PARTEC CLUSTER COMPETENCE CENTER

MANAGED HPC

ParaStation MPI

MPICH BoF · SC16 · Salt Lake City
Nov 15, 2016

Norbert Eicker
Technical Lead – ParaStation Consortium

ParaStation Cluster Suite

- ParaStation ClusterTools
  - *Provisioning and Management*
- ParaStation HealthChecker & TicketSuite
  - *Automated error-detection & handling*
  - *Ensuring integrity of the computing environment*
  - *Error prediction*
  - *Keeping track of issues*
  - *Powerful analysis tools*

Maximize job throughput
Minimize administration

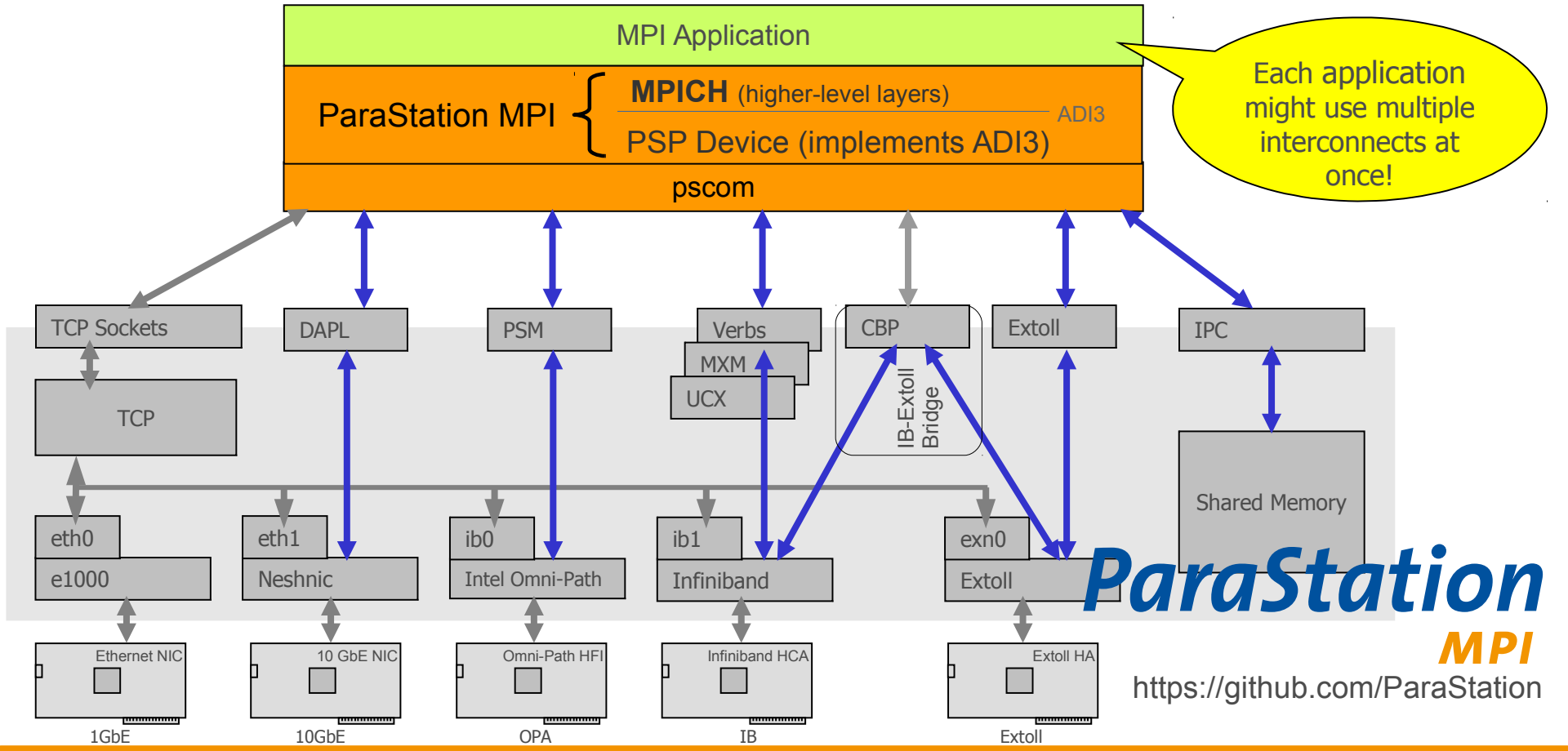
- ParaStation MPI & Process Management
  - *Runtime environment specifically tuned to the largest distributed memory supercomputers*
  - *Scalable & mature software setup*
  - *Keep control over processes*
  - *Full batch system integration (Torque, SLURM)*

