# HashiCorp Nomad

## STORAGE DAY

HashiCorp Nomad Technical Introduction





LIN•BIT Brian Hellman COO



HashiCorp Thomas Harris Sr. Solutions Engineer





## **Introducing Nomad**

Cloud application automation





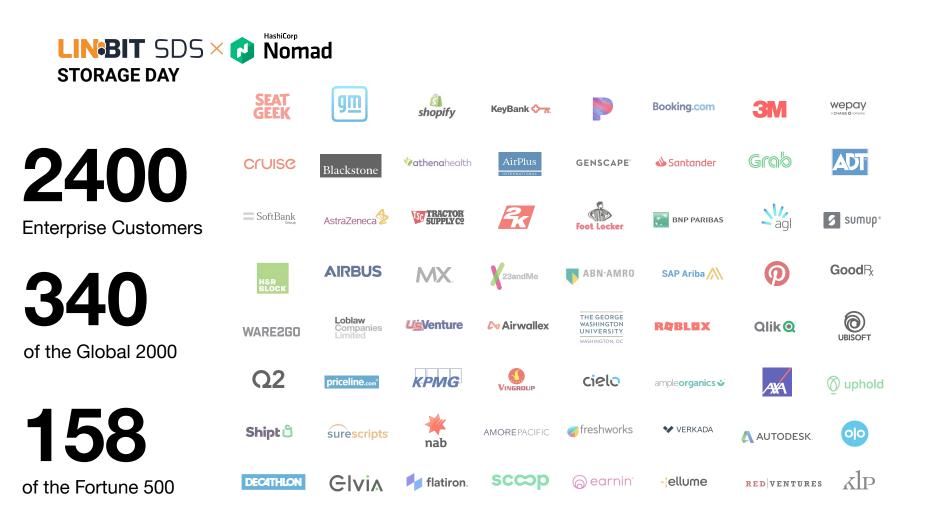
#### Leading cloud infrastructure automation

Our software stack enables the **provisioning**, **securing**, **connecting**, and **running** of apps and the infrastructure to support them.

#### We unlock the cloud operating model for

every business and enable their digital transformation strategies to succeed.









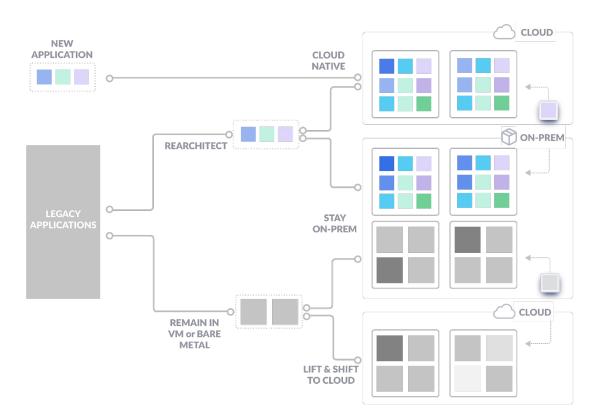
# Why Orchestration?





#### The Move to Containers and Clouds

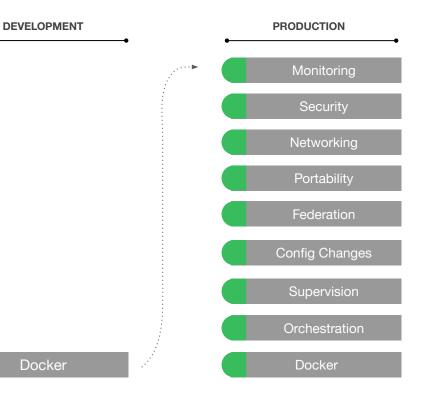
- Adopt containers and microservices for new applications
- Re-architect & migrate existing applications
- Segment of legacy applications remains the same





# **Complexity** is a top challenge in using and deploying containers

40% of surveyed organizations cited "complexity" as the number one challenge in container deployment.

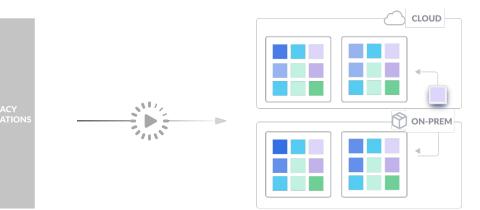


https://www.cncf.io/blog/2018/08/29/cncf-survey-use-of-cloud-native-technologies-in-production-ha s-grown-over-200-percent/



#### Challenges of Moving to Containers

- Lack of budget or time to refactor existing applications
- Increased complexity to support mixed systems and workflows
- Limited time to meet IT/Business requirements - Incremental vs Overhaul







# Why Nomad?





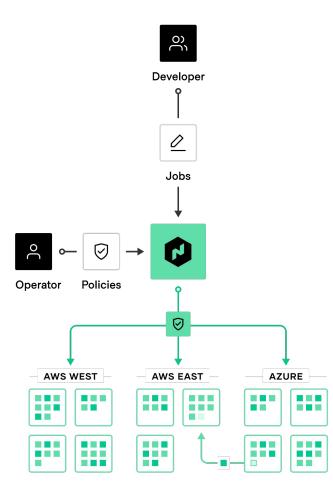
### Nomad

Provides the foundation for cloud application automation by enabling simple and flexible workload orchestration in the cloud operating model.

