# JURECA Booster

## Introduction

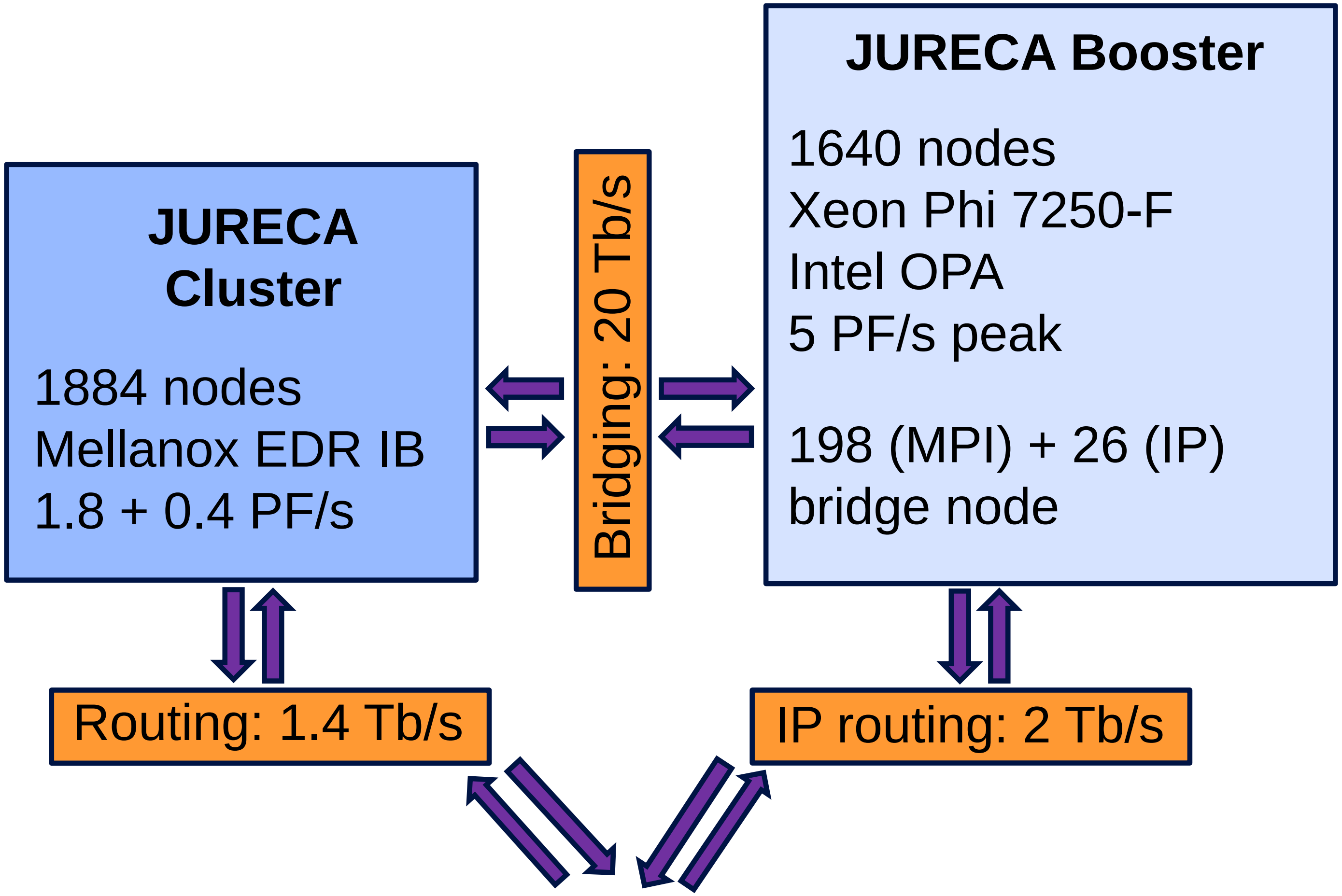
2017-11-24 | Dorian Krause

# JURECA Booster

# JURECA Booster Hardware

- Extension of the JURECA (**Jü**lich **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



