Genomics on Cisco Metacloud + SwiftStack

Technology is a large component of driving discovery in both research and providing timely answers for clinical treatments. Advances in genomic sequencing have driven up demands of compute, networking, and storage. As a result, High Performance Computing is undergoing the same evolution to cloud-native technologies as the rest of Enterprise IT.

The main driver of this transformation is an increase in accessibility and availability of IT resources to users within the organization. The applicability of cloud computing to science and data analytics is clearly indicated by the increased use of public cloud providers such as AWS, Azure, and Google Compute in this area.

However, because public cloud providers tend to tailor to a more generic use case, running HPC workloads on public clouds can be cost prohibitive. In addition, the cost to transfer and store data in public clouds can be substantial (on average a 30% uplift on top of compute resource costs). These factors actually pull dollars away from research programs that could be more effectively spent on private cloud resources.

Private cloud is the best of both worlds – the flexibility of public cloud but with the performance characteristics of on-premises HPC infrastructure.

Next-generation sequencing produces an enormous amount of data, and private cloud solutions can dramatically reduce the turnaround time analyzing that data. The faster genomes can be compared, the faster scientific research can happen, the faster a clinician can get the results, and the faster a patient can get a diagnosis.

HPC-optimized Private Cloud

Traditionally, HPC applications have been run on special purpose hardware administered by staff with specialized skills. Furthermore, most HPC software stacks are rigid and fundamentally different from other more widely adopted environments requiring a special skillset by the researchers that want to run the applications, often needing to become programmers or part-time sysadmins themselves.

The adoption of cloud-native technologies drastically increases the productivity of your research organization by making both its activities and underlying resources more flexible, scalable, efficient and portable. Cloud platforms such as OpenStack provide a way to collapse multiple silos into a single private cloud platform while making those resources more accessible through self-service portals, CLIs, and APIs. Using OpenStack, multiple workloads can be distributed amongst the resources in a granular fashion increasing overall utilization and reducing cost.

While virtualization has not been embraced by the HPC community, a new class of embarrassingly parallel applications work well in this environment. Furthermore, open source virtualization technologies, such as Kernel Virtual Machine (KVM), bring hypervisor overhead in the below 2%. Even with this overhead, however, the cost savings from increased utilization and economies of scale allow institutions to offer computing resources to a wider audience of researchers. While the time-to-solution for a single computational job might take a fraction longer in a cloud environment, overall time to scientific discovery is actually reduced, because more compute capacity is available due to decreased cost and increased access, flexibility, and efficiency to infrastructure.

With the ever-increasing volume of data, current storage architectures are burdened with bottlenecks and limitations. They do not scale out efficiently and therefore create silos of storage pools resulting in operational overhead. The industry has seen how powerful an object storage system can be. Amazon had popularized object storage with their wildly-successful Simple Storage Service (S3). However, for data-intensive storage workflows, public cloud can be cost prohibitive and data transfer times can be hard to overcome. Object storage platforms such as those from OpenStack, enable compute jobs to run at “full-tilt.” High throughput, highly parallel access can take advantage of wide networking capacity, which enables the scaling-out of compute resources, shortening processing time.

