DeiC Super Computing 2019 Report

Fact Finding Tour at Super Computing 19, November - Denver, Colorado

Boye, Mads; Nielsen, Ole Holm; Bortolozzo, Pietro Antonio; Hansen, Niels Carl; Happe, Hans Henrik; Madsen, Erik; Visling, Jannick

Publication date:
2020

DeiC Super Computing 2019 Report

Fact Finding Tour at Super Computing 19, November - Denver, Colorado

Boye, Mads - mb@its.aau.dk
Bortolozzo, Pietro Antonio - pbor@dtu.dk
Hansen, Niels Carl - ncwh@phys.au.dk
Happe, Hans Henrik - happe@nbi.ku.dk
Madsen, Erik B. - erikm@sdu.dk
Nielsen, Ole Holm - ole.h.nielsen@fysik.dtu.dk
Visling, Jannick - janvi@sdu.dk

2020-02-24

Contents

1 Preface

The SC19 International Conference for High Performance Computing, Networking, Storage and Analysis was held in Denver, Colorado where a Danish delegation represented the five Danish universities:

- Aarhus University (AU)
- Technical University of Denmark (DTU)
- University of Copenhagen (KU)
- University of Southern Denmark (SDU)
- Aalborg University (AAU)

The purpose of the present document is to briefly report the delegation's findings. The trip was subsidized in part by the Danish e-infrastructure Cooperation (DeiC). Prior to SC19 conference, parts of the delegation attended various events, such as a DELL EMC visit in Round Rock, HP-CAST 33 conference and the INTEL HPC Developer Conference 2019. During these conferences the delegation attended a number of Birds of a Feather sessions (BoF), Talks, Workshops, and Technical Sessions. In addition, the delegation had arranged meetings with a number of vendors to gain knowledge of hardware and software developments within the High Performance Computing (HPC) field. Information obtained during vendor meetings was mostly under Non-disclosure agreements (NDA), and cannot be disclosed in this report.

2 Satellite Events

Prior to the SC19 conference, it was possible to visit various events, which is not part of the technical program for SC19. The delegation covered these events, based on interest. The knowledge obtained at these events are covered in the following section.

2.1 DELL EMC Visit in Round Rock

The visit at DELL EMC had a day with NDA briefing at DELL EMC headquarters in Round Rock, Austin, TX. Without breaking the NDA it is safe to say that the future of HPC is very power hungry, which much of our talks revolved around. And we had insights into how DELL EMC deals with that.

We also got some interesting information about performance of the current CPUs. Especially the newly released AMD EPYC 7002 series (Rome). DELL has published some of the performance results, together with curiosities about the CPU architecture[13]. Especially the "InfiniBand bandwidth and message rate" section of the "AMD Rome is it for real? Architecture and initial HPC performance"[13] article is important for HPC.

DELL EMC also presented upcoming additions to their HPC storage portfolio, which already include Lustre, BeeGFS, NFS and Isilon.

The Data Accelerator from University of Cambridge was also introduced. This work shows how to make a cost-effective burst buffer with standard NVMe servers[14].

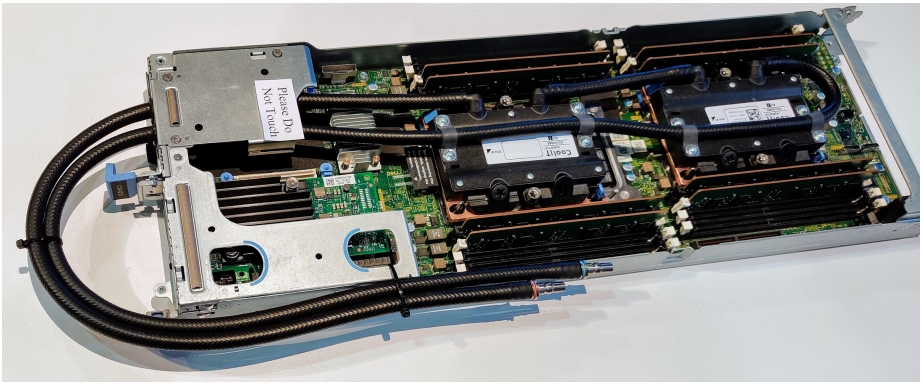

Figure 1: Dell PowerEdge C6420 server with custom water cooling for CPU.

The DELL EMC HPC and AI Innovation Lab has published a lot about their AI findings[15].

2.2 HP-CAST 33

HP-CAST

Figure 2: HP-CAST33 Logo

This year was permeated by the fact that Hewlett Packard Enterprise (HPE) bought CRAY almost two years ago and the merge became ratified a year ago.

All of the sessions was presented by at least two people on the scene, one from the old CRAY, and one from the HPE part of the consortia. The same message was repeated again and again: "We will not let any costumer down. We will support everything you bought, and in the future build products with the best of both worlds."

There were some interesting words for us university people.

HPE see a clear path for the CPU market in the upcoming years. As a fact they believe suppliers will be capable of producing CPUs down to 2 nm technology, maybe down to 1.6 nm technology, within the next two decades, if you are to believe HPE, the market needs new technologies, and here they ask for help on technologies like analog, quantum, biologic, or other principles that can get Flops pr. Watt at a higher level that the transistor can bring to the market.

