

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*
- 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)*

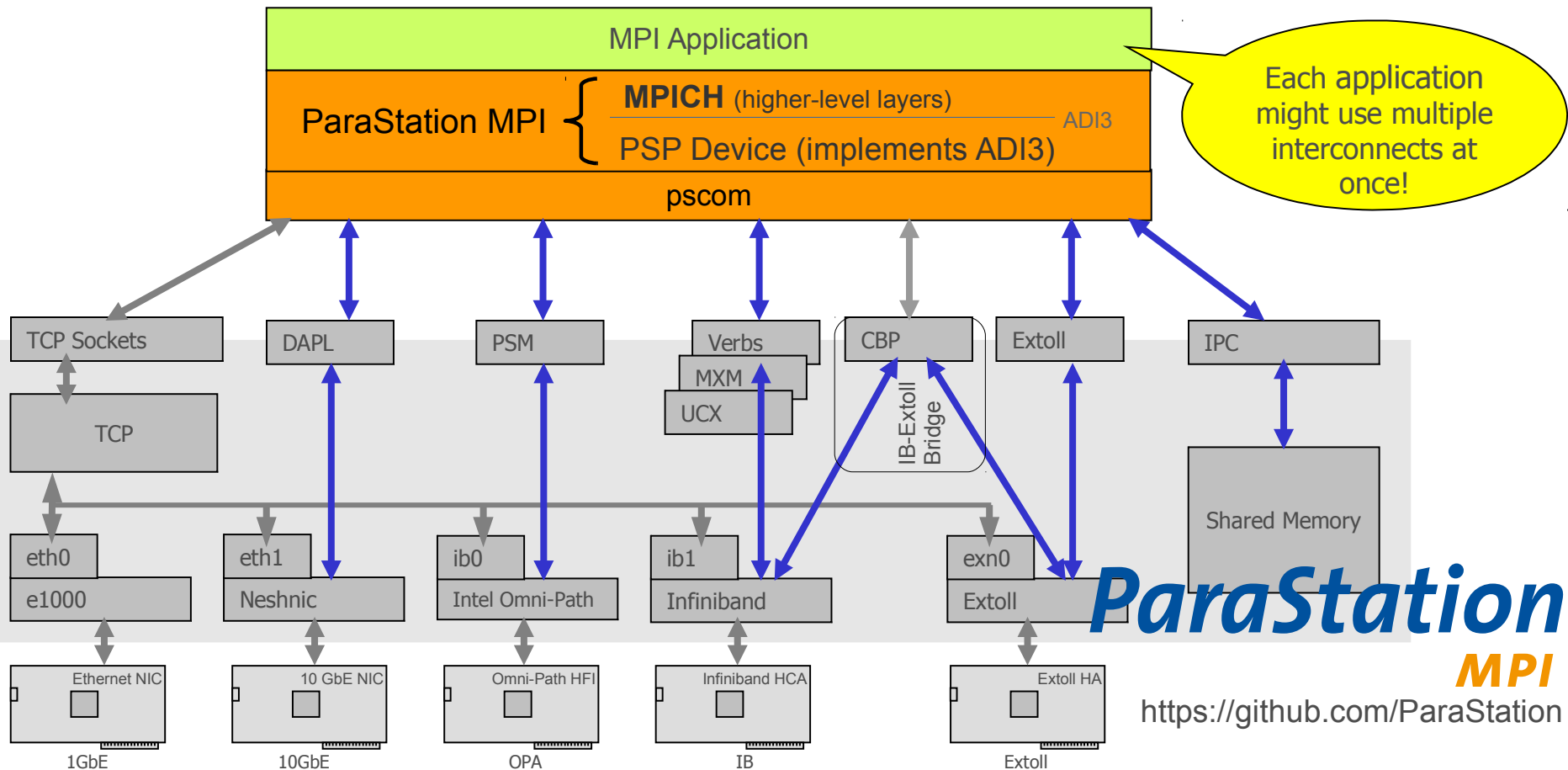
Maximize job throughput
Minimize administration

