Secure Private Clouds for HPC

Technical University of Denmark’s approach to virtualizing HPC supercomputing for life science and healthcare research
Table of contents

Accelerating Health and Life Science Research ................................................................. 3
  The DTU Supercomputing center for life sciences 3
  Goals for Computerome 3

Computerome Solution Architecture Overview ................................................................. 5
  Logical domains 6
  Delivering efficient HPC services in a flexible manner 8

Cloud Computing Technologies used in Computerome ....................................................... 9
  Abstracting compute infrastructure 9
  Cluster deployment and configuration 9
  Networking at supercomputing cluster scale 10
  Integrated network virtualization and security services 10
  Cluster life cycle automation using Infrastructure as Code 12

Cloud and virtualization deliver additional benefits for HPC ........................................... 14

Glossary ............................................................................................................................... 16

Authors ............................................................................................................................... 16
Secure Private Clouds for HPC

Accelerating Health and Life Science Research

In healthcare, time is of the essence. The research that advances healthcare must also move fast—while still preserving the privacy of patients’ sensitive medical data. Technical University of Denmark (DTU) took an innovative approach to virtualize and use modern cloud technologies to implement a secure High-Performance Computing (HPC) environment tailored specifically for health and life science research. The university achieved its security goals and operational requirements by leveraging VMware infrastructure.

The DTU Supercomputing center for life sciences

DTU supports its research efforts with a world-class supercomputer center named Computerome. Researchers need advanced computing capabilities for projects as diverse as personalized medicine and drug development, the latter of which is based on new understanding of the mechanisms of diseases and disease biology. As part of Denmark’s national e-infrastructure, DTU offers access to all the country’s researchers in the field of life sciences, including other Danish universities, hospitals, and pharmaceutical companies. Thus, Computerome is a vital resource for DTU to realize its mission of providing value for society.

Computerome 2.0 is a second-generation supercomputer installed at the center. The original Computerome ranked number 121 among the TOP500 supercomputing sites when it opened in 2014. But in 2019, DTU made a significant capital investment to upgrade the university’s computing power. It augmented Computerome with the latest cloud technology. Computerome 2.0 consists of approximately 800 servers providing 31,760 CPU cores and 221 TB of memory, which gives life sciences research teams the ability to process large volumes of patient data securely.

Today, Computerome’s users in industry and academia number in the thousands, making it a highly coveted—and highly utilized—computing resource for life sciences and healthcare research in the Nordics.

Goals for Computerome

The DTU supercomputing leadership team’s vision and goals were driven primarily by the unique security and data processing needs of its multidisciplinary researchers. Security is critical to the broad spectrum of users, including university and hospital researchers working with personal health information. A traditional batch processing HPC environment, where the overall system is optimized and tuned for solving a singular set of scientific problems, could not meet the necessary security requirements.

In addition to the security concerns, the Computerome team knew that one size does not fit all when it comes to users’ needs. The university needed to provide customized compute profiles with different software stacks, while supporting large and diverse datasets at the same time.
Computerome 2.0 offers key capabilities over a traditional HPC architecture:

a) **Secure, customized HPC resources for researchers**

Computerome took a revolutionary approach when it leveraged the latest cloud infrastructure solutions to provide the best of both worlds—traditional HPC and cloud computing—to its users. Research teams that need specialized computing resources with very specific software and security requirements can request access to a dedicated virtual computer cluster with customized software and security policies for their needs. DTU refers to this compute resource as a “secure private cloud.” The infrastructure hosts multiple secure private clouds, each uniquely customized and accessible by its tenants.

