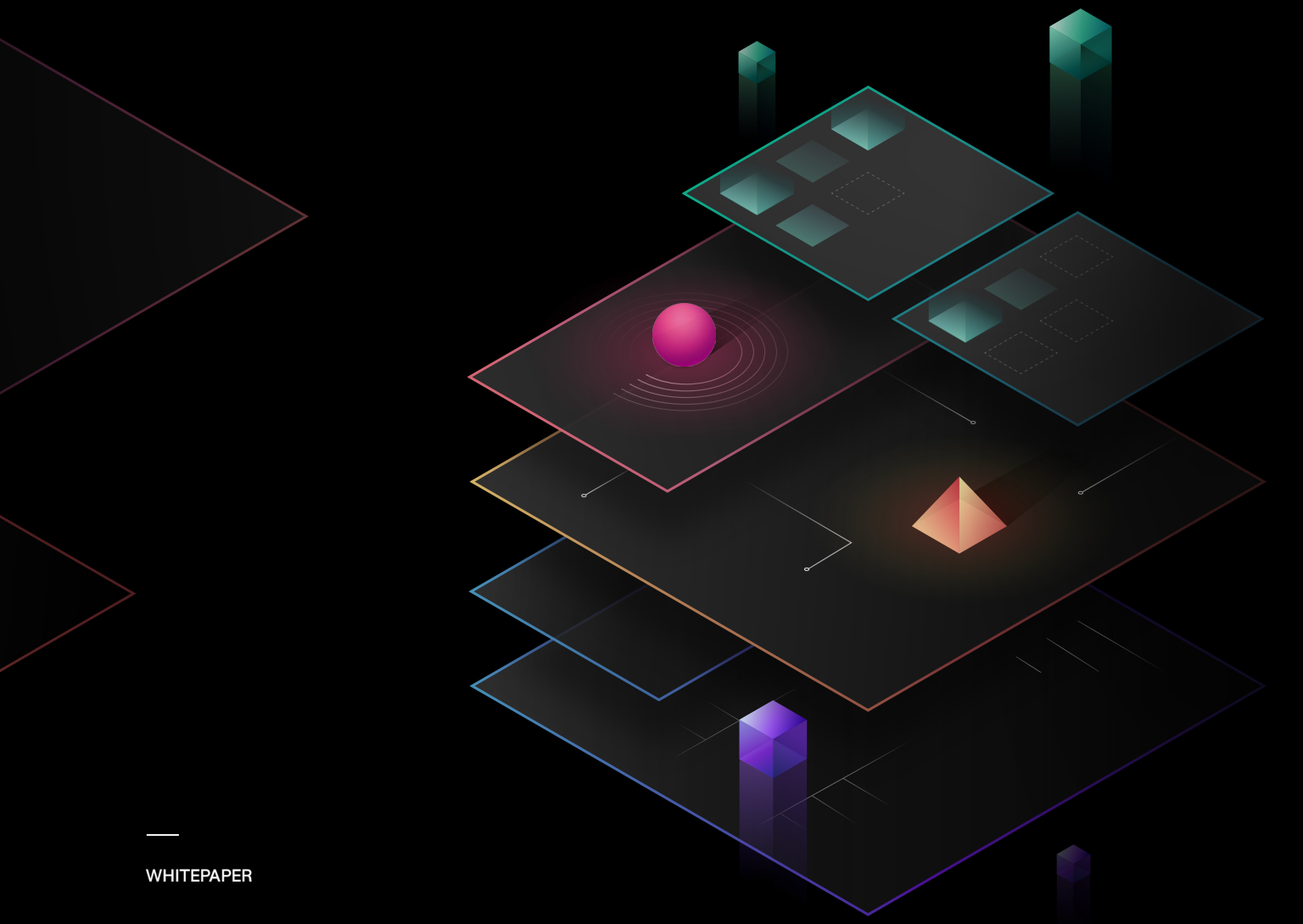
HashiCorp | aws

Unlocking the Cloud Operating Model

Achieving the fastest path to value
in a modern, hybrid-cloud datacenter



WHITEPAPER

Executive Summary

To thrive in an era of hybrid-cloud architecture, driven by digital transformation, Enterprise IT must evolve from ITIL-based gatekeeping to enabling shared self-service processes for DevOps excellence.

We talk to organizations of all sizes about their infrastructure plans and how they're adopting the cloud operating model as they navigate the transition to building new applications to differentiate their business. For most enterprises, digital transformation efforts mean delivering new business and customer value more quickly, and at a very large scale. The implication for Enterprise IT is navigating the shift from cost optimization models to speed-optimization models. The cloud is an inevitable part of this shift as it presents the opportunity to rapidly deploy on-demand services with limitless scale. To unlock the fastest path to value on AWS, enterprises must consider how to industrialize the application delivery process across each layer of the cloud, embracing the cloud operating model, and tuning people, process, and tools to it.

This white paper covers the motivations for adopting AWS, the transition from a static to dynamic environment on AWS, and lastly, a deep-dive into infrastructure, security, networking and application delivery in an AWS world.

Motivating Changes and Business Drivers for Cloud Adoption

Operational Costs

Operational costs are the costs of running your infrastructure. They include the unit price of infrastructure, matching supply and demand, investment risk for new applications, markets, and ventures, employing an elastic cost base, and building transparency into the IT operating model.

Workforce Productivity

Workforce productivity is how efficiently you are able to get your services to market. You can quickly provision AWS services, which increases your productivity by letting you focus on the things that make your business different; rather than spending time on the things that do not, like managing data centers. With over 170 services at your disposal, you eliminate the need to build and maintain these independently.

Cost Avoidance

Cost avoidance is setting up an environment that does not create unnecessary costs. Eliminating the need for hardware refresh and maintenance programs is a key contributor to cost avoidance. Customers tell us they are not interested in the cost and effort required to execute a big refresh cycle or data center renewal and are accelerating their move to the cloud as a result.

Operational Resilience

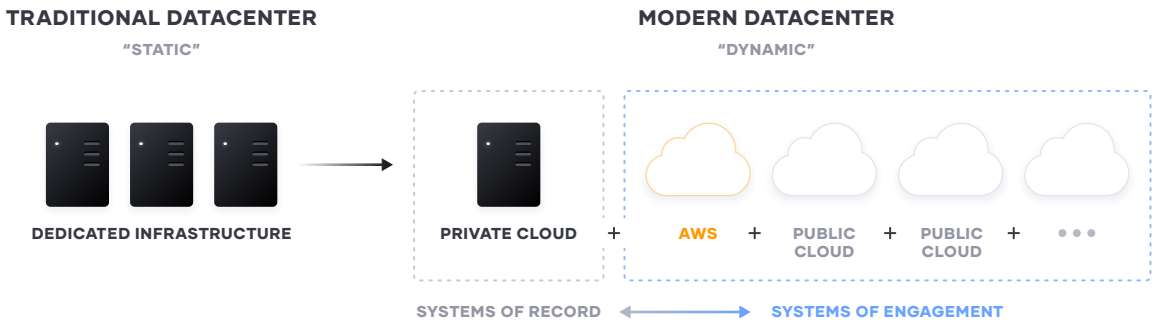
Operational resilience is reducing your organization's risk profile and the cost of risk mitigation. With 22 Regions comprising 69 Availability Zones (AZs), With AWS, you can deploy your applications in multiple regions around the world, which improves your uptime and reduces your risk-related costs. After migrating to AWS, our customers have seen improvements in application performance, better security, and reduction in high-severity incidents.

