



CONTRACT NUMBER 508830

DEISA
**DISTRIBUTED EUROPEAN INFRASTRUCTURE FOR
SUPERCOMPUTING APPLICATIONS**

European Community Sixth Framework Programme
RESEARCH INFRASTRUCTURES
Integrated Infrastructure Initiative

SA4 Third Annual Activity Report

Deliverable ID: DEISA-DSA4-5
Due date: April 30, 2006
Actual delivery date: May 25th, 2007
Lead contractor for this deliverable: IDRIS-CNRS, France

Project start date: May 1st, 2004
Duration: 4 years

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Table of Content

Table of Content.....	1
1. Executive Summary.....	2
2. Introduction.....	3
3. DEISA Common Production Environment.....	4
4. Software monitoring for the DEISA Common Production Environment.....	6
5. Support of scientific Joint Research Activities.....	8
5.1 JRA 1 (Material Sciences).....	8
5.2 JRA 2 (Cosmology Applications).....	8
5.3 JRA 3 (Fusion Research).....	8
5.4 JRA 4 (Life Sciences).....	8
5.5 JRA 5 (Industrial CFD).....	9
5.6 JRA 6 (Coupled Applications).....	9
6. DEISA Training Sessions.....	11
7. Other activities.....	12
7.1 Trouble Ticket System.....	12
7.2 DEISA Cluster Resource Management Package.....	12
7.3 DEISA Test Users.....	12
7.4 User documentation.....	12
8. References and Applicable Documents.....	14
9. List of Acronyms and Abbreviations.....	15

1. Executive Summary

The main objective of this service activity is to deploy all the actions required to allow scientific users to adopt and use the DEISA supercomputing infrastructure. This is mainly done by providing and maintaining a *Common Production Environment* on all the platforms of the infrastructure, and also providing documentation on its usage, the organization of training sessions in various European countries and a decentralised *Help Desk* service.

This document gives a detailed overview of all the tasks achieved or currently undertaken in the *User Support and Applications Service Activity* (SA4) during the reporting period (PM 31 to PM 36). This constitutes the report of the third annual activities, in complement to the D-SA4-4.2 deliverable [8] which covered the first six months.

This document is publicly available.

2. Introduction

During the last six months, the *Applications and User Support Service* Activity has mainly concentrated its efforts on the following tasks:

- Maintenance of the *DEISA Common Production Environment* (DCPE).
- Migration to the completely new version 2.0 of the INCA software, used to monitor all of our software environment.
- Continued provision of help to the users of the scientific *Joint Research Activities*.
- Organization of the third DEISA training session and preparation of the fourth one.

3. DEISA Common Production Environment

As explained in the previous deliverables, the unified *DEISA Common Production Environment* (DCPE) is a major feature in DEISA, as in all distributed infrastructures of this kind, which defines a coherent set of software accessible on the various sites of the infrastructure. It offers both a common interface to the users, independent of the target platform really used, and the ability to migrate jobs between different supercomputers of the same architecture. Of course, the level of coherence is not the same everywhere in the infrastructure, ranging from very high inside each subgroup of homogeneous computers to a lower level across the other subgroups.

Three main components define the DCPE:

- a coherent set of software packages divided into six categories: environment (which does not include any software, but defines the environment needed by a DEISA user), shells, compilers, libraries, tools and applications;
- a uniform interface to access the software, provided by the *Modules* software [18];
- a framework to monitor the software (this will be described in section 4).

As reported in the deliverables D-SA4-3.2 [6] and D-SA4-4 [7], the DCPE has run in full production mode on all the seven AIX sites for over a year, and for the core sites for more than 18 months. During the last period, the only modifications were some minor updates, according to the commonly agreed evolutions of the software packages. These updates were to both the software stack and the *Modules* environment, as required.

As reported in the deliverable D-SA4-4.2 [8], two non AIX sites (SARA which operates a SGI Altix system and BSC an IBM/PowerPC supercomputer under a Linux operating system) were also integrated during the second year. SARA reached the full production mode status at the end of the second year and BSC during the preceding reporting period (PM 25 to PM 30).

