AccessHPC

Delivering ready-to-run, cloud enabled HPC

PENGUIN COMPUTING

Contents

Solution-at-a-Glance

Features

- A comprehensive, proven compute architecture for on-prem, cloud-enabled HPC.
- Workload-optimized compute servers.
- Rapid image-based provisioning of bare metal servers.
- High throughput, low latency networking.
- Flexible infrastructure to support heterogeneous compute.
- High performance workload containerization and virtualization.
- Intuitive and centralized cluster and cloud management, reporting, and diagnostics.
- Browser-based, HTML5 remote desktops.
- Enhanced support for secure environments (SELinux with MLS, FIPS 140-2).
- Professional support by HPC, cloud, and cluster management experts.

Benefits

- Jump-start on-prem initiatives with ready-to-run, cloud-enabled enterprise HPC.
- Deliver the power and flexibility of cloud-native portability to your local users.
- Build true hybrid clouds across private and public domains.
- Integrate and deploy apps faster in containers and virtual machines.
- Create a private cloud with public cloud-like scalability.
- Improve the performance, adaptability, and accessibility of your data platforms.
- Accelerate adoption of rapidly changing technologies.
- Customize and manage your HPC and cloud environment with a unified management framework.
- Free up administrators so they can focus on tasks specific to your business.

The Cloud-Enabled HPC Challenge

The HPC landscape has changed dramatically over the past two decades. Technology changes almost daily, but the move of compute environments to the Cloud is one of the biggest changes. Researchers have grown accustomed to the flexibility of the cloud and seek that same ease of use in their on-premises platforms. New entrants to the market are leveraging these new strategies to push the boundaries of the discipline. At the same time, the underlying technology is evolving at an increasingly rapid pace, enhancing the capabilities of organizations that can adapt and improving their competitiveness.

Whether cloud-enabled HPC is a new concept for your organization or you are a seasoned practitioner looking for a performance edge, consider the following when adopting these technological advances:

Platform Complexity: Designing the right on-premises, cloud-enabled HPC platform for targeted workloads is a complex task. Ensuring that the compute, storage, and networking subsystems are well designed individually and function in a balanced manner together is critical. There are many degrees of freedom in the process, and new variables are introduced with each technological advance in these subsystems. A poor design choice at any point in the process can negatively impact performance and reliability, and can significantly reduce the value of any HPC investment.

Cloud Workload Portability: Creating value depends on your ability to deliver applications and workloads faster anywhere, either at the core or in the cloud. Cloud-native technologies like containerization and virtualization are becoming the preferred way to build new software experiences and modernize existing applications, workloads, and workflows at scale from on-premises to public clouds and multi-cloud.

Management, Control, & Customization: The productive output of an HPC platform relies primarily on its architecture and secondarily on the effectiveness of the management layer and how it is customized to an organization's workflows. Tools that automate the basic management functions of an HPC environment in a manner that is conducive to the workflows of the user base are critical. Proper selection, implementation, and ongoing management of this layer boosts productivity for users and administrators alike and frees both to excel the more complex aspects of their work.

Secure Remote Workload Access: In today's environment, you need to be able to securely deliver real-time interactive enterprise class visualization to users throughout the enterprise. Many use cases now require a remote visualization tool that allows you to manage multi-monitor high resolution graphics effectively.

Risk Mitigation: The complexity of HPC platforms can increase points of failure. Sometimes these failures can range from subtle performance degradation that is difficult to identify to catastrophic failures of key elements of the architecture that result in measurable loss of productivity. Fortunately, elements of high availability can be designed into HPC architectures in order to avoid the worst failures.

Hardware Abstraction: HPC users are moving closer to their workloads and farther from the underlying hardware that runs them. This very effective approach to HPC leverages cloud-native technologies to accelerate innovation by allowing domain experts to do what they do best: deliver research results. However, the fact that engineers are less concerned with hardware — combined with the fact that those hardware platforms are increasingly more complex — places an additional burden on research infrastructure teams to bridge that gap. Adopting HPC for the first time, augmenting an existing HPC platform with emerging technologies, or repatriating cloud-based HPC resources all drive a need for new tools and skill sets.

Exponential Data Growth: While HPC users have grappled with the challenges of processing, moving and extracting insight from data for decades, those challenges have been pushed to the fore as our global data production increases. These challenges are further compounded with the move to cloud and now the emergence of powerful hybrid and multi-cloud architectures. Grappling with issues such as data silos, expensive bandwidth, the drive to real-time insight, and a geographically distributed workforce can be daunting without the powerful new tools and innovative strategies required to succeed.

Data Center Infrastructure: HPC platforms have always tested the infrastructure limits of contemporary data centers. That has never been more true than today as we push the envelope with new demands on connectivity, space, cooling, and power densities that reach up to 100kW per rack. The symbiosis between the compute platforms and the data centers that house them call for a unified approach to the complete HPC architecture from the platform outward. Even if a new data center is not in your plans, there are many power and cooling technologies that can enable existing data centers to accommodate the new standards in HPC.

You need a secure, cloud-enabled HPC infrastructure that is optimized for your unique workloads and cloud needs and is engineered by a partner who knows HPC and who knows the cloud. And to manage and customize it all, you need cloud-native tools that are created specifically for administrators by engineers who manage HPC and all types of cloud environments.

Penguin Computing AccessHPC

Penguin Computing AccessHPC™ combines decades of HPC design experience with the best in cloud-native technologies to provide proven, streamlined, cloud-enabled HPC architectures. The platform delivers the flexibility and portability of cloud-native technologies to your on-premises resources for cloud-like ease of use locally.

