cerit scientific cloud

Contact

CERIT Scientic Cloud
Institute of Computer Science
Masaryk University
Botanická 68a
602 00 Brno
Czech Republic

www.cerit-sc.cz

Director • Prof. Luděk Matyska
Project Manager • Roman Čermák
support@cerit-sc.cz

Mission

Within the context of the national e-Infrastructure, Center CERIT-SC focuses on highly flexible real time computing and storage resources development and provisioning. Virtualized computing and storage resources, available through novel interfaces combining grid and cloud environment, represent a unique installation.

Provision of these resources is complemented with extensive research activities, carried both independently and in cooperation with user communities, targeting the e-infrastructure as well as the other scientific domains.

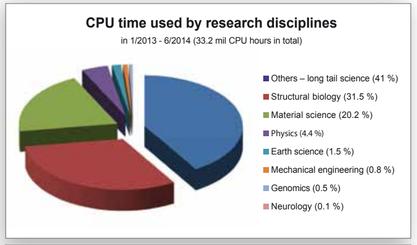
Joint research projects are run, targeted at ICT-assisted solutions of specific scientific problems, as well as novel infrastructure architecture and its configuration, and resulting also in publications with both-sides authorship.

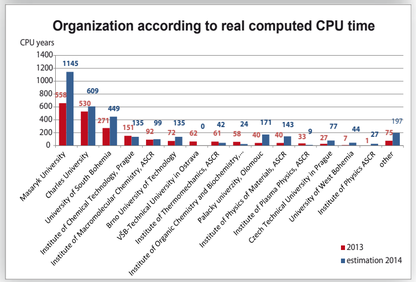
Funding

Transformation of SCB into CERIT-SC was supported by a project of the 3rd axis of the RD&I Operational Programme with overall budget 5 MEUR. CERIT-SC is included in the Roadmap for Large Research, Development and Innovation Infrastructures in the Czech Republic.

CERIT-SC in Numbers

- 4512 CPU cores
- 864 users from all major Czech academic and research institutions
- usage 2.5 mil. CPU hours per month





Equipment

Storage

Storage facilities of two types are available. Standard disk arrays to keep data used for computations on the clusters (630 TB available), and HSM (hierarchical storage management) system (4 PB) built on MAID (massive array of idle disks), intended for large scale and archival storage.

Computing

Most of the resources are virtualized, supporting cloud interfaces (using primarily OpenNebula, but being open to experimental setup of other cloud management systems) and highly flexible access methods. The hardware is composed from computation nodes of three basic types:

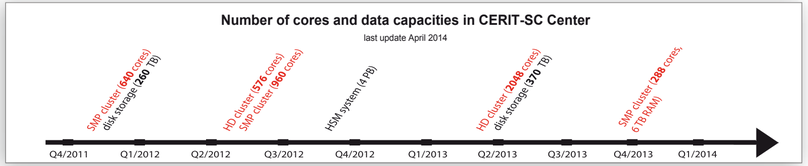
- HD (High Density)—usually with two CPUs (i.e., 8-20 cores) in shared memory, offering maximal computation power/price ratio. They cover the needs of applications with limited or no internal parallelism that can make use of running many simultaneous instances (high-throughput computing)
- SMP (Symmetric MultiProcessing)—clusters with more CPUs (80 cores) in shared memory (up to 1 TB per node), oriented towards memory-demanding applications and applications requiring larger numbers of CPUs communicating with shared memory (suitable for finer parallelisation)
- NUMA (Non-Uniform Memory Architecture) server (SGI UV2)- with extremely large shared memory (6TB) and reasonably high number of CPU cores (currently 288). This machine is available for extremely demanding applications like climate modeling.

High-speed Networking

- 10 Gbit/s uplinks of all the clusters and storage resources to the NREN backbone
- 40 Gbit/s low-latency InfiniBand local interconnect

Software

CERIT-SC provides access to a complex software system based on Debian Linux but covering other Linux flavours and MS Windows, wide range of compilers and developer tools, and application software (both freely available and commercial). Nowadays we have hundreds of applications from tens research areas available , through the whole National Grid Infrastructure.

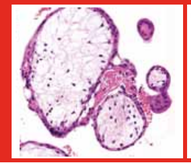
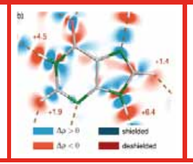


Center CERIT-SC

The national Center CERIT-SC (CERIT Scientific Cloud) continues the tradition of the Supercomputing Centre at the Masaryk University in Brno, providing flexible storage and computing resources and related services for both production and experimental use. Provision of these resources is complemented with extensive research activities, carried both in cooperation with the user communities and in the e-Infrastructure area itself.

EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND
INVESTING IN YOUR FUTURE

OP Research and Development for Innovation


cerit scientific cloud

Interdisciplinary Research

Besides running computing and storage resources, the CERIT-SC role lies in a close collaboration with end users and research teams, either continuing already established research collaborations or developing new ones.

CERIT-SC users become research partners, participating on the research and development activities. Both short-term collaboration over a specific problem and especially a long-term research collaboration (including joint third party funded projects) is supported.

CERIT-SC researchers collaborate with user communities on continuous improvement of services, methods and tools, and they develop new algorithms, programs, forms and means to use the computing and storage capacities, regardless those are owned by CERIT-SC, CESNET, or other parties, including the end users themselves.

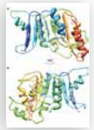
All these research activities include involment of both pre- and postgraduate students. CERIT-SC works closely with the Faculty of Informatics and other parts of the Masaryk University. The research work evolves in a doctoral school with student participation from both ICT and application areas.

