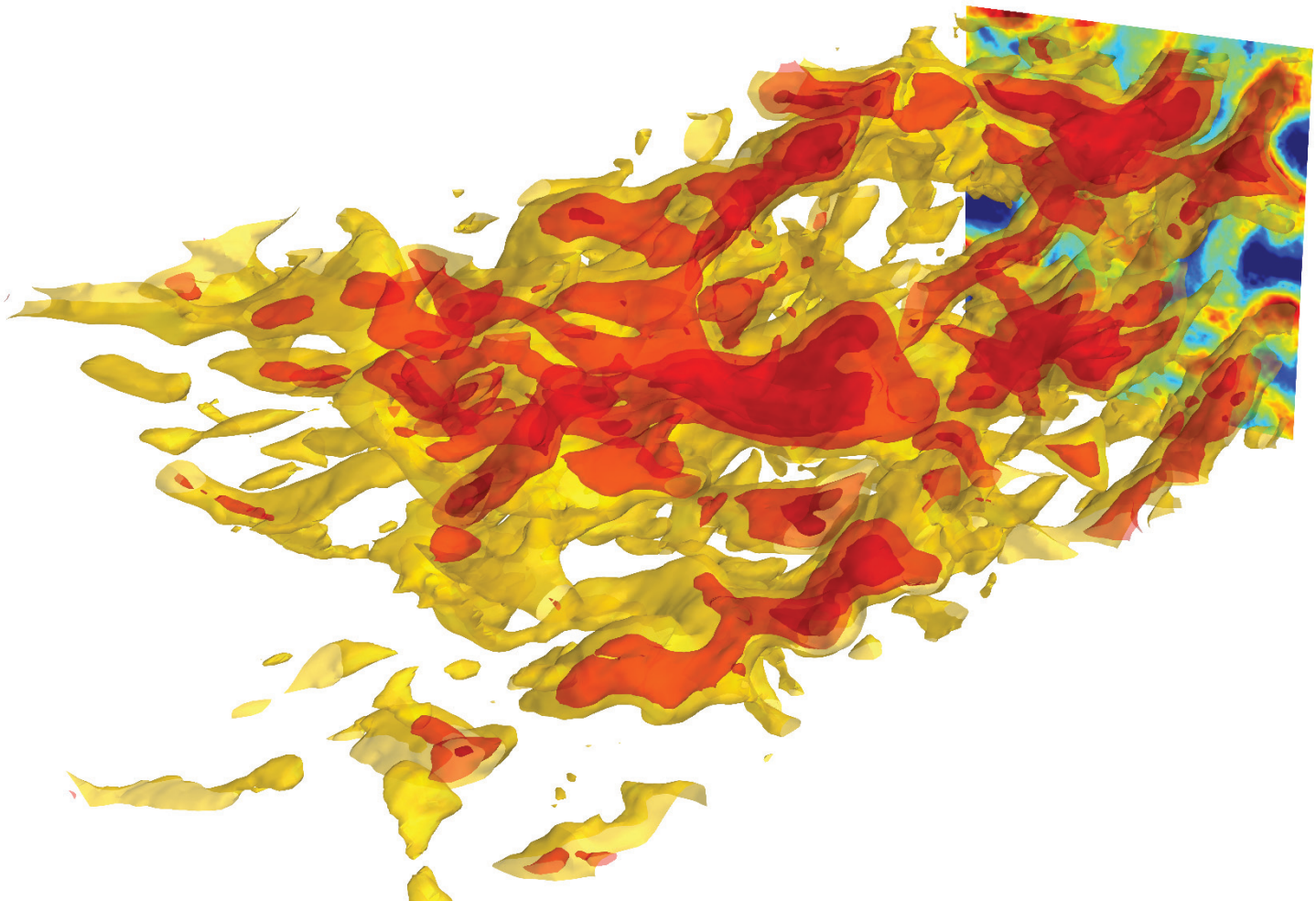
PDC Center for
High Performance Computing

no:1 2013

PDC Newsletter



Editorial

Erwin Laure,
Director PDC & HPCViz

Published by PDC at CSC, KTH

PDC operates leading-edge, high-performance computers as easily-accessible national resources. These resources are primarily available for Swedish academic research and education. PDC, which is hosted by KTH, is one of the six centres in the Swedish National Infrastructure for Computing (SNIC).

Editor: Erwin Laure
Editorial staff: Genet Edmondson
Layout: Maria Malmqvist & Genet Edmondson
Graphic Design: Maria Malmqvist
E-mail: pdc-newsletter@pdc.kth.se
ISSN 1401-9671

Cover

Simulating fusion

The image shows the time-evolution of a cross-section of the electrostatic potential, from the linear regime at the onset of a simulation, to fully developed turbulent dynamics. This is based on data from non-linear gyrokinetic simulations.

HPC architectures are becoming ever larger and more powerful, but at the same time more complex to use due to their increasingly heterogeneous design and the sheer number of processing elements reaching over a million cores. It has thus been widely recognized that serious efforts in software improvements, both on the algorithmic and programming levels, are needed to be able to use current HPC systems efficiently. A recent e-IRG report provides a good overview of this issue (for more information, see www.e-irg.eu/news/news/478/e-irg-policy-paper-on-scientific-software-published.html).

PDC has a long history of working with its users to improve their code base and is contributing to several large projects in this field. The ScalaLife project focuses on life science software, including GROMACS, DALTON, and DISCRETE – read about recent developments for Dalton, and a CECAM workshop that ScalaLife organized, in this newsletter. Significant scaling efforts are also being made in the context of the PRACE projects and we report on the results achieved by working with the computational mechanics code “Alya” in this issue. The CRESTA project is exploring paths towards exascale and PDC is particularly working with GROMACS and NEK5000 in this project – more about these efforts in the next issue of our newsletter.

Increasing computational power also opens up new opportunities in fusion research and our cover article by Andreas Skyman and his colleagues discusses the usage and prospects of HPC in this exciting field.

But high-speed computing is only one side of the coin – dealing with the ever-increasing amounts of scientific data is becoming an equally important problem in the scientific community. In this issue we describe how the EUDAT project is helping various communities with their data management problems. One of the underlying technologies used by EUDAT is iRODS, a rule-based federated data management system, which is now being provided by PDC and is also under evaluation as a suitable technology for SweStore. Read more about the functionalities and opportunities provided by iRODS in this issue.

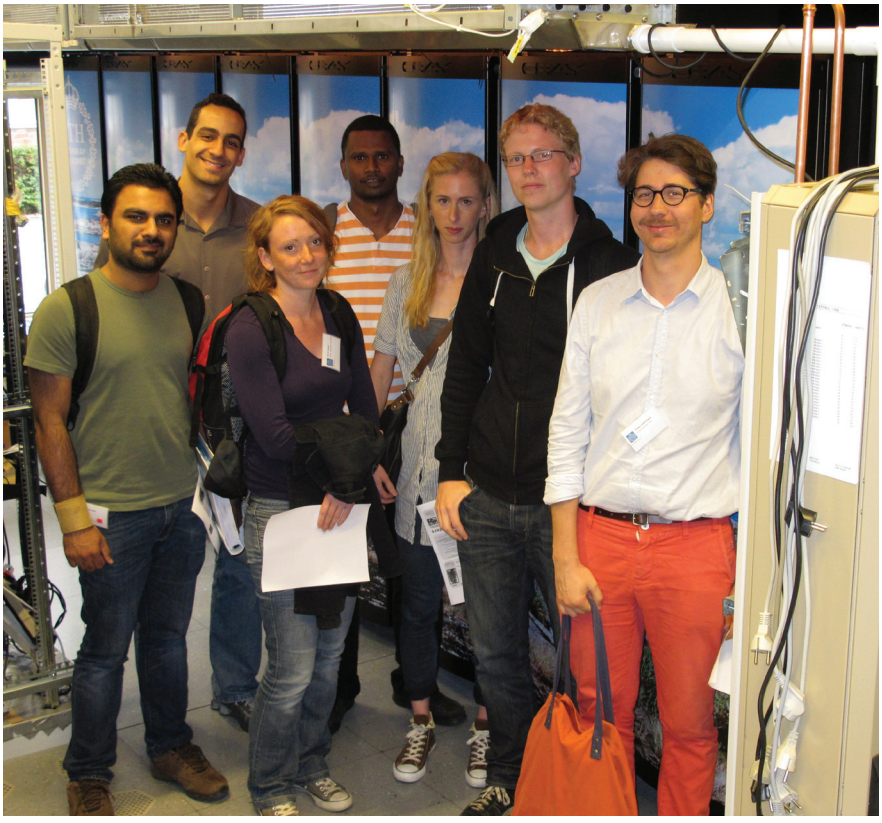
From coding and data, we move to considering hardware, where the latest news is that PDC’s own hardware base has undergone significant re-arrangements in the past couple of months. After Ekman reached its end-of-service in December 2012, we gradually replaced

our old Ferlin hardware with hardware "recycled" from Ekman so that now all of Ferlin is composed of more powerful hardware. We also re-purposed our energy-efficiency prototype Povel to become a fully-fledged pre- and post-processing system for Lindgren. All Lindgren users automatically get allocations on Povel for these purposes.

On the Nordic level, a new organization has been formed to foster collaborations on e-infrastructure: the Nordic e-Infrastructure Collaboration (NeIC). PDC's Michaela Barth has recently been appointed as coordinator of NeIC's generic technologies area and provides an overview of NeIC in this issue.

Finally, our cloud group at PDC recently started a SNIC-funded project to provide the bioinformatics workflow system "Galaxy" to Swedish bioinformatics researchers as a platform as a service (PaaS). We look forward to hearing more about this project after the summer – in the meantime, enjoy your (hopefully cloud-free) summer holidays!

Erwin Laure, Director PDC and HPCViz



PDC Summer School 2012 - Visit to the PDC computer hall

In This Issue

How to Contact PDC

Visiting address:
Teknikringen 14, "Plan 4",
KTH, Stockholm
Mail: PDC, KTH,
SE-10044 Stockholm, Sweden
E-mail: support@pdc.kth.se
www: http://www.pdc.kth.se
Phone: +46 8 790 7800
Fax: +46 8 247 784

Staff Focus

Ann Seares

Ann Seares recently joined PDC as a department administrator. She was already well-acquainted with KTH's administrative systems from her previous work at the Department of Media Technology and Interaction Design. Ann grew up in the Philippines and graduated with a Bachelor of Science in Business Administration from the University of San Jose-Recolletos. Before coming to work at KTH, Ann worked in Taiwan at the ASUS company.