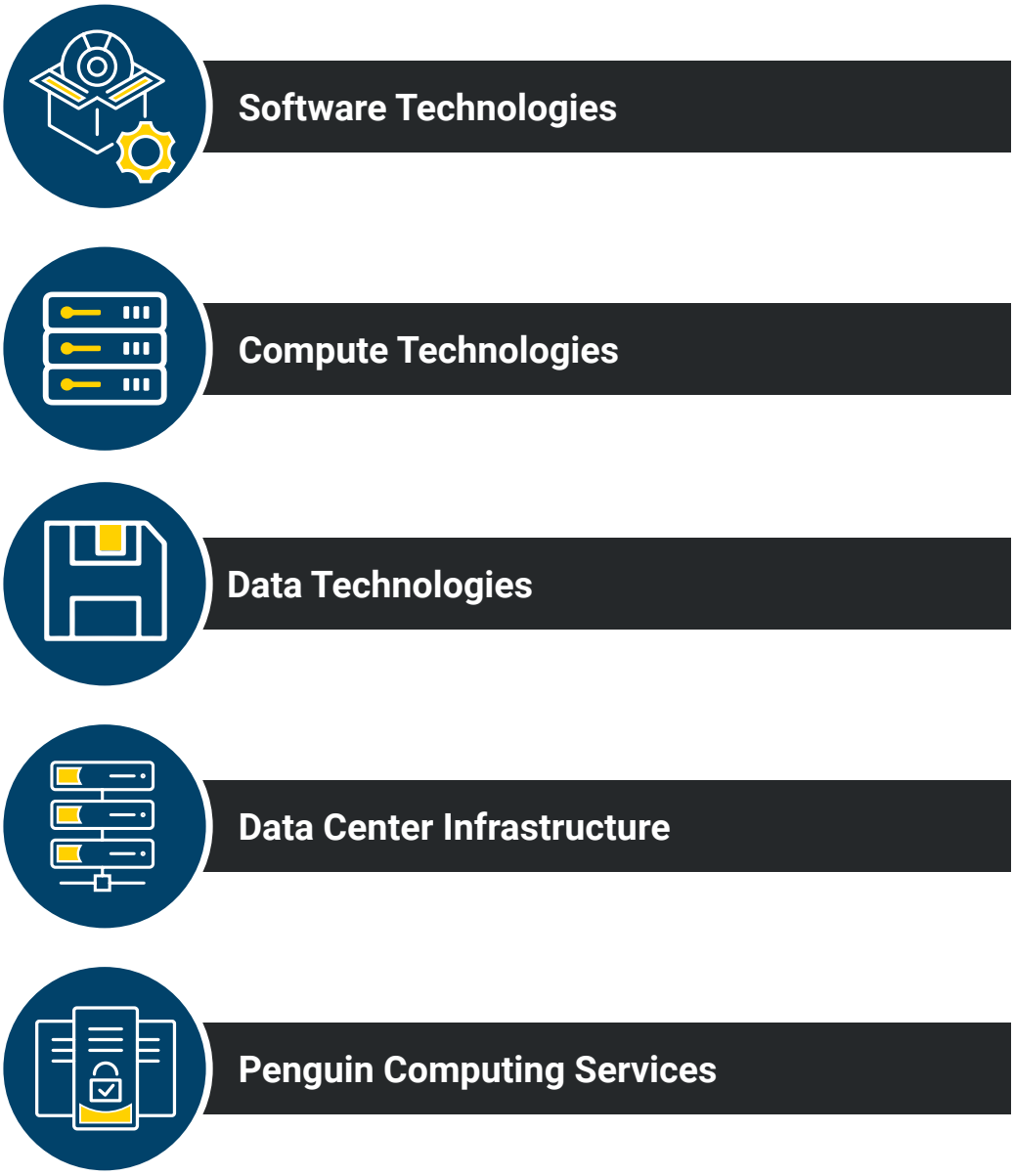
AccessHPC provides a complete software, hardware, and management platform built on our compute-optimized hardware, Red Hat Cloud Suite technologies, and Penguin Computing Scyld cloud and cluster orchestration and management software. This out-of-the-box, cloud-enabled HPC solution also leverages high-performance, low-latency networking and storage technologies to deliver optimized HPC for your workloads.

AccessHPC allows you to get the most performance out of your underlying HPC cluster from day one. With the powerful cloud and cluster management of Scyld ClusterWare® and Scyld Cloud Manager and expert management from the Penguin Computing services teams, your end users can engage with these resources with the same ease and even better performance than they need to keep innovating.