ParaStation V5

∑ ParaStation Cluster Suite

ParaStation MPI

- Based on mpich-3.1.4
- Powered by “pscom”: efficient, low-level communication library
- Supporting a wide range of interconnects – even in parallel

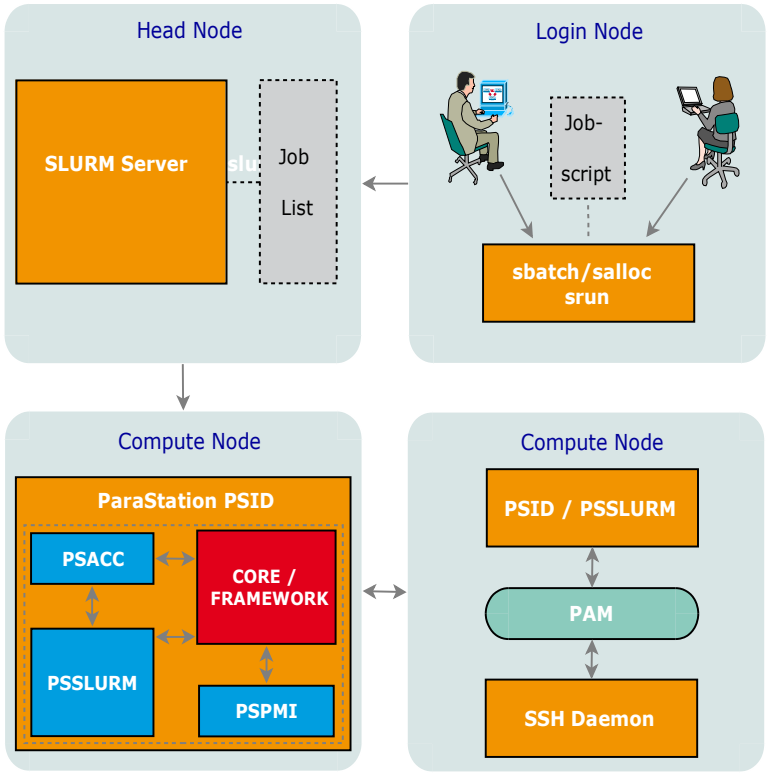


ParaStation MPI

https://github.com/ParaStation

ParaStation MPI

- Proven to scale to 3,000+ nodes and 86,000+ processes per job
- Network of MPI management daemons on computational nodes:
  - *Process startup and control, I/O forwarding, …*
  - *Precise resource monitoring*
- PSSLURM and PSMOM: Full integration for SLURM & TORQUE
- Preliminary CUDA support
  - *CUDA awareness*



- ParaStation MPI is ideal to develop and implement new Exascale concepts!

ParaStation MPI

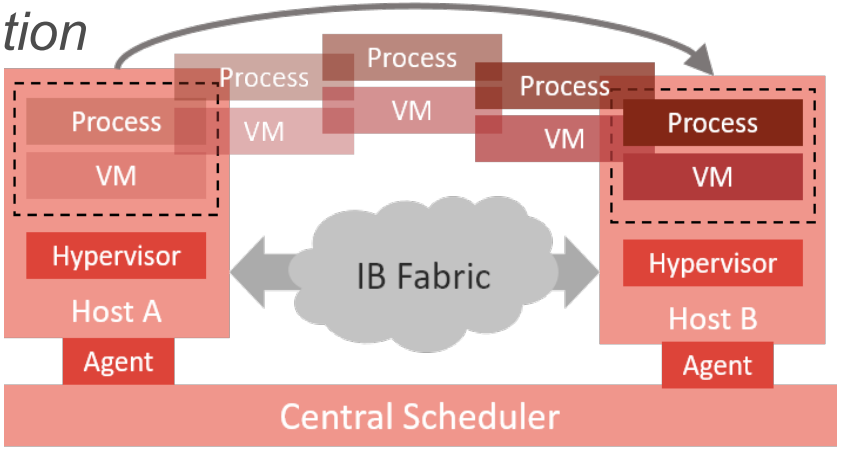
https://github.com/ParaStation

ParaStation MPI in FAST FOR EXASCALE

http://www.en.fast-project.de/

FaST Project:

- *Find a Suitable Topology for Exascale*
- *Joint Research Project funded by German Federal Ministry of Education and Research (BMBF)*
- *Goal: MPI process migration for truely dynamic co-scheduling*
- *Support application-transparent VM migration in ParaStation MPI:*
  - ✔ *Shutdown/Reconnect support for pscom library*
  - ✔ *MQTT-based interface to process migration framework*
  - ✔ *Working prototype for InfiniBand with little to no runtime overhead*

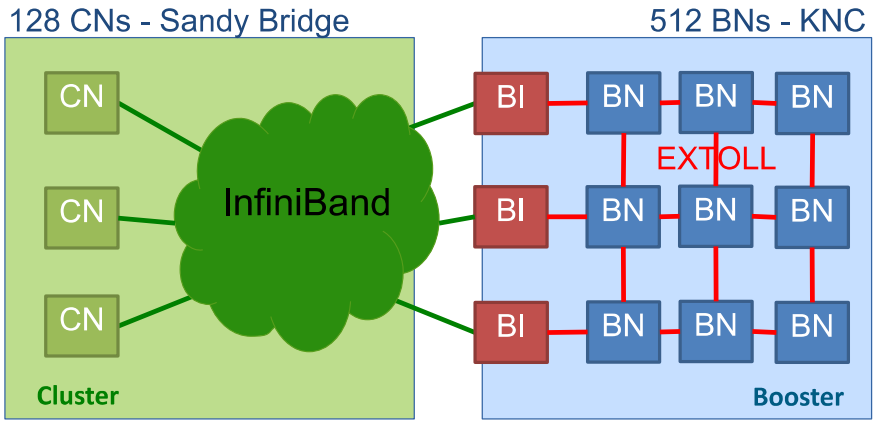


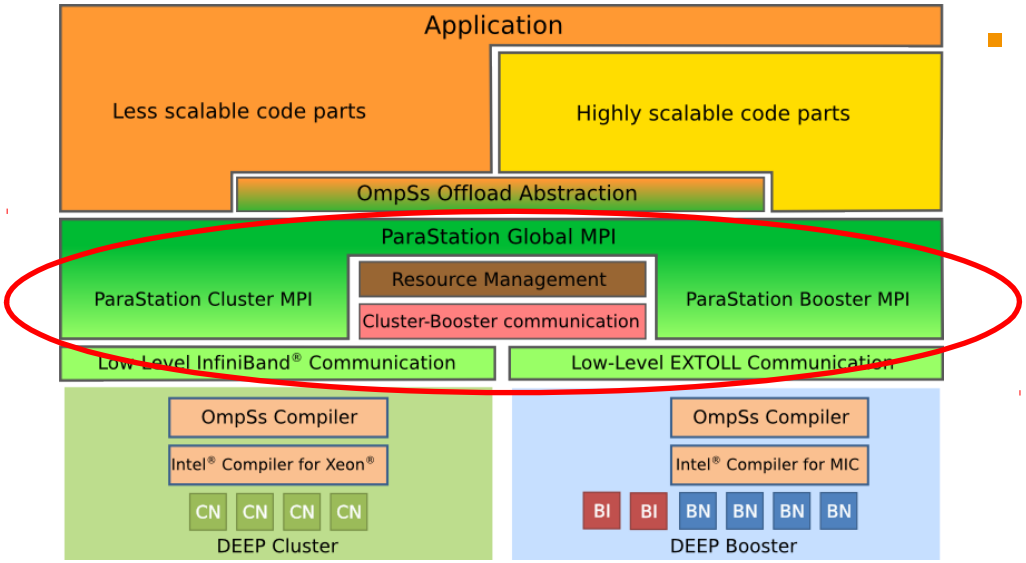
ParaStation MPI

https://github.com/ParaStation

ParaStation Global MPI in DEEP-ER

- Heterogeneous system: Cluster-Booster architecture
  - *Cluster Nodes (CN) with Intel Xeon multi-core CPUs*
  - *Booster Nodes (BN) with Intel MIC many-core CPUs*
  - *Booster Interfaces (BI) connecting Cluster and Booster*

