transtec
accelerate productivity

IBM

Simulation
Risk Analysis
High Throughput Computing
Automotive
Price Modelling
Engineering
CAE
Aerospace
CAD
Big Data Analytics
Life Sciences

HIGH PERFORMANCE COMPUTING 2012/13

TECHNOLOGY COMPASS IBM SPECIAL

TECHNOLOGY COMPASS

TABLE OF CONTENTS AND INTRODUCTION

MORE THAN 30 YEARS OF EXPERIENCE IN SCIENTIFIC COMPUTING

1980 marked the beginning of a decade where numerous startups were created, some of which later transformed into big players in the IT market. Technical innovations brought dramatic changes to the nascent computer market. In Tübingen, close to one of Germany's prime and oldest universities, transtec was founded.

In the early days, transtec focused on reselling DEC computers and peripherals, delivering high-performance workstations to university institutes and research facilities. In 1987, SUN/Sparc and storage solutions broadened the portfolio, enhanced by IBM/RS6000 products in 1991. These were the typical workstations and server systems for high performance computing then, used by the majority of researchers worldwide.

In the late 90s, transtec was one of the first companies to offer highly customized HPC cluster solutions based on standard Intel architecture servers, some of which entered the TOP500 list of the world's fastest computing systems.

Thus, given this background and history, it is fair to say that transtec looks back upon a more than 30 years' experience in scientific computing; our track record shows nearly 500 HPC installations. With this experience, we know exactly what customers' demands are and how to meet them. High performance and ease of management – this is what customers require today. HPC systems are for sure required to peak-perform, as their name indicates, but that is not enough: they must also be easy to handle. Unwieldy design and operational complexity must be avoided or at least hidden from administrators and particularly users of HPC computer systems.

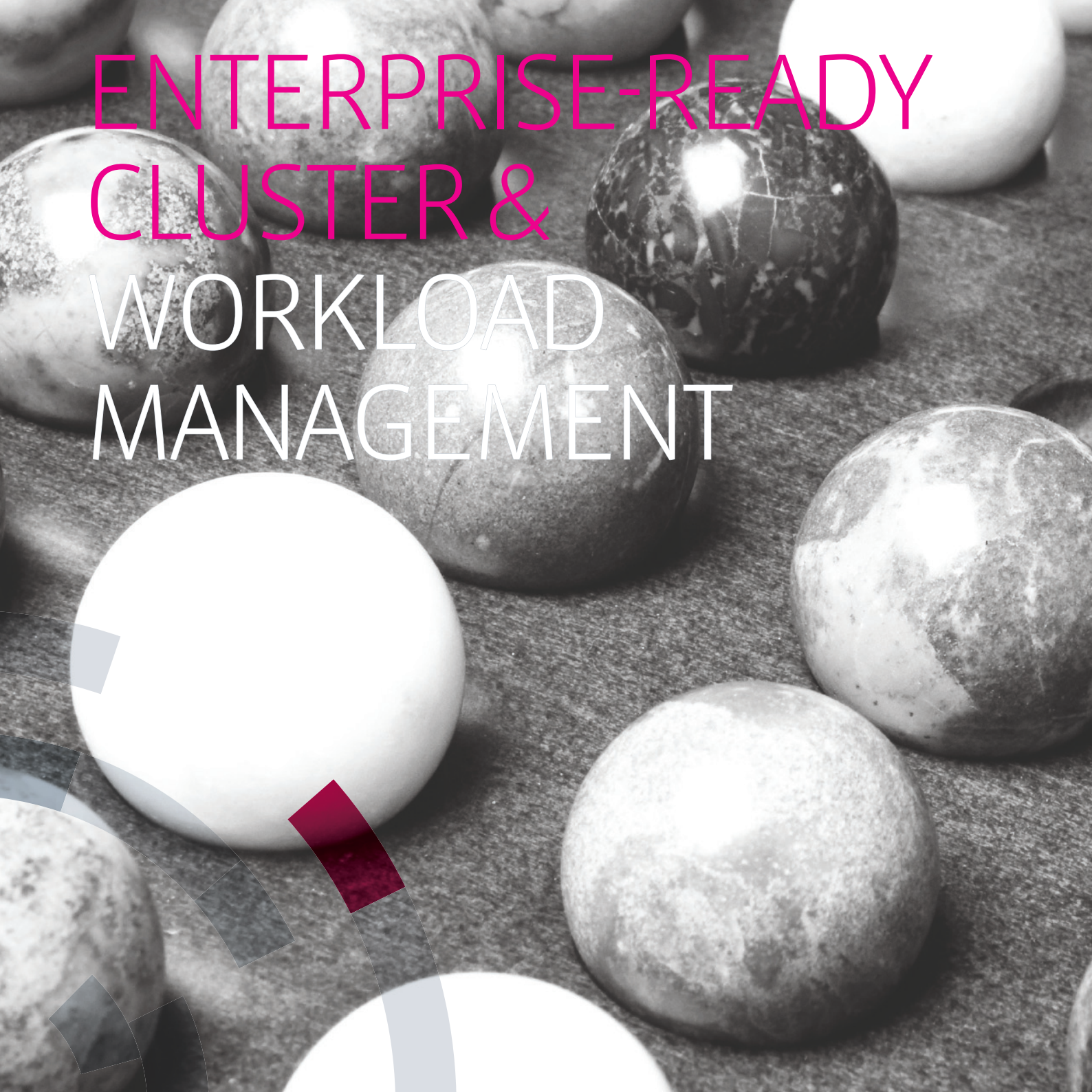
transtec HPC solutions deliver ease of management, both in the Linux and Windows worlds, and even where the customer´s environment is of a highly heterogeneous nature. Even the dynamical provisioning of HPC resources as needed does not constitute any problem, thus further leading to maximal utilization of the cluster.

transtec HPC solutions use the latest and most innovative technology. Their superior performance goes hand in hand with energy efficiency, as you would expect from any leading edge IT solution. We regard these basic characteristics.

In 2010, transtec entered into a strategic partnership with IBM, surely one of the biggest players in the HPC world with a very strong brand. The flexibility and long-year experience of transtec, combined with the power and quality of IBM HPC systems constitute a perfect symbiosis and provide customers with the most optimal HPC solution imaginable. IBM iDataPlex systems are highly optimized for HPC workload in datacenter environments, regarding performance, flexibility, and energy, space and cooling efficiency. Platform HPC and LSF are both enterprise-ready HPC cluster and workload management solutions and are widespread in all kinds of industrial HPC environments.

Last but not least, your decision for a transtec HPC solution means you opt for most intensive customer care and best service in HPC. Our experts will be glad to bring in their expertise and support to assist you at any stage, from HPC design to daily cluster operations, to HPC Cloud Services.

Have fun reading the transtec HPC Compass 2012/13 IBM Special!


ENTERPRISE-READY
CLUSTER &
WORKLOAD
MANAGEMENT

High performance computing (HPC) is becoming a necessary tool for organizations to speed up product design, scientific research, and business analytics. However, there are few software environments more complex to manage and utilize than modern high performance computing clusters. Therefore, addressing the problem of complexity in cluster management is a key aspect of leveraging HPC to improve time to results and user productivity.

ENTERPRISE-READY CLUSTER & WORKLOAD MANAGEMENT

IBM PLATFORM HPC

INTRODUCTION

Clusters based on the Linux® operating system have become increasingly prevalent at large supercomputing centers and continue to make significant in-roads in commercial and academic settings. This is primarily due to their superior price/performance and flexibility, as well as the availability of commercial applications that are based on the Linux OS.

Ironically, the same factors that make Linux a clear choice for high performance computing often make the operating system less accessible to smaller computing centers. These organizations may have Microsoft Windows administrators on staff, but have little or no Linux or cluster management experience. The complexity and cost of cluster management often outweigh the benefits that make open, commodity clusters so compelling. Not only can HPC cluster deployments be difficult, but the ongoing need to deal with heterogeneous hardware and operating systems, mixed workloads, and rapidly evolving toolsets make deploying and managing an HPC cluster a daunting task.

These issues create a barrier to entry for scientists and researchers who require the performance of an HPC cluster, but are limited to the performance of a workstation. This is why ease of use is now mandatory for HPC cluster management. This paper reviews the most complete and easy to use cluster management solution, Platform HPC, which is now commercially available from Platform Computing.

