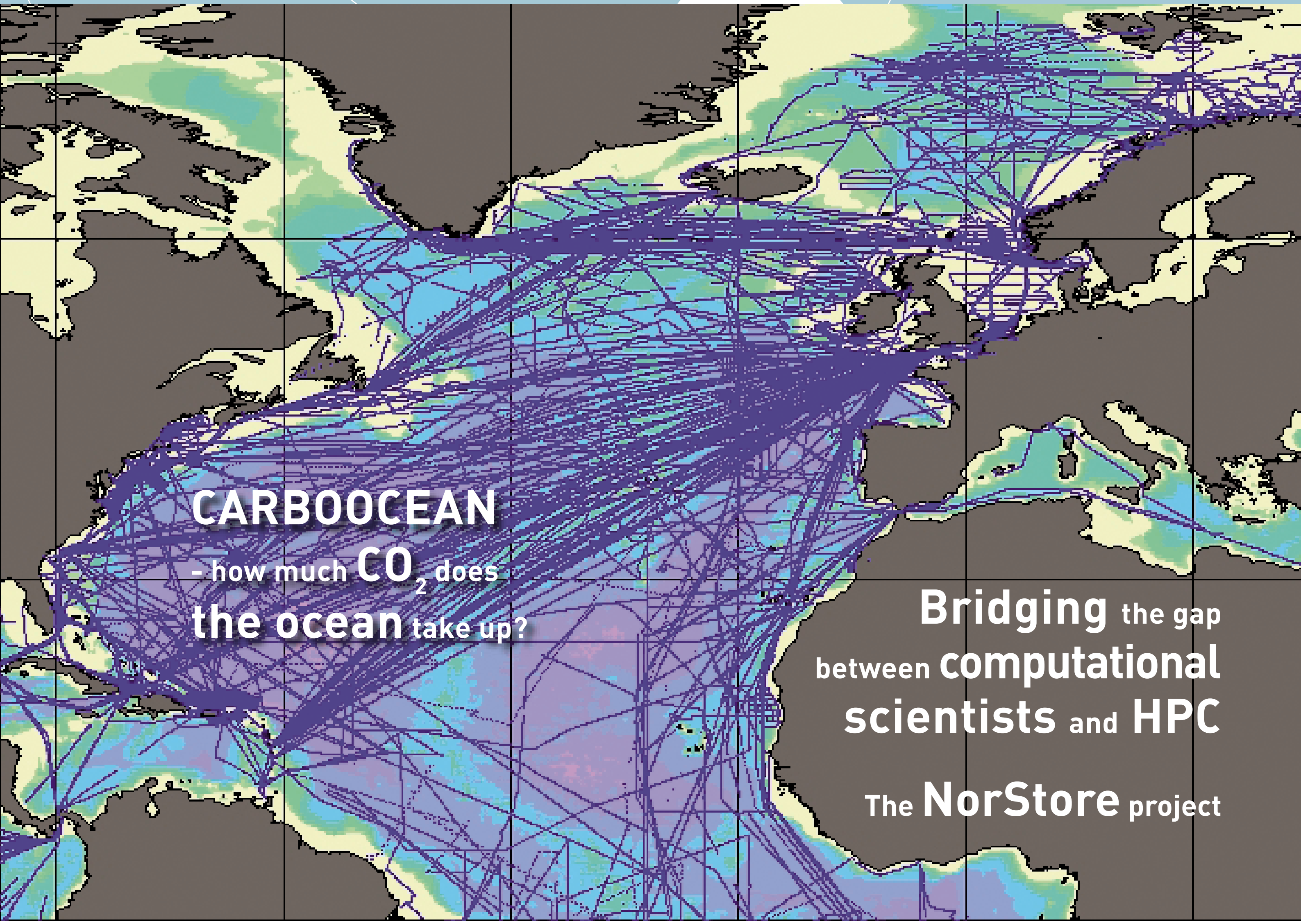
Number 3 : 2007

meta

A magazine published by the Notur II project



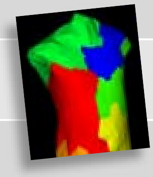
CARBOOCEAN – how much CO_2 does the ocean take up?

Bridging the gap between computational scientists and HPC

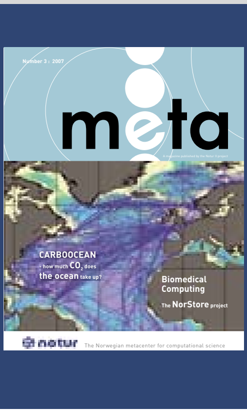
The NorStore project

notur
The Norwegian metacenter for computational science

CONTENTS



Cover picture:Cruise tracks from voluntary observing ships and research vessels which provide semi-continuous measurements of the ocean surface CO_2 partial pressure. This map includes data collected before and during the CARBOOCEAN project (Source: Benjamin Pfeil, University of Bergen).



Number 3 : 2007
A magazine published by the Notur II project – The Norwegian metacenter for computational science

Editorial Director: Jacko Koster
Contributing Editor: Eva Irene Haugen
Subscription: An electronic version is available on www.notur.no. Here you can download a PDF-file of the magazine and subscribe or unsubscribe to the magazine.
ISSN: 1890-1956

Reproduction in whole or in part without written permission is strictly prohibited.
e-mail address: sigma@uninett.no. Phone: +47 73 55 79 00. Fax: +47 73 55 79 01.
Print: Skipnes

EDITORIAL

In earlier issues of the META magazine this year, the new resources were presented that will increase the computational capacity provided by the Notur project by a factor of ten between October 2007 and March 2008. In addition, new storage resources with more than one PetaByte capacity will be taken into operation as part of the new NorStore project early 2008. These new resources, and the operations and support associated with it, will add major value to the Norwegian infrastructure for computational science.

Despite these recent developments on the national level, the quality and competitiveness of the national infrastructure must eventually be measured by where it stands internationally. A popular exercise for example is to see where the largest national compute resources rank in the list of largest compute resources world-wide, even though everybody agrees that this is a meaningless metric. A more relevant measure is how the national infrastructure engages in international efforts as infrastructure for computational science is becoming increasingly more international. European strategies and policies on infrastructure will eventually impact the strategies and plans for the national infrastructure and as such the national infrastructure must pay considerable attention and contribute at an early stage to what is happening internationally.

One of the processes on the European level that has gained considerable momentum in 2007 is the design, preparation and implementation of the different layers in the European ecosystem (or performance pyramid) to provide European research with first-class computing and data handling environments. The top layer in this European ecosystem includes a small number of compute resources in the Petaflop range, while the middle and lower layers consist of a federation of national and local resources. The implementation of these layers is being targeted by large European infrastructure projects such as PRACE, DEISA, and EGEE. Also the coupling between the layers in the performance pyramid must be given considerable attention. The planned collaboration among the National Grid Initiatives (with NorGrid in Norway) through the European Grid Initiative (EGI) to establish a sustainable grid infrastructure will also add major value to this ecosystem.

For Norwegian science to benefit from such a layered European ecosystem requires that the national infrastructure is well integrated in the European landscape. It is for example highly non-trivial to migrate a scientific application from a local resource with some hundreds of processors to a national resource (or grid of resources) with thousands of processors and eventually lift it up to a European resource (or grid of resources) that includes tens of thousands of processors. It is therefore important that the national infrastructure actively adopts the technologies and policies that will be used on a European scale and engages in (as well as contributes to) the efforts that drive the development of these technologies and policies.

Jacko Koster, Project Coordinator Notur II, Managing Director UNINETT Sigma AS

Copyright: Justyna Furmanczyk, Poland

FEATURES

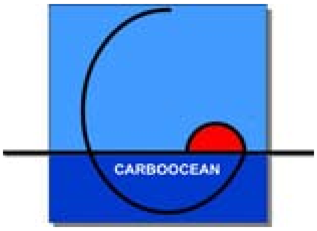
CARBOOCEAN – how much CO_2 does the ocean take up?

Since the beginning of the industrial revolution at around year 1750, mankind has increasingly released additional carbon dioxide (CO_2) to the atmosphere. This excess CO_2 is the main contributor to a human induced climate change through increasing the atmosphere's greenhouse effect. The atmospheric CO_2 level is now about 1/3 higher than the preindustrial level and in the coming decades we expect a further steep increase in human caused CO_2 emissions from fossil fuel burning, land use, and also cement production.

AUTHOR

Christoph Heinze
CARBOOCEAN coordinator,
University of Bergen,
Geophysical Institute and
Bjerknes Centre for Climate
Research. Co-authors: the
CARBOOCEAN consortium

At present, the ocean takes up about 25% of the annual CO_2 emissions. The EU funded FP6 Integrated Project CARBOOCEAN (contract no. 511176 GOCE) is dedicated to an improved quantification of the oceanic uptake of anthropogenic CO_2. CARBOOCEAN runs from 2005-2009 and includes about 200 scientists from 15 countries (from Europe, Morocco, USA, and Canada), see Figure 1. CARBOOCEAN is highly policy relevant, as the source/sink distribution for CO_2 has to be well known in order to frame policies aiming at CO_2 emission reductions and measures to enforce such policies. In the recent IPCC 4th Assessment Report of WGI, mostly physical climate models have been used for the climate change scenarios until 2100. However, the report states, that the carbon cycle feedback to climate change is quantitatively important. It has to be taken into account properly for an improved prediction of the atmospheric greenhouse gas forcing in such scenarios.