- Simple Container Orchestration to easily deploy and manage containers at any scale.
- Solution Non-Containerized Application Orchestration to modernize legacy applications without rewrites.



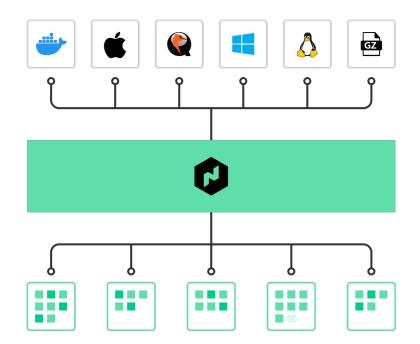




LINBIT SDS × C Nomad STORAGE DAY

#### Guiding principle: Orchestrate any application

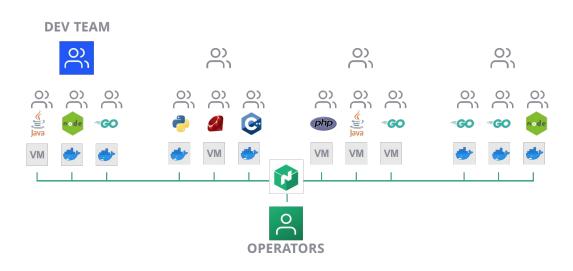
- Bringing modern orchestration benefits to all, containerized, non-containerized and batch applications.
- A simple, lightweight layer that can be integrated with any existing infrastructure.
- A single, unified workflow to accelerate incremental application modernization.





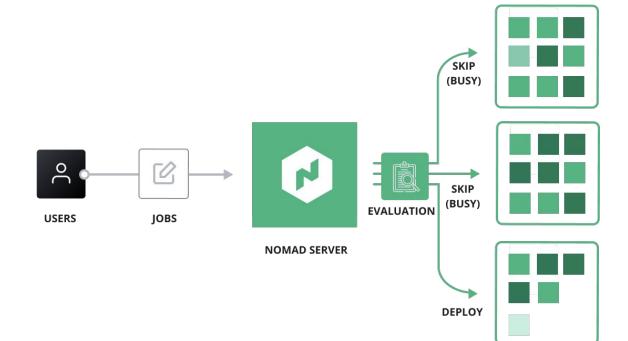
# Easy to deploy and maintain by system operators

- Accelerate time-to-cluster by using your normal software deployment methodology
- **Reduce cost** by utilising the skills you already have to run and maintain the clusters.
- **Don't reinvent the wheel** by using your current infrastructure.





Automated Deployment Workflow

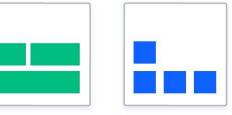




# Improved resource utilization

Improve resource utilization by densely scheduling applications over underutilized resources.