Meanwhile, the traditional HPC batch system remains available to all researchers whose needs can be met by a standard HPC environment.

b) **Secure end-to-end computing environment for highly sensitive healthcare data**

Computerome houses petabyte-size volumes of data that are highly sensitive and covered by the Danish Data Protection Act, so data security is a top priority. Both security measures and security guarantees are built into the infrastructure to prevent unauthorized access from other users of the same infrastructure. One data owner cannot access another owner’s data unless both parties have entered into a specific agreement for collaboration.

c) **Dynamic and flexible environment optimized for life sciences**

Computerome, like many supercomputing sites, is built as an infrastructure project which the university makes available to researchers across all disciplines. Life science projects, however, place unique demands on the system, as they involve many different specialist subfields as well as diverse data: genomes, high-resolution medical imaging, and even text data in the form of patient records, to name a few.

Automated workflows (workflows managed by Infrastructure-as-Code) allow the university’s IT teams and administrators to service their diverse research user base efficiently, so they can focus on improving Computerome capabilities. Without leveraging virtualization and VMware private cloud capabilities, IT personnel would have to spend significant time and effort to build each custom environment as new requests came in.
Secure Private Clouds for HPC

Computerome Solution Architecture Overview
The Computerome HPC platform architecture aims to provide a scalable and distributed platform that offers increased security, performance, and flexibility for highly demanding HPC workloads.

Researchers sometimes access shared resources by submitting job requests to a batch system, consisting of Adaptive Computing’s Moab Workload Manager and Torque Resource Manager. This is a traditional HPC approach, where jobs enter a queue and execute when cluster resources are available.

Other times, researchers can request access to an individual customized secure private cloud based on Linux or Windows compute nodes. The systems reside in an internal network that is only accessible via a virtual desktop with two-factor authentication. The Linux or Windows clusters are deployed with HPC middleware and a workload scheduler requested by the secure private cloud owner. At any time, there are multiple secure private clouds of various sizes hosted on Computerome.

The compute, storage, and network resources used in Computerome are representative of resources used for higher-end supercomputers. To accommodate the range of compute, memory, and I/O requirements in life science research, the system offers three types of compute nodes along with high-speed networking capabilities (Table 1).

<table>
<thead>
<tr>
<th>TABLE 1. COMPUTEROME INFRASTRUCTURE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General Compute Node</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Fat Compute Node</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>GPU Compute Node</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Interconnect</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Storage</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

The system administrators balance Computerome 2.0’s physical resources between the traditional HPC clusters and secure virtual clusters based on incoming demand by researchers. Figure 1 shows the different cluster types. The infrastructure footprint for each type changes based on user demand.
Computerome used an innovative approach by utilizing enterprise-proven computing and network virtualization technologies to meet its specific needs around security and diverse user software stack requirements. Adopting modern cloud technologies enabled Computerome administrators to offer enterprise-grade HPC services for secure provisioning of custom HPC environments and execution of life sciences workflows with access to secure curated data sets.

**Logical domains**

Three logical domains help Computerome administrators deliver HPC services to its users.

**Management domain**

The management domain is a central function for hosting all administrative components required for the virtual infrastructure platform, as well as components for adding platform monitoring and the cloud management portal. The management domain is deployed on a dedicated management cluster, configured to tolerate multiple host failures. It’s also important to note that loss or disconnection of the management domain will not impact running jobs within a workload domain. However, it will prevent the deployment of new cluster resources that were being provisioned.

**Workload domain**

The workload domain consists of virtual resource clusters, which abstract pools of physical hardware such as storage, compute and network capacity for the compute node instances to execute HPC workloads. The resource clusters are optimized for execution of the HPC and research workloads, by ensuring that the virtual compute nodes of the secure cloud are in close network proximity to each other for optimal performance.

**Virtual Desktop Infrastructure (VDI) domain**

The VDI domain consists of multiple virtual resource clusters dedicated to providing virtual desktops to Computerome 2.0 users. These virtual desktop pools are tightly integrated with their respective secure cloud environments—that is, they are created and connected to the same logical switches during cloud deployment, providing direct access to all secure cloud resources that adhere to the same security model.