On its side, LRZ prepared during the second year its flavour of the DCPE on a Linux cluster used as a test system, integrating the DEISA *Modules* environment as a special case of their general *Modules* environment, which they decided, according to the DEISA promoted way, to use as the standard interface for all their users to access all the software available on their system. This is why they designed and built, in the year preceding the installation of their awaited new supercomputer, a complete *Modules* environment, which includes the DEISA one. They installed this environment in pre-production mode during the summer of 2006 when their new SGI Altix system was installed. During the last months, they consolidated their environment and were able to use it in full production mode.

We can therefore say that since several months ten sites operate the DCPE in full production mode. As initially planned in the two layer infrastructure of DEISA, there is, as required, a very strong homogeneity of the software stack inside each common set of platforms (IBM AIX and SGI Altix), with only minor discrepancies on the patch level of the software included, and a weaker one across the different set of platforms. Nevertheless, there is a general coherence and many common items. A complete list of the components of the software stacks for the three architectures currently supported can be found in the Primer [15].

The CVS server which has been set up at IDRIS to distribute the configuration files, allowing each partner to easily find the changes made since their last update and to download only the modified files, continued to be heavily used.

Our main objective now is obviously to keep a stable framework. The only changes foreseen are those required by the evolutions of the software stacks, according to both the evolution of some software when new versions are released and to requirements for the inclusion of a few new software packages, especially when they will be used by some DECI projects on several sites. Nevertheless, we recently introduced a change in the way that we handle some third party applications, to be conformant with their specific license management. The issue is that some software (for instance the CMPD and NAMD applications used in molecular dynamics research) are freely usable, without any fee, for an academic usage but however require that each user has personally signed a license (most often an electronic signature on a Web page is enough), in particular to declare that he will not redistribute the software by himself. To take care of this problem we require that all our DEISA users must now notify that they have signed the appropriate license if they want access to one of these applications. For this purpose we have introduced a mechanism in the corresponding *modulefiles* to check that the user has explicitly made the declaration that they have signed the license.

It must be also noticed that the evolution of the hardware infrastructure of some partners (CSC with a CRAY XT4 system and SARA with an IBM Power5 system under Linux) will require the port and installation of the DCPE. The second case is foreseen to be reasonably straightforward as it is very close to what has already been done, however the first case will be a complete new one, with known technical difficulties due to the CRAY software environment management. This issue obviously needs to be carefully studied in the near future.

Otherwise, as previously stated, in the second half of 2007 we plan to make a deep analysis of the behaviour of our framework to benefit from all of the accumulated experience, and to look at this time for possible improvements for its usage in the future, in both its design as well as in its implementation. Some preliminary informal discussions and exchanges have already started.

4. Software monitoring for the DEISA Common Production Environment

It is an absolutely necessity that each centre's staff have both an up to date overview and a detailed status of the software environment, especially in a distributed infrastructure like DEISA. The first feature required in tools which provide this overview is the ability to verify the accessibility of the various software packages installed on the different computers, the expected version level of these packages and also, when possible, their current behaviour. It is also important that these tools allow the verification of the status after software upgrades and offer the administrators and the user support services an updated view of what is installed and available on the various computers of the distributed infrastructure, alerting them of possible deficient components.

As explained in the deliverable D-SA4-3 [5], the INCA tool (*Test Harness and Reporting Framework*) [17], developed by the San Diego Supercomputer Center and the Argonne National Laboratory for the TeraGrid project [20], has been chosen after an evaluation process. As described in the previous deliverables (D-SA4-3.2 [6], D-SA4-4 [7] and D-SA4-4.2 [8]), this framework was initially deployed on the AIX core sites to monitor the availability of the software of the AIX DCPE with the required version levels (with what is called *version reporters*), then extended to the other sites in the first half of 2006. In parallel, another kind of sensor, called *unit reporters*, have been deployed and operated, first on the core sites then later on the others, to check periodically the correct behaviour of the critical software components. They are based on the *DEISA Applications Test Suite* (DATS), described in the deliverables D-SA4-4 [7] and D-SA4-4.2 [8].

