

ParaStationV5

ParaStationV5 with their integrated parts **ParaStationMPI** and **GridMonitor** form ParTec's compute cluster OS environment designed to deliver 'High Performance' and 'Maximum Availability' on today's commodity based supercomputers. **ParaStationMPI** has demonstrated scalability in excess of 24,000 processes over more than 3,000 compute nodes – making it a true high-performance MPI suitable for environments requiring proven scalability. Stability is another crucial requirement of the cluster environment. The ability to identify and resolve resource and hardware issues before they cause job failures is fundamental. The **GridMonitor** provides administrators with the early warning mechanism and visibility of cluster parameters required to maintain cluster availability.

ParaStationV5 is a comprehensive middleware-solution for compute clusters and farms. It provides a stable, reliable platform for commodity clusters and produces measurable productivity gains. User-friendly interfaces and cluster provisioning tools greatly simplify cluster administration.

Communication

ParaStationV5 provides a standard interface for parallel applications - MPI (MPI-2). Intra-node communication is handled via a shared memory model (shmem), while inter-node communication can be achieved using any of the following **interconnects**:

- Myrinet,
- Fast Ethernet, Gigabit Ethernet,
- 10G-Ethernet
- Infiniband, InfiniPath

Optimized protocol-stack for Ethernet communication (P4Sock)

The P4Sock protocol for Ethernet was developed and optimized for cluster communication. It provides a secure, high performance, low latency inter-process data transfer protocol.

TCP-Bypass provides higher performance

TCP-Bypass routes standard TCP socket based communication directly to the high-performance P4Sock protocol layer. It is not necessary to modify or re-link the application.

The job-clean-up functionality cancels a job automatically if a hardware problem occurs. All processes belonging to a dedicated job are terminated and no orphan processes remain.

The pre-emption functionality allows operators to suspend running jobs, and allow higher priority job to proceed. The suspended jobs can be restarted by operator request.

Features:

- A single system view for managing multiple clusters
- Capable of managing heterogeneous clusters
- Single-point of cluster management
- Process management for parallel and serial jobs
- MPI-2 (Ethernet, Myrinet, Infiniband, P4sock, shmem, QsNet and iwarp)
- Administration command console (psadmin)
- Parallel shell (psh)
- High-performance parallel file copy
- Resource management for load balancing
- Job queuing system (first-in first-out)
- Test utility for communication path
- Installation-Management (Cluster installer)
- Job accounting information
- Extensible system monitoring tools (GridMonitor) collecting cluster parameters and device values from compute nodes, switches, UPS, RAID etc.
- Comprehensive graphical user interface
- Event notification for „out-of-range“ system values
- Multi platform support
- Multi vendor support
- Automated system installation tool
- Multi-core aware
- Integration with standard batch systems (PBS/moab)
- Independent of Linux distribution



The **virtual nodes** concept eliminates the need for hard-coded node lists. The pool of nodes is managed by distributed daemons. Non-operational nodes are ignored. Users can, however, work with node-lists if they wish to target specific nodes.

Serial Applications

Nonparallel applications can also be started under **ParaStationV5** to ensure full process control and resource management (Farming).

Single-point of cluster management

- Reduced software complexity
- Smooth & secured upgrade path to future releases
- Multi cluster support
- Multi platform support

ParaStationV5 supports cluster **partitioning**, or the exclusive allocation of processors or nodes, enabling specific jobs or users to access dedicated resources.

Daemon concept

- **ParaStationV5** daemon runs on each node (no rsh / ssh login)
- Sophisticated **process-management** to start, monitor and close processes
- Distributed management logic
- Job suspend and resume
- fault tolerant – failed nodes are configured out§



Monitoring

ParaStationV5 allows cluster administrators to work with the command line (PSIadmin) or use a browser based user interface (GridMonitor) to monitor the following cluster state:

- Displays all running jobs and applications
- Displays the current load on the cluster
- Displays all relevant hardware parameters, devices, users and processes
- IPMI compatible
- Event notification

HW-Platforms

Intel / AMD x86-32, x86-64, IA64,
IBM PowerPC and Power based cluster

SW-Platforms

SUSE Linux Enterprise Server (SLES),
RedHat and other major Linux distributions

Proven Reliability and Performance

ParaStationV5 in global Top10 (see top500.org)

JuRoPA cluster at Jülich Supercomputer Centre (Germany):

3288 compute nodes (26304 cores)
274,8 Teraflops Linpack performance (June 2009)
QDR Infiniband

ParaStationV5 Operating and Management

ParaStationMPI

SLES11 Linux

Intel Compiler and Development Tools

MOAB Scheduler

Lustre Filesystem

ParaStationMPI delivered proven scalability running more than 25,000 MPI tasks without any threading library with parallel efficiencies in excess of 91,6%.

Certifications



Intel Cluster Ready