HPE also has a clear view on an exascale computer model, and the most interesting idea here is the network. HPE believes that the interconnect is ethernet with a speed that can be compared with InfiniBand, but with new technologies, called Slingshot to prevent congestion.[1]

HPE is still looking at the old concept of The Machine[19], but nowadays it is in the form of Gen-Z, where HPE is working on a bus structure as the heart of the super computer and not the CPU. They are now capable of producing light emitter switches that can be used for this, and HPE believe, that we soon will see different forms of products designed to be used under standards of Gen-Z e.g HPEs Cadet project[2].

2.3 Intel HPC Developer Conference

Figure 3: Presentation slide from Intel HPC Developer Conference 2019

This year's Developer Conference was more of an excuse than a tell of how the HPC developers world will be the next coming years. The few workshops compared to earlier years, had a bad setup where only few people were capable hearing the speech of the speaker. Seen from a developers sight, the only new thing was OneAPI. An idea of a compiler that chooses the right hardware from whatever the compiler is offered.

Intel gave a sneak peek of the future. Intel is coming up with a new packing of CPUs (Intel Next Gen 7 nm Process & Foveros packing), where CPUs will be built in layers. And a few slides later Intel also revealed a new GPU (code name Ponte Vecchio) under the name Xe. This is what Intel needs to build one of the exascale computers in 2021 named Aurora.

A really interesting point of the HPC Developer conference is that Intel does not mention OmniPath or a new fabric product to follow.

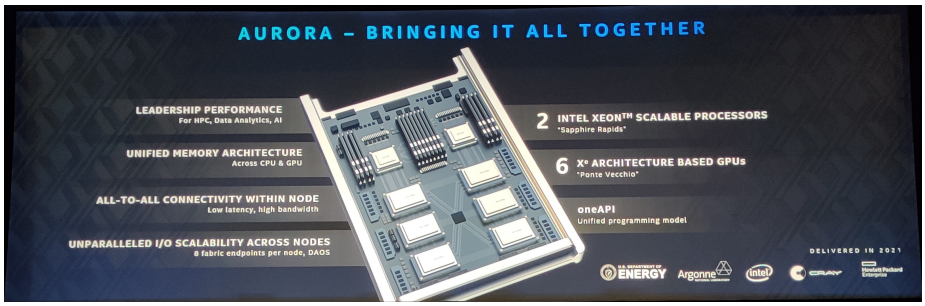

Figure 4: Presentation slide of how a compute unit of the Aurora Exascale super computer will designed.

2.4 SC19 workshops

Super Computing 19, provides a wide range of workshops, covering all areas of HPC professions. The attended workshop and the main conclusions is summed up in the list below:

- Reproducible computations in exascale.

  There is a need of better verification methods that match large data set.

  Computational modules must be robust, as some fields of science have much noise in data set.

  Large distributed jobs must be analyzed as global summation and rounding, can have a larger impact.

- Training of system professionals and users.

  A common challenge for HPC centers, is that the usage of HPC is growing to new fields of science, where there is no history for programming. This causes much bad code to run on cluster, which can cause the cluster to run inefficient.

  This problem was solved in different ways for the presenters:

  Oregon State University: Provides courses for scientific staff on how to use collaboration tools, source code licensing, structuring of python packages, object oriented programming among other things.

  Los Alamos National Labs: They provided a long training session, where the attendees would have two weeks during the summer, to setup a HPC cluster, starting form a bunch of servers, cables, and an empty rack, to a running system. The following weeks, there would be different tasks for learning to manage, bench marking, and running applications on the cluster. The feedback they have from the attendees was, that having a better understanding of the underlying hardware, helped them a lot in designing future code and applications.

- HPC and cloud.

  Cloud seems to be more adopted at many HPC site, and has gotten a bigger presence at the super computing conference. Most sites use it

as an addition to HPC, for offloading parts of jobs pipelines, which is not distributed computations. Furthermore it allows for self-service and testing facilities for the users. Cloud does not solve all issues, to name a few of the challenges to be solved with cloud is: Security in containers, and data synchronizations.

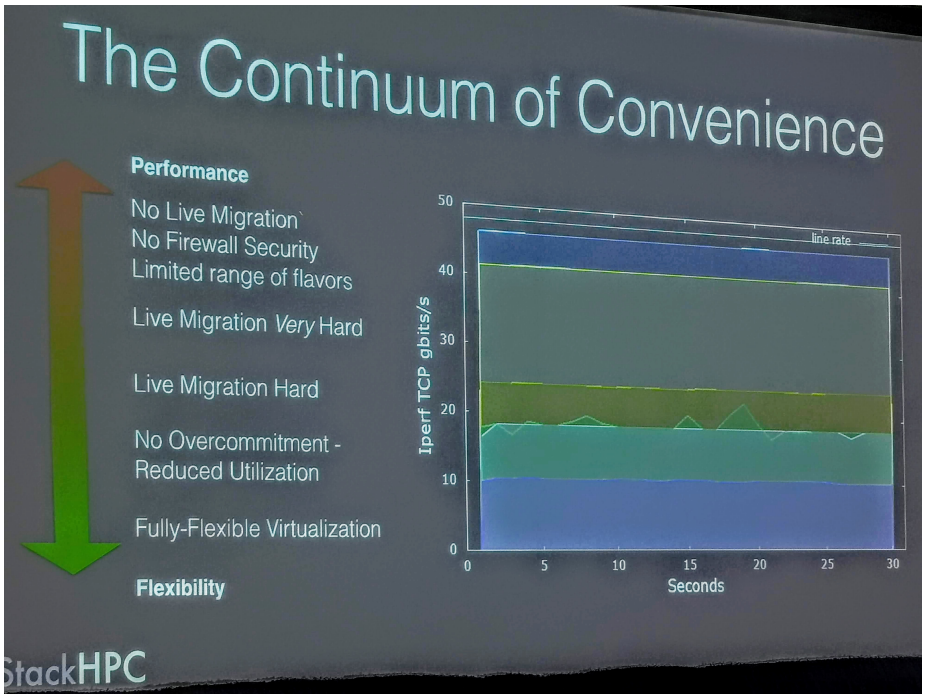

Figure 5: Presentation slide from *Cloud and Open Infrastructure Solutions to Run HPC Workloads* workshop.