Within CARBOOCEAN, field measurements, process studies and advanced ocean modelling are carried out in order to better quantify the amount of anthropogenic CO_2 which has been taken up by the ocean since pre-industrial times, how much is taken up at present, and how much will be taken up in future. We are considering a time span of -200 to +200 years from now. We aim at narrowing down the uncertainties in the marine CO_2 uptake rates. We focus in particular on the Atlantic Ocean (including the Arctic) and the Southern Ocean (Antarctic Ocean). Backbones for our research are semi-continuous measurements of ocean surface CO_2 partial pressure from voluntary observing ships (Figure 2), deep section measurements from research vessels, case study experiments such as mesocosm studies (controlled ecosystem perturbation experiments) and high-end coupled biogeochemical-physical climate modelling using supercomputers (Figure 3).

CARBOOCEAN is now in its fully operational phase. Three major issues emerge:

1. From careful analysis of biogeochemical tracers in the ocean it can be deduced, that so far per unit area (let us say 1 m^2) the North Atlantic Ocean was a quite efficient sink for anthropogenic carbon. This is due to the fact, that deep water production at northern high latitudes pumped considerable amounts of anthropogenic CO_2 downwards to greater depth. Now, however, this formerly efficient CO_2 sink seem to

RELEVANT LINKS

- http://www.carboocean.org
- http://www.carboschools.de

Figure 1: The CARBOOCEAN consortium at their annual meeting in Amsterdam, November 2005 (Source: Robert Key, Princeton).

weaken. This is indicated by a decrease of the air-sea difference in CO_2 partial pressure in the Northern North Atlantic region. Currently, different hypotheses on the specific cause for this trend are followed and we expect to receive more concrete information about the associated cause-effect links soon. Recently, also evidence for a weakening of the anthropogenic CO_2 uptake in the Southern Ocean was found. Should the trend of a weakening oceanic sink continue, potentially greenhouse gas emission scenarios for a stabilisation of climate change may have to be corrected downwards (implying a necessity for more severe savings in human energy consumption).

2. Though the oceanic CO_2 uptake of additional CO_2 emissions may slow down, the global net flux of anthropogenic CO_2 is expected to be into the ocean during the coming decades. This has the negative side effect of a change in pH value. The pH value indicates how alkaline or acid an aqueous solution is. The CO_2 added to the ocean decreases the pH value of the ocean and hence makes it less alkaline than it was before ("ocean acidification"). Many organisms in the marine environment depend on a certain pH range for optimal survival. Ocean acidification therefore has the potential to change life in the oceans. Particularly prone to ocean acidification are organisms which build shells made of calcium carbonate. The process of ocean acidification could be documented from repeated field measurements and parts of the deep ocean previously oversaturated with respect to calcium carbonate become now undersaturated. Controlled ecosystem experiments have been carried out with changes in CO_2 partial pressure and hence pH value. The results are evaluated at present. The net effect of ocean acidification on the marine CO_2 uptake is still difficult to quantify as all effects (inorganic chemistry, change in calcium carbonate production as well as re-dissolution, change in marine particle flux) have to be combined, but estimates on the range of potential changes can already be made.

3. In future scenarios, CARBOOCEAN scientists investigate how the changes in oceanic CO_2 sources and sinks will be modified under climatic change. We try to implement the major marine processes known from the field observations and then see how the combined effect evolves during the coming decades under prescribed CO_2 emissions. So far, the combined carbon cycle climate feedback is expected to be positive, i.e. it reinforces climate change. This is important information, as projections on future climatic change so far have mostly based

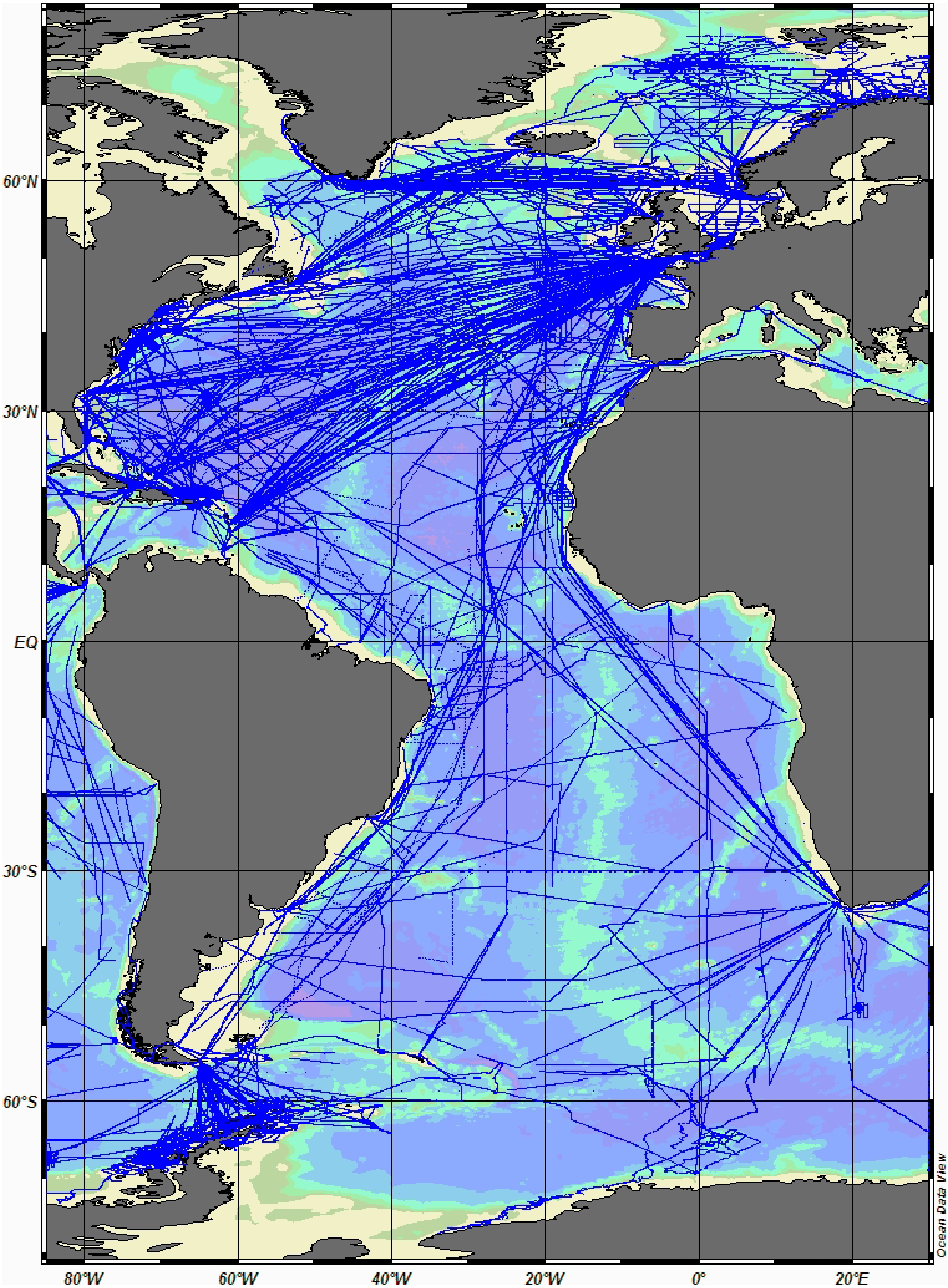


Figure 2: Cruise tracks from voluntary observing ships and research vessels which provide semi-continuous measurements of the ocean surface CO_2 partial pressure. This map includes data collected before and during the CARBOOCEAN project (Source: Benjamin Pfeil, University of Bergen).

on purely physical models. We employ different Earth system models and combine them with the marine field observations which have been collected. Thus we have a clearer view on how reliable the models are for the present time and how well they may project future changes.

These Earth system model runs are among the most complicated undertakings in computational science. Different model components for each Earth system component (such as atmospheric physics, ocean physics, ocean biogeochemistry, land carbon cycle, but optionally also atmospheric chemistry, ice sheets, or even the antroposphere) are coupled together. Each model component per se already is a highly complex computer programme. The models have to be spun up before they can be used in predictive scenarios or any other experiments, so that experiments start from controlled model situations which are in reasonable equilibrium. These spin-ups can take several months of computing time and cannot be repeated very often. Climate change scenario runs usually start at the preindustrial and then continue over several hundreds of years. In order to validate the models, results have to be written in form of time series (daily, monthly, annually) and the model results have to be compared with observations. Also for the predicted time span (often until year 2100, in CARBOOCEAN until 2200), results have to be written in regular intervals in order to allow a detailed analysis after the model run has been completed. Next to extremely fast high-end central processing units, fast disks and huge archiving systems are needed to store the results properly. This is an essential part of the entire scenario computations. The climate change simulations with Earth system models are expensive research experiments and their results need to be conserved, also for sharing them with other research groups nationally and internationally.