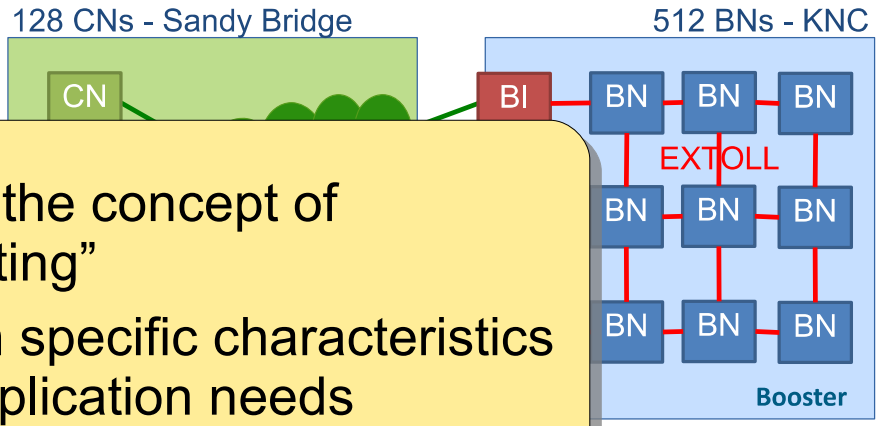




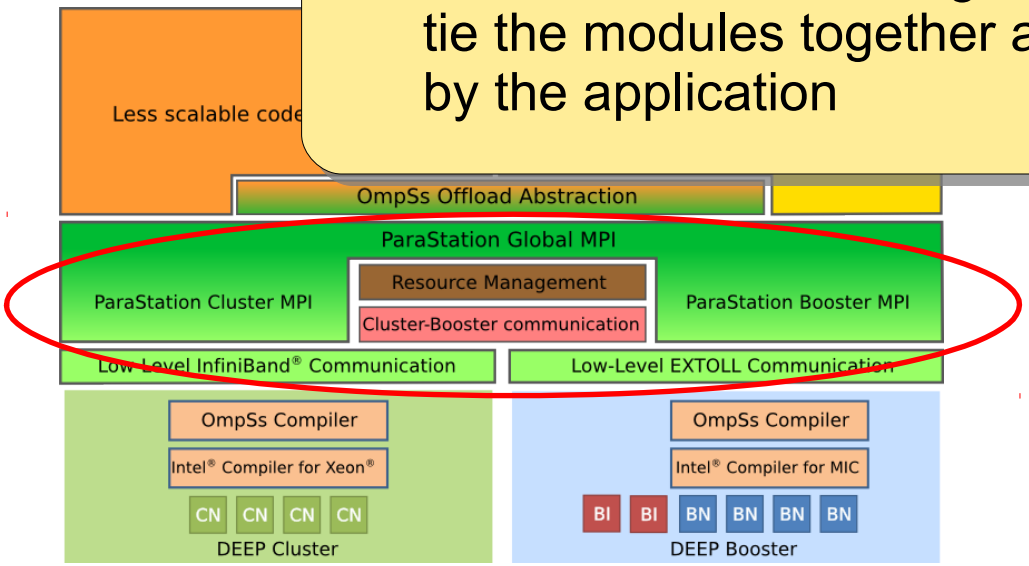
- ParaStation MPI powers DEEP's MPI-based offloading mechanism
  - *Using inter-communicators based on MPI_Comm_spawn()*
  - *Offloading of highly-scalable code parts to the Booster*
  - *Enable transparent Cluster-Booster data exchange via MPI*

ParaStation Global MPI in DEEP-ER

- Heterogeneous system: Cluster-Booster architecture
  - *Cluster Nodes (CN) with Intel*
  - *Boos Intel*
  - *Boos conn*



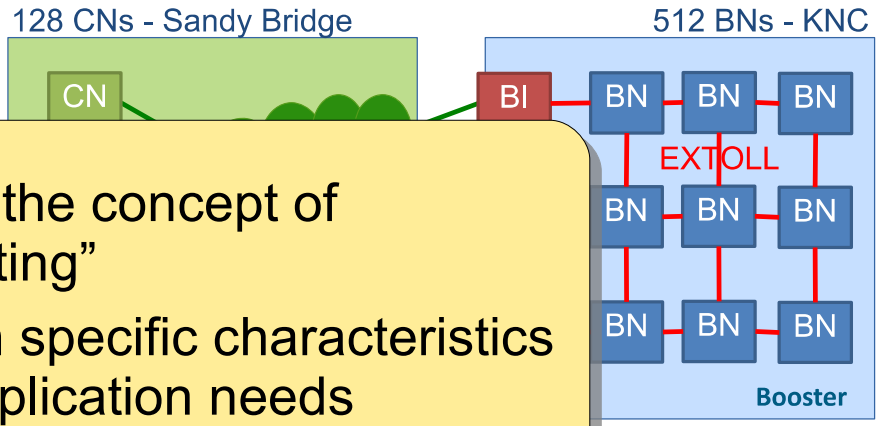
- Logically extensible to the concept of "Modular Supercomputing"
- Combine modules with specific characteristics adapted to different application needs
- ParaStation then acting as sort of a "glue" to tie the modules together and ease their use by the application



