

Transform your business with high-performance computing

Executive summary

High-performance computing (HPC) and supercomputing trace their history back as far as the 1920s. In the early modern era, HPC was used by monolithic supercomputers, such as those produced by Cray Research beginning in the 1970s. This technology was largely limited to national research laboratories due to the high cost and need for specialized facilities and expertise.

By the 1990s, computers built from industry-standard components began to emerge alongside proprietary machines. This trend marked the beginning of HPC based on open standards hardware and open source software, which continues to provide HPC capabilities for mainstream organizations.

HPC is now applied to a wide scope of problems – and across industries – due to its ability to solve large-scale computational problems within reasonable time and cost parameters. Greater compute power or efficiency allows the same problem to be completed more quickly, allowing more iterations of a simulation – and generating better results.

In the coming decade, the distinctions between HPC and general-purpose enterprise computing will continue to dissipate. In place of monolithic, on-premise supercomputers, the industry will continue to move toward more flexible infrastructure based on open standards, with workloads decoupled from specific hardware.

Technologies associated with mainstream enterprises, as opposed to proprietary, special-purpose technologies, will make this trend possible. In particular, open source approaches to hybrid cloud and virtualization using both containers and hypervisors will deliver economic advantages as organizations apply HPC capabilities to distill insights from big data. Forward-looking companies should recognize that this shift can deliver greater value from data while making their HPC environments more flexible.

The rarified roots of supercomputing

High-performance computing (HPC) and supercomputing, which have become essentially synonymous, trace their history back as far as the 1920s. In the early modern era, HPC was the province of monolithic supercomputers, such as those produced by Cray Research beginning in the 1970s. The original CRAY-1, illustrated in Figure 1, was introduced in 1976. That machine was capable of 100-160 megaflops, weighed more than five tons, and was priced between US\$5 and \$8.8 million for the system alone.^{1,2} The need for specialized facilities and installation expertise drove the actual price far higher than that. The cost and other requirements for HPC in this era placed it outside the realm of possibility for most organizations. This class of compute power was reserved for rarified customers, such as national research laboratories.



facebook.com/redhatinc
@redhat
linkedin.com/company/red-hat

redhat.com

¹ The Open Compute Project, "Cray 1." March 14, 2013. https://www.thocp.net/hardware/cray_1.htm

² Cray Research, Inc., "The CRAY-1 Computer System." 1977. <http://s3data.computerhistory.org/brochures/cray.cray1.1977.102638650.pdf>.



Figure 1. Cray-1 supercomputer housed at the Deutsches Museum in Munich.³

Today, mainstream gaming consoles perform multiple teraflops – a million times more than the Cray-1. At the high end, the Sunway TaihuLight supercomputer in Wuxi, China topped the Top500 list in 2016 at 93 petaflops,⁴ or roughly a half-billion times faster.

By the 1990s, clusters of massively parallel computers built from industry-standard components began to emerge alongside proprietary machines. This trend marked the beginning of HPC based on open standards hardware and open source software, which continues to provide democratized HPC capabilities for mainstream organizations.

³ Pfeiffer, Clemmons. "Cray-1-deutsches-museum.jpg." Wikimedia Commons, November 25, 2006 <https://commons.wikimedia.org/wiki/File:Cray-1-deutsches-museum.jpg>.

⁴ TOP500.org, "November 2017 TOP500 List." November, 2017. <https://www.top500.org/lists/2017/11/>.

The mainstream transition and expansion of HPC

As access to large-scale computing capabilities has expanded, so has the scope of problems to which they are applied. HPC used to be used exclusively for scientific and technical computing, but now fields as diverse as sociology and economics now commonly use these technologies. Similarly, HPC is used across industries – from oil and gas exploration to agriculture.

HPC has found its way into everything from modeling the climate to simulating subatomic interactions to designing golf clubs, and society as a whole has become more data-driven. Everyday matters from business decisions to election strategies now draw on big-data analytics, and our collective appetite for this intelligence continues to grow.