ParaStation
V5

Σ ParaStation Cluster Suite

ParaStation MPI

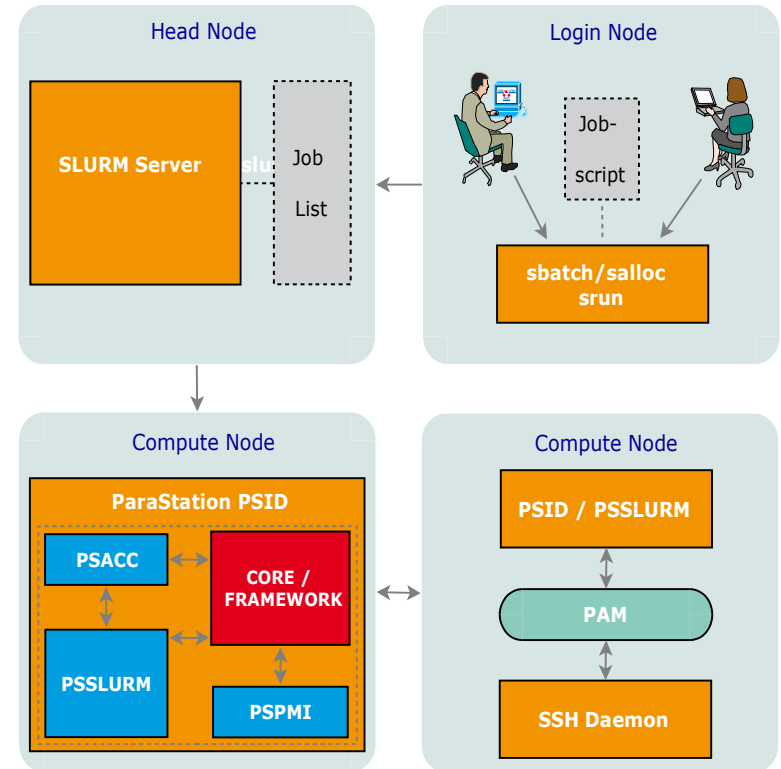
- Based on mpich-3.1.4
- Powered by “pscom”: efficient, low-level communication library
- Supporting a wide range of interconnects – even in parallel



ParaStation MPI

- 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!

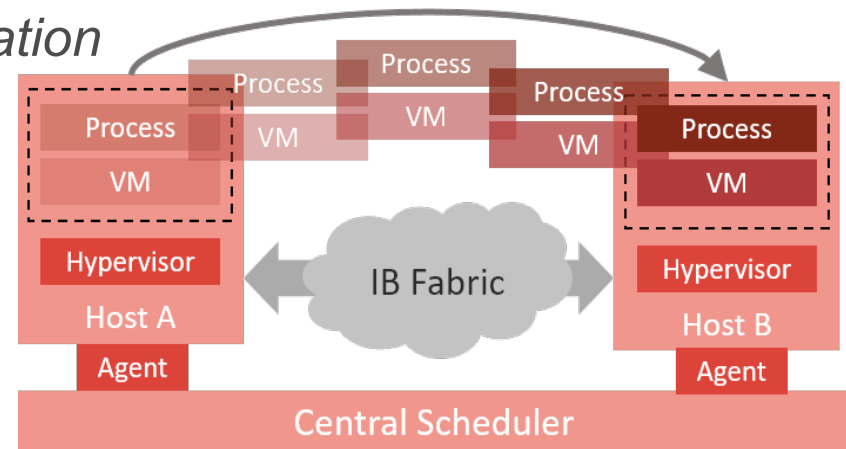


ParaStation
MPI

<https://github.com/ParaStation>

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 truly 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*