Ann is a very keen gardener, and has lots of orchids in her house - she has extended her administrative duties to include taking good care of the PDC pot plants! (We wonder if Ann and Gert can come up with a project to use excess heat from the supercomputers to warm a greenhouse and grow mangoes.) Ann also enjoys dancing, and playing golf with her family in the summer.

Turbulence, Fusion and Clean Energy

by Andreas Skyman, Hans Nordman and Pär Strand, Department of Earth and Space Sciences, Chalmers University of Technology

Fusion 101

In 1926 Sir Arthur Eddington published his treatise *The Internal Constitution of the Stars*, the first comprehensive work on fusion, and with its publication the vision of fusion as a power source was kindled. Since then, taming the nuclear furnace and bringing the power of the Sun to Earth has been the ambition of generations of physicists and engineers. With the ITER experiment (www.iter.org) planned for 2020, the goal seems within reach, appropriately around the centennial of Sir Arthur's theory.

Fusion relies fundamentally on the same physical principle as fission: transmuting elements, in such a way that the resulting elements are nearer to iron, which yields a net excess of energy. In fusion processes, however, lighter and normally stable elements are fused to form heavier elements. Fusion reactions release vastly higher amounts of energy per nucleon than fission reactions and the fuel is available globally, anywhere on Earth. Fusion thus has an enormous potential as a clean and environmentally friendly power source. Although fission occurs spontaneously for several radioactive elements found on Earth, fusion requires more exotic conditions: intense pressure or fantastically high temperatures. In the Sun, pressures far beyond those that have been achieved on Earth facilitate the fusion of protons, but the most favourable route to fusion for power production relies on the latter alternative, namely extreme temperatures. In the largest currently operational fusion experiment, the *Joint European Torus* (JET), temperatures of 100,000,000 K are routinely achieved, which is what is necessary to efficiently fuse the hydrogen isotopes deuterium (D) and tritium (T). JET and ITER are examples of *tokamaks*, which are the most widely researched class of devices for controlling the fusion-plasma using magnetic confinement.

Turbulence in tokamaks

The dynamics of fusion plasmas are generally divided into two categories: stability and transport. In this context, stability refers to global-scale dynamics, affecting the whole bulk of the fuel plasma, whereas transport encompasses smaller-scale phenomena, acting within the plasma. The global modes can often be understood and studied using Magneto Hydro Dynamics (MHD), whereas the latter require more advanced physical models, closer to the first principles of plasma dynamics as described by the kinetic Boltzmann–Vlasov equation. Although stability is a prerequisite for magnetic confinement fusion, when it comes to deciding the broad character

of the feasible operating scenarios for future fusion power plants, transport is no less crucial and is a much harder problem.

A common measure of the quality of the confinement is the *energy confinement time*. It is defined as the quotient of the energy content E and the power input P_{in} needed to sustain the plasma at that level of energy: $\tau_E = E/P_{in}$, which has the dimension time. This is a measure of how quickly the energy would be lost, if power were not supplied. For the temperature regimes valid for fusion, the condition for power break-even is summarised by the condition that the fusion triple product: $n_i T_i \tau_E$ supersedes a critical value. This points to three avenues available for increasing the efficiency of fusion devices, however, two of those are severely constrained. The optimal ion temperature (T_i) is defined by the cross-section of the fusion process, and the ion density (n_i) cannot be increased beyond a certain value for reasons of global stability. The route ahead therefore lies mainly in increasing the confinement time, which is closely related to understanding and controlling transport.

Transport in a tokamak plasma is generally turbulent. It originates from a multitude of small-scale instabilities, which are driven by the free energy available in the steep gradients of temperature and pressure within the plasma. The effects of this turbulence on confinement are complex and highly non-linear and – crucially – turbulent transport is unavoidable. However, if transport can be better understood, complex non-linear phenomena, such as internal transport barriers, may be controlled to counter the adverse effects of turbulence.

Understanding transport in fusion plasmas, difficult though it may be, is a very promising prospect. An increase of the confinement time would relax the condition on the density value of the fusion triple product, meaning that a smaller machine with a weaker magnetic field could produce the same net power output. A doubling of τ_E would mean that a machine with a lesser volume would suffice, with a corresponding decrease in the required monetary investment.

Plasma turbulence modelling

One characteristic of turbulence is the involvement and interaction of a multitude of time and length scales. In a plasma, long-range electromagnetic forces add a new category of phenomena to the picture, making the already numerically-stiff problem yet more difficult.

For small-scale turbulence – the kind that is considered responsible for transport in fusion plasmas – the situation is typically that a plasma instability is feeding free energy (from global gradients in temperature or density) to smaller scales. The wave numbers and frequencies involved are particular to the instability at hand. If there is a positive feedback, the induced small-scale perturbations turn turbulent, cascading energy to higher and/or lower wave numbers. Several such mechanisms can co-exist, some of the most common being the ion and electron temperature gradient modes (ITG and ETG) and the trapped electron mode (TEM). Though these are all plasma-specific instabilities, they are in many ways analogous to the fluid dynamic Rayleigh–Taylor instability, where a denser fluid is supported by a lighter one under the influence of an effective gravity.

The separation of magnitudes, scales and frequencies between driving instabilities has often been exploited with great success, but reduced models are too simplistic for a realistic model of a functioning fusion power plant. The non-linear interactions between modes and scales are crucial for both predictive capability and theoretical understanding.

The argument of scale, however, is still very important. Different models are appropriate for studying different phenomena, but at the core of the matter, a more simplistic model is only as valuable as its agreement with more fundamental models. Within the fusion community, therefore, models are routinely benchmarked not only against experiments, but also against first principle models.

A common feature used in fully non-linear models of fusion plasmas is that ions and electrons are confined to move along magnetic field

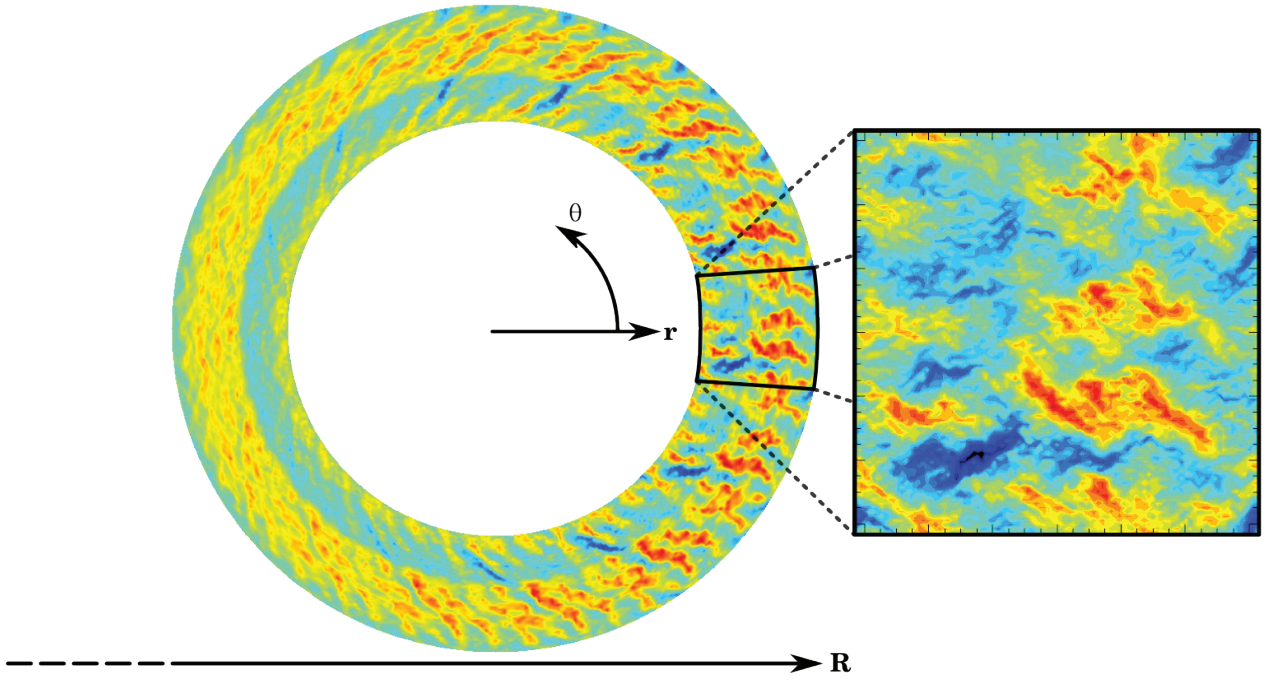
lines, and their gyrating motion around the field lines is small and fast compared to the transport phenomena of interest. This leads to the gyrokinetic approach, which reduces the phase space of the fundamental problem from six to five dimensions. Even so, the calculations for a well-resolved first principle simulation (with impurities and full realism) still requires around 300,000–500,000 CPU hours and several gigabytes of shared memory.

It's not all hydrogen

The plasma in a fusion device consists mainly of hydrogen isotopes or, in some experiments, of helium. There are, however, many other species of atoms in a plasma. These are subjected to the same turbulent transport mechanisms, but affect the confinement of the plasma in widely different ways.

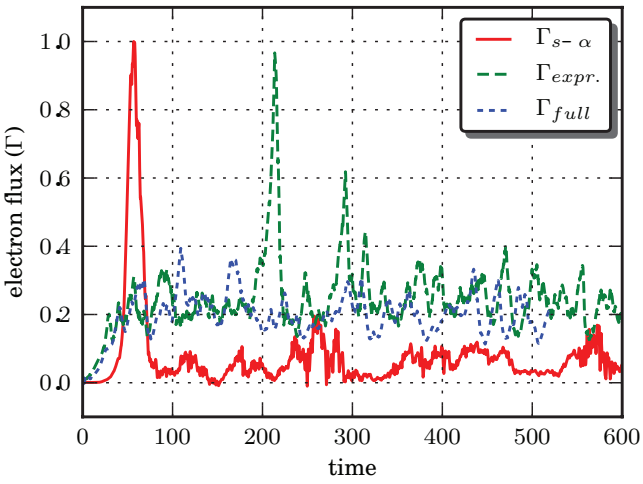
There are three main sources of impurities: the first being the walls of the reactor chamber. Due to the different roles played by different parts of the walls, they contribute both light and heavy impurities. The *divertors*, for instance, need to withstand heavy power loads from energetic particles, and are therefore made of heavy metals such as Tungsten (W, nuclear charge $Z=74$). Such heavy elements will not be fully ionised, even in the extreme temperatures of a burning plasma. This leads to what is called line radiation, where remaining electrons bound to the impurity respond to a collision by jumping to a higher electron orbit. As the electrons relax, returning to the lower energy levels, they lose the energy gained in the collision, which is released in the form of radiation. For heavy elements, this process can continue indefinitely, leading to severe power losses. Because of the danger of line radiation, using an element as heavy as Tungsten is not practical for all of the reactor chamber, and hence lighter candidates with high heat-resilience are used elsewhere. For example, at JET the new "ITER-like wall project" was recently initiated, testing the feasibility of using a coating of the light metal Beryllium (Be, nuclear charge $Z=4$) on the plasma-facing first wall of the reactor chamber.