#### WITHOUT BIN PACKING



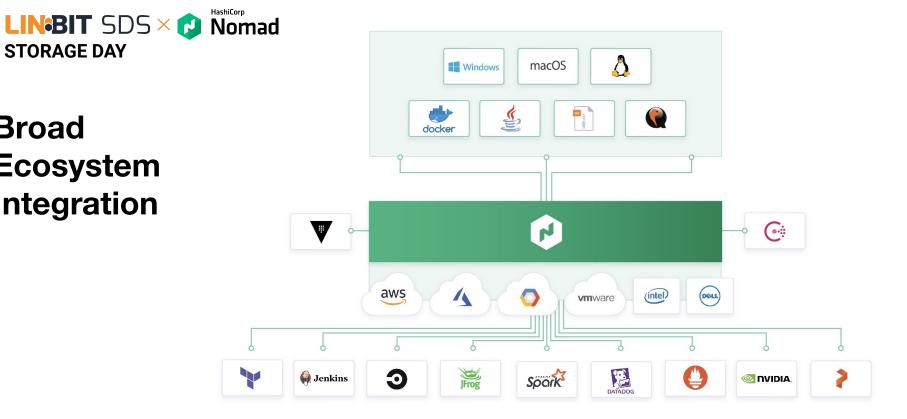


#### WITH BIN PACKING



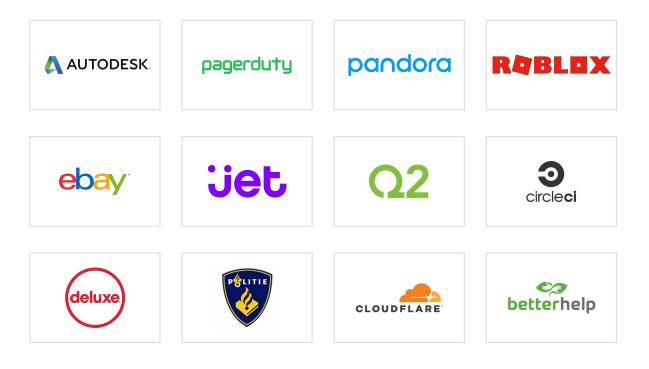
**Broad Ecosystem** Integration

**STORAGE DAY** 





#### Public Users of Nomad



## How organizations leverage Nomad



Multiple orchestrators for multiple use cases

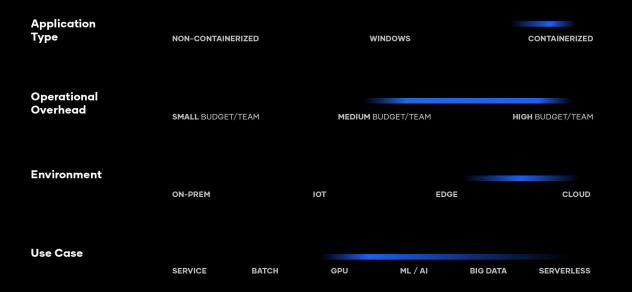


Based on interview data from HashiCorp users and customers running self-hosted Kubernetes and do not apply to managed Kubernetes services (EKS, AKS, GKE, etc.)





Multiple orchestrators for multiple use cases



Based on interview data from HashiCorp users and customers running self-hosted Kubernetes and do not apply to managed Kubernetes services (EKS, AKS, GKE, etc.)

#### How organizations leverage Nomad and Kubernetes



Multiple orchestrators for multiple use cases

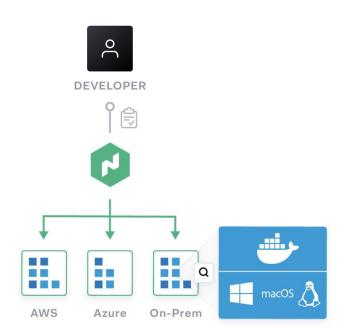


Based on interview data from HashiCorp users and customers running self-hosted Kubernetes and do not apply to managed Kubernetes services (EKS, AKS, GKE, etc.)



#### **Container Orchestration**

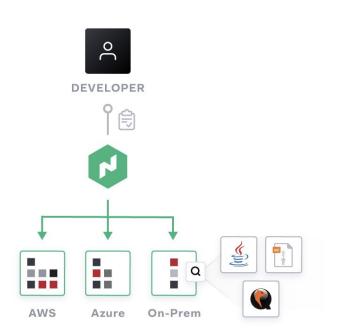
- Accelerate time-to-value with fast time-to-deploy and seamless integration with Consul and Vault
- Increase adoption by providing simplest onboarding experience to developers
- Reduce operational overhead with a lightweight, robust, and easy-to-manage platform





#### Non-Containerized Application Orchestration

- Accelerate time-to-value by integrating Nomad with existing infrastructure
- **Reduce cost** by densely scheduling applications to underutilized resources.
- Zero downtime deployment by bringing the benefits of modern deployment strategies to legacy applications





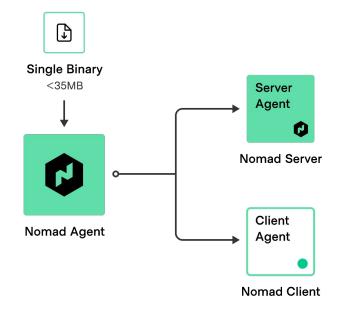


## **Technical Overview**





### Lightweight single binary



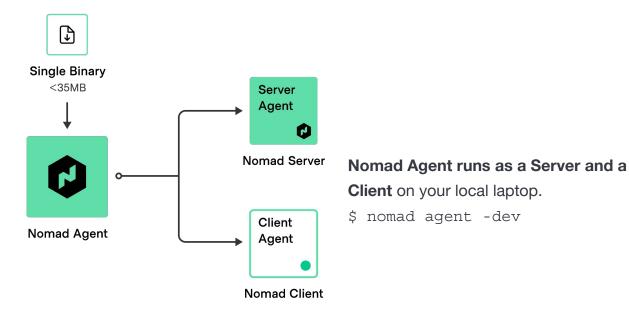
**Nomad Server** forms the control plane for scheduling.

\$ .nomad -config=server.hcl

Nomad Client runs on the node which registers with the servers, watching for any work to be assigned and execute tasks \$ .nomad -config=client.hcl