**JURECA Cluster**

1884 nodes
Mellanox EDR IB
1.8 + 0.4 PF/s

Bridging: 20 Tb/s

**JURECA Booster**

1640 nodes
Xeon Phi 7250-F
Intel OPA
5 PF/s peak

198 (MPI) + 26 (IP)
bridge node

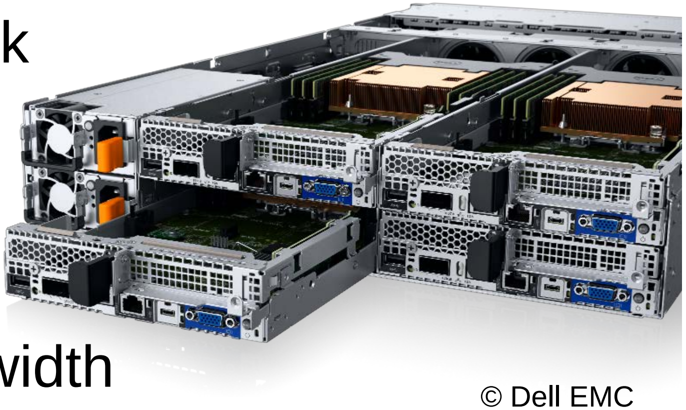
Routing: 1.4 Tb/s

IP routing: 2 Tb/s

**JUST: `$WORK` and `$HOME`**

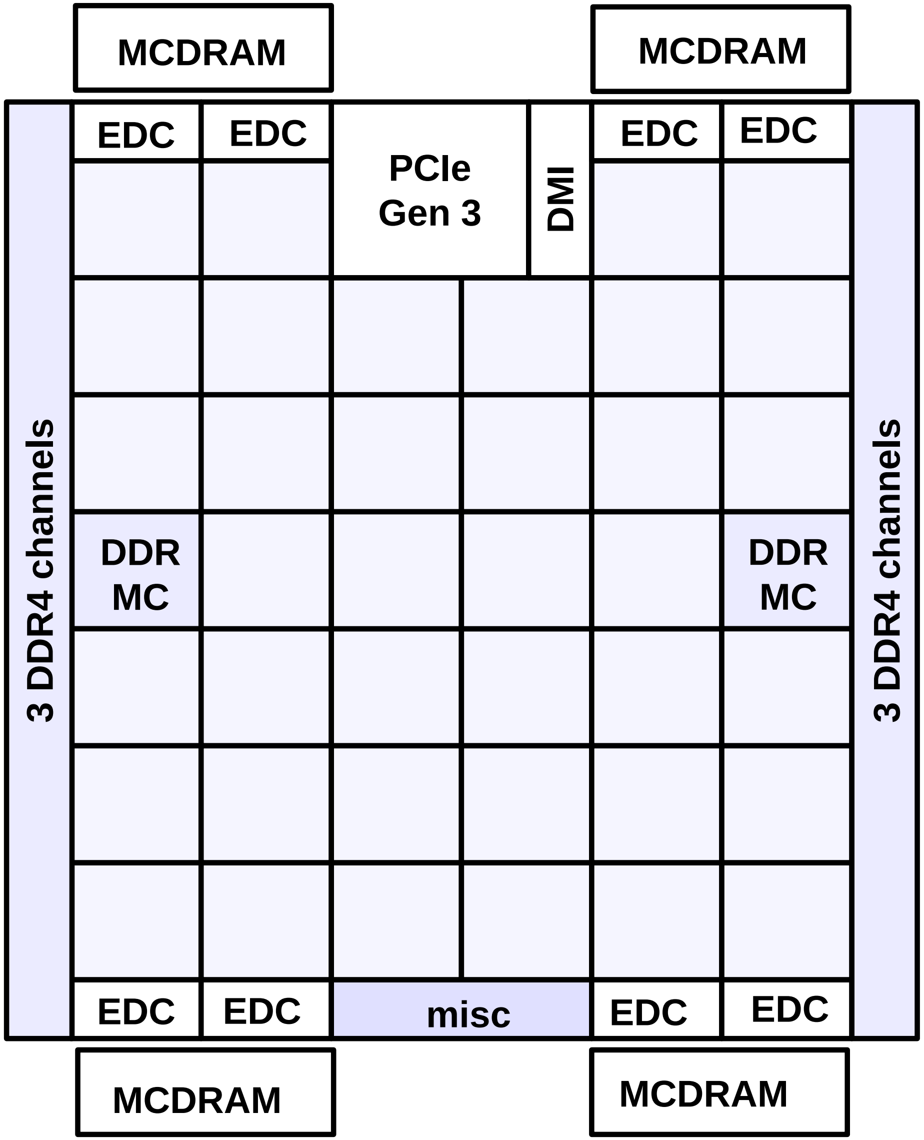
# JURECA Booster Hardware

- 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
- 157 TiB main memory + 26 TiB MCDRAM
- Peak performance: 5 PF/s
- Intel Omni-Path Architecture network
  - 100 Gbps per link and direction
  - Full fat-tree topology
- Design for 200 GBps storage bandwidth



© Dell EMC

# Intel Knights Landing Architecture


MCDRAM
MCDRAM
EDC
EDC
PCIe
Gen 3
DMI
EDC
EDC
3 DDR4 channels
DDR
MC
DDR
MC
3 DDR4 channels
EDC
EDC
misc
EDC
EDC
MCDRAM
MCDRAM

- 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

# JURECA Booster 101

- 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

# Application Software Environment (1/3)

- 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 will be provided
- To browse the Booster SW from the login nodes:
  - `ml Architecture/KNL`

# Application Software Environment (2/3)

- **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

# Application Software Environment (3/3)

- Job submission
  - Use booster partition (names will be published soon on the web page
  - 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)

# Booster Timeline

- General access for admissible users planned for End of November
    - Information about partitions will be published online
    - Initial support for different KNL NUMA (Quadrant, SNC-2) and MCDRAM (Flat, Cache, Hybrid50) modes planned for testing
- Initially weekly maintenances
- Support for heterogeneous Haswell and KNL jobs
    - Available at prototype-level
    - Targeting spring 2018 for release

# Further Information

- JURECA (incl. Booster) **`motd`**: message of the day
  - Information about preventive and emergency maintenances
  - Information bout 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