THE CLUSTER MANAGEMENT CHALLENGE

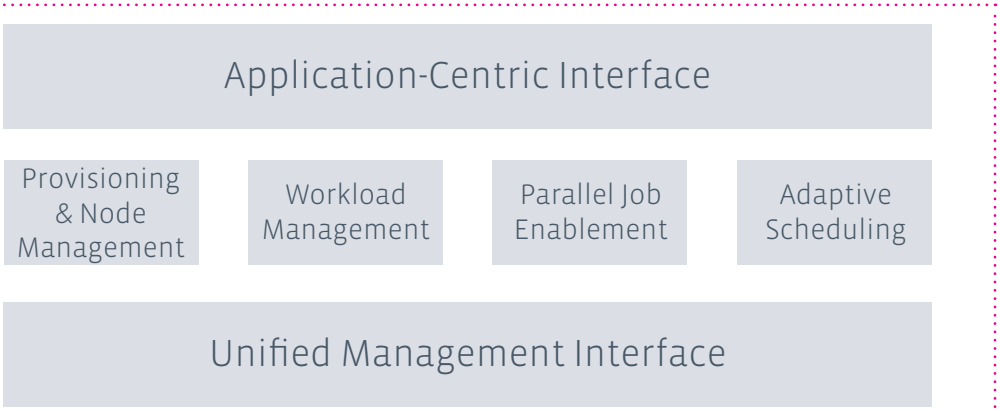
To provide a proper HPC application environment, system administrators need to provide a full set of capabilities to their users, as shown below. These capabilities include cluster provisioning and node management, application workload

management, and an environment that makes it easy to develop, run and manage distributed parallel applications.

Modern application environments tend to be heterogeneous; some workloads require Windows compute hosts while others require particular Linux operating systems or versions.

The ability to change a node's operating system on-the-fly in response to changing application needs - referred to as adaptive scheduling - is important since it allows system administrators to maximize resource use, and present what appears to be a larger resource pool to cluster users.

ESSENTIAL COMPONENTS OF AN HPC CLUSTER SOLUTION



Learning how to use a command line interface to power-up, provision and manage a cluster is extremely time-consuming. Administrators therefore need remote, web-based access to their HPC environment that makes it easier for them to install and manage an HPC cluster. An easy-to-use application-centric web interface can have tangible benefits including improved productivity, reduced training requirements, reduced errors rates, and secure remote access.

While there are several cluster management tools that address parts of these requirements, few address them fully, and some tools are little more than collections of discrete open-source software components.

Some cluster toolkits focus largely on the problem of cluster provisioning and management. While they clearly simplify cluster deployment, administrators wanting to make changes to node configurations or customize their environment will quickly find themselves hand-editing XML configuration files or writing their own shell scripts. Third-party workload managers and various open-source MPI libraries might be included as part of a distribution. However, these included components are loosely integrated and often need to be managed separately from the cluster manager. As a result the cluster administrator needs to learn how to utilize each additional piece of software in order to manage the cluster effectively.

Other HPC solutions are designed purely for application workload management. While these are all capable workload managers, most do not address at all the issue of cluster management, application integration, or adaptive scheduling. If such capabilities exist they usually require the purchase of additional software products.

Parallel job management is also critical. One of the primary reasons that customers deploy HPC clusters is to maximize

ENTERPRISE-READY CLUSTER & WORKLOAD MANAGEMENT

IBM PLATFORM HPC

> *"Platform HPC and Platform LSF have been known for many years as a highest-quality and enterprise-ready cluster and workload management solutions, and we have many customers in academia and industry relying on them."*
>
> **Dr. Oliver Tennert** Director Technology Management & HPC Solutions

application performance. Processing problems in parallel is a common way to achieve performance gains. The choice of MPI, its scalability, and the degree to which it is integrated with various OFED drivers and high performance interconnects has a direct impact on delivered application performance. Furthermore, if the workload manager does not incorporate specific parallel job management features, busy cluster users and administrators can find themselves manually cleaning up after failed MPI jobs or writing their own shell scripts to do the same.

Complexity is a real problem. Many small organizations or departments grapple with a new vocabulary full of cryptic commands, configuring and troubleshooting Anaconda kick start scripts, finding the correct OFED drivers for specialized hardware, and configuring open source monitoring systems like Ganglia or Nagios. Without an integrated solution administrators may need to deal with dozens of distinct software components, making managing HPC cluster implementations extremely tedious and time-consuming.

RE-THINKING HPC CLUSTERS

Clearly these challenges demand a fresh approach to HPC cluster management. Platform HPC represents a "re-think" of how HPC clusters are deployed and managed. Rather than addressing only part of the HPC management puzzle, Platform HPC addresses all facets of cluster management. It provides:

- A complete, easy-to-use cluster management solution
- Integrated application support
- User-friendly, topology-aware workload management
- Robust workload and system monitoring and reporting

- Dynamic operating system multi-boot (adaptive scheduling)
- GPU scheduling
- Robust commercial MPI library (Platform MPI)
- Web-based interface for access anywhere

Most complete HPC cluster management solution

Platform HPC makes it easy to deploy, run and manage HPC clusters while meeting the most demanding requirements for application performance and predictable workload management. It is a complete solution that provides a robust set of cluster management capabilities; from cluster provisioning and management to workload management and monitoring. The easy-to -use unified web portal provides a single point of access into the cluster, making it easy to manage your jobs and optimize application performance.

Platform HPC is more than just a stack of software; it is a fully integrated and certified solution designed to ensure ease of use and simplified troubleshooting.

Integrated application support

High performing, HPC-optimized MPI libraries come integrated with Platform HPC, making it easy to get parallel applications up and running. Scripting guidelines and job submission templates for commonly used commercial applications simplify job submission, reduce setup time and minimize operation errors. Once the applications are up and running, Platform HPC improves application performance by intelligently scheduling resources based on workload characteristics.

Fully certified and supported

Platform HPC unlocks cluster management to provide the easiest and most complete HPC management capabilities while reducing overall cluster cost and improving administrator productivity. It is based on the industry's most mature and robust workload manager, Platform LSF, making it the most reliable solution on the market.

Other solutions are typically a collection of open-source tools, which may also include pieces of commercially developed software. They lack key HPC functionality and vendor support, relying on the administrator's technical ability and time to implement. Platform HPC is a single product with a single installer and a unified web-based management interface. With the best support in the HPC industry, Platform HPC provides the most complete solution for HPC cluster management.

COMPLETE SOLUTION

Platform HPC provides a complete set of HPC cluster management features. In this section we'll explore some of these unique capabilities in more detail.

Easy to use, cluster provisioning and management

With Platform HPC, administrators can quickly provision and manage HPC clusters with unprecedented ease. It ensures maximum uptime and can transparently synchronize files to cluster nodes without any downtime or re-installation.

Fast and efficient software Installation – Platform HPC can be installed on the head node and takes less than one hour using three different mechanisms:

- Platform HPC DVD
- Platform HPC ISO file
- Platform partner's factory install bootable USB drive

ENTERPRISE-READY CLUSTER & WORKLOAD MANAGEMENT

WHAT'S NEW IN IBM PLATFORM LSF 8

Installing software on cluster nodes is simply a matter of associating cluster nodes with flexible provisioning templates through the web-based interface.

Flexible provisioning – Platform HPC offers multiple options for provisioning Linux operating environments that include:

- Package-based provisioning
- Image based provisioning
- Diskless node provisioning

Large collections of hosts can be provisioned using the same provisioning template. Platform HPC automatically manages details such as IP address assignment and node naming conventions that reflect the position of cluster nodes in data center racks.

Unlike competing solutions, Platform HPC deploys multiple operating systems and OS versions to a cluster simultaneously. This includes Red Hat® Enterprise Linux, CentOS, Scientific Linux, and SUSE® Linux Enterprise Server. This provides administrators with greater flexibility in how they serve their user communities and means that HPC clusters can grow and evolve incrementally as requirements change.

WHAT'S NEW IN IBM PLATFORM LSF 8

Written with Platform LSF administrators in mind, this brief provides a short explanation of significant changes in Platform's latest release of Platform LSF, with a specific emphasis on scheduling and workload management features.

ABOUT IBM PLATFORM LSF 8

Platform LSF is the most powerful workload manager for demanding, distributed high performance computing environments. It provides a complete set of workload management capabilities, all designed to work together to reduce cycle times and maximize productivity in missioncritical environments.

