

JURECA Booster Module

Introduction

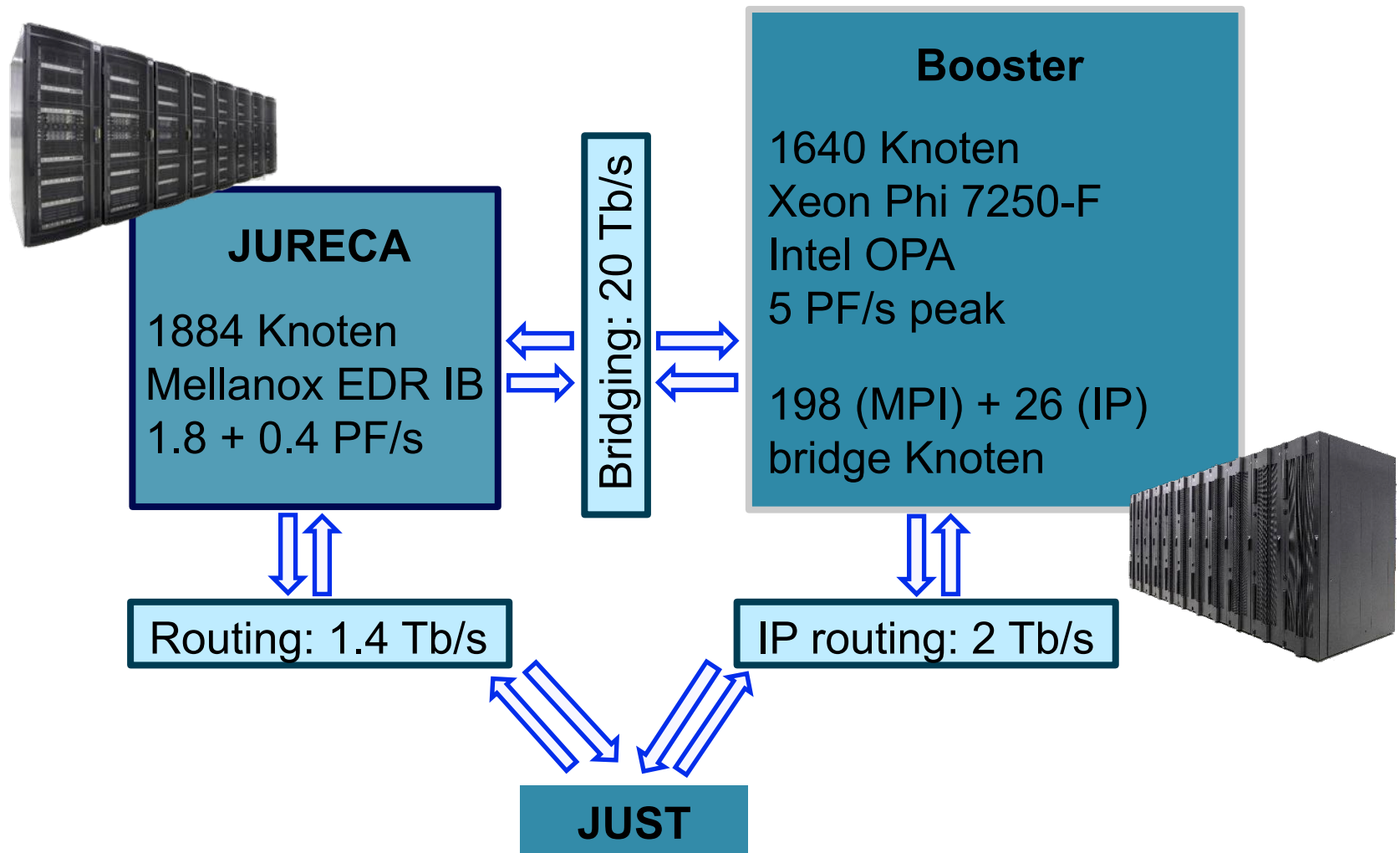
2018-05-28 | Philipp Thörnig | HPS Division @ JSC

JURECA Booster



- Extension of the JURECA (**J**ülich **R**esearch on **E**xascale **C**luster **A**rchitecture)
 - Augments Cluster module with a highly-scalable component
- Designed for capability workloads
 - System integrators: Intel with Dell
- Compute time allocation
 - Primarily for scientists from Jülich and Aachen
 - Available for admissible researchers at German universities for a two-year interim period via NIC
- First implementation of a **Modular Supercomputer** at Petascale

JURECA Cluster-Booster Architecture



198 bridge nodes

- Dell PowerEdge R430
- 1× Xeon E5-2690v4
- 1× Intel OPA HFI
 - 1× Mellanox ConnectX-4 HCA

