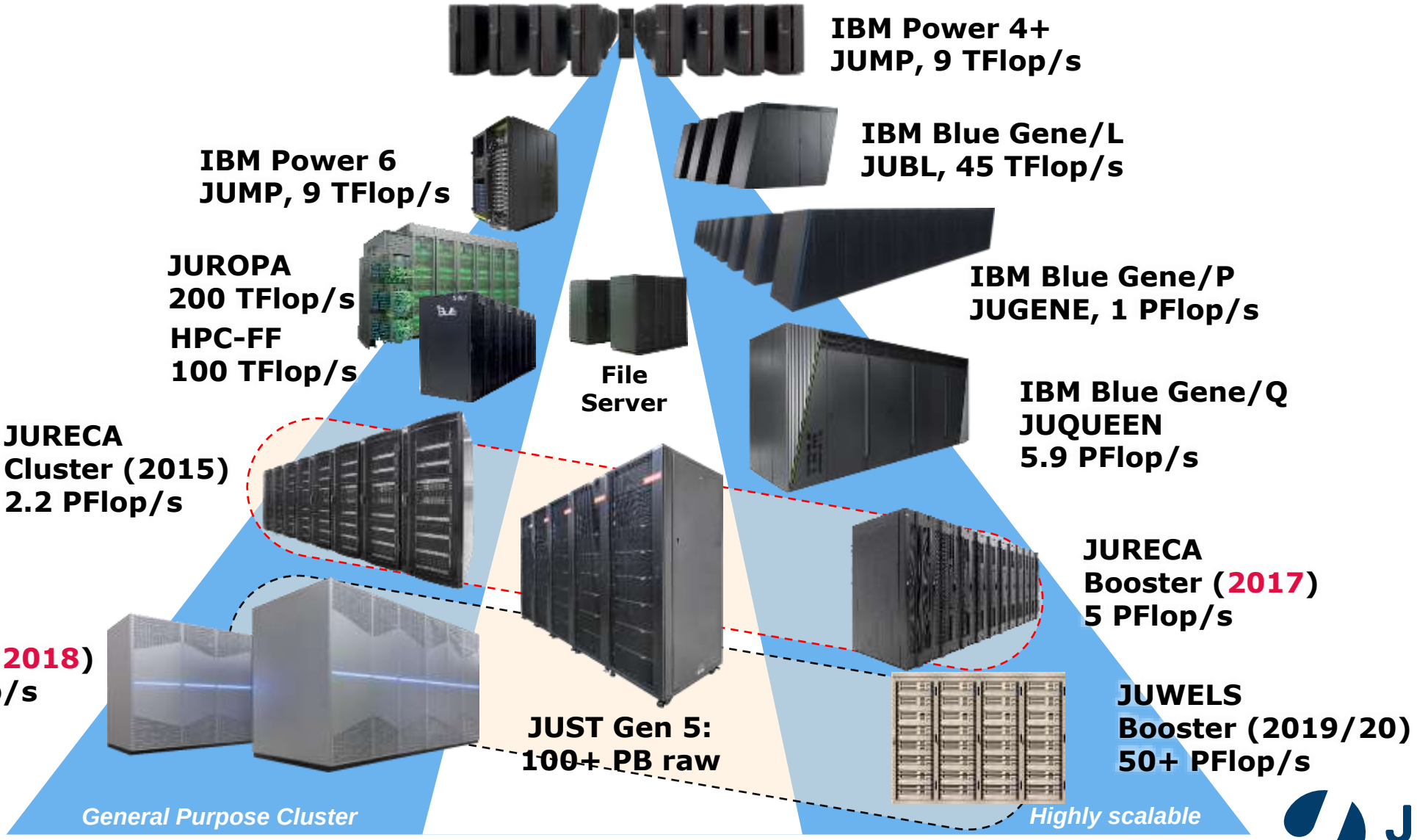
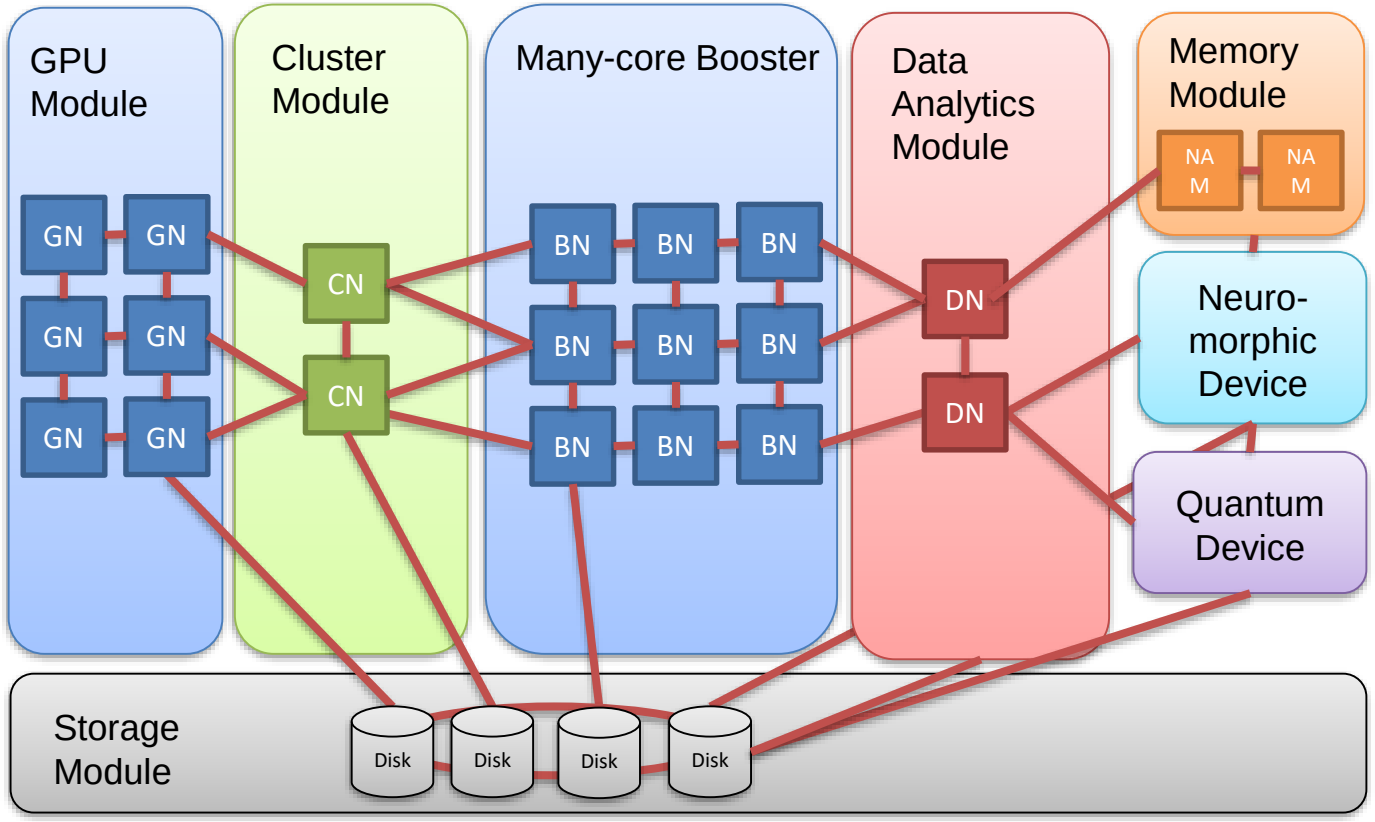
UPDATE ON JÜLICH SUPERCOMPUTING CENTRE (FZJ/JSC)