Figure 2 shows the three domains built on top of the virtual infrastructure, as well as how they are accessed by the HPC users and cluster administrators.
Delivering efficient HPC services in a flexible manner

For HPC infrastructure administrators, building a dynamic environment at the scale of Computerome places enormous demands on IT resources. With VMware technology, they were able to:

- Provide a highly dynamic environment without repetitive IT cycles and redundant resources
- Maintain multitenancy, data privacy, and security in software
- Automate workflows to deliver Infrastructure as Code

On Computerome, compute hosts may boot into either traditional HPC or secure private cloud environments. Administrators configure the networking capabilities of the secure private cloud based on user preferences. For example, an environment can simply exist without any data upload or download capabilities, or it can be configured with a proxy server to access the Internet for downloads.

The secure private cloud platform mitigates technical risks related to software attacks and cyberattacks two ways: First, management policies ensure users can only access groups for which they are authorized. Second, network and user group segmentations ensure that research groups are not able to access or compromise other research groups or their data.

In addition to flexibility, performance is a key requirement for Computerome. The HPC-specific performance requirements for the platform are based on the following principles:
Secure Private Clouds for HPC

- The ability to place compute nodes in proximity to each other to ensure performance for tightly coupled workloads that rely on Message Passing Interface to scale out.
- A secure cloud deployment that provides dedicated resources to specific users.
- The ability to provide InfiniBand for workloads to ensure low latency and high bandwidth.

Cloud Computing Technologies used in Computerome

Abstracting compute infrastructure
Managing compute infrastructure at scale is always difficult, and even more so when that infrastructure must support multiple independent research teams and organizations across the European Union. To address challenges around scale and complexity, DTU turned to vSphere software, which has become the de facto platform for managing both traditional applications and modern applications built with containers and microservices.

Through virtualization, the software transforms siloed clusters into aggregated computing infrastructures that include CPU, storage, and networking resources. vSphere software manages these infrastructures as a unified operating environment and provides organizations with the tools to administer the data centers that participate in that environment.

vSphere lifecycle management makes the maintenance of software upgrades, patching, and firmware updates easier and less disruptive. Combined with improved hardware utilization and streamlined operations, this results in reduced capital costs and operating costs, as well as increased business agility.

Cluster deployment and configuration
In recent years, organizations such as DTU have started capitalizing on virtualization for HPC workloads and have begun realizing many of the same operational benefits as they might with enterprise applications—all while maintaining performance akin to purely physical supercomputers. In fact, DTU has realized much better hardware utilization and can now run a diverse set of workloads across Computerome 2.0, all while reducing costs through more effective resource sharing.

It starts with deploying and configuring the compute infrastructure. DTU relies on vSphere host profiles to provide a consistent, automated, and centrally managed mechanism for host configuration. In combination with vSphere Auto Deploy, this setup streamlines the initial deployment of new HPC clusters by reducing reliance on repetitive, manual tasks. This is also essential for maintaining configuration compliance, which is key to remaining aligned with statutory data acts in the European Union.

Each host profile captures the configuration of a reference host and stores this configuration as a managed object with a catalog of information to configure networking, storage, security, and other host-level parameters. These profiles can then be applied to individual hosts or entire clusters, depending upon the needs of any particular research team.

Checking the compliance of hosts or clusters against their respective host profiles validates the configuration across the environment. Any configuration drift is flagged and reported through operational dashboards.
With a tested and validated set of host profiles defined, DTU uses Auto Deploy and a Preboot eXecution Environment infrastructure to provision and customize hosts across Computerome at scale. No state information is stored on the host profiles. Instead, the Auto Deploy server manages state information for each host, saving significant time during initial deployment and as new projects replace old ones.

With a physical infrastructure in place, DTU needed an easy way to deploy virtual infrastructure—this is where VM templates offered a solution. VM templates are master copies of virtual machines (VMs) that describe parameters such virtual disks, devices, and other settings, and are used to streamline the repeated deployment of virtual machines.

Whereas installing and configuring a Linux operating system on a physical computer takes several minutes, an entire cluster of VMs can be cloned from VM templates and deployed across a virtual cluster in seconds. DTU manages a small library of VM templates configured to address the various configurations required by Computerome 2.0 customers.