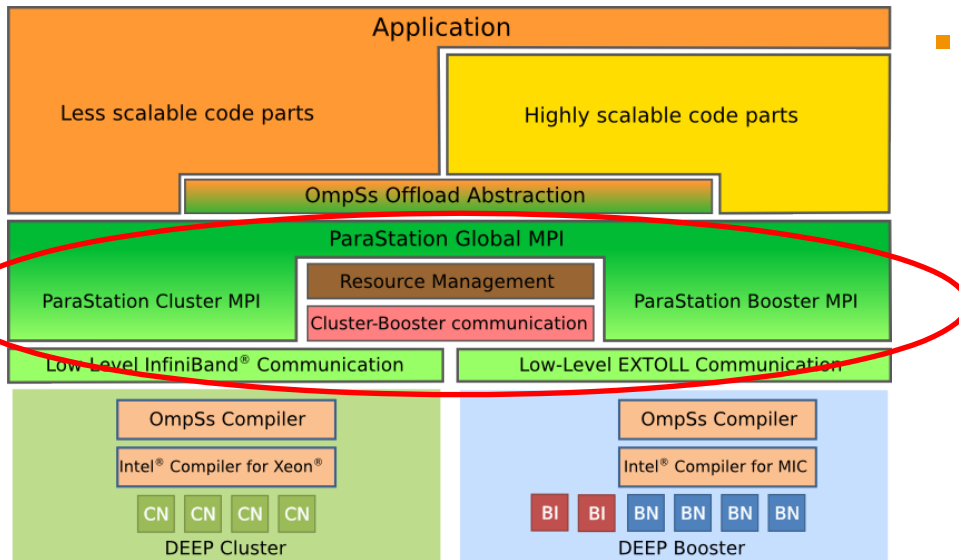
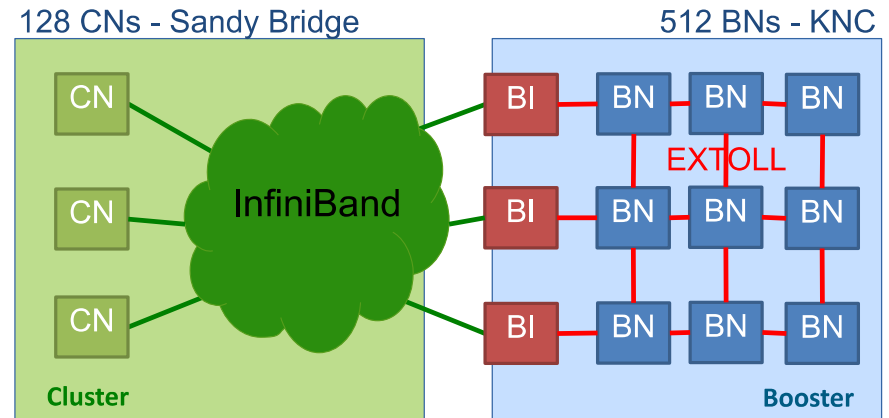


ParaStation

MPI

<https://github.com/ParaStation>

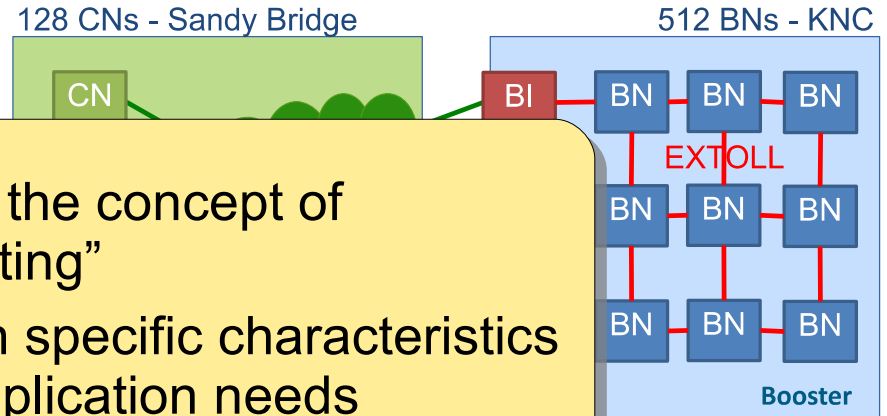
- Heterogeneous system: Cluster-Booster architecture
 - Cluster Nodes (CN) with Intel Xeon multi-core CPUs
 - Booster Nodes (BN) with Intel MIC many-core CPUs
 - Booster Interfaces (BI) connecting Cluster and Booster