The observing and prediction systems on the marine carbon cycle which have been developed under CARBOOCEAN will be essential tools to quantify the marine uptake of CO_2 also in future. A dedicated educational programme CarboSchools - developed jointly by the terrestrial EU FP6 IP CarboEurope and the marine EU FP6 IP CARBOOCEAN - will enable teachers and students of secondary schools to understand current research in carbon cycling and motivate the next generation of young researchers for this scientific field. ●

CONTACT INFORMATION:

Christoph Heinze
CARBOOCEAN coordinator
email: christoph.heinze@gfi.uib.no

Andrea Volbers
CARBOOCEAN scientific project manager
email: andrea.volbers@bjerknes.uib.no

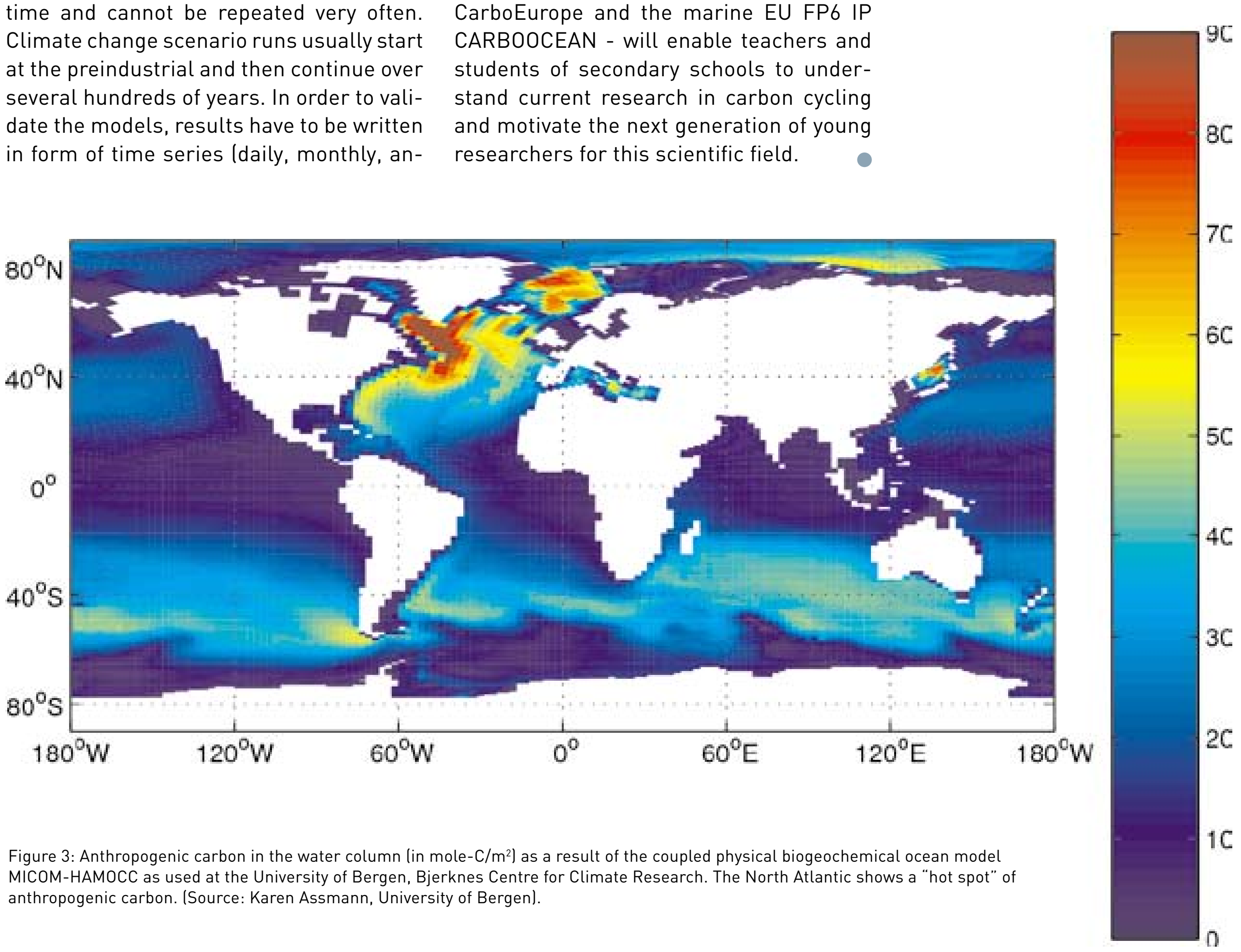


Figure 3: Anthropogenic carbon in the water column (in mole-C/m²) as a result of the coupled physical biogeochemical ocean model MICOM-HAMOCC as used at the University of Bergen, Bjerknes Centre for Climate Research. The North Atlantic shows a "hot spot" of anthropogenic carbon. (Source: Karen Assmann, University of Bergen).

A steady growth in the need for compute resources

Since a number of years, the demand for computing resources by the research community in Norway has been growing more rapidly than the total available resources. This article summarizes some of the usage statistics for the compute facilities operated by the Notur project in the period October 2006 - September 2007.

AUTHOR

Eva I. Haugen
Information Advisor
UNINETT Sigma

Available resources

The Notur project provides high-performance computing resources and services for research and education at the Norwegian universities, colleges and research institutes. The four university partners in the project (NTNU, UiB, UiO, UiT) host and operate the high-performance computing facilities. In the period October 2006 – September 2007, there were essentially four facilities in operation in the Notur project (magnum – 64 processors, njord – 896 processors, snowstorm – 400 processors, and tre – 96 processors). In November 2006, njord replaced gridur as the system that is used a number of times per day by the Meteorological Institute for operational weather forecasting.

All research projects that are financed through the Research Council of Norway or the Ministry of Education and Research can apply for access to one or more of these facilities. Access is typically granted for a 6- or 12-month period (allocation period) that can be renewed when necessary. Applications for access are received by UNINETT Sigma and are evaluated by the Resource Allocation Committee that is appointed by the Research Council. Once an application for access has been approved, the applicant gets assigned a certain number of processor hours (quota) on one or more of the compute facilities. The quota is shared by all users that are connected to this project.

As the facilities have processors with different speed (i.e, different peak performance) the total available processor hours are counted in allocation units. These are processors hours multiplied by a factor that reflects relative processor speed. In 2007, one allocation unit was defined as one processor hour on njord (with power5+ 1.9 Ghz processor, 7.6 Gflops/s).

Due to the cost-sharing of the facility between the hosting university and the Research Council of Norway (through UNINETT Sigma), the total available allocation units on the facility are divided into a 'national' part and a 'local' (or university) part. The university has the local part to its disposal (primarily for research activity strategic to that university), whereas researchers at all Norwegian research organizations can apply for access to the national part of the facility. The statistics given in this article correspond to the national part of the facilities.

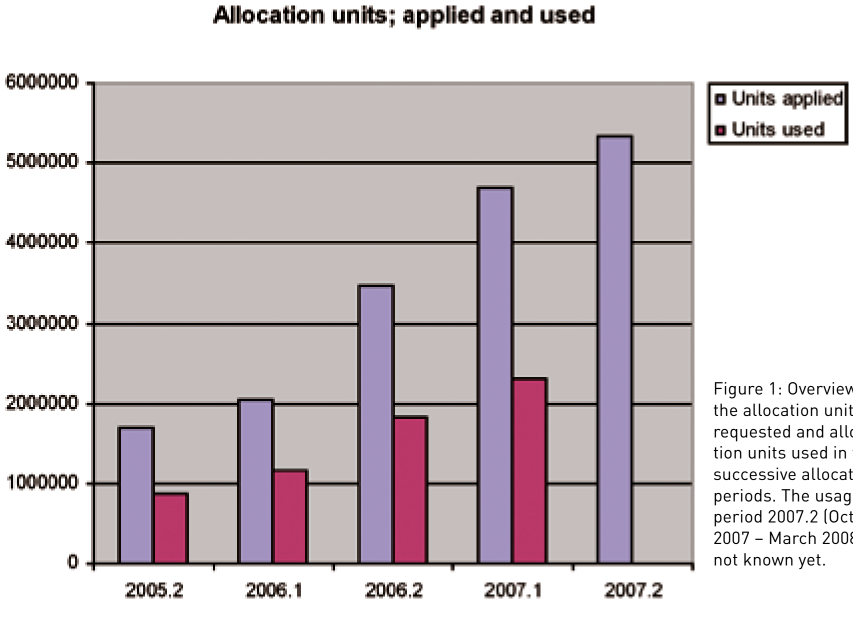


Figure 1: Overview of the allocation units requested and allocation units used in five successive allocation periods. The usage for period 2007.2 (October 2007 – March 2008) is not known yet.

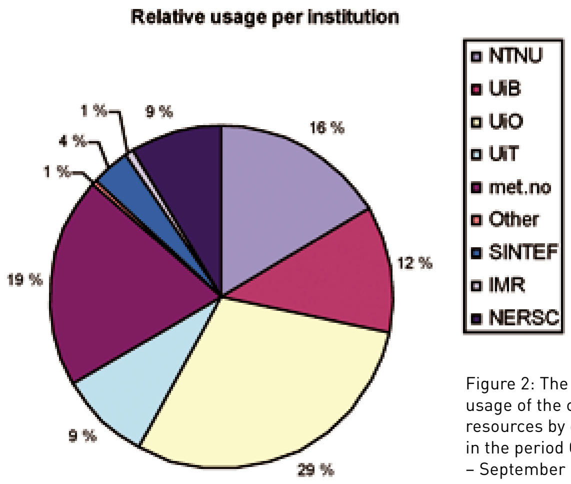


Figure 2: The relative usage of the computing resources by organization in the period October 2006 – September 2007.