Not all impurities, however, are contaminants. The second main source of plasma impurities is the injection of particles for control



Above: Cross-section of the toroidal annulus formed by the flux-tube as it wraps around the tokamak following the magnetic field-lines, showing fluctuations in the electrostatic potential. The major and minor radii (R and r), and the poidal angle (θ), have been indicated, and a cross-section of the actual flux-tube has been high-lighted. This is based on data from non-linear gyrokinetic simulations.

Right: Time-series showing the net-flux of electrons in non-linear gyrokinetic simulations using two different magnetic geometry-models: the simplified s-α model and the magnetic equilibrium calculated for the actual experiment. Additionally, data for a simulation with several added degrees of realism (most notably collisions) is shown.



purposes. Here the cooling mechanisms are beneficial to the operation of the fusion reactor. By injecting inert gases such as Argon (Ar, *Z=18*), that radiate energy in the right locations, the heat load on components such as the divertors can be spread out, thereby protecting them from wearing out too quickly. Impurities are also injected for experimental purposes, in order to study their transport properties.

Finally, the fuel ions will, in a working power plant, be diluted by the steady production of α-particles (sometimes referred to as "helium ash") through fusion reactions.

Current work

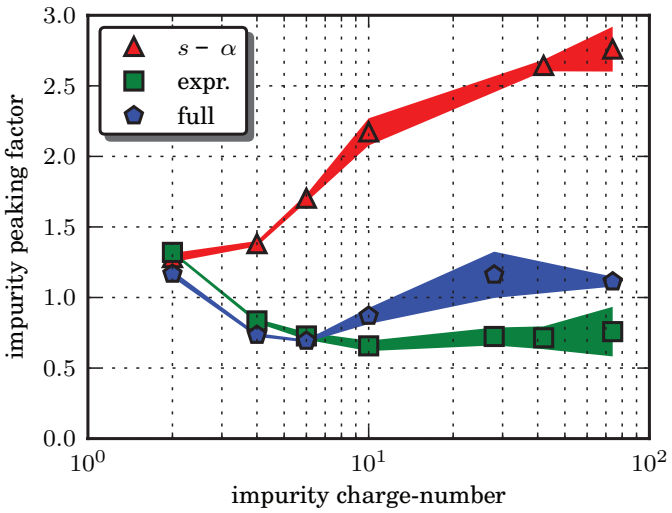
A major concern for fusion devices such as ITER, is whether the different species of ions will exhibit a net inward or outward particle transport. Determining this requires theoretical, experimental and numerical efforts. This is a major topic for the Plasma Physics and Fusion Energy group at Chalmers.

Working in collaboration with JET, the first principles code GENE (http://gene.rzg.mpg.de) has been used to interpret results of impurity injection experiments. The main focus has been on finding scaling properties of steady-state impurity transport with various experimental parameters. The bulk of the effort has concerned fully non-linear simulations of the turbulence in a limited sub-domain of the plasma called a *flux-tube*. These simulations have been performed on the PDC Lindgren supercomputer, which is currently the only Swedish HPC resource capable of meeting the computational demands of such simulations. A cutaway-view of the electrostatic potential from one such simulation can be seen in the diagram on the previous page. The time evolution of the potential is illustrated in image on the front page.

Time-series of physical quantities such as average particle- and heat-fluxes can be obtained by taking averages over the five-dimensional simulation volume. One example is shown above. By performing many such simulations, scalings such as those shown on the following page, can be obtained. These scalings can then be compared with scalings from experiments, less computationally intensive models, and theoretical results. The non-linear results highlight some of the key differences between the different models that are available.

During the last simulation-campaign, we investigated the importance of the shape of the plasma on impurity transport. In the simplest geometry-models, it is assumed that the cross-section of the plasma is circular, but in experiments it is elongated and slightly D-shaped. The real magnetic geometry was calculated using data obtained from the JET-experiments under study, and this was used in the simulations. The results show that the choice of geometry-model has important effects on the impurity transport.

When using a flux-tube domain, we implicitly assume that the interactions between large scale modes (MHD stability) and small scale modes (transport) are limited. Fortunately, the flux-tube is often a good approximation, especially where transport in the core of the plasma is concerned. Since global simulation-domains are still on the brink of being feasible, if they are to resolve both the large- (or intermediate-) and small-scale dynamics, this is a necessary restriction, but a



Left: Scaling of impurity "peaking factor" (a measure of the balance between diffusion and advection at steady state) with impurity charge-numbers for the cases in the previous figure (of time series showing the net-flux of electrons in non-linear gyrokinetic simulations). Estimated standard-errors of $1 - \sigma$ *have been indicated. Each data-point is calculated from a multiple non-linear time-series of impurity fluxes.*

restriction nonetheless. In future work, we aim to investigate the connection between local and global dynamics. However, if this path is to be scientifically fruitful and internationally competitive, it will require access to computational resources with a larger shared memory and more cores than are currently available.

In order to pave the way for the next generation of large scale gyrokinetic simulations, our group is also involved in code development. Current efforts in this area mainly concern optimising the hybrid use of MPI and OpenMP – two key parallelisation schemes – in the code GENE.

Toward a virtual tokamak

The overarching goal of the European and international modelling communities is the vision of a *virtual tokamak*, with enough predictive capability to accurately model a future fusion power plant. For future fusion experiments, most notably for ITER, the economy and time constraints will dictate what experimental setups are considered. This will most likely mean that all experiments will be preceded by extensive predictive modelling to support the experimental undertaking.

Currently work is being done to implement "large eddy" gyrokinetic models, where dissipation at smaller scales is modelled, rather than simulated. This may, in the near future, pave the wave for feasible first-principle simulations of global plasma dynamics. Another promising avenue is the multi-scale approach, where first principle and reduced transport models are combined, in order to access longer time-scales with enhanced accuracy.

For *modellers*, the virtual tokamak is considered a grand challenge. Progress in high-performance computing is making ever-wider areas of phenomena accessible for study. However, even with the technical and theoretical progress, combining the many coexistent scales important to fusion plasma dynamics is a humbling prospect. A recent estimate is that a well-resolved global gyrokinetic simulation, resolving the ion-scale turbulence sufficiently beyond the turbulence saturation time, will require in excess of 100 petaflops, with a total memory requirement in the range of several TB (www.prace-ri.eu). Including electron-scale turbulence will increase both these estimates by an order of magnitude or more.

Such HPC infrastructure is on the horizon, but what can be modelled is not merely a matter of hardware. The codes employed also have to be first-class to perform well on super-computers that are orders of magnitude larger than today's state-of-the-art machines. GENE and other gyrokinetic codes have been proven to scale well using up to 100,000 cores, but peta- and exa-scale computing will increase the number of cores to 10,000,000 and beyond. Overcoming the bottlenecks associated with communication between nodes and cores is therefore one of the major issues for the fusion modelling community. However, that this challenge is even being contemplated seriously is a testament to the fact that the complex interactions between different time and length scales are no longer an obstacle, but are instead rapidly turning into a field of scientific opportunity.

Alya System – Large-Scale Computational Mechanics

by Guillaume Houzeaux, Dept. of Computer Applications in Science and Engineering, Barcelona Supercomputing Center (BSC-CNS), Barcelona, Spain, and Jing Gong, PDC

Introduction

The Alya System, which was developed at the Barcelona Supercomputing Center (BSC), is a Computational Mechanics (CM) code with two main features. Firstly, it is specially designed to run with the highest efficiency standards in large-scale supercomputing facilities. Secondly, it is capable of solving different physics problems, each one with its own modelling characteristics, in a coupled way. These two main features are intimately related, which means that any complex coupled problems solved by Alya will still be solved efficiently.

Alya is organized in modules that solve different physical problems. There are modules included in the code for handling incompressible and compressible flows, non-linear solid mechanics, species transport equations, excitable media, thermal flows, n-body collisions, electro-magnetics, quantum mechanics, and Lagrangian particle transport.

The Alya code has been present in the PRACE benchmark suite since the start of the PRACE projects and is currently involved in the PRACE-2IP-WP7/8/9 and PRACE-3IP-WP7 projects. The code has been thoroughly tested and proven to run efficiently on many supercomputers such as the BlueGene (Jugene at JSC, Germany), the Bull Cluster (Curie at CEA, France), and the PowerPC cluster (MareNostrum at BSC-CNS, Spain). As a result of these efforts, last year the Alya team was allocated more than 20 million CPU hours for a project to solve a biomechanics problem on the Fermi supercomputer at CINECA (Italy) through a regular PRACE call.

The development team is currently putting a lot of effort into enhancing the performance of Alya, mainly in the context of PRACE work packages. In this process, we are mainly addressing algorithmic issues - namely iterative solvers, mesh multiplication, and parallel I/O - along with some OpenMP implementation concerns. This article discusses aspects of the implementation in Alya and presents some speedup results obtained in the context of the PRACE-2IP-WP7 project.

