DeIC SC15 Fact Finding Tour

Austin, Texas, USA, November 2015

In November 2015 representatives from 5 Danish universities (AAU, AU, DTU, KU, SDU) participated in the DeIC-sponsored Fact Finding tour to the *ACM/IEEE Supercomputing Conference* (SC15) in Austin, Texas. Before the conference, most of us attended the HP-CAST conference, and the Intel Developer conference, both also held in Austin.

During the conference several private meetings were held with various vendors of HPC equipment.

The aim of this report is briefly to report the group's findings. Please observe that a major part of our meetings were held under Non Disclosure Agreements (NDA), implying that we cannot reveal issues discussed or information shown to us during these meetings.

Technology update

CPUs

Intel is expected to launch the 4th generation Xeon server processors code-named *Broadwell* in Q1 2016 for the 2P (dual-CPU) versions. Broadwell should feature up to 22 CPU cores per processor, up to about 50 MB of on-die cache, and an increased DDR4 RAM memory speed of up to 2400 MHz. The memory speed options will be more restrictive than in the current processors, and the comparable core clock frequencies will be slightly lower.

Future Intel Xeon as well as Xeon Phi "Knights Landing" many-core processors were outlined under NDA restrictions (see also further below).

AMD is working on a new server processor Opteron series with a redesigned x86 core code-named "Zen". Impressive design parameters were described under NDA restrictions. AMD also mentioned their future ARM "K12" 64-bit processor.

Accelerators

Wrt. accelerators there were interesting announcements from both Intel and AMD.

Intel

Intel announced the next edition of the Xeon Phi coprocessor which will be released in both a PCI-e edition, and a socket edition. The number of cores will be 72 instead of previous 61. Memory will remain 16 GB. The socket edition will also be able to address up to 384 GB of RAM on the motherboard. The next generation of Xeon Phi will also have an option of having Intel's Omni-Path socket on die. The socket edition of the Phi will be an interesting addition to the accelerator market, as the standalone system can have advantages over the PCI-e system both in terms of performance and manageability.

AMD

The most interesting announcement from AMD was their *Boltzmann Initiative*, which is a project to build a heterogeneous compute compiler (HCC). The key benefit will be the ability

to compile and run C++ and CUDA application directly on AMD accelerators providing much more flexibility to researchers to test out platforms. In terms of new accelerators there was nothing exciting. Performance should be on par with similar Nvidia Products, however with a lower power consumption. AMD will be interesting to follow in the coming year or two, as the plan to re-enter x86 CPU market, release an ARM platform, and with new release of the FirePro series, but it is to early to say if the can pull it off with success.

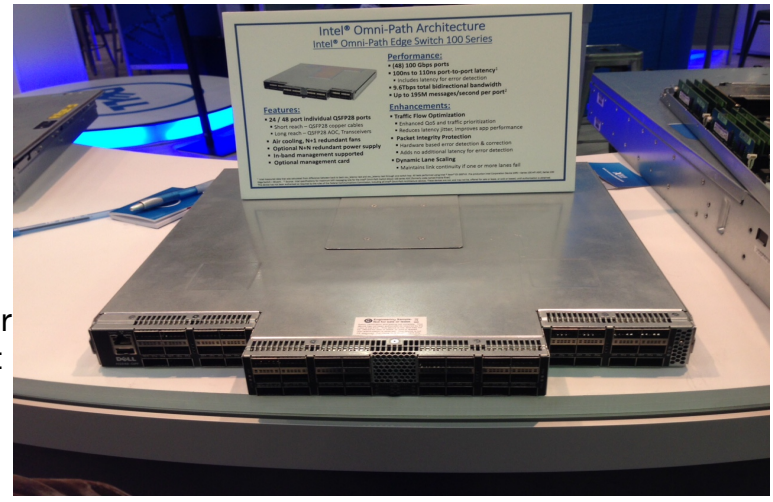
Nvidia

Nvidia had no new announcements at SC15 for HPC. Prior to SC, M40 and M4 was announced; both targeted for machine learning. Nvidia is planning to skip the Maxwell architecture for HPC accelerators, which is likely the cause for no new accelerator for HPC. Of the clusters with accelerators in the TOP500 list approx. 70% of them use Nvidia, i.e., Nvidia is very strong in this field and we will likely see new HPC accelerators in 2016 with the Pascal architecture.

Flash technologies

Flash drives are still getting faster with higher capacity, but they still have problems with write durability. This is the reason that vendors use memory to build drives (Battery backed DRAM, NVRAM, NVMe). Further, Intel's 3D-XPoint technology based on NVMe memory looks interesting with over 4x higher speed, and 1000x longer durability compared to high-end SSD's. This product will enter the market in 2016.

Interconnect



The main players within interconnects are Mellanox (with InfiniBand) and Intel (with their new Omni-Path). Mellanox introduced the possibility to offload some MPI instructions from the CPU to the network card. This should result in lower latency and free up CPU cycles for other purposes.

On the other side, we find Intel with their new Omni-Path. The key features about Omni-Path are:

- Interconnect on die
- 48 ports switches
- Fat Tree configuration

Intel will use their advantage and integrate Omni-Path on their new lines of CPUs - thus avoiding some of the limitations to the bandwidth of the PCIe bus but at the same time adding complexity to the CPU. With 48 port switches, they add 33% extra ports per unit in the rack (compared to InfiniBand). As Omni-Path only supports a Fat Tree configuration, the extra ports might just match the 2:1 sparse tree structure a lot of places use to save on interconnect.