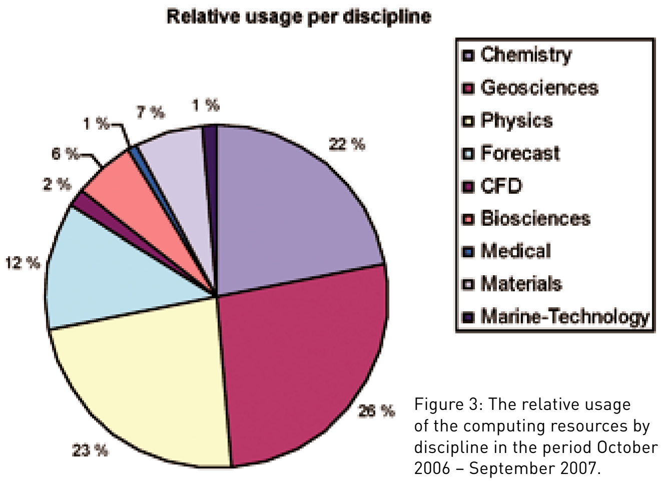


Figure 3: The relative usage of the computing resources by discipline in the period October 2006 – September 2007.

To obtain high utilization of the resources, the Resource Allocation Committee uses a certain amount of overbooking on each system. In practice, there are always some projects that do not use the quota that they received. This can e.g., be due to unexpected delays in the availability of software or (input) data. Therefore, the Resource Allocation Committee allocates on each system more allocation units than there are available on the facility. Projects can also apply for extra allocations during an ongoing period in case they have exhausted their current allocation. Extra allocations are also meant for new projects that are not aware of the allocation procedures and deadlines or for projects that would like to investigate the use of other facilities.

The demand for resources

Figure 1 shows the amount of allocation units applied for and the amount of allocation units used in five successive (6-month) allocation periods. One can observe a clear increase in the demand for computing time in successive periods. The figure also shows that the gap between the total demand and total usage has been increasing. Also the amount of allocation units used increases per allocation period but at a lower rate. The increase in usage is obviously limited by the total capacities of the available systems.

Since the demand for resources is larger than the available resources, the Resource Allocation Committee typically must reduce applications significantly in size or (if possible) allocate projects on other systems than they applied for.

Usage by institution

Figure 2 shows the relative usage of computing time per institution. Similar to earlier statistics, UiO and the Meteorological Institute (met.no) are the organizations that make most use of the Notur facilities, with respectively 29% and 19% of the total usage. About 2/3 of the computing time used by met.no (or 12% of the total usage) is used for the operational weather forecasting models that are run on njord at NTNU.

Usage by discipline

Figure 3 shows that the systems are being used by a variety of disciplines. Chemistry, geosciences and physics are the disciplines that make most use of the facilities. Researchers at the universities typically use the local facility (at their home institution) more than the facilities at other institutions.

Advanced User Support

Researchers can also apply to the Notur project for higher-level (advanced) application user support. Advanced user support goes beyond regular user support and aims at improving application performance and functionality, thereby also improving the utilization of the expensive hardware resources. This is typically done by analysing the performance, optimization, tuning and parallelization of applications, by enabling complex applications, and by providing new services to experienced as well as new users of the facilities.

Also the applications for advanced user support are evaluated by the Resource Allocation Committee. In 2007, a total of seven applications were approved. The descriptions of these projects can be found on the Notur web pages.

2008

The old systems tre and magnum were taken out of the Notur system in October and December 2007, respectively. The compute cluster titan at UiO has been expanded with 224 nodes (896 processor cores) and is included in Notur since October. It will be further upgraded late 2007. UiT has installed a large compute cluster, stallo, with 704 nodes (5632 processor cores) that will be available for use in January 2008. UiB will install a new Cray XT4 system with 1388 nodes (5552 processor cores) in January 2008. The three facilities will increase the computational capacity in the Notur project by a factor of ten and it can be expected that this will (only temporarily) bridge the gap between the demand and availability of resources. ●

Relevant links

Notur hardware resources:
www.notur.no/hardware

Allocation procedures:
www.notur.no/quotas/apply

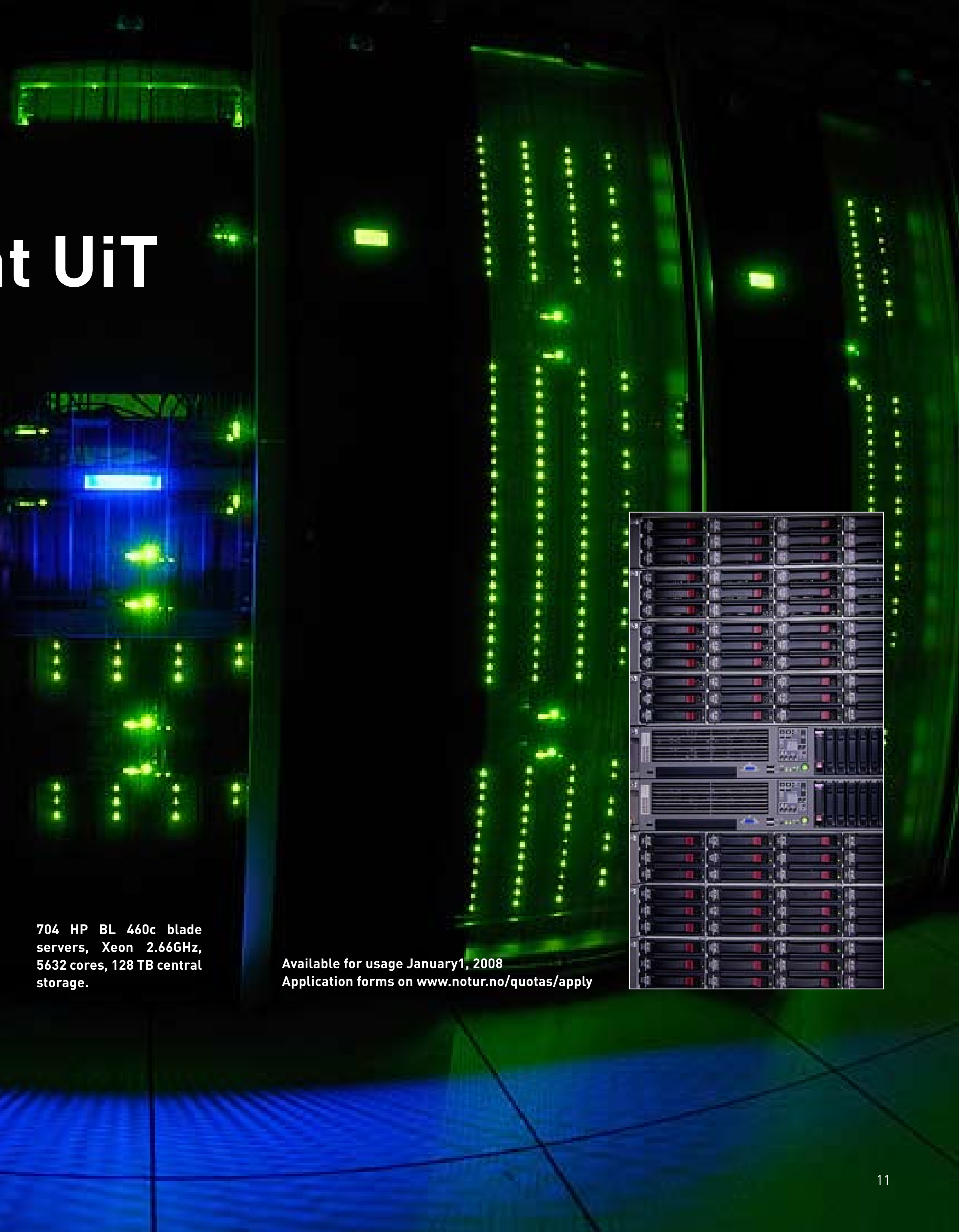
Advanced user support:
www.notur.no/support/advanced

STALLO
- The new supercomputer a

notur

UNIVERSITETET I TROMSØ

t UiT



704 HP BL 460c blade servers, Xeon 2.66GHz, 5632 cores, 128 TB central storage.

Available for usage January1, 2008
Application forms on www.notur.no/quotas/apply

Bridging the gap between computational scientists and HPC

Computer simulation is becoming a new paradigm in many branches of Science. In an everlasting pursuit of more realism and accuracy, by means of complicated systems with millions or even billions of degrees of freedom, high performance computing (HPC) on parallel computers is often the only viable approach. For computational scientists who are not computer scientists by training, however, the threshold of using a multi-processor computing platform can be high. This is rather because of the lack of suitable parallel software than due to the lack of access to parallel computers.

AUTHOR

Xing Cai
Center for Biomedical Computing
Simula Research Laboratory

Let us look at a typical situation for a computational scientist, who has already some fast single-processor code for solving small-scale problems. If she or he is lucky, the serial code may have been based on some numerical software library that has a parallel version. Otherwise, some form of manual parallelization is mandatory, because there exist (at least for the moment) no versatile tools that can automatically convert any serial code to run efficiently on a multi-processor platform. Our aim is thus to simplify and standardize the parallelization procedure, bridging the gap between computational scientists and HPC.