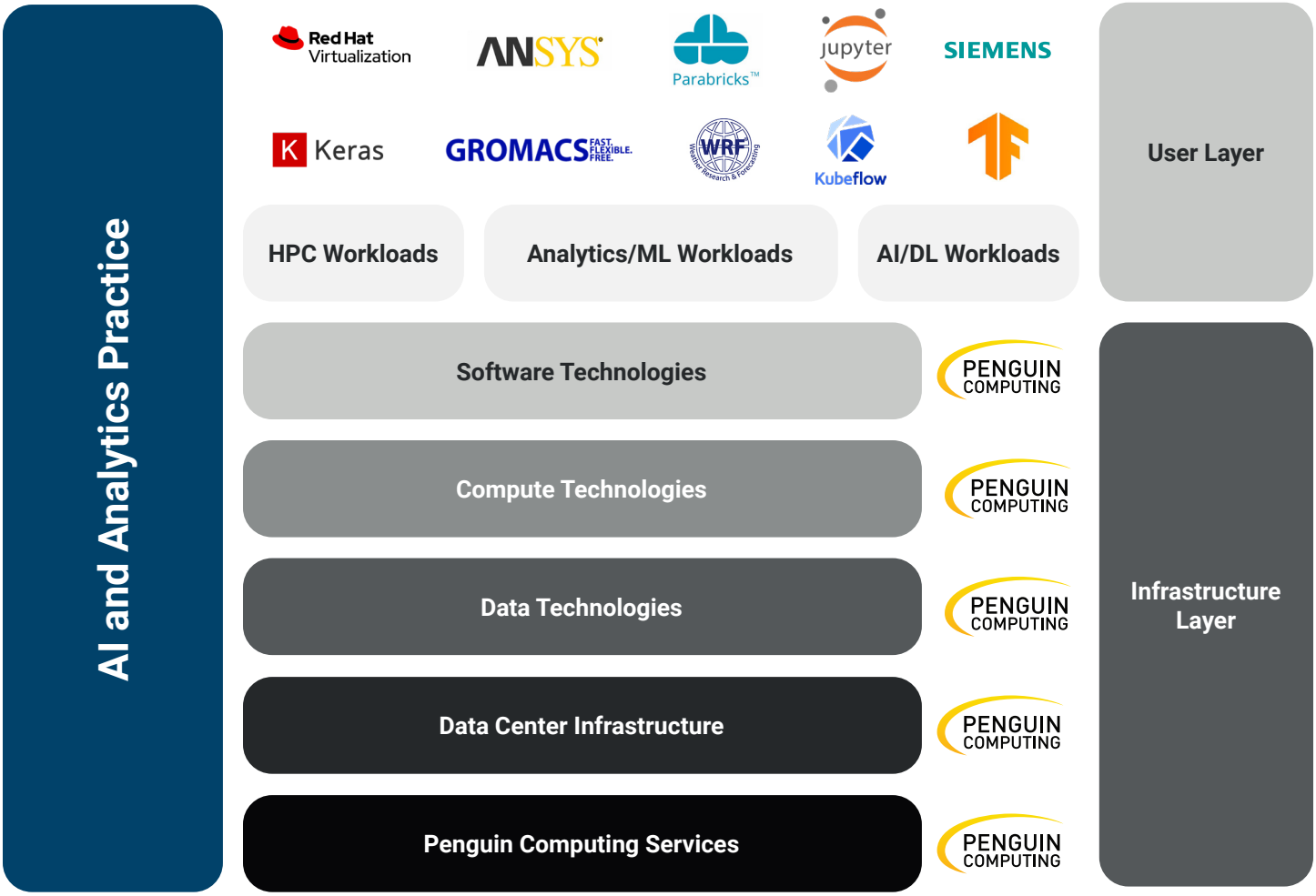
You can combine AccessHPC with other Penguin Computing reference architectures for Data, Cloud, and AI/Analytics. When you combine the benefits from these different technologies, you can easily

build complex, high-performance environments across many facets of your IT infrastructure.

Access HPC includes:



Penguin Computing AccessHPC Components

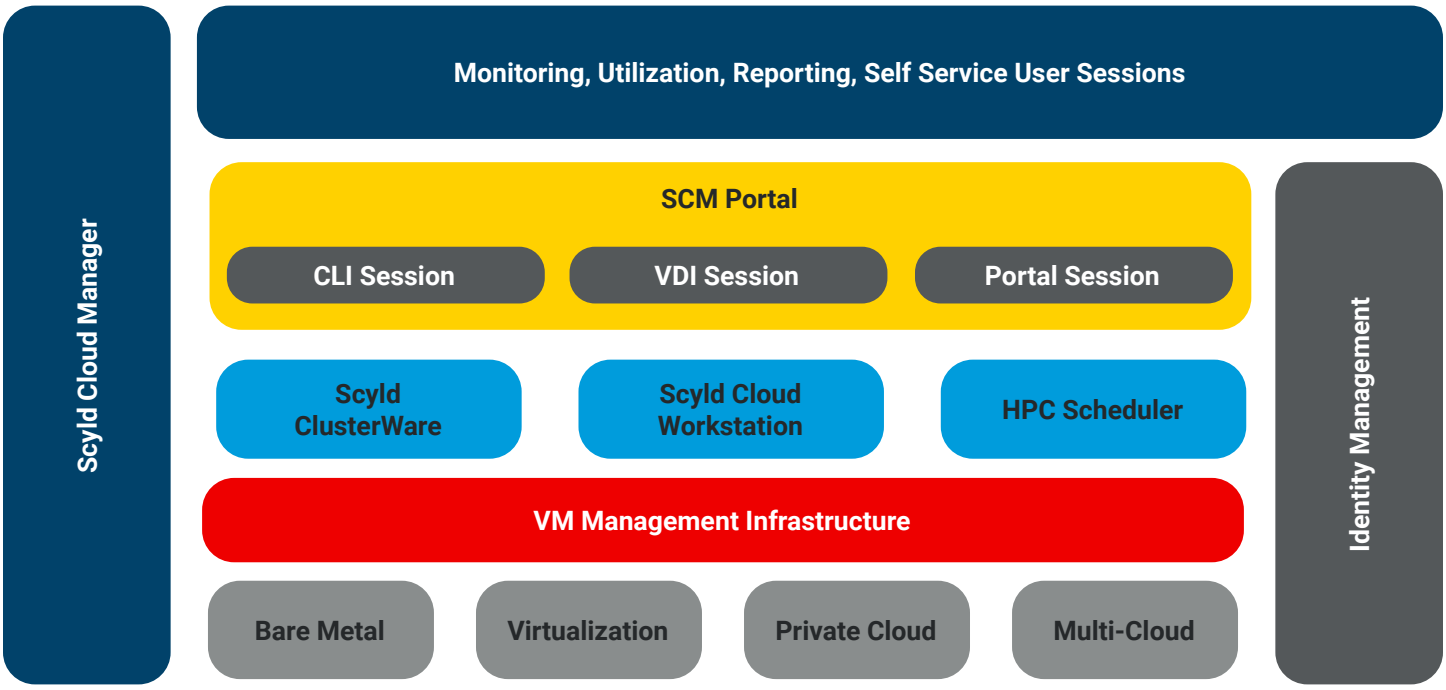


Software Technologies

Scyld Cloud Manager

HPC Cloud Portal

AccessHPC leverages Penguin Computing's Scyld Cloud Manager cloud software suite to enable a complete HPC cloud solution. Scyld Cloud Manager (SCM) provides administrators and users with a single pane of glass for registration, access management, system provisioning, and more. Managers can access and view cluster utilization metrics in a single location, giving them insight into what's happening on their cluster and who's using it, enabling them to plan environment upgrades and expansions for the next year or next three years. In addition to powering AccessHPC, Scyld Cloud Manager also powers Penguin Computing's public HPC cloud: Penguin On-Demand (POD).



