HashiCorp Vault

Vault Self-Managed Getting Started & Technical Overview

October 2021

Agenda

1. Overview
2. Architecture
3. Deployment Patterns
4. Operations
5. Next Steps

Vault Enterprise Path to Production

Weeks 1-3

Vault Deployed

- Kick Off Webinar
- Architectural Deep Dive Workshop
- Terraform starter code modules
- Office Hours
- Schedule Success Planning Session

Weeks 4-5

Identity & Access Management

- Namespaces, ACL Policies, Auth Methods Workshop
- Office Hours
- 1:1 Success Planning

Weeks 6-7

Operational Considerations

- Telemetry & Monitoring and DR Operations Workshop
- HCDiag & Vault Metrics Lunch & Learn
- Office Hours
- 1:1 Success Planning

Weeks 8-9

Secrets Management

- How to consume Vault into applications Workshop
- Dynamic Secrets Lunch & Learn
- Office Hours
- 1:1 Success Planning

Weeks 10-11

Ready for Production!

- Success Plan Reviews
- Exit Ramp & Operational Readiness Workshop

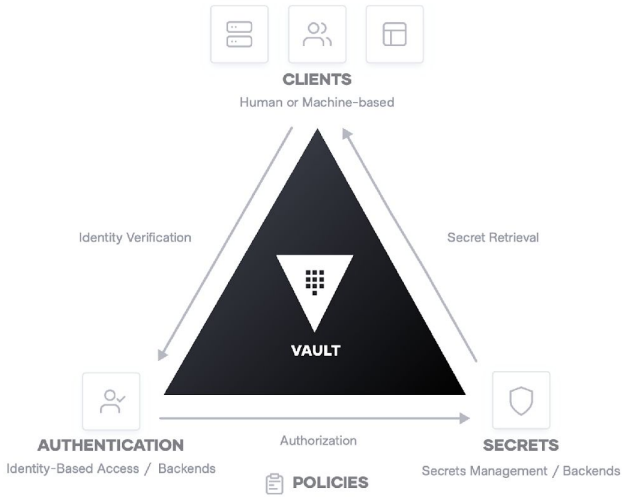
01

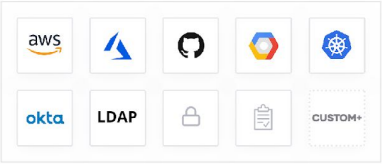
Vault Overview

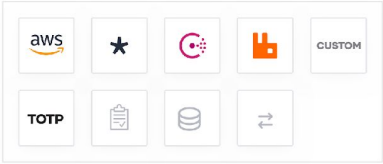
Overview

Vault tightly controls access to secrets and encryption keys by authenticating against trusted sources of identity such as Active Directory, LDAP, and cloud identity platforms.