Bridging for ParaStation MPI
based on ParTec's ParaStation
psgwd prototype

- Development as part of Intel,
ParTec, JSC cooperation



JURECA Booster Computes

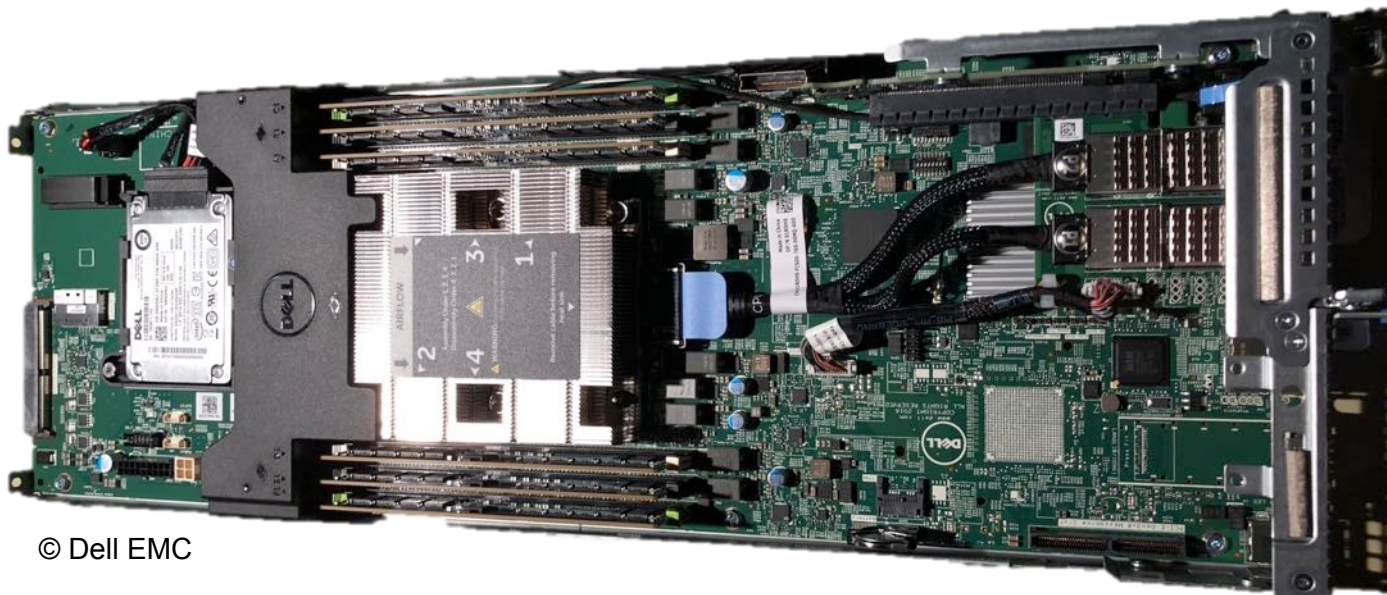
Dell PowerEdge C6320P specifications

Dell PowerEdge C6320P solution

- Intel Xeon Phi “Knights Landing” 7250-F
68 cores @ 1.4 GHz
- 96 GiB main memory, 16 GiB MCDRAM
- On-package Intel Omni-Path Architecture network interface



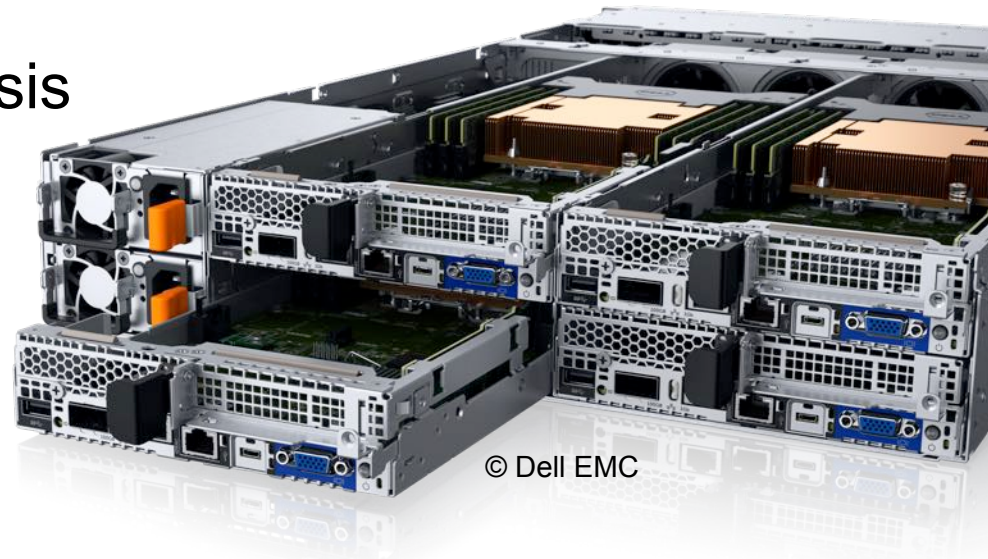
© Dell EMC



© Dell EMC

JURECABooster Computes

- 1 Chassis houses 4 KNL Blades
- 1 Rack houses 18 Chassis
- 23 Racks equipped with KNL-Chassis
 - 1640 KNL Systems
 - 157 TiB main memory + 26 TiB MCDRAM
 - Peak performance: 5 PF/s