To give a concrete idea of the type of problems being addressed, the next figures show two examples of the kind of work that is produced with Alya. Here the turbulent and free-surface flows over an obstacle and inside a flushing toilet were simulated. The computa-

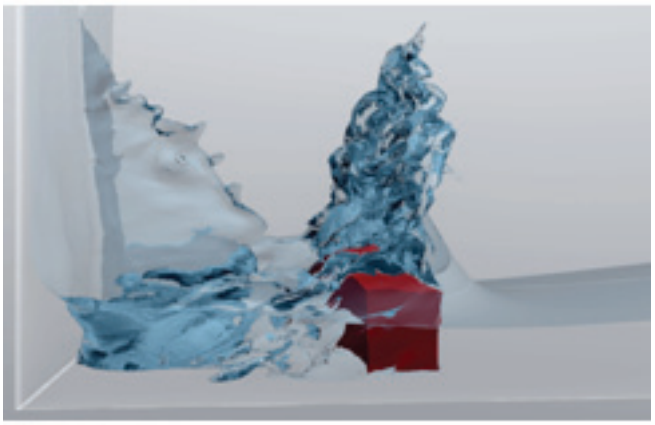
Staff Focus

Jing Gong

Jing Gong joined PDC in January 2012 as an application expert in computational fluid dynamics (CFD). He completed his Ph.D. in scientific computing on “Hybrid Methods for Unsteady Fluid Flow Problems in Complex Geometries” at Uppsala University in 2007. He also holds an M.Sc. in scientific computing from KTH, Stockholm, and an M. Eng. in Mechatronics from Beihang University, Beijing, China. At PDC, Jing also works part-time on the CRESTA and PRACE projects.

Above: Lindgren, the CRAY XE6 at PDC, was used for some of the fusion and computational mechanics simulations described in the first two articles of this newsletter.

Below: Two examples of turbulent free surface flows: dam break (top) and flushing toilet (bottom)



tions for these particular simulations were run on the MareNostrum supercomputer at BSC-CNS.

Numerical strategy

There are different physical modules within Alya that are part of the PRACE benchmark suite. These modules solve incompressible flows, solid mechanics and excitable media problems. This article focuses on a particular module (called Nastin) which is used for solving the incompressible Navier-Stokes equations. This section summarizes the numerical models and strategies employed in the Nastin module.

Discretization method

The numerical model on which Nastin is based is a stabilized finite element method. The stabilization is based on the Variational MultiScale (VMS) method. The formulation is obtained by splitting the unknowns into grid scale and subgrid scale components. This method was introduced in 1995 and established a remarkable mathematical basis for understanding and developing stabilization methods. In the present formulation of Alya, the subgrid scale is, in addition, tracked in time and in space, thereby giving more accuracy and more stability to the numerical model [5].

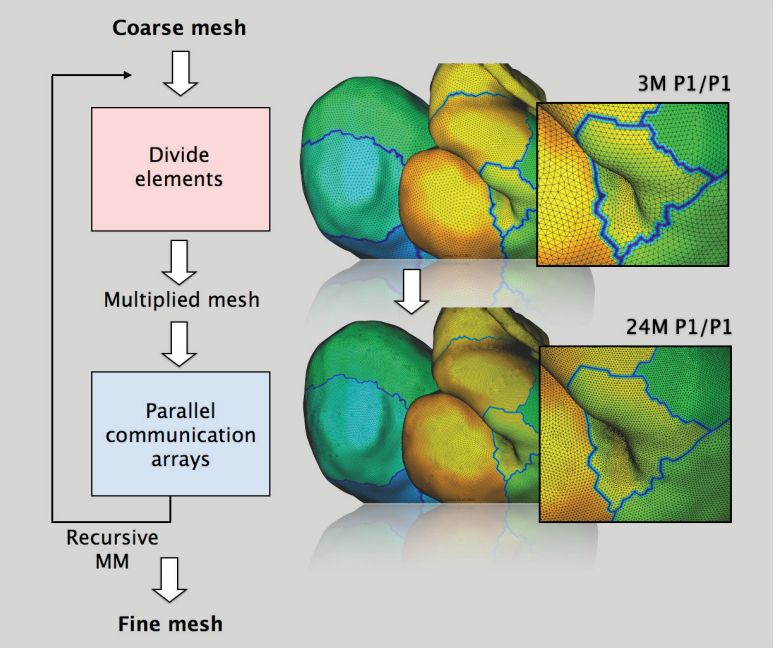
Solution strategy

The discretization of the Navier-Stokes equations yields a coupled algebraic system to be solved at each linearization step within a time loop. Algebraic solvers to solve this coupled system are not robust enough; the system is therefore split to solve the momentum and continuity equations independently. This is achieved by applying an iterative strategy, namely the Orthomin (1) method for the Schur complement of the pressure [2]. At each linearization step it is necessary to solve the momentum equation twice and the continuity equation once. The momentum equation is solved using the GMRES or BICGSTAB method (diagonal and Gauss-Seidel preconditioners are usually efficient), and the continuity equation is solved using the Deflated Conjugate Gradient method [3] together with a linelet preconditioner well-suited for boundary layers.

Mesh Multiplication

In petascale applications, the pre- and post-processing tasks are becoming a bottleneck in the complete simulation cycle. Techniques like parallel I/O have been introduced to mitigate these effects in post-processing, but these are only effective within a limited range. Mesh multiplication (MM) was introduced as an alternative. This technique consists of refining the mesh uniformly, recursively, on-the-fly and in parallel. For tetrahedra, hexahedra and prisms, each level multiplies the number of elements by eight, while a pyramid is divided into ten new elements. This technique is also very useful for studying mesh convergence as well as weak and strong scalability [1]. The figure at the top of the next page sketches the recursive MM algorithm. As an example of the efficiency of the algorithm, a mesh of 3 billion elements was obtained in 1 second on 16,384 CPUs, starting from a mesh of 3 million elements. These particular results were obtained during a PRACE-2IP type C project.

Below: Outline of the recursive mesh multiplication (or MM) algorithm



Parallelization

Full details about the code parallelization can be found in [4]. Briefly speaking, the parallelization is based on a master-slave strategy for distributed memory supercomputers, using MPI as the message-passing library. The master reads the mesh and performs the partition of the mesh into submeshes, or subdomains, using METIS (an automatic graph partitioner). Each process will then be in charge of a subdomain. These subdomains are the slaves. The slaves build the local element matrices and the local right-hand sides, and are in charge of solving the resulting system solution in parallel.

In the assembling tasks, no communication is needed between the slaves, and the scalability depends only on the load balancing. In the iterative solvers, the scalability depends on the size of the interfaces and on the communication scheduling.

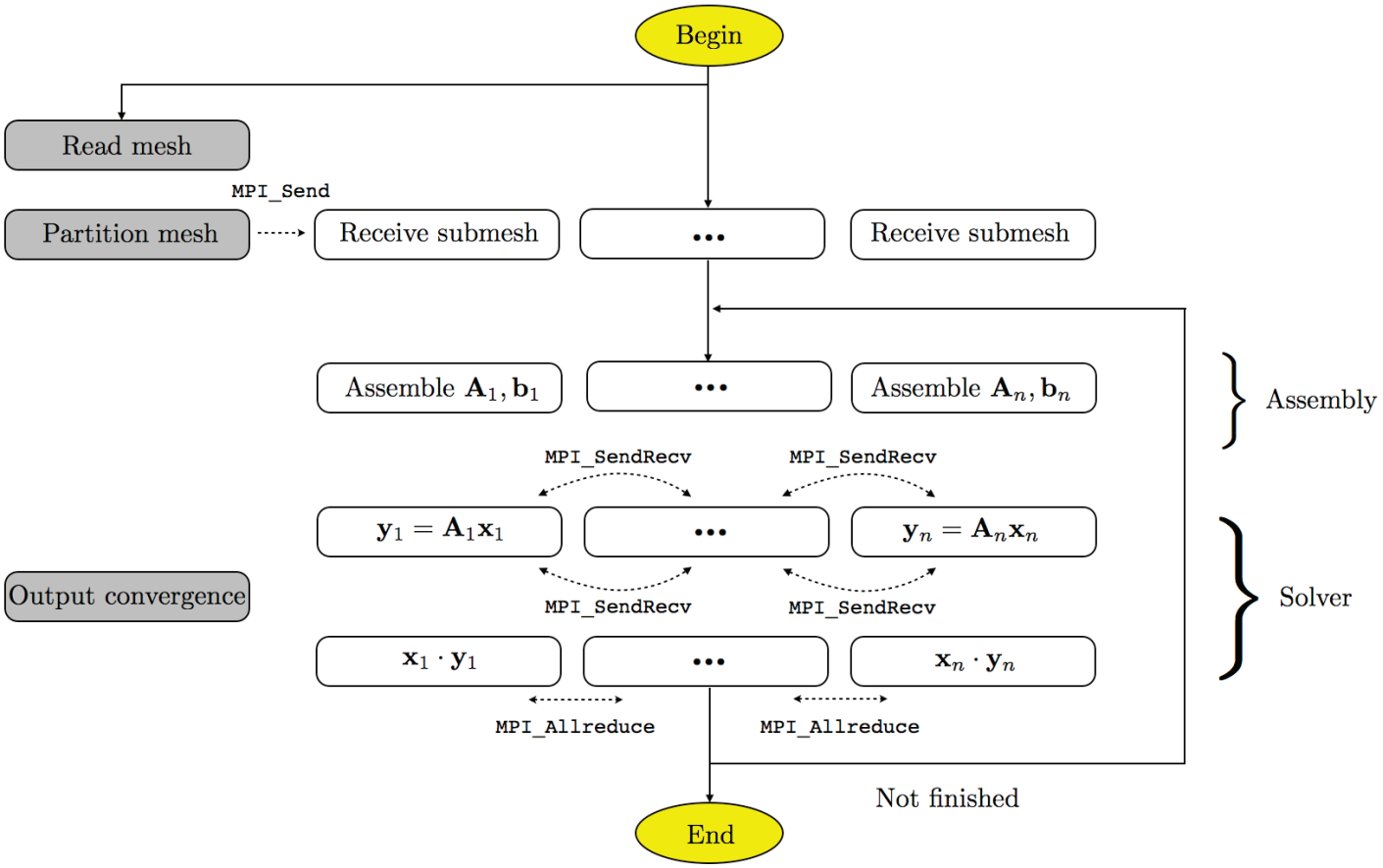
As mentioned previously, the momentum and continuity equations are solved with unsymmetric and symmetric iterative solvers respectively. During the execution of the iterative solvers, two main types of communications are required:

- global communications via MPI_AllReduce, which are used to compute residual norms and scalar products, and
- point-to-point communications via MPI_SendRecv, which are used when sparse matrix-vector products are calculated.