Business Agility

Business agility is the ability to react quickly to changing market conditions. Migrating to the AWS Cloud helps increase your overall operational agility. You can expand into new markets, take products to market quickly, and acquire assets that offer a competitive advantage. You also have the flexibility to speed up divestiture or acquisition of lines of business. Operational speed, standardization, and flexibility develop when you use DevOps models, automation, monitoring, and auto-recovery or high-availability capabilities.

Transitioning to a Hybrid-Cloud Datacenter

The transition to AWS, and hybrid-cloud environments is a generational shift for IT. This transition means shifting from largely dedicated servers in a private datacenter to a pool of compute capacity available on demand. While most enterprises began with one cloud provider, there are good reasons to use services from others and inevitably most Global 2000 organizations will use more than one, either by design or through mergers and acquisitions.



The cloud presents an opportunity for speed and scale optimization for new "systems of engagement" — the applications built to engage customers and users. These new apps are the primary interface for the customer to engage with a business, and are ideally suited for delivery in the cloud as they tend to:

- Have dynamic usage characteristics, needing to scale loads up and down by orders of magnitude during short time periods
- Be under pressure to quickly build and iterate. Many of these new systems may be ephemeral in nature, delivering a specific user experience around an event or campaign

For most enterprises though, these systems of engagement must connect to existing "systems of record" — the core business databases and internal applications, which often continue to reside on infrastructure in existing data centers. As a result, enterprises end up with a hybrid — a mix of multiple public and private cloud environments.

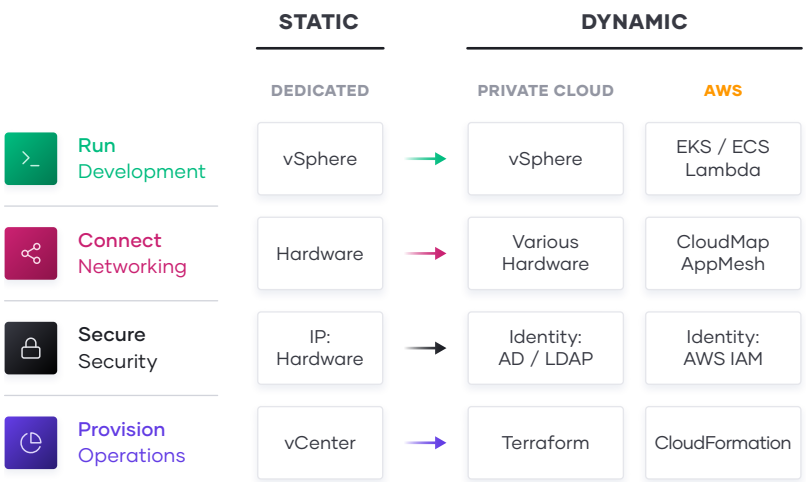
Implications of the Cloud Operating Model

The essential implication of the transition to the cloud is the shift from "static" infrastructure to "dynamic" infrastructure: from a focus on configuration, and management of a static fleet of IT resources, to provisioning, securing, connecting, and running dynamic resources on demand.

	STATIC		DYNAMIC
Run	Dedicated Infrastructure	→	Scheduled across the fleet
Connect	Host-based Static IP	→	Service-based Dynamic IP
Secure	High trust IP-based	→	Low trust Identity-based
Provision	Dedicated servers Homogeneous	→	Capacity on-demand Heterogeneous

Decomposing this implication, and working up the stack, various changes of approach are implied:

- **Provision.** The infrastructure layer transitions from running dedicated servers at limited scale to a dynamic environment where organizations can easily adjust to increased demand by spinning up thousands of servers and scaling them down when not in use. As architectures and services become more distributed, the sheer volume of compute nodes increases significantly.
- **Secure.** The security layer transitions from a fundamentally "high-trust" world enforced by a strong perimeter and firewall to a "low-trust" or "zero-trust" environment with no clear or static perimeter. As a result, the foundational assumption for security shifts from being IP-based to using identity-based access to resources. This shift is highly disruptive to traditional security models.
- **Connect.** The networking layer transitions from being heavily dependent on the physical location and IP address of services and applications to using a dynamic registry of services for discovery, segmentation, and composition. An enterprise IT team does not have the same control over the network, or the physical locations of compute resources, and must think about service-based connectivity.
- **Run.** The runtime layer shifts from deploying artifacts to a static application server to deploying applications with a scheduler atop a pool of infrastructure which is provisioned on-demand. In addition, new applications have become collections of services that are dynamically provisioned, and packaged in multiple ways: from virtual machines to containers.



To address these challenges those teams must ask the following questions:

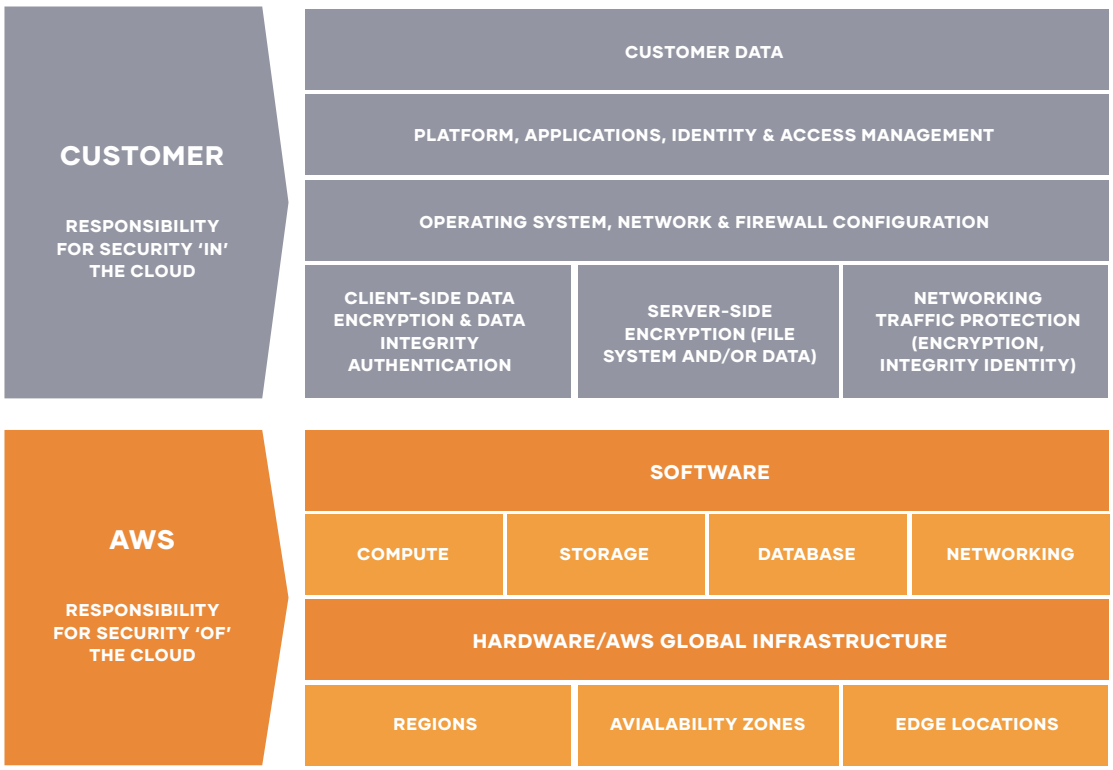
- **People.** How can we enable a team for a hybrid-cloud reality, where skills can be applied consistently regardless of target environment?
- **Process.** How do we position central IT services as a self-service enabler of speed, versus a ticket-based gatekeeper of control, while retaining compliance and governance?
- **Tools.** How do we best unlock the value of the available capabilities of the cloud providers in pursuit of better customer and business value?