The monitoring of the DEISA software environment has been fully operational for a long time now, progressively integrating all the sites. Nevertheless, a large step forward has been expected for nearly a year. As explained in the deliverables D-SA4-4 [7] and D-SA4-4.2 [8], SDSC have been working since the middle of 2005 on a completely new version of INCA, rewritten from scratch, which will have many improvements and advantages. It has an improved architecture, new functionalities and enhanced customization opportunities. The main changes have already been described in the previous deliverables.

The migration to this new version was first scheduled in the summer of 2006, but needed to be postponed several times, first in December 2006 then in April 2007. The first reason was that SDSC postponed several times the release of the first *production* version, initially planned around May/June 2006, but eventually only being made available in February 2007. In the meantime, we have obviously tested and experimented with the numerous *Alpha* then *Beta* versions in the testbed that we set up in March 2006 between LRZ, IDRIS and CINECA, after the dedicated workshop organized at SDSC in February 2006. However, we decided at the very beginning not to switch our production environment to the new version before the availability of the first official version, in the meantime obviously continuing to use the previous version 1.0.

The second reason was that our leading expert on INCA at LRZ decided to change their job and to leave LRZ at the end of 2006, which prevented us from being able to perform the migration as soon as the production version had been released. A new INCA administrator arrived at LRZ last March and, allowing him the time to learn the various related issues required by the INCA framework, we started to operate the INCA 2.0 framework in production at the end of April. This will be fully deployed on all

the sites in the first weeks of May (but it must be noted that the initial framework will still be operated and usable until the migration is completely accomplished).

In addition to the *version reporters* and to the *unit reporters* related to the *DEISA Applications Tests Suite*, we also introduced at this time the *unit reporters* for one of the other *Test Suites* developed in the various service activities of DEISA. In particular, SA3 developed a dedicated test suite for the IBM LoadLeveler batch manager, the *DEISA LoadLeveler Tests Suite* (DLLTS), and it was decided several months ago that the best way to operate this would be to integrate it in the INCA framework. This will ensure a regular and automatic run of the tests, and will also offer a convenient way to collect and consult their results. This was prepared during the first months of 2007, after that SA4 developed the *unit reporters* required to run each of these tests, and it is now operated in the INCA 2.0 framework.

We have also prepared the possibility, as required for those partners that have particular security constraints (half of them), to install and operate a local *Reporter Manager*. This feature was not initially expected in the INCA software, and required important efforts from us in order for it to be set up and tested on our testbed.

Some enhancements are planned in the coming months, in particular to the *data consumer* component of INCA, which is the presentation layer of the results extracted from the data base filled by the *reporters*. We have already identified some important improvements that we would like to see, and that we will have to develop in the next period.

It must be emphasized that the close relationship with the INCA development team, established for more than two years, has been very beneficial for the two sides. The direct contact with the project manager and the developers enables us to receive a very fast and efficient support, both during our experiments on our testbed and then later during the installations on the production systems. And, in return, we have helped them to improve the portability and robustness of their software, especially providing them with feedback and problem reports from our diverse platforms and operating systems, which is particularly important since they made all their developments in a Linux environment.

5. Support of scientific Joint Research Activities

All the projects included in these scientific JRAs have had a direct and strong connection with one of the core partners (RZG for JRA 1 and 3, IDRIS for JRA 4 and 6, CINECA for JRA 5).

5.1 JRA 1 (*Material Sciences*)

In JRA 1, work was focused during the reporting period on the full integration of the simulation codes LAMMPS and NAMD (see the deliverable D-JRA1-6 [9]). Both software packages were already included in the DCPE during the last period and the required *modulefiles* were developed (see the deliverable D-SA4-4.2 [8]), and recently these two packages were also fully integrated, using dedicated plug-ins, into the Web application of the materials science portal. Otherwise, according to a problem reported by the LIAMS DECI project which discovered a major problem in the NAMD software initially distributed in our DCRMP distribution (see paragraph 7.2), a new version of the software was built, which solved the problem using a more recent version of the Charm++ parallelisation library, and the DCRMP distribution was updated.