For many scientific applications, the simulation results are sought in a spatial domain over a period of simulation time. A natural way of dividing the work between multiple processors, which is the essence of parallel computing, is to "cut up" the global spatial domain into a set of subdomains. The purpose is to let each processor be responsible for one such subdomain. The global problem can then be solved by asking the processors to (1) solve their local subdomain problems and (2) collaborate with neighboring subdomains through exchanging information.

The coarse-grain parallelization approach mentioned above has a solid mathematical foundation, generally referred to as domain decomposition methods. One subcategory of these methods, named overlapping domain decomposition methods, suit particularly well for parallelizing many types of scientific computing codes. The algorithmic structure is rather simple; every subdomain approaches its true local solution in an iterative procedure, where information exchange is carried out within the overlap zones during each iteration. One particular advantage with respect to software code is that the subdomain local problems are of exactly the same type as the global problem, so that an existing serial code has a very good chance of reuse as a subdomain solver.

To program a parallel simulation code following the strategy of overlapping domain decomposition, there need to be three main components: subdomain solver, global administrator, and inter-subdomain communicator. As we have mentioned above, the subdomain solver can (to a great extent) reuse an existing serial solver. Besides, the global administrator and inter-subdomain communicator can be pre-programmed as generic library components, independent of specific scientific problems. Therefore, the actual programming effort needed in a specific parallelization case consists simply of wrapping up an existing serial code plus some other minor programming work, so that the newly developed subdomain solver fits with the pre-programmed generic administrator and communicator.

In respect of implementation, object orientation is the technique capable of producing a generic framework in which a pre-defined common interface of all possible subdomain solvers is implemented to match the generic global administrator and inter-subdomain communicator. In the terminology of object-oriented programming, all these three generic components can be programmed as classes. When a concrete subdomain solver wants to reuse an existing serial solver, the programming work is simply to wrap up the serial solver with the generic subdomain interface. In other

RELEVANT LINKS

Python: http://www.python.org/
The F2PY wrapping tool:
http://www.scipy.org/F2py
The FEniCs project:
http://www.fenics.org/
Simula Research Laboratory:
http://www.simula.no/

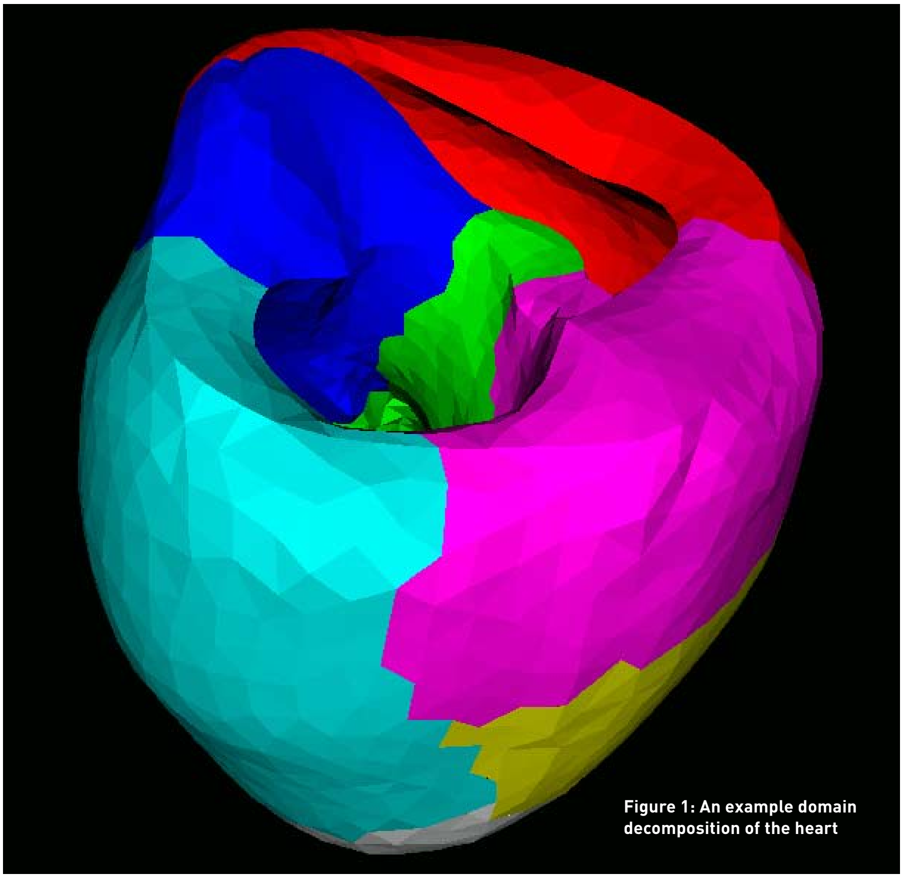


Figure 1: An example domain decomposition of the heart

words, a light-weight subclass needs to be derived. The same idea of programming light-weight subclass is also applicable in case the generic class of global administrator needs to be adjusted for special situations. All in all, a user's programming effort is limited and the actual communication commands (such as MPI calls) can be hidden completely inside the generic communicator component.

An object-oriented software toolbox for parallel overlapping domain decomposition methods was implemented and tested within the Diffpack computational environment (http://www.diffpack.com). Two examples can illustrate the advantage of this software strategy of parallelization. The first example concerns the propagation of electrical signals inside the heart and throughout the body. The physical process can be modeled by a system of partial and ordinary partial differential equations. Analytical solutions to these equations in 3D realistic geometries are not possible by pencil and paper, so numerical simulation is mandatory. Due to some extreme behavior of the signal propagation inside the heart, very fine mesh resolution is needed. More specifically, a spacing of 0.2mm is desired between two computational mesh points. For an average adult heart, the 3D volume needs approximately 50 million mesh points in total. In addition, the body exterior to the heart also needs sufficient numerical resolution, meaning millions of mesh points there. Aiming at such numerical resolutions, an existing serial heart simulation code was parallelized using the software strategy of overlapping domain decomposition. (Figures 1 and 2 show an example of partitioning the heart and body domain into a number of subdomains.) The resulting parallel heart simulator was able to reuse more than 90% of the serial code, requiring only a small amount of new code lines. On an early NOTUR machine (Origin 3000) we were able to run parallel simulations on realistic 3D unstructured heart and body meshes that consist of over 81 million degrees of freedom for involved partial differential equations, and more than 250 million degrees of freedom for involved ordinary differential equations.

The second example actually carries the concept of overlapping domain decomposition further. The physical problem of interest is the propagation of tsunami waves across an entire ocean. Different bathymetry profiles numerically mean that advanced mathematical models should be used in certain "difficult" regions of an ocean, while the remaining vast regions can adopt simpler mathematical models. Consequently, some regions require very high resolution of the computational mesh, but not for other regions. In short, some form of hybridness should be incorporated into the numerical computations, in order to effectively utilize the total computational resource. We divided the Indian Ocean into a set of subdomains to simulate the tragic tsunami of 2004. A few small subdomains used unstructured local meshes of very high resolution, while the other subdomains used structured local meshes of moderate resolution. Accordingly, a sophisticated finite element C++ code was wrapped up as the subdomain solver for unstructured local meshes, whereas a legacy Fortran 77 finite difference code was wrapped up as the subdomain solver for structured local meshes. Parallel simulations of tsunami propagation were realized (see Figure 3 for a simulation snapshot), despite that some processors use one type of subdomain solver, whereas others use a totally different solver. In respect of programming, this hybridness did not impose any extra difficulty, when using the generic library components of overlapping domain decomposition as explained above.

