	Rating Average	Response Count
Multi-core cache optimisation	**55.6% (30)**	37.0% (20)	7.4% (4)	0.52	54
Multi-core memory management	**57.4% (31)**	37.0% (20)	5.6% (3)	0.48	54
Multi-core bandwidth management	**64.8% (35)**	33.3% (18)	1.9% (1)	0.37	54
				answered question	**54**
				skipped question	**65**

Training programmes to exploit multi-core processors were assessed by Questions 17 and 18. **Only one-fifth of respondents had received some formal training in multi-core programming**, and of these 20 respondents, at least 17 described their proficiency as *basic* or *none* in each of cache optimization, memory management, and bandwidth management techniques. Nine respondents provided information regarding training and/or training materials that they found particularly useful, and where possible, this information will be added to the PRACE training materials repository[11]. It should be noted that despite Question 18 being designed to assess proficiency only amongst those who had formal training in multi-core techniques, an additional 34 respondents answered the question to make a total response count of 54. The vast majority of these 34 respondents described their proficiency as *none* for each technique.

As chip manufactures continue to promote dual, quad, 6 and 8-core processors over the next couple of years, it will necessary for scientific programmers to embrace multi-core programming to effectively exploit the hierarchical memory structures of next generation supercomputing architectures. The results obtained in Questions 17 and 18 indicate that training needs in this area will be very important over the next few years, particularly if users wish to fully exploit the performance of future PRACE Tier-0 leadership-class architectures.

3.3 Parallel Programming Methodologies

This section investigated user proficiency and training experience and satisfaction in the parallel programming methodologies that are indicative of current supercomputing architectures and hardware.

[11] See Annex 5.2 for further details.

19. Describe your proficiency in developing highly-optimised codes on the following architectures:					
	None	Basic	Advanced	Rating Average	Response Count
Vector Processors	37.4% (34)	**41.8% (38)**	20.9% (19)	0.84	91
Massively Parallel (MPPs)	17.6% (16)	**49.5% (45)**	33.0% (30)	1.15	91
Symmetric Multiprocessing (SMP)	40.0% (34)	**42.4% (36)**	17.6% (15)	0.78	85
Heterogeneous (mixed-architecture) systems	**60.5% (52)**	32.6% (28)	7.0% (6)	0.47	86
Novel architectures (eg. Cell, FPGA, GPU)	**89.7% (78)**	8.0% (7)	2.3% (2)	0.13	87
			Other (please specify architecture and rate proficiency)		1
				answered question	**91**
				skipped question	**28**

20. Do you believe you have received adequate training to maximise the resources of the architectures you selected above?			
		Response Percent	Response Count
Yes		15.4%	14
No		**84.6%**	77
		Additional comments:	7
		answered question	**91**
		skipped question	**28**

Respondents in this section were initially asked in Questions 19 and 20 to rate their proficiency in developing highly optimized codes for the leading classifications of high-performance architecture: Vector, MPP, SMP, mixed-architecture and novel architectures (e.g. Cell, FPGA). The significant findings were:

1. **For each architecture at least two-thirds of all proficiency ratings were *basic* or *none***
2. **61% of respondents have no proficiency optimizing code for heterogeneous (mixed-architecture systems) and a further 33% described their proficiency as only *basic*.**
3. **90% of respondents have no proficiency optimizing code for novel architectures; seven of 87 respondents rated their proficiency as *basic*, and just two as *advanced*.**
4. **Proficiency is highest for MPP architectures, for which one-third of respondents answered *advanced*, followed by vector processors.**
5. **Only 15% of respondents believe that they have received adequate training to maximize the resources of the architectures mentioned.**

The results from these questions highlight an urgent need for users to obtain satisfactory training and education on code optimisation for high-performance computers. It is quite likely that many of the PRACE HPC Infrastructure systems that will be deployed over the next few

years will contain architectures not dissimilar to those assessed in Question 19. These results indicate that top tier users are relatively inexperienced in developing codes that fully exploit the performance of modern architectures, particularly those that follow a heterogeneous or cell/accelerator based design. Future PRACE training programmes should begin to remedy this imbalance so that users are fully prepared to exploit Tier-0 systems when they become available.

Table 3 outlines some comments that were submitted along with the responses to Question 20 expressing the views of some of the respondents:

- "…not found courses on highly optimized code beyond 'Use Fortran! Use complier flags"
- "…there is not enough training on novel architectures."
- "…trawling the web for multi-core typically turns up Use the compiler! Use OpenMP!"
- "Wiki's, best practice exchange forums, workshops etc… are likely useful as in my experience those who are most skilled in exploiting these machines are too busy to perform formal teaching…and course material is normally too basic."

Table 2: Views of some respondents on HPC architecture optimization

21. In which order would you consult the following channels to find information on parallel programming methodologies and techniques?

	1st	2nd	3rd	4th	5th	6th	Rating Average	Response Count
Internet	**70.3% (64)**	14.3% (13)	6.6% (6)	7.7% (7)	1.1% (1)	0.0% (0)	5.45	91
Face-Face Training Courses/Workshops	11.0% (10)	19.8% (18)	**25.3% (23)**	14.3% (13)	17.6% (16)	12.1% (11)	3.56	91
Online Training Resources	4.4% (4)	19.8% (18)	20.9% (19)	20.9% (19)	**22.0% (20)**	12.1% (11)	3.27	91
Journals	0.0% (0)	5.5% (5)	9.9% (9)	16.5% (15)	27.5% (25)	**40.7% (37)**	2.12	91
Books	13.2% (12)	**30.8% (28)**	20.9% (19)	16.5% (15)	16.5% (15)	2.2% (2)	4.01	91
University/College Courses	1.1% (1)	9.9% (9)	16.5% (15)	24.2% (22)	15.4% (14)	**33.0% (30)**	2.58	91
						Other (please specify)		7
						answered question		**91**
						skipped question		**28**

22. Do you believe you are being sufficiently informed about the latest parallel programming methodologies and techniques?		Response Percent	Response Count
Yes		20.9%	19
No		79.1%	72
	answered question		91
	skipped question		28

In Question 21 respondents were asked to rank the order in which they would consult various channels to find information on parallel programming methodologies and techniques. The channels were ranked in decreasing priority producing as a result the following ordering: the Internet, books, face-to-face training courses/workshops, online training resources, university/college courses, and journals. In addition to these channels several respondents cited discussions with experienced users and specialists. Clearly any PRACE Training and Education programme should exploit these preferences by not only ensuring that high-quality face-to-face training sessions are provided but that all material should be made readily available via online methods along with links to further information sources.

To further support the argument introduced in Section 3.2 (see Questions 10-14) **80% of respondents to Question 22 felt that they were not sufficiently informed about the latest parallel programming methodologies and techniques**. It is recommended that PRACE facilitate the education of users by providing information channels (via the project website, RSS feeds and newsletters) that provides ongoing updates and articles on the latest HPC trends and technological developments within the community.

23. Please rate your understanding of the following fundamental HPC principles:	None	Basic	Advanced	Rating Average	Response Count
Scalability (including distinction between weak and strong)	9.9% (9)	**58.2% (53)**	31.9% (29)	1.22	91
Efficiency	5.5% (5)	**60.4% (55)**	34.1% (31)	1.29	91
Load-balancing	7.8% (7)	**57.8% (52)**	34.4% (31)	1.27	90
Communication/Computation Overlapping	18.9% (17)	**55.6% (50)**	25.6% (23)	1.07	90
Data Decomposition	23.3% (21)	**52.2% (47)**	24.4% (22)	1.01	90
Task Decomposition	25.6% (23)	**54.4% (49)**	20.0% (18)	0.94	90
			answered question		91
			skipped question		28

Question 23 asked respondents to rate their understanding of the following fundamental HPC principles: scalability, efficiency, load-balancing, overlapping of communication and computation, data decomposition and task decomposition. **In each case, the majority understanding of responses was *basic***. Notably, the understanding of data decomposition and task decomposition were rated as *advanced* by only 24% and 22% of respondents

respectively; it is worthy of mention that 57 of the 91 respondents who answered question 23 have six or more years experience in HPC.