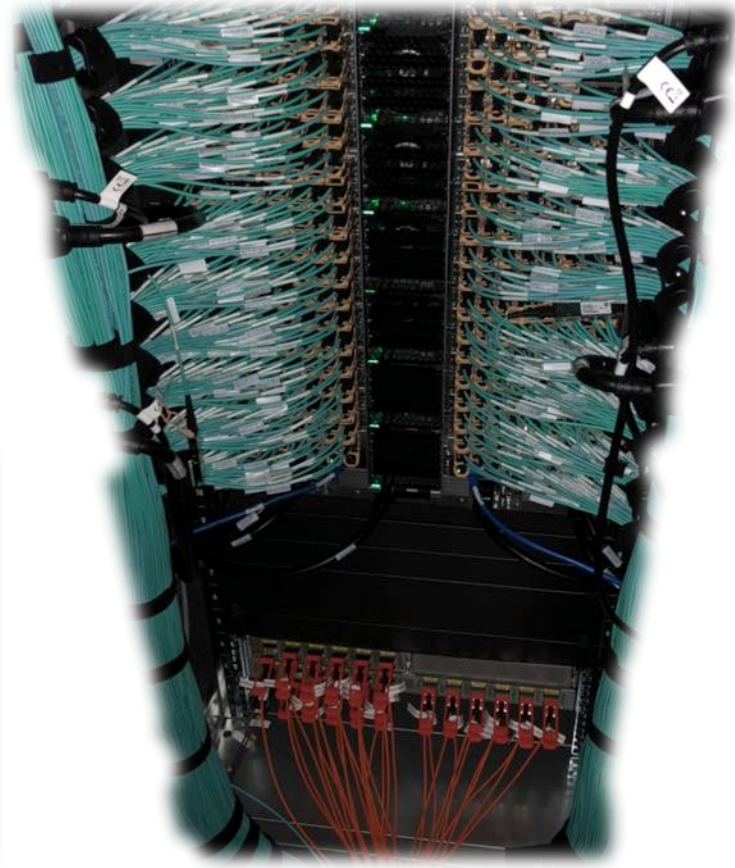
© Dell EMC



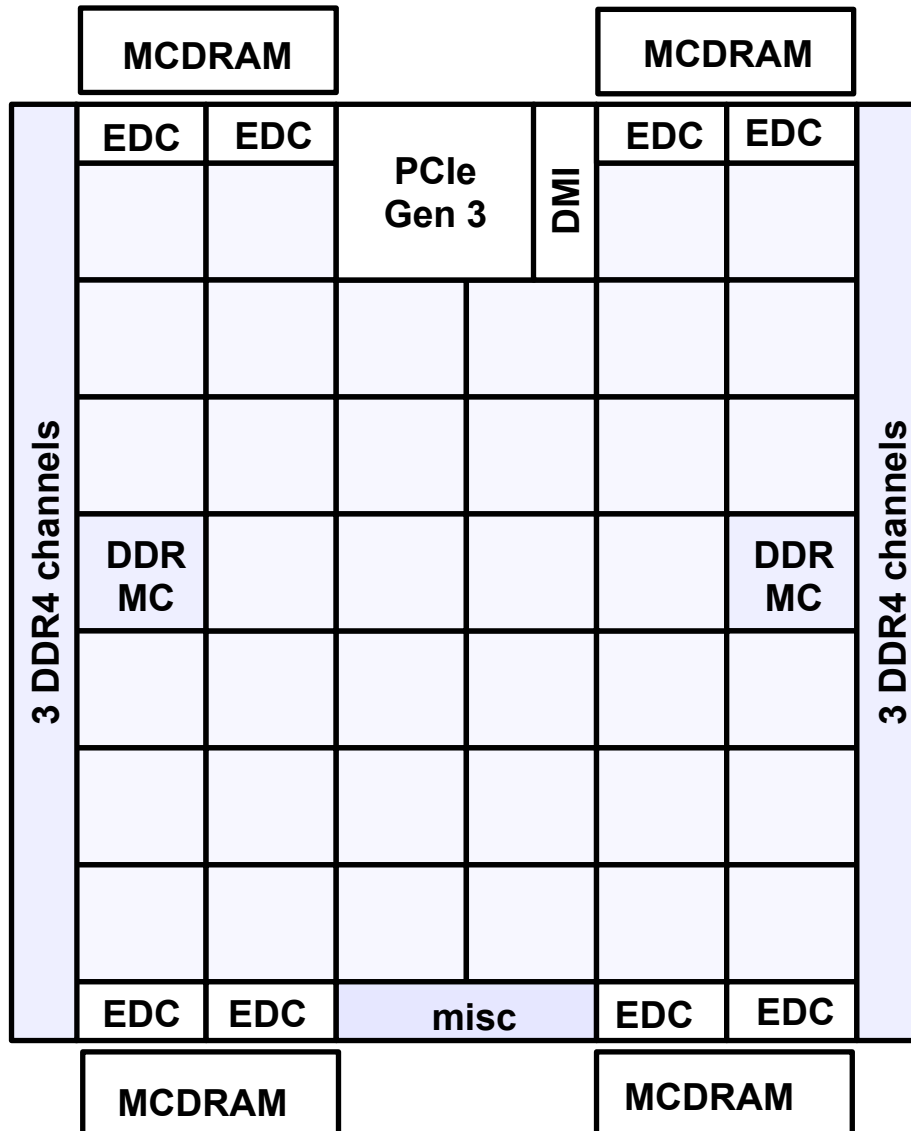
© Dell EMC

JURECA Booster Interconnect

- Intel Omni-Path Architecture network
 - 100 Gbps per link and direction
 - Full fat-tree topology
- Design for 200 GBps storage bandwidth



Intel Knights Landing Architecture



- 36 tiles, 2-dim mesh
- tile = 2 cores +
2 VPU/core +
1 MB L2
- 4 threads per core
- AVX-512 ISA extension
- 16 GiB MCDRAM
 - High bandwidth
 - O(500) GB/s
- 6 DDR4 channels
 - O(100) GB/s

- Same login nodes and file systems as Cluster module
`ssh <user>@jureca.fz-juelich.de`
`ssh <user>@jureca[01-12].fz-juelich.de`
- Same file systems (`$HOME`, `$WORK`)
- Same system software environment as JURECA
 - CentOS 7.X
 - GNU, Intel Compiler
 - ParaStation MPI, Intel MPI
- One workload management system: Slurm/ParaStation
 - Separate partitions for Booster nodes
 - Similar to handling of e.g., GPU-equipped nodes

- The Booster has essentially the same software environment than JURECA
- Key differences:
 - Less software, due to its more specialized nature
 - Slightly different ISA (AVX-512): incompatible with Haswell nodes in most cases
 - Interactive sessions need extra care:
 - `srun --pty --cpu_bind=none -mpi=none /bin/bash {-l|-i}`
 - Test and compile partition in Slurm is **develbooster**
- To browse the Booster SW from the login nodes:
 - **m1 Architecture/KNL**

- **Option 1:** Compilation on KNL nodes
 - Get an interactive session on a Booster node, with **-l** or **-i**
 - Will load the Booster SW environment