AWS Shared Responsibility Model

Security and Compliance is a shared responsibility between AWS and the customer. This shared model can help relieve the customer's operational burden as AWS operates, manages and controls the components from the host operating system and virtualization layer down to the physical security of the facilities in which the service operates.

AWS responsibility "Security of the Cloud" - AWS is responsible for protecting the infrastructure that runs all of the services offered in the AWS Cloud. This infrastructure is composed of the hardware, software, networking, and facilities that run AWS Cloud services.

Customer responsibility "Security in the Cloud" – Customer responsibility will be determined by the AWS Cloud services that a customer selects. This determines the amount of configuration work the customer must perform as part of their security responsibilities.



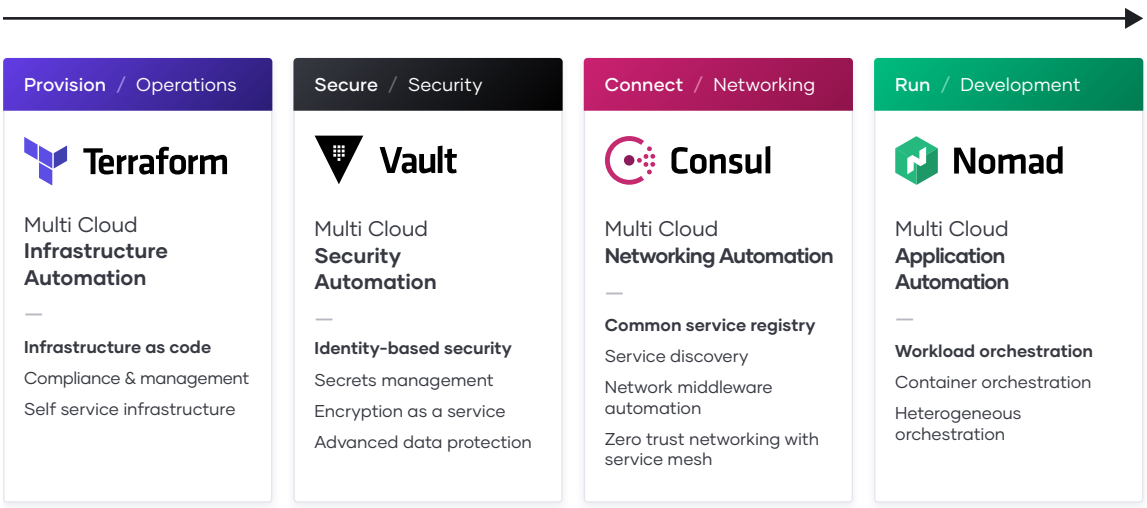
Unlocking the Cloud Operating Model on AWS

As the implications of the cloud operating model impact teams across infrastructure, security, networking, and applications, we see a repeating pattern amongst enterprises of establishing central shared services — centers of excellence — to deliver the dynamic infrastructure necessary at each layer for successful application delivery.

As teams deliver on each shared service to for the cloud operating model, IT velocity increases. The greater cloud maturity an organization has, the faster its velocity.

For self-service IT, the decoupling of the template-creation process and the provisioning process greatly reduces the time taken for any application to go live since developers no longer need to wait for operations approval, as long as they use a pre-approved template.

EXPANDING USE OF THE HASHICORP STACK INCREASES MATURITY AND VELOCITY FOR OUR CUSTOMERS

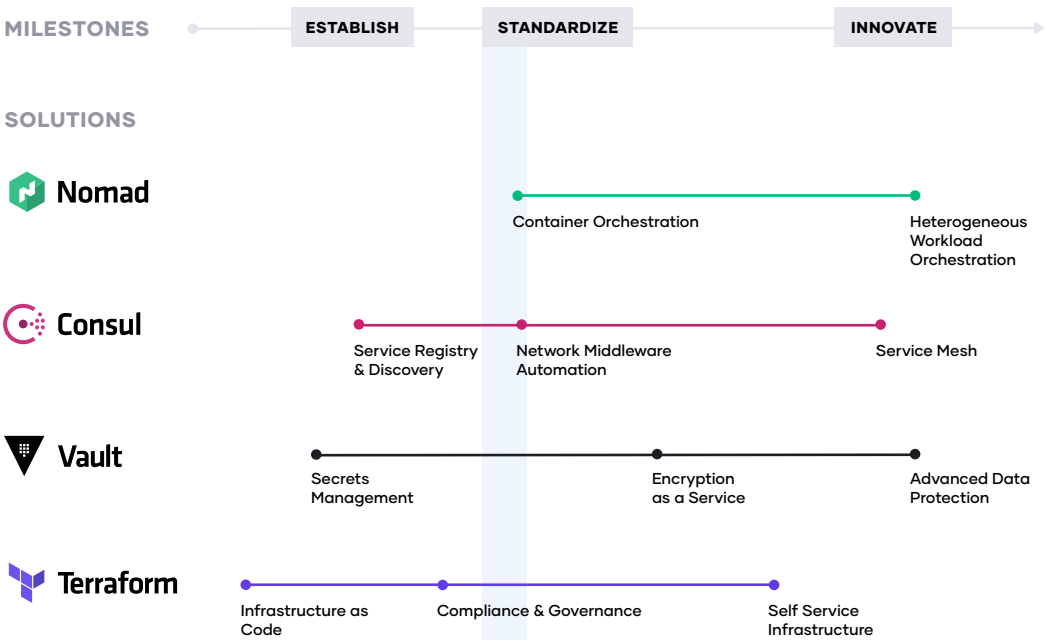


The typical journey we have seen customers adopt, as they unlock the cloud operating model on AWS, involves three major milestones

1. **Establish the cloud essentials** - As you begin your journey to the cloud, the immediate requirements are provisioning the cloud infrastructure typically by adopting infrastructure as code and ensuring it is secure with a secrets management solution. These are the bare necessities that will allow you to build a scalable and truly dynamic cloud architecture that is futureproof.

2. **Standardize on a set of shared services** - As cloud consumption starts to pick up, you will need to implement and standardize on a set of shared services so as to take full advantage of what the cloud has to offer. This also introduces challenges around governance and compliance as the need for setting access control rules and tracking requirements become increasingly important.