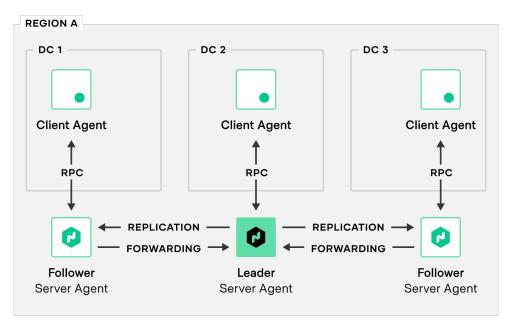


#### **Nomad Architecture - Laptop**



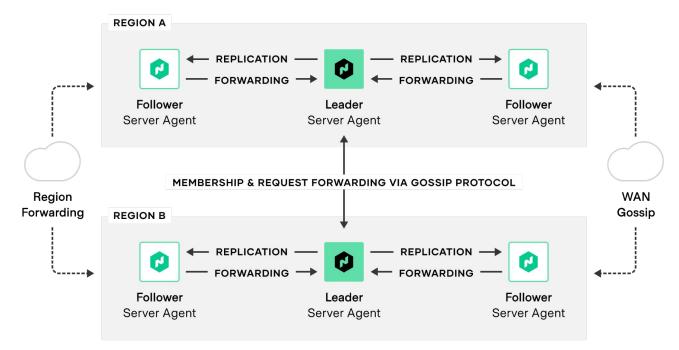


#### Nomad architecture - single region



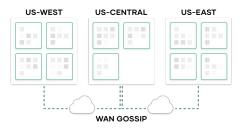


#### Nomad architecture - multi-region



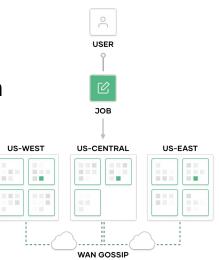
## LINBIT SDS × Normad STORAGE DAY

The first and only orchestrator on the market with complete and fully-supported federation capabilities for production.



#### **Multi-Cluster Visibility**

- Single command to join Nomad clusters together.
- Single interface to monitor job status across any cluster.



#### **Multi-Cluster Deployment (ENT)**

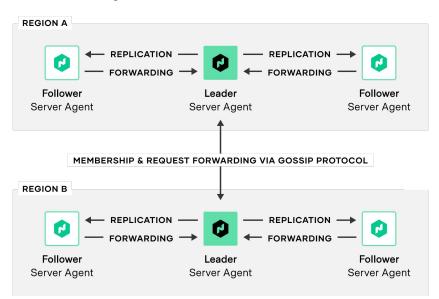
- Single job spec to deploy applications across multiple clusters.
- Orchestrate configurable rollout and rollback across multiple clusters.
- New in 0.12



# Native federation capabilities

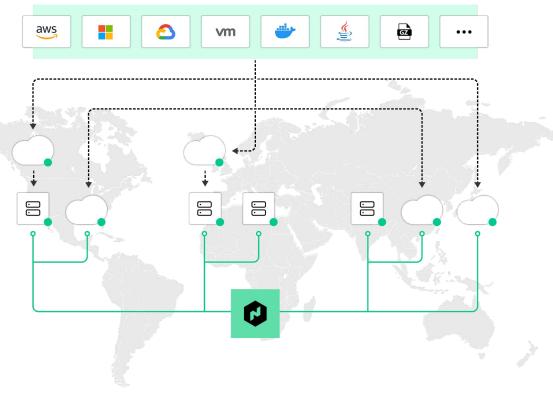
- Federating is as simple as joining the servers.
- A single command to join a server in a remote region.
  - \$ nomad server join
  - X.X.X.X:XXXX
- Submit jobs to any region or query the state through any server in any region transparently.

#### Nomad Multi-Region Architecture





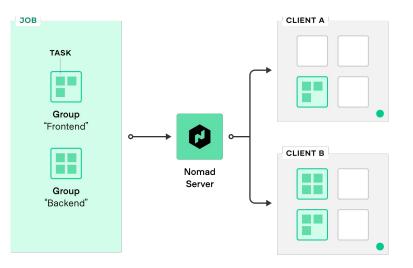
Single monitoring and control plane with federation





## **Core Concepts**

- TaskThe smallest unit of scheduling work. It<br/>could be a Docker container, a Java<br/>application, or batch processing.
- Group A series of tasks that should be co-located on the same Nomad client. Tightly-coupled tasks in the same group can share the same network/storage
- Job The declarative that defines the deployment rules for applications
- Application The instance of a task group that are running on client



## LINBIT SDS X D Nomad STORAGE DAY

#### Application deployment as code

- Nomad Job Declarative specification using HCL (like Terraform) or JSON.
- Set the deployment rules for applications fast and intuitively. Define tasks, images, resources, priorities, constraints, service registrations, secrets and other information required to deploy the application.

```
TERMINAL
job "my_job" {
  region = "us"
  datacenters = ["us-west-1", "us-east-1"]
  type = "service"
  group "web" {
    count = 5
    task "frontend" {
      driver = "docker"
      config {
        image = "hashicorp/web-frontend"
      resources
               = 500 # MHz
        cpu
        memory = 128 # MB
        network {
          mbits = 100
          port "http" {}
          port "https"
            static = 443
```

## LINBIT SDS × C Nomad STORAGE DAY

#### Intuitive deployment rules

#### job

# define the deployment rules for applications

#### \\_ group

# defines a series of tasks that should be co-located on the same Nomad client

#### \\_ task

# defines a command, service, application or "set of work" to execute, such as a docker container, webapp or batch processing. Tasks are executed by their driver

```
TERMINAL
      job "my_job" {
\odot
        region = "us"
        datacenters = ["us-west-1", "us-east-1"]
        type = "service"
        group "web" {
\odot
          count = 5
\odot
          task "frontend" {
            driver = "docker"
            config {
              image = "hashicorp/web-frontend"
            resources
                      = 500 \# MHz
               cpu
              memory = 128 # MB
              network {
                mbits = 100
                port "http" {}
                port "https"
                   static = 443
```

## LINBIT SDS × P Nomad STORAGE DAY

#### Enable flexibility with extensible drivers

- **Task Drivers** execute tasks on the Nomad Client and provide resource isolation.
- First-class support of a broad set of workloads across all major operating systems.