Networking at supercomputing cluster scale

Another key challenge for Computerome was networking at scale and configuring security overlays for individual clouds. With hundreds of hosts and VMs, DTU needed a way to manage everything from one place. This was accomplished through a combination of the vSphere Distributed Switch (vDS) and NSX Data Center, VMware’s virtual networking and security platform.

vDS provides a centralized interface from which DTU can configure, monitor, and administer VM access switching for all of Computerome. It provides simplified VM network configuration, enhanced network monitoring, and troubleshooting capabilities.

vDS also separates the data plane and management planes. The management functionality of the distributed switch resides on the vCenter Server system that lets DTU administer the networking configuration of Computerome on a data center level. The data plane remains locally on every host that is associated with the distributed switch. The data plane section of the distributed switch is called a host proxy switch. The networking configuration DTU creates on vCenter Server (the management plane) is automatically pushed down to all host proxy switches (the data plane).

As highlighted previously, InfiniBand interconnect provides high-performance, low-latency communication among the nodes in a cluster for tightly coupled traditional scale out HPC workloads. VMware and Mellanox Technologies Ltd. introduced support for Single-Rooted I/O Virtualization (SR-IOV) for InfiniBand adapters in vSphere 6.5, which enables the hardware resources of the physical device to be presented as many virtual functions. Each of these can be individually configured and passed through into a different VM.

Another usage of SR-IOV is to allow one or more VMs on a host to access high-performance cluster file systems, such as a Lustre file system, by sharing the single, physical IB connection between the host and storage system. By using SR-IOV from the beginning, Computerome can provide extreme performance with very low latency while simultaneously serving the diverse needs of multiple VMs on a host.

Integrated network virtualization and security services

While the vDS dramatically simplified networking at scale, DTU needed to take a software-defined approach to networking. Choosing NSX as the networking platform for Computerome, they were able to bring networking and security closer to the high-performance workloads, regardless of whether they are running in virtual machines, containers, or even traditional physical machines. Like the operational
model of VMs, networks can be provisioned and managed independent of underlying hardware. NSX reproduces the entire network model in software, creating and provisioning Computerome network topology—including switches, routers, load balancers, and firewalls—in seconds. Moreover, system administrators can use VMware’s Cloud Management Platform to automate the deployment of networks and security policies as new clusters are provisioned.

Given the nature of personal health information that resides on Computerome, enacting zero trust security is a top priority for DTU—whether it’s locking down specific workloads, creating logical demilitarized zones (DMZs), or reducing the attack surface of a virtual desktop environment used by researchers. Computerome takes a defense in depth security approach that operates in three main firewall layers: traditional perimeter firewall, NSX-T edge gateway, and distributed firewall.

Virtualizing Computerome networking and security services enabled faster provisioning and deployment of full stack applications by removing the bottleneck of manually managed networking and security services and policies. NSX also integrates natively with vRealize Automation to empower DTU developers and IT teams to provision, deploy and manage workloads much faster and more efficiently.

The Computerome architecture contains various components that must be integrated to make them work well together as well as provide defense in depth. These components include networking and security components, infrastructure services, storage devices, and applications.

There are seven key networking and security components:

- External network: the network external to the NSX domain. Connectivity to and from the external network is provided by the NSX Edge cluster.
- Physical perimeter firewall: the entry point to the Computerome 2.0 environment and the first level of filtering on traffic from the external environment. This firewall enforces security control of traffic going to external non-trusted networks. However, it does not enforce security on the application traffic flows in the internal network or legacy trust zone flows.
- Physical router: sits between the perimeter firewall and HPC environment, creating a DMZ for the physical HPC environment.
- NSX-T tier-0 logical router: a gateway that handles traffic entering and leaving the NSX domain by way of BGP router peering. The firewall monitors states on tier-0, unless the edge nodes are configured for ECMP, in which case the rules for the firewall are automatically converted to stateless.
- NSX-T tier-1 router: a logical router optimized for forwarding in the virtualized space—that is, communication between VMs.
- Internal networks: the main workload networks connected to a tier-1 distributed logical router that sits behind the NSX Edge cluster tier-1 router. These networks are overlay segments instantiated and maintained by NSX. They are completely decoupled from the underlying network, without the need for any configuration in the physical infrastructure.
- NSX-T Distributed Firewall: a firewall distributed across the hypervisors in a transport zone. It provides granular firewalling for communication among the VMs. VMware’s NSX Distributed Firewall product operates at the VM virtual network interface card level, meaning that a VM is always protected, regardless of how it connects to the logical network.
Cluster life cycle automation using Infrastructure as Code

Turnover and reconfiguration of high-performance computing resources from one project to another have traditionally been extremely time consuming. DTU needed a way to seamlessly orchestrate provisioning of compute, storage, network, and security. It also needed a way to manage users and groups to efficiently deploy (and destroy) private clouds for the various research teams using Computerome.

To accomplish this, the Computerome team chose VMware vRealize Automation (vRA)—a modern infrastructure automation platform that increases productivity and agility by streamlining and eliminating manual tasks. This provided both IT operations and research teams the environments and resources they required, while maintaining security and control needed to meet strict compliance standards. It also reduced the provisioning time of HPC workloads from days or weeks to minutes.

Replacing human tasks and individual scripts with a fully automated pipeline not only accelerates deployment, but significantly reduces mistakes and ensures configuration compliance. Furthermore, Computerome 2.0 takes advantage of vRA’s tight and continuous integration with Workspace ONE Access and Active Directory for identity management.
At its core, vRA provides DTU with two primary capabilities. First is the design canvas, a visual drag-and-drop tool that allows system administrators to create blueprints of entire cloud deployments complete with compute, storage, networking, security policies, and other infrastructure. Second is a self-service portal that allows end users—in this case, the DTU team—to select from a catalog the appropriate blueprint and configuration specs to be provisioned.

Another key component of the platform that Computerome relies on heavily is vRealize Orchestrator—an extremely capable and extensible workflow engine, built with an open and flexible architecture that system administrators and IT operations staff can use to streamline tasks and integrate functions with third-party software.

Computerome used vRealize Orchestrator to create and manage workflows like the secure cloud deployment shown in Figure 4. Once a member of the system administrator team initiates the deployment process from the vRealize Automation self-service portal, this workflow executes a series of steps to stand up Active Directory groups, subnets, IP addresses, compute, networks, storage, and security policies—a total of 26 individual tasks altogether. The end-to-end automation of these otherwise manual processes enabled DTU’s small team of systems administrators to focus on critical issues and lean forward on innovation.

**FIGURE 4:** Automating life cycle operations using vRA.
In addition to leveraging vRealize Automation and Orchestrator for setting up Secure Private Cloud infrastructure, the system administrator team developed a set of predefined services that can be requested by Computerome users, including:

- Cluster with Moab Queuing System
- Kubernetes Cluster
- OpenStack Cluster
- SFTP server to import and export data in a secure fashion
- Ubuntu/CentOS and Windows standalone Virtual Machines

Users have several options to customize their cluster as well:

- Choice of operating system for the compute nodes
- Choice of precompiled bioinformatician software stacks
- Installation of custom software solutions not available in the shared HPC cluster
- Alternate container runtime environments

**Cloud and virtualization deliver additional benefits for HPC**

Computerome is a key tool for DTU researchers to conduct technical-scientific research which contributes to the development of society. Many of the research projects running on the supercomputer system involve highly sensitive as well as scientifically and financially valuable data. Data across research projects are often unstructured data from several sources - population DNA, patient journals, medical imaging, and published research. Of course, security and data protection are key priorities when it comes to clinical and patient data. The shared infrastructure on traditional HPC clusters doesn’t offer ways to enforce strict security boundaries within a shared cluster. Computerome drew upon compute and network virtualization techniques, along with workflow automation to augment traditional HPC capabilities.