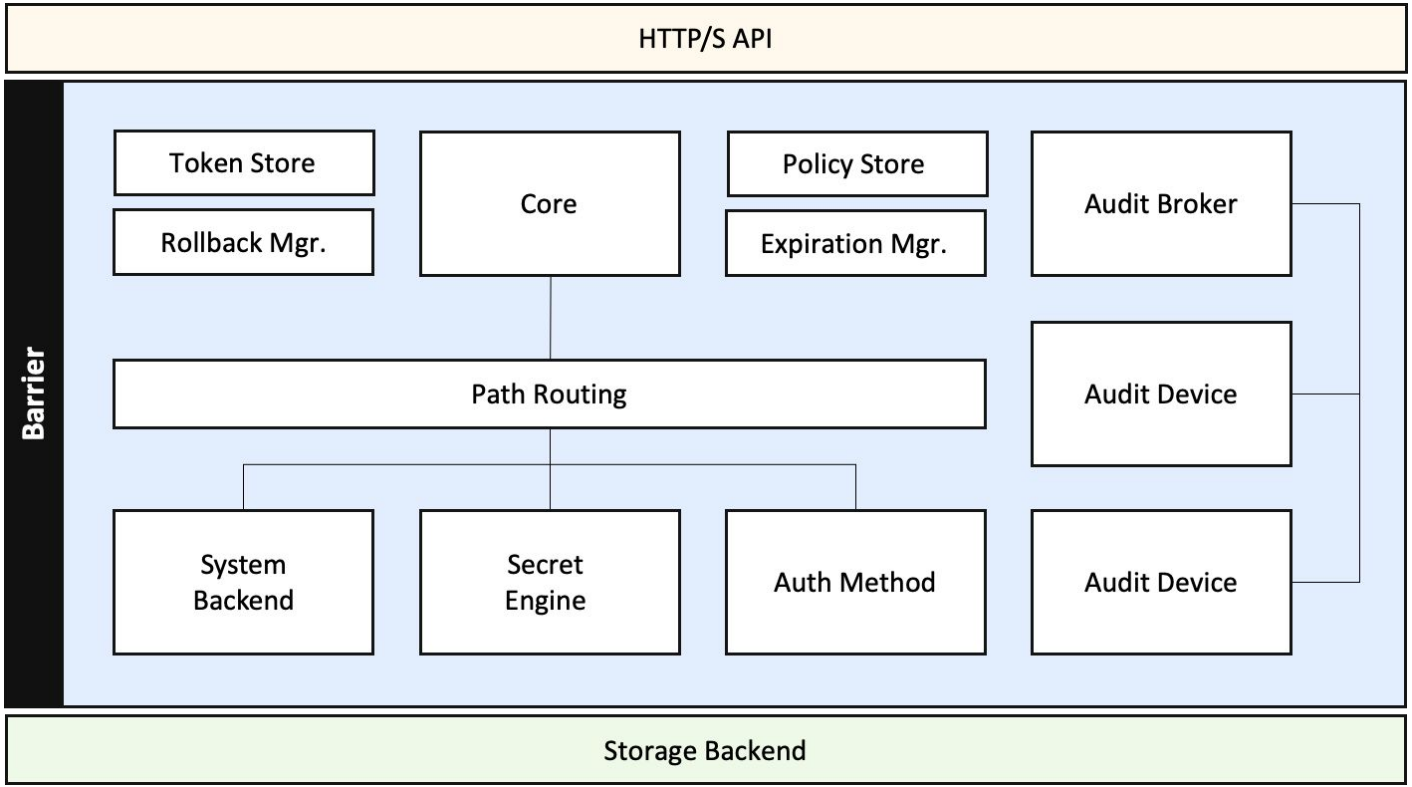
Vault enables fine grained authorization of which users and applications are permitted access to secrets and keys.







Architecture & Cryptographic Barrier



Vault Security Model

- **It's all about access to the Encryption Key.**
- Configuring "`cap_ipc_lock=+ep`", `LimitNOFILE, and LimitMEMLOCK` prevent Memory Swapping to Disk, so secrets are not written in plain text to disk.
- The Vault Encryption Key is stored in memory in **PLAIN TEXT**
  - This is done for performance
  - Root access to an unlocked vault server could compromise this
  - Isolation technologies which allow reading of memory could compromise this (VM issues, but principally Kubernetes)
- Master Root Key protects the Encryption key, so it also must be secure
- Auto-Unseal shifts risk profile

Cryptography Security Model

- **Vault uses publicly available cryptographic technologies**
- P vs NP - Good cryptographic algorithms are exponential in difficulty to solve but polynomial in difficulty to validate answers for.
- Numerous algorithms (SHA1) were exposed to have defects that allowed them, or a subset of them to be reduced to polynomial difficulty problems
- Short encryption keys and faster computers has made brute-forcing older encryption standards possible.
- Software based random number generations suffer from a lack of randomness.