3 Central Processing Unit (CPU)

3.1 AMD

The AMD EPYC processors are now making an impact in the HPC arena with currently 4 systems on the TOP500 list and many more expected during 2020.

The EPYC processors are comprised of multiple chiplets containing 1-8 CPU cores and connected to a central memory/IO die for a total of up to 64 CPU cores. The latest AMD Rome Zen 2 CPU is now competing head to head with the Intel Xeon processors, and for some applications AMD Rome is the winner with more CPU cores, more and faster memory channels, and more and faster PCIe Gen4 bus lanes. With PCIe Gen4 AMD Rome supports 200 Gbit/s network fabrics based on InfiniBand or Ethernet technology. The power usage may increase up to 280 W per processor.

The delegation held an NDA meeting with AMD and received detailed information about the next-generation Milan Zen 3 and the future Genoa Zen 4 processors.

The future US exascale system Frontier to be delivered in 2021 by CRAY will be based upon AMD CPUs and GPUs.

AMD delivers the AMD Optimizing C/C++ Compiler (AOCC) which is based on LLVM, and is optimized for new AMD CPU's. The compiler is open source, and AMD is working together with Spack, to have it integrated. AMD would also look into support with EasyBuild.

3.2 Intel

The delegation had an NDA meeting with Intel and received information about Intel Xeon CPUs for HPC. Detailed information was given about the current Cascade Lake processor as well as the coming Cooper Lake and Ice Lake processors which are planned for 2020. New CPU instructions for AI/ML were described. The chips will go from 14 nm to 10 nm technology, but the power usage may still increase up to 270 W per processor, making liquid cooling an important consideration.

The future Sapphire Rapids CPU architecture, planned to debut in the first US exascale system Aurora in late 2021, was described under NDA. However, some details are publicly announced[5].

4 Accelerators

4.1 AMD

AMD is releasing the AMD Instinct MI100, which is their accelerator platform. A challenge for them is that most Artificial Intelligence (AI) software is build on top of the CUDA software stack. AMD puts a lot of effort into their ROCm compiler which helps cross compile CUDA code to OpenCL for execution on non Nvidia GPUs. A new addition is the Heterogeneous-Compute Interface (HIP) which target to make portable C++ programs which can run on both AMD and other GPU's.

4.2 Intel

Developing a new GPU is an ambitious goal for Intel, especially considering its failed Larrabee project in 2010, but Intel is back into the graphics game. The company will debut the X^e graphics cards in late 2020 (code name Arctic Sound), which is Intels first new plug-in graphics card since the Intel740 released back in 1998.

The technology behind this GPU and its potential performance remain hidden in mystery, but as time goes on details are coming to surface. If it proves a viable alternative to Nvidia and AMD, this will be an important event in the graphics card industry for years. At the Intel booth they mentioned two variants of the X^e architecture: one optimized for consumer clients (and thus, probably gaming) and the second optimized for data centers.

But the overall opinion among HPC evangelists is that Nvidia and AMD will not fear the competition from Intels first appearances in the GPU market for years. The next generation Ponte Vecchio might be a turning point for Intel. For now, information on the technology is close to zero but according to Jim Keller, Intel chief architect, Ponte Vecchio will be manufactured on Intels 7 nm

technology and will be Intels first X^e-based GPU optimized for HPC and AI workloads. Ponte Vecchio will also leverage Intels Foveros 3D (three-dimensional integrated circuit (3D IC) face-to-face-based packaging technology) and EMIB (embedded multi-die interconnect bridge packaging), high-bandwidth memory (HBM), Compute Express Link interconnect among other technologies.

Along with Jim Keller, Intel have assembled an expert team like Raja Koduri and Chris Hook highly skilled architects and graphics designers in the development process.

4.3 Nvidia

At Nvidia it is business as usual. Nvidia has been rolling out graphics cards non-stop, and is working on the next-generation GPU technology. Nvidia will launch its 7 nm GPU architecture (code name; Ampere) in 2020. It will be a whole new architecture compared to the Titan Volta technology with faster tensor core, support of FP16 & bfloat16, faster and more memory, PCIe Gen 4, New hardware engine.

Compared with todays Titan Volta V100 cards we will see a 2 times AI training speedup, 2 times performance increase with FP64 flops, 8 times larger L2 cache, 2 times NVlink bandwidth, 2 times PCIe bandwidth, JPEG HW decoder and Direct Storage a technology that transfer data direct from storage to GPU memory and vice versa.