5.2 JRA 2 (*Cosmology Applications*)

JRA 2 finished on PM 30 (October 2006).

5.3 JRA 3 (*Fusion Research*)

During the reporting period, the work of JRA 3 (see deliverable D-JRA3-6 [10]) was first devoted to the finalization of the enabling work of the EUTERPE plasma physics code, now used in the DECI project Gyro3d. In parallel, a new code, GEM, has been selected and analysed in detail. After the identification of several bottlenecks, various enhancements have been made and the single processor performance has been significantly improved. However, a large amount of work is still planned to get an efficient parallelisation level of this code.

Additionally, various enhancements were done to the already supported codes TORB, ORB5 and GENE. For the latter two, improvements were made to their scalability, and impressive results were obtained, especially for the GENE code.

5.4 JRA 4 (*Life Sciences*)

During the last six months, JRA 4 has been continuing its efforts in the comprehension and handling of the needs of the life sciences community. For this purpose, SA 4 was asked to update and upgrade in the relevant sites the software that it had already made available and to keep track of and test the newly parallelized applications. For example, the latest version of NCBI'S BLAST (v2.2.15) is running on the IDRIS IBM P690+ supercomputer twice as fast as the previous version.

JRA 4 is also collaborating in the realization of the DEISA life sciences portal by the eDEISA activities. Concerning the first version of the portal, three applications have been selected: NCBI's BLAST (genomics), NAMD (molecular dynamics simulations),

and RAxML (phylogeny). These applications have been installed and tested by the user support teams on the three architectures that will be part of the first implementation of the portal, respectively at the IDRIS (IBM SP4), the BSC (IBM PPC) and the LRZ (SGI Altix). The generic account on the DEISA infrastructure has been set up, with its new extended *Modules* environment including the *modulefiles* for the applications accessible with the portal, and now the first version of the interface realized by NICE is awaited. The NAMD and RAxML applications are now fully available to the users of the three sites. For details, see the deliverable D-JRA4-7 [11]).

After the first version of the portal, JRA 4 will extend the number of software packages available through the portal to all of the good candidates from the existing toolbox. It may also ask for help from SA 4 to test the new applications and their access through the portal, and to promote the portal in the user community.

5.5 JRA 5 (Industrial CFD)

JRA 5 is involved in modelling for aero-acoustic (CAA) and aerodynamic (CFD) simulations for some relevant automotive applications. The software used is CFD++ from Metacomp, a commercial package. The support activities to the JRA, during the last semester of the project, were mainly related to the following (see the deliverable D-JRA5-7 [13] for more details):

- changes to the installation and configuration of the CFD++ license servers in order to make them reachable from the internal DEISA network (see the deliverable D-JRA5-6 [12]),
- submission of production cases for infrastructure assessment.

5.6 JRA 6 (Coupled Applications)

During the reporting period, the JRA 6 activities have been mainly devoted to continuing the production of the coupled applications of the second set and to achieving the latest optimisations. This current set of coupled projects finished at the end of the reporting period (April 2007). For more details, see the deliverable D-JRA6-6 [14].

We have had to face several blocking or/and embarrassing problems during this period. They have been solved with the help of the *DEISA user support* teams at IDRIS and HLRS. Here are the most important:

- The production of the coupling DECI project (called FOCUS) went on running large coupled applications until the end of 2006. Particular attention was paid to manage the CPU resources optimally during this period and required, with the DEISA user support help, to continuously survey the coupled application driven by end users. In parallel, changing the coupling technology used in this application was planned and the DEISA user support groups also helped us to install different MPI versions and to obtain an optimal configuration to test the new coupling layer.
- The 3D Natural Convection / Radiation coupled application has been numerically validated and intermediate configurations have been launched during this reporting period. In this project, the DEISA user support group at IDRIS helped us to build a more convenient context to launch coupled applications.

- A new optimisation phase has been achieved into the coupling module called KOP3D and in the way to distribute the sub-domains. During some weeks, daily contacts have been made to solve problems related to the porting of the PACX-MPI coupling technology, and to teach users unfamiliar with them particular AIX features such as compiler options.