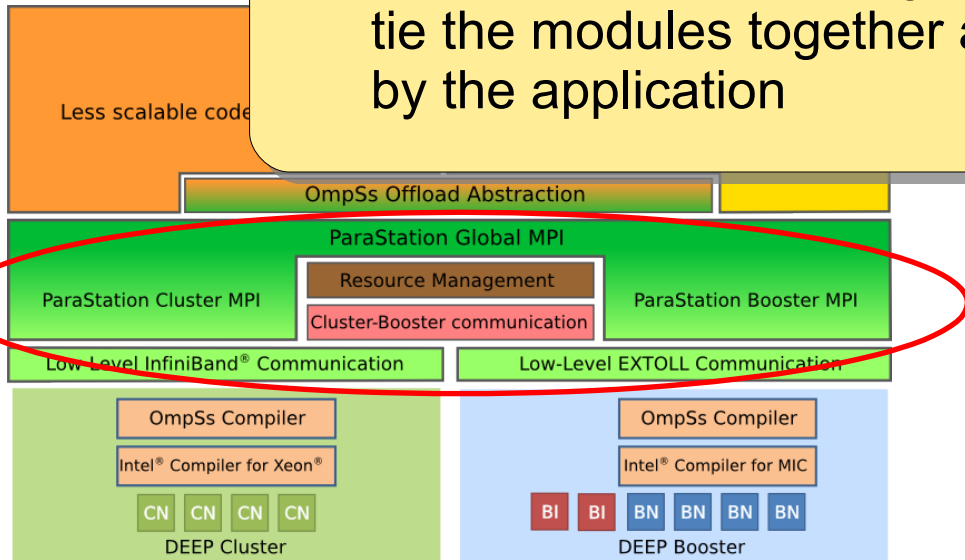
- ParaStation MPI powers DEEP's MPI-based offloading mechanism
 - Using inter-communicators based on `MPI_Comm_spawn()`
 - Offloading of highly-scalable code parts to the Booster
 - Enable transparent Cluster-Booster data exchange via MPI

- Heterogeneous system: Cluster-Booster architecture

- Cluster Nodes (CN) with Intel Xeon
- Booster Nodes (BN) with Intel Xeon
- Booster Nodes (BN) with Intel Xeon



- 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



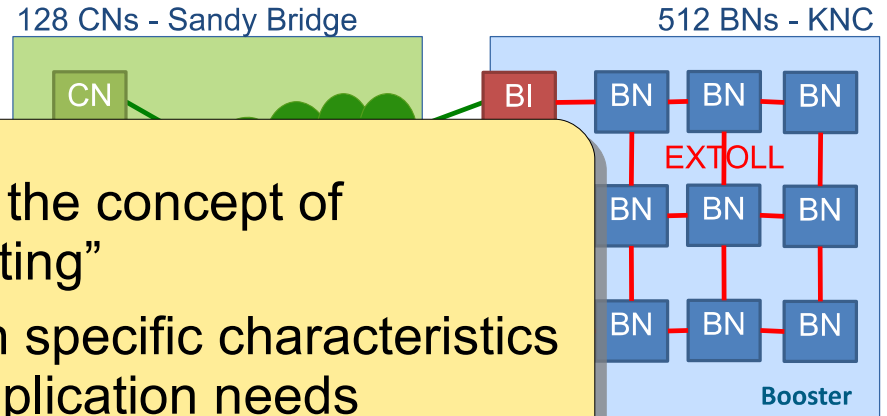
...ers DEEP's mechanism

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

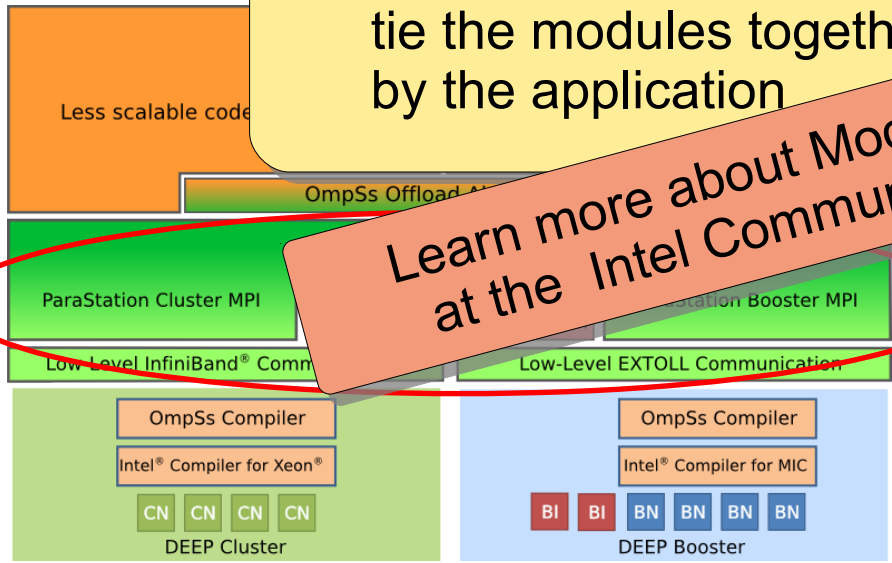
- Heterogeneous system: Cluster-Booster architecture

- Cluster Nodes (CN) with Intel Xeon
- Booster Nodes (BN) with Intel Xeon Phi
- Booster Nodes (BN) with Intel Xeon Phi



- 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 enable scaling by the application

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



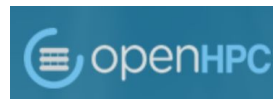
- DEEP-ER uses DEEP’s offloading mechanism based on MPI_Comm_spawn()
- Offloading of highly-scalable code parts to the Booster
- Enable transparent Cluster-Booster data exchange via MPI

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*
- Unrivalled 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*



Questions?

