Unlocking the Cloud Operating Model: Infrastructure

Contents

Overview

To thrive in an era of multi-cloud architecture, driven by digital transformation, Enterprise IT must evolve from ITIL-based gatekeeping to enabling shared self-service processes for DevOps excellence.
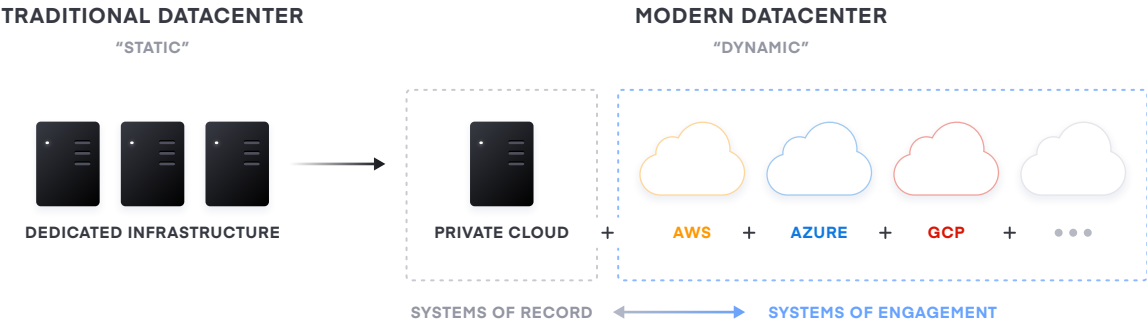
For most enterprises, the goals of digital transformation efforts mean delivering new business and customer value more quickly, and at a very large scale. The implication for Enterprise IT then is a shift from cost optimization to speed optimization. The cloud is an inevitable part of this shift as it presents the opportunity to rapidly deploy on-demand services with limitless scale.

To unlock the fastest path to value of the cloud, 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.

In this white paper, we look at the implications of the cloud operating model, and present solutions for IT teams to adopt this model across the infrastructure provisioning layer.

——

Transitioning to a Multi-Cloud Datacenter

The transition to cloud, and multi-cloud, environments is a generational transition 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.

TRADITIONAL DATACENTER
"STATIC"

DEDICATED INFRASTRUCTURE

MODERN DATACENTER
"DYNAMIC"

PRIVATE CLOUD + AWS + AZURE + GCP + • • •

SYSTEMS OF RECORD ◄———► SYSTEMS OF ENGAGEMENT

- 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.
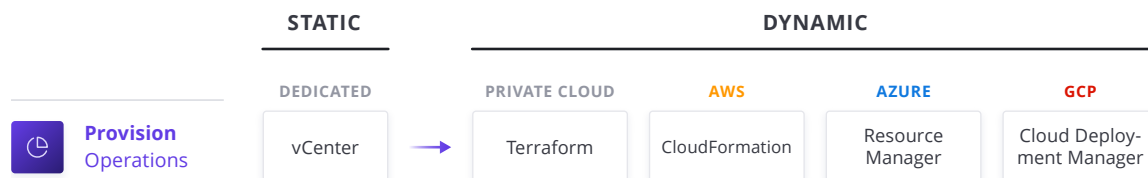
Implications of the Cloud Operating Model

The essential implications 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
Provision	**Dedicated** servers Homogeneous	Capacity **on-demand** Heterogeneous
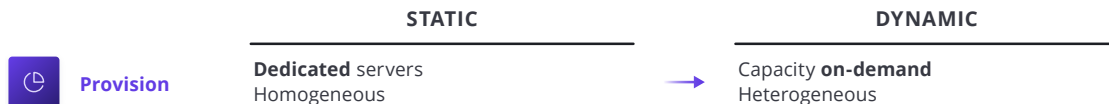
Decomposing these implications, various changes of approach are implied: The infrastructure layer transitions from running dedicated servers at limited scale to a dynamic environment where organizations can easily adjust to increased demand by spinning up thousands of servers and scaling them down when not in use. As architectures and services become more distributed, the sheer volume of compute nodes increases significantly.

Additionally, each cloud provider has its own solution to these challenges. For enterprise IT teams, these shifts in approach are compounded by the realities of running on hybrid- and multicloud infrastructures and the varying tools each technology provides.

	STATIC	DYNAMIC		
	DEDICATED	PRIVATE CLOUD	AWS	AZURE	GCP
Provision Operations	vCenter	Terraform	CloudFormation	Resource Manager	Cloud Deployment Manager

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

- **People.** How can we enable a team for a multi-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?
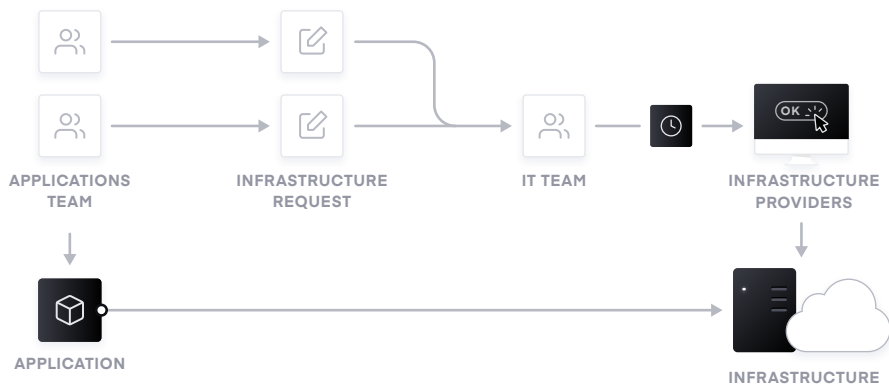
Unlocking the Cloud Operating Model

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.
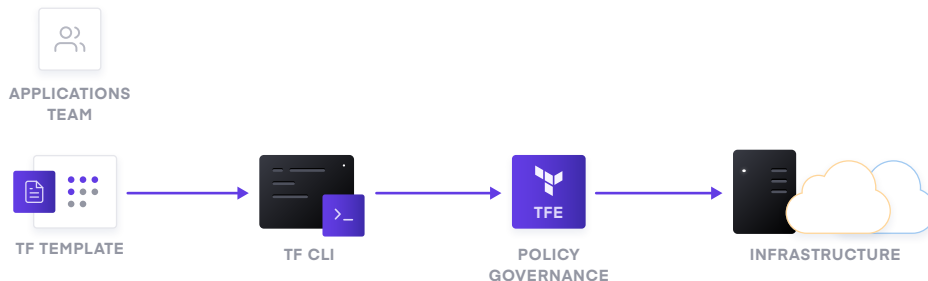
Multi-cloud Infrastructure Provisioning

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

BEFORE TERRAFORM

APPLICATIONS TEAM — INFRASTRUCTURE REQUEST — IT TEAM — INFRASTRUCTURE PROVIDERS

APPLICATION — INFRASTRUCTURE

AFTER TERRAFORM

APPLICATIONS TEAM

TF TEMPLATE — TF CLI — POLICY GOVERNANCE — 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 public and private cloud.

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.

——

Summary

A common cloud operating model is an inevitable shift for enterprises aiming to maximize their digital transformation efforts. The HashiCorp suite of tools seeks to provide solutions for each layer of the cloud to enable enterprises to make this shift to the cloud operating model.

Enterprise IT needs to evolve away from ITIL-based control points with its focus on cost optimization, toward becoming self-service enablers focused on speed optimization. It can do this by delivering shared services across each layer of the cloud designed to assist teams deliver new business and customer value at speed.