Open source platforms such as Apache Hadoop MapReduce, Apache Cassandra, and Apache Spark, have been created to help parse and interpret this data into insights. To power increasingly sophisticated algorithms, high-performance data analysis (HPDA) has emerged as a new segment that applies the resources of HPC to big data. In addition, supercomputing is enabling deep learning and neural networks to advance artificial intelligence (AI).

Regardless of the problem domain, these efforts all draw on a common root concern: the ability to solve large-scale computational problems within reasonable time and cost parameters. By applying more compute power – or applying a given amount of compute power more effectively – data-intensive problems can be run using larger datasets, in the same amount of time. This ability allows problems to be described and examined at higher resolution, larger scale, or with more elements. Greater compute power or efficiency can also allow the same problem to be completed more quickly, allowing more iterations of a simulation – and generating better results.

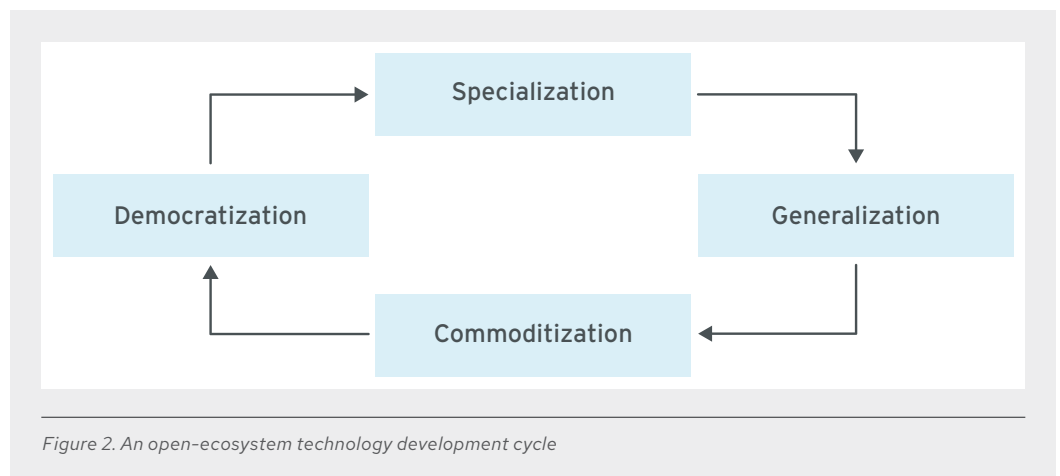
The compute resources being applied to these problems are expanding beyond the on-premise compute clusters that are typically associated with HPC. Increasingly, these problems are drawing on the elastic resources available from public cloud services. Specialized hardware is also expanding, such as tensor processing units (TPUs), a specialized type of application-specific integrated circuit (ASIC) that can be deployed on-premise or in the cloud to accelerate machine learning and other tasks.

Together, these factors illuminate a future of HPC that is defined by diverse hardware and open source software in place of yesterday's proprietary, special-purpose architectures. This vision includes cloud computing and HPC-as-a-Service instead of depending on massive on-premise compute resources. It also embraces community-based innovation for tools and techniques to deliver knowledge at the speed of imagination.

Moving the HPC ecosystem forward with open technologies

Much of the power behind HPC's ongoing development has always been community-driven. Collaboration between industry, academia, and public entities, such as the U.S. Department of Energy and the European Commission, continue to develop tools and technologies to expand existing capabilities and generate new ones.

Technology capabilities are often developed for specialized purposes as the beginning of a cycle that also includes generalization, commoditization, and democratization, as shown in Figure 2. For example, a private company that takes the source code of an internally developed tool and opens it to the community benefits from contributions made by others, which improves the original tool beyond what that company could have done on its own. At the same time, others adapt the tool to their own problem domains, expanding its capabilities to a more generalized set of usages. Members of the growing ecosystem work around its dependencies on proprietary technologies, enabling commoditization. The resulting lower cost and functional variation makes the capabilities more ubiquitous, democratizing technology. That sets the stage for the development of new, specialized functionality that begins the cycle anew.



