

Deres/Your ref.:

Vår/Our ref.:

Trondheim, 2016-11-14

Norwegian academic sector HPC and Storage procurements - Information to suppliers/vendors autumn 2016

This memo is meant to give general technical background information to prospective suppliers/vendors. It may be freely distributed.

UNINETT Sigma2 AS is the Norwegian e-infrastructure for research and education. Procurements discussed here are under the laws covering Norwegian public sector procurements. If you have any questions regarding this, or the contents of this memo, please contact UNINETT Sigma2 at sigma2@uninett.no.

PROCUREMENT BACKGROUND INFORMATION

One of the important objectives for the UNINETT Sigma2 is to have a strong focus on procurement of new HPC and storage equipment. The current HPC production facilities in the e-infrastructure for universities and colleges became operational during 2012. About 50% of the capacity (2 of 4 systems) will be phased out during the first half of 2017, while the last two systems are expected to deliver CPU cycles throughout 2018. A new system is under installation at the Arctic University of Norway in Tromsø end of this year, and will start production in April 2017. A tender for a second HPC replacement system is scheduled for publication in December 2017. This second system will be hosted by the Norwegian University of Science and Technology (NTNU) in Trondheim. The size of this investment, and thus the size of the system, is currently not known, but the size is likely to exceed that of the last investment.

The national storage infrastructure for research is also currently being renewed, with storage facilities in Tromsø and Trondheim. Each storage facility is closely connected to the corresponding HPC facility, and data will in addition be geo-replicated between the two storage sites. A storage procurement is in its final phase, with operationalization of the new infrastructure planned for March 2017.

The current strategy is to continue this dual HPC system service, with a new system being installed every second year, assuming a four year lifetime of each system. No decision has yet been made on where to locate the systems following the upcoming installation in Trondheim.



HPC INFRASTRUCTURE INFORMATION

As of 2016, the Norwegian academic HPC infrastructure consists of four systems, located in Tromsø, Trondheim, Bergen and Oslo (<https://www.notur.no/hardware>). In aggregate, this infrastructure provides approximately 500M core hours per year for national research computations within universities, colleges and to publicly funded research at research institutes. Yearly growth in demand depends heavily on the support services provided, in particular advanced user support, but will most probably be in the range 10-25% per year.

Usage of the national allocation of these resources is spread across over 200 scientific projects, where the three largest projects consumed 59M, 32M and 19M core hours during the last year of allocation. Average yearly consumption per project is 1.6M core hours, and the median is at 0.16M core hours. The infrastructure services more than 200 software applications, with workloads ranging from sequential to distributed memory parallel applications using up to 6000 cores.

Currently, the HPC infrastructure is almost exclusively CPU based, except for 16 nodes with 2 NVIDIA K20x GPUs each and 4 nodes with 2 Xeon Phi 5110P each in the HPC system in Oslo. The load on the accelerator resources has been high for the last two years. Forthcoming systems must thus provide a suitable number of accelerator nodes initially, and we will need flexibility in expanding with more accelerator nodes, depending on uptake. Our current view is that Intel MIC nodes and NVIDIA GPUs are the most interesting accelerator technology for our application portfolio.

At present, our default production compute nodes provide 32 or 64 GiB of memory per node, with 1, 2 or 4 GiB per core. We consider 2 GiB per core sufficient for the default compute nodes. In addition, we anticipate a need (approximately 10% of total) for medium range nodes of approximately 512 GiB, or four times the default compute node memory size, and a very few very large capacity nodes in the 6+ TiB range. In this respect, a road map for possible memory sizes is important.

High performance global parallel storage local to each HPC resource is considered important. Both Lustre and BeeGFS are in use in our systems today, and we have prior experience with GPFS. Metadata operations are an issue with parallel file systems and should be addressed. Our present systems have 100-300 TiB file systems for user's home (with backup), and 300-1500 TiB higher performing scratch/work file systems (w/o backup) to serve running compute jobs. We will prefer solutions where the storage is directly connected to the interconnect fabric, using native interconnect with RDMA or similar, and not to employ a secondary interconnect like a SAN.

From an inquiry about user software done in 2014, the top ten applications in aggregate across all the present national systems are given in the table below:

Rank	Application	Usage [%]
1	VASP	11.26
2	CCSM/NorESM	9.85
3	Gaussian	6.99
4	Bifrost	6.46
5	LAMMPS	5.83
6	Dalton/LSDalton	5.79
7	ATLAS	4.15
8	NAMD	3.88
9	Harmonie (NWP)	3.58
10	ADF	3.05

UNINETT

sigma2

Of the top ten applications, two are codes developed and maintained within the two largest projects (stellar astrophysics (Bifrost) and chemistry (Dalton/LSDalton)), and two are commercial (ADF and Gaussian). The most used application is the TU Vienna VASP. Other important high volume applications are CCSM/NorESM (NorESM is a derivative from UCAR CCSM), LAMMPS, CERN ATLAS, NAMD and the Harmonie NWP model. All the tabulated commercial or restricted applications offer access to program source code, making it possible to build the software for particular, supported architectures, without involvement from external developers.

The code for our largest project is currently being ported to run on GPUs and other projects will look into options for other codes as well, including Open POWER.

STORAGE INFRASTRUCTURE INFORMATION

UNINETT Sigma2 is presently in the completing phase of a procurement of a new infrastructure for research data. The resulting Norwegian Infrastructure for Research Data (NIRD) is planned to be in operation from March 2017. While the contract covering this investment will provide the storage infrastructure with resources for the next 4 - 5 years through multiple upgrades, the investments in HPC resources that will be going on in parallel and are expected to integrate with this infrastructure.

The national e-Infrastructure is adopting a data-centric architecture with this new storage procurement. All research data in the national system will reside on NIRD, and services, including HPC, will be built and operated around it. The chosen technical solution for NIRD allows for tight coupling the HPC resources.

More information about NIRD will be released on www.sigma2.no as soon as the current procurement process is completed

Sincerely,
UNINETT Sigma2