Computerome users benefit from the fast, flexible, and secure infrastructure and the ability to combine different types of sensitive data and perform analysis in a timely manner. The performance and agility provided by VMware cloud software is helping DTU predict people’s medical trajectory and develop precision medicine for life threatening diseases.

Today, research organizations are asked to provide HPC services for a growing set of diverse use cases, many of which cannot be met by traditional architectures. Frequently, users demand a more agile and flexible infrastructure platform that aligns with modern cloud principles and user experiences. If your organization is facing similar challenges, consider going back to the drawing board to leverage the latest cloud virtualization and services from VMware to build the solution.
Additional case studies and best practices for HPC on VMware are available at www.vmware.com/go/hpc.
## Glossary

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workload Manager</td>
<td>HPC cluster resource manager and orchestration platform to schedule and manage batch jobs on distributed compute nodes.</td>
</tr>
<tr>
<td>Host</td>
<td>A physical computer that uses virtualization software to run virtual machines. Also called the host computer, host machine, or host system.</td>
</tr>
<tr>
<td>Micro-Segmentation</td>
<td>Network virtualization technology to create increasingly granular secure zones in data centers which isolate each individual workload and secure it separately.</td>
</tr>
<tr>
<td>Virtual Desktop Infrastructure (VDI)</td>
<td>Use of virtual machines to provide and manage virtual desktops. VDI hosts desktop environments on a centralized server and deploys them to users on request.</td>
</tr>
<tr>
<td>Secure Private Cloud</td>
<td>Computing resources dedicated to a single user organization with highest levels of isolation, dedicated resources, privacy, and security.</td>
</tr>
<tr>
<td>Distributed Firewall (DFW)</td>
<td>A software-defined Layer 7 firewall enabled at each workload to segment east-west traffic and block lateral movement of threats.</td>
</tr>
<tr>
<td>InfiniBand (IB)</td>
<td>Computer networking communications standard used in high-performance computing that features very high throughput and very low latency.</td>
</tr>
<tr>
<td>Border Gateway Protocol (BGP)</td>
<td>Gateway protocol that enables the internet to exchange routing information between autonomous systems.</td>
</tr>
<tr>
<td>SR-IOV</td>
<td>A specification maintained by the PCI Special Interest Group (SIG) that allows a PCIe device to appear to be multiple separate physical PCIe devices.</td>
</tr>
<tr>
<td>Higher Performance Computing (HPC)</td>
<td>High-performance computing is the use of super computers and parallel processing techniques for solving complex computational problems.</td>
</tr>
<tr>
<td>Infrastructure as Code</td>
<td>Infrastructure as Code is the managing and provisioning of infrastructure through code instead of through manual processes.</td>
</tr>
<tr>
<td>Software Containers</td>
<td>A form of virtualization where applications run in isolated user spaces, called containers, while using the same shared operating system.</td>
</tr>
</tbody>
</table>

## Authors

**Ramesh Radhakrishnan**

Ramesh is a Technical Director in the HPC/ML team within VMware OCTO, where he leads a team focused on HPC and Machine Learning workloads. He has presented his work at key industry events such as NVIDIA GTC, VMworld, Dell Technologies World, and ARM TechCon. He has 15 published patents and received his PhD in Computer Science and Engineering from the University of Texas at Austin.

**Shawn Kelly**
Shawn is a Staff Engineer in San Diego who works with VMware customers to develop and deploy innovative cloud solutions. He supports VMware’s Office of the CTO in High Performance Computing and Machine Learning projects and leads a team of Machine Learning SMEs in the field. Shawn earned Master of Science degrees in both Software Engineering and Information Technology Management from the Naval Postgraduate School in 2010 and subsequently served at Marine Corps Systems Command implementing solutions for the Marine Corps. In the decade preceding, he served as a Marine Helicopter Pilot.