Accounting & Chargeback

Many large organizations depend on internal chargeback and tracking models to ensure resources are being consumed by the users that help generate revenue or budget for the environment. AccessHPC provides organizations with the same billing system that Penguin Computing uses to run Penguin Computing On-Demand — extending the same billing capabilities into their own environments. Organizations can track job completion time, job resource utilization, and who ran the job, giving them the ability to bill by user, group, or project.

Virtualization & Bare Metal Computing

AccessHPC provides the best features of a cloud environment with the best features of an on-premises HPC cluster. Head nodes, login nodes, and other administrative nodes are virtualized to enable process redundancy and flexibility across the infrastructure. Compute nodes are provisioned as bare metal servers to ensure HPC jobs run optimally and consistently in the cluster. Users can expect predictable job performance from the cluster, and administrators can expect flexible and rapid provisioning across the environment. AccessHPC is designed to integrate with workload-optimized computing solutions, such as those available in Penguin Computing's TrueHPC™ solution.

Integration & Security

AccessHPC is tested and delivered as an integrated solution with security and support as paramount components. AccessHPC can integrate into existing customer environments, providing the flexibility

needed to integrate with many different networks, identity managers, authentication policies, and end user workflow models. The AccessHPC architecture provides a reliable cloud-enabled foundation with endpoints that are integrated into the private HPC networks and services as well as enterprise networks and services. The architecture is built on a solid cloud-enabled OpenStack framework providing a secure platform to integrate with and strengthen security compliance.

Scyld Cloud Manager Features

- Fully integrated HPC cloud solution for administrators
- Guided self-registration web portal to on-board users onto HPC resources
- Web portal interface allows access rights management and system monitoring
- Live and historical usage reporting for billing and chargeback management
- Web service APIs for reporting, resource management, and access control
- Flexible VM configurations that support remote 3D visualization desktops

Scyld Cloud Workstation

Scyld Cloud Workstation is a remote desktop solution designed to deliver real-time interactive enterprise class visualization through a standard browser without plugins. Users simply connect to the remote environment from virtually any device running Firefox, Internet Explorer, Edge, Chrome, or Safari.

3D Accelerated Engineering VDI

Scyld Cloud Workstation provides engineering class visualization for HPC and AI/ML engineers running graphical applications such as CFD, CAE, or FAE codes. Scientists in weather and chemistry are another class of users that commonly need to visualize data. Traditional HPC and AI/ML environments require users to download large data files to on-premises workstations for pre/post processing, model development, and data analysis offline from the computing resource and centralized storage. This is a time-consuming process that makes it hard to create an efficient workflow with predictable time to results.

Scyld Cloud Workstation offers significant time savings by moving pre- and post-processing to a workstation with direct access to a cluster's data storage, eliminating the need to download large data files. Users can use the same GUI tools as on their local workstations, ensuring continued productivity.

Remote Collaboration on Shared Desktops

Scyld Cloud Workstation enables multi-user collaboration and remote desktop access for up to ten temporary or permanently authorized users. Desktop control can be passed from one user to another while our QoS algorithm intelligently adjusts the frame rate per client to ensure an optimal experience. Through Scyld Cloud Workstation, customers can deploy large scale remote desktop environments using open source or commercial virtual machine platforms for provisioning.

Expansive OS Support

Scyld Cloud Workstation can run in any bare-metal or virtualized environment, enabling you to couple advanced VDI solutions into an existing environment. Whether you are enabling high-end 3D

Scyld Cloud Workstation Features

- Browser-based, HTML5 remote desktops
- No client installation or plug-in needed
- Supported on macOS®, Linux, and Windows®
- Enables 3D accelerated interactive workflows
- Operates in a wide variety of network conditions
- Secure HTTPS authentication from anywhere
- Collaborative, secure multi-user sessions
- Supports multiple monitors
- Intelligent QoS for minimizing network usage
- Dual channel stereo audio support
- Up to 4K, 3840x2160 resolutions at 30fps
- Outperforms traditional VDI

accelerated desktops, deploying GPU-enabled workstations for AI/ML, enabling remote access to software suites for content creators, or building a platform to enable thousands of users with intuitive secure access to a familiar remote desktop, Scyld Cloud Workstation is able to meet your needs.

With support for Linux®, Microsoft Windows®, and Apple macOS®, integration with existing workflows allows for frictionless enablement of your user base into a familiar environment.

Secure Remote Workforce

Scyld Cloud Workstation supplies secure access through HTTPS, requiring no additional ports through the firewall. This unique architecture saves bandwidth, simplifies implementation for IT departments, improves image quality, and ensures near-universal accessibility for users. IT departments can use custom SSL certificates and couple authentication into centralized identity managers through Scyld Cloud Workstation's ability to pass authentication onto the operating system.

Lossless Remote Desktops

For users that require pixel accuracy, Penguin Computing delivers an optional client that delivers two classes of high-quality video outputs at up to 4K UHD:

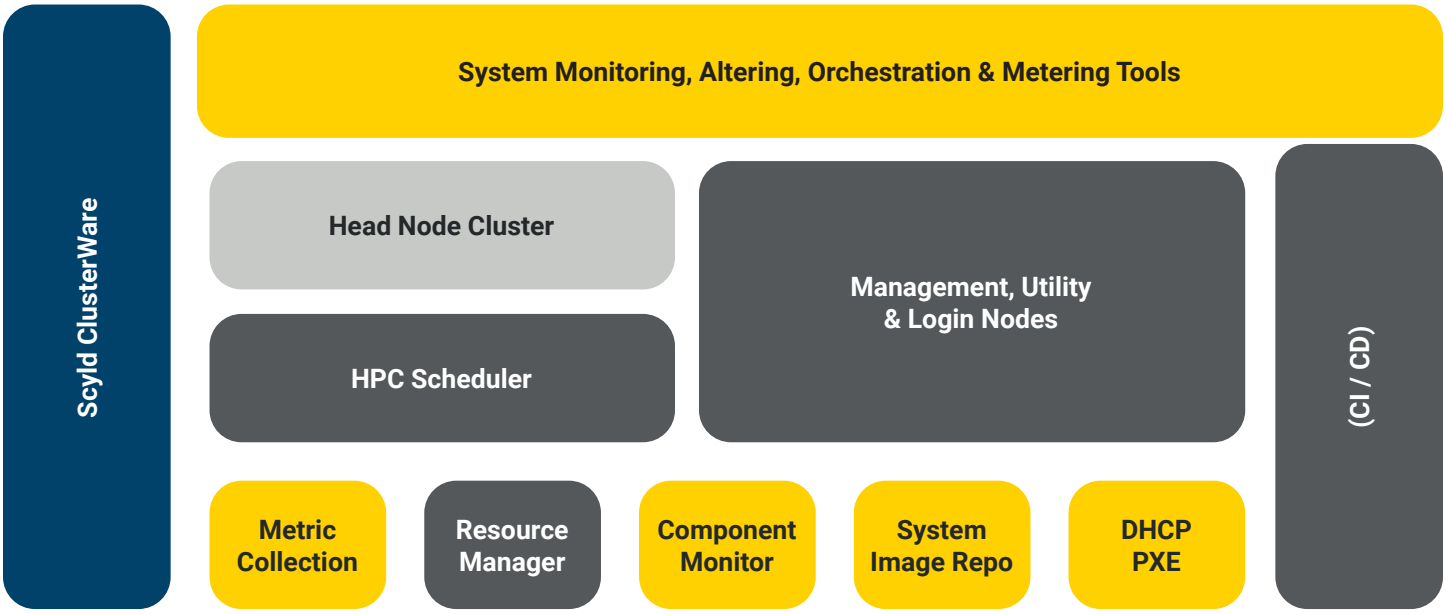
- Visually Lossless provides high fidelity video at near lossless compression to optimize your network.
- Alternatively, Lossless Video delivers uncompressed video streams.

Users can easily toggle between visually lossless and completely lossless interactive sessions.

Scyld ClusterWare

HPC Cluster Orchestration

AccessHPC leverages Penguin Computing's Scyld ClusterWare cluster orchestration software. Scyld ClusterWare provides a complete HPC environment that supports Slurm, OpenPBS, and TORQUE to handle the scheduling and queueing of HPC jobs. Scyld ClusterWare also provisions the hardware to operate as a single, unified cluster by booting compute nodes using PXE, establishing IP address using DHCP, monitoring node health, and collecting metric data across the cluster.

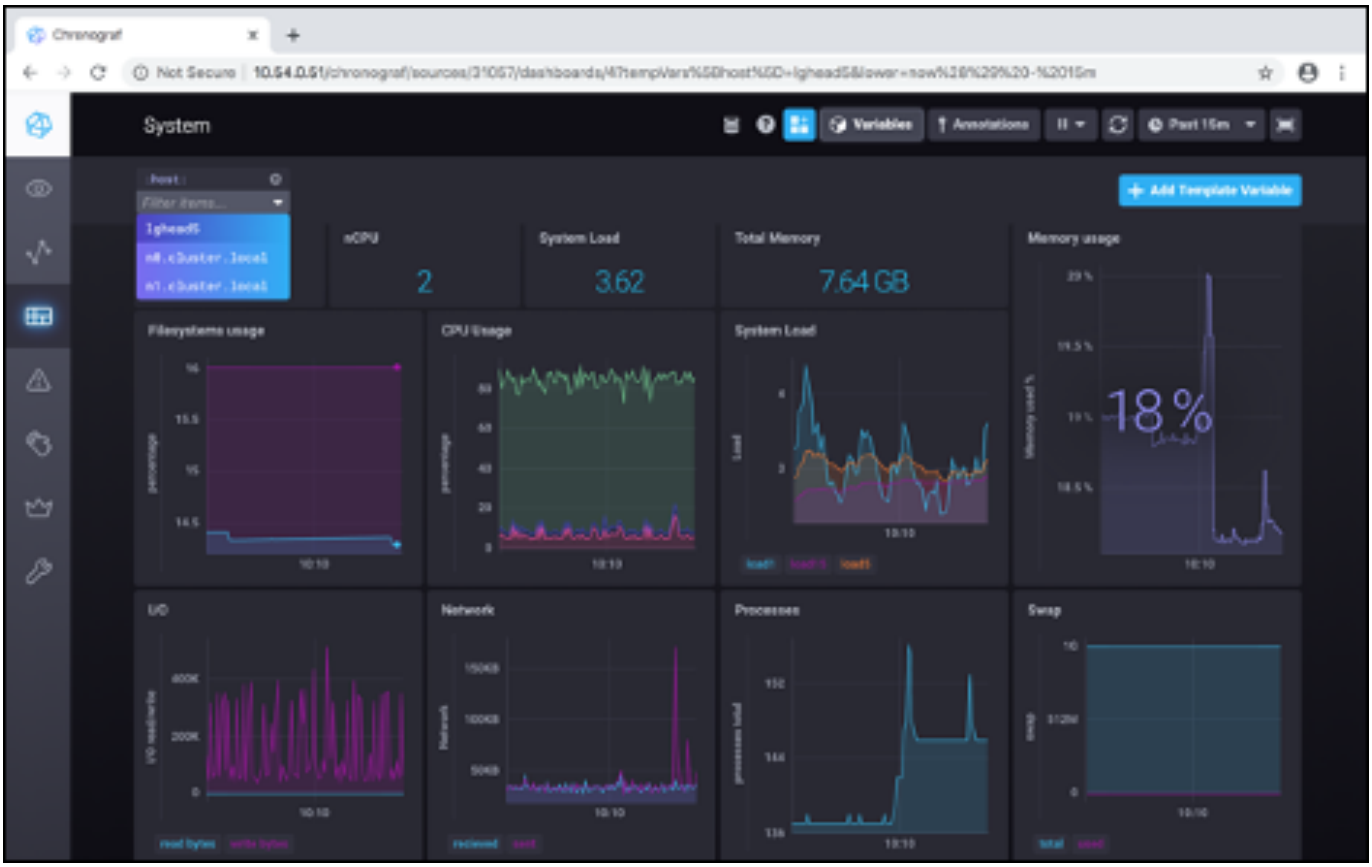


Alerting and Monitoring

Scyld ClusterWare supports alerting features that can integrate into enterprise communication tools, such as email, Slack, PagerDuty, and more to send out important alerts to different groups within an organization regarding the current status of an AccessHPC implementation. This information is also centralized to the head node, providing system administrators with a single pane of glass displaying the status of every node in the cluster.

