Jülich Supercomputing Centre starts deployment of a Booster for JURECA

Since its installation in autumn 2015, the JURECA ("Jülich Research on Exascale Cluster Architectures") system at the Jülich Supercomputing Centre (JSC) has been available as a versatile scientific tool for a broad user community. Now, two years after the production start, an upgrade of the system in autumn 2017 will extend JURECA's reach to new use cases and enable performance and efficiency improvements of current ones. This new "Booster" extension module, utilizing energy-efficient many-core processors, will augment the existing "Cluster" component, based on multi-core processor technology, turning JURECA into the first "Cluster-Booster" production system of its kind.

The "Cluster-Booster" architecture was pioneered and successfully implemented at prototype-level in the EU-funded DEEP and DEEP-ER projects [1], in which JSC has been actively engaged since 2011. It enables users to dynamically utilize capacity and capability computing architectures in one application and optimally leverage the individual strengths of these designs for the execution of sub-portions of, even tightly coupled, workloads. Lowly-scalable application logic can be executed on the Cluster module whereas highly-scalable floating-point intense portions can utilize the Booster module for improved performance and higher energy efficiency.

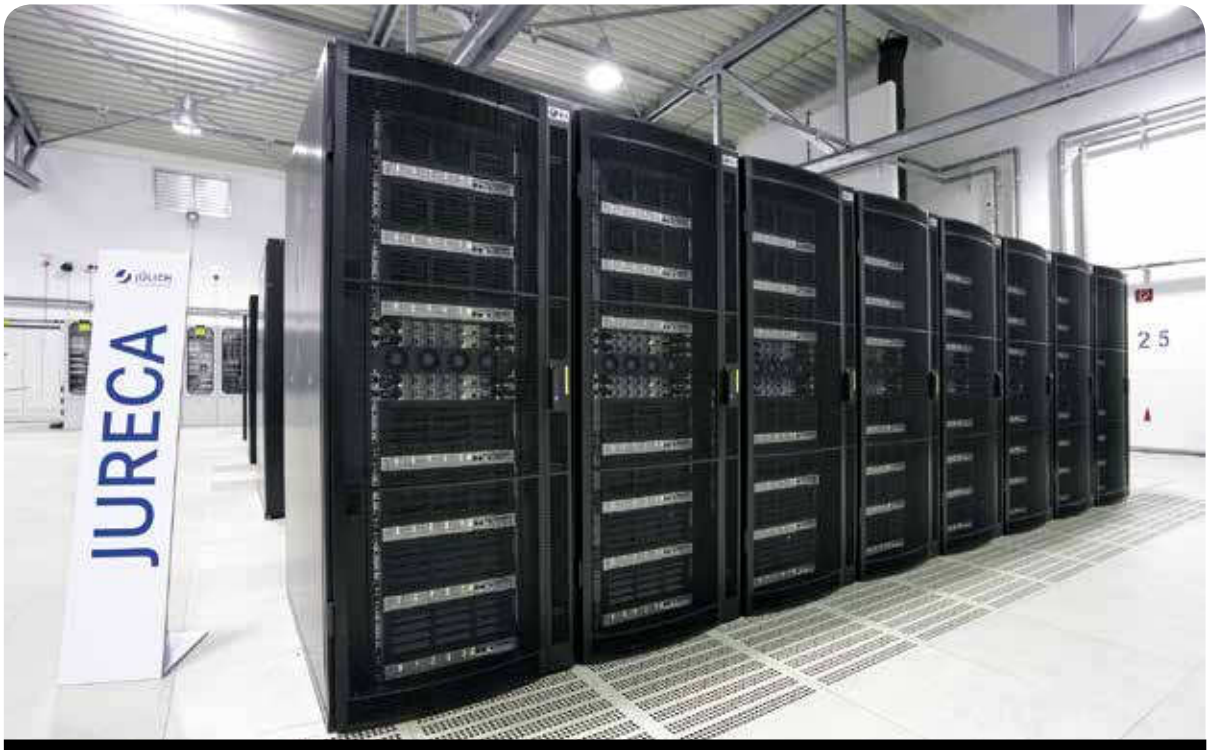

Fig. 1: The JURECA Cluster module at Jülich Supercomputing Centre.