This latest Platform LSF release delivers improvements in performance and scalability while introducing new features that simplify administration and boost user productivity. This includes:

- **Guaranteed resources** – Aligns business SLA's with infrastructure configuration for simplified administration and configuration
- **Live reconfiguration** – Provides simplified administration and enables agility
- **Delegation of administrative rights** – Empowers line of business owners to take control of their own projects
- **Fairshare & pre-emptive scheduling enhancements** – Fine tunes key production policies

PLATFORM LSF 8 FEATURES

Guaranteed Resources Ensure Deadlines are Met

In Platform LSF 8, resource-based scheduling has been extended to guarantee resource availability to groups of jobs. Resources can be slots, entire hosts or user-defined shared resources such as software licenses. As an example, a business unit might guarantee that it has access to specific types of resources within ten minutes of a job being submitted, even while sharing resources between departments. This facility ensures that lower priority jobs using the needed resources can be pre-empted in order to meet the SLAs of higher priority jobs.

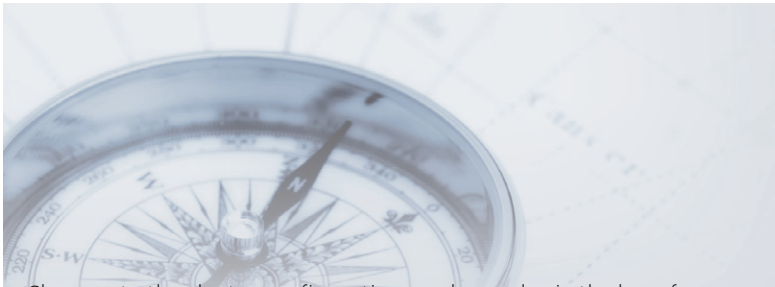
Because jobs can be automatically attached to an SLA class via access controls, administrators can enable these guarantees without requiring that end-users change their job submission procedures, making it easy to implement this capability in existing environments.

Live Cluster Reconfiguration

Platform LSF 8 incorporates a new live reconfiguration capability, allowing changes to be made to clusters without the need to re-start LSF daemons. This is useful to customers who need to add hosts, adjust sharing policies or re-assign users between groups "on the fly", without impacting cluster availability or running jobs.

ENTERPRISE-READY CLUSTER & WORKLOAD MANAGEMENT

WHAT'S NEW IN IBM PLATFORM LSF 8



Changes to the cluster configuration can be made via the bconf command line utility, or via new API calls. This functionality can also be integrated via a web-based interface using Platform Application Center. All configuration modifications are logged for a complete audit history, and changes are propagated almost instantaneously. The majority of reconfiguration operations are completed in under half a second.

With Live Reconfiguration, down-time is reduced, and administrators are free to make needed adjustments quickly rather than wait for scheduled maintenance periods or non-peak hours. In cases where users are members of multiple groups, controls can be put in place so that a group administrator can only control jobs associated with their designated group rather than impacting jobs related to another group submitted by the same user.

Delegation of Administrative Rights

With Platform LSF 8, the concept of group administrators has been extended to enable project managers and line of business managers to dynamically modify group membership and fair-share resource allocation policies within their group. The ability to make these changes dynamically to a running cluster is made possible by the Live Reconfiguration feature.

These capabilities can be delegated selectively depending on the group and site policy. Different group administrators can manage jobs, control sharing policies or adjust group membership.

More Flexible Fairshare Scheduling Policies

To enable better resource sharing flexibility with Platform LSF 8, the algorithms used to tune dynamically calculated user priorities can be adjusted at the queue level. These algorithms can vary based on department, application or project team preferences. The Fairshare parameters ENABLE_HIST_RUN_TIME and HIST_HOURS enable administrators to control the degree to which LSF considers prior resource usage when determining user priority. The flexibility of Platform LSF 8 has also been improved by allowing a similar "decay rate" to apply to currently running jobs (RUN_TIME_DECAY), either system-wide or at the queue level. This is most useful for customers with long-running jobs, where setting this parameter results in a more accurate view of real resource use for the fairshare scheduling to consider.

Performance & Scalability Enhancements

Platform LSF has been extended to support an unparalleled scale of up to 100,000 cores and 1.5 million queued jobs for very high throughput EDA workloads. Even higher scalability is possible for more traditional HPC workloads.

Specific areas of improvement include the time required to start the master-batch daemon (MBD), bjobs query performance, job submission and job dispatching as well as impressive performance gains resulting from the new Bulk Job Submission feature. In addition, on very large clusters with large numbers of user groups employing fairshare scheduling, the memory footprint of the master batch scheduler in LSF has been reduced by approximately 70% and scheduler cycle time has been reduced by 25%, resulting in better performance and scalability.

More Sophisticated Host-based Resource Usage for Parallel Jobs

Platform LSF 8 provides several improvements to how resource use is tracked and reported with parallel jobs. Accurate tracking of how parallel jobs use resources such as CPUs, memory and swap, is important for ease of management, optimal scheduling and accurate reporting and workload analysis. With Platform LSF 8 administrators can track resource usage on a per-host basis and an aggregated basis (across all hosts), ensuring that resource use is reported accurately. Additional details such as running PIDs and PGIDs for distributed parallel jobs, manual cleanup (if necessary) and the development of scripts for managing parallel jobs are simplified. These improvements in resource usage reporting are reflected in LSF commands including bjobs, bhist and bacct.

Improved Ease of Administration for Mixed Windows and Linux Clusters

The lspasswd command in Platform LSF enables Windows LSF users to advise LSF of changes to their Windows level passwords. With Platform LSF 8, password synchronization between environments has become much easier to manage because the Windows passwords can now be adjusted directly from Linux hosts using the lspasswd command. This allows Linux users to conveniently synchronize passwords on Windows hosts without needing to explicitly login into the host.

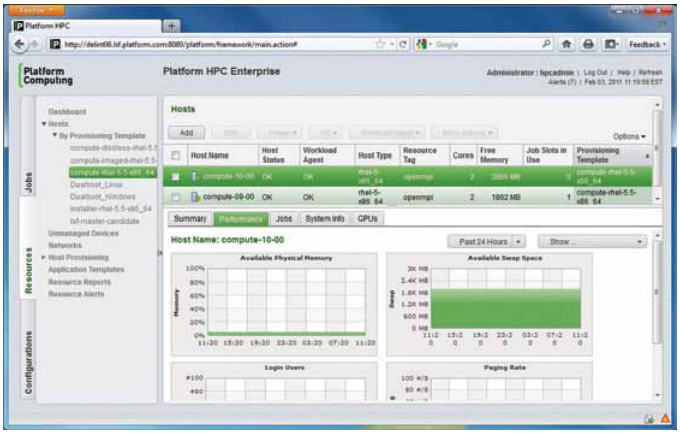
Bulk Job Submission

When submitting large numbers of jobs with different resource requirements or job level settings, Bulk Job Submission allows for jobs to be submitted in bulk by referencing a single file containing job details.

ENTERPRISE-READY CLUSTER & WORKLOAD MANAGEMENT

IBM PLATFORM HPC

RESOURCE MONITORING



Simplified configuration changes – Platform HPC simplifies administration and increases cluster availability by allowing changes such as new package installations, patch updates, and changes to configuration files to be propagated to cluster nodes automatically without the need to re-install those nodes. It also provides a mechanism whereby experienced administrators can quickly perform operations in parallel across multiple cluster nodes.

Repository snapshots / trial installations – Upgrading software can be risky, particularly in complex environments. If a new software upgrade introduces problems, administrators often need to rapidly "rollback" to a known good state. With other cluster managers this can mean having to re-install the entire cluster. Platform HPC incorporates repository snapshots, which are "restore points" for the entire cluster. Administrators can snapshot a known good repository, make changes to their environment, and easily revert to a previous "known good" repository in the event of an unforeseen problem. This powerful capability takes the risk out of cluster software upgrades.

New hardware integration – When new hardware is added to a cluster it may require new or updated device drivers that are not supported by the OS environment on the installer node. This means that a newly updated node may not network boot and provision until the head node on the cluster is updated with a new operating system; a tedious and disruptive process. Platform HPC includes a driver patching utility that allows updated device drivers to be inserted into existing repositories, essentially future proofing the cluster, and providing a simplified means of supporting new hardware without needing to re-install the environment from scratch.