02

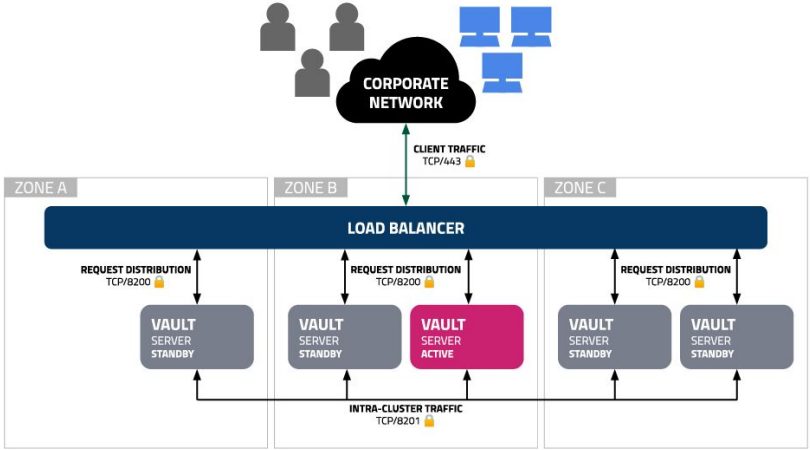
Vault Architecture

Vault Cluster Architecture

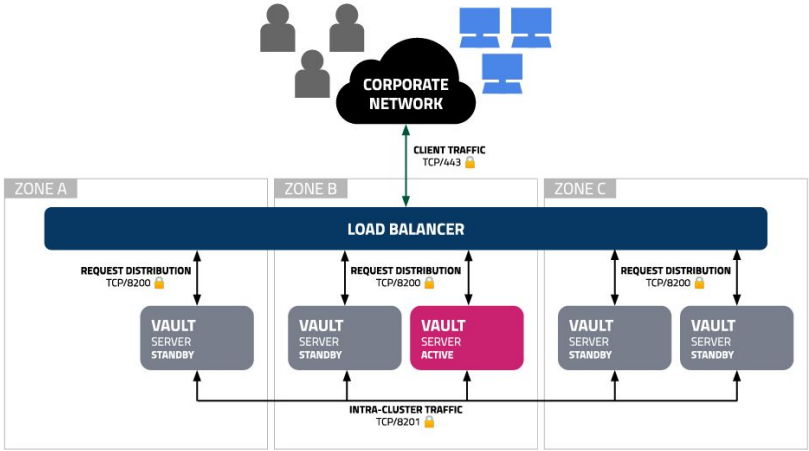
Integrated Storage Reference Architecture

This architecture provides a highly resilient and scalable Vault deployment.

- Able to withstand loss of two nodes or an entire Availability Zone (AZ).
- Integrated Storage requires high-bandwidth, low latency (<8ms RT) connection.



Vault Integrated Storage Architecture



5 Vault Servers, 3 Availability Zones

On-premise Considerations

A single region in a On-premise Deployment is considered a geographically separate collection of one or more Availability Zones.

- An AZ would be considered a single failure domain containing it's own compute, storage, network, power, etc.
- A high bandwidth, low latency (<8ms RT) connection should exist between each AZ.

Sizing

Per instance sizing recommendations

	Small (Dev/Test/Staging /QA)	Large (Production)
CPU	2 - 4 Core	4 - 8 Core
Memory	8 - 16 GB RAM	32 - 64 GB RAM
Disk Capacity	100+ GB	200+ GB
Disk IO	3000+ IOPS	10000+ IOPS
Disk Throughput	75+ MB/s	250+ MB/s

Cloud Instance Sizing

Provider	Size	Instance/VM Types	Disk Volume Specs
AWS	Small	m5.large, m5.xlarge	100+ GB gp3, 3000 IOPS, 125 MB/s
	Large	m5.2xlarge, m5.4xlarge	200+ GB gp3, 10000 IOPS, 250 MB/s
Azure	Small	standard_d2s_v3, standard_d4s_v3	1024 GB Premium_LRS
	Large	standard_d8s_v3, standard_d16s_v3	1024 GB Premium_LRS
GCP	Small	n2-standard-2, n2-standard-4	500 GB pd-balanced
	Large	n2-standard-8, n2-standard-16	1000 GB pd-ssd

Performance Considerations

Profile Workloads
As you scale the adoption of Vault throughout your organization, you will have varying workloads access Vault. Telemetry monitoring should be leveraged to ensure proactive monitoring of Vault Cluster resources. Additionally, as you onboard new applications/services/teams/users to Vault, take time to profile the usage patterns to ensure optimal authentication and consumption patterns are used.