The JURECA system currently consists of an 1,872-node compute cluster based on Intel "Haswell" E5-2680 v3 processors, NVidia K80 GPU accelerators and a Mellanox 100 Gb/s InfiniBand EDR (Extended Data Rate) interconnect [2]. The system was delivered by the company T-Platforms in 2015 and provides a peak performance of 2.2 PFlops/s. The new Booster module will add 1,640 more compute nodes to JURECA and increase the peak performance by five PFlops/s. Each compute node is equipped with a 68-core Intel Xeon Phi "Knights Landing" 7250-F processor and offers 96 GiB DDR4 main memory connected via six memory lanes and additional 16 GiB of high-bandwidth MCDRAM memory. As indicated by the "-F" suffix, the utilized processor model has an on-package Intel Omni-Path Architecture (OPA) interface which connects the node to the 100 Gb/s OPA network organized in a three-level full-fat tree topology. The Booster, just as the Cluster module, will connect to JSC's central IBM Spectrum Scale-based JUST ("Jülich Storage") cluster. The storage connection, realized through 26 OPA-Ethernet router nodes, is designed to deliver an I/O bandwidth of up to 200 GB/s. In addition, 198 bridge nodes are deployed as part of the Booster installation. Each bridge node features one 100 Gb/s InfiniBand EDR HCA and one 100 Gb/s OPA HFI, in order to enable a tight coupling of the two modules' high-speed networks. The Booster is installed in 33 racks directly adjacent to the JURECA cluster module in JSC's main machine hall. JSC and Intel Corporation co-designed the system for highest energy efficiency and application scalability. Intel delivers the system with its partner Dell, utilizing Dell's C6320 server design (see Figure 2). The group of partners is joined by the software vendor ParTec, whose ParaStation software is one of the core enablers of the Cluster-Booster architecture. The Cluster and Booster module of JURECA will

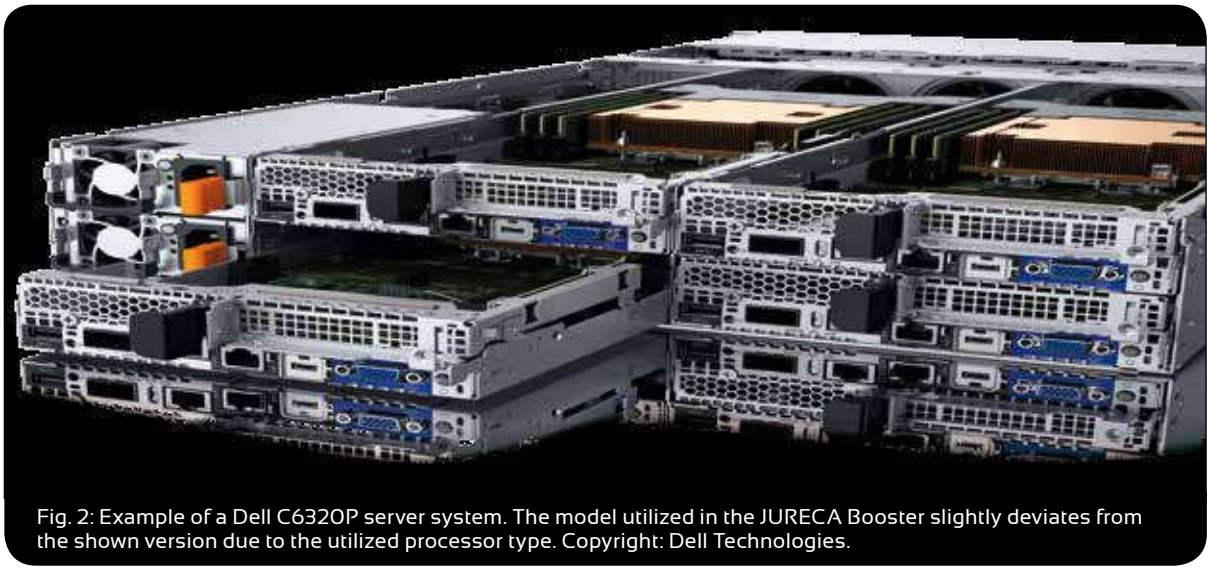

Fig. 2: Example of a Dell C6320P server system. The model utilized in the JURECA Booster slightly deviates from the shown version due to the utilized processor type. Copyright: Dell Technologies.

Fig. 3: The JURECA Booster module at the Jülich Supercomputing Centre. The Cluster module is visible at the left border of the photograph.

be operated as a single system with a homogeneous global software stack.

As part of the deployment, the partners engage in a cooperative research effort to develop the necessary high-speed bridging technologies that enables high-bandwidth, low-latency MPI communication between Cluster and Booster compute nodes through the bridge nodes. The development will be steered by a number of real-world use cases, such as earth systems modeling and in-situ visualization.

The compute time on the Booster system will be made available primarily to scientists at Forschungszentrum Jülich and RWTH Aachen University. During a two-year interim period, all admissible researchers at German universities can request computing time by answering the calls of the John von Neumann Institute for Computing (NIC) until the second phase of the JUQUEEN successor system has been fully deployed.

The realization of the Cluster-Booster architecture in the JURECA system marks a significant evolution of JSC's dual architecture strategy as it brings "general purpose" and highly-scalable computing resources closer together. With the replacement of the JUQUEEN system in 2018, JSC intends to take the next step in its architecture roadmap and, in phases, deploy a Tier-0/1 "Modular Supercomputer" that tightly integrates multiple, partially specialized, modules under a global homogeneous software layer.

References

[1]: **DEEP and DEEP-ER projects:**
http://www.deep-projects.eu

[2] **Jülich Supercomputing Centre, JURECA:**
General-purpose supercomputer at Jülich Supercomputing Centre. Journal of large-scale research facilities, Volume 2, A62.
http://dx.doi.org/10.17815/jlsrf-2-121, 2016

Written by Dorian Krause
Jülich Supercomputing Centre (JSC)
Contact: d.krause@fz-juelich.de