The future in GPUs is exciting and there is no doubt that the big HPC GPU manufactures is following each other very closely. It is clear that Intel is the vendor the highest risk. A lot depends on Ponte Vecchio. Nvidia does not disclose much information, but the fact that Ampere is replaced by new technology at the entrance to the exascale period puts even more pressure on Intel. In an interview where Jensen Huang, CEO of Nvidia, was asked about Intels GPU plans, Jensen said:

"Im anxious to see it, just like you probably are. I enjoy looking at other peoples products and learning from them. We take all of our competitors very seriously, as you know. You have to respect Intel. But we have our own tricks up our sleeves. We have a fair number of surprises for you guys as well. I look forward to seeing what theyve got."[10]

4.4 FPGA

At the Intel booth one stated:

"The development of FPGAs has evolved like the Swiss Army Knife. Meaning, the fabric (LUT[11] - are building blocks of reconfigurable computing fabrics, providing a 2 x 1 bit memory capable of realizing an m-input Boolean logic function) is the one thing that uniquely gives FPGAs their superpowers. Taking full advantage of FPGAs requires digital logic to be designed, and the technology are not yet at the point where FPGAs can be optimally used without at least some degree of hardware expertise in the design process."

There are a lot questions to that statement. First, after years of dealing with the huge (and ever-increasing) complexity of FPGA design, FPGA vendors have come up with a few tricks to mitigate this issue. The easiest and most significant is the use of predesigned single-function accelerators and applications.

The second is raising the level of design abstraction and design expertise. The lowest level of design in the FPGA is place and route. This process takes a net list of interconnected LUTs, arranges it on the chip in a near optimal fashion, and makes the required connections through the programmable interconnect fabric. Place and route is the heart and soul of FPGA implementation and the better the place and route algorithms perform, the fewer routing resources are required to complete most designs.

At the same time, Intel reminded us to keep aware that "FPGA developers will need to modify their code for optimal performance on the FPGA or the use of libraries that have been pre-optimized for the FPGA architecture."

5 Interconnect/Fabric

5.1 Mellanox

For Mellanox it seems to be business as usual. They are continuing their road maps with 200G and 400G InfiniBand / Ethernet. Mellanox is in a good spot, with Intel discontinuing their OmniPath network solution in near future and has yet to introduce a CPU supporting PCIe Gen4 it seems that Mellanox will keep their upper-hand for the time being.[7]

PCIe Gen4 is the next step for the interconnect. Mellanox is already shipping 200G which requires the CPU to have PCIe Gen4 giving AMD the advantage for now to sustain a system with 200G network.

A possible challenge for Mellanox, could be ethernet, as many hardware vendors is focused on this technology, including Mellanox themselves. HPE is now working together with CRAY and introducing their own network fabric/switches called Slingshot.[8]

Intel will have PCIe Gen4 soon and will introduce a network platform for Ethernet. The race is on since Mellanox is already working towards 400G InfiniBand / Ethernet. They also have a gateway product to connect the two network types. If the interconnect will switch to Ethernet in near future you will still be able to keep and connect your current InfiniBand network with it. With 400G InfiniBand / Ethernet, we will likely see the need for PCIe gen.

Cooling of the optical transceivers and NICs becomes necessary, as more speed means higher temperatures. Mellanox reports that max wattage should be in the range of $25W \pm 3W$ so air cooling should be possible also in the future.

Mellanox's merger with Nvidia could also have a big impact on the interconnect work. It will be exiting to see which new development will come of this. Announcements is expected during 2020.

Mellanox is also continuing the road map for the system on a chip project called Bluefield. Bluefield 2 is released with support of PCIe Gen4, an array of ARM processors and the newest Connect-X6 NIC with 200G InfiniBand / Ethernet.[9]

5.2 OmniPath

The Intel OmniPath 100 Gbit fabric has given competition to Mellanox's InfiniBand HPC network fabrics, however, Intel has stated that they have canceled the next-generation OmniPath 200 Gbit products.

The current OmniPath 100 Gbit products will still be sold for some time, and installed products will continue to be supported. Intel has recently bought the Ethernet silicon and software company Barefoot Networks, leading the market to speculate if Intel plans to launch HPC fabrics based upon Ethernet technology in the future. This leaves Mellanox as the sole provider of HPC network fabrics, at least for some time.

6 Servers

6.1 Huawei

The Huawei company is a large IT provider based in China. In Denmark Huawei has recently joined the SKI government procurement agreement.

The delegation had a meeting with a representative from Huawei USA, although Huawei was not present on the SC19 show floor owing to recent events between USA and China.

Huawei offers a number of HPC products:

1. Intel-based servers with a roadmap for Cooper Lake, Ice Lake, and Sapphire Rapids Xeon processors. Huawei offers direct liquid cooling solutions for CPU and RAM, covering about 75% of the cooling needs.

2. ARM v2 processors are produced by Huawei and integrated into a number of solutions, including also HPC servers with a PCIe Gen4 bus. Huawei is working on the next-generation ARM v3 with improved floating-point capabilities.

3. Huawei has launched the AI processor "Ascend 910" and an AI computing framework "MindSpore", supporting also TensorFlow and Pytorch.

Huawei offers standard HPC products with NVIDIA GPUs and Mellanox InfiniBand, but they do not plan to offer any AMD-based servers.

Huawei is offering Open Compute Project (OCP) OpenRack based solutions where servers are blades in a water-cooled rack infrastructure called "FusionPod" with rack power levels up to 50 kW.

6.2 Lenovo

The Lenovo product line continues to be based primarily on Intel CPUs. For now Lenovo continues to offer both Mellanox and OmniPath interconnects, as well as support for the Nvidia GPUs.