#### $\mathbf{O}$

 $\odot$ 

```
job "my job" {
  region = "us"
  datacenters = ["us-west-1", "us-east-1"]
  type = "service"
  group "web" {
    count = 5
    task "frontend" {
      driver = "docker"
      config {
        image = "hashicorp/web-frontend"
      resources {
        сри
               = 500 \# MHz
        memory = 128 \# MB
        network {
          mbits = 100
          port "http" {}
          port "https"
            static = 443
```

TERMINAL

#### LINBIT SDS × P Nomad STORAGE DAY

# Deploy interdependent applications

Deploy interdependent applications in their sequential orders with lifecycle hooks in the Task stanza:

- Startup & Initialization
  - PreStart
  - PostStart
- Sidecar
  - Boolean parameter
  - Deploy long-lived tasks within batch job
  - Run for duration of allocation
- Post deployment
  - PostStop





# Schedulers to run all types of workloads

• **Batch Scheduler** is optimized for fast placement for short-lived workload.

Example: Daily reports, transactions, billing invoices.

Service Scheduler is optimized for long-running workloads.

**Example:** Business-critical applications, customer facing webapps, database.

- **System Scheduler** is optimized for background tasks. *Example:* Logging/monitoring, security, background processes.
- **System Batch** is optimized for background tasks. *Example:* Same as System but in Batch Style.

```
TERMINAL
job "my_job" {
 region = "us"
 datacenters = ["us-west-1", "us-east-1"]
 type = "service"
 group "web" {
   count = 5
   task "frontend" {
     driver = "docker"
     config {
        image = "hashicorp/web-frontend"
      resources
               = 500 \# MHz
        cpu
        memory = 128 # MB
       network {
          mbits = 100
          port "http" {}
          port "https"
            static = 443
```

 $\bigcirc$ 

# Robust application update strategy

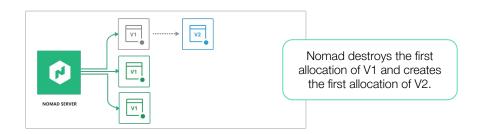
Automate job update and migration to minimize down time.

- Rolling Updates
- Blue/Green deployments
- Canaries deployments
- Updates can be gated on Consul health checks and automatically reverted.

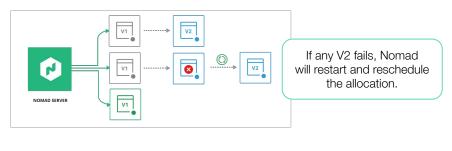
					TERMIN	AL
	job "my_job" {					
	update {					
	<pre>max_parallel</pre>	=	3			
	health_check	=	"check	s"		
	min_healthy_time	=	"10s"			
	healthy_deadline	=	"10m"			
	auto_revert	=	true			
>	canary	=	1			
	stagger	=	"30s"			
	}					
	}					



## Zero downtime deployment



- Reduce downtime disruption by leveraging modern deployment strategies, such as rolling updates, blue/green, and canary deployments
- **Example:** Rolling Deployment

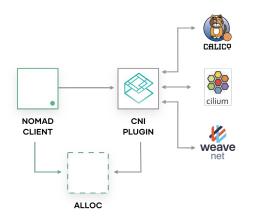


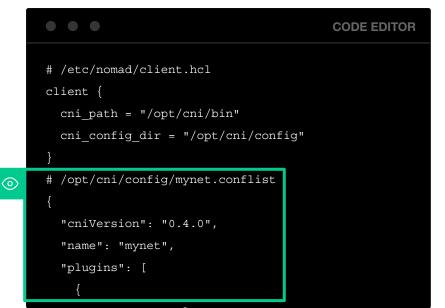


### **LINBIT** SDS × Normad STORAGE DAY Modern networking with simplicity

Deploy containerized applications on Nomad with their desired network configurations from any third-party vendors via the Container Networking Interface (CNI) plugin system.

- Operators set directories of CNI configurations,
- Nomad fingerprints and validate CNI networks on the host.
- Developers define network mode as "CNI/xxxx" in the job spec to use the CNI plugins.





# Multi-interface networking

Allow applications to bind to multiple network interfaces to support communications across multiple networks.

- Define multiple host networks in Nomad Client configuration.
- Define host\_network to the port block to select a named host network.
- New network env variables to support multiple interfaces.