3. **Innovate using a common logical architecture** - As you fully embrace the cloud and depend on cloud services and applications as the primary systems of engagement, there will be a need to create a common logical architecture. This requires a control plane that connects with the extended ecosystem of cloud solutions and inherently provides advanced security and orchestration across services and multiple clouds.

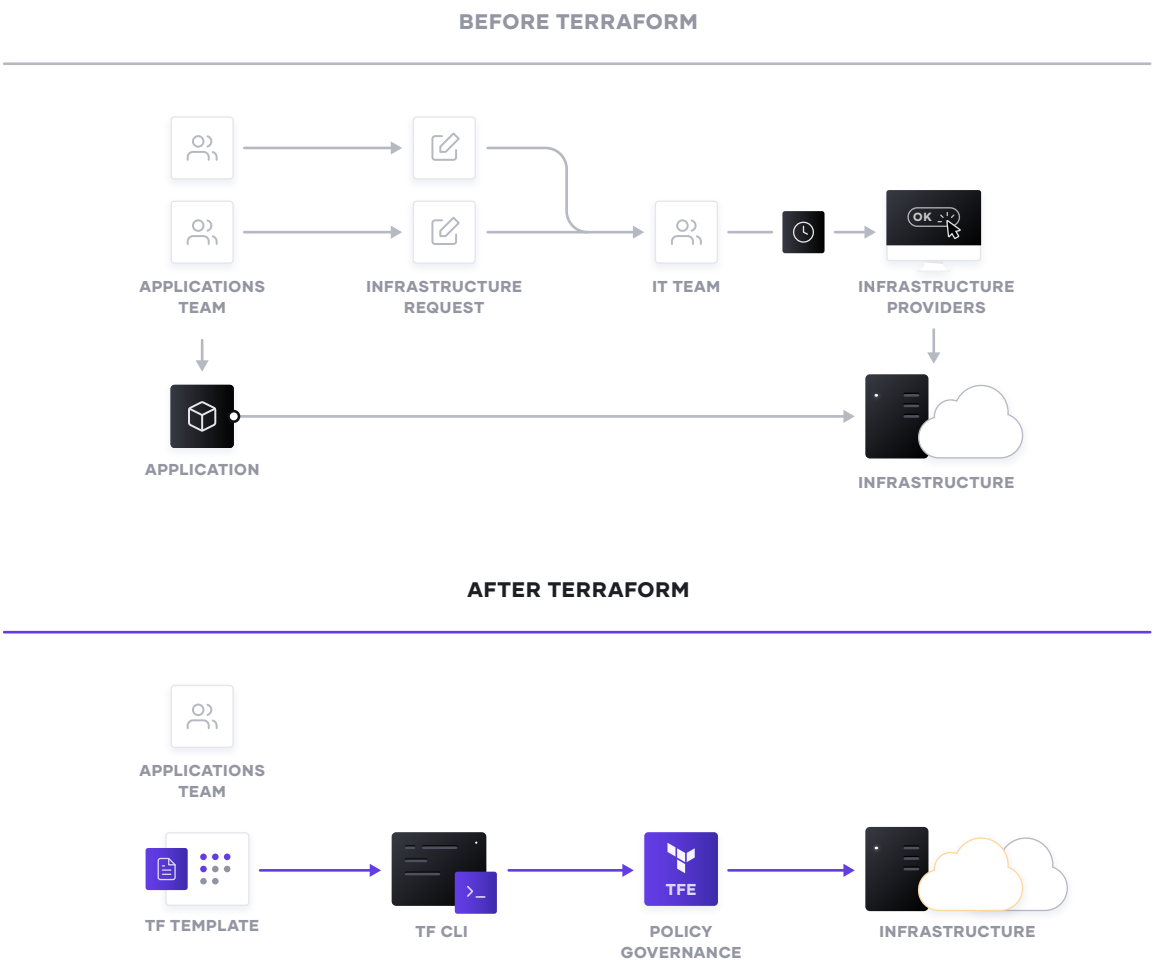


What follows is the step-by-step journey that we have seen organizations adopt successfully.

Step 1: Hybrid-Cloud Infrastructure Provisioning

The foundation for adopting the cloud is infrastructure provisioning. HashiCorp Terraform is one of the world's most widely used cloud provisioning products and can be used to provision infrastructure for any application using an array of providers for any target infrastructure.

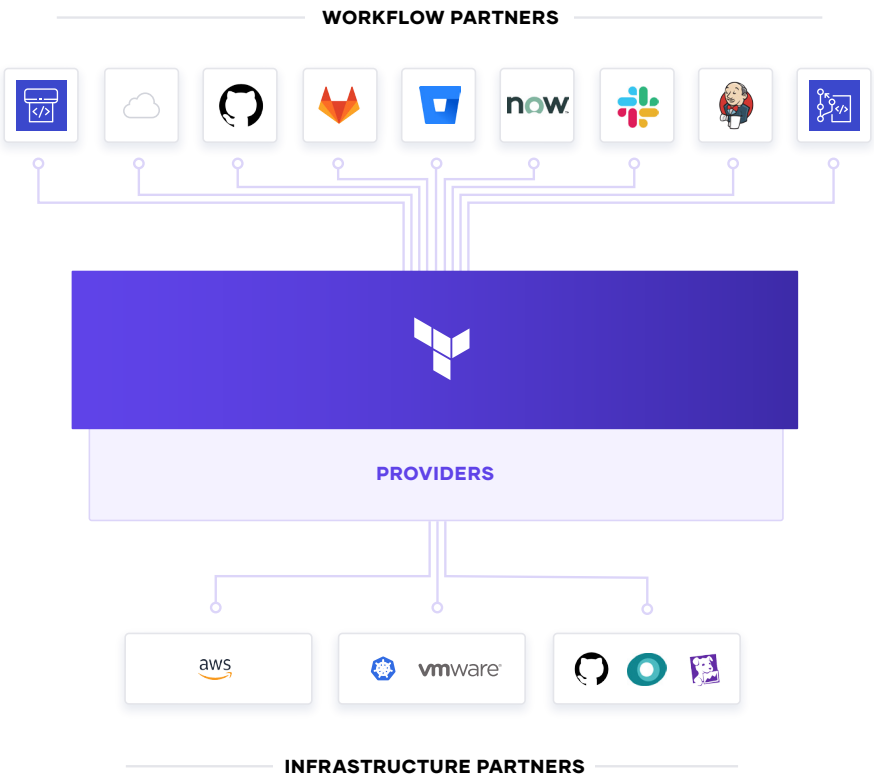
To achieve shared services for infrastructure provisioning, IT teams should start by implementing reproducible infrastructure as code practices, and then layering compliance and governance workflows to ensure appropriate controls.



Reproducible Infrastructure as Code

The first goal of a shared service for infrastructure provisioning is to enable the delivery of reproducible infrastructure as code, providing DevOps teams a way to plan and provision resources inside CI/CD workflows using familiar tools throughout.

DevOps teams can create Terraform templates that express the configuration of services from one or more cloud platforms. Terraform integrates with all major configuration management tools to allow fine grained provisioning to be handled following the provisioning of the underlying resources. Finally, templates can be extended with services from many other ISV providers to include monitoring agents, application performance monitoring (APM) systems, security tooling, DNS, and Content Delivery Networks, and more. Once defined, the templates can be provisioned as required in an automated way. In doing so, Terraform becomes the lingua franca and common workflow for teams provisioning resources across AWS, private clouds, or other infrastructure.