OCTOBER 1, 2018 I THOMAS LIPPERT, NORBERT ATTIG

DUAL HARDWARE STRATEGY AT JSC



… AND EVOLUTION TO A MODULAR SUPERCOMPUTING ARCHITECTURE



JURECA BOOSTER MODULE

JURECA BOOSTER MODULE: TECHNICAL DATA

Name	JURECA booster
Background	The JURECA booster augments the JURECA cluster module with a highly-scalable component. It is designed for capability workloads. JURECA serves as a pilot system for the modular concept developed at Jülich. It is funded by Helmholtz, not by GCS. The concept was proven in the EU-funded technology projects DEEP, DEEP-ER and DEEP-EST.
Vendor	Intel, Dell (system integrator)
Architecture	Intel Xeon Phi “Knights Landing” (KNL)
Node Architecture	One Intel Xeon Phi 7250-F Knights Landing CPU per node with 68 cores @ 1.4 GHz each, Intel Hyperthreading Technology (Simultaneous Multithreading), AVX-512 ISA extension, 96 GiB memory plus 16 GiB MCDRAM high-bandwidth memory
Node Quantities	1.640 nodes + 198 (MPI) and 26 (IP) bridge nodes
Network	Intel Omni-Path Architecture high-speed network with non-blocking full fat tree topology
Cooling	Passive water cooling
Performance	5 PF/s peak
Power consumption	~0.5 MW

JURECA BOOSTER ACHIEVEMENTS AND CHALLENGES

- General user availability since Nov 2017
- Successful cluster + booster Linpack for Nov 2017 Top500 demonstrated capabilities of system software at prototype level
  - 3.78 PF/s across cluster and booster: InfiniBand + OPA / Haswell + KNL mixed
- Initial challenge: I/O stability and performance
  - Storage design based on IP-over-OPA and Ethernet-router: Instable and underperforming under high load
  - Significant development effort on OPA software on router nodes by Intel
  - Current situation: stable 100+ GB/s I/O performance, per-node performance limited by KNL serial core performance
- User base initially limited to few (power) users, but: Broadening since Spring 2018
- Intel shift away from Xeon Phi will clearly impact code adaptation and optimization efforts in communities
- On-going effort: System software maturation for cluster + booster support (MPI, Workload Manager, UI improvements) and application/workload porting

JUWELS: CLUSTER MODULE

JUWELS CLUSTER MODULE: TECHNICAL DATA

Name	JUWELS cluster
Background	JUQUEEN will be succeeded by a modular system JUWELS. The first component of JUWELS is a cluster. A booster module will be added in 2019/2020. Designed for optimal total cost of ownership (TCO) for a representative set of application benchmarks. The cluster consists of a GCS partition and a HGF partition for Earth system sciences.
Vendor	Atos (Bull)
Architecture	Sequana
Node Architecture	Dual-socket Skylake Platinum 8168 (24C, 2.7 GHz), 96 GB (88.7%) and 192 GB (9.4%), 4 Volta GPUs in 1.9% of nodes
Node Quantities	2.559 nodes, 10 Sequana cells
Network	EDR InfiniBand, fat-tree interconnect
Cooling	Direct hot-water cooling
Performance	12 PF/s peak (10.6 PF/s without GPUs).
Power consumption	~1.6 MW