Selected Collaborations

Life Sciences

Reconstruction of molecular structures from combined experiments

CEITEC - Central European Institute of Technology



Existing software package CYANA was considerably extended to include molecular-dynamics simulation of the reconstructed system while favouring constraints which reflect the NMR inputs. The result is far more realistic geometry of the reconstructed macromolecule. Parallel processing on multiple CPU cores is used extensively, GP-GPU acceleration is considered.

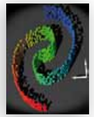
Improvements of robustness and scalability of the Echo corrector of sequencing errors

Department of Experimental Biology, Faculty of Science, MU Brno

The original implementation of Echo, despite of being based on very robust algorithm, was designed to deal with small sized sequencing data. We re-implemented the algorithm to deal with large data and to leverage parallelism—speedup up to 40x and 10x reduction of memory footprint was achieved. The new implementation allows handling large data sets which was not feasible before. The work carries on with defining methods of thorough evaluation of the implementation correctness which is not trivial for a randomized algorithm. Results were published in a journal paper.

Speedup of molecular metadynamics

Department of Biochemistry and Microbiology, Institute of Chemical Technology, Prague



Modified Plumed + Gromacs implementation was deeply analyzed and due to rewriting critical code sections significant speedup was achieved. Opportunities for parallel processing are also opened. The achieved results were submitted as a journal paper.

Transport Research

Vulnerability analysis methods for road networks

Transport Research Centre (CDV)

Road network belongs to important lifelines of the modern society, however, it is often affected by natural hazards whose impacts are both direct (e.g., road damage by a landslide) and indirect (e.g., widespread service disabilities and considerable travel delays affecting population living far from places originally hit by the disaster). The collaboration addresses the problem of evaluating the vulnerability of road network by simulating impacts of simultaneous closure of multiple roads/links. The CERIT-SC team has participated on a development of a sophisticated algorithm focused on identification of network break-ups and evaluating them by various indices. Leveraging several principles from graph theory, the proposed algorithm is able to cope with considerably larger networks and identify the break-ups of requested size for further evaluation in reasonable time, thus greatly improving the previous highly-limited analyse.

Astronomy

Exo-planet photometry

Department of Theoretical Physics and Astrophysics, Faculty of Science, MU Brno

SuperWASP is a recognized extra-solar planet detection tool. It consists of two robotic observatories that operate continuously all year around, allowing to cover both hemispheres of the sky. Data of the SuperWASP archive (more than 3 TB) were rearranged and indexed for efficient search and retrieval.

Cybernetic Security

Cybernetic proving ground

KYPO, Faculty of Informatics, MU Brno

Facing the threats posed by contemporary attackers requires to develop techniques and methods for detection and prevention of attacks continually. However, studying attacks in a real environment is not viable and therefore it is necessary to find other methods to examine their nature. The CERIT-SC team collaborates on the development of a testbed which provides a generic way to simulate and study a wide range of cyber attacks, and it facilitates establishment of isolated virtual environments that researches can use to pursue controlled analysis of attacks. Using virtualization and cloud techniques we managed to provide a controlled/monitored environment where it is possible to configure any common network configuration using a novel networking approach and thus fullfiling needs of various security scenarios.

Environmental Sciences

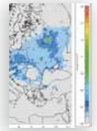
Three-dimensional tree reconstructions from LiDARsScans

CzechGlobe

A unique fully-automated method for tree reconstructions from laser scan measurements (LiDAR) was developed in order to improve the quality of further studies, particularly an estimation of the total chlorophyll content of forest canopies from remote sensing data by means of coupling the leaf-level and the canopy-level radiative transfer models. The algorithm is very general and it can deal with sparse and non-uniform 3D points cloud. The program takes a cloud of 3D points (the tree scan) as input and it produces structure consisting of the trunk, branches, and foliage.

Enhancing regional air quality models to study fate and transport of persistent organic pollutants

Research Centre for Toxic Compounds in the Environment – RECETOX



The collaboration focuses on understanding the processes, important for geographical distribution of persistent organic pollutants in an environment, and time trends of their concentration. A self-learning automated system for iterative parametrization of chemical transport model (CMAQ) is being developed by CERIT-SC, which aims to improve an estimation of initial burden of additional compartments (e.g., soil, water). The overall goal is to improve the correlation of the modelling system results with real measurements performed by active monitoring stations, and thus to understand the pollution processes better.

Industry Collaborations

Large scale simulations for power grids

CEZ/CzechGlobe/Mycroft Mind

CEZ, one of the major providers Czech power grid, looks quite far into the future for infrastructure innovation. Wide introduction of smartmeters and accompanying technology requires large simulations to evaluate its impact on the power grid infrastructure. Recently, we used our testbed to run a simulation of 3.5 million smartmeters, evaluating impact of various behaviour patterns of non-reliable data transfers. The currently approved TACR project extends our collaboration towards processing detailed weather forecast information to predict production of fotovoltaic and wind power plants. This information is used to control the power accordingly, minimizing negative impact of weather changes. The results of the collaborations bring immediate benefits to CEZ, while the experience in running such large scale simulations find their way into independent CERIT-SC publications.

Optimization of large scale Ansys computations

SVS FEM s.r.o.

Large, high-fidelity simulations allow engineering teams to innovate with a high degree of confidence, having accurate and detailed insight into the performance of a proposed design. We prepare and tune-up a distributed computing environment, which is able to run the largest models possible (current tests aim to compute/determine the flow around a large two-shaft gas turbine model). Besides the system optimization itself, the tools making the run of Ansys computations within the environment more comfortable are adapted, namely the Ansys Remote Solve Manager is deployed in a specific way.