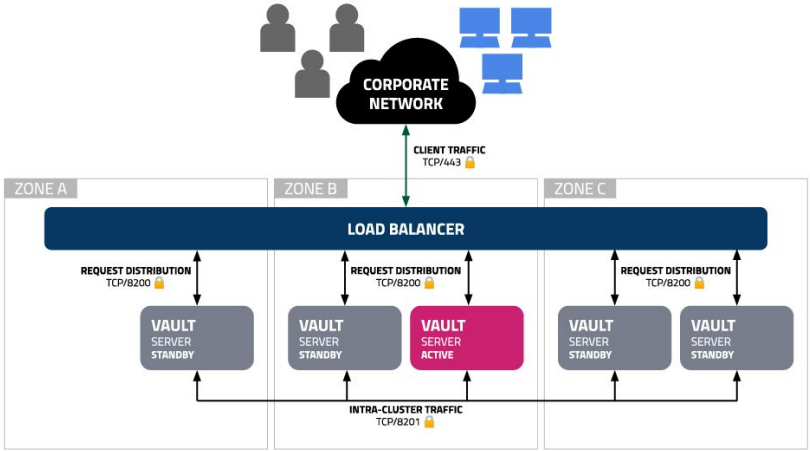
External Systems
Depending on the Authentication Methods and Secrets Engines used by your organization, you will likely have dependency on other external systems for Vault requests to be completed. Ensure telemetry is enabled on those services and proactively monitor for performance issues.

Networking

Networking Considerations

Integrated Storage can enable you to simplify your networking requirements compared to other storage backends.

- <8ms RT network connection required to ensure Raft Storage remains consistent across all Vault Nodes.
- Restrict communication to only required ports and CIDRs
- Standard HTTPS TLS encryption should be used to protect network traffic



Networking Requirements

Source	Destination	Port	Protocol	Direction	Purpose
Client Machines	Load Balancer	443	tcp	incoming	Request distribution
Load Balancer	Vault Servers	8200	tcp	incoming	Vault API
Vault Servers	Vault Servers	8200	tcp	bidirectional	Cluster Bootstrapping
Vault Servers	Vault Servers	8201	tcp	bidirectional	Raft, replication, request forwarding
Vault Servers	External Systems	various	various	varios	External APIs

High Availability

HA Strategy for Client Access
Vault does not include built in load balancing capabilities. To ensure Vault can meet the highest levels of reliability and stability, either an external load balancer or Consul should be used to distribute client requests.

- TLS should terminate at Vault and not the load balancer to ensure end-to-end encryption
- Use Vault's health endpoint to determine active node and node health https://<vaultnode>:8200/v1/sys/health

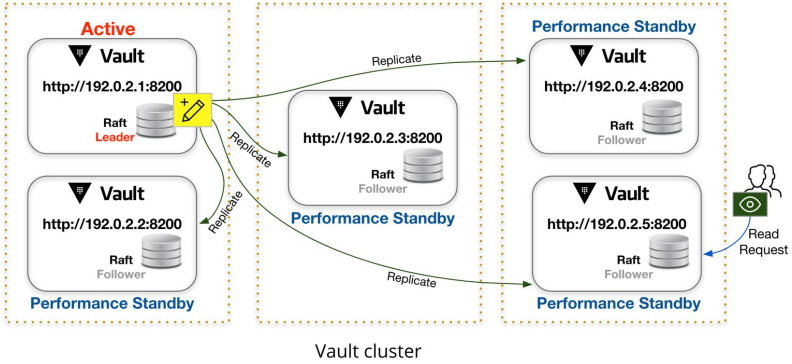
Scaling Considerations

Managed scaling services should be leveraged when deploying in a cloud environment to ensure the Vault cluster remains populated with health nodes.

- Additional nodes will not increase performance. Scaling services will increase your resiliency but not improve performance.
- Do not replace all instances at once in a scaling group otherwise you will have to restore from a snapshot

Cloud	Managed Auto Scaling Service
AWS	Auto Scaling Group
Azure	Virtual Machine Scale Sets
GCP	Managed Instance Groups