Clearly the majority of users (even those who would be regarded as experienced) are not familiar with some of the fundamental principles of HPC programming and practice. This deficiency needs to be addressed through subsequent PRACE training programmes.

24. Describe your competency with the following traditional HPC parallel programming methodologies:

	None	Basic	Advanced	Rating Average	Response Count
Serial (single processor) Optimisation	5.5% (5)	**49.5% (45)**	45.1% (41)	1.40	91
Basic MPI (Point-Point Communication, Collective Communication, Communicators, Topologies)	8.8% (8)	**46.2% (42)**	45.1% (41)	1.36	91
Advanced MPI (One-sided communications, Process Creation, Parallel I/O)	33.0% (30)	**54.9% (50)**	12.1% (11)	0.79	91
OpenMP	41.1% (37)	**44.4% (40)**	14.4% (13)	0.73	90
Hybrid (Mixed-Mode) MPI-OpenMP	**66.3% (59)**	29.2% (26)	4.5% (4)	0.38	89
				answered question	**91**
				skipped question	**28**

25. Please describe the quality of the training you received for the HPC parallel programming methodologies discussed above:

	Poor	Satisfactory	Good	Excellent	None Taken	Rating Average	Response Count
Serial Optimisation	16.5% (15)	20.9% (19)	**28.6% (26)**	5.5% (5)	**28.6% (26)**	1.32	91
Basic MPI	9.9% (9)	18.7% (17)	**31.9% (29)**	9.9% (9)	29.7% (27)	1.59	91
Advanced MPI	20.0% (18)	17.8% (16)	14.4% (13)	4.4% (4)	**43.3% (39)**	1.06	90
OpenMP	23.1% (21)	13.2% (12)	18.7% (17)	2.2% (2)	**42.9% (39)**	1.00	91
Hybrid MPI-OpenMP Programming	31.9% (29)	14.3% (13)	4.4% (4)	0.0% (0)	**49.5% (45)**	0.46	91
If possible, please provide specific details (courses, locations, URLs) of training courses/training materials you found particularly useful:							7
						answered question	**91**
						skipped question	**28**

In Questions 24 and 25 respondents were asked to rate their competency and the quality of training they have received in parallel programming methodologies. The highest competency ratings were given for serial optimization and basic MPI, where nearly half of all respondents

rated their proficiency as *advanced* and nearly all as at least *basic*. In contrast, competency in advanced MPI, and OpenMP, was rated as *advanced* by only 12% and 14%, respectively. **Moreover, two-thirds of all respondents considered that they had no competency in hybrid (mixed-mode MPI-OpenMP) programming whatsoever**.

With respect to the quality of training received, ratings fell between *satisfactory* and *good* in all topics **except hybrid programming, which had an overall rating between *poor* and *satisfactory* and was rated as *poor* by nearly a third of all respondents**. OpenMP and Advanced MPI training were also rated as *poor* by more than 20% of respondents. There were very few ratings of *excellent* across the board. Seven respondents provided information regarding training and/or training materials that they found particularly useful, and where possible this information will added to the PRACE training materials repository (see Section 5.2).

From these results it is clear to see that the quality of training material across the fundamental HPC skill-sets has much-needed room for improvement. Effort must be made to ensure existing PRACE training programmes are of the highest quality both in terms of content and presentation. Furthermore, it maybe necessary to investigate the development of a noteworthy mixed-mode MPI/OpenMP training course as part of PRACE training package.

26. Have you developed or contributed to the development of an application that achieves sustained teraflop/s performance on a given architecture?

	Response Percent	Response Count
Yes	37.4%	34
No	33.0%	30
Not sure	29.7%	27
If yes, please describe the application (if possible):		20
answered question		91
skipped question		28

In Question 26 respondents were asked if they had developed or contributed to the development of an application that achieves teraflop/s performance. Responses were divided roughly evenly between *Yes*, *No* and *Don't Know*. Codes cited included lattice QCD, CASINO, VASP, PWSCF, Conquest and other linear scaling quantum chemistry, classical and ab initio molecular dynamics, hybrid particle-mesh methods, GROMACS, hydrodynamic multi-D, magneto-hydrodynamics, ELMFIRE, codes using BOIC framework, molecular quantum dynamics, Lanczos implementation for Hubbard models, and astrophysical/cosmological computational fluid dynamics.

27. What is the maximum number of cores you have managed to scale a code to?

	Response Percent	Response Count
<32	13.2%	12
32	6.6%	6
64	4.4%	4
128	**17.6%**	16
256	8.8%	8
512	9.9%	9
1024	12.1%	11
2048	6.6%	6
4096	6.6%	6
8192	3.3%	3
16384	5.5%	5
>16384	5.5%	5
Please briefly describe the application (if possible):		33
answered question		91
skipped question		28

The final question in section 2 asked respondents to quantify the maximum number of cores they were successful in scaling a code to. **40% of those surveyed have achieved scaling to 1024 cores or more, but only 11% to 16384 or more**. Codes cited as scaling to 8192 cores or more were lattice QCD codes, a hybrid particle-mesh method for incompressible flows, a path integral ab-initio molecular dynamics code, a Lanczos implementation for Hubbard models, a Quantum Monte Carlo solver for DMFT, and an AMR code for self-gravitating magnetized fluid dynamics.

It is hoped that the recommendations made in this report will result in a PRACE training and education infrastructure that will lead to an increase in the number of users who will develop terascaling and/or petascaling codes in the near future.

3.4 Programming Tools and Libraries

This section covered the use of code development tools and numerical libraries, and the requirements for training in specific classes of numerical computing.

28. Do you currently use version control systems to manage your source code tree?

	Response Percent	Response Count
Yes	**60.9%**	53
No	39.1%	34
	answered question	**87**
	skipped question	**32**

29. If you use version control systems, please describe your understanding of the following tools:

	None	Basic	Advanced	Rating Average	Response Count
SVN	**40.3% (25)**	**40.3% (25)**	19.4% (12)	0.79	62
CVS	17.6% (12)	**58.8% (40)**	23.5% (16)	1.06	68
				Other (please specify)	5
				answered question	**70**
				skipped question	**49**

30. Do you believe you would benefit from formal training on version control systems for your code development?

	Response Percent	Response Count
No	17.2%	15
Yes (probably)	**64.4%**	56
Yes (definitely)	18.4%	16
	answered question	**87**
	skipped question	**32**

31. Do you document your code with source code documentation tools? (e.g. Doxygen, ROBOdoc)

	Response Percent	Response Count
Yes	18.4%	16
No	**81.6%**	71
	answered question	**87**
	skipped question	**32**

32. Do you believe you would benefit from formal training on documentation tools for source code development and maintenance?	Response Percent	Response Count
No	21.8%	19
Yes (probably)	**55.2%**	48
Yes (definitely)	23.0%	20
answered question		**87**
skipped question		**32**

Questions 28-32 assessed the current use of version control systems (e.g. SVN and CVS) and source code documentation tools (e.g. Doxygen, ROBOdoc). More than half of all respondents use version control systems and **83% of respondents believe that they would benefit from formal training in version control systems**. In contrast, source code documentation tools are being used by only 18% of those surveyed, however 23% thought that they would definitely benefit, and a further 55% would probably benefit, from formal training in these tools.