Software updates with no re-boot – Some cluster managers always re-boot nodes when updating software, regardless of how minor the change. This is a simple way to manage updates. However, scheduling downtime can be difficult and disruptive. Platform HPC performs updates intelligently and selectively so that compute nodes continue to run even as non-intrusive updates are applied. The repository is automatically updated so that future installations include the software update. Changes that require the re-installation of the node (e.g. upgrading an operating system) can be made in a "pending" state until down-time can be scheduled.

USER-FRIENDLY, TOPOLOGY AWARE WORKLOAD MANAGE-MENT

Platform HPC includes a robust workload scheduling capability, which is based on Platform LSF - the industry's most powerful, comprehensive, policy driven workload management solution for engineering and scientific distributed computing environments. By scheduling workloads intelligently according to policy,
Platform HPC improves end user productivity with minimal system administrative effort. In addition, it allows HPC user teams to easily access and share all computing resources, while reducing time between simulation iterations.

GPU scheduling – Platform HPC provides the capability to schedule jobs to GPUs as well as CPUs. This is particularly advantageous in heterogeneous hardware environments as it means that administrators can configure Platform HPC so that only those jobs that can benefit from running on GPUs are allocated to those resources. This frees up CPU-based resources to run other jobs. Using the unified management interface, administrators can monitor the GPU performance as well as detect ECC errors.

UNIFIED MANAGEMENT INTERFACE

Competing cluster management tools either do not have a web-based interface or require multiple interfaces for managing different functional areas. In comparison, Platform HPC includes a single unified interface through which all administrative tasks can be performed including node-management, job-management, jobs and cluster monitoring and reporting. Using the unified management interface, even cluster administrators with very little Linux experience can competently manage a state of the art HPC cluster.

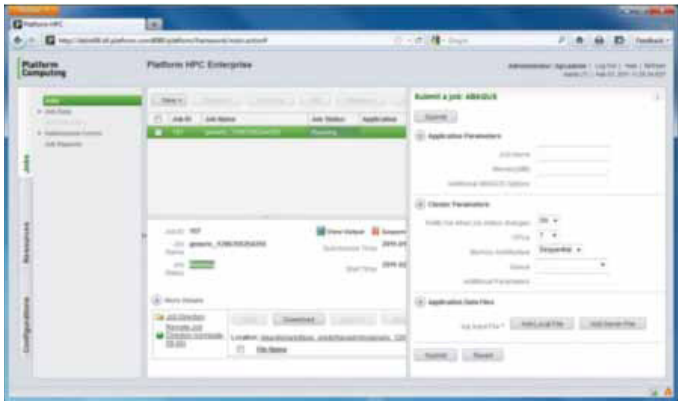
Job management – While command line savvy users can continue using the remote terminal capability, the unified web portal makes it easy to submit, monitor, and manage jobs. As changes are made to the cluster configuration, Platform HPC automatically re-configures key components, ensuring that jobs are allocated to the appropriate resources. The web portal is customizable and provides job data management, remote visualization and interactive job support.

Workload/system correlation – Administrators can correlate workload information with system load, so that they can make timely decisions and proactively manage compute resources against business demand. When it's time for capacity planning, the management interface can be used to run detailed reports and analyses which quantify user needs and remove the guess work from capacity expansion.

ENTERPRISE-READY CLUSTER & WORKLOAD MANAGEMENT

IBM PLATFORM HPC

JOB SUBMISSION TEMPLATES



Simplified cluster management – The unified management console is used to administer all aspects of the cluster environment. It enables administrators to easily install, manage and monitor their cluster. It also provides an interactive environment to easily package software as kits for application deployment as well as pre-integrated commercial application support. One of the key features of the interface is an operational dashboard that provides comprehensive administrative reports. As the image illustrates, Platform HPC enables administrators to monitor and report on key performance metrics such as cluster capacity, available memory and CPU utilization. This enables administrators to easily identify and troubleshoot issues.

The easy to use interface saves the cluster administrator time, and means that they do not need to become an expert in the administration of open-source software components. It also reduces the possibility of errors and time lost due to incorrect configuration. Cluster administrators enjoy the best of both worlds – easy access to a powerful, web-based cluster manager without the need to learn and separately administer all the tools that comprise the HPC cluster environment.

ROBUST COMMERCIAL MPI LIBRARY

Platform MPI – In order to make it easier to get parallel applications up and running, Platform HPC includes the industry's most robust and highest performing MPI implementation, Platform MPI. Platform MPI provides consistent performance at application run-time and for application scaling, resulting in top performance results across a range of third-party benchmarks.

Open Source MPI – Platform HPC also includes various other industry standard MPI implementations. This includes MPICH1, MPICH2 and MVAPICH1, which are optimized for cluster hosts connected via InfiniBand, iWARP or other RDMA based interconnects.

INTEGRATED APPLICATION SUPPORT

Job submission templates – Platform HPC comes complete with job submission templates for ANSYS Mechanical, ANSYS Fluent, ANSYS CFX, LS-DYNA, MSC Nastran, Schlumberger ECLIPSE, Simulia Abaqus, NCBI Blast, NWChem, ClustalW, and HMMER. By configuring these templates based on the application settings in your environment, users can start using the cluster without writing scripts.

Scripting Guidelines – Cluster users that utilize homegrown or open-source applications, can utilize the Platform HPC scripting guidelines. These user-friendly interfaces help minimize job submission errors. They are also self-documenting, enabling users to create their own job submission templates.

Benchmark tests – Platform HPC also includes standard benchmark tests to ensure that your cluster will deliver the best performance without manual tuning.

FLEXIBLE OS PROVISIONING

Platform HPC can deploy multiple operating system versions concurrently on the same cluster and, based on job resource requirements, dynamically boot the Linux or Windows operating system required to run the job. Administrators can also use a web interface to manually switch nodes to the required OS to meet application demands, providing them with the flexibility to support special requests and accommodate unanticipated changes. Rather than being an extracost item as it is with other HPC management suites, this capability is included as a core feature of Platform HPC.

COMMERCIAL SERVICE AND SUPPORT

Certified cluster configurations – Platform HPC is tested and certified on all partner hardware platforms. By qualifying each platform individually and providing vendor-specific software with optimized libraries and drivers that take maximum advantage of unique hardware features, Platform Computing has essentially done the integration work in advance. As a result, clusters can be deployed quickly and predictably with minimal effort. As a testament to this, Platform HPC is certified under the Intel Cluster Ready program.

Enterprise class service and support – Widely regarded as having the best HPC support organization in the business, Platform Computing is uniquely able to support an integrated HPC platform. Because support personnel have direct access to the Platform HPC developers, Platform Computing is able to offer a higher level of support and ensure that any problems encountered are resolved quickly and efficiently.

SUMMARY

Platform HPC is the ideal solution for deploying and managing state of the art HPC clusters. It makes cluster management simple, enabling analysts, engineers and scientists from organizations of any size to easily exploit the power of Linux clusters. Unlike other HPC solutions that address only parts of the HPC

ENTERPRISE-READY CLUSTER & WORKLOAD MANAGEMENT

IBM PLATFORM HPC

management challenge, Platform HPC uniquely addresses all aspects of cluster and management including:

- Easy-to-use cluster provisioning and management
- User-friendly, topology aware workload management
- Unified management interface
- Robust commercial MPI library
- Integrated application support
- Flexible OS provisioning
- Commercial HPC service and support

By providing simplified management over the entire lifecycle of a cluster, Platform HPC has a direct and positive impact on productivity while helping to reduce complexity and cost. The comprehensive web-based management interface, and features like repository snapshots and the ability to update software packages on the fly means that state-of-the-art HPC clusters can be provisioned and managed even by administrators with little or no Linux administration experience.