 - Set your flags correctly to enable AVX-512 (Intel: **-xHost**, GNU: **-march=native**)

- **Option 2:** Cross-compilation
 - Load the **Architecture/KNL** module
 - Set your flags correctly to enable AVX-512 (Intel: **-xMIC-AVX512** GNU: **-march=knl -mtune=knl**)
 - Can fail if the build process requires to execute binaries compiled with these flags

- Job submission
 - Use booster partitions:
booster, develbooster, modetestbooster
 - Make sure you have the right environment for the job:
 - either when submitting the job (**Architecture/KNL** loaded)
 - or inside the job (load **Architecture/KNL** in the script submitted)

- General access for admissible users possible since End of November 2017
 - Support for different KNL NUMA (Quadrant, SNC-2) and MCDRAM (Flat, Cache, Hybrid50) modes available at **modetestbooster** partition
 - Information about partitions published online at “**JSC⇒Expertise⇒Supercomputers⇒JURECA⇒UserInfo⇒QuickIntroduction**”
- Initially weekly maintenances published at **motd**
- Support for heterogeneous Haswell and KNL jobs
 - Available at prototype-level
 - Targeting Mid 2018 for release

- JURECA (incl. Booster) **motd**: message of the day
 - Information about preventive and emergency maintenances
 - During Booster Module offline maintenance you are able to access your files as usual via Login Systems
 - Information about system configuration changes
- On-line documentation
 - <http://www.fz-juelich.de/ias/jsc/jureca>
- User support at FZJ
 - sc@fz-juelich.de
 - Phone: 02461 61-2828