33. Please rate your proficiency with the following numerical libraries:	None	Basic	Advanced	Rating Average	Response Count
BLAS	**47.1% (41)**	33.3% (29)	19.5% (17)	0.72	87
LAPACK	33.3% (29)	**49.4% (43)**	17.2% (15)	0.84	87
PBLAS	**79.3% (69)**	18.4% (16)	2.3% (2)	0.23	87
ScaLAPACK	**65.5% (57)**	29.9% (26)	4.6% (4)	0.39	87
Trilinos	**100.0% (87)**	0.0% (0)	0.0% (0)	0.00	87
PETSc	**92.0% (80)**	4.6% (4)	3.4% (3)	0.11	87
SuperLU	**97.7% (84)**	2.3% (2)	0.0% (0)	0.02	86
FFTW	**59.8% (52)**	28.7% (25)	11.5% (10)	0.52	87
answered question					**87**
skipped question					**32**

34. Please describe the quality of the training you received for (any) of the following libraries:							
	Poor	Satisfactory	Good	Excellent	N/A	Rating Average	Response Count
BLAS	17.2% (15)	9.2% (8)	1.1% (1)	2.3% (2)	**70.1% (61)**	0.62	87
LAPACK	20.7% (18)	11.5% (10)	1.1% (1)	2.3% (2)	**64.4% (56)**	0.58	87
PBLAS	18.4% (16)	8.0% (7)	0.0% (0)	0.0% (0)	**73.6% (64)**	0.30	87
ScaLAPACK	18.4% (16)	8.0% (7)	1.1% (1)	1.1% (1)	**71.3% (62)**	0.48	87
Trilinos	21.8% (19)	3.4% (3)	0.0% (0)	0.0% (0)	**74.7% (65)**	0.14	87
PETSc	22.1% (19)	2.3% (2)	1.2% (1)	0.0% (0)	**74.4% (64)**	0.18	86
SuperLU	21.8% (19)	3.4% (3)	0.0% (0)	0.0% (0)	**74.7% (65)**	0.14	87
FFTW	19.5% (17)	6.9% (6)	2.3% (2)	1.1% (1)	**70.1% (61)**	0.50	87
If possible, please provide specific details (courses, locations, URLs) of training/training materials you found particularly useful:							2
						answered question	**87**
						skipped question	**32**

35. Please indicate your requirements for comprehensive formal training in the following numerical libraries:					
	Not important	Somewhat important	Very important	Rating Average	Response Count
BLAS	**52.9% (45)**	37.6% (32)	9.4% (8)	0.56	85
LAPACK	**47.7% (41)**	41.9% (36)	10.5% (9)	0.63	86
PBLAS	**59.0% (49)**	32.5% (27)	8.4% (7)	0.49	83
ScaLAPACK	**54.2% (45)**	37.3% (31)	8.4% (7)	0.54	83
Trilinos	**79.5% (66)**	16.9% (14)	3.6% (3)	0.24	83
PETSc	**67.9% (57)**	26.2% (22)	6.0% (5)	0.38	84
SuperLU	**74.7% (62)**	19.3% (16)	6.0% (5)	0.31	83
FFTW	**54.8% (46)**	31.0% (26)	14.3% (12)	0.60	84
If you would like training in another library, please specifiy:					0
				answered question	**87**
				skipped question	**32**

Questions 33-35 reveal the proficiency and requirements for training in the use of numerical libraries. **Average proficiency ratings fell between *none* and *basic* for each library**.

Proficiency was rated highest for linear algebra libraries, particularly BLAS and LAPACK, followed by FFTW and ScaLAPACK. **Proficiency in the PETSc and SuperLU libraries was rated particularly low**.

Of those who had undertaken formal training in specific numerical libraries, **the vast majority rated the quality of the training as poor**; the average rating was below satisfactory for every library. The need for formal training was rated highest for LAPACK, FFTW, BLAS, ScaLAPACK and PBLAS.

It is interesting to note that no user had experience with the Trilinos library[12]. In general, there seems to be little formal training undertaken in numerical libraries. The authors are concerned that lack of training could encourage ignorance of highly optimised and robust numerical libraries that are readily available to solve central tasks in numerical computation; one author knows many code developers who invest many hours in developing solvers that are already tested, scalable and available in leading numerical libraries. It is important that PRACE users are acquainted with the standard numerical libraries before they determine whether they need to develop their own custom routines.

36. Describe your understanding of the following seven classes of scientific computing:

	None	Basic	Advanced	Rating Average	Response Count
Dense Algebra	35.3% (30)	**44.7% (38)**	20.0% (17)	0.85	85
Sparse Algebra	35.3% (30)	**43.5% (37)**	21.2% (18)	0.86	85
Structured Grids	28.7% (25)	**41.4% (36)**	29.9% (26)	1.01	87
Unstructured Grids	**54.8% (46)**	32.1% (27)	13.1% (11)	0.58	84
Monte Carlo	19.8% (17)	**43.0% (37)**	37.2% (32)	1.17	86
FFT	16.3% (14)	**62.8% (54)**	20.9% (18)	1.05	86
N-body Calculations	**44.7% (38)**	27.1% (23)	28.2% (24)	0.84	85
				answered question	**87**
				skipped question	**32**

[12] The Trilinos library is object-oriented software framework for the solution of large-scale, complex multi-physics engineering and scientific problems, developed at Sandia National Labs. The Trilinos library received a 2004 R&D 100 Award, given out yearly by R&D Magazine to recognize the "100 most technologically significant products introduced in the past year."

37. For which of the following seven classes of scientific computing would you like to receive formal training?					
	Not important	Somewhat important	Very important	Rating Average	Response Count
Dense Algebra	**51.8% (43)**	36.1% (30)	12.0% (10)	0.60	83
Sparse Algebra	**41.0% (34)**	31.3% (26)	27.7% (23)	0.87	83
Structured Grids	32.9% (28)	**42.4% (36)**	24.7% (21)	0.92	85
Unstructured Grids	34.5% (29)	**40.5% (34)**	25.0% (21)	0.90	84
FFT	40.5% (34)	**44.0% (37)**	15.5% (13)	0.75	84
Monte Carlo	38.1% (32)	**44.0% (37)**	17.9% (15)	0.80	84
N-body Calculations	**51.8% (43)**	33.7% (28)	14.5% (12)	0.63	83
	If you require training in another application class, please specifiy:				1
				answered question	**87**
				skipped question	**32**

Questions 36 and 37 investigated the understanding and training requirements of users with respect to the seven classifications of scientific computing (the *seven dwarves* as defined in [3]). Understanding was highest for Monte Carlo and Structured Grids, where the average rating fell between *basic* and *advanced*, whereas the average rating fell between *none* and *basic* in the other five classes. The need for training was considered as very important by more than 25% of respondents for each of Sparse Algebra, Structured Grids and Unstructured Grids. The majority of respondents considered training at least somewhat important in each class of scientific computing except Dense Algebra and N-body calculations.

3.5 Debugging, Profiling and Optimisation Tools

This section investigates the utilisation of tools and techniques for profiling, optimisation and debugging.