All solvers need both these types of communication, but, when using complex solvers like the DCG, additional operations may be required, such as the MPI_AllGatherv functions explained in [3]. In the current implementation of Alya, the solution obtained in parallel is, up to round-off errors, the same as the sequential one all the way through the computation. This is because the mesh partition is only used for distributing work without in any way altering the actual sequential algorithm. This would not be the case if one considered more complex solvers, like the primal/dual Schur complement solvers, or more complex preconditioners, like linelet or block LU.

The next figure is a schematic flowchart for the execution of a simulation using Alya. The tasks that the master process is responsible for are shown on the left side of the figure with a grey background. The master process performs the first steps of the execution, namely reading the file and partitioning the mesh. Afterwards, the master sends the corresponding subdomain information to each slave process; then the master and the slaves enter the time and linearization loops, represented as one single loop.

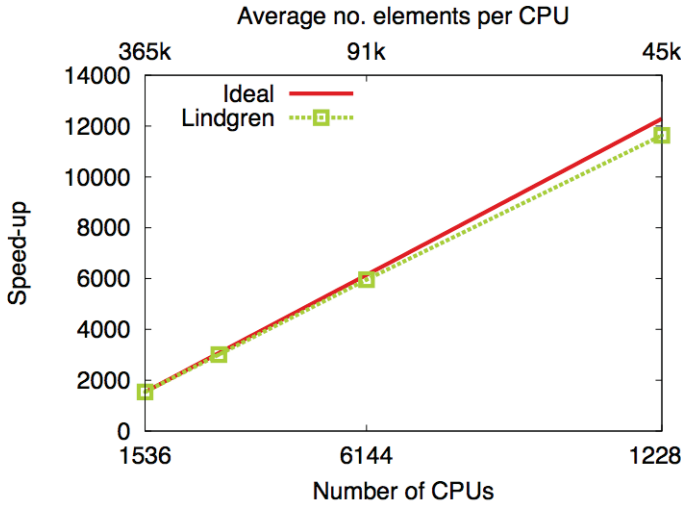
Below: Schematic flowchart for the execution of a simulation using Alya



The scalability test for the benchmark suite starts using a 1.1 million element mesh (a 0-level mesh) that is uniformly refined in parallel to a mesh of 552.9 million elements (a 3-level mesh) using the mesh multiplication algorithm presented previously. The tests for this were carried out on Lindgren at PDC - a Cray XE6 system that is based on AMD Opteron 12-core processors and Cray Gemini interconnect technology, and that has 1,516 computer nodes and a total of 36,384 cores.

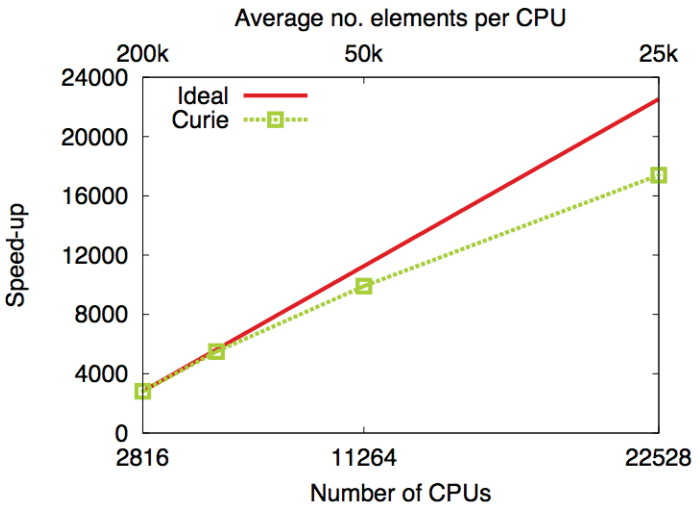
The plot in the figure to the right shows the speed-up [1] on Lindgren. The speed-up was measured by keeping the problem size constant while increasing the number of processors. The scheme's hard scalability shows an almost linear behaviour on the Cray XE6 system.

Tests with the same mesh size have also been performed on Curie at the Centre de Calcul Recherche et Technologie (CCRT). Curie is a Bull BullX system with thin nodes specification. It is composed of 10,080 eight-core processors (Intel Xeon Next Generation) which gives a total of 80,640 cores. The speed-up on the thin nodes of Curie is shown in the figure at the top of the next page.



Above: Speed-up achieved on Cray XE6 Lindgren at PDC

Below: Speed-up achieved on BullBullX Curie at CCRT



Summary

In order to fully benefit from the increasing number of CPUs available on new supercomputers, Alya developers have been constantly upgrading and improving the code, since it is part of the PRACE benchmark suite. The code is being used in production on a regular basis using up to 30,000 CPUs. The barrier of running on more than 100,000 CPUs should be broken this year, thanks to the implementation of OpenMP pragmas in the code.

Acknowledgements

The second author greatly appreciated the scholarship from COST IT-0805 STSM and SeRC that supported his visit to BSC. We also gratefully acknowledge the computer time for running these simulations that was made available by SNIC and PRACE-2IP-WP7.4, and we thank all the PRACE team for supporting the recent developments in Alya.

References

[1] G. Houzeaux, R. de la Cruz, H. Owen, M. Vázquez, "Parallel uniform mesh multiplication applied to a Navier-Stokes solver", In press, Computers & Fluids, 2013.

[2] G. Houzeaux, R. Aubry, M. Vázquez, "Extension of fractional step techniques for incompressible flows: The preconditioned Orthomin(1) for the pressure Schur complement", Computers & Fluids, 44, 297–313, 2011.

[3] R. Lohner, F. Mut, J. Cebral, R. Aubry, G. Houzeaux, "Deflated Preconditioned Conjugate Gradient Solvers for the Pressure-Poisson Equation: Extensions and Improvements", Int. J. Num. Meth. Eng., 87(1-5), 2-14, 2010.

[4] G. Houzeaux, M. Vázquez, R. Aubry, J. Cela, "A Massively Parallel Fractional Step Solver for Incompressible Flows", J. Comput. Phys., 228(17), 6316–6332, 2009.

[5] G. Houzeaux, J. Principe, "A Variational Subgrid Scale Model for Transient Incompressible Flows", Int. J. Comp. Fluid Dyn., 22(3), 135–152, 2008.

PDC Hosts Dalton Developer Meeting

by Olav Vahtras, PDC

In mid-October last year, PDC and the ScalaLife project (scalalife.eu) hosted a meeting for Dalton developers in Stockholm.

Dalton (daltonprogram.org) is a quantum chemistry program geared towards calculations of molecular properties for most common approximate wave-function models: Hartree-Fock, density-functional theory, multi-configuration self-consistent field theory and coupled-cluster theory. The program is available free of charge and has a world-wide user base with over 2,000 licenses issued.

The Dalton project is one of the most long-lived Scandinavian academic collaborations with roots that go back to the 1980s. Among other things, the agenda of the meeting included a presentation of recent Dalton-related ScalaLife activities, preparations for the new release of Dalton (which is scheduled to appear in 2013), and preparations for a Dalton publication (to appear in WIREs Computational Molecular Science).

Staff Focus

Gert Svensson

After finishing an M.Sc. in Engineering Physics at KTH in 1981, Gert Svensson worked for a period at the Radio Lab of the Swedish public telephone company Televerket. Gert soon returned to the Telecommunication and Computer Systems department at KTH where he became an early and enthusiastic user of the newly developed operating system UNIX. From the beginning, Gert understood the importance of open software and was a board member of the Swedish branch of the EurOpen organisation for many years. In the mid-eighties, he installed the first internet connection at KTH.

Gert's research interests developed to parallel computing and the department soon acquired a Sequent computer with 10 CPUs. In fact, Gert was one of the founders of PDC when a group of scientists were awarded a grant for a massively parallel Connection Machine in 1989. Gert was the first employee of PDC and has since then worked in different positions at PDC and is now the deputy director. Over the years Gert has been involved in, initiated and coordinated many research projects, mostly on the European scale, in areas like high performance computing, networking, grid technology and virtual

The Niceties of NeIC

by Michaela Barth, PDC

The Nordic e-Infrastructure Collaboration (NeIC) is an organization consisting of experts from academic IT-centres in the Nordic countries. These experts work together to facilitate the development of advanced IT tools and services in areas that are important to Nordic researchers. For example, NeIC is involved with the provision of both data storage and high-performance computing resources for scientific research. NeIC promotes collaboration between Nordic IT experts and centres, which helps the Nordic countries to avoid costly duplication of work in each individual country, and also enables these countries to jointly produce a more efficient and responsive e-Infrastructure for research than could be achieved by any of the countries individually. One practical example of this type of Nordic collaboration is the Nordic Data Grid Facility (or NDGF). The NGDF is a network between Denmark, Finland, Norway, and Sweden serving scientific communities such as the Nordic High Energy Physics research community through the operation of the Nordic Tier-1 – which is now maintained by NeIC.

In actual fact, NeIC grew out of the NDGF pilot project. Back in 2003, the research funding agencies in the Nordic countries decided to contribute to developing a Nordic Data Grid Facility. The NDGF was designed to be a distributed computing infrastructure that could primarily, but not only, be integrated into the Worldwide LHC Grid Collaboration (WLCG). This Nordic pilot project was highly successful and lead to the deployment of the first (and only) distributed Tier-1 service within the WLCG collaboration (a great example of Nordic innovation and collaboration). The Nordic Tier-1 service provides computing and data storage services for the Large Hadron Collider experiments ATLAS, ALICE and CMS at CERN, which means that it stores and processes the data produced by those experiments.

In 2009, the Nordic e-Science Initiative (eNORIA) group received a proposal for a new Nordic collaboration aiming to provide services and support for scientific communities in the Nordic countries (for which access to a distributed computing infrastructure and research infrastructures with a strong international dimension is essential). One of the major objectives of the proposal was to provide a shared pool of competence for developing and deploying pan-Nordic and pan-European services, in particular to support Nordic participation in existing, emerging and planned strategic European research infrastructures.

In response to this, eNORIA proposed the continuation of NDGF which lead to NDGF and its possible continuation being evaluated by a panel in 2010. The panel recommended that the NDGF activities be continued, but with an improved funding scheme which would ensure that the NDGF had enough independence to direct its work and also that there would be sufficient and sustained funding on the Nordic level, with most of the partner countries making approximately equal contributions.

Later, at the start of 2012, NeIC was formally established as a unit under NordForsk in Oslo. NordForsk is an organization dedicated to enabling Nordic researchers to perform research of the highest possible quality. NeIC helps to implement NordForsk's strategy by providing a common e-Infrastructure for the Nordic lands, and thus lays a foundation for increased Nordic research collaboration. However, the current aims of NeIC are wider! In the coming years, NeIC will focus on innovation in Nordic e-Infrastructure solutions, in addition to maintaining selected services (such as the important Nordic Tier-1 service for the WLCG community).