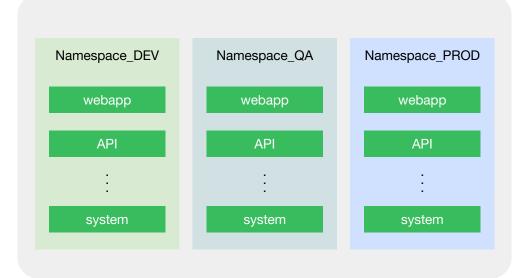
**CODE EDITOR** client { host\_network "public" { cidr = "203.0.113.0/24"// lb.nomad network { mode = "bridge" port "admin" { host network = "private" to = 8080port "https" { host network = "public" to = 443



### Namespaces

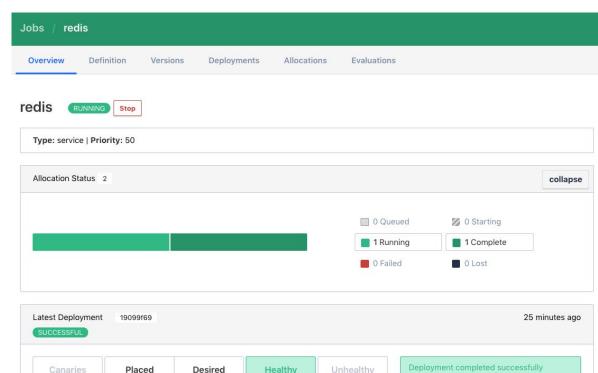
Allow many teams and projects to share a single multi-region Nomad deployment without conflict.

- ACL policies provide enforcement of namespaces.
- Job IDs are required to be unique within namespaces but not across namespaces.
- Namespaces are automatically replicated across regions for easy, centralized administration at global scale.





View job with deployment status



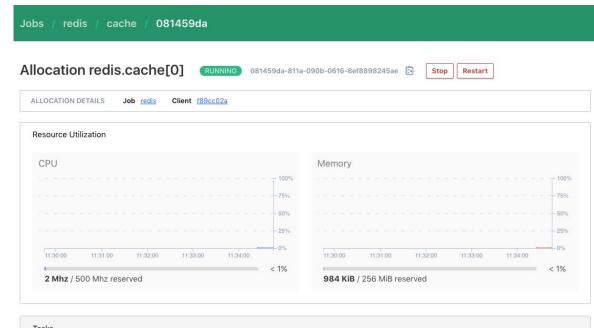
0

1

1



View task status



Tasks						
Name 🌲	State	Last Event	Time	Addresses	CPU	Memory
redis	running	Task started by client	Jan 10, '20 11:05:30 -0500	<b>db:</b> <u>127.0.0.1:26203</u>		

## Secrets management with Vault

- Automatic Vault token retrieval.
- Automatic Vault token renewal.
- Automatic secret retrieval and renewal via template stanza.

retrieve



## TERMINAL job "my\_job" { group "example" { task "server" { vault { policies = ["cdn", "frontend"] change mode = "signal" change\_signal = "SIGUSR1"



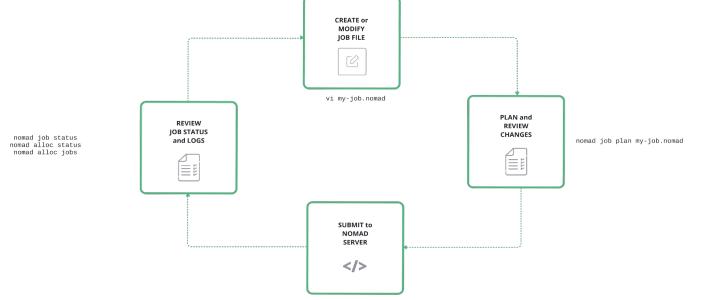
# Service discovery with Consul

Built-in service discovery, registration, and health check monitoring for all applications deployed under Nomad.

#### • dc1 Services Nodes ACL Intentions Documentation Settings Key/Value Services Q Passing (29) A Warning (0) Critical (0) All (29) Search by name Service Node Health Tags 🚺 blue-green 10 global web-server 1 consul 6 nomad rpc 2 http nomad-client web-server 10 global web-server



# Consistent workflow for job lifecycle management





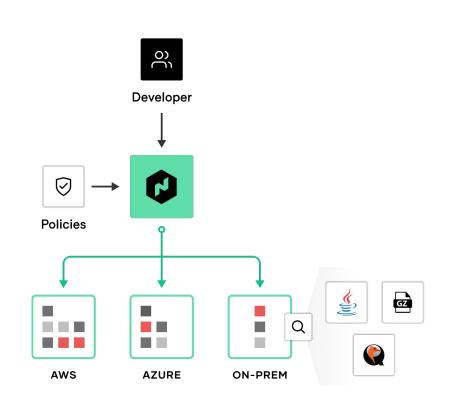


# Orchestrate non-containers as containers



## Non-Containerized Application Orchestration

- Achieve immediate results by integrating Nomad
   with existing infrastructure.
- **Reduce cost** by densely scheduling applications to underutilized resources.
- Zero downtime deployment by bringing the benefits of modern deployment strategies to legacy applications.