38. Which of the following compiler suites do you regularly use?			
		Response Percent	Response Count
PGI		42.0%	34
Pathscale		16.0%	13
GCC		64.2%	52
Intel		**74.1%**	60
IBM		50.6%	41
Other (please specify)		7.4%	6
		answered question	**81**
		skipped question	**38**

39. Please rate your proficiency with compiler optimisations using the following compiler suites:					
	None	Basic	Advanced	Rating Average	Response Count
PGI	**41.9% (31)**	**41.9% (31)**	16.2% (12)	0.74	74
Pathscale	**73.6% (53)**	22.2% (16)	4.2% (3)	0.31	72
GCC	25.7% (19)	**52.7% (39)**	21.6% (16)	0.96	74
Intel	17.7% (14)	**53.2% (42)**	29.1% (23)	1.11	79
IBM	**43.4% (33)**	**43.4% (33)**	13.2% (10)	0.70	76
			Other (please specify compiler and rate proficiency)		1
				answered question	**81**
				skipped question	**38**

40. Do you believe you would benefit from specialised training material on compiler optimisation techniques?			
		Response Percent	Response Count
No		12.3%	10
Yes (probably)		**55.6%**	45
Yes (definitely)		32.1%	26
		answered question	**81**
		skipped question	**38**

The most commonly used compiler suites are Intel, GCC, IBM and PGI in that order. Also used, but to a much lesser extent are Pathscale, g95, NAG, Sun and MS Visual C++. Of those who regularly use a particular compiler, the most common rating of proficiency was *basic*, with less than a third rating their proficiency as *advanced* in every case.

A significant 88% of respondents considered that they would benefit from specialised training material on compiler optimization techniques.

41. Please rate your proficiency with memory debugging techniques:			
		Response Percent	Response Count
Poor		**46.9%**	38
Satisfactory		35.8%	29
Good		11.1%	9
Excellent		6.2%	5
		answered question	**81**
		skipped question	**38**

42. Do you believe you would benefit from specialised training material on memory debugging techniques?

	Response Percent	Response Count
No	6.2%	5
Yes (probably)	**58.0%**	47
Yes (definitely)	35.8%	29
	answered question	**81**
	skipped question	**38**

43. Please rate your proficiency with with the following parallel debugging techniques:

	None	Basic	Advanced	Rating Average	Response Count
Breakpoints	27.2% (22)	**58.0% (47)**	14.8% (12)	0.88	81
Instruction Stepping	38.8% (31)	**47.5% (38)**	13.8% (11)	0.75	80
Watchpoints	**51.9% (41)**	39.2% (31)	8.9% (7)	0.57	79
Call Stack Examination	**64.6% (51)**	26.6% (21)	8.9% (7)	0.44	79
Observing variables, structures and arrays	39.2% (31)	**48.1% (38)**	12.7% (10)	0.73	79
			answered question		**81**
			skipped question		**38**

44. Please rate your proficiency with with the following debugging tools:

	None	Basic	Advanced	Rating Average	Response Count
Allinea DDT	**94.9% (75)**	3.8% (3)	1.3% (1)	0.06	79
TotalView	**55.6% (45)**	39.5% (32)	4.9% (4)	0.49	81
			Other (please specify tool and rate proficiency)		3
			answered question		**81**
			skipped question		**38**

45. Do you believe you would benefit from specialised training material on parallel debugging tools and techniques?		Response Percent	Response Count
No		7.4%	6
Yes (probably)		**61.7%**	50
Yes (definitely)		30.9%	25
		answered question	**81**
		skipped question	**38**

Questions 41-45 assessed user competencies, training and satisfaction in code debugging. Nearly half of all respondents rated their proficiency in memory debugging as *poor*, and **94% believe that they would likely benefit from specialised training material on memory debugging techniques**. In terms of parallel debugging, average ratings for proficiency fell between *none* and *basic* for each of the listed techniques. Less than 15% rated their proficiency as *advanced* in any of the listed techniques; proficiency was particularly low for call stack examination and watchpoints. Nearly half of all respondents considered that they had at least a *basic* proficiency using the TotalView parallel debugger, but only 5% rated their proficiency as *advanced*.

93% of respondents believed that they would likely benefit from specialised training in debugging tools and techniques.

46. Please rate your proficiency with analysing and diagnosing the following performance areas:	None	Basic	Advanced	Rating Average	Response Count
I/O model and volume measurements	**53.1% (43)**	43.2% (35)	3.7% (3)	0.51	81
Requirements for dynamic data repartitioning	**75.3% (61)**	22.2% (18)	2.5% (2)	0.27	81
Load imbalance	40.7% (33)	**45.7% (37)**	13.6% (11)	0.73	81
Communication patterns (global, local, message sizes, number of messages)	42.0% (34)	**44.4% (36)**	13.6% (11)	0.72	81
Performance bottlenecks and metrics	37.0% (30)	**46.9% (38)**	16.0% (13)	0.79	81
Memory usage (cache misses, TLB misses, latency, bandwidth)	**45.7% (37)**	43.2% (35)	11.1% (9)	0.65	81
Interconnect performance (bandwidth and latency)	**55.6% (45)**	30.9% (25)	13.6% (11)	0.58	81
Indirect addressing	**79.0% (64)**	17.3% (14)	3.7% (3)	0.25	81
Flop rates	**44.4% (36)**	42.0% (34)	13.6% (11)	0.69	81
				answered question	**81**
				skipped question	**38**

47. Do you routinely use performance analysis tools to study your code performance?

	Response Percent	Response Count
Yes	27.2%	22
No	**72.8%**	59
	answered question	**81**
	skipped question	**38**

48. Please rate your proficiency with with the following performance optimisation tools:

	None	Basic	Advanced	Rating Average	Response Count
PAPI	**91.3% (73)**	7.5% (6)	1.3% (1)	0.10	80
Prof/Gprof	**57.5% (46)**	31.3% (25)	11.3% (9)	0.54	80
Intel Trace Analyzer	**81.3% (65)**	18.8% (15)	0.0% (0)	0.19	80
CrayPat and Apprentice2	**82.7% (67)**	12.3% (10)	4.9% (4)	0.22	81
Vtune	**86.1% (68)**	13.9% (11)	0.0% (0)	0.14	79
Vampir	**86.1% (68)**	11.4% (9)	2.5% (2)	0.16	79
Tau	**98.8% (79)**	1.3% (1)	0.0% (0)	0.01	80
Kojak	**97.5% (78)**	1.3% (1)	1.3% (1)	0.04	80
IPM	**98.7% (78)**	1.3% (1)	0.0% (0)	0.01	79
Allinea OPT	**95.0% (76)**	5.0% (4)	0.0% (0)	0.05	80
				Other (please specify)	1
				answered question	**81**
				skipped question	**38**

49. Do you believe you would benefit from specialised training material on performance optimisation techniques?

	Response Percent	Response Count
No	4.9%	4
Yes (probably)	**61.7%**	50
Yes (definitely)	33.3%	27
	answered question	**81**
	skipped question	**38**

50. Do you believe existing training material on performance optimisation techniques is adequate?	Response Percent	Response Count
No	40.7%	33
Yes (probably)	**55.6%**	45
Yes (definitely)	3.7%	3
answered question		**81**
skipped question		**38**

Respondents to Questions 46-50 were asked to rate their proficiency in analysing and diagnosing a number of performance areas such as load imbalance, communication patterns, memory usage (cache/TLB misses), and flop rates. Average proficiency ratings fell between *none* and *basic* in every area, and the most common rating was *none* in each area except load imbalance, communication patterns, and performance bottlenecks and metrics. Nearly three-quarters of respondents do not routinely make use of performance analysis tools, and corresponding, proficiency in using specific tools was extremely limited.

95% of respondents believe they would likely benefit from formal training material on performance optimization techniques, and 40% think that existing training material is inadequate.