One of NeIC's tasks is to create a dialogue between the highly skilled experts at the Nordic e-Infrastructure providers and the research communities. The purpose of this is to identify areas where new e-Infrastructure services are required, and to then establish project teams to develop and provide those services. Such projects will be kept focused by dividing the work into well-defined concrete blocks (corresponding to deliverables and milestones) with a short running time, typically between 12 and 18 months. Innovation into new e-Infrastructure solutions is also likely to be a major focus of the next framework programme of the EC (Horizon 2020). Thus, NeIC's focus on innovation will support and prepare the Nordic e-Infrastructure organizations for future participation in Horizon 2020.

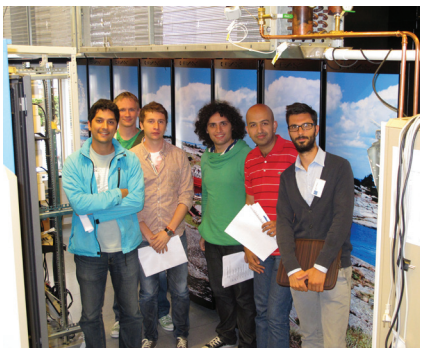
This year NeIC received an extra 5 million SEK from NordForsk to support projects that will contribute to producing innovative e-infrastructure solutions for the Nordic countries. This grant will be used to help provide base funding for such projects: 30-50% of the funding for each project will come from the NeIC, with co-funding (either in cash or in kind) from the infrastructure providers or other institutions covering the rest of the costs. The purpose of co-funding the projects is to ensure that the services that are developed will be relevant to the needs of the research communities, and to make sure that the co-funding institutions will be able to sustain the provision of the services in the longer term. NeIC is well aware that it is not enough just to develop a service, but that provision must be made for the service to continue to be available to the research communities for as long as it is needed.

Staff Focus

reality. His interests lately also include energy savings and heat re-use at computer centres.

When not working, Gert enjoys dancing swing dances, cycling and walking in the mountains of northern Sweden.

PDC Summer School 2012
Visit to the PDC computer hall



The NeIC particularly wants to encourage both innovation and collaboration. For NeIC, collaboration means not just bringing together IT-experts from the Nordic countries, but also involving both technology experts and the researchers who will use the e-Infrastructure services in the development process. This kind of cross-fertilization is vital if the resulting services are to be successful.

To promote collaboration between the Nordic countries, NeIC has chosen to use itself as a model and have its coordinators spread out over the Nordic countries. Gudmund Høst, the NeIC director, and Erlendur Helgason, the administrative area coordinator, are both from NordForsk in Oslo. Mattias Wadenstein, the Tier-1 area coordinator is from HPC2N, Umeå University, and Joel Hedlund of the National Supercomputer Centre (NSC), Linköping, is the Bio- and Medical Sciences area coordinator. And since March this year, Michaela Barth, from PDC-KTH in Stockholm has been the Generic Area coordinator.

Michaela is one of the people helping NeIC to foster a spirit of Nordic collaboration and innovation. She will take care of things that are of joint Nordic interest such as facilitating the development of tools and services applicable to a broad range of users. In her work Michaela will be advised by a small technical provider forum consisting of members representing the Nordic infrastructure providers.

Michaela emphasises that e-infrastructures are not just about hardware and high-performance computing resources, but also about people. The competence of the people we already have within the Nordic countries is definitely an asset, as long as we are able to utilize their abilities effectively. We aim for a smoothly functioning ecosystem of e-infrastructures based on high-quality solutions. Then we will not have to shy away from comparisons on European or international levels.

As a start, Michaela will work on some concrete tasks like planning and organizing an external evaluation of the Nordic HPC pilot project (www.nhpc.hi.is). In this project, Denmark, Norway and Sweden have decided to place a joint supercomputer in Iceland to study the organizational and technical challenges of joint procurement, administration and operation of a computational infrastructure for science. Another thing very high on the agenda of the Generic Area coordinator is to identify and diminish the administrative and political hurdles to cross-border resource-sharing, which is urgently needed by the Nordic e-Science Globalisation Initiative, NEGI.

As far as plans for NeIC generally, work is underway establishing various projects. The NeIC Board decided that new strategic areas for NeIC would be selected based on an open process, whereby potential research communities would submit letters of interest. A call for submissions was launched last year and 12 letters of interest were received in response. Of these, the BMS area was given priority and we are in the process of defining projects for this area.

Currently we are considering making a contribution to the coupling between dCache and iRODs which would make it easier to access the dCache data. From the Swedish point of view, this would be a natural step to take and would provide cheaper data storage – which is exactly what the BMS communities really need! Michaela hopes that we will soon be seeing results from this project that are meaningful and valuable on European, and even international, levels.

CECAM Workshop

"High Performance Computing in Computational Chemistry and Molecular Biology: Challenges and Solutions Provided by the ScalaLife Project"

by Rossen Apostolov, PDC

PDC is the coordinating partner in the EU-funded project ScalaLife (scalalife.eu). This project works on improving the performance and scalability of several widely-used packages for bio-molecular simulations, such as Gromacs and Dalton.

In October 2012, PDC members helped the ScalaLife project to organize the international workshop "High Performance Computing in Computational Chemistry and Molecular Biology: Challenges and Solutions provided by the ScalaLife Project". The workshop was held in Lausanne, Switzerland and was sponsored and hosted by CECAM (cecam.org) – an organization which funds very high-profile training events for researchers in mathematics and the natural sciences. More than 30 people attended the three day event.

The workshop started with a presentation outlining the current status quo of computing infrastructures along with the performance and scalability problems that life science applications are facing nowadays. In subsequent sessions, leading scientists discussed the latest advances and crucial pending issues in the development of fast algorithms for molecular simulations, as well as techniques for the efficient utilization of HPC resources through ensemble and hybrid computing. Other topics were also taken up including large-scale modelling and long time-scale simulations, and the prediction of molecular structures and their interactions.

The workshop included presentations of key software packages (Gromacs, Dalton and Discrete) for the quantum and classical mechanics modelling of molecular systems, as well as a novel framework for ensemble computing, Copernicus. In order to improve the adoption of the latest techniques and best practices for application usage, a major part of the workshop included hands-on tutorial sessions with users of the three packages. The pilot Competence Centre for Life Science software (which is being developed by the ScalaLife project) was presented as a long-term infrastructure for the provision of user support.

If you need help developing code for bio-molecular simulations, or if you would like help with installing or using Dalton (or other simulation packages supported by ScalaLife), please visit us at www.scalalife.eu/support.

Introducing EUDAT

by Genet Edmondson, PDC

The data tsunami – background to EUDAT

Once upon a time researchers collected information and produced data from their experiments and projects relatively slowly and in relatively small quantities. The resulting material could be stored fairly simply on paper, or, later, using the fairly basic computer storage facilities at the researchers' universities. But nowadays, calculations that used to take days (or even years) by hand can be performed in seconds or fractions of a second by supercomputers. As a result of this and the increasingly complexity of scientific measuring equipment, more and more digital data is being collected and produced by researchers. This explosion of data is sometimes referred to as the data tsunami.

Some of the waves in this tsunami of information come from medically-oriented researchers who, for example, generate vast quantities of data about genes, and from climate researchers and astrophysicists collecting enormous quantities of measurements relating to the Earth and other astronomical bodies. Linguists even generate huge volumes of information from common types of streaming data (for example, audio and video data), along with other types of time-series data (such as eye- or gesture-tracking and brain-imaging data) to study how human language works. Mathematical modellers use supercomputers to make more and more complex models of real systems - for example, modelling fluid flows (to help produce more fuel-efficient aeroplanes), or modelling how neurones in the human body work (to help find better medical treatments or discover new drugs), or modelling the climate and the weather (in order to predict weather and to help investigate climate change).

Staff Focus

Laeeq Ahmed

Laeeq started his computer science studies in 1999 as an undergraduate at Edwardes College, Peshawar, Pakistan. After finishing his B.Sc., Laeeq studied at the University of Liverpool, England, where he was awarded an M.Sc. in Distributed Systems in 2005. During his studies, Laeeq worked as a database application developer at different organizations. In 2006, Laeeq joined the University of Engineering and Technology, Peshawar, as a lecturer in 2006 and later, in 2008, he was appointed as an assistant professor.

Laeeq started his PhD in cloud computing at PDC in August 2012. His thesis topic is "Programming models for large datasets". In this work, Laeeq is using Spark for the virtual screening of chemical libraries. Spark is a tool that is used for "Big Data" Analytics with in-memory processing and fault tolerance capabilities. Virtual screening is a technique in which chemical libraries are searched for specific types of molecules, which are helpful for drug discovery. Currently Laeeq is looking into the behaviour of caching while these datasets are kept and processed from memory.

Other than sitting in front of the computer for hours, Laeeq likes to play and watch cricket.

These vast quantities of data pose problems! Researchers want to store the data that has been collected or produced so that it can be used later – by themselves or by others. Often this data is valuable not only to other researchers, but also to decision-making bodies, for example, to governmental departments that need to make decisions about environmental issues in relation to climate change. But all these enormous quantities of data raise hard-to-answer questions such as:

- **Where can all of this data be stored so as to be available in the future?**
  Many researchers or institutions may not have space to store the large files resulting from complex simulations or collections of large quantities of measurements in the longer term. Researchers whose projects are granted time to run simulations on supercomputers will not necessarily have space to store the complex results on the smaller computer systems at their local institutions.
- **How can researchers find data collected or produced by others that would be useful to their own research now?**
  This can apply to the issue of geographically-distant researchers (for example, at different universities or at institutions in different countries) who may not be aware of one another's research data. This question also needs to be addressed for the situation where data generated in one discipline may be useful to researchers in another apparently-unrelated area (even though the researchers might be based at the same location).
- **If researchers find some useful data produced by others, how can those researchers get access to and use that data?**
  Transferring huge data files between systems in different institutions or countries can be difficult.

What is EUDAT and what does it do?

The European Data Infrastructure (EUDAT) Project was started in October 2011 in order to address these types of questions and to provide a solution, particularly in relation to researchers within Europe. The aim of the project (which received initial funding for three years) is to produce a Collaborative Data Infrastructure (CDI) for European researchers that will meet current and future researchers' needs for digital data services in a sustainable way.