In total, there is no clear picture about the future for interconnects - will it be offloading jobs to the interconnect itself - or will it be everything on die?

Servers

Huwai

Chinese Huawei is entering the HPC-market. More HPC centers around the world, e.g., two University HPC-centres in Poland, have selected Huawei-servers as platform for their HPC supercomputer.

Until now, Huawei is mostly known for providing solutions for the Telecom technology market (ICT) - and their smartphones. Huawei offers an interesting product portofolio, including the well known 4-servers-in-2U concept and a blade server platform. All servers can be equipped with the latest Intel E5-26xx v3 (Haswell) generation chips. Huawei manufacture the motherboard themselves, but most other components in the servers are standard devices from Mellanox (InfiniBand), Hitachi (disk drives), Nvidia (GPUs), etc.

Lenovo

IBM has transferred all their x86 products to Lenovo (also a Chinese company). Lenovo offers various products for the HPC-market, including rack, high-density and blade servers. The storage product previously known as GPFS will be marketed as "Lenovo GSS". Lenovo presented their roadmap, including support for the new generations of CPUs from Intel (Broadwell, Skylake) in their products.

HP/HPe

Hewlett-Packard (HP) recently split into two companies, HP inc. and HP Enterprise. HP will sell PCs and printers, while HP Enterprise will sell commercial (HPC-)computer systems, software, and tech. services. HPe continues the well known server families for HPC, the Apollo 2000 and Apollo 6000.

Dell

Dell now offers a PowerEdge R930-system with up to 4 CPUs and 6TB memory. This server is capable of dealing with the most memory demanding applications. The C6320 product line will be supporting Intel’s new Broadwell CPUs, which is expected to be generally available in the first half of 2016.

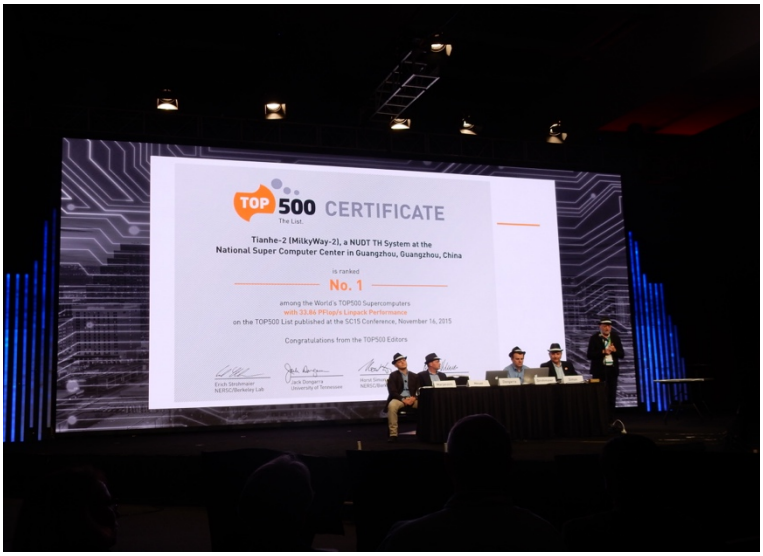
Storage

Storage is still moving towards Software Defined Storage (SDS) due to flexibility, transfer speed and not least the issue to capacity on a single drive. Capacity per drive is an issue that drives vendors to using new technologies like filling hard disk drives with helium to prevent the spinning disks to vibrate, and laser to heat the surface giving the ability to store more data.

Middleware / Software

The Slurm scheduler is becoming increasingly popular in particular among HPC sites who previously used Torque/MAUI and want to switch to a maintained, open-source scheduler. The same development is seen at our own Danish universities.

Among interesting talks, we heard about software module systems (LMOD) and package building systems (Easybuild and Spack), both looking very interesting to smaller HPC sites. Further, Intel has now introduced an interesting Intel MKL optimised python distribution with included NumPy and SciPy.



Announcements

Selected entries from the November 2015 TOP500 list:

Rank on TOP500	Name	Country	Performance
1	Tianhe-2	China	33.8 Pflops
2	Titan	ORNL, USA	17.6 Pflops
3	Sequoia	LLNL, USA	17.1 Pflops
4	K computer	RIKEN, Japan	10.5 Pflops
7	Piz Daint	CSCS, CH	6.2 Pflops
236	Computerome	DTU, DK	410 Tflops
267	Abacus 2.0	SDU, DK	363 Tflops
500	Qpace2	Germany	206 Tflops

USA is clearly the leading customer of HPC systems with 199 of the 500 systems although its share has dropped sharply to an all time low in the new November 2015 list. The European share (107 systems compared to 141 last time) has fallen and is now lower than the Asian share of 173 systems, up from 107 in June 2015. Dominant countries in Asia are China with 109 systems (up from 37) and Japan with 37 systems (down from 39). In Europe, Germany is the clear leader with 32 systems followed by France and the UK at 18 systems each.
In one year Computerome has fallen from 121 to 236.
The last entry on the list is 53 Tflops faster than last year's number 500. Last year this machine would have been number 295.

Green500
The purpose of the Green500 is to provide a ranking of the most energy-efficient supercomputers in the world. Several rules apply in order to get onto the list, see

http://www.green500.org
Abacus 2.0 is number 81 on the GreenTop500 list which is the best ranking among supercomputers in the Scandinavian countries.

Next SC

The next Supercomputing Conference, SC16, will be held in Salt Lake City, Utah in November 2016.