The success of this ongoing cycle inherently depends on open, shared architectures and application programming interfaces (APIs) that all parties can access and use to advance the state of the underlying projects. Therefore, open source software, open standards hardware, and community-driven standards are critical to the ongoing development of HPC technology. Private industry also has a substantial role to play as connections between self-interest and community effort are key to this sustainability.

The community must ensure the enterprise-readiness of the code, systems, and standards being developed, making them suitable for business-critical usages. Participants in this cycle include private-sector community members, such as Red Hat, that specialize in hardening open source technologies for security, stability, and support.

It is also important to ensure that the next generation of programmers embraces the future sustainability of a vibrant HPC ecosystem. According to HPCwire, the talent shortage in this area exists because new university graduates “lack basic HPC competency and because supercomputing is sometimes mistakenly seen as an old technology.”⁵ To remedy this shortage, industry has started to offer new internship offerings and on-the-job training.

⁵ HPCwire, “ISC 2017 student day tackles the worldwide HPC workforce shortage.” June 12, 2017. <https://www.hpcwire.com/2017/06/12/isc-2017-student-day-tackles-worldwide-hpc-workforce-shortage/>.

The changing roles of private companies in the future of HPC reveal new relationships and interdependencies between for-profit entities and the rest of the community. The technologies that are moving HPC forward require collaborative development between commercial, academic, and public sector participants.

Technology approaches to meet emerging challenges

The idea of a collaborative path forward that brings together the interests of all players is at once artificial and inevitable. That is, while there is obviously no overarching plan laid out by a central authority, the ecosystem is nevertheless coherent because of the interdependencies outlined above and the tendency for advances to produce generalized benefits to all.

Organizations such as the Society of HPC Professionals (SHPCP) also play a role in this development, bringing together the interests of different groups. For example, the U.S. Federal Bureau of Investigation (FBI) attended one of the Society's recent annual technical meetings, to which it lent a strong emphasis on cybersecurity. As a result, the potential for attacks either on or using HPC systems was elevated in the SHPCP's proceedings and therefore within the ecosystem as a whole.

Similar cross-pollination of interests occurs from participation by people and organizations that work in areas as diverse as precision agriculture, oil and gas, and medical devices. Each has a unique take on the needs of the community that nevertheless resonate more generally, regardless of specific problem domains. These shared interests below the surface provide the lasting basis for industry groups and standards to bring together the divergent interests within the community.

In addition, crossover of technologies and capabilities from technology areas outside HPC play a significant role in the next generation of HPC. General-purpose computing and broader trends from enterprise technology are helping to shape HPC's future.

Continuing on the path away from the monolithic on-premise supercomputers of the past, a substantial thrust of progress in the coming years is to decouple computation entirely from specific hardware infrastructure. Technologies that will play a role in this evolution include hybrid cloud,

Linux® containers, and hypervisor-based virtualization. These approaches have become increasingly mainstream in general-purpose computing, and adapting HPC workloads to take advantage of these approaches will be a significant area of focus in the coming years.

Simplified, robust hybrid cloud architecture

Cloud adoption for HPC is central to the transition of workloads from an on-premise-only approach to one that is decoupled from specific infrastructure or location. This vision takes advantage of the elasticity of public cloud computing, providing resources on demand beyond the scope of owned capital equipment. Providing this ability is both cost-effective and conducive to maximum flexibility, even for large, sporadic workloads.

Smooth interoperation between on-premise resources and public cloud services, such as Amazon Web Services, Google Cloud Platform, or Microsoft Azure, can be provided by a hybrid cloud approach. The architecture of internal clouds are more similar to public clouds than other on-premise infrastructures, enabling large-scale HPC modeling, simulations, and analytics to treat the combination of internal and external clouds as a single, coherent resource. Red Hat Cloud Suite – a container-based application development platform that includes Red Hat CloudForms®, Red Hat OpenStack® Platform, Red Hat Virtualization, Red Hat OpenShift® Container Platform, and Red Hat Satellite – provides a comprehensive toolchain to support this approach.