In practice, such a CDI physically consists primarily of high-performance computing (HPC) and data storage facilities at different locations that are linked by communication networks. However, for researchers to be able to effectively utilize such a CDI, there also need to be suitable interfaces and tools available for manag-

ing and working with the data, and hence these also constitute an important part of the CDI. Not surprisingly, one of EUDAT's initial tasks was to interview research communities to ascertain the main data management tools and services that were needed for research purposes!

In fact, EUDAT is actually a consortium consisting of both research communities and centres that provide data-processing or storage facilities. There were five main research communities initially involved with EUDAT:

- CLARIN (Linguistics),
- EPOS (Earth sciences),
- ENES (Climate sciences),
- LIFEWATCH (Environmental sciences), and
- VPH (Biological and medical sciences).

These communities consist of a number of research institutions and universities in different countries. Since the inception of the project, further institutions and organizations, such as the International Neuroinformatics Coordination Facility (INCF), have become involved at different levels (for example, with "Observer" or "Associate Partner" status).

The EUDAT project receives funding from the EU's Seventh Framework Programme (FP7/2007-2013) and is being co-ordinated by the CSC – IT Center for Science, Finland. The people involved in the project come from various institutions within Europe and are organised into seven work packages, each of which tackles a different aspect of developing the CDI. Broadly-speaking, the work is divided up as follows.

- WP1 is responsible for the overall coordination of the work within the project, including budget matters.
- WP2 is tasked with ensuring that the EUDAT CDI services will continue to be available in the long-term, in particular by finding means to fund the services in the future.
- WP3 is mainly concerned with communication between the project members, potential users and industry.
- WP4 ensures that the CDI services created by EUDAT actually meet the needs of the research communities.
- WP5 is largely responsible for designing the CDI services and investigating the suitability of the technologies and tools that are available to do so.
- WP6 is responsible for operating the infrastructure and deploying the services on this infrastructure.
- WP7 performs research to support the longer term growth and development of the EUDAT infrastructure.

EUDAT's first year

As mentioned earlier, one of EUDAT's first steps was to work with the research communities that joined EUDAT initially to analyse the types of data services that those communities needed. From this work, the following five core services were identified as being in demand.

- **Safe data replication** makes it possible to replicate small- and medium-sized data repositories from one storage site to another for preservation purposes, for optimizing access to users, or for bringing data closer to relevant instruments such as computers.
- **Data staging** lets researchers move large quantities of data from storage sites to HPC facilities (so that computations can be performed) and then moves the results back to storage.
- **Simple store** provides a facility to researchers for storing relatively small quantities of data.
- **Metadata** allows researchers to catalogue stored data in a consistent way so that researchers from all over Europe can search in the resulting catalogue for relevant data.
- **AAI (Authentication and Authorization Infrastructure)** essentially makes it possible for data to be handled securely across the network of computational and storage facilities in the CDI.

Prototypes for the first two services (Safe replication and Data staging) were launched at the end of 2012 and are currently being tested and refined by user communities. Prototypes for the Metadata and Simple store services are expected to be available by July 2013.

PDC and SNIC involvement in EUDAT

PDC represents SNIC in the EUDAT infrastructure. As such, PDC contributes to the federated EUDAT storage infrastructure and works closely with the neuroinformatics community, via the INCF, in order to integrate them into the EUDAT infrastructure. The work on EUDAT's storage infrastructure is also beneficial for the national Swedish storage infrastructure (SweStore) as EUDAT technologies are being deployed on SweStore too.

PDC is also leading the Simple Store Task Force that is developing a simple storage service, based on CERN's Invenio technology, for the "long tail of researchers with small to medium storage needs". PDC staff also provide editorial assistance to WP3 and the Publications Editorial Committee (a subgroup of WP3 that is responsible for overseeing all publications produced by EUDAT).

Where is EUDAT going now?

During this second year of the project, EUDAT is continuing to focus on completing the establishment of the initial services, (and offering training for people who use the services), along with starting work on a second-round of data services.

As mentioned previously, the EUDAT project received funding for three years from the EC FP7 programme. However one of the central issues for EUDAT has been setting up a means to store and access digital data that will be sustainable. Consequently, part of the work for the remaining years of the project is finding ways to ensure that the established services will continue to be available in the long-term.

For further information about EUDAT, see www.eudat.eu.

Sharing and Preserving Scientific Data with iRODS

by Carl Johan Håkansson, PDC

What is iRODS?

Today's stampeding growth of scientific data is causing an ever-increasing demand for the safe and efficient storage of large data sets, along with everything that entails – such as back-ups, transfers, data management, and the long-term archival of both scientific data and the metadata related to it. At the same time, we are experiencing an increasing demand for data accessibility, both geographically and over time. The publication of scientific reports generally comes with a requirement for accessible research data sets. Global collaboration and data sharing within the science community also require accessible data. Increasing data set sizes, increasing total amounts of data, and increasing needs for accessible data all in turn increase the demands and challenges put on data management systems.

iRODS, the integrated Rule-Oriented Data System, aspires to provide the next generation data management infrastructure for the scientific community. As such, it provides end users and data administrators with a uniform and easily accessible facility for the preservation of scientific data, which is separated from system administration and physical storage. This means that data distributed over diverse storage resources in a potentially global data grid can be seamlessly integrated and presented for users as though the users were just working with the data on their own workstations - where some of the data handled by iRODS may very well reside.

In technical terms, iRODS provides a data grid middleware with server side support for replication and federation, and with command-line, graphical or web-based interfaces for end-users and administrators. The "Rule-Oriented" data management policy implementation provides a high level of automation which facilitates admin-

istration, and further helps by separating data management from system administration.

In this article we will take a look at the ongoing development of iRODS, how iRODS is structured and can be used, and finally peek a little bit into its current and future use at PDC.

The development of iRODS

iRODS is available in the form of BSD-licensed open source code (see www.irods.org). The core development team is part of the DICE (Data Intensive Cyber Environments) research group, who are currently based at the University of North Carolina at Chapel Hill. DICE was the original developer group for iRODS. They based the development of iRODS on their previous experience developing the Storage Resource Broker (SRB), which has been in widespread use since 1997 and can be regarded as a predecessor to iRODS.

The Renaissance Computing Institute (RENCI) – a research unit at the University of North Carolina at Chapel Hill – is another major contributor to iRODS, particularly when it comes to client development. RENCI is also leading the development of e-irods, or "Enterprise iRODS". This is a hardened binary enterprise distribution of iRODS for which RENCI provides diverse service agreements and professional support.

Many universities and research institutions all over the world have been contributing to developing iRODS and the development community remains very active, even though the core iRODS has reached a stable state nowadays. The current iRODS development roadmap is inspired by the RedHat-Fedora model with DICE governing the community core source code releases and RENCI planning to release enterprise packages every 18 months.

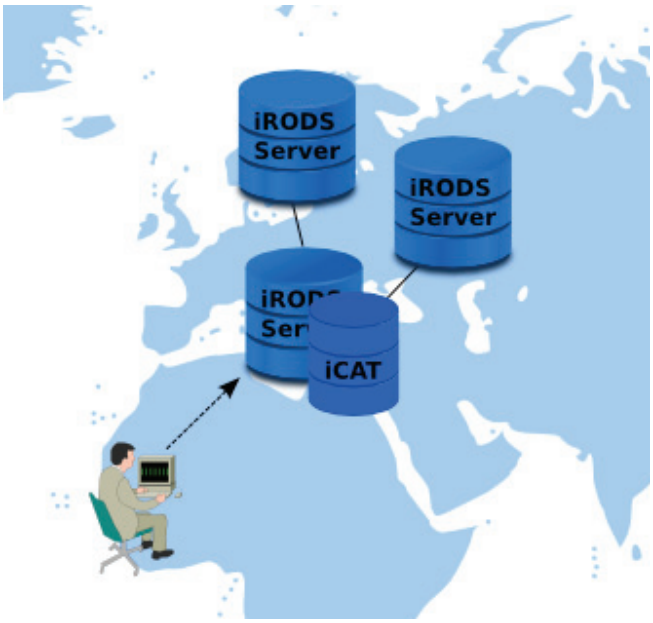
Architectural overview of iRODS

The development of iRODS has very much been driven by the desire to share scientific data over space and time. For scientific data, sharing pretty much requires data preservation as very large data sets often need to be accessible over a considerable length of time. iRODS is therefore designed with data sharing and long-term preservation in mind. It is also designed to provide service throughout the life cycle of the scientific data, from ingestion to permanent archiving or disposal. Workflow management is based on policy-driven data management implemented by procedural rules, either pre-defined or defined by system or data administrators, and triggered by events such as data upload, the download or replication of data, or time-driven events such as the archival of data after a certain time. Policy-driven data management also separates data management from storage administration (as policies are defined within the community). This type of management enables data administrators and end-users to manage data to a large extent without help from system administrators, and to apply policies on a grid level, independently of the actual physical location of the data. Thus, in architectural terms, iRODS is a data grid middleware, as well as being a rule-oriented data management infrastructure.

iRODS always presents data to the end-users as logical data collections, independently of where the actual data resides, whether it be in files on a local disk, on a remote disk array or tape archive, in a database system, in an Amazon S3 cloud or somewhere else. This makes it easy for users to manage and share data independently of the actual physical location of the data. iRODS' metadata catalogue, the iCAT, handles the mapping between the logical data collections and the physical files. The iCAT also handles users, roles, access control, and the management of other metadata. Via iCAT, iRODS manages its own password login system, although iRODS also supports several other authentication systems such as PAM, GSI and Kerberos.

Multiple iRODS servers may share the same iCAT, and storage resources can thus be shared between sites to form a single logical namespace for file storage, which is known in iRODS terms as a "zone". Parallel transfer mode ensures that iRODS transfers files with high performance, but files can also be automatically replicated between sites in order to improve accessibility and

Below: Multiple sites in a single iRODS zone
An iRODS zone is a complete data grid consisting of one or more iRODS servers connected to a single metadata catalogue (iCAT). Users can access scientific data anywhere in the zone via a nearby iRODS server. Metadata (such as user rights, file names and location) is handled within the zone via the iCAT database.