Price Modelling

Automotive

Big Data Analytics

CAE

Engineering

CAPABILITY / FEATURE	PLATFORM HPC
Cluster Provisioning and Management	x
Initial cluster provisioning	x
Multiple provisioning methods	x
Web-based cluster management	x
Node updates with no re-boot	x
Repository snapshots	x
Flexible node templates	x
Multiple OS and OS versions	x
Workload Management & Application Integration	x
Integrated workload management	x
HPC libraries & toolsets	x
NVIDIA CUDA SDK support	x
Web-based job management	x
Web-based job data management	x
Multi-boot based on workload	x
Advanced parallel job management	x
Commercial application integrations	x
MPI Libraries	x
Commercial grade MPI	x
Workload and system monitoring, reporting and correlation	x
Workload monitoring	x
Workload reporting	x
System monitoring & reporting	x
Workload and system load correlation	x
Integration with 3rd party management tools	x

ENTERPRISE-READY CLUSTER & WORKLOAD MANAGEMENT

IBM PLATFORM MPI 8.1

IBM PLATFORM MPI 8.1

BENEFITS

- Superior application performance
- Reduced development and support costs
- Faster time-to-market
- The industry's best technical support

FEATURES

- Supports the widest range of hardware, networks and operating systems
- Distributed by over 30 leading commercial software vendors
- Change interconnects or libraries with no need to re-compile
- Seamless compatibility across Windows® and Linux® environments
- Ensures a production quality implementation

IDEAL FOR:

- Enterprises that develop or deploy parallelized software applications on HPC clusters

- Commercial software vendors wanting to improve applications performance over the widest range of computer hardware, interconnects and operating systems

THE STANDARD FOR SCALABLE, PARALLEL APPLICATIONS

Platform MPI is a high performance, production–quality implementation of the Message Passing Interface (MPI). It is widely used in the high performance computing (HPC) industry and is considered the de facto standard for developing scalable, parallel applications.

Platform MPI maintains full backward compatibility with HP-MPI and Platform MPI applications and incorporates advanced CPU affinity features, dynamic selection of interface libraries, superior workload manger integrations and improved performance and scalability.

Platform MPI supports the broadest range of industry standard platforms, interconnects and operating systems helping ensure that your parallel applications can run anywhere.

FOCUS ON PORTABILITY

Platform MPI allows developers to build a single executable that transparently leverages the performance features of any type of interconnect, thereby providing applications with optimal latency and bandwidth for each protocol. This reduces development effort, and enables applications to use the “latest and greatest” technologies on Linux or Microsoft Windows without the need to re-compile and re-link applications.

Platform MPI is optimized for both distributed (DMP) and shared memory (SMP) environments and provides a variety of flexible CPU binding strategies for processes and threads, enabling better performance on multi–core environments. With this capability memory and cache conflicts are managed by more intelligently distributing the load among multiple cores.

With support for Windows HPC Server 2008 and the Microsoft job scheduler, as well as other Microsoft operating environments, Platform MPI allows developers targeting Windows platforms to enjoy the benefits of a standard portable MPI and avoid proprietary lock-in.

ENTERPRISE-READY CLUSTER & WORKLOAD MANAGEMENT

IBM PLATFORM MPI 8.1

SUPPORTED OPERATING SYSTEMS

- Red Hat Enterprise Linux 4.6, 5.x and 6.x
- SUSE Linux Enterprise Server 10 and 11
- CentOS 5.3
- Microsoft Windows® XP/Vista, Server 2003/Server 200 8/HPC Server 2008, Windows 7

SUPPORTED INTERCONNECTS AND PROTOCOLS

Myrinet (Linux)	GM & MX on X86-64 and Itanium2
InfiniBand (Linux)	OFED, PSM, uDAPL on X86-64 and Itanium2 OFED 1.1, 1.2, 1.3, 1.4, 1.5 SDR, DDR, QDR, ConnectX and ConnextX2 Mellanox FCA
GigE (Linux)	RDMA, uDAPL, TCP/IP
Infiniband (Windows)	WinOF 2.x, IBAL, WSD, SDR, DDR, QDR, ConnectX(2)
GigE (Windows)	TCP/IP on x86-64

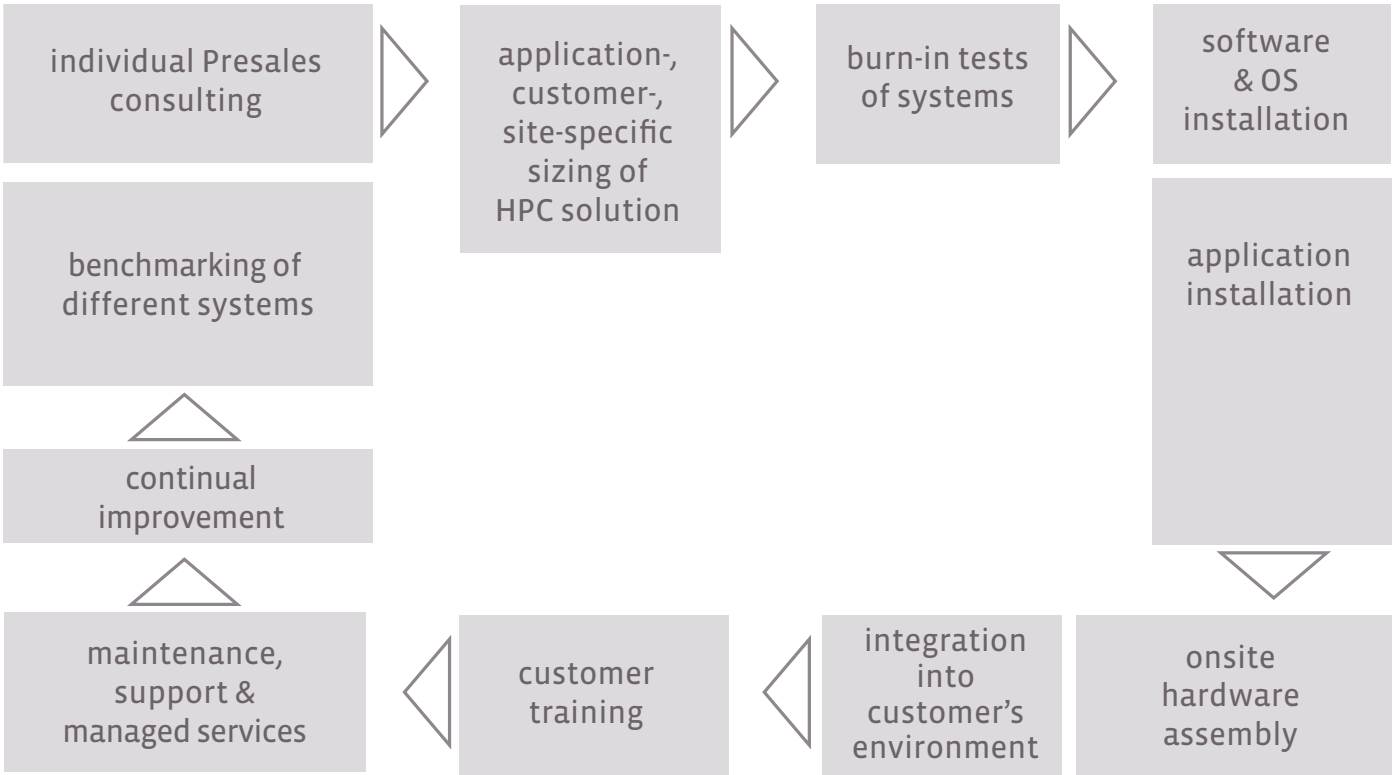
FEATURES AND BENEFITS

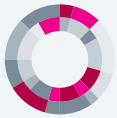
	Features	Benefits
Simplicity	▪ Fully complies with the MPI 2.2 standard, providing dynamic processes, one–sided communications, extended collectives, thread safety, and updated ROMIO ▪ Complete debugging, diagnostic and profiling tools ▪ Auto-detection of interconnects and dynamic loading of libraries ▪ No re-link required for debugging and profiling ▪ Supported by the largest dedicated HPC support organization	▪ Applications port easily to other platforms ▪ Protects ISV software investment ▪ Reduces time-to-market ▪ Increased robustness and quality of applications ▪ Technical problems resolved quickly and efficiently
Performance	▪ Improved shared memory performance, incorporating code and methods from Platform MPI 5.6 (Scali MPI) ▪ 75% reduction in job startup and shutdown at scale ▪ Scalability to 17,000 ranks ▪ RDMA message progression & coalescing enhancements ▪ Flexible CPU binding options maximize cache effectiveness and balance applications to minimize latency ▪ Automated benchmarking of collective operations	▪ Takes maximum advantage of available hardware ▪ Reduced latency for better performance ▪ Performance improves without explicit developer action ▪ Better message throughput in streaming applications ▪ Easier to optimize application performance
Compatibility	▪ Common source-code base between Linux and Windows ▪ Binary compatible with applications developed for HP-MPI ▪ MPICH-2 compatibility mode ▪ Linux Standard Bindings ensure full compatibility across all major Linux distributions ▪ Scheduler agnostic with workload manager integrations for Windows HPC, Platform LSF, PBS Pro, SLURM and other popular schedulers and resource managers	▪ Avoid the cost of separate releases for different platforms ▪ Easily used with existing MPI applications ▪ Common mpirun syntax between Linux and Windows ▪ Customers avoid proprietary “lock-in” ▪ Avoid floating point issues causing inconsistent results
Flexibility	▪ Supports the widest variety of networks and interconnects ▪ Select interconnects at run-time with no need to re-compile ▪ Write applications once and deploy across multiple OS and hardware topologies ▪ CPU binding features well suited to GPU-aware applications	▪ Develop applications that will run on more platforms ▪ Reduce testing, maintenance and support costs ▪ Enjoy strategic flexibility