51. Please select which of (if any) of the following testing methodologies you routinely apply to your code development:	Response Percent	Response Count
Unit testing (black box,white box)	**66.7%**	20
Integration testing	46.7%	14
Regression testing	40.0%	12
answered question		**30**
skipped question		**89**

52. Do you routinely implement comprehensive exception handling routines in your codes?	Response Percent	Response Count
Yes	24.7%	20
No	**75.3%**	61
answered question		**81**
skipped question		**38**

53. Do you believe you would benefit from specialised training material on code testing techniques?	Response Percent	Response Count
No	9.9%	8
Yes (probably)	**72.8%**	59
Yes (definitely)	17.3%	14
answered question		**81**
skipped question		**38**

The final three questions in this section were related to code testing techniques. Approximately one half of all respondents routinely apply unit testing, integration testing and regression testing, however only 25% implement comprehensive exception handling.

90% of respondents believe that they would likely benefit from specialised training material on code testing techniques.

3.6 Parallel I/O and Fault-Tolerance

This section of the survey investigated user proficiency and requirements for implementing parallel I/O and fault-tolerance within codes.

54. Do you implement parallel I/O in your codes?	Response Percent	Response Count
Yes	38.8%	31
No	**61.3%**	49
answered question		**80**
skipped question		**39**

55. Please rate your proficiency with the following parallel I/O libraries:	None	Basic	Advanced	Rating Average	Response Count
HDF5	**82.3% (65)**	13.9% (11)	3.8% (3)	0.22	79
Parallel NetCDF	**92.4% (73)**	7.6% (6)	0.0% (0)	0.08	79
MPI-IO	**65.0% (52)**	33.8% (27)	1.3% (1)	0.36	80
Other (please specify library and rate proficiency)					2
answered question					**80**
skipped question					**39**

56. Do you believe existing training material on parallel I/O is adequate?	Response Percent	Response Count
No	38.8%	31
Yes (probably)	**60.0%**	48
Yes (definitely)	1.3%	1
answered question		**80**
skipped question		**39**

57. Do you believe you would benefit from specialised training material on parallel I/O techniques?	Response Percent	Response Count
No	20.0%	16
Yes (probably)	**50.0%**	40
Yes (definitely)	30.0%	24
answered question		**80**
skipped question		**39**

Nearly two-thirds of all respondents implement parallel I/O in their codes, however less than 5% rated their proficiency as *advanced* in any of the MPI-IO, HDF5 or Parallel NetCDF libraries. Proficiency in MPI-IO was rated highest, but even for this library the majority of responses were *none*.

80% of respondents believe that they would likely benefit from formal training in parallel I/O techniques, and 40% feel that existing training material is inadequate.

58. Do you implement checkpointing/restart in your codes?	Response Percent	Response Count
Yes	**57.5%**	46
No	42.5%	34
answered question		**80**
skipped question		**39**

59. Do you believe existing training material on checkpointing is adequate?

	Response Percent	Response Count
No	28.8%	23
Yes (probably)	**66.3%**	53
Yes (definitely)	5.0%	4
answered question		**80**
skipped question		**39**

60. Do you believe you would benefit from specialised training material on checkpointing techniques?

	Response Percent	Response Count
No	23.8%	19
Yes (probably)	**62.5%**	50
Yes (definitely)	13.8%	11
answered question		**80**
skipped question		**39**

Checkpointing is implemented in codes by slightly over half of all respondents. 75% of respondents consider that they would benefit from specialised training in checkpointing techniques. The majority believe that existing training material is probably adequate, so it is likely that information transfer is of immediate importance.

3.7 Third-Party Scientific Applications

This section focused on the use of third-party scientific applications, and was answered by all respondents.

61. What is the largest number of cores you have used with a third-party scientific application?

	Response Percent	Response Count
<32	21.1%	20
32	5.3%	5
64	9.5%	9
128	9.5%	9
256	8.4%	8
512	6.3%	6
1024	6.3%	6
2048	3.2%	3
4096	1.1%	1
8192	1.1%	1
16384	0.0%	0
>16384	0.0%	0
N/A	**28.4%**	27
	Application name:	24
	answered question	**95**
	skipped question	**24**

Question 61 was designed to analyse which, if any, third-party codes might be suitable candidates for the Petascale-computing regime. The results showed that very few codes were being run on thousands of processors; three respondents have utilized up to 2048 cores, one respondent up to 4096 cores, and one respondent running VASP code up to 8192 cores.

62. Are you responsible for building and maintaining the third-party scientific applications that you use on HPC resources?

	Response Percent	Response Count
Yes	23.2%	22
No	**76.8%**	73
	answered question	**95**
	skipped question	**24**

63. Do you believe you would benefit from formal training in the use of specific third-party scientific applications, especially with respect to making the best use of high-end HPC resources?		Response Percent	Response Count
No		**48.4%**	46
Yes (probably)		41.1%	39
Yes (definitely)		10.5%	10
	If so, please list the scientific applications here:		10
	answered question		**95**
	skipped question		**24**

Three-quarters of those surveyed were not responsible for the maintenance of third-party applications. Approximately half of those surveyed believe that they would benefit from formal training in the use of specific applications. There were five requests for training in VASP, three in CPMD, and one request for each of CP2K, SIESTA, WIEN2K, CFD, ADF, Ocean, ABINIT, quantum-ESPRESSO, Gaussian, CRYSAL and TURBOMOLE. **It is interesting to note the predominance of materials science codes in the requests for training.**

3.8 Basic Linux/Linux Tools and Techniques for HPC

This section of the survey focused on training requirements with respect to Unix/Linux system commands, shell scripting, build tools and batch job management.

64. Have you received any formal training on fundamental Unix/Linux system commands, shell scripting, editor skills and build tools (e.g. GNU make, autctools etc.)?		Response Percent	Response Count
Yes		29.5%	28
No		**70.5%**	67
	answered question		**95**
	skipped question		**24**

65. Do you believe you would benefit from formal training on fundamental Unix/Linux system commands, shell scripting, editor skills and build tools (eg. GNU make, autotools etc.)?		Response Percent	Response Count
No		32.6%	31
Yes (probably)		**44.2%**	42
Yes (definitely)		23.2%	22
	answered question		**95**
	skipped question		**24**

66. Have you received any formal training on job submission, job scripting and queue management tools?		Response Percent	Response Count
Yes		27.4%	26
No		**72.6%**	69
	answered question		**95**
	skipped question		**24**

67. Do you believe you would benefit from formal training on job submission, job scripting and queue management tools?		Response Percent	Response Count
No		35.8%	34
Yes (probably)		**51.6%**	49
Yes (definitely)		12.6%	12
	answered question		**95**
	skipped question		**24**

70% of the respondents have not received formal training in fundamental Unix/Linux commands, scripting and build tools (Makefiles, autotools etc), but two-thirds feel that they would probably benefit from this training. Trends were very similar with respect to training in job submission, job scripting and queue management tools, with slightly fewer respondents answering that they would definitely benefit from formal training.

3.9 Distributed Systems Access and Grid Middleware

This section focused on training requirements with respect to distributed systems and Grid middleware.

68. Do you have a Grid certificate?		Response Percent	Response Count
Yes		26.3%	25
No (but I know what a Grid Certificate is)		24.2%	23
No (and I don't know what a Grid certificate is)		**49.5%**	47
	answered question		**95**
	skipped question		**24**

69. Are you member of a Grid Virtual Organization (VO)?		Response Percent	Response Count
Yes		18.9%	18
No (but I know what a VO is)		20.0%	19
No (and I don't know what a VO is)		**47.4%**	45
Don't know		13.7%	13
		answered question	**95**
		skipped question	**24**

Respondents were first asked in Questions 68 and 69 about their experiences with Grid certificates and virtual organisations (VOs). About one quarter of those surveyed have a Grid certificate, and a further quarter know what a Grid certificate is but do not possess one. 20% are members of VOs, and a further 20% know what VOs are but are not members. The majority either do not know what a VO is or are not sure whether they belong to one.