6. DEISA Training Sessions

Six training sessions were scheduled between mid-2006 and early 2008 in different European countries (in France, Germany, Spain, Finland, Italy and Germany). The purpose of the training is to enable fast development of user skills and know-how needed for the efficient utilisation of the DEISA infrastructure.

Each session (except the first one, which did not have the second part) is divided in three parts:

- The first part gives a global description and introduction to the DEISA infrastructure and describes the general middleware services, the usage of the *DEISA Common Production Environment*, and the detailed utilisation of UNICORE and DESHL.
- The second part consists of a *hands-on* session, which allows each user to experiment by himself with the main services of the infrastructure (Global File Systems, UNICORE, DESHL, DCPE).
- The third part is dedicated each time to a different HPC topic, itself usually divided in two parts, one technical and one based on *use cases*, to give pedagogical summaries of experiments in different scientific fields, mostly coming from completed DECI projects.

The attendance is always limited to 30 participants and DEISA takes responsibility for the travel and living expenses of up to 15 participants travelling from outside the country which hosts the event.

The first session, organized by IDRIS, occurred in Paris in July 3-6 and the second one, organized by FZJ, was held in Jülich in October 23-25 (see deliverable D-SA4-4.2 [8]).

During the reporting period, the third session occurred in Barcelona in March 7-9, organized by BSC. It was a full success, with 33 European scientists, 6 staff members and 9 speakers, with again a fully positive feedback (see the deliverable D-eN2-1 [16]). The presentations were well appreciated, both in the general and specialized parts, and there was a general agreement amongst participants to emphasize the importance of the hands-on session, where people can really experiment by themselves with what was previously explained. As this was very similar to the feedback received after the second session, we improved and extended the exercises given to the participants in all areas, including UNICORE (with a new workflow exercise and also one dedicated to the opportunities offered by the global file systems), DESHL (with various sub-exercises), and we also extended the required usage of the DCPE.

Also, following the practise first introduced during the second session in October, each participant to a training session will now become a *DEISA Test User* (see paragraph 7.4) and will have a DTU account opened to continue to experiment with the infrastructure for a period of three months after the session that he attended.

During the reporting period, the fourth session was prepared, which will be organized by CSC and will occur in Helsinki from May 30th to June 1st. The two remaining sessions are scheduled in October 2007 in Bologna (organized by CINECA) and in February 2008 in Stuttgart (organized by HLRS).

7. Other activities

7.1 *Trouble Ticket System*

As explained in the deliverable D-SA4-4 [7], the centralised *Trouble Ticket System*, based on the *Request Tracker* [19] software, set-up and operated for the DEISA purposes at RZG, and now used as a centralised service for all the DEISA service activities, especially the user support and operations teams, obviously continued to be used.

7.2 *DEISA Cluster Resource Management Package*

The *DEISA Cluster Resource Management Package* (DCRMP) helps the administrators of the DEISA sites in the installation and maintenance of their software environments, both for system and user oriented components. This activity, led by CINECA as an SA3 one, has already involved SA4 because, according to the work already achieved in SA3 to distribute some middleware components and the potential interest of it for SA4 activities, it was decided to use it also for some application oriented packages (see the deliverables D-SA4-4 [7] and D-SA4-4.2 [8]).

Using the work done in SA3 for this task, especially building the dedicated tool called *DCRMP Package Creator* (DPC) (see the deliverables D-SA3-5 [1] and D-SA3-7 [2]), several software packages included in the AIX DCPE were packaged and distributed in this way, to lower for the partners the deployment time to install these packages. The latest software handled in this way were two public domain application packages in the Molecular Dynamics field (NAMD and LAMMPS). A few months ago, the LIAMS DECI project which heavily used NAMD detected a major problem in the executable generated and distributed, which failed on one of their data sets. After investigation, it was found that the problem occurred in the underlying Charm++ parallelisation library used, and a more recent version of it was installed to generate a working version of the NAMD executable, included in an updated DCRMP distribution of this software.