For self-service IT, the decoupling of the template-creation process and the provisioning process greatly reduces the time taken for any application to go live since developers no longer need to wait for operations approval, as long as they use a pre-approved template.

Compliance and Management

For most teams, there is also a need to enforce policies on the type of infrastructure created, how it is used, and which teams get to use it. HashiCorp's Sentinel policy as code framework provides compliance and governance without requiring a shift in the overall team workflow, and is defined as code too, enabling collaboration and comprehension for DevSecOps.

Without policy as code, organizations resort to using a ticket based review process to approve changes. This results in developers waiting weeks or longer to provision infrastructure and becomes a bottleneck. Policy as code allows us to solve this by splitting the definition of the policy from the execution of the policy.

Centralized teams codify policies enforcing security, compliance, and operational best practices across all cloud provisioning. Automated enforcement of policies ensures changes are in compliance without creating a manual review bottleneck.

Step 2: Hybrid-Cloud Security

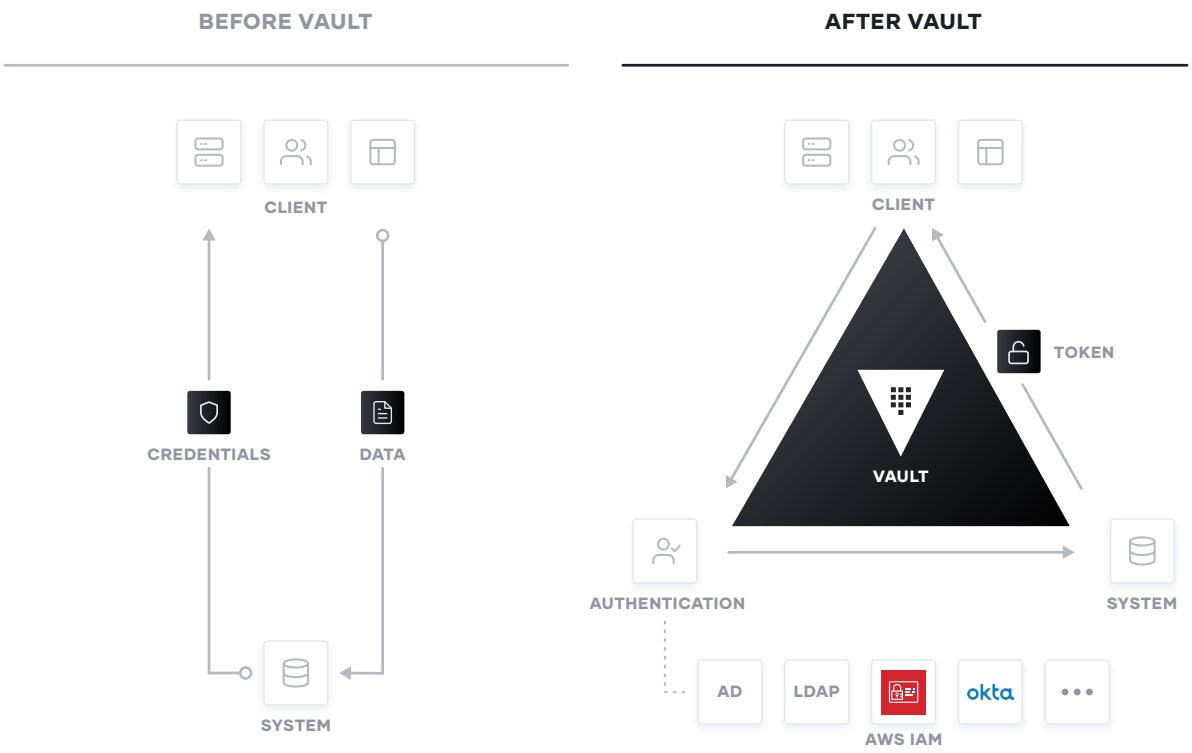
Dynamic cloud infrastructure means a shift from host-based identity to application-based identity, with low- or zero-trust networks across multiple clouds without a clear network perimeter.

In the traditional security world, we assumed high trust internal networks, which resulted in a hard shell and soft interior. With the modern "zero trust" approach, we work to harden the inside as well. This requires that applications be explicitly authenticated, authorized to fetch secrets and perform sensitive operations, and tightly audited.

HashiCorp Vault enables teams to securely store and tightly control access to tokens, passwords, certificates, and encryption keys for protecting machines and applications. This provides a comprehensive secrets management solution. Beyond that, Vault helps protect data at rest and data in transit. Vault exposes a high level API for cryptography for developers to secure sensitive data without exposing encryption keys. Vault also can act like a certificate authority, to provide dynamic short lived certificates to secure communications with SSL/TLS. Lastly, Vault enables a brokering of identity between different platforms, such as AWS IAM and Active Directory on premises to allow applications to work across platform boundaries.

Vault is widely used including across stock exchanges, large financial organizations, hotel chains, and everything in between to provide security in the cloud operating model.

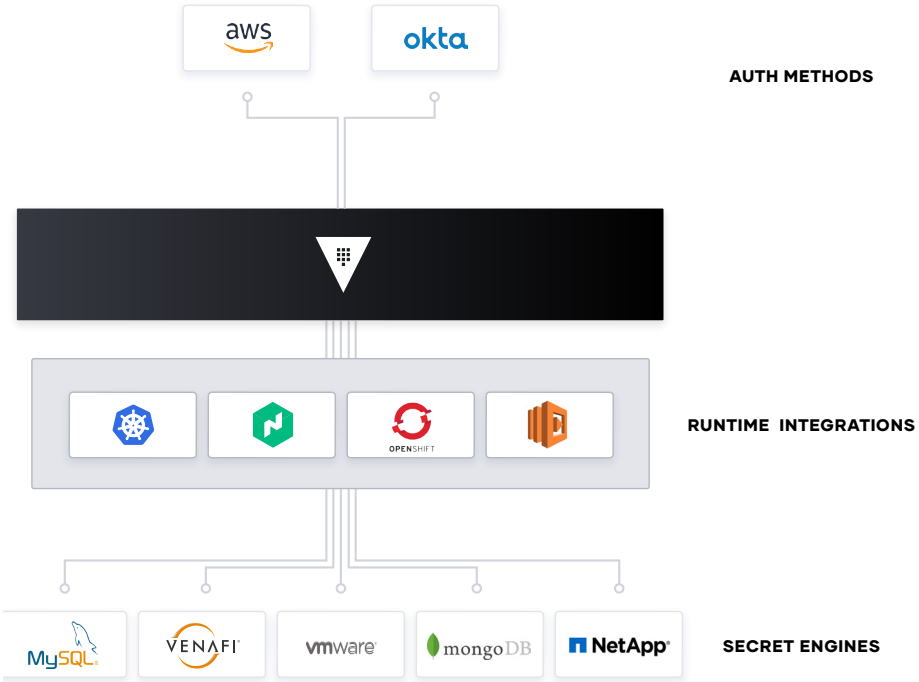
To achieve shared services for security, IT teams should enable centralized secrets management services, and then use that service to deliver more sophisticated encryption-as-a-service use cases such as certificate and key rotations, and encryption of data in transit and at rest.



