Deres/Your ref.: Vår/Our ref.: Trondheim, 2018-06-05

Norwegian academic sector HPC and Storage procurements – Information to suppliers/vendors 1H 2018

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.

NEW AND UPDATED INFORMATION SINCE PREVIOUS RELEASE

Usage statistics, together with user surveys and discussions with stakeholders, have revealed needs for computational resources with very different characteristics. On one side we have data-driven (memory, IO) applications with low degree of parallelism, and on the other side, a growing need for resources with capabilities suitable for large (tightly coupled) applications. To serve both needs with high efficiency and low TCO, we have decided to invest in two separate technical solutions, and to do so through two separate procurements.

The project name for the capability system is “ANS 2018 - B1”, and for the capacity system “ANS 2018-C1”. There will be two separate competitions for the two systems.

The procurement of C1 will be aimed at resulting in a minimum amount of CPU with the required specifications and be ahead in time relative to B1, the capacity system.

The current general time-schedule is as follows:

C1 - capacity system

- Publication of the Competition: Q2 2018
- Procurement Process: Q3 – Q4 2018
- Installation and Commissioning: Q4 2018 – Q1 2019

B1 - capability system

- Publication of the Competition: Q2 2018
- Procurement Process: Q3 2018 – Q1 2019
- Installation and Commissioning: Q2 – Q3 2019

Postadresse/Postal address
UNINETT Sigma2 AS
Postboks 4769 Sluppen
NO-7465 Trondheim

Besøksadresse/Visiting address
Abels gate 5
Trondheim, Norway

Foretaksnr./Company reg. no.: 814 864 332
Bankkonto/Bank account: 4212.03.37054

E-post/email: sigma2@uninett.no
Tel.: +47 73 55 79 00

Both systems will be installed at NTNU in Trondheim (more info below).

PROCUREMENT BACKGROUND INFORMATION

One of the important objectives for UNINETT Sigma2 is to have a strong focus on procurement of new HPC and storage resources. 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) has been phased out during 2017, while the last two systems are expected to deliver CPU cycles throughout 2018.

The first round of investments in 2016/2017 resulted in a new HPC system and storage infrastructure. The new HPC system, "Fram", (1.1 PFLOP/s, net. 450 MW) was installed at the Arctic University of Norway in Tromsø (UiT) and the storage infrastructure, "NIRD", (approx. 6 PiB) was installed at UiT and Norwegian University of Science and Technology (NTNU) in Trondheim. Both were put into operation in 2017.

The process for the second round of investments, to allow for the replacement of the remaining old systems, is well under way and RFPs to be released before summer 2018. The funding for the second round of investments has been secured.

The placement of the next HPC system will be NTNU in Trondheim. This will complete the intended data-centric architecture for the national e-infrastructure for research, with two storage facilities closely connected to a corresponding HPC facility, and data geo-replicated between the two sites.

The current strategy is to continue with new investments every second year, assuming a four-year operational lifetime of each major HPC system. No decision has yet been made on where to locate the systems following the upcoming investment to be installed in Trondheim.

HPC INFRASTRUCTURE INFORMATION

As of 2016, the Norwegian academic HPC infrastructure consisted of four systems, located in Tromsø, Trondheim, Bergen and Oslo (https://www.sigma2.no/content/hpc-hardware-resources). In aggregate, this infrastructure provided 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 is typically in the range 10-25% per year.

Usage of the national allocation of these resources is spread across over hundreds of 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 8000 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 HPC systems today, while the national infrastructure for research data (NIRD)

is based on Spectrum Scale (formerly GPFS). Metadata operations are an issue with parallel file systems and critical to address for some of our workloads. 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

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 ARM and Open POWER.

STORAGE INFRASTRUCTURE INFORMATION

UNINETT Sigma2 has completed the installation of a new infrastructure for research data, the Norwegian Infrastructure for Research Data (NIRD). 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 are expected to integrate with this infrastructure.

The national e-Infrastructure has adopted a data-centric architecture with NIRD. All research data in the national system will reside on here, and services, including HPC and cloud technology based resources, will be built and operated around it. The chosen technical solution for NIRD, as well asthe location of HPC systems allows for tight coupling of the resources.

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