<http://www.par-tec.com>



The screenshot shows the ParTec website with the following content:

- Navigation:** Home, Company, News, Products, Co-Operations, Contact, Support, Login, Inprint.
- ParTec's Cluster Competence Center:**
 - ParTec's Cluster Competence Center offers the software, consultancy and support services necessary to achieve new insights in productivity and availability of today's commodity-based supercomputers.
 - ParaStation, an open source project developed and maintained by ParTec, has unique features specifically designed to address the challenges of scalability and reliability on extremely large clusters.
 - ParTec's vast range of expertise in the field of parallelisation, consultancy and support has made it the partner of choice in some of the leading HPC sites across Europe.
- Industry Sectors:** Computer Aided Engineering, Life Science, Financial, Scientific, Energy, Government.
- Customer Reference: Research Center Jülich:**
 - In June 2009, ParaStation VMS propelled the 3200 node JuRoPA cluster at the Jülich Supercomputing Centre to an impressive 374.8 TeraFlops of sustained performance with a parallel efficiency of 95%.
 - At that time JuRoPA Ranked No. 10 in the World and No. 1 in Europe of the most powerful general purpose supercomputers (top500.org).
 - ParaStation MPI was shown to scale to more than 25,000 MPI processes (without the need for OS to be modified).
 - JuRoPA was inaugurated May 26, 2009 by the German Federal Minister for Education and Research, Prof. Dr. Annette Schavan, the Prime Minister of North Rhine-Westphalia, Dr. Jochen Rüttge, and Prof. Dr. Achim Huckert, Chairman of the Board of Directors at Research Center Jülich as well as high-ranking international guests from academia, industry and politics.
- ParTec - Your Trusted Partner delivering - High Availability - Scalability - Maximum Utilization:**
 - ParTec understands that in today's economy it is crucial to utilize the largest HPC installations with demanding requirements in areas such as power consumption and cooling efficiency. However, effective use of resources demands that utilization is able to sustain availability, system utilization and parallel efficiency goals.
 - ParTec has developed software, administration tools, maintenance, backup and archiving policies which maximize machine utilization and let our customers maximize return on their investment.
- Right Side Navigation:**
 - Latest News:**
 - 16.12.2010: Foundation of The ESSEF (ESA Euro Flight Systems) @ FZJ DLR
 - 22.10.2010: Conference on International Electronic Warfare Protect
 - 18.10.2010: Manager's Update - ParTec's National Energy Cluster Computing
 - 22.07.2010: Researchers Seeking the Fourth Dimension of Big Data Using JuRoPA Cluster
 - 17.06.2010: Julia - One Year On - Single File for Sale at SC 2010 (Dresden)
 - Jobs - We're Hiring:** Software Developer
 - Upcoming Events:**
 - 02.03.2011: Meet our booth 651 on ISC Conference June 19-23 Hannover, Germany
 - Our Products:**
 - ClusterSupport
 - ParaStation
 - ParaStation MPI
 - ParaStation
 - ParaStation MPI
 - Latest versions of documentation
 - ParaStation is Intel Cluster Ready
 - Learn More
 - Running Projects:**
 - ESSEF: Integrated systems and application specific for massive parallel computer clusters
 - DLR: DLR's Deutsche Ozean-Flotte
 - ee-Cluster: ee-Cluster Energy Efficient Cluster Computing
 - ee-Cluster: ee-Cluster Energy Efficient Cluster Computing
 - ee-Cluster: ee-Cluster Energy Efficient Cluster Computing
 - Services Overview:**
 - Technical consulting from our team of HPC engineers, programmers and administrators
 - Procurement, installation and commissioning from a single source
 - Project management and contract negotiation services
 - Vendor neutral hardware and software selection
 - Custom software development
 - Flexible service and support packages