- *…rs DEEP's …mechanism*
- *Using inter-communicators based on MPI_Comm_spawn()*
- *Offloading of highly-scalable code parts to the Booster*
- *Enable transparent Cluster-Booster data exchange via MPI*

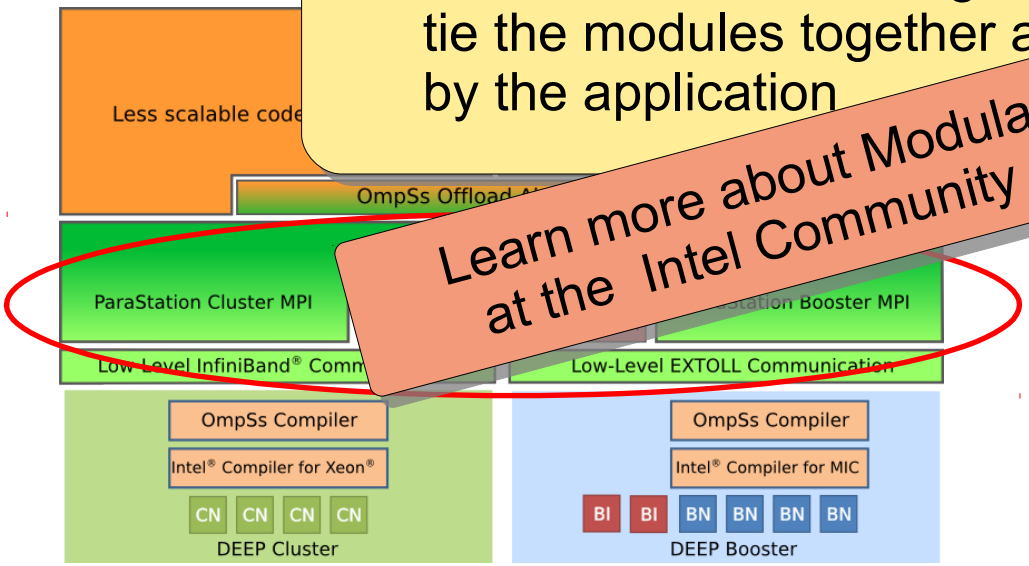
ParaStation Global MPI in DEEP-ER

- Heterogeneous system: Cluster-Booster architecture
  - *Cluster Nodes (CN) with Intel*
  - *Boos Intel*
  - *Boos conn*



- Logically extensible to the concept of "Modular Supercomputing"
- Combine modules with specific characteristics adapted to different application needs
- ParaStation then acting as sort of a "glue" tie the modules together and e by the application

Learn more about Modular Supercomputing at the Intel Community Hub on Thursday



rs DEEP's
mechanism
based on MPI_Comm_spawn()

- *Offloading of highly-scalable code parts to the Booster*
- *Enable transparent Cluster-Booster data exchange via MPI*

ParTec enables HPC!

- Strong general purpose cluster specialist for more than a decade
  - *Spin-off of the University of Karlsruhe*
  - *Working as SME since 1999 in the fields of cluster computing*
- Unrivaled expertise in developing cluster software
- ParTec was elected as the partner of choice by some leading HPC sites across Europe, especially by the Jülich Supercomputing Centre (JSC) for the JuRoPA and JURECA projects
- Engagement and active development in projects towards Exascale

Technical activities

- *JuRoPA family*
- *JURECA*
- *ExaCluster Lab*
- ***DEEP Project***
- ***DEEP-ER Project***
- *OpenHPC Initiative*
- *ParaStationConsortium*

Political activities

- *EOFS & Exascale10*
- *PROSPECT e.V.*
- *ETP4HPC*

Thank you!

Questions?

http://www.par-tec.com