# Single workflow for any application



0

## Enable flexibility with extensible driver

- Task Drivers execute tasks on the Nomad Client and provide resource isolation.
- First-class support of a broad set of workloads across all major operating systems.

#### **TERMINAL** task "legacy" { driver = "exec" source = "https://legacy.com/wrapper.zip" } artifact { source = "https://legacy.com/banking.zip" } artifact { artifact · source = "https://legacy/accounts.db" destination = "data" config { command = "wrapper" args = ["\${NOMAD ADDR http}"] **S** Microsoft macOS

lava



## Improved visibility - Consolidated control plan

📢 Nomad							Documentation   ACL Tokens	
	J	obs						
WORKLOAD Jobs	1	Search jobs						
CLUSTER Clients		Name	Status	Туре	Priority	Groups	Allocation Status	
Servers		MongoDB	RUNNING	service	50	1		
		Redis	RUNNING	service	50	1		
		HDFS	RUNNING	service	50	2		
		syslog	RUNNING	system	50	1		
							1–4 of 4	





# **Nomad Enterprise**

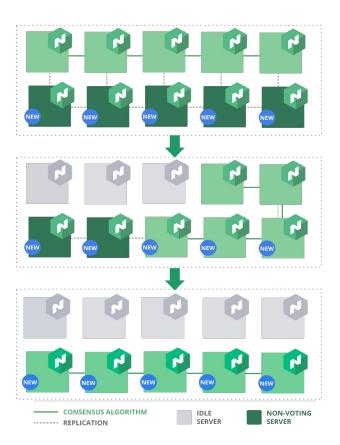




### **Automated upgrade**

Prevent downtime with automated server upgrade.

- New servers with higher version auto-join as non-voting servers.
- Autopilot waits until enough new servers have joined, and then starts to demote the old servers and promotes the new versions until the transition is complete.



### **Automated** backups

Enables automatic snapshotting and storing of Nomad cluster state locally or to a cloud storage of choice, like Amazon S3 or Azure Blob Storage.

HashiCorp

The snapshot agent subcommand starts a process that takes snapshots of the state of the Nomad servers and saves them locally, or pushes them to an optional remote storage service.

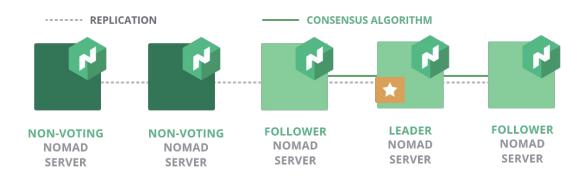
	• • •		TERMINAL
>	snapshot {		
	interval	= "1h"	
	retain	= 30	
	stale	= false	
	service	= "nomad-snapshot"	
	deregister_after	= "72h"	
	lock_key	<pre>= "nomad-snapshot/lock"</pre>	
	max_failures	= 3	
	name_prefix	= "nomad"	
	}		
	<pre>log {     level     enable_syslog     syslog_facility }</pre>	= false	



# Enhanced read scalability

Increase the scale that a Nomad server group can handle for read intensive operations, without impacting performance.

- Non-voting servers receive the replication streams but do not vote in the quorum.
- Non voting servers can act as scheduling workers to improve scheduling throughput.

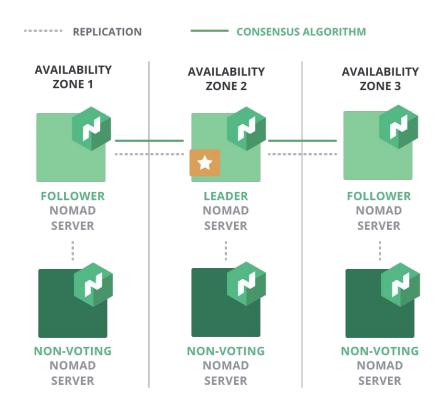




## **Redundancy zones**

Enable redundancy deployment in each availability zone, without impacting performance.

- One voting, one non-voting server per Zone.
- Auto-promote the non-voting server to a voting server in the cluster, when a voting server is lost.

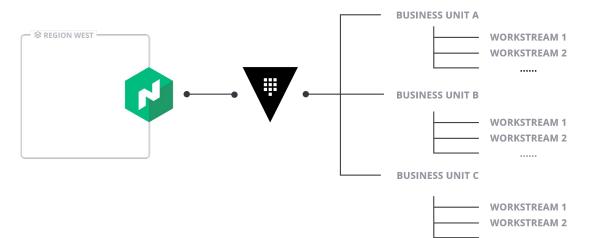




## Multi-Vault Namespaces

Allow one Nomad cluster to span multiple Vault namespaces for simplified topology and fleet consolidation

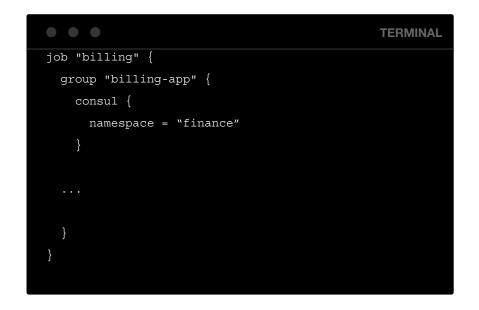
Users can specify a Vault namespace in their job file





## **Consul Namespace Support**

- Enable applications running in Nomad to register their Consul services in specific Consul Namespaces via Nomad job files.
- Improve the interoperability between Nomad and Consul while simplifying the adoption of hierarchical networks powered by Consul Enterprise.