70. Please rate your proficiency with the following Grid Middleware stacks:	None	Basic	Advanced	Rating Average	Response Count
UNICORE	**87.4% (83)**	10.5% (10)	2.1% (2)	0.15	95
LCG/gLite	**95.7% (90)**	4.3% (4)	0.0% (0)	0.04	94
Globus GT4	**84.0% (79)**	16.0% (15)	0.0% (0)	0.16	94
NorduGrid ARC	**93.6% (88)**	6.4% (6)	0.0% (0)	0.06	94
			Other (please specify middleware and rate proficiency)		0
			answered question		**95**
			skipped question		**24**

Question 70 asked respondents to rate their proficiency with specific Grid middleware stacks, namely UNICORE, Globus Toolkit, LCG/Lite and NorduGrid ARC. **Over 84% of all respondents rated their proficiency as *none* for each middleware stack**. Only two respondents rated their proficiency as advanced, and in both cases this was for the UNICORE middleware.

71. Have you received any formal training on Grid computing from the Grid/middleware provider organization (e.g., EGEE, NorduGrid, DEISA)?		Response Percent	Response Count
Yes		9.5%	9
No (there were no such events I was aware of)		35.8%	34
No (I didn't attend training events)		**54.7%**	52
		answered question	**95**
		skipped question	**24**

72. Have you received any formal training on Grid computing from your VO?		Response Percent	Response Count
Yes		5.6%	5
No (there were no such events I was aware of)		40.0%	36
No (I didn't attend training events)		**54.4%**	49
		answered question	**90**
		skipped question	**29**

Just one tenth of all respondents have received training in Grid computing from their Grid middleware provider organisation, and 6% from their VO. Over one third of respondents were unaware of training events organised by these groups.

73. Do you believe you would benefit from specialized training material on distributed computing and Grid access?		Response Percent	Response Count
No		33.7%	32
Yes (probably)		**54.7%**	52
Yes (definitely)		11.6%	11
		answered question	**95**
		skipped question	**24**

74. Do you believe existing training material on Grid access is adequate?	Response Percent	Response Count
No	28.4%	27
Yes (probably)	**69.5%**	66
Yes (definitely)	2.1%	2
answered question		**95**
skipped question		**24**

75. If you believe existing Grid Access training is NOT adequate, then please describe what topics you think are not adequately covered by existing training material? (Tick all applicable items):	Response Percent	Response Count
Grid job submission and control	**82.1%**	23
Data management	67.9%	19
Programming/adapting code for use on the Grid	71.4%	20
Other (please specify)	3.6%	1
answered question		**28**
skipped question		**91**

Two thirds of all respondents believe that they would likely benefit from specialised training material on distributed computing and Grid access. The majority believe that existing training material is probably adequate, but for those who do not, most considered there is a need for improved training material on job submission and control, data management, and programming/adapting code for the Grid.

3.10 Scientific Visualisation

This section focused on user training requirements and satisfaction in the field of scientific visualisation.

76. Are you a user of scientific visualization tools?	Response Percent	Response Count
Yes	**65.3%**	62
No	34.7%	33
answered question		**95**
skipped question		**24**

Two-thirds of those surveyed responded that they are users of scientific visualisation tools, and they were required to answer the questions in this section. The remaining users were skipped directly to the last section of the survey.

77. Describe your proficiency in using the following scientific visualization tools:

	None	Basic	Advanced	Rating Average	Response Count
Client-server applications	33.9% (20)	**54.2% (32)**	11.9% (7)	0.78	59
Parallel Visualization applications	**75.0% (45)**	21.7% (13)	3.3% (2)	0.28	60
CPU-GPU Coupled Computations	**86.7% (52)**	13.3% (8)	0.0% (0)	0.13	60
Remote Image Delivery	**86.7% (52)**	13.3% (8)	0.0% (0)	0.13	60
Batch-mode Parallel Animation Generation?	**95.0% (57)**	5.0% (3)	0.0% (0)	0.05	60
Multi-wall Displays (CAVES etc.)	**91.7% (55)**	8.3% (5)	0.0% (0)	0.08	60
				Other (please specify tool and rate proficiency)	5
				answered question	**60**
				skipped question	**59**

Respondents were first asked to rate their proficiency in using various visualisation tools and techniques. The average ratings for proficiency fell between *none* and *basic* for all categories, with use of client-server applications rated highest (12% answered *advanced*). The most weakness was displayed in the areas of batch-mode parallel animation generation, the use of multi-wall displays (CAVES etc.), CPU-GPU coupled computations, and remote image delivery.

78. Which of the following area(s) do you perceive as bottlenecks to large-scale visualization?

	Response Percent	Response Count
Polygon Rendering	29.3%	12
Volume Rendering	48.8%	20
Data I/O	**63.4%**	26
Feature Extraction	26.8%	11
Other (please specify)	7.3%	3
	answered question	**41**
	skipped question	**78**

79. Please describe the quality of the training you received for the visualization areas discussed above:							
	Poor	Satisfactory	Good	Excellent	None Taken	Rating Average	Response Count
Polygon Rendering	31.7% (19)	8.3% (5)	1.7% (1)	0.0% (0)	**58.3% (35)**	0.28	60
Volume Rendering	30.0% (18)	11.7% (7)	0.0% (0)	0.0% (0)	**58.3% (35)**	0.28	60
Data I/O	31.7% (19)	15.0% (9)	3.3% (2)	0.0% (0)	**50.0% (30)**	0.43	60
Feature Extraction	32.2% (19)	6.8% (4)	3.4% (2)	1.7% (1)	**55.9% (33)**	0.42	59
If possible, please provide specific details (courses, locations, URLs) of training courses/training materials you found particularly useful:							0
						answered question	**60**
						skipped question	**59**

Question 78 asked users which areas are perceived as bottlenecks to large-scale visualisation. Data I/O was ranked first, followed by volume rendering, polygon rendering and feature extraction.

Question 79 further asked respondents to rate the quality of training they had received in each area. **In each case the vast majority rated the training they had received as *poor*, and no respondents were able to provide details of training courses or materials that they found particularly useful. Of a total of 106 individual ratings there were only five ratings of *good*, and a single rating of *excellent*.**

80. Describe your understanding of the following fundamental visualization principles:					
	None	Basic	Advanced	Rating Average	Response Count
Screen-space Partitioning	**78.3% (47)**	20.0% (12)	1.7% (1)	0.23	60
Object-space Partitioning	**83.3% (50)**	15.0% (9)	1.7% (1)	0.18	60
Sort-Last Rendering	**86.7% (52)**	11.7% (7)	1.7% (1)	0.15	60
Level-of-detail Rendering	**78.3% (47)**	20.0% (12)	1.7% (1)	0.23	60
Multi-threaded Rendering	**83.3% (50)**	16.7% (10)	0.0% (0)	0.17	60
Other (please specify and rate proficiency)					0
				answered question	**60**
				skipped question	**59**

81. Please indicate your requirements for comprehensive formal training in the following fundamental visualization principles:					
	Not important	Somewhat important	Very important	Rating Average	Response Count
Screen-space Partitioning	43.1% (25)	**50.0% (29)**	6.9% (4)	0.64	58
Object-space Partitioning	44.1% (26)	**49.2% (29)**	6.8% (4)	0.63	59
Sort-Last Rendering	46.6% (27)	**48.3% (28)**	5.2% (3)	0.59	58
Level-of-detail Rendering	41.7% (25)	**46.7% (28)**	11.7% (7)	0.70	60
Mult-threaded Rendering	**44.8% (26)**	41.4% (24)	13.8% (8)	0.69	58
If you would like training in another technique, please specify:					1
				answered question	**60**
				skipped question	**59**

Question 80 and 81 asked respondents to rate their understanding and training requirements for a series of fundamental visualisation principles. **In each case over 75% of respondents described their understanding as *none***, and there was at most a single response of *advanced*. The majority considered training somewhat or very important for each visualisation principle, but there was no clear requirement for training in any one particular area over the others.