Cluster Customization

Scyld ClusterWare provides a robust HPC orchestration capability for system administrators to provision and manage an HPC environment. This capability has been designed to provide a familiar, intuitive, and documented method for configuring your cluster and executing change management. Scyld ClusterWare enables experienced Linux administrators to use familiar tools, methodologies, and best practices to customize AccessHPC.



System Image by Application

Scyld ClusterWare supports custom system image deployment. Users can save a compute node image into a repository managed by the Scyld ClusterWare head node. System images can be completely different from the operating system that the head node uses. The head node could be running RHEL 8, while the compute nodes are running RHEL 6, RHEL 7, RHEL 8, Ubuntu 16.04, Ubuntu 18.04, or some combination of operating systems and system versions across the cluster.

Eliminating Compute Silos

Scyld ClusterWare's ability to dynamically deploy system images allows system administrators to centralize their compute infrastructure and eliminate compute silos that would otherwise require a dedicated workstation or compute cluster to support very specific operating systems that some number of user applications would require. The ability to centralize compute resources into one dynamic environment relieves administrative headaches and allows IT teams to focus their efforts on supporting a centralized environment where updates, patches, fixes, and upgrades have a greater impact across their organization.

High Availability

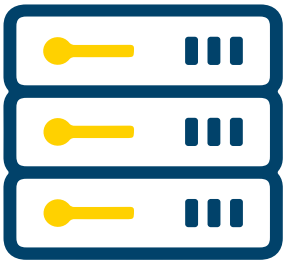
Scyld ClusterWare provides a robust high availability HPC cluster that allows the system to remain operational when unexpected failures occur and during scheduled maintenance windows. The HA feature of Scyld ClusterWare enables system administrators to upgrade a kernel or reboot a head node without disrupting the functional use of the environment to end users.

Enterprise Security

Scyld ClusterWare supports deploying compute nodes with SELinux enforced, and with an MLS policy enabled. Scyld ClusterWare can also be deployed in a manner that meets FIPS 140-2 requirements.

Scyld ClusterWare Features

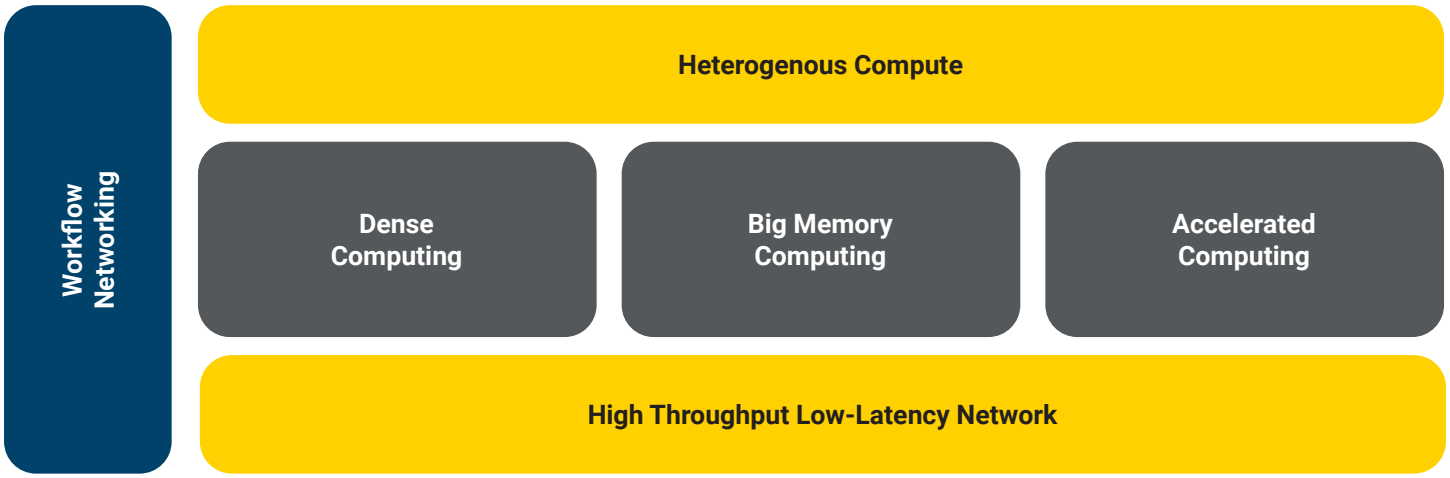
- Rapid provisioning for technical computing environments
- Designed to manage optimized HPC clusters and their coupled enterprise services
- Single source for tested HPC middleware (MPI implementations, HPC schedulers)
- Image based node management facilitates simplified change management
- Flexible provisioning options (for example, diskless, diskfull, network mounted)
- Robust high availability architecture prevents downtime when unexpected failures occur
- Supports SELinux in MLS mode and FIPS 140-2 implementations
- Monitoring GUI for visualizing system telemetry and building custom dashboards
- Notification and alerting integration with email, Slack, and PagerDuty



Compute Technologies

Heterogeneous Compute

AccessHPC leverages industry-leading technologies from Intel, AMD, NVIDIA, and other technology providers to enable a complete technology ecosystem that supports many different workloads. HPC workloads often require high core count, high clock speed, high memory bandwidth, low latency communication, and/or accelerated computing using GPUs, FPGAs, and ASICs. AccessHPC supports heterogeneous computing environments within a single architecture using workload-optimized server building blocks for many types of high performance workloads.