Scaling Performance Standby Nodes



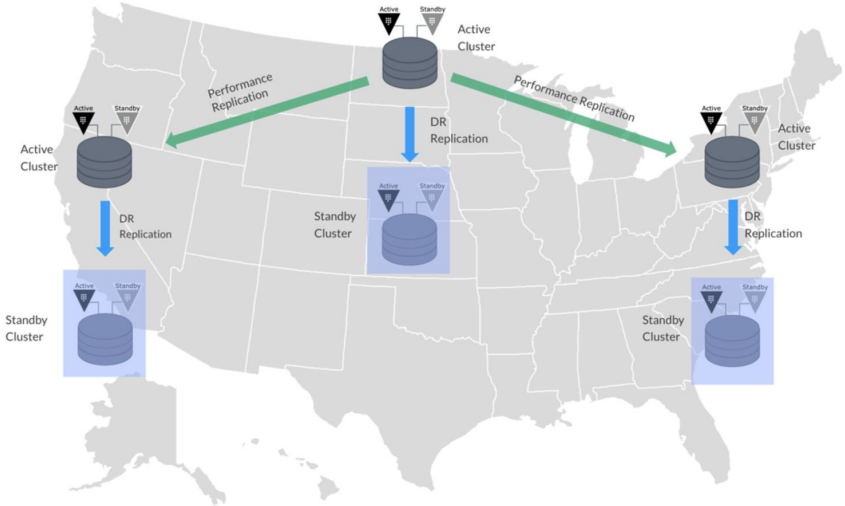
Horizontal scalability for read requests
Performance Standby Nodes can be used to respond to read-only requests. Performance Standby Nodes are enabled by default and process read-only requests locally. Write requests that modify shared state in Vault will be forwarded to the singular Active Node.

- Integrated Storage uses eventual consistency and data may not be available across all nodes immediately
- Vault 1.7+ includes multiple methods to control how requests are handled

Multiple Regions

Vault Replication

Vault can be extended to multiple regions using replication. The primary cluster uses asynchronous replication to ship data to the secondaries. Multiple replication modes can be combined to provide resilience and performance.



Replication Types

Disaster Recovery

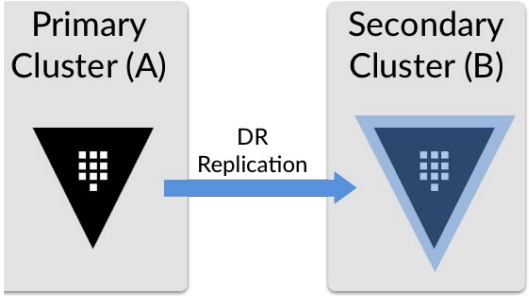
Provides a warm standby cluster that contains all data from the primary Vault cluster. **It is strongly recommended to deploy at least one DR cluster.**

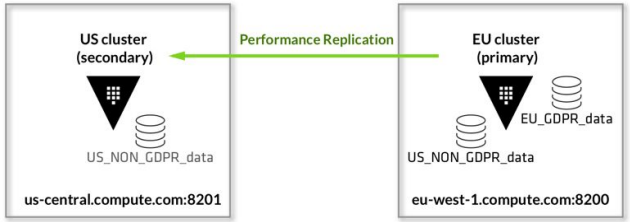
Performance Replication

Provides an active Vault cluster with shared state of the primary. This includes secrets, authentication methods, policies, and other shared data. Token and leases are not replicated to performance secondaries.

Replication

Determine Replication Mode





DR Replication

- RPO/RTO requirements
- Tier 0 or 1 apps be accessing secrets from Vault

Performance Replication

- Source of Requests
- Performance requirements for request and response
- Types of workloads