In modern HPC-installations the power consumption per rack increases dramatically and Lenovo offers some very interesting solutions regarding liquid cooling of nodes and racks. On the Lenovo booth, the "Liebert XDU"from the company Vertiv was shown. The Liebert XDUprovides liquid cooling in a air-cooled environment thus saving costs to retrofit the data center with water-cooling.

Figure 6: The Liebert XDU provides liquid cooling in an air cooled environment.

7 Storage

7.1 BeeGFS

BeeGFS is planing a rewrite of the entire software stack. This is big news, but not much else was presented. It will be interesting to see what will come of this.

7.2 Ceph

Ceph is seeing a larger adaptation at HPC installations. The Ceph BoF was well attended. The main take home messages were:

- Next version of Ceph (Octupus), will have a new orchestrator API, and a new dashboard.
- Pg_autoscaler will be enabled as default. Pg_autoscaler is a tool for sizing the placement group of the Ceph pools. This has been a manual process when creating Ceph pools, and it has only been possible to scale up, not down.
- Ceph will now have an option for sending crash reports to the developer. It is an opt in.
- QOS will be implemented in RADOS.
- The bluestore queue will be reimplemented.
- Better cephFS file creation.

7.3 Lustre

Lustre is moving forward as an open-source project after DDN took over last year.[16]

The roadmap is laid out and new releases and bug fix releases come out at a steady pace[17].

CRAY and DDN are the main vendors delivering Lustre systems. They even offer all flash or tiered setups to cope with the high IOPS produced by AI and life-science workloads. Other vendors left us with the complaint that DDN has ended the level 2+3 support that Intel used to supply.

7.4 DAOS

A big announcement during the Intel Developer conference was that Intel would use Distributed Asynchronous Object Storage (DAOS) as the storage system for the Aurora exascale supercomputer in 2021. It should be able to sustain 25TB/s of IO. Conceptually DAOS is an open-source project that looks a lot like Ceph, but is much more focused on performance. With this commitment it could become the Ceph of HPC[18].

8 Middleware / Software

8.1 Slurm

The Slurm developers held an SC19 BoF session with an estimated 2-300 participants. A review of the latest Slurm 19.05 release was given, highlights being a much better handling of GPU resources through a new "cons_tres" plugin. Also, the X11 graphics has been improved.

The upcoming Slurm 20.02 will feature "config-less" Slurm for large clusters, some additions to Burst Buffer handling, and a new interface to prolog/epilog scripts. A new REST API service will be added in 20.02 or 20.11.

8.2 OneAPI

As described in the section on Intel HPC Developer Conference, Intel has launched a new cross-architecture programming interface named OneAPI, for transparently writing code to multiple types of hardware including CPUs, GPUs, FPGAs, and other accelerators. The OneAPI will be needed for the "Aurora" exascale machine described above comprising new CPU and GPU hardware. Intel is making OneAPI available for its competitors, proposing it as a unified programming interface across all platforms.

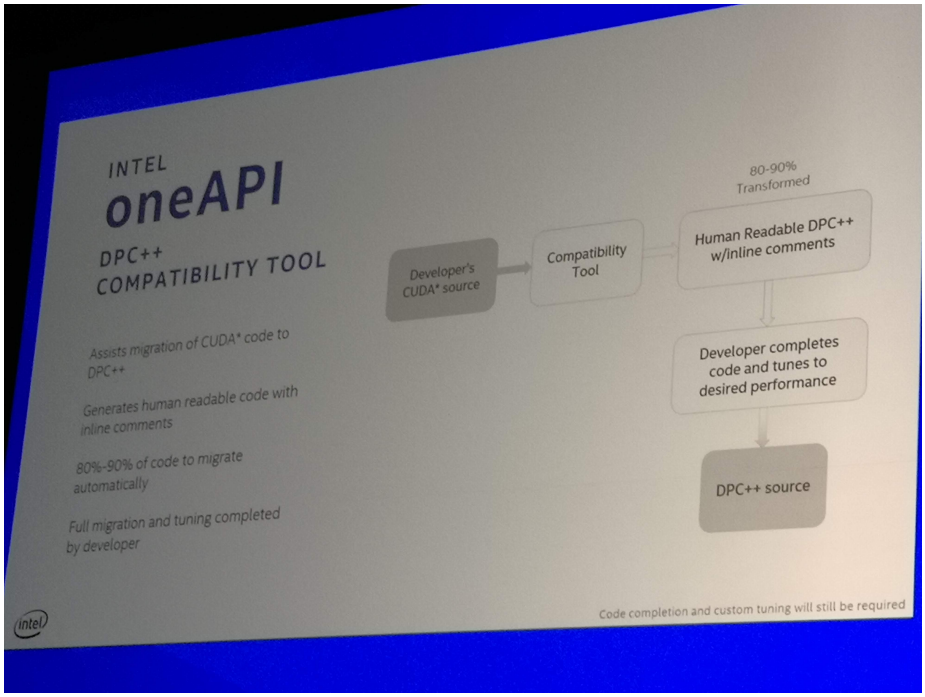

Figure 7: Presentation slide from the Intel oneAPI announcement.

A OneAPI[6] Beta Program is available today for developers. Programmers will use the Intel oneAPI DPC++ Compiler (data parallel C++) for direct programming of code targeting accelerators. CUDA code can be translated into DPC++ using a code migration tool. Support libraries will include Intel oneAPI Math Kernel Library, among others.

Programming examples in DPC++ were demonstrated by Intel, but the immediate impression was a fairly complex language syntax required by the generality of OneAPI.