Secrets Management

The first step in cloud security is typically secrets management: the central storage, access control, and distribution of dynamic secrets. Instead of depending on static IP addresses, integrating with identity-based access systems such as AWS IAM and AAD to authenticate and access services and resources is crucial.

Vault uses policies to codify how applications authenticate, which credentials they are authorized to use, and how auditing should be performed. It can integrate with an array of trusted identity providers such as cloud identity and access management (IAM) platforms, Kubernetes, Active Directory, and other SAML-based systems for authentication. Vault then centrally manages and enforces access to secrets and systems based on trusted sources of application and user identity.



Enterprise IT teams should build a shared service which enables the request of secrets for any system through a consistent, audited, and secured workflow.

Encryption as a Service

Additionally, enterprises need to encrypt application data at rest and in transit. Vault can provide encryption-as-a-service to provide a consistent API for key management and cryptography. This allows developers to perform a single integration and then protect data across multiple environments.

Using Vault as a basis for encryption-as-a-service solves difficult problems faced by security teams such as certificate and key rotation. Vault enables centralized key management to simplify encrypting data in transit and at rest across clouds and datacenters. This helps reduce costs around expensive Hardware Security Modules (HSM) and increases productivity with consistent security workflows and cryptographic standards across the organization.

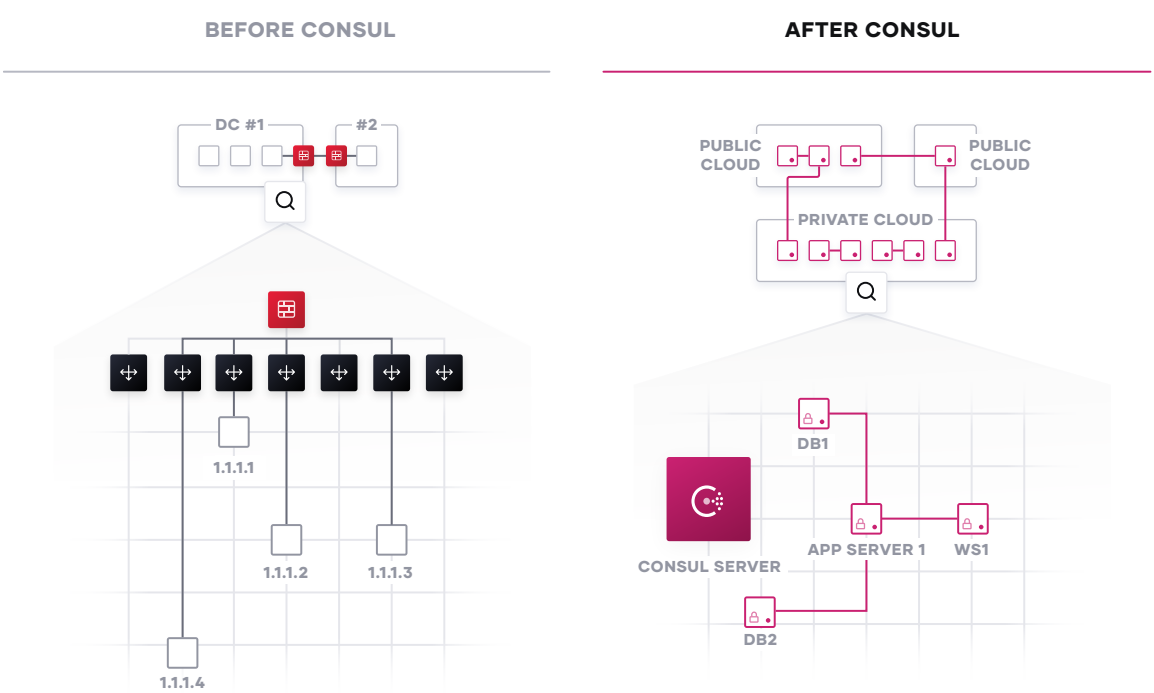
While many organizations provide a mandate for developers to encrypt data, they don't often provide the "how" which leaves developers to build custom solutions without an adequate understanding of cryptography. Vault provides developers a simple API that can be easily used, while giving central security teams the policy controls and lifecycle management APIs they need.

Step 3: Hybrid-Cloud Service Networking

The challenges of networking in the cloud are often one of the most difficult aspects of adopting the cloud operating model for enterprises. The combination of dynamic IP addresses, a significant growth in east-west traffic as the microservices pattern is adopted, and the lack of a clear network perimeter is a formidable challenge.

HashiCorp Consul provides a hybrid-cloud service networking layer to connect and secure services. Consul is a widely deployed product, with many customers running significantly greater than 100,000 nodes in their environments.

Networking services should be provided centrally, where by IT teams provide service registry and service discovery capabilities. Having a common registry provides a "map" of what services are running, where they are, and their current health status. The registry can be queried programmatically to enable service discovery or drive network automation of API gateways, load balancers, firewalls, and other critical middleware components. These middleware components can be moved out of the network by using a service mesh approach, where proxies run on the edge to provide equivalent functionality. Service mesh approaches allow the network topology to be simplified, especially for hybrid-cloud and multi-datacenter topologies.



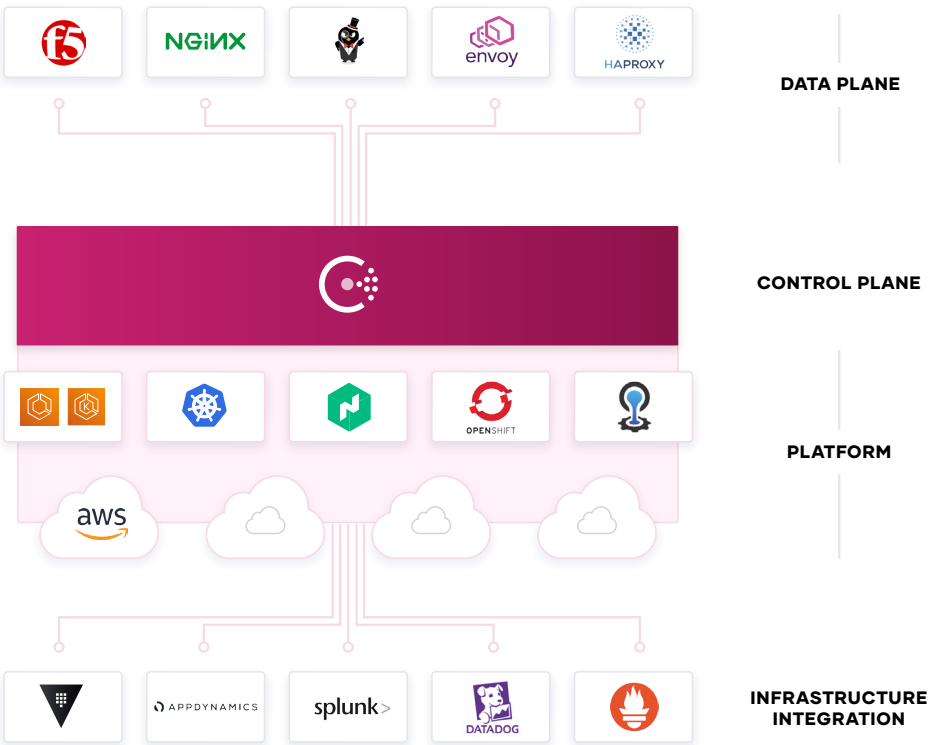
Service Registry & Discovery

The starting point for networking in the cloud operating model is typically a common service registry. This would integrate health checks and provide DNS and API interfaces to enable any service to discover and be discovered by other services.