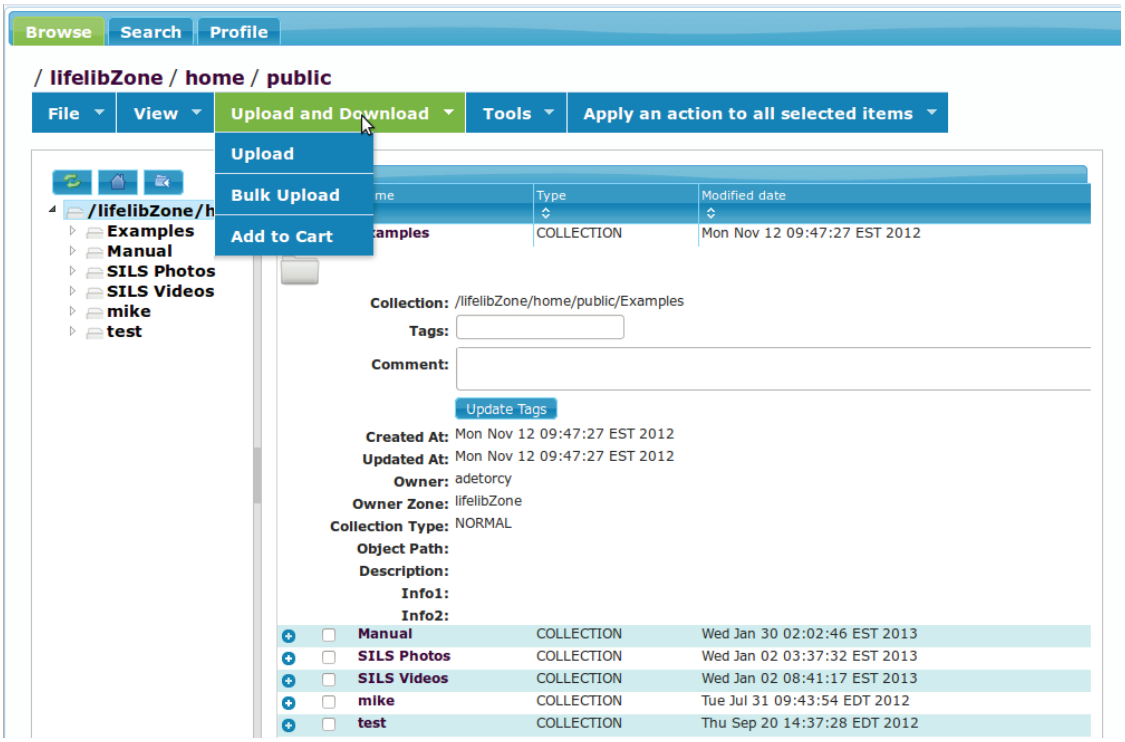
decrease transfer latency by bringing the data nearer to the computing centre where it will be used.

There are several user interfaces to choose from on the client-side: the command line based icommands, different web-based clients (see screenshot below for an example), and also some graphical Java-based interfaces. There is even some support for using iRODS via native file managers or mounting iRODS resources as a local file system. For the best performance, transfers between clients and servers using iRODS' transfer protocol may be parallelised for single files, and multiple files can be bundled for a single transfer.

Federation - sharing between zones

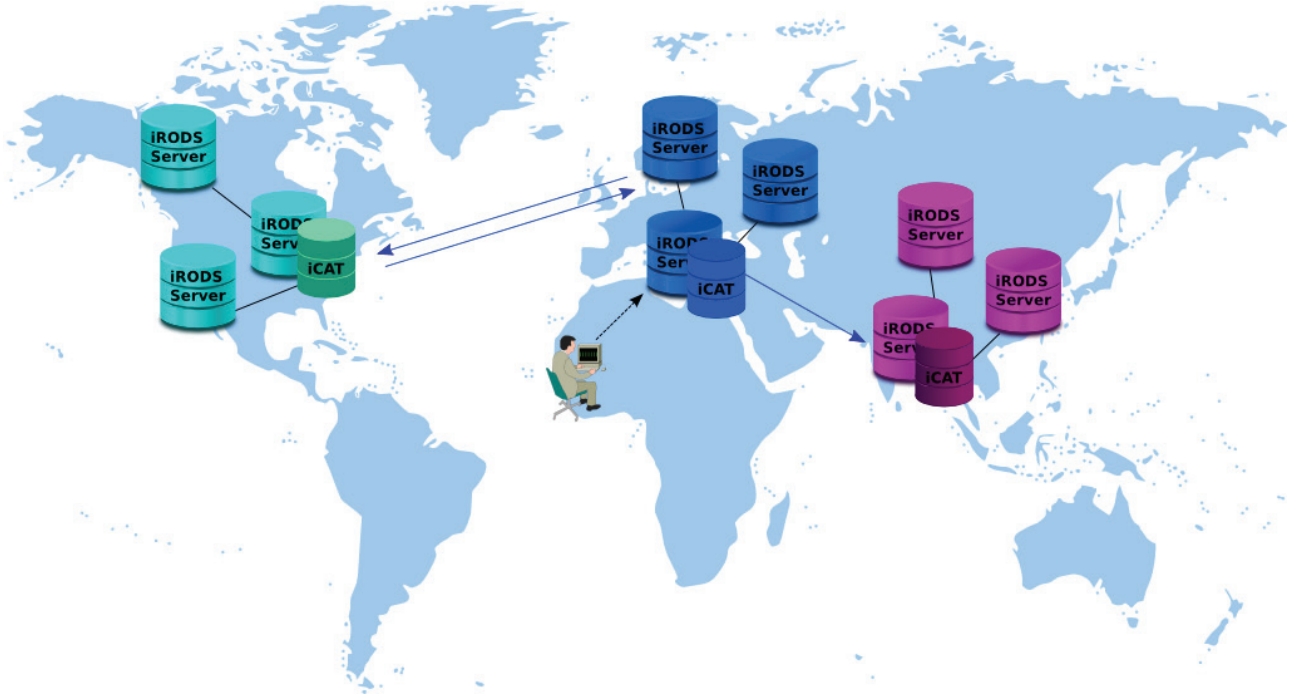
iRODS systems may be assembled - or subdivided - into zones, where each zone corresponds to a unique iCAT database, although the database may be replicated or physically distributed in some other way in order to achieve high availability, to balance loads or to improve network and I/O performance. Federation can be used between physical sites, for example, between different research centres or institutions, as well as between zones storing data in the same physical location.

With federation, it is possible to easily share data between research projects while maintaining local access control, security and user administra-



Above: Screenshot illustrating the use of the iDrop web interface to access data

Below: Example of federation between iRODS zones
Independent iRODS zones, each of which is a complete data grid with its own iCAT, can be federated.



tion. Federation also allows transparent sharing of data between sites using completely different hardware, system architecture or administration policies. User authentication can be handled by the users' home institution, while data access control is managed by the site where the data resides. Federation thus separates data management tasks from system administration and allows architectural and administrative differences between sites, or between projects with data residing at the same site.

iRODS at PDC

PDC is in the process of deploying iRODS on our systems and, although it is early days yet, several projects are already up and running at PDC.

(a) PDC iRODS

PDC supplies iRODS for local users, that is, for users working on research projects that have been granted time allocations on PDC's HPC-system. At present, the first users are testing this system.

(b) SNIC iRODS

Many research groups that are already using SNIC's SweStore for storing their data also want to be able to use iRODS. iRODS has been evaluated for provision within SNIC and currently the SNIC centres (including PDC) are collaborating on testing an iRODS system, which is intended to be in use later in 2013.

(c) EUDAT iRODS

The EUDAT project is working on developing a pan-European collaborative data infrastructure which supplies a range of services including replication of data, and data-management based on auditable policy rules and persistent identifiers (as offered by the European Persistent Identifier Consortium, EPIC). EUDAT's replication and data management is handled with iRODS via federated zones. PDC is running one of the zones within the EUDAT e-infrastructure and is currently federated with CSC in Finland and the INCF.

Thus a basis for iRODS usage on local, national and European levels is already being established, and, as larger research projects may like to run their own zones and federate on any level, there will be more coming up in the field of iRODS data management.

The New SNIC Galaxy Project

by Åke Edlund, PDC and HPCViz

In March this year, the PDC Cloud Group, together with UPPMAX and UPPNEX, started a new one-year project called "SNIC Galaxy". The goal of the project is to deliver Galaxy as a service, using the Galaxy cloud management platform, Cloudman, on local cloud installations (that is, on private clouds).

For those of you who may not have encountered it before, Galaxy (galaxyproject.org) is a platform for scientific workflow, data integration, data and analysis persistence, and publishing that aims to make computational biology accessible to research scientists who do not have computer programming experience. Although Galaxy was initially developed for genomics research, it is largely domain independent and is now used as a general bioinformatics workflow management system.

Adding Galaxy onto the SNIC Cloud Infrastructure – a private cloud environment – will benefit our life science users in a number of ways, especially with respect to storage (as it is not feasible to upload many terabytes of data to Amazon on a daily basis) and privacy (as the data will reside in Sweden).

In addition to adding Galaxy to the SNIC Cloud service, we will implement the Galaxy Cloudman service, enabling the user to elastically scale his or her applications with the necessary resources. Through the Galaxy Cloudman, it will also be easier for user groups to administer the allocated resources between the users.

In general, the groups that use this service will be from SNIC, with special interest being shown by SciLifeLab, UPPNEX, ScalaLife, and ELIXIR (Sweden and UK) as well as by Nordic collaborations (through NeIC).

For more information about the SNIC Galaxy project, you are welcome to contact Åke Edlund, the PDC contact for the project:

edlund@pdc.kth.se

PDC-Related Events

PDC Summer School 2013: Introduction to High-Performance Computing

19-30 August 2013, KTH Main Campus

www.pdc.kth.se/education/summer-school

HPC Sources

We recommend the following sources for other interesting HPC opportunities and events:

CERN
cerncourier.com/cws/events

cdsweb.cern.ch/collection/Conferences?ln=en

EGI
www.egi.eu/about/events

HPC University
www.hpcuniv.org/events/current

HPCwire
www.hpcwire.com/events

Linux Journal
www.linuxjournal.com/events

Netlib
www.netlib.org/confdb

PRACE
www.prace-project.eu/prototype-access

www.prace-project.eu/hpc-training-events

www.prace-project.eu/news

SNIC
www.snic.vr.se/news-events

US Department of Energy
hpc.science.doe.gov

XSEDE
www.xsede.org/conferences-and-events