HIGH PERFORMANCE COMPUTING

PERFORMANCE TURNS INTO PRODUCTIVITY

SERVICES AND CUSTOMER CARE FROM A TO Z





transtec
accelerate productivity

HPC @ TRANSTEC: SERVICES AND CUSTOMER CARE FROM A TO Z

transtec AG has over 30 years of experience in scientific computing and is one of the earliest manufacturers of HPC clusters. For nearly a decade, transtec has delivered highly customized High Performance clusters based on standard components to academic and industry customers across Europe with all the high quality standards and the customer-centric approach that transtec is well known for.

Every transtec HPC solution is more than just a rack full of hardware – it is a comprehensive solution with everything the HPC user, owner, and operator need.

In the early stages of any customer's HPC project, transtec experts provide extensive and detailed consulting to the customer – they benefit from expertise and experience. Consulting is followed by benchmarking of different systems with either specifically crafted customer code or generally accepted benchmarking routines; this aids customers in sizing and devising the optimal and detailed HPC configuration.

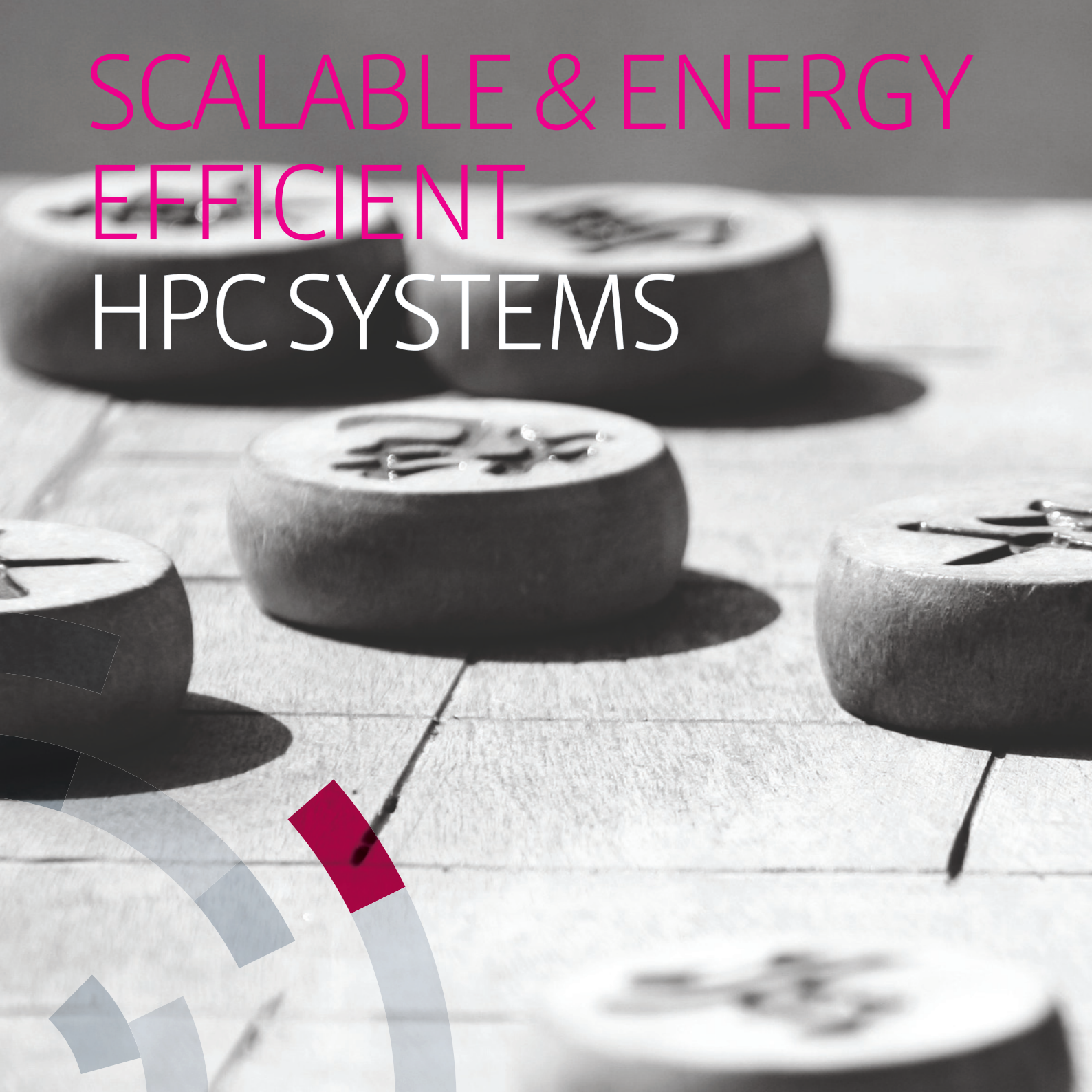
Each and every piece of HPC hardware that leaves our factory undergoes a burn-in procedure of 24 hours or more if necessary. We make sure that any hardware shipped meets our and our customers' quality requirements. transtec HPC solutions are turnkey solutions. By default, a transtec HPC cluster has everything installed and configured – from hardware and operating system to important middleware components like cluster management or developer tools and the customer's production applications. Onsite delivery means onsite integration into the customer's production environment, be it establishing network connectivity to the corporate network, or setting up software and configuration parts.

transtec HPC clusters are ready-to-run systems – we deliver, you turn the key, the system delivers high performance. Every HPC project entails transfer to production: IT operation processes and policies apply to the new HPC system. Effectively, IT personnel is trained hands-on, introduced to hardware components and software, with all operational aspects of configuration management.

transtec services do not stop when the implementation projects ends. Beyond transfer to production, transtec takes care. transtec offers a variety of support and service options, tailored to the customer's needs. When you are in need of a new installation, a major reconfiguration or an update of your solution – transtec is able to support your staff and, if you lack the resources for maintaining the cluster yourself, maintain the HPC solution for you. From Professional Services to Managed Services for daily operations and required service levels, transtec will be your complete HPC service and solution provider. transtec's high standards of performance, reliability and dependability assure your productivity and complete satisfaction.

transtec's offerings of HPC Managed Services offer customers the possibility of having the complete management and administration of the HPC cluster managed by transtec service specialists, in an ITIL compliant way. Moreover, transtec's HPC on Demand services help provide access to HPC resources whenever they need them, for example, because they do not have the possibility of owning and running an HPC cluster themselves, due to lacking infrastructure, know-how, or admin staff.

SCALABLE & ENERGY EFFICIENT HPC SYSTEMS

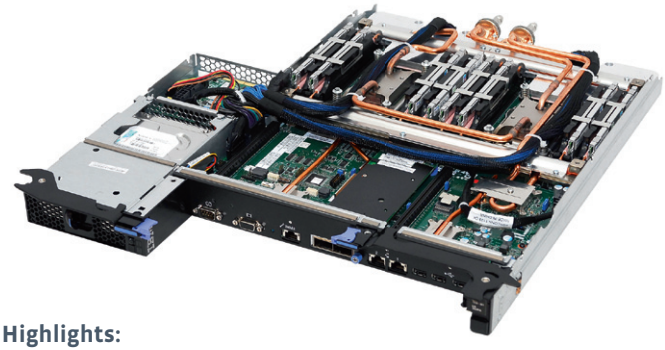


Data Centers today increasingly face challenges to keep pace with growth and varying business demands that require rapid scaling and technology evolution. While each technology refresh cycle delivers a significant performance boost at roughly the same price, these data centers face rising facilities costs and mounting energy bills for power and cooling. The operational costs for labor and the ongoing equipment maintenance costs are other key components of the Data Center Total Cost of Ownership (TCO). In order to get the most out of their current and future IT investments, Data Center managers must minimize TCO while achieving desired efficiencies, computing performance and effective manageability of massive scale-out environments.