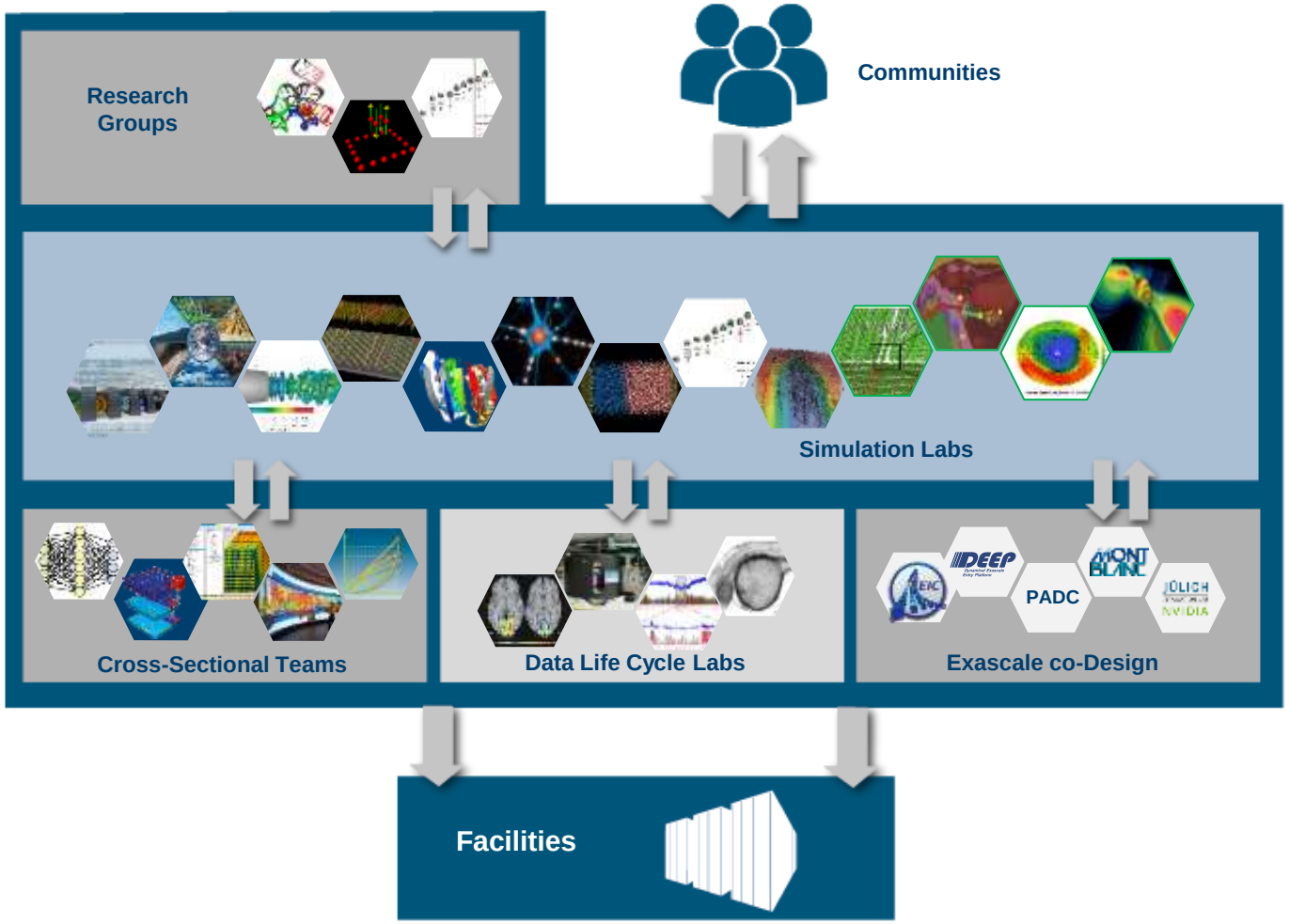
JUWELS CLUSTER ACHIEVEMENTS AND CHALLENGES

- General user availability since Jul 11, 2018
  - No overlap with JUQUEEN operations due to facility cooling limits
- Official inauguration on Sep 18, 2018
  in the presence of Federal Minister Anja Karliczek and NRW Prime Minister Armin Laschet
- Entered #23 in Jun 2018 Top500: Fastest system in Germany at the time, #6 in Europe
- Largest Sequana installation outside France
  - Learning experience for all project partners: Handling of highly specialized and optimized Sequana components is non-trivial
  - Additional novelty: Entirely custom software stack designed for modularity (central piece: ParaStation)
- Successfully in operation with 34 °C inlet temperature since July
  - Currently utilizing central cold water supply, free cooling equipment expected to be operational in 2019
- Related activity: Replacement of storage infrastructure with newer Lenovo DSS-G system (JUST5) incl. facility Ethernet fabric update to 100 Gb/s Ethernet in spring
  - Successful online migration

FURTHER ACTIVITIES

- Extension of data storage infrastructure for further use cases through new storage tiers
- Drafting of technical requirements for the JUWELS booster has started: Targeting installation in late 2019 or beginning 2020
- GCS application support with respect to national projects is becoming sync'ed, strengthened and streamlined between the GCS centres
- The GCS High-Level Support Team (HLST) of PRACE is in operation since spring 2018; close collaboration between the three teams of the HLST at the three GCS centres and coordination with the other PRACE HLSTs. Focus is on
  - PRACE Tier-0 production projects on the GCS systems
  - Code refactoring for potential future applications in coordination with the scientific communities.
  - Supporting community codes important to PRACE (e.g. OpenFOAM, GROMACS, …) → HLST coordinated action is highly desirable.
  - Enabling of applications from proposals with lacking HPC readiness
  - Setup of support collaborations in case of community code support between HLSTs and support structures at PRACE general partner sites.

SUPPORT AND RESEARCH LANDSCAPE AT JSC



SIMLAB STRUCTURE