### Memory Oversubscription

- Allow tasks to use more than their scheduled amount of memory without OOM errors
- Improve cluster efficiency and bin packing
- Add additional guards to protect against client-level OOM errors

Memory Max is the resource threshold a task can never exceed Memory is the requested resource that the task is guaranteed to get. **Capacity Available** Max Requested Used



😥 N

Jobs Optimize

Storage

Clients

Topology

## **Dynamic Application Sizing**

Enables Nomad to right-size applications to their most efficient levels of resource consumption.

Reduce spend and drive cost savings intelligently and non-disruptively at scale.

ad	Q Jump to			
	Recommendations			
I	Resource Recommendation example / cache		redis task	Min 0.06494829058647156
ΓΑ	Namespace: default Current 3192 MHz 1024 MiB Difference -2388 MHz Recommended 804 MHz 1024 MiB % Difference -75%	+0%	Current 3192 MHz 1024 Recommended 804 MHz 1024	
	Applying the selected recommendations will save 2388 MHz of 0	CPU. CPU Mem	↓ CPU	Current 3192
	Accept Dismiss		>	- Jok Odreik

## Multi-cluster deployment

Example:

- Simple tweaks to existing single cluster job spec.
- Configurable rollout and rollback strategies.

## TERMINAL multiregion { strategy { max parallel = 1 on failure = "fail all" region "A" { count = 3region "B" { count = 6

### **Resource quotas**

Enable governance across tenants by limiting resource consumption on a per namespace basis.

- Resources limits are defined independently per region.
- Single resource quota an be applied to multiple namespaces.

 $\odot$ 

 Automatic replication across regions for easy, centralized administration at global scale.

	TERMINAL
\$ nomad quota init Example quota specification written to spec.hcl	
Example quota specification written to spec.net	
\$ cat spec.hcl	
name = "default-quota"	
description = "Limit the shared default namespace"	

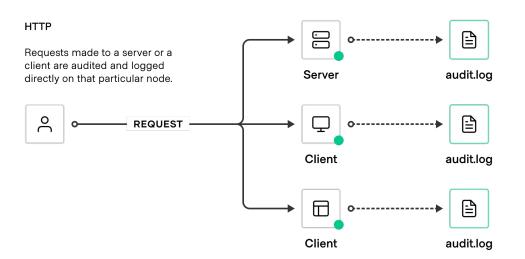
# Create a limit for the global region. Additional limits may
# be specified in-order to limit other regions.
limit {
 region = "global"
 region\_limit {
 cpu = 2500
 memory = 1000
 }
}



## **Audit logging**

Secure clusters by providing administrators with a complete set of records for all user-issued actions.

- HTTP API-based logging system.
- Event broker that runs on all Nomad agents (clients & servers).
- Writes events to an audit log file local to each node.
- JSON response and request objects of each API transaction.
- Identical mental model as API for operator ease in digestibility.



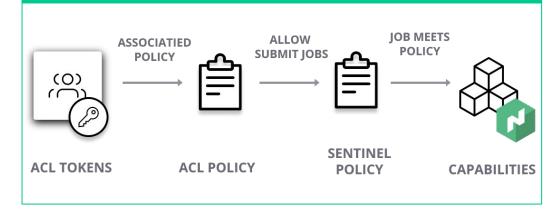


## **Sentinel policies**

Enforce fine-grained policy as code with Sentinel integration.

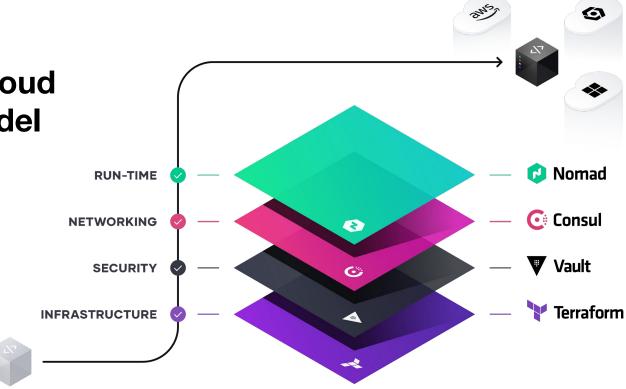
- Integrate with the Nomad ACL system.
- Policies are enforced on job submission and job update.
- Sentinel policies are automatically replicated across regions for easy, centralized administration at global scale.

#### SENTINEL POLICY WORKFLOW





### A Common Cloud Operating Model









# 

# HashiCorp

#### **LINBIT HA-Solutions Gmbh**

Vivenotgasse 48 1120 Vienna, Austria sales@linbit.com HashiCorp Germany GmbH team-emea-dach@hashicorp.com