03

Deployment Patterns

Recommended Patterns

Immutable Builds

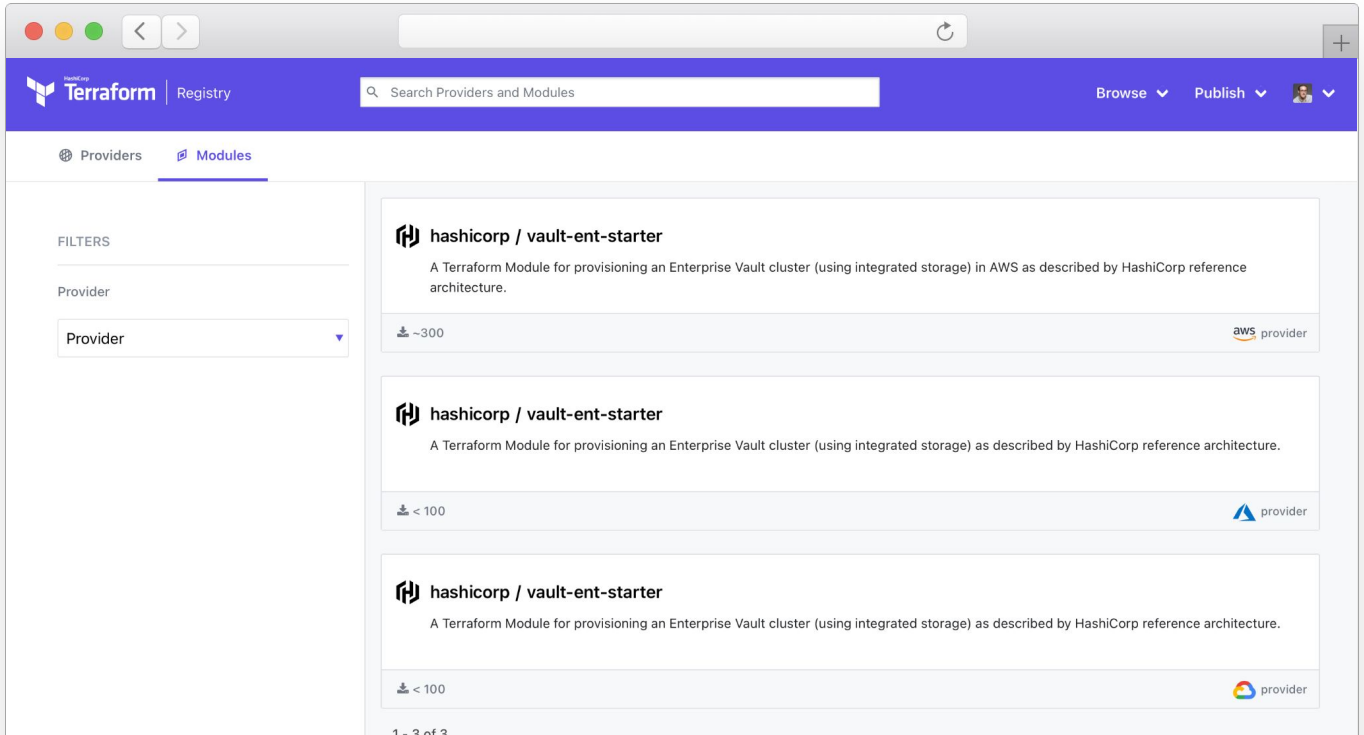
Tooling like Packer can be used to build immutable images of Vault and perform blue/green deployment using your existing CI/CD orchestration. This can streamline your lifecycle processes. However, when using integrated storage, you will need to take measures to ensure quorum is maintained as new image versions are introduced to the cluster.

Configuration Management

For organizations who have not adopted the above pattern, Vault can be integrated into your configuration management patterns to install, upgrade, and configure Vault.

Terraform Modules

Quickly deploy Vault cluster based on reference architecture



Vault Helm Chart

Deploy Vault Reference Architecture inside Kubernetes

Search or jump to... / Pull requests Issues Marketplace Explore

hashicorp / vault-helm Public

Watch 50 Star 622 Fork 515

Code Issues 98 Pull requests 35 Actions Projects Security Insights

main 4 branches 25 tags Code

tvoran fix chart publish job (#620) ... 9758666 9 days ago 437 commits

.circleci	fix chart publish job (#620)	9 days ago
.github	Update jira sync github action (#411)	11 months ago
templates	Adding support for the old leader-elector (#607)	23 days ago
test	vault-helm 0.16.1 release (#619)	9 days ago
.gitignore	feature: Support configuring various properties as YAML directly. (#5...	3 months ago
.helmignore	Ignore bin dirs	3 years ago
CHANGELOG.md	vault-helm 0.16.1 release (#619)	9 days ago
CONTRIBUTING.md	vault-helm default branch is now `main` (#618)	11 days ago
Chart.yaml	vault-helm 0.16.1 release (#619)	9 days ago

About

Helm chart to install Vault and other associated components.

Readme

MPL-2.0 License

Releases 24

v0.16.1 Latest
9 days ago

+ 23 releases

Packages

No packages published

Upgrades

Major releases of Vault are released quarterly and we release minor releases monthly and as needed

The process to update Vault should be automated as much as possible to ensure Vault remain patched with the latest bug and security fixes. We ensure upgrade compatibility with N-2 major releases.