SCALABLE, ENERGY EFFICIENT HPC SYSTEMS

IBM IDATAPLEX DX360 M4 HPC CLUSTERS

IBM SYSTEM X IDATAPLEX DIRECT WATER COOLED DX360 M4



Highlights:

- 40% efficiency advantage over air cooled systems by eliminating 90% of heat per node
- Provides up to 10% power advantage over air-cooled systems
- Two water cooling loops per node to cool processors and memory dimms remaining components air cooled
- 90% heat recovery can be reused for economical purposes, such as warming other buildings and facilities
- Typical operating conditions: Ambient air inlet @ 27-35°C and water inlet @ 40-45°C to the nodes

INTRODUCTION

As a leading provider of High Performance Computing (HPC) solutions across a wide range of application and industry segments over the last few decades, IBM has constantly innovated and helped alleviate customer pain points through its rich portfolio for HPC. The new system design of IBM's iDataPlex dx360 M4 servers leverages Intel's industry-standard highly scalable, powerful processors to address energy efficiency and significantly reduce carbon emission footprints of next generation data centers. Intel's Xeon processor E5-2600 product family using 32nm Sandy Bridge processors power the IBM iDataPlex dx360 M4 air-cooled half depth servers to provide more rack density per square feet of floor space in the data center enabling higher compute power for less energy per square foot of data center floor-space.

Overcoming Data Center Challenges

Business users, engineers and scientists attempting to solve challenging problems in engineering, financial markets, media, public sector, oil & gas and the life and earth sciences, rely heavily on HPC systems. These users are some of the most demanding IT solution clients. Their insatiable demand for computational performance continues to drive HPC platforms towards increasing system scalability, larger computing density, lower power consumption, efficient system cooling, easier and more reliable system management, and adaptable programming environments.

Multiple industry studies indicate that revenue from HPC servers will continue to grow much faster than overall server revenues with clusters being the dominant platform – over

75 percent of HPC servers. The battle for HPC leadership will become increasingly global and strategic. More real-world applications will run at trans-petaflop speeds. However, these studies also suggest that the escalating costs of server management, power and cooling, reliability, availability and serviceability (RAS) of HPC clusters with thousands of racks and facilities management at HPC data centers will far outpace the costs of buying new servers. This has caused a severe crisis in HPC data centers. IT solution providers such as IBM have responded by designing innovative solutions that address these issues while retaining all the attractive attributes of industry-standard cluster architectures for HPC.

HPC Cluster Differentiators: iDataPlex Innovations in Packaging, Power and Cooling

IBM's iDataPlex dx360 M4 server can significantly reduce data center costs arising from power consumption. It also helps reduce floor space requirements for cold aisles through innovations in cooling technologies and smart system designs that require less floor space. Data centers can deploy the next generation dx360 M4 with Intel Xeon processors E5-2600 product family and obtain more computing density while consuming the same or lower power. They also benefit from new packaging and cooling methods – traditionally used in very large enterprise class systems and proven very effective for designing energy-efficient clusters where water is pumped through the rack and into node-level cooling loops, which will provide data centers with un-paralleled advantages.

The iDataPlex dx360 M4 has a hybrid air and direct water-cooled design. The direct water-cooled solution is innovative node level cooling that pumps water through the rack and into node-level cooling loops. This water-cooled design eliminates the need for fans in a 2U chassis and significantly boosts energy efficiency. This allows operations that are more reliable and better placement of large clusters at HPC data centers – significantly reducing the overall power, cooling and facilities management costs for the entire data center.

iDataPlex: Scalable, Energy Efficient and Economical HPC Workhorse

With direct water-cooling technology, the iDataPlex can help lower power consumption by up to 40% and can greatly reduce the air conditioning need in data centers. Another advantage is reuse of 90% of the heat removed using direct water-cooling technology from the data center for community purposes such as heating buildings or water. The direct effect of this innovative architecture is that it economically increases the compute density in the data center by a factor of five while retaining many of the attractive attributes of current integrated rack cluster solutions from IBM. As compared to air-cooling, direct water-cooling technology also consumes much less energy.

IBM Continues Active Collaboration with HPC Developers and Partners

Since the iDataPlex can provide a large scale-out cluster using industry standard components, the large portfolio of HPC applications that have been mapped, migrated, and optimized for the IBM Cluster architecture will benefit greatly from this increased density, scalability, performance and cost-effectiveness. This helps to protect customer's prior investment in application enablement and related. The iDataPlex provides an efficient, reliable and cost-effective building block for supercomputing HPC clusters.

SCALABLE, ENERGY EFFICIENT HPC SYSTEMS

IBM IDATAPLEX DX360 M4 HPC CLUSTERS

IDATAPLEX: FOR HIGH PERFORMANCE AND A FLEXIBLE DATA CENTER

To address the new data center challenges, a new way of designing data centers and the server infrastructure that goes into the data centers is essential. The design must encompass data center power, cooling, management, and acquisition as well as operating costs as the chief design goals, in addition to the performance, cost, and scalability requirements at the server level. While IBM has years of experience designing server technologies for scale-up and scale-out that primarily focus on performance and scalability as the fundamental requirements, the iDataPlex solution focuses on a different set of goals:

- Reduce the initial hardware acquisition costs and on-going maintenance costs for data center owners
- Improve efficiency in power consumption
- Eliminate data center cooling requirements
- Achieve higher server density within the same footprint as the traditional rack layout
- Simplify manageability for massive scale-out environments
- Reduce the time to deployment through pre-configuration and full integration at manufacturing

As is evident, these design goals go far beyond a single server or a single rack level; they are goals for the entire data center. With this new philosophy and the new design, the iDataPlex solution promises to address the data center challenges at various levels:

- an innovative rack design for higher node density
- a flex node chassis based on industry standard components
- shared power and cooling components to improve node as well as rack level efficiencies
- optional rear-door heat exchangers that virtually eliminates traditional cooling and lastly,
- optional direct water cooling technology that can lower power consumption by up to 40%

THE FOLLOWING FIGURE HIGHLIGHTS EVOLUTION OF COOLING TECHNOLOGIES IN IBM'S IDATAPLEX SOLUTION

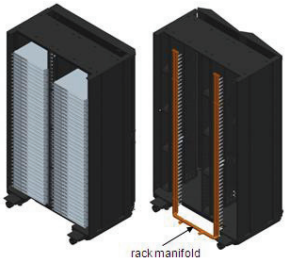
TODAY

Cooling at Rack Level with Rear Door Heat Exchanger

&

TOMORROW

Cooling at Node Level with Direct Water Cooling



- Increase data center density by eliminating hot/cold aisles
- Eliminate rack heat exhaust
- Same dimensions as standard iDataPlex rear door 4" deep
- Liquid cooling at the rack is up to 95% more efficient than air cooling by a CRAC
- No electrical or moving parts
- No condensation
- Chilled water

- Node-level cooling loops cool processors and memory, all other components air cooled
- Water manifolds for intra-rack distribution
- 40% efficiency advantage over air cooled servers
- 90% heat recover per node
- Heat removed from the datacenter can be reused for community purposes (ex. heating buildings or water)

SCALABLE, ENERGY EFFICIENT HPC SYSTEMS

IBM IDATAPLEX DX360 M4 HPC CLUSTERS

iDataPlex Architecture: Cost Effective, Flexible and Scalable

With major cooling innovations, the direct water-cooled iDataPlex promises to be a unique offering for HPC data centers. IBM's enterprise experience and traditional HPC needs combine to create a low cost platform with no redundant hardware, easily available industry standard parts, extreme compute density, increased performance/watt, and quick and large-scale deployment. With a more efficient standard form factor, the iDataPlex combines servers, chassis, switches, PDUs, management appliances, and efficient rear door heat exchanger and latest direct water-cooling technology to address many HPC needs such as flexible configurations for nodes, storage and I/O capability including power-optimized architectures.