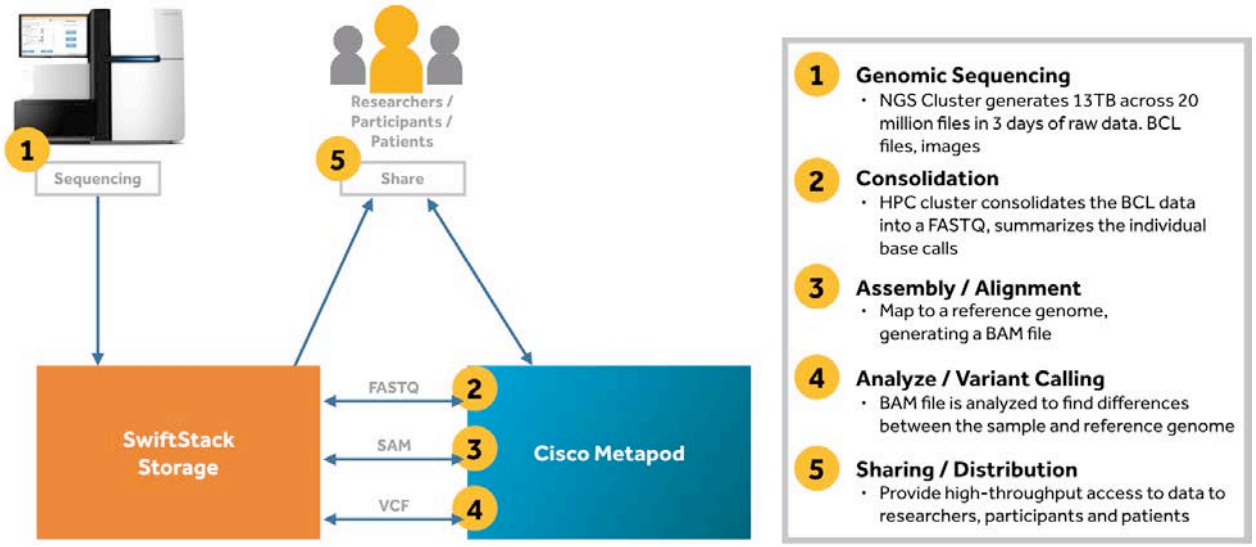
Linux containers are another form of cloud technology that can enable HPC workloads. Containers offer both a convenient and portable form of packaging and way to run multiple, independent workloads on the same server without the use of a hypervisor. This provides two important capabilities to HPC applications. First, the lack of a hypervisor provides greater overall performance and better access to special purpose hardware such as GPUs. Second, the packaging benefits of containers allow applications to be more portable and, as a result, alleviate the need for HPC centers to create environments to accommodate every possible workload their customers might require. This lowers the burden on operational staff and allows for HPC resources to be more accessible by researchers in the organization or greater scientific community. For institutions in the business of creating the applications and toolkits used by the broader scientific community, containers additionally allow a more agile method for developing and distributing software.

Solutions for private cloud

Cisco Metacloud is a production-ready, OpenStack-based, on-premises solution that Cisco engineers, deploys, and remotely operates 24x7x365 on your behalf. Cisco Metacloud provides an experience like public cloud, but in your data center.

SwiftStack is an object storage system, which includes an unmodified, 100% open-source release of OpenStack Swift at the core. SwiftStack provides a complete object storage solution with extensive functionality for deploying, integrating, upgrading, and managing single and multi-region Swift clusters, supported by SwiftStack 24x7. SwiftStack can also be an integrated component of a managed Cisco Metacloud deployment.

Genomics on Cisco Metacloud + SwiftStack



Genomics is a complex operation encompassing a sequence of steps, each with demanding requirements. Cisco Metacloud + SwiftStack provides a flexible infrastructure to meet the needs of the entire lifecycle of genomics research. Beginning with the analysis of genetic material, the sequencers write out BCL files to SwiftStack Object Storage through an SMB mount point provided by a SwiftStack Drive storage. Next, virtual compute nodes are spawned on the Metacloud Managed OpenStack Private Cloud that read the BCL files from the object storage via Object APIs and consolidate them into FASTQ files that summarize the individual base calls and their associated quality scores. In parallel, more virtual compute nodes read the indexes of the FASTQ files created in the previous step to align with a reference genome to create a SAM file with aligned sequence data and metadata that can be compressed into binary BAM files. Finally, the BAM file is analyzed for differences in the sequenced genome and a reference genome and the variants are written out to a VCF file. All of the data produced in these steps can be stored long-term in the SwiftStack cluster for further analysis. This data can then be shared with researchers, participants, and patients using web front-ends, application servers, databases, etc., all running on the Metacloud infrastructure.