The delegation asked GPU competitors NVIDIA and AMD whether they were considering adopting Intel's OneAPI interface. Unfortunately, the response was negative at this point in time.

8.3 Cloud bursting from local HPC clusters

One member of the delegation investigated the addition of HPC Cloud resources ("Cloud bursting") to local on-premise Linux clusters. Specifically, the Slurm

batch system developers are collaborating with Cloud providers, in particular Google Cloud Platform (GCP). We found out, for example, that the US national San Diego Supercomputer Center (SDSC) is already offering their users cloud bursting access to the Amazon cloud service.

Discussions about the detailed implementation of cloud bursting were made with the Slurm developers, Google, and Microsoft. Other cloud providers such as Amazon and Oracle also offer HPC cloud facilities.

Private clouds, like OpenStack seems to be adopted on more sites, to handle workloads that are not in need of super computer. This also provides a platform for running new technologies such as Kubernetes in a controlled environment.

8.4 EasyBuild & Spack

The BoF session **Getting scientific software installed** hosted both news from EasyBuild and Spack, which are the two most adopted build tools for HPC. A survey was conducted and from SC13 where the BoF was held the first time to SC19 a growth in adaptation of build tool among the BoF participants, are noticeable[20], EasyBuild being the number one and Spack the second.

8.5 EasyBuild

There was not a lot of news from EasyBuild which had released version 4 in September 2019. EasyBuild is ready for deprecation of Python 2, with full support of Python 3, and is now fully based on standard python libraries. There is a a new *SYSTEM* tool chain, which should be used instead of the deprecated dummy tool chain.

8.6 Spack

Spack has implemented a lock feature which makes it possible to keep the build information for a particular software stack, making it possible to rebuild for reproduceability purposes. Another feature coming is a multi-user build function. Spack is a single user tool which means a lot of wasted storage when multiple users compile and use the same software in their home directory. In collaboration with a Spack site, they have managed to extend Spack to better handle multi-users, which is a nice addition. Spack has been re-licensed the entire project from LGPL to Apache-2.0/MIT license.

9 Liquid Cooling

It is time to start planning for liquid cooling, and preparing the data centers and server rooms for liquid cooling, as both CPU and GPU are getting faster and warmer. Nvidia V100 uses up to 300 watts, and next generations could double the power consumption. Intels Cascade Lake AP demands liquid cooling at 350 watts, whereas "normal CPUs are in the range of 150 350 watts.[21]

When combining the hardware in 4, 8 or 16 pieces and sticking them all in the smallest numbers of rack Us possible, problems will arise. As total rack power goes above 30 kW to 35 kW pr. rack, liquid cooling must at least be considered.

In the past years we have seen different approaches and vendors for supplying the demand of liquid cooling.This seems to be narrowing down. The industry leaders like Apple, Dell EMC, HPE, Intel, AMD, NVIDIA, Supermicro, and others are now in a collaboration with CoolIT, who provides custom cooling solutions for CPU, GPU, Memory cold plates, and everything else from Rack manifolds to CDU. The collaboration is to keep cost low, but will hopefully introduce some standards for liquid cooling.

Lenovo is developing and building their own solution with cobber tubing and cold plates only.

CoolIT was showcasing their solutions for rack cooling on the exhibition floor:

- Rack DCLC CHx80: To place in the bottom of the rack. This is good for smaller installations offering 80 kW of cooling capacity.
- Rack DCLC CHx750: To place next to the racks, offering up to 750 kW of cooling capacity in a single unit with Redundancy. This is good for bigger installations and will come of a price similar to about 3 x the Rack DCLC CHx80.[22]

Motivair was showcasing their solutions for rack cooling on the exhibition floor. They also offered CDU solutions with up to 1.25 MW cooling capacity. One of their products was named ChilledDoor. This is an active rear door heat exchanger capable of removing server densities up to 75 kW per rack.[23]