82. Which of the following tools do you use?			
		Response Percent	Response Count
Paraview		25.6%	11
VisIT		18.6%	8
AVS		14.0%	6
OpenDX		25.6%	11
EnSight		7.0%	3
Other (please specify)		**55.8%**	24
		answered question	**43**
		skipped question	**76**

83. Please indicate your requirements for comprehensive formal training in the following tools:					
	Not important	Somewhat important	Very important	Rating Average	Response Count
AVS	**67.8% (40)**	25.4% (15)	6.8% (4)	0.39	59
OpenDX	**64.4% (38)**	25.4% (15)	10.2% (6)	0.46	59
Paraview	**57.6% (34)**	30.5% (18)	11.9% (7)	0.54	59
VisIT	**70.7% (41)**	20.7% (12)	8.6% (5)	0.38	58
EnSight	**71.9% (41)**	22.8% (13)	5.3% (3)	0.33	57
If you would like training in another technique, please specify:					2
				answered question	**60**
				skipped question	**59**

Around a third of respondents consider their requirements for formal training in a specific tool as being either somewhat or very important, with OpenDX and Paraview rating the highest.

3.11 Training Requirements

The final section of the survey investigated user preferences with respect to the delivery and presentation of training materials.

84. Rate your overall current satisfaction with the formal HPC training you have received:			
		Response Percent	Response Count
Poor		18.3%	17
Satisfactory		**46.2%**	43
Good		21.5%	20
Excellent		1.1%	1
N/A		12.9%	12
		answered question	**93**
		skipped question	**26**

The first question in this section asked respondents to rate their overall satisfaction with the formal training they had undertaken in HPC. Of those who had undertaken formal training, over half rated the training as satisfactory, 25% as good and 21% as poor. **Just a single respondent gave a rating of excellent**.

85. Do you believe there is a need for improved HPC training programmes?		
	Response Percent	Response Count
Yes	**89.2%**	83
No	10.8%	10
	answered question	**93**
	skipped question	**26**

86. Do you believe your local HPC Centre provides adequate training content?		
	Response Percent	Response Count
Yes	47.3%	44
No	**52.7%**	49
	answered question	**93**
	skipped question	**26**

90% of surveyed users considered that there is a need for improved HPC training programmes, and over half do not think that their local HPC centre provides adequate training content.

87. Would you make use of a pan-European centralised repository of high-quality training material that is regularly updated and available to European HPC users?		
	Response Percent	Response Count
Yes	**88.2%**	82
No	11.8%	11
	answered question	**93**
	skipped question	**26**

Nearly 90% of respondents said that they would make use of a pan-European centralised repository of high-quality training material that was be regularly updated and available to European HPC users.

88. What mode of traditional training do you find most suitable when learning new HPC techniques and methodologies (please rank):									
	1st	2nd	3rd	4th	5th	6th	7th	Rating Average	Response Count
Face-Face Classes	**33.3% (31)**	16.1% (15)	14.0% (13)	11.8% (11)	12.9% (12)	6.5% (6)	5.4% (5)	5.04	93
Electronic Slides	7.5% (7)	18.3% (17)	**21.5% (20)**	17.2% (16)	15.1% (14)	15.1% (14)	5.4% (5)	4.19	93
Online web tutorials	**31.2% (29)**	19.4% (18)	14.0% (13)	20.4% (19)	9.7% (9)	5.4% (5)	0.0% (0)	5.26	93
Interactive computer-based training courses	11.8% (11)	16.1% (15)	17.2% (16)	11.8% (11)	**21.5% (20)**	14.0% (13)	7.5% (7)	4.13	93
Reading User Guides	9.7% (9)	17.2% (16)	19.4% (18)	**21.5% (20)**	**21.5% (20)**	6.5% (6)	4.3% (4)	4.35	93
Books	6.5% (6)	12.9% (12)	12.9% (12)	10.8% (10)	14.0% (13)	**35.5% (33)**	7.5% (7)	3.51	93
Journals	0.0% (0)	0.0% (0)	1.1% (1)	6.5% (6)	5.4% (5)	17.2% (16)	**69.9% (65)**	1.52	93
								Other (please specify)	0
								answered question	**93**
								skipped question	**26**

89. Rank in order of preference the following novel modes of training that you would like to see supplement traditional training methods:							
	1st	2nd	3rd	4th	5th	Rating Average	Response Count
Virtual learning environments	**55.1% (49)**	11.2% (10)	13.5% (12)	10.1% (9)	10.1% (9)	3.91	89
Multi-cast remote training (Access Grid classes)	3.4% (3)	**31.8% (28)**	12.5% (11)	28.4% (25)	23.9% (21)	2.63	88
Live web-broadcasts with teleconference audio	14.8% (13)	11.4% (10)	**36.4% (32)**	15.9% (14)	21.6% (19)	2.82	88
Flash Documentation	20.2% (18)	22.5% (20)	15.7% (14)	**23.6% (21)**	18.0% (16)	3.03	89
Screencasting (narrated software movies)	10.2% (9)	**23.9% (21)**	21.6% (19)	20.5% (18)	**23.9% (21)**	2.76	88
						Other (please specify)	4
						answered question	**93**
						skipped question	**26**

Question 88 and 89 investigated preferences for traditional and novel modes of training when learning new HPC techniques and methodologies. **Online web tutorials and face-to-face classes were rated first and second**, followed by user guides, electronic slides, interactive computer-based courses, books, and journals. In terms of novel modes of training, **virtual learning environments was rated first**, followed by Flash documentation, live web-

broadcasting, screen casting, and Access Grid classes. **It should be noted that several respondents suggested that none of these novel modes of training were at all suitable.**

90. Are you a member of the following professional technology bodies?

	Response Percent	Response Count
ACM	33.3%	1
IEEE	0.0%	0
Other (please specify)	**66.7%**	2
	answered question	**3**
	skipped question	**116**

Professional technology bodies provide members with access to vast repositories of learned information and publications regarding their profession. Many bodies organise local chapters which meet regularly, and host workshops on new technogolgies and methodologies. The two largest professional computing bodies (ACM and IEEE) are a valuable resource for all members of the computing community. In this survey only one user stated that they were a member of such an organisation.