Workload-optimized HPC workloads typically require a large number of cores and high core clock speeds to achieve the best performance possible. These workloads also require high performance interconnects because many HPC workloads span multiple servers, requiring constant node-to-node communication that benefits from high-throughput and low-latency network technologies. Optimized server building blocks for HPC workloads need to provide many cores with high clock speeds and low-latency, high-throughput interconnects to provide the best application performance possible.

Memory-centric workloads call for additional server memory resources to support applications that require extreme read and write performance and extremely low latency. Optimized server building blocks for big memory computing require as much memory bandwidth, capacity, and clock speed as possible to provide the best application performance.

Accelerated computing workloads require enterprise accelerators, such as GPUs, FPGAs, and ASICs, to drastically improve the performance of certain applications. Optimized server building blocks for accelerated computing require in-system, device-to-device communication optimizations to ensure that accelerators can communicate with CPUs, SSDs, NICs, and other accelerators without communication bottlenecks in order to provide the best application performance possible.

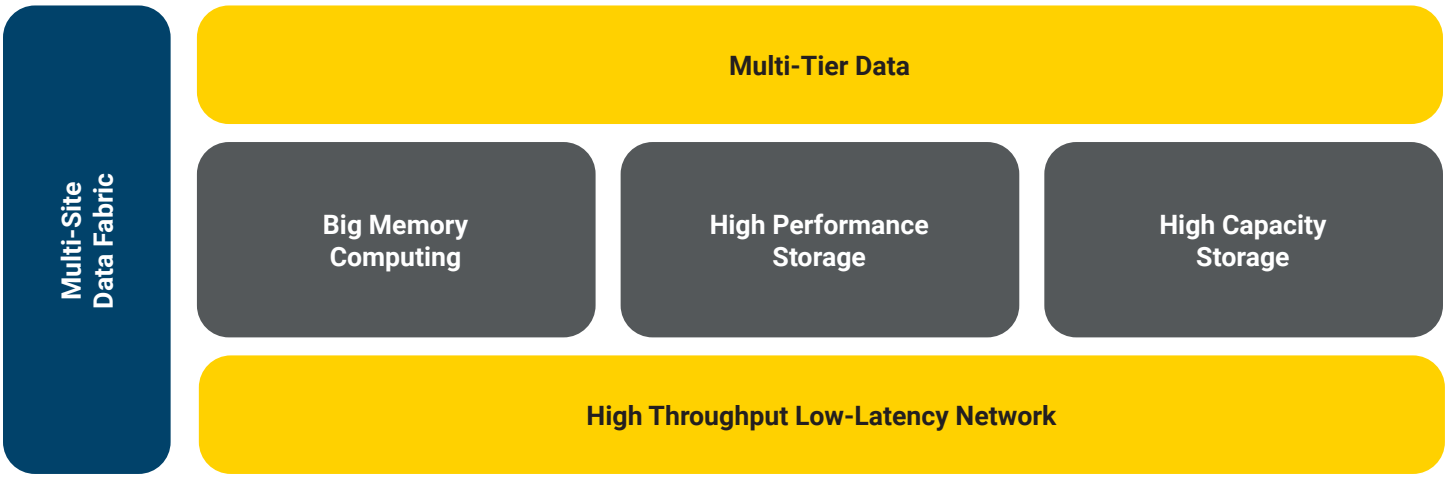
High Speed Low-Latency Interconnects

AccessHPC supports the leading high-throughput, low-latency networking interconnects that help maximize the performance of an HPC cluster for certain workloads.

Data Technologies

HPC is moving toward data-driven workloads that consume and generate large amounts of data. This data growth drives the need for data solutions that can scale to exabyte capacities. HPC environments have data requirements that create data workflow and infrastructure challenges related to management and orchestration.

Data I/O requirements weigh heavily on the overall success of an HPC solution. I/O patterns and performance vary across different tiers of storage in the environment. AccessHPC integrates with the data solutions in Penguin Computing's Data Practice, which cover the entire spectrum of I/O — from memory, to flash, to cold storage — to support the entire data lifecycle.



Big Memory Computing

Some HPC workloads can require massive memory storage or high memory performance. By utilizing the LiveData™ solution, AccessHPC can support memory-centric workloads that require high memory capacity, persistent memory, and high memory tier performance.

High Performance Storage

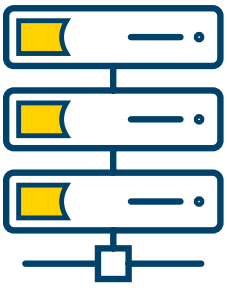
HPC workloads require high-performance storage that can fulfill the ingress and egress demands of high performance workloads. AccessHPC can be paired with the ActiveData™ solution to provide high performance storage for data-heavy computing workloads. Just as with the AccessHPC, the ActiveData solution leverages the industry-leading technologies that are tuned for specific customer workloads.

High Capacity Storage

HPC workloads often ingest or output massive amounts of data that must be kept in a general purpose storage environment when not being used for computing. AccessHPC can be paired with the DeepData™ solution to provide scale-out, capacity-optimized storage tier best suited for storing long-term data.

Multi-Site Data Fabric

Some HPC environments require connectivity to the cloud or another site. Some workloads might require cross-organizational collaboration on datasets that span multiple locations. AccessHPC solution can access data sets across the world as if they were local using the DataNexus™ solution.



Data Center Infrastructure

AccessHPC can be built using both a traditional 19" rack platform and a modern 21" OCP (Open Compute Project) platform. Traditional 19" rack infrastructures are supported in almost every data center worldwide and in a variety of dimensions. Modern 21" OCP rack infrastructures require data centers that can support the most demanding physical and power densities. Penguin Computing has partnered with leading data center facility pioneers who can support the demanding characteristics of today's HPC platforms.