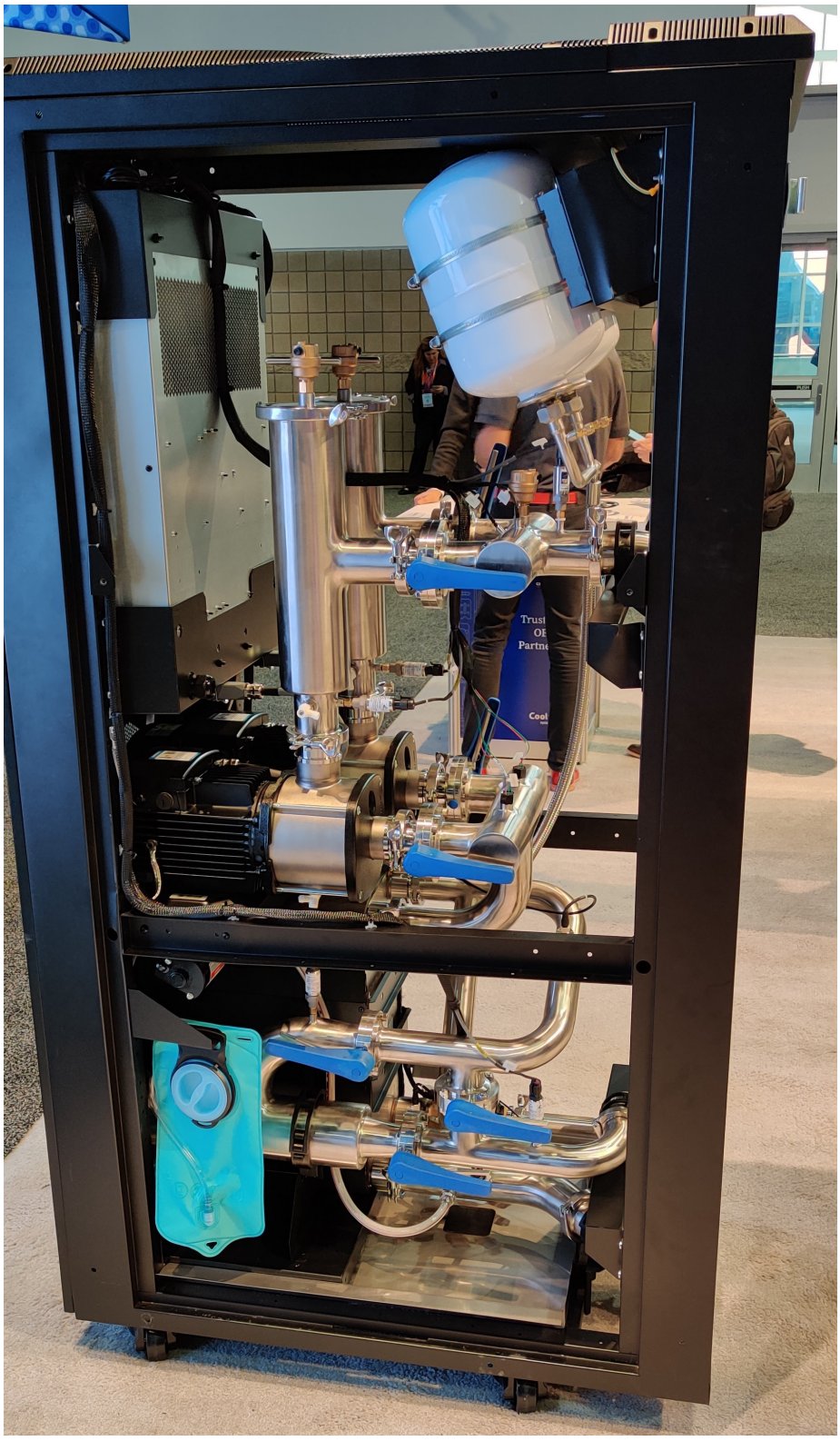

Figure 8: CoolIT's RACK DCLC CHx750 on the show floor at SC 19.

10 Artificial Intelligence

Last year at Dallas we saw the tip of the AI iceberg[12]. AI has been around for decades but big data, increased computing capabilities, and user demand has created an increased interest where AI applications are evolving at a rapid pace. AI is to solve business challenges, risk and compliance, sales/marketing, finance, accounting, and other areas where processes are inefficient, time-consuming, costly, and data-driven decision-making. This year we saw deep learning algorithms in neural networks to generate predictive models, and information like never before. There is a significant learning curve to be able to master the various directions of AI such as Natural Language Processing, Machine Learning, Data Science, and others. It was made clear that developing capabilities in AI technologies requires significant investment and commitment.

11 Top 500 Announcements

During SC19 the 54th edition of the TOP500 list was released. Compared to the previous list from June 2019 there were no changes in the top of the list, the first new system appears on position 24. This means that the two US located pre-exascale IBM-systems Summit and Sierra still holds place 1st and 2nd place. The 3rd and 4th systems are two Chinese installations. The Frontera system at Texas Advanced Computer Center comes in at a 5th place, and at 6th place appears the largest European supercomputer, the Swiss Piz Daint. Denmark has currently no systems in the Top500 list, Norway and Sweden have each 2 systems. Compared to US, China has the most systems on the list, 46% vs. 23% , but US has lead of aggregated performance, 37% vs. 32%. To enter the TOP500 list, a performance of at least 1.14 Pflops is now necessary.[3]

The Green500-list ranks the top 500 supercomputers by energy efficiency, and this list was published together with Top500. Normally, to appear on the Green500-list, some sort of accelerator (eg. GPU) to aid the CPU with dense computations is needed. But interestingly, no. 1 on the list, the Japanese Fugaku-prototypesystem, equipped with the A64FX manycore ARM-CPU, doesn't have any accelerator.[4]

12 Conclusion

The exascale race intensifies, and more vendors, both hardware and software is chiming in with their solution. There is till major challenges to solve within the next 1-2 year, before the first exascale system will be placed in to production by the end of 2021, if road maps are to be trusted and everything goes according to plan. One of the major issues that all hardware vendors, whether it is, CPU, GPU, or interconnect is cooling. Higher clock, cores count, throughput, and higher bandwidth uses more power, which demands more cooling, for the entire system. It will no longer be enough to cool the CPU and GPU, memory DIMMs and interconnect must be cooled as well.

System vendor, is challenged to make as dense system as possible, and it seems like the default high density solution with 4 servers in 2 rack U, will change, to even more dense designs.

There is an increased focus on accelerators, as it is impossible to achieve an exaflop with existing CPU technologies. This is why we see Intel entering the GPU domain. Nvidia is still dominant here. There will be a lot of competition in the accelerator domain the next couple of years, which will provide HPC installations with more different accelerators. Nvidia has an edge here as CUDA is the most utilized programming language for GPU and AI/ML/DL for a long period of time. It will be interesting to see if AMD's work with ROCm and HIP can change this.