Prior to performing any upgrade it is recommended to review the changelog and version specific upgrade guide, test against version in QA environment, and ensure up to date snapshots are available in case you need to restore.

04

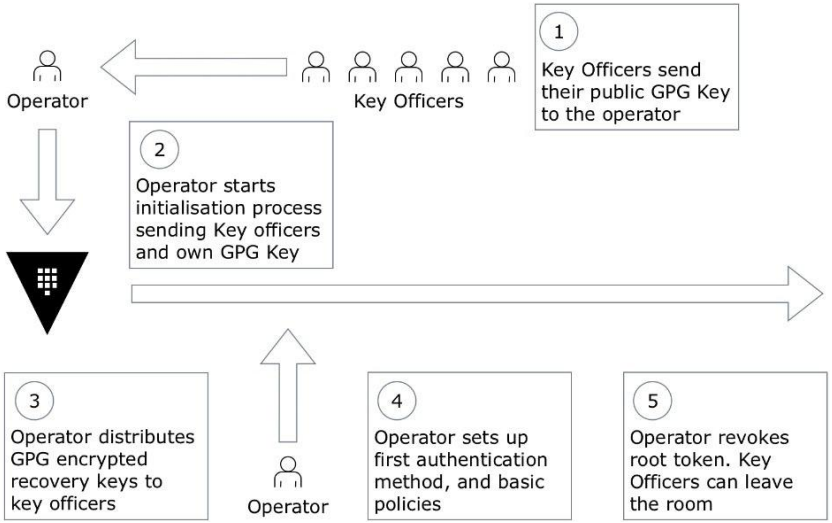
Operations

Initialization Ceremony

Vault Cluster Initialization Process

The initialization process only occurs once in a Vault installation. During this process, unseal or recovery keys and root token are returned to the operator. Keys are then distributed to key officers who are the guardians of Vault trust.

- Recommended to complete initialization process in single room with Vault operator and key holders until root token is revoked.
- Recovery Keys are returned when Vault is initialized while using Cloud KMS or HSM. Recovery keys should be protected the same as unseal keys.



TERMINAL

```
> vault operator init
Unseal Key 1: Ly7wgNFzKVcw95nv6fLTQ/lsf49Wn4JaIEYGPm15pSzn
Unseal Key 2: JWeteKjgpFXI2wY2I16j8JCCy92PO4GxGCyXvLCoHp1L
Unseal Key 3: zLkMbO9Lcr3QRwIgwe7KBPy5jRD9aUttt0l0HZ4dusvx
Unseal Key 4: 0J5fD29c5ZisKl1jL13K0XOmIWu66PfA6NBV3UEK7f/f
Unseal Key 5: ahR0lB203KzxvOa0HgBLUDmByxhFdeVOVeA3l6PMIKMn

Initial Root Token: s.dZlm13ORBFkFOrQeWtLF3uiA

Vault initialized with 5 key shares and a key threshold of 3. Please securely
distribute the key shares printed above. When the Vault is re-sealed, restarted, or
stopped, you must supply at least 3 of these keys to unseal it before it can start
servicing requests.

Vault does not store the generated master key. Without at least 3 key to reconstruct
the master key, Vault will remain permanently sealed!

It is possible to generate new unseal keys, provided you have a quorum of existing
unseal keys shares. See "vault operator rekey" for more information.
```

Root Token Generation

Root Token Handling Practices

The root token is returned to the operator during the initialization ceremony. This token can do **anything** in Vault and its usage should be closely monitored.

- Once operator has configured a secondary authentication method and granted administrators sudo access, almost all operations can be performed.
- Best practice is **NOT** persisting the root token.
- Generate a root token only when absolutely necessary.

Production Readiness

Critical items to have in place before production go-live

Backup

Automated Integrated Storage Snapshots, included in your enterprise subscription, can take periodic snapshots of your data.

- Determine where snapshot files will be stored
- Configure based off your RPO/RTO requirements.
- If snapshot is restored, the unseal keys that were valid at the time of the snapshot will be used to unseal.