Lightweight Linux containers

Traditionally, HPC workloads have been created, tuned, and maintained with a specific set of hardware dependencies in mind. This approach is a direct reflection of on-premise infrastructure models, where detailed assumptions can be made about the known, steady systems environment. By contrast, organizations must begin to prepare workloads for agile movement among cloud-based resources operated by multiple entities at multiple locations. In this modality, flexibility is highly valued, as a given workload must be able to operate on any available general-purpose computing equipment, regardless of the specific resources or instruction set architecture.

Using Linux containers, a workload's hardware dependencies can be packaged up with it, in a self-contained unit for deployment as needed. Moreover, because containers are designed to be lightweight, they enable flexibility with low levels of overhead—improving performance and cost.

Red Hat OpenShift simplifies container adoption with an environment built on top of the Kubernetes container cluster manager. Red Hat OpenShift enables unified development and deployment on nearly any infrastructure, either public or private.

Hypervisor-based virtualization

While containers have the clear advantage of low overhead, they are not suited to every implementation. A key consideration in this area is the isolation of data between workloads, which is addressed strictly at the software level with containers. That is, containers are an instance of operating-system (OS)-level virtualization, where data isolation between containers is handled by the kernel. Software-only data protection is inherently vulnerable to software-based attacks, which can be a disqualifying consideration for particularly sensitive or regulated workloads.

Hypervisor-based virtualization, by contrast, isolates data within the virtual machine (VM) where a workload executes, using hardware-based measures. The trade off is that the hypervisor itself carries a degree of overhead, as does the copy of the OS that is contained within each VM. Therefore, compared with containers, hypervisor-based virtualization provides more robust data protection, but lower efficiency. The use of both approaches in tandem can therefore allow for greater flexibility than using either approach on its own. Red Hat Virtualization is an enterprise virtualization platform based on Kernel-based virtual machine (KVM) that can help organizations increase the efficiency and cost effectiveness of virtualization, while also delivering the benefits of a true open source solution with high performance and security.

Preparing for the world ahead

IT departments should consider how they can deliver HPC using their general enterprise infrastructures. Continuing the pattern of moving beyond the massive, monolithic on-premise machines of decades past, emerging HPC strategies should include general-purpose computing approaches such as hybrid cloud, Linux containers, and hypervisor-based virtualization.

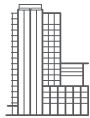
Aside from infrastructure, this generalized approach also draws from mainstream enterprise skills, such as parallel programming. In a world where new college graduates in the science, technology, engineering, and math (STEM) fields increasingly pursue the next big thing in areas such as social computing analysis and AI, sensitivity to how those fields can inform a more vibrant future for HPC is a wise course for the community.

Conclusion

In the coming decade, the distinctions between HPC and general-purpose enterprise computing will continue to dissipate. In place of monolithic, on-premise supercomputers, the industry will continue to move toward more flexible infrastructure based on open standards, with workloads decoupled from specific hardware.

Technologies associated with mainstream enterprises, as opposed to proprietary, special-purpose technologies, will make this trend possible. In particular, open source approaches to hybrid cloud and virtualization using both containers and hypervisors will deliver economic advantages as organizations apply HPC capabilities to distill insights from big data. Forward-looking companies should recognize that this shift can deliver greater value from data while making their HPC environments more flexible.

About Red Hat



Red Hat is the world's leading provider of enterprise open source software solutions, using a community-powered approach to deliver reliable and high-performing Linux, hybrid cloud, container, and Kubernetes technologies. Red Hat helps customers integrate new and existing IT applications, develop cloud-native applications, standardize on our industry-leading operating system, and automate, secure, and manage complex environments. Award-winning support, training, and consulting services make Red Hat a trusted adviser to the Fortune 500. As a strategic partner to cloud providers, system integrators, application vendors, customers, and open source communities, Red Hat can help organizations prepare for the digital future.



facebook.com/redhatinc
@redhat
linkedin.com/company/red-hat

North America
1 888 REDHAT1
www.redhat.com

**Europe, Middle East,
and Africa**
00800 7334 2835
europe@redhat.com

Asia Pacific
+65 6490 4200
apac@redhat.com

Latin America
+54 11 4329 7300
info-latam@redhat.com