Cisco Metacloud + SwiftStack on Cisco UCS

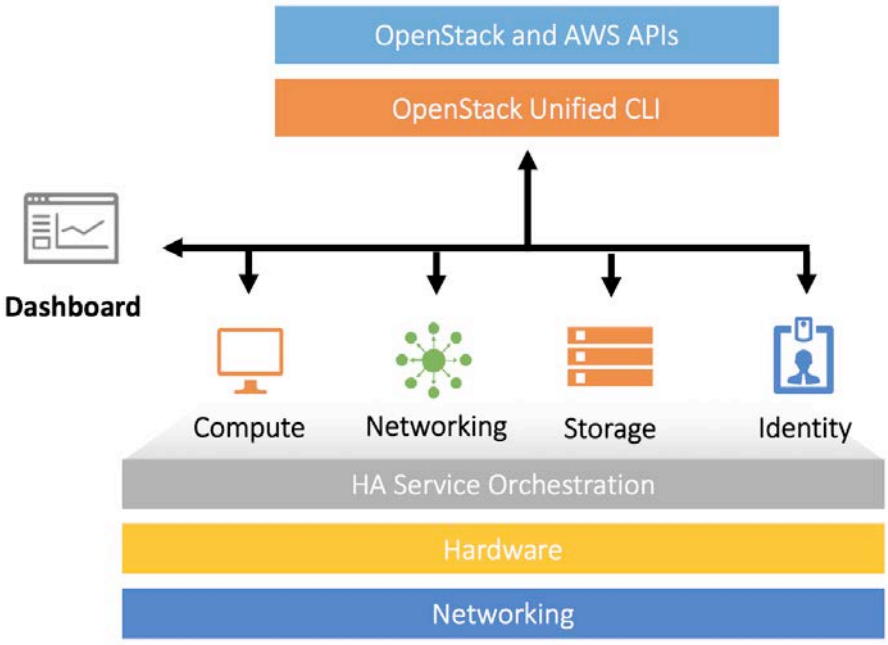
Software and operational costs (including staff) are among the largest costs that are incurred in running most data centers. Open source software such as OpenStack has several benefits including cost, flexibility, freedom, security and accountability. However, It is essential that you select the right solution including software, hardware, infrastructure, and operational services to run it in an efficient and cost effective manner. Cisco addresses this with Cisco Metacloud and SwiftStack deployed with the Cisco Unified Computing System (UCS) converged cloud platform.

Cisco UCS reduces the cost of staff, infrastructure, power, space, and cooling with a versatile platform that converges compute, network, and storage into a unified platform with a single management interface. The compute, network, and storage components of this infrastructure can be scaled independently, avoiding islands of stranded capacity. In addition to providing a common interface to a converged infrastructure, it also provides a single programmatic API for the software-defined compute, network, and storage provided by OpenStack to

decrease administrative burden on staff and increase operational efficiency and response to users. Cisco UCS also provides a wide range of options, including a dense form factor, that can reduce the number of components and data center space needed for efficient cloud computing.

Cisco Metacloud

Cisco Metacloud is a remotely engineered and operated private cloud designed with developers and researchers in mind. With Metacloud, Cisco can handle the operational needs of your organization so your experience mirrors the ease and stability of a public cloud with the security that comes from a solution behind your own firewall. Cisco Metacloud can be deployed in a matter of weeks with a 99.99% uptime SLA.