With the Intel Xeon processor E5-2600 product family powered iDataPlex M4, Petascale HPC clusters and data centers can achieve over 55% reduction in overall TCO and as much as 90% savings in power and cooling through innovations in air cooling, rear door heat exchanger chassis, energy efficient Intel processors, rack, and optional liquid cooling in an iDataPlex.

iDataPlex Features and Benefits at a Glance

The iDataPlex is a result of a decade of system innovation by IBM and promises to close the loop between low-end inexpensive volume servers and enterprise class SMP and blade servers. This massive, scalable computing workhorse includes a wide-range of flexible options that make it a great choice for HPC applications and other large-scale applications. The figure below depicts the salient system details and summarizes the iDataPlex features and benefits.

BENEFITS	IDATAPLEX FEATURES
Deployable	▪ Racks fully integrated by IBM manufacturing at the factory site ▪ Delivered, installed, and powered up at the Data Center before customer acceptance ▪ Minimal on-site time required – shipping dock to image-load
Serviceable	▪ All front access – eliminates need for accessing rear of rack ▪ Blade-like design with chassis docking into power connector ▪ Innovative cable layout provides highly efficient cable routing ▪ Flexible support options from self maintenance to 24x7x4 response time
Efficient	▪ Lower power consumption by up to 40% per rack ▪ 90% extracted heat can be reused for community purposes (heating buildings) ▪ Liquid cooling at the rack is up to 95% more efficient than air cooling by a CRAC ▪ Energy expenses reduced by up to 40% ▪ Minimize heat exhaust with optional Rear Heat Exchanger
Flexible	▪ Excellent flexibility in node and rack configuration ▪ Factory integrated racks delivered to the Data Center ▪ 3rd party options and rack support
Affordable	▪ Shared infrastructure designs cut cost out ▪ Non-redundant components ▪ TCO benefits -- Electricity cost ~90% lower, IT Cap Ex ~60% higher but 55% less overall TCO with iDataPlex.
Manageable	▪ Blade like thinking in design ▪ Scalable system management software for large-scale configurations ▪ Rack management appliance for several racks ▪ Ease in servicing from complete front access for upgrades ▪ RAS advantages

SCALABLE, ENERGY EFFICIENT HPC SYSTEMS

IBM IDATAPLEX DX360 M4 HPC CLUSTERS

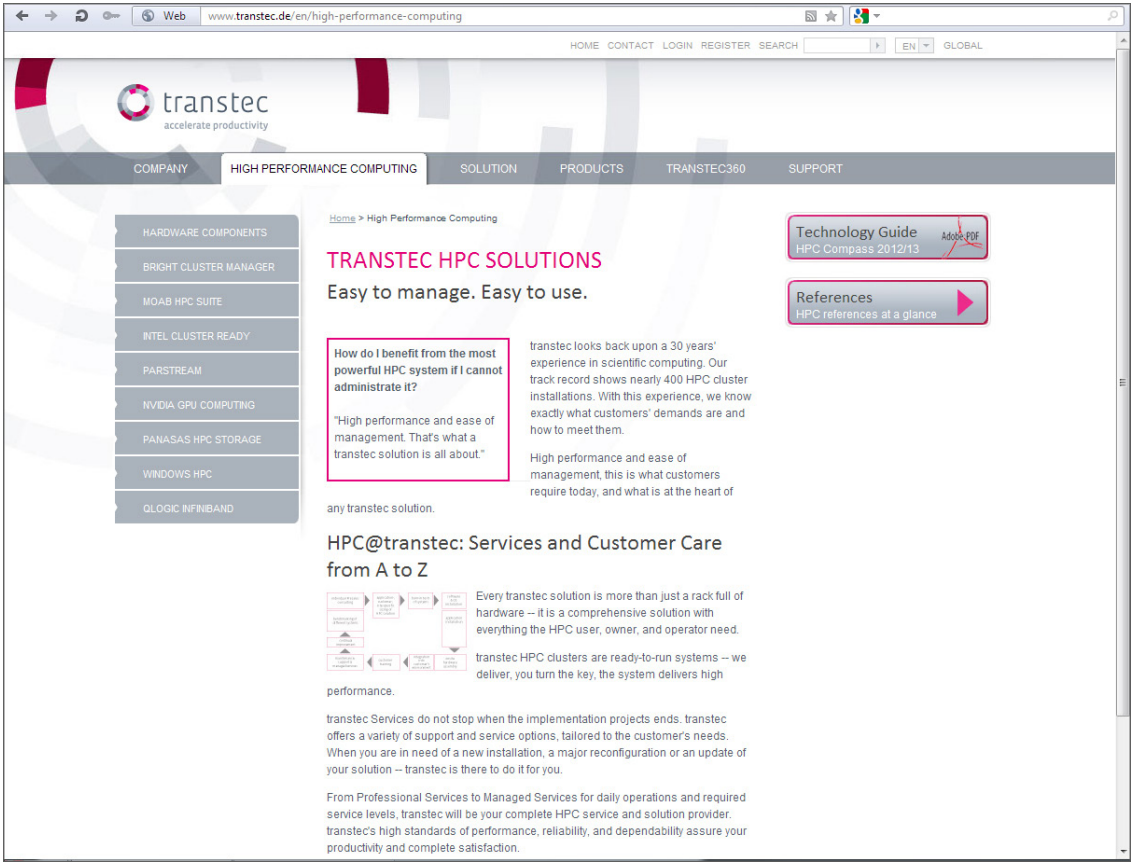
THE IDATAPLEX: TAILORED FOR HPC APPLICATIONS

HPC users have always demanded computing solutions that have the best performance, price/performance, and now are increasingly demanding energy-efficient platforms. The iDataPlex with its economical, large scale-out architecture is an excellent workhorse for a wide-range of HPC applications in engineering, life sciences, upstream petroleum, financial services, and scientific research. Often, these applications do not require hardware redundancy for failure protection, and software fault-tolerance built into the application and/or system software is more than adequate for practical purposes.

Over the last decade, in conjunction with application developers, IBM has made substantial investments to optimize and tune HPC applications to take advantage of the IBM Cluster architecture. Together they have achieved substantial performance tuning by careful load balancing, maximizing single socket performance, maintaining data locality, minimizing cache misses, and maximizing the computation-to-communication ratio. In many cases, application performance is further enhanced by the use of optimized mathematical libraries and high-performance I/O solutions like the IBM General Parallel File System (GPFS). All these investments translate readily to the iDataPlex M4.

NOTES


Risk Analysis
Automotive
CAD
Life Sciences
Price Modelling
High Throughput Computing
Aerospace
Engineering
ice Modelling
CAE
Simulation
Aerospace
High Throughput Computing
CAD
Life Sciences
Risk Analysis
Engineering
Big Data Analytics
Data Analytics
Simulation
CAE
Automotive



ALWAYS KEEP IN TOUCH WITH THE LATEST NEWS

Visit us on the Web!

Here you will find comprehensive information about HPC, IT solutions for the datacenter, services and high-performance, efficient IT systems.

Subscribe to our technology journals, E-News or the transtec newsletter and always stay up to date.

www.transtec.de/en/hpc

transtec
accelerate productivity

Automotive
Risk Analysis
High Throughput Computing
Simulation
CAE
Engineering
Aerospace
Big Data Analytics
Price Modelling
Life Sciences
Risk Analysis
CAD

transtec Germany
Tel +49 (0) 7071/703-400
transtec@transtec.de
www.transtec.de

transtec Switzerland
Tel +41 (0) 44/818 47 00
transtec.ch@transtec.ch
www.transtec.ch

transtec United Kingdom
Tel +44 (0) 1295/756 500
transtec.uk@transtec.co.uk
www.transtec.co.uk

ttec Netherlands
Tel +31 (0) 24 34 34 210
ttec@ttec.nl
www.ttec.nl

transtec France
Tel +33 (0) 3.88.55.16.00
transtec.fr@transtec.fr
www.transtec.fr

Texts and concept: Dr. Oliver Tennert, Director Technology Management & HPC Solutions | Oliver.Tennert@transtec.de
Layout and design: Stefanie Gauger, Graphics & Design | Stefanie.Gauger@transtec.de