Consul can be integrated with other services that manage existing north-south traffic such as a traditional load balancers, and distributed application platforms such as Kubernetes, to provide a consistent registry and discovery service across multi-data center, cloud, and platform environments.

Service Mesh

In a sophisticated environment, Consul provides a distributed service mesh to connect, secure, and configure services across any runtime platform and cloud. Consul provides an API driven control plane, which integrates with proxies such as Envoy, HAProxy, and Nginx for the data plane. This allows critical functionality like naming, segmentation and authorization, and routing to be handled by proxies at the edge rather than using centralized middleware.

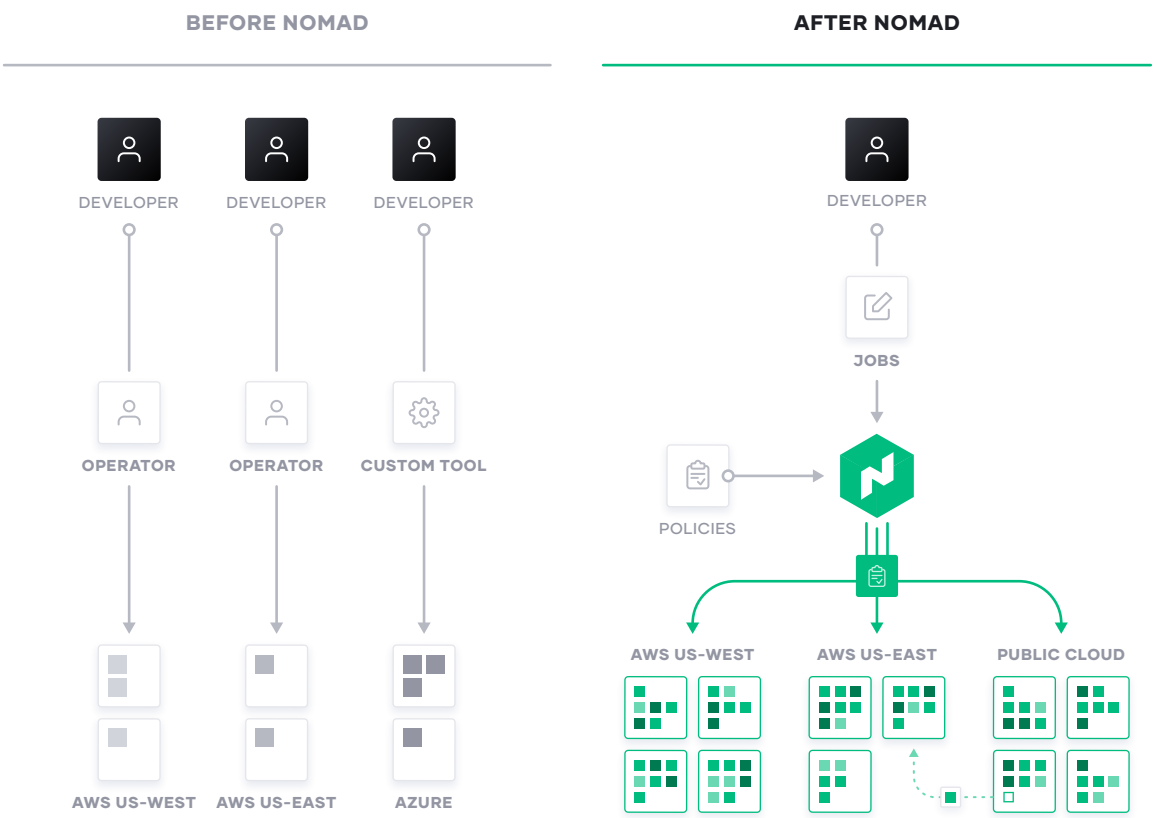


Consul enables fine grained service segmentation to secure service-to-service communication with automatic TLS encryption and identity-based authorization. Consul can be integrated with Vault for centralized PKI and certificate management. Service configuration is achieved through API-driven Key/Value store that can be used to easily configure services at runtime in any environment.

Step 4: Hybrid-Cloud Application Delivery

Finally, at the application layer, new apps are increasingly distributed while legacy apps also need to be managed more flexibly. HashiCorp Nomad provides a flexible orchestrator to deploy and manage legacy and modern applications, for all types of workloads: from long running services, to short lived batch, to system agents.

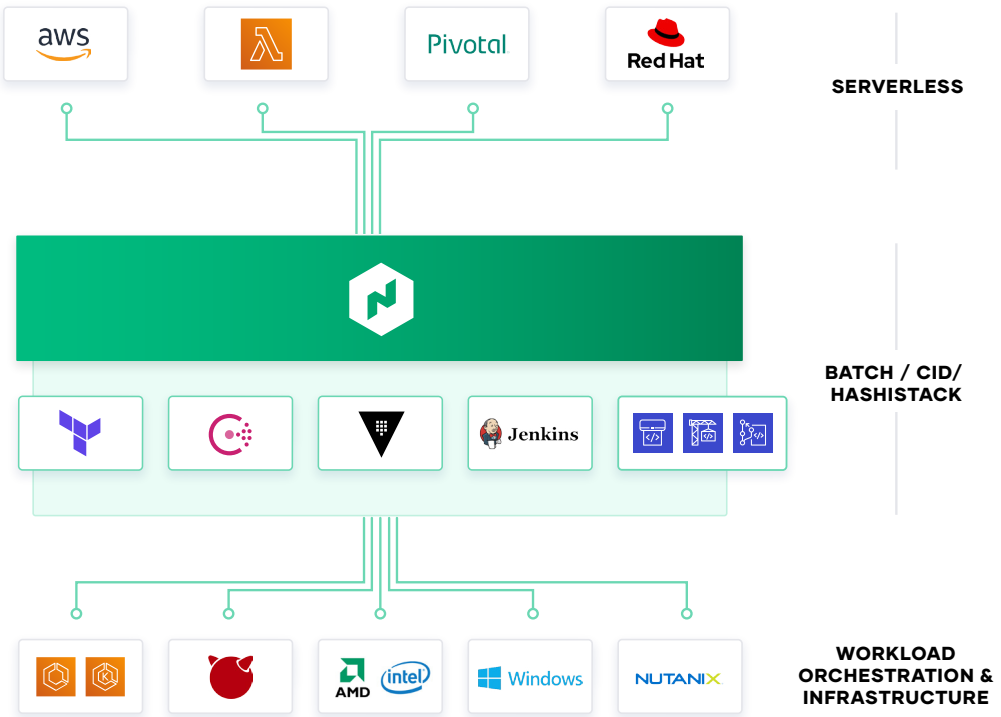
To achieve shared services for application delivery, IT teams should use Nomad in concert with Terraform, Vault, and Consul to enable the consistent delivery of applications on cloud infrastructure, incorporating necessary compliance, security, and networking requirements, as well as workload orchestration and scheduling.