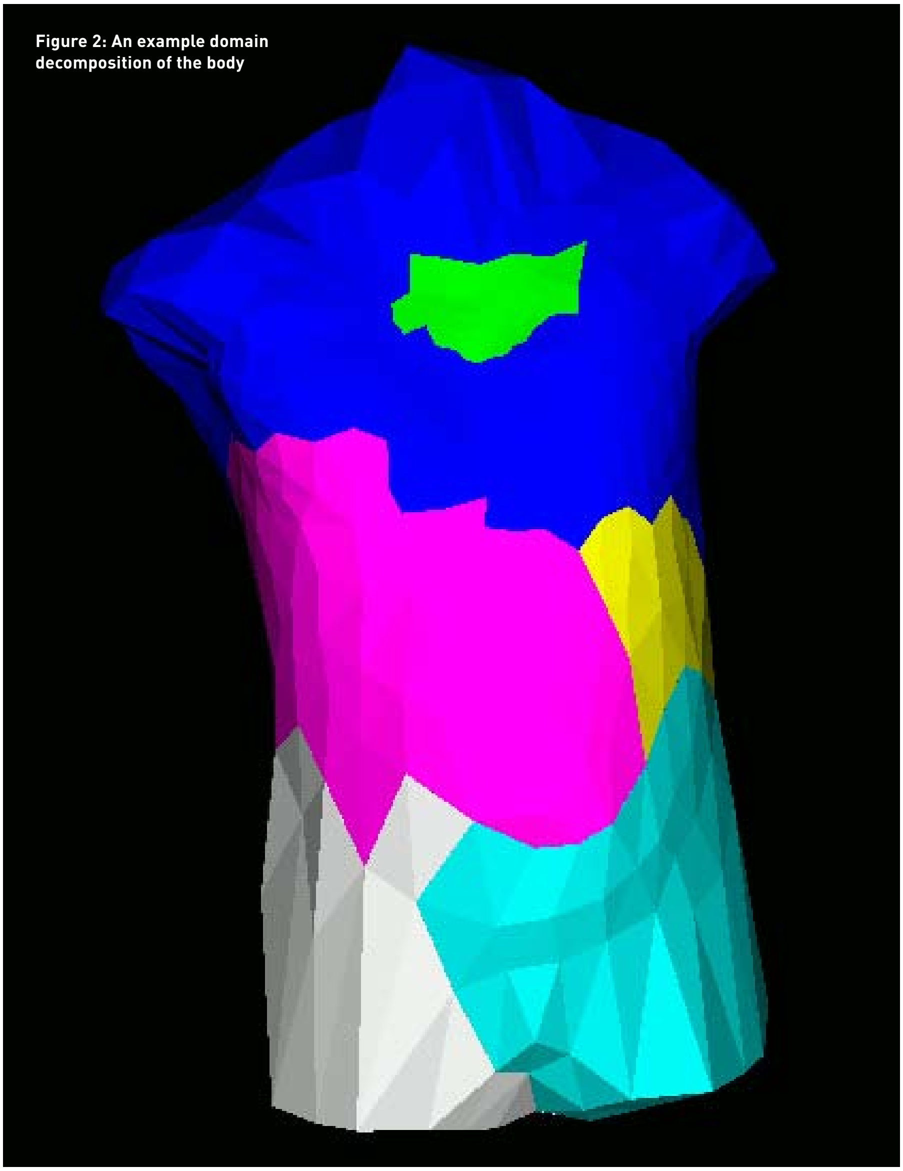


Figure 2: An example domain decomposition of the body

Currently, we are working with an alternative software approach to parallelization. The new key word is Python, which is a modern programming language embracing many attractive features. Due to Python's extremely flexible and dynamic syntax, programming can be further simplified. In contrast with implementing lightweight subclasses in C++, the same effect can be obtained by simply specifying a Python function to work as the subdomain solver. Of course, computational efficiency is by no means the strength of Python. We therefore have to use traditional compiled programming languages such as Fortran/ C/C++ in the underlying numerical code. Nevertheless, wrapping up a Fortran subroutine (or C/C++ function) with a Python interface can in many cases be done automatically by such as the F2PY wrapping tool. We envision this Python-enabled approach to be more user-friendly, while the underlying numerical code written in a compiled language ensures the overall computation speed.

Looking into the future, we believe that the subdomain-based coarse-grain parallelization strategy fits well the upcoming parallel architecture which will heavily involve multicore-processors. A multicore-processor can work as a compute unit, responsible for one subdomain. The challenge is to incorporate parallelism also within each subdomain (possibly thread-based). Anyway, this is no more difficult than parallelizing an entire serial code for a multicore-processor, because each subdomain solver is in essence an individual piece of software.

Take the things a step even further into future, the software gap between computational scientists and HPC may be completely eliminated. This can be true if efforts such as the FEniCs project which aims to develop automated solution of differential equations, adopt an inherent parallel data structure and parallel solution strategy in its underlying software implementation.

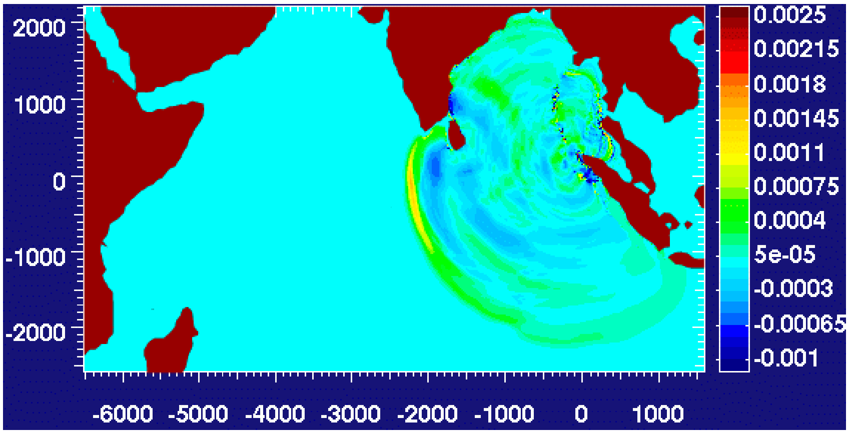


Figure 3: A snapshot of a parallel simulation of tsunami wave propagation

The world's largest grid to pinpoint the universe's smallest fragments

The Large Hadron Collider at CERN in Geneva, Switzerland, could be the most ambitious scientific undertaking ever. It is beyond doubt that the results of the LHC experiments will probably change our fundamental knowledge of the universe and provide a lot of new knowledge about the elementary particles and how the forces between them work. It takes however a lot of hardware, software, data and man effort at CERN and spread across the globe to analyse the results of these experiments. When LHC begins operations, it will produce roughly 15 Petabytes (15 million Gigabytes) of data annually, which thousands of scientists around the world will access and analyse.

AUTHORS

Hans A. Eide, USIT, (UiO)
Csaba Anderlik, BCCS, (UiB)
Jon E. Strømme, UNINETT Sigma

The LHC is being built in a circular tunnel 27 km in circumference and around 50 to 175 meter underground. The LHC is designed to collide two counter rotating beams of protons or heavy ions. It is planned to circulate the first beams in May 2008 and the first collisions at high energy are expected mid-2008. The proton beams are injected into the LHC with energy of 450 GeV and then accelerated to 7 TeV (tera electron volt). The beam moves around the LHC ring inside continuous vacuum chambers which pass through a large number of magnets that bend the beam. The momentum of the beam is very high and these magnets have to produce a very strong magnetic field. The beams will be stored at high energy for ten to twenty hours. Every second, the particles make over 11000 revolutions around the ring. During this time, collisions take place inside the four main LHC experiments which are gigantic underground detectors that try to capture the trajectories of tens of millions of tiny particles that are generated during these collisions. The total LHC power consumption totals around 120 MW.

Smashing protons moving at almost the speed of light into each other, recreates the conditions a fraction of a second after the Big Bang. The LHC experiments try and work out what happened. In the storm of particles generated in the collisions, the physicists hope to find the signature of the Higgs boson. If the Higgs boson exists, the LHC will be able to make this particle detectable. If the particle is found, this would be a strong confirmation of the "standard theory", the way we currently look at our universe. In case the Higgs boson will be found, but has properties different from those predicted, brand new theories are needed that probably will lead to major breakthroughs in many areas of research. Maybe even a unified model of our universe that includes gravity, and can explain a series of cosmological phenomena observed that the current standard model cannot.

ATLAS (A Toroidal LHC ApparatuS) is one of the LHC experiments. ATLAS itself is a complex sensor meant to be a general purpose detector system in the context of the LHC. It will be able to study potentially all interesting processes in proton-proton collisions at LHC energies (14 TeV). It is optimized for the discovery of the Higgs boson, but will also do high precision studies of abundantly produced particles such as B-mesons, top quarks, W- and Z-bosons.

ALICE - A Large Ion Collider Experiment is an LHC experiment aiming to study the properties of Quark-Gluon Plasma (a state of matter where quarks no longer confine into hadrons). In the experiment, heavy ions (such as led Pb or gold Au) are accelerated to velocities close to that of light and then collided. The resulting "Little Bangs" will produce a "hot" enough state to form Quark-Gluon Plasma.

Once LHC is in full production in 2008, the data produced by the four experiments will amount to up to 1 Gigabyte per second. CERN does not have the capacity to store this amount of data. The data is therefore continuously transmitted to facilities around the world in near real-time. At the various locations, the data must be stored reliably and remain accessible for decades. Considerable investments in network, compute, and storage capacity are planned in the coming years. Also the software, tying the heterogeneous resources together, is a major component. To this end, an infrastructure to distribute and manage the data has been developed and is being deployed: the World-wide LHC Computing Grid (WLCG).

TOTAL	2007	2008	2009	2010
CPU (kSI2K)	266	634	1083	1376
Disk (TB)	123	339	527	756
Tape (TB)	0	120	376	726

Table 1: The pledged Norwegian contribution to the processor, disk and tape capacity in the Nordic Tier-1. kSI2K are kilo-SPECINT2000 benchmark units.

To transfer the enormous amounts of data, CERN relies on a layered network, a part of WLCG. CERN itself is the inner Tier-0 layer. There currently exist eleven Tier-1 centres world-wide that must be able to receive and store data from CERN at 50 to 200 Megabyte per second sustained rate. Also some of the raw data processing is done at the Tier-1 centres. Reliability and availability are key components. Data from CERN will be transmitted to two Tier-1 centres at any given time, in addition to maintaining a local cache of a few weeks' worth of data. The Tier-1 centres will make data available to Tier-2 centres, each consisting of one or several collaborating computing facilities, which can store sufficient data and provide adequate computing power for specific analysis tasks. Individual scientists will access these facilities through Tier-3 computing resources, which can consist of local clusters in a university department or even individual PCs, and which may be allocated to LCG on a regular basis. One such network of PCs is implemented at the University of Oslo.