Automated Integrated Storage Snapshots

TERMINAL

```
> vault write \
     sys/storage/raft/snapshot-auto/config/testsnap \
    storage_type=local \
    file_prefix=testsnappy \
    interval=120m \
    retain=7 \
    local_max_space=1000000 \
    path_prefix=/opt/vault/
```

Production Readiness

Critical items to have in place before production go-live.

Monitoring

Vault should be monitored closely to ensure the service remains healthy and available in production.

- Telemetry - Export telemetry data to solution that can analyze and identify trends overtime.
- Log Analytics - Capture app logs and system logs and perform analysis on the log files for useful signals.
- Active Health Checks - Query health endpoints to get the health of nodes and route traffic to active node.

Production Readiness

Critical items to have in place before production go-live.

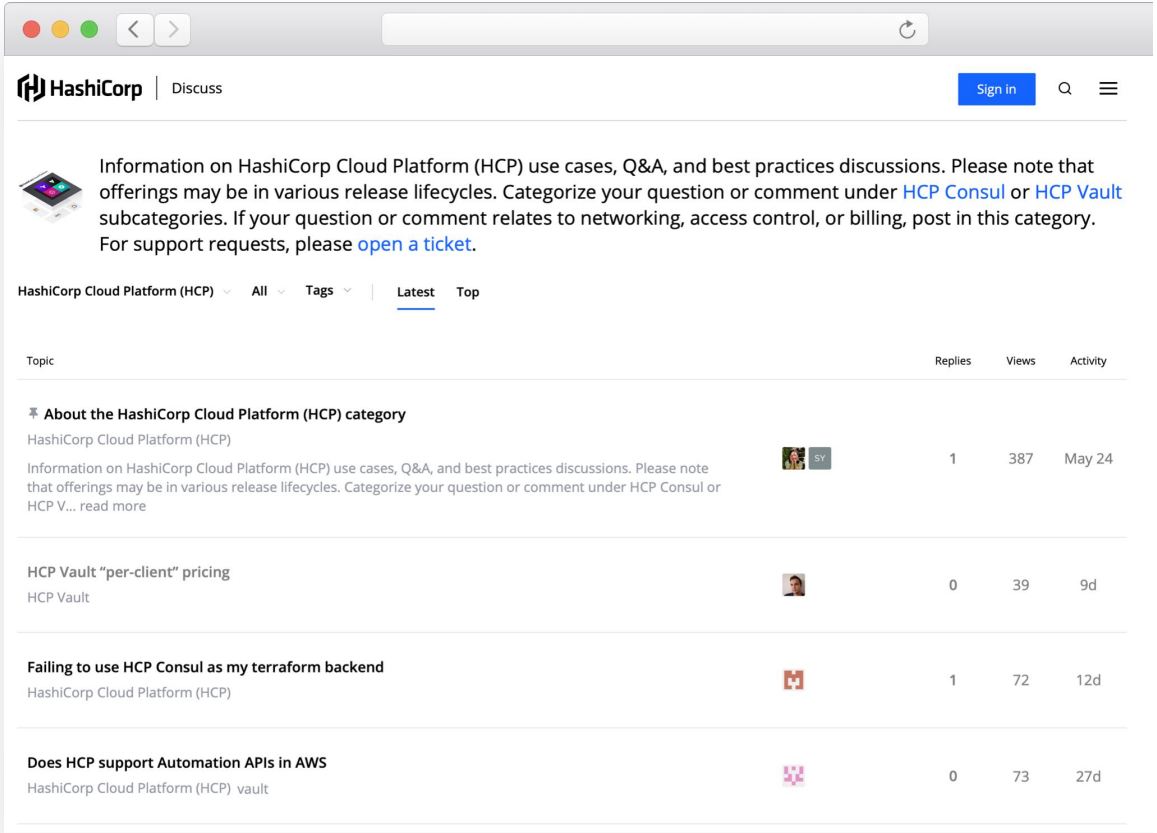
Auditing

Vault sends audit information to a SIEM system or logging backend.

- Determine audit devices that will be used.
- Vault will not respond if the audit device is unavailable, use multiple audit devices to ensure Vault remains available.
- Sensitive fields are HMAC, Selectively determine if any HMAC fields need to be exposed

05

Next Steps