91. Do other commitments limit your ability to attend face-face training courses (eg. teaching commitments)?

	Response Percent	Response Count
Yes	**67.7%**	63
No	32.3%	30
	answered question	**93**
	skipped question	**26**

92. When do you find it easiest to attend face-face training courses? Please rank in order of ease.

	1st	2nd	3rd	4th	Rating Average	Response Count
Summer	**42.9% (39)**	5.5% (5)	8.8% (8)	**42.9% (39)**	2.48	91
Autumn	27.8% (25)	**31.1% (28)**	24.4% (22)	16.7% (15)	2.70	90
Winter	22.5% (20)	**34.8% (31)**	24.7% (22)	18.0% (16)	2.62	89
Spring	9.0% (8)	30.3% (27)	**40.4% (36)**	20.2% (18)	2.28	89
				answered question		**93**
				skipped question		**26**

93. Do you dislike travelling to undertake face-face training programmes?		Response Percent	Response Count
Yes		46.2%	43
No		**53.8%**	50
		answered question	**93**
		skipped question	**26**

94. What is your preferred duration for attending a face-face training course?		Response Percent	Response Count
1 day		9.7%	9
2 days		25.8%	24
3 days		**47.3%**	44
4 days		5.4%	5
5 days		8.6%	8
>5 days		3.2%	3
		answered question	**93**
		skipped question	**26**

Questions 91-94 sought to gain an understanding of the potential barriers to training. Two-thirds of respondents were limited in their ability to attend face-to-face training courses by other commitments, and nearly half dislike travelling to undertake face-to-face training. Autumn was ranked as the most convenient season to attend face-to-face training courses, followed by winter, summer and spring. **However, it is important to note that summer was rated as most convenient and least convenient by the highest number of respondents. Regarding the duration of face-to-face training courses, nearly half of all respondents considered three days as optimal, with a further 25% preferring two-day courses. Just 3% had a preference for training courses longer than 5 days.**

The final question of the survey asked respondents to describe the training they considered particularly urgent, and which training they anticipate that they will require in the future.

4 Conclusions and Recommendations

The data presented in Section 3 was obtained from the most comprehensive evaluation of user training requirements and satisfaction undertaken across top European HPC users. The principal rationale for this survey is to provide recommendations which will steer future training and education programmes within the PRACE HPC Research Infrastructure. In this section the authors present their recommendations for training needs based on the requirements identified in Section 3.

To facilitate a comprehensive roadmap which incorporates both the user requirements and PRACE project resources[13], we present our recommendations within a progressive framework of fulfillment. This approach provides a clear structure for the implementation of a sustained and high-quality education and training programme that will begin to remedy the significant user requirements identified in this survey. Our recommendations are classified into immediate requirements (those requirements that we believe should be implemented at the earliest opportunity), short term requirements (those requirements we believe should be developed by the end of the preparatory PRACE project) and long-term requirements (those requirements that while identified as important within the context of this survey, may not necessarily indicate an urgent need). We also provide a section on general comments that are not determined by time but reflect general principles that should be applied when possible.

4.1 Immediate Requirements

i. A significant need for mixed-mode (hybrid) programming was identified during the survey. Over two-thirds of respondents indicated that they had no competency in hybrid programming techniques. A third of respondents revealed that the level of training for hybrid programming techniques was poor.
ii. An initiative should be undertaken to promote knowledge of Partitioned Global Address Space (PGAS) languages via articles and other dissemination means. 93% of users were unfamiliar with this continuously important parallel programming approach.
iii. 80% of users believed they were inadequately informed about the latest developments in parallel programming and high-performance computing. The PRACE HPC training infrastructure should ensure that users are regularly informed, via the PRACE website, articles and comprehensive links to the best online HPC information resources.
iv. An initiative should be undertaken to promote knowledge of numerical libraries via articles and other publications. Proficiency and awareness of standard numerical libraries was rated as *low* among most users.
v. High-quality training material on Visualization should begin to be developed. The vast majority of respondents claimed that Visualization training they received had been poor.

4.2 Short-Term Requirements

i. Formal training courses on modern Fortran programming should be delivered. Over half of respondents indicated that this was a very important training requirement in competition with other traditional HPC languages. This course should also cover more

[13] In particular those resources available within WP3 for training and education

sophisticated constructs of the language that promote modern software engineering principles.

ii. The presentation of training material on code optimization, debugging tools and code testing is highly recommended. In each case approximately 90% of users indicated that they would benefit from training in these specialized fields.

iii. Content on version control systems and parallel I/O was sought after by 80% of users. 66% of respondents indicated that they implemented parallel I/O yet the vast majority had basic or no competence in parallel I/O techniques. 40% of respondents believed existing training material was inadequate. Clearly, the PRACE education and training infrastructure should seek to improve the availability and quality of these important HPC tools.

iv. Over 70% of respondents would benefit from training material in standard Unix/Linux and HPC system skills. A comprehensive PRACE training programme should ensure that users have competency with the fundamental skills and tools for interfacing with high-performance computing architectures.

v. Only one-fifth of respondents had received some formal training in multi-core programming, and of these 20 respondents, at least 17 described their proficiency as basic or non-existent. For users to exploit the complex memory hierarchies of modern multi-core processors it is imperative that training materials be made available to the community which explains this complex topic.

4.3 Long-Term Requirements

i. Users should be kept informed about developments in next-generation HPC programming languages such as Chapel, X10 and Fortress. From the survey it was revealed that 96% of users were unfamiliar with these technologies.

ii. 60% of respondents considered training in scripting languages as somewhat or very important. While the full benefit of scripting languages in the HPC domain remains to be determined it is important that these technologies are not ignored in favour of mainstream languages. PRACE education and training personnel should continue to monitor the impact of scripting languages within the HPC community and develop appropriate training material when required.

iii. The emergence of Grid technology as a fundamental component of a distributed computing infrastructure will require users to become familiar with Grid tools to exploit the work-flow of the PRACE HPC Research Infrastructure. It is important that appropriate training material is developed for PRACE users as the technology matures. Two thirds of all respondents believe that they would likely benefit from specialised training material on distributed computing and Grid access.

4.4 General Requirements

i. Users recommended that face-to-face training should be delivered by experts in the field of the given HPC topic area. This would influence their willingness to attend face-to-face training sessions and their confidence in the transfer of knowledge. PRACE education and training programmes should ensure they adopt the best educators and renowned experts in the corresponding HPC field, during the delivery of face-to-face schools and workshops.

ii. Face-to-face training sessions were ranked the most important channel for training delivery. The PRACE education and training programme should certify that high-quality training schools and workshops are developed and maintained. Training

material should also be made widely available online via the PRACE website and training materials repository.

iii. 90% of surveyed users considered that there is a need for improved HPC training programmes. There was no training course, designated as excellent, by any user. This suggests that more effort is required to develop and maintain excellent high-quality training material and content.

iv. 95% of all respondents agreed that they would benefit from a pan-European centralized training repository. It is important that PRACE training and education infrastructure implements and maintains such a repository of high-quality world-class training content and material.

v. 50% of users were not adequately served by their local HPC centre. The PRACE training and education programme should work with local centers to ensure that training needs and expectations are met for all users.

vi. Users recommended that face-to-face training sessions should have an ideal duration of 3 days. There is little agreement on the best time throughout the year to host such events.

Final Note:

As of writing, the programme committee for the PRACE Summer School (to be held in August 2008), has already made significant progress in addressing many of the requirements set out in Sections 4.1-4.4. In particular, this school has a major focus on mixed-mode (MPI-OpenMP) training, along with further lectures on performance optimisation and porting. Further information can be found at [6]. It is planned that the Winter School in Greece (2009) will continue to further address the requirements set out in this document.

References and Applicable Documents

[1] http://www.prace-project.eu
[2] PRACE - Grant Agreement Number: RI-211528 - Annex I: Description of Work (currently not publically available)
[3] P. Colella, “Defining Software Requirements for Scientific Computing,” 2004.
[4] Horst D. Simon. Petascale Computing in the U.S. ACTS Workshop, August 2006
[5] http://surveymonkey.com/s.aspx?sm=_2f7D2YUJfkH9Dua0PP_2f30_2fQ_3d_3d
[6] http://www.pdc.kth.se/systems_support/training/2008/PRACE-P2S2/