The Tier-1 centres are large, centralized operations, with one exception: the Nordic Data Grid Facility (NDGF) hosted by NORDUnet. NDGF is a Tier-1 centre that is distributed between the four Nordic countries and resources are hosted by seven sites: CSC in Finland, NBI in Denmark, HPC2N, PDC and NSC in Sweden, and UiB and UiO in Norway. In some sense, NDGF is a grid within the world-wide WLCG grid.

The data transfer, storage and processing between the layers in WLCG is provided through grid middleware, adapted and developed by the EGEE project. The EGEE middleware is gLite that is based on the much-deployed international grid middleware Globus. In the Nordic countries, parts of the EGEE middleware are replaced by the "Advanced Resource Connector" middleware (ARC). The ARC middleware is interoperable with the middleware used elsewhere in WLCG, and suited to the distributed structure of the Nordic Tier-1. The storage solution is managed using the dCache distributed storage abstraction software. Resource usage is accounted using SGAS (SweGrid Accounting System). The EU financed project KnowARC has seven researchers working on middleware development. The KnowARC project aspires to improve and extend the existing state-of-the-art technology found in the ARC middleware.

According to the so-called "Memory of Understanding" between the participating countries and CERN, NDGF is pledging to provide about 5% of the total Tier-1 processing and storage capacity. The Norwegian contribution to NDGF constitutes 20% of NDGF or 1% of the world-wide Tier-1 capacity. This does not mean that the Norwegian contribution is small. It means that the handling and processing of LHC data is an enormous task. By 2010, Norway shall have stored ca. 1.5 Petabyte LHC data. See Table 1.

The goal is that the Tier-1 resources at UiB and UiO support all experiments that the

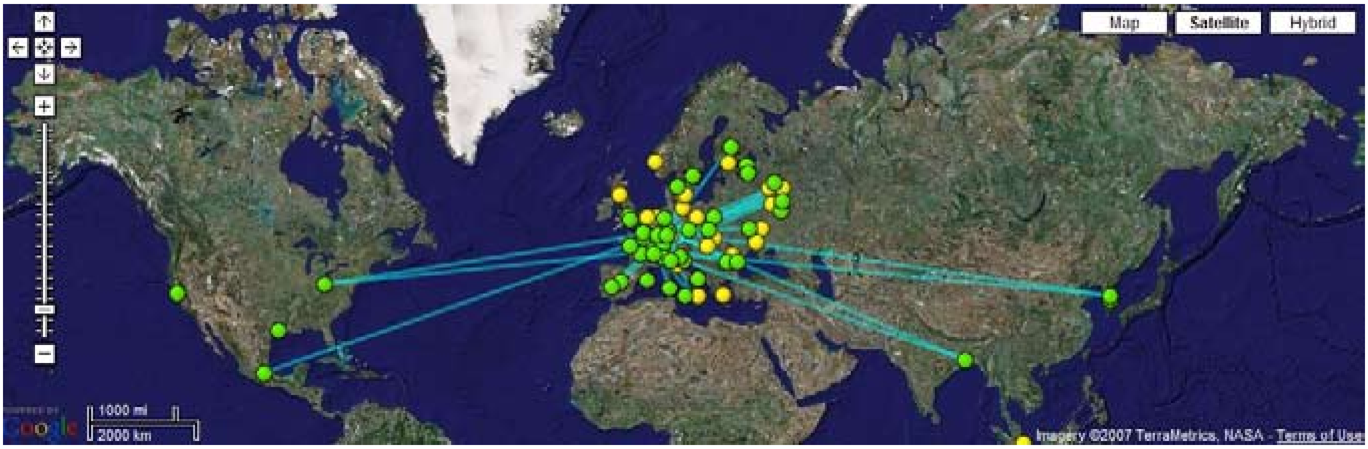


Figure 1: Over 50 sites contributed to distributed production of Monte Carlo data from 2001.

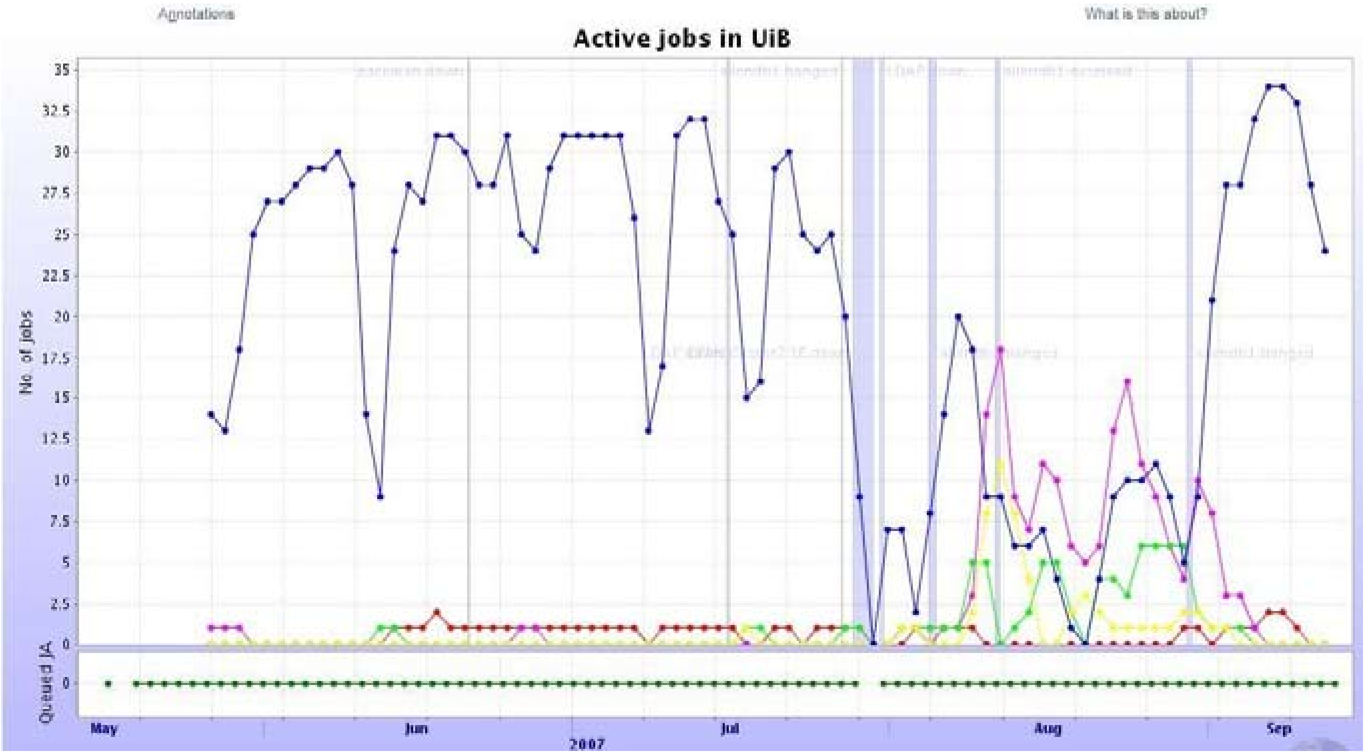


Figure 2: The different states of the jobs submitted to the resources at UiB.

NDGF collaboration is expected to take part in, which will be a great challenge to the software developers. Currently, computations related to the ATLAS experiments are performed at UiO and ALICE computations are performed at UiB. The Tier-1 storage pool at UiO has been operational since early 2007, and was initially the only pool that was operational in NDGF. The storage system has 10 gigabit per second connections all the way to CERN, and the system will in theory be capable of receiving data at 8 gigabit per second sustained.

The dedicated Tier-1 computational resources at UiO are included in titan, a 448-node 1792-core compute cluster which is a joint venture between the various research entities at UiO and the Notur project. Because of the varying demand for computational resources, the NDGF grid computations have benefited greatly from idle processor capacity which can be used by anyone. Thus, the actual Tier-1 and Tier-2 usage of the cluster may at times greatly exceed the currently 128 cores dedicated for this purpose. In the autumn of 2007, the number of ATLAS production jobs at one time was 720, at that time more than half of the entire Nordic contribution. At those load-levels, the stress on various cluster subsystems is high, especially for the networked file system that is used to connect the compute nodes to the storage system.

The ALICE Grid analysis system is based on AliEn (Alice Environment, an Open Source Grid framework). The jobs are controlled by an intelligent workload management system. The analysis starts with a metadata selection in the AliEn file catalogue, followed by a computation phase. Analysis jobs are sent to the sites where the data is located, thus minimizing network traffic. The system was first deployed for ALICE users at the end of 2001, for distributed production of Monte Carlo data, detect simulation and reconstruction, at over 50 sites located on four continents. See Figure 1. The system uses a monitoring tool called MonaLisa. Figure 2 shows the different states of the jobs submitted to the resources at UiB. UiB has contributed to creating the interface between AliEn and ARC to enable the execution of ALICE jobs on the distributed Nordic Tier-1 facility. An intelligent information collector was developed, which transforms information extracted from MonaLisa about the resource utilization of ALICE jobs at NDGF, and publishes this to the NDGF central SGAS service.