From a software perspective challenges with regards to ensure software stacks and application produce correct results when computing at exascale.

Super Computing in 2020 will be in Atlanta, GA from the 15-20th of November.

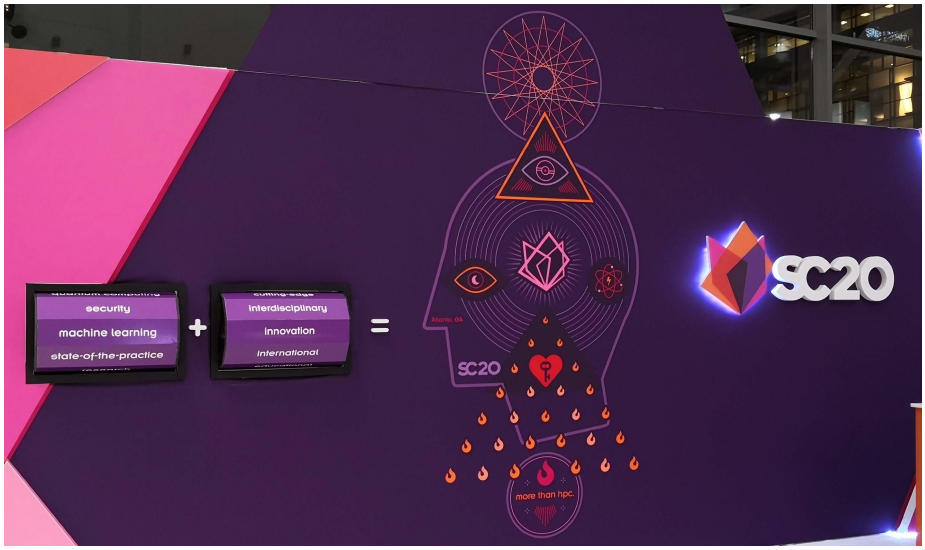

Figure 9: Super Computing 2020 - More than HPC.

Bibliography

[1] SLINGSHOT: THE INTERCONNECT FOR THE EXASCALE ERA https://www.cray.com/sites/default/files/Slingshot-The-Interconnect-for-the-Exascale-Era.pdf

[2] INSIDE HPES GEN-Z SWITCH FABRIC https://www.nextplatform.com/2019/09/09/inside-hpes-gen-z-switch-fabric/

[3] Top500 List - November 2019 https://www.top500.org/list/2019/11/

[4] Green500 List - November 2019 https://www.top500.org/green500/lists/2019/11/

[5] Intels 2021 Exascale Vision in Aurora: Two Sapphire Rapids CPUs with Six Ponte Vecchio GPUs https://www.anandtech.com/show/15120/intels-2021-exascale-vision-in-aurora-two-sapphire-rapids-cpus-with-six-ponte-vecchio-gpus

[6] Intel® oneAPI Toolkits https://software.intel.com/en-us/oneapi

[7] ConnectX® Ethernet Adapters https://www.mellanox.com/products/connectx-smartnic/

[8] HPE to acquire supercomputing leader CRAY https://www.hpe.com/us/en/newsroom/press-release/2019/05/hpe-to-acquire-supercomputing-leader-cray.html

[9] High-Performance Programmable SmartNICs https://www.mellanox.com/products/smartnic/

[10] Jensen Huang interview Nvidia can shake off rivals that have complicated and untested AI solutions https://venturebeat.com/2019/11/17/jensen-huang-interview-nvidia-can-shake-off-rivals-that-have-complicated-and-untested-ai-solutions/

[11] Kaushik, Brajesh Kumar. *Nanoscale Devices: Physics, Modeling, and Their Application*. CRC Press, 2018.

[12] DeIC HPC TekRef group report on the Supercomputing 2018 conference https://www.deic.dk/sites/default/files/uploads/SC18_report_DeIC_HPC_TekRef_0.pdf

[13] AMD Rome is it for real? Architecture and initial HPC performance https://www.dell.com/support/article/dk/da/dkbsdt1/sln319015/amd-rome-is-it-for-real-architecture-and-initial-hpc-performance?lang=en

[14] The Data Accelerator https://www.hpc.cam.ac.uk/research/data-acc

[15] High Performance Computing https://www.dell.com/support/article/dk/da/dkbsdt1/sln311501/high-performance-computing?lang=en

[16] STORAGE LEADER DDN ACQUIRES LUSTRE FILE SYSTEM CAPABILITY FROM INTEL https://www.ddn.com/press-releases/storage-leader-ddn-acquires-lustre-file-system-capability-intel/

[17] Lustre Roadmap http://lustre.org/roadmap/

[18] DAOS Storage Engine https://github.com/daos-stack/daos

[19] Memory-Driven Computing Explained https://www.hpe.com/us/en/newsroom/blog-post/2017/05/memory-driven-computing-explained.html

[20] Getting Scientific Software Installed https://github.com/easybuilders/easybuild/wiki/SC19-BoF-session-Getting-Scientific-Software-Installed

[21] Intel Cascade Lake Xeon Platinum 8280, 8268, and Gold 6230 Review: Taking The Fight to EPYC https://www.tomshardware.com/reviews/intel-cascade-lake-xeon-platinum-8280-8268-gold-6230-amd-epyc,6058-6.html

[22] MODULAR, RACK-BASED LIQUID COOLING https://www.coolitsystems.com/rack-dlc/

[23] ChilledDoor https://www.motivaircorp.com/products/chilleddoor/