7.3 *DEISA Test Users*

As explained in the deliverable D-SA4-4.2 [8], it was decided by the *DEISA Executive Committee* in the summer of 2006 to offer a new kind of access to the infrastructure, to allow any European scientist to test specific services like the global file systems and UNICORE, and to enable evaluation of the new opportunities that the DEISA infrastructure can offer to them, encouraging them to answer the next calls for DECI proposals. For this purpose, a new category of users, the *DEISA Test Users* (DTU), was defined and set up, and in particular all the participants to the training sessions now have an account of this kind opened, starting from the second session last October. This was obviously done too for the third session in March.

7.4 *User documentation*

The current public version (1.4) of the *Primer* [15] is available since March 2007. Compared to the previous version 1.3, the major improvements were first to fully describe the situation of the SGI Altix and Linux/PowerPC sites, with some additional information about the characteristics of the usage of their own environments inside the DEISA infrastructure and especially the specificities of their own DCPE, and

second to give the small amount of specific information useful for the new kind of DEISA users recently introduced, the *DEISA Test Users*.

8. References and Applicable Documents

- [1] DEISA D-SA3-5 deliverable: *Second release of DCRMP*
- [2] DEISA D-SA3-7 deliverable: *Third release of DCRMP*
- [3] DEISA D-SA4-1 deliverable: *SA4 Service Definition and Operation*
- [4] DEISA D-SA4-2 deliverable: *Basic DEISA Infrastructure Documentation*
- [5] DEISA D-SA4-3 deliverable: *First SA4 Annual Report*
- [6] DEISA D-SA4-3.2 deliverable: *SA4 Mid-term Activity Report*
- [7] DEISA D-SA4-4 deliverable: *SA4 Second Annual Activity Report*
- [8] DEISA D-SA4-4.2 deliverable: *SA4 Mid-term Activity Report*
- [9] DEISA D-JRA1-6 deliverable: *Support of a set of important material sciences codes with grid-enabled features*
- [10] DEISA D-JRA3-6 deliverable: *Final report on EUTERPE enabling and further new applications*
- [11] DEISA D-JRA4-7 deliverable: *Annual report and update of work program*
- [12] DEISA D-JRA5-6 deliverable: *Assessment of DEISA Grid Functionality*
- [13] DEISA D-JRA5-7 deliverable: *Medium heavy test cases results: engineering and IT results*
- [14] DEISA D-JRA6-6 deliverable: *Final report on technological and scientific impact of the second set of projects. Presentation of the third set of projects selected.*
- [15] DEISA Primer Documentation: <http://www.deisa.org/userscorner/primer.php>
- [16] eDEISA D-eN2-1 deliverable: *First Outreach Activity Report*
- [17] INCA (Test Harness and Reporting Framework): <http://inca.sdsc.edu/>
- [18] Modules: <http://modules.sourceforge.net/>
- [19] Request Tracker: <http://www.bestpractical.com/rt>
- [20] TeraGrid: <http://www.teragrid.org/>
- [21] UNICORE (UNiform Interface to COmputing REsources): <http://unicore.sourceforge.net/>

9. List of Acronyms and Abbreviations

ATaskF	Applications Task Force
CNRS	Centre National de la Recherche Scientifique
CPMD	Car-Parrinello Molecular Dynamics
DATS	DEISA Applications Test Suite
DCPE	DEISA Common Production Environment
DCRMP	DEISA Cluster Resource Management Package
DECI	DEISA Extreme Computing Initiative
DESHL	DEISA Services for the Heterogeneous Management Layer
DLLTS	DEISA LoadLeveler Test Suite
DPC	DCRMP Package Creator
DSU	DEISA Standard User
DTU	DEISA Test User
Execution site	The site where the jobs of a user are executed
Home site	The site where a user usually works, logs in and submits jobs
JRA	Joint Research Activity
NAMD	NAnoscale Molecular Dynamics
RMIS	Resource Management Information System
SDSC	San Diego Supercomputer Center
TTS	Trouble Ticket System