MORE INFORMATION:

WLCG: http://lcg.web.cern.ch/LCG

LHC: http://lhc.web.cern.ch/lhc

NDGF: http://www.ndgf.org

KnowARC: http://www.knowarc.eu

NorStore: infrastructure for the curation of scientific data

eVITA, the research programme on e-Science from the Research Council of Norway, has recently established a new project - NorStore – that will deploy a sustainable infrastructure for the curation, archiving and preservation of scientific data.

AUTHOR

Jacko Koster
Managing Director,
UNINETT Sigma

Data curation

Modern collaborative and interdisciplinary science relies on increased sharing of expertise, instruments and computing resources, and, crucially, increasing access to collections of primary research data and information. The sharing of scientific data provides a knowledge base that creates new opportunities and horizons for research and discovery. Researchers rely on the availability of modern information technology tools that assist in the creation, transformation, discovery, re-exploitation and presentation of data. However, these tools evolve rapidly and the flexibility in using these tools puts the data itself at risk. The survival of digital scientific information depends on a hierarchy of constantly shifting technologies – hardware, storage media, operating systems, data formats, applications software and middleware. It also relies on tacit knowledge that is external to the data. In practice, much remains to be done at all levels to keep data usable and valid to future researchers.

At the same time, the sizes of scientific data collections have increased to the Terabyte scale. Large-scale, standardized and quality-controlled data infrastructures are emerging in areas like biology, physics and earth sciences. A well-known example is the Large Hadron Collider (LHC) at CERN that will enter production in 2008. Norway participates in the world-wide collaboration for the analysis of the data that will be generated by LHC and it is expected that only Norway will already store about 1.5 PetaByte (1 500 TeraByte) of data before the end of 2010. Also in other disciplines, high-resolution data is collected from real-time instruments (e.g., sensors) and large complex distributed databases are used. Observational data often concerns unique events and the measured data needs to be stored for a long period (or even forever) as it cannot be recreated. In other scientific areas, large quantities of data are being generated during long computer simulations. Also computer-generated data sets often cannot be recreated easily, as both hardware technologies and the complex layers of software technologies, standards and interfaces evolve continuously.

In today's digital environments, it is not obvious to have trust in data which has been passed on. Trust in data can be enhanced by the existence of qualified domain specialists who curate the data and deal with issues of security, confidentiality and privacy, ownership, provenance, authenticity, integrity, as well as the quality of the primary data and associated metadata. Besides trust, aspects that impact quality include discoverability (e.g., how to find data in foreign domains or out-dated archives), access management, heterogeneity of data formats, and complexity of composite data (possible including links to external objects and external dependencies).

The long-term validity of data collections also crucially depends on the existence of policies and awareness on what to keep and how. Selection of data for use and retention introduces uncertainties concerning who sets the selection criteria, how to assess selection, when, and by whom. In the ideal case, the creators of data are aware of such considerations and accompany data with metadata and proper guidelines for the preservation and curation of the data.

As a consequence, modern science demands increasingly advanced levels of data curation, i.e., strategy, policy and practice regarding the creation, management, and long-term care of data. Curation is about maintaining and adding value to digital information for contemporary and future use. Scientific data collections are not merely stored or archived anymore, but are subject to frequent revision and enhancement. Data repositories which are actively curated to ensure that data fits its purpose and is available for discovery and reuse, have become a reality.

NorStore

The eVITA programme on e-Science from the Research Council of Norway has recently established a new project - NorStore - which shall meet the needs for data storage and data management from several fields within the natural sciences. The primary objective of the project is to establish and maintain a nationally coordinated infrastructure to support the curation, archiving and preservation of digital scientific data. The aim is that the infrastructure shall make scientific processes that rely on access to foreign data sets more efficient and ultimately, support multidisciplinary communities and improve the cross-fertilisation of scientific results. The infrastructure will facilitate the creation and use of digital scientific repositories that satisfy internationally accepted standards and protocols. The sciences that are targeted include, but are not limited to, earth sciences, biosciences, chemistry, physics, material sciences, fluid dynamics, and the medical sciences.

The project will operate large scale data storage resources, provide support for individuals and groups that have a need for storage capacity, digital repositories and curation services, and promote a set of standard services and best practices that aim to improve the reuse and reusability of scientific data. The infrastructure will support easy, secure and transparent access to geographically distributed databases and repositories, provide large aggregate capacities for storage and data transfer, and optimize the utilization of the overall resource capacity that is available in the infrastructure.

In parallel to the operational part, the infrastructure will have a more experimental activity that studies and evaluates new technologies and validates new services before they are put on the production resources.

The infrastructure will be an integrated part in the national e-Infrastructure and will be connected to resources that are located at several major research centers in Norway, including heterogeneous computing systems, networks, other data storage systems, and possibly scientific instruments. The envisaged infrastructure must be sustainable, cost-efficient and allow efficient utilization of the available resources, services and competencies.

The project shall engage into collaborations with parties that have similar objectives, interests and needs for services. National and international co-operations will be established with organizations that have an interest in infrastructure for scientific data and a need for standardization and interoperability of services.

The strategic responsibility for the envisaged infrastructure as well as the prioritization of the user communities that will be granted access lies within the Research Council of Norway. The project will be a broad and nationally coordinated effort. UNINETT Sigma has the operational responsibility. The initial project consortium includes UNINETT and the four universities UiO, UiB, UiT and NTNU. The actual operation of the production systems will be the responsibility of centres that have the expertise, competencies, and the required infrastructure to provide a comprehensive service to meet the challenging demands of academic user groups.

International infrastructure for data

Also internationally there is increasing awareness that there is a need to establish infrastructure for scientific databases and repositories that is deployed according to strategies, standards, policies, and community needs. The European Strategy Forum for Research Infrastructures (ESFRI) has identified a number of strategic infrastructures for European scientists and engineers to remain competitive internationally and to maintain or regain leadership [1]. Several of these infrastructures crucially depend on the availability of large scale storage resources and curation services. A high-performance distributed data infrastructure is an indispensable tool to support the solution of challenging large-scale problems in emerging European infrastructures for large scale supercomputing (e.g., Partnership for Advanced Computing in Europe - PRACE) and federated computing environments (e.g., Enabling Grids for E-SciencE - EGEE).

Recently, the Nordic Council of Ministers recognized the necessity to establish a common Nordic e-Science strategy. A working group that was given the task to draft a first strategy recommended amongst others the establishment of a Nordic infrastructure for digital databases and repositories [2].

2008

The NorStore project envisages the creation of a permanent infrastructure that needs to be designed with care and in collaboration with many parties. The start-up of the project must address models for management, collaboration, operation, support, usage and financing. The project must also define policies and best practices for establishing and maintaining data repositories, define a core set of services, standards and interfaces that shall be maintained across the infrastructure and establish a peer review process to prioritize and support leading-edge science and optimal use of the infrastructure.

Another important activity in the start-up of the project is the establishment of the initial physical infrastructure and in particular the storage resources and services. The storage resources from the first procurement will be installed early 2008. It is envisaged that the infrastructure will be upgraded regularly and at least once a year. The upgrade will consist of expanding the storage capacity and adding new services. Existing resources can be included as well. The criteria for upgrades of the infrastructure will be defined every year and will primarily be driven by user needs. ●

MORE INFORMATION:

NorStore:

http://www.norstore.no/

REFERENCE:

[1] European Roadmap for Research Infrastructures – Report 2006. ESFRI. ISBN-92-79-02694-1. http://cordis.europa.eu/esfri/

[2] Nordic eScience. Research, Education, and Sustainable Infrastructure Services. A strategy document for the Nordic Council of Ministers 2007-07-17.

notur www.notur.no

Return address: UNINETT Sigma AS, 7465 Trondheim, NORWAY

Important dates for the Notur user community

January 15: Deadline for applications for LARGE CPU-hour grants on the Notur facilities

February 28: Deadline for applications for NORMAL CPU-hour grants on the Notur facilities

Both deadlines are for the allocation period 2008.1 that starts April 1, 2008.

PARA'08- Workshop on Parallel HPC

May 13-16, 2008. Trondheim, Norway.

Tentative dates:
February 8: Deadline for extended abstracts
March 7: Notification of acceptance
May 2: Deadline for full papers
May 13: PARA'08 Tutorials
May 14-16: PARA'08 Workshop

For more information see:
http://para08.idi.ntnu.no/

The Notur user survey 2007

A survey was carried out in June 2007 among the user community of the HPC facilities of the Notur project. The full report with the results of the survey can be downloaded from the Notur web pages:
http://www.notur.no/publications/

eVITA winter school and scientific meeting

January 20-25, 2008. Geilo, Norway. http://www.sintef.no/evita

The Notur administration wishes you all
a Merry Christmas and a Happy New Year!

The Notur II project provides the infrastructure for computational science in Norway. The infrastructure serves individuals and groups involved in education and research at Norwegian universities, colleges and research institutes, operational forecasting and research at the Meteorological Institute, and other groups who contribute to the funding of the project. Consortium partners are UNINETT Sigma AS, the Norwegian University of Science and Technology (NTNU), the University of Bergen (UiB), the University of Oslo (UiO), the University of Tromsø (UiT), and the Meteorological Institute (met.no). The project has a 10-year duration (2005-2014). The project is funded in part by the Research Council of Norway (through the eVITA programme) and in part by the consortium partners.