Mixed Workload Orchestration

Many new workloads are developed with container packaging with the intent to deploy to Kubernetes or other container management platforms. But many legacy workloads will not be moved onto those platforms,nor will future Serverless applications Nomad provides a consistent process for deployment of all workloads from virtual machines, through standalone binaries, and containers, and provides core orchestration benefits across those workloads such as release automation, multiple upgrade strategies, bin packing, and resilience.

For modern applications — typically built in containers — Nomad provides the same consistent workflow at scale in any environment. Nomad is focused on simplicity and effectiveness at orchestration and scheduling, and avoids the complexity of platforms such as Kubernetes that require specialist skills to operate and solve only for container workloads.



Nomad integrates into existing CI/CD workflows to provide fast, automatic application deployments for legacy and modern workloads.

High Performance Compute

Nomad is designed to schedule applications with low latency across very large clusters. This is critical for customers with large batch jobs, as is common with High Performance Computing (HPC) workloads. In the million container challenge, Nomad was able to schedule one million instances of Redis across 5,000 machines in three data centers, in under 5 minutes. Several large Nomad deployments run at even larger scales.

Nomad enables high performance applications to easily use an API to consume capacity dynamically, enabling efficient sharing of resources for data analytics applications like Spark. The low latency scheduling ensures results are available in a timely manner and minimizes wasted idle resources.

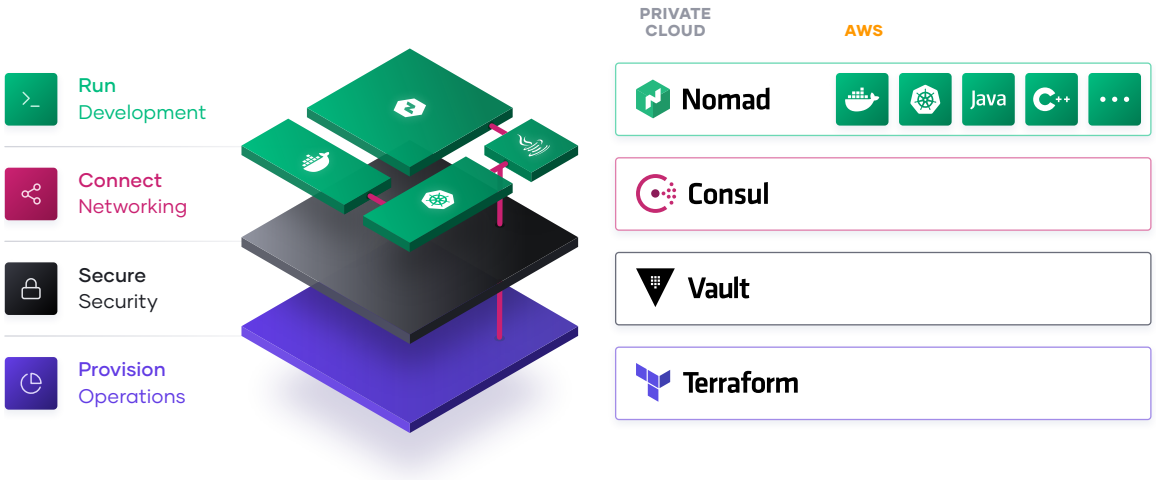
Multi-Datacenter Workload Orchestration

Nomad is multi-region and hybrid-cloud by design, with a consistent workflow to deploying any workload. As teams roll out global applications in multiple data centers, or across cloud boundaries, Nomad provides orchestrating and scheduling for those applications, supported by the infrastructure, security, and networking resources and policies to ensure the application is successfully deployed.

Step 5: Industrialized Application Delivery Process

Ultimately, these shared services across infrastructure, security, networking, and application runtime present an industrialized process for application delivery, all while taking advantage of the dynamic nature of each layer of the cloud.

Embracing the cloud operating model enables self-service IT, that is fully compliant and governed, for teams to deliver applications at increasing speed.



AWS Well-Architected Framework Pillars

The Well-Architected Framework has been developed to help cloud architects build secure, high-performing, resilient, and efficient infrastructure for their applications. Based on five pillars — operational excellence, security, reliability, performance efficiency, and cost optimization — the Framework provides a consistent approach for customers and partners to evaluate architectures, and implement designs that will scale over time.

Operational Excellence. The operational excellence pillar focuses on running and monitoring systems to deliver business value, and continually improving processes and procedures. Key topics include managing and automating changes, responding to events, and defining standards to successfully manage daily operations.

Security. The security pillar focuses on protecting information & systems. Key topics include confidentiality and integrity of data, identifying and managing who can do what with privilege management, protecting systems, and establishing controls to detect security events.

Reliability. The reliability pillar focuses on the ability to prevent, and quickly recover from failures to meet business and customer demand. Key topics include foundational elements around setup, cross project requirements, recovery planning, and how we handle change.

Performance Efficiency. The performance efficiency pillar focuses on using IT and computing resources efficiently. Key topics include selecting the right resource types and sizes based on workload requirements, monitoring performance, and making informed decisions to maintain efficiency as business needs evolve.

Cost Optimization. Cost Optimization focuses on avoiding un-needed costs. Key topics include understanding and controlling where money is being spent, selecting the most appropriate and right number of resource types, analyzing spend over time, and scaling to meet business needs without overspending.

Conclusion

In conclusion, the transition to AWS, and hybrid-cloud environments is a generational transition for IT. Each organization's cloud journey is unique. In order to successfully execute adoption, it is important to understand the current static-state, the target dynamic-state, and how to traverse this shift.

Leveraging the cloud operating model with HashiCorp and AWS will help accelerate transformation, the shifting of skills across people, and establish new self-service IT processes through the right tools designed for a new, dynamic environment.

Our collaboration with AWS

HashiCorp and AWS have a long standing relationship driven by both the companies and the community built around their tools. The proactive engagement of the open source community enables many HashiCorp products to have immediate support for new services provided by AWS. HashiCorp is an active member of the Amazon Partner Network and currently an Advanced Tier Technology Partner. Additionally, HashiCorp holds a DevOps & Containers Competency which certifies both technical proficiency and proven customer success. Organizations of all sizes trust HashiCorp tools to provision, secure, run, and connect any application running in AWS.

About HashiCorp

HashiCorp is the leader in multi-cloud infrastructure automation software. The HashiCorp software suite enables organizations to adopt consistent workflows to provision, secure, connect, and run any infrastructure for any application. HashiCorp open source tools Vagrant, Packer, Terraform, Vault, Consul, and Nomad are downloaded tens of millions of times each year and are broadly adopted by the Global 2000. Enterprise versions of these products enhance the open source tools with features that promote collaboration, operations, governance, and multi-data center functionality. The company is headquartered in San Francisco and backed by Mayfield, GGV Capital, Redpoint Ventures, True Ventures, IVP, and Bessemer Venture Partners. For more information, visit **www.hashicorp.com** or follow HashiCorp on Twitter **@HashiCorp**.