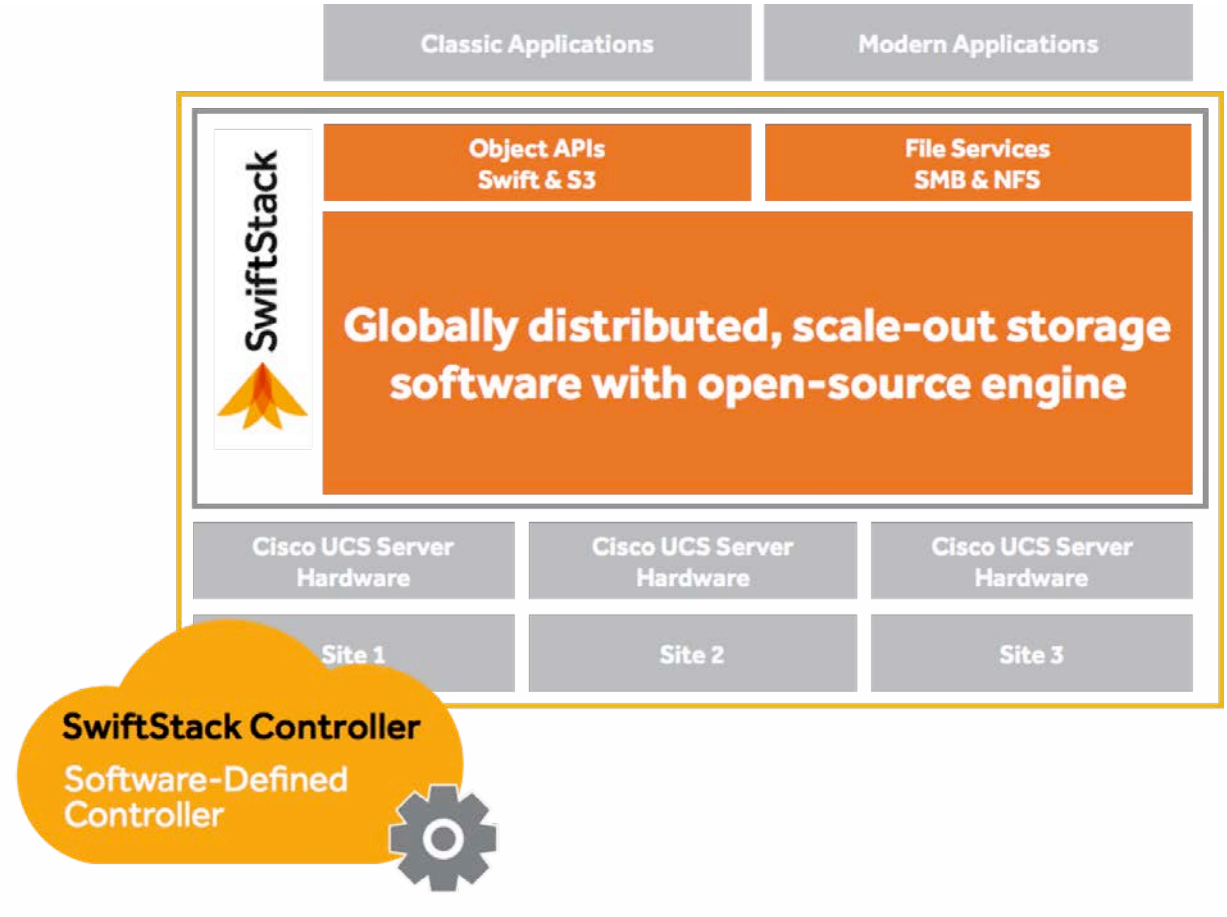


SwiftStack Object Storage

SwiftStack provides object storage software for high-capacity workloads. The SwiftStack solution enables on-premises storage to be delivered “as-a-service” like the public cloud with OpenStack Swift APIs, AWS S3 APIs, and filesystem access. A key characteristic is presenting file data as objects that enables workflows to transition from filesystem to object APIs.

The SwiftStack Controller decouples the control, management, and configuration of storage nodes from the physical hardware. This enables non-disruptive capacity additions, rolling upgrades, and multi data-center management.

SwiftStack directly integrates with Cisco Metacloud and Cisco UCS. The solution can be delivered either as a managed service with Cisco Metacloud or self-managed. SwiftStack provides 24x7 support and hardware integration for Cisco UCS.



Cisco Metacloud + SwiftStack Features:

- **Central Management Interface:** Cisco Metacloud + SwiftStack lets you manage all of your object storage clusters from a single centralized management interface, reducing operational complexity, conserving IT resources, and saving costs.
- **No Hardware Lock-In:** Cisco Metacloud + SwiftStack runs on any industry-standard x86 hardware available from any vendor. You can mix-and-match hardware from multiple vendors, matching your capital investments directly to your storage needs and reusing legacy hardware.
- **Self-Service Portal:** Cisco Metacloud + SwiftStack provides access to compute, storage, and network resources through an open and extensible web portal, API, and command line.
- **Multi-tenancy:** Native multi-tenancy with separate network, compute cells, and storage quotas to avoid resource contention or provide secure isolation of workloads and data
- **Security:** Advanced identity and role management, project quotas, and separate L2 and L3 network domains with firewalling
- **Supports Existing Application Workflows:** Apps access SwiftStack storage via a RESTful HTTP API and a broad set of client libraries for PHP, Python, Ruby, Java, and other languages. And we also support the S3 API and file system protocols like NFS and CIFS.

- **Flexible Data Protection:** Storage policies let you create multiple tiers of storage within the same namespace, giving admins and applications full control over where their data is located and how many replicas or erasure-coded segments are kept for availability.
- **Enterprise Integrations:** Cisco Metacloud + SwiftStack easily integrates with your existing enterprise authentication infrastructure, including Active Directory and LDAP, speeding up the provisioning and management of storage users.
- **Non-disruptive Upgrades:** The SwiftStack Controller lets you update your storage cluster while it is running, with no disruption or downtime. One click and all of the steps needed to update nodes happens automatically in the background.
- **Monitoring and Alerts:** Cisco Metacloud + SwiftStack proactively monitors the health of each storage node and drive to ensure availability and performance. Alerts can be integrated with external monitoring tools such as Nagios and via SNMP.
- **Automatic Data Distribution:** A SwiftStack object storage cluster can be spread over multiple, geographically distributed data centers with a single namespace. This means built-in disaster recovery and more flexibility for your data.
- **Elasticity:** Cisco Metacloud virtual and bare-metal clusters can scale up and down to match demand and reduce under utilization.
- **24x7 Enterprise Support**

For More Information

Visit the Cisco Metacloud website: www.cisco.com/metacloud
Visit the SwiftStack website: www.swiftstack.com

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**