Power

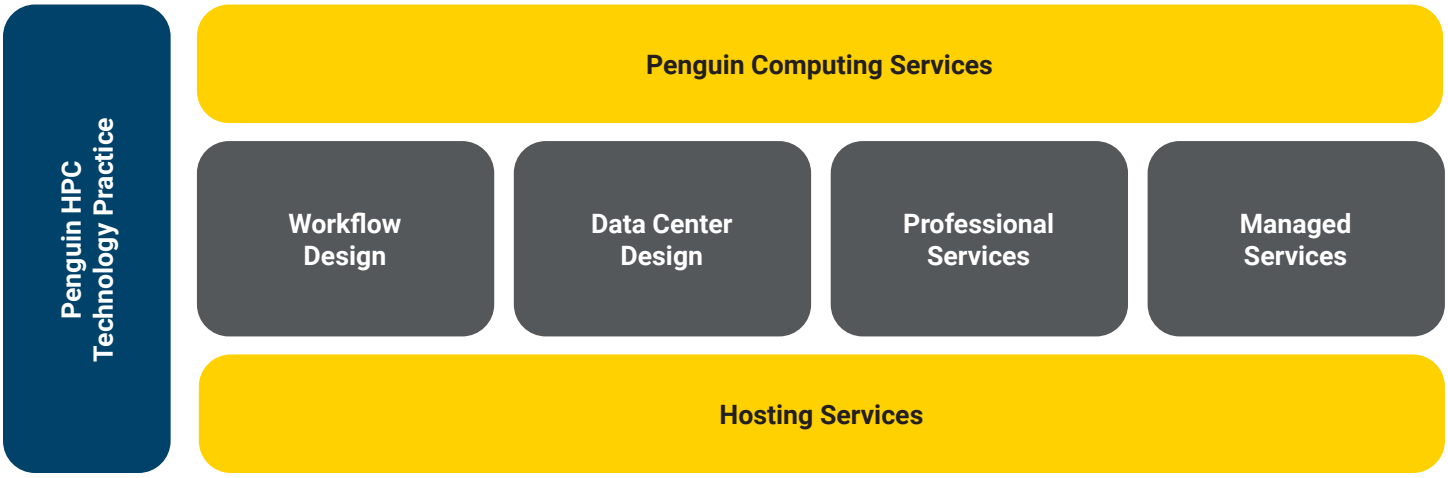
AccessHPC supports three-phase 50A or 60A, 208V, 277V, or 480V power options as well as A+B fully redundant power, or N+1 redundant power. 21" OCP also supports 12V or 48V power delivered directly to the servers, which enable much higher power density per rack.

Cooling

AccessHPC can be air cooled with traditional HVAC equipment. Penguin Computing recommends using a combination of air cooling and liquid cooling when deploying AccessHPC into a data center not designed for high-power equipment. Rear Door Heat Exchangers capture hot air exhaust at the rear of the rack, and can be deployed on most 19" and 21" rack infrastructures. AccessHPC is also designed to integrate Direct-To-Chip cooling options that capture heat directly from the CPU block. This cooling solution removes 85% of server heat before it's transferred into the air, and can be used in select 21" infrastructures.

Penguin Computing Services

AccessHPC is a comprehensive, end-to-end solution that organizations can leverage to jump-start their HPC initiatives. In some cases, the solution will directly meet the needs of the organization, right out of the box. However, most often there will be additional design, deployment, integration, and hosting considerations that need to be addressed.



Penguin Computing provides services that consider rack and floor space, how to scale the environment, maximum rack power consumption, power phase balance, efficient heat removal, and the optimal networking topologies when using low-latency, high throughput interconnects.

AccessHPC is supported by Penguin Computing engineering services, including Design Services, Professional Services, Managed Services, and Hosting Services.

Data center hosting services are offered through Penguin Computing's strong partnerships with data center service providers. Our partners can provide the space, power, and cooling AccessHPC needs — as a service.

Design Services

Workflow Design

- Software Orchestration
- Compute Performance
- Multi-Node Communication
- Data Storage and Data Tiering
- Data Ingest and Egest
- Environment Sizing

Data Center Design

- Rack and Floor Space
- Environment Scalability
- Maximum Power Consumption
- Power Phase Balance
- Efficient Cooling and Heat Removal
- Optimal Networking Topologies

Professional Services

Stand Up and Initialization

- System Burn-In Testing
- Racking and Cabling
- Software Installation & Tuning
- On-Site Deployment and Integration

Hosting Services

Data Center Hosting

- Penguin Data Center
- Customer Data Center
- Power, Space, and Cooling Management
- Monthly or Annual Billing (As-A-Service)

Managed Services

System Administration:

- Complete Hands-Off Experience
- Augment Existing IT Capabilities
- Collaborate with Penguin Support
- Tens to Thousands of Servers
- Terabytes to Exabytes of Data
- Multi Data Center Support

Conclusion

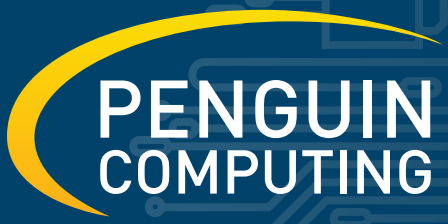
Penguin Computing AccessHPC provides a single, secure, complete end-to-end reference architecture for private or hybrid cloud HPC that includes a flexible, scalable, workload-optimized compute infrastructure and cluster management tools that allow you to not only monitor your clusters, users, and workload utilization, but easily provide a self-service private HPC cloud.

The Penguin Computing AccessHPC solution frees organizations from having to focus valuable time and human resources on creating a cloud-enabled HPC architecture and cluster management solution from scratch, allowing them to lower TCO, reduce risk, and accelerate time-to-innovation.

Penguin Computing can apply our decades of experience to create quality, integrated solutions for our clients. We offer a wide range of professional and managed services that can quickly bring your HPC initiatives to production.

Contact Us

Use this **form** or call Penguin Computing today at 1-888-736-4846 to find out how you can deploy ready-to-run, cloud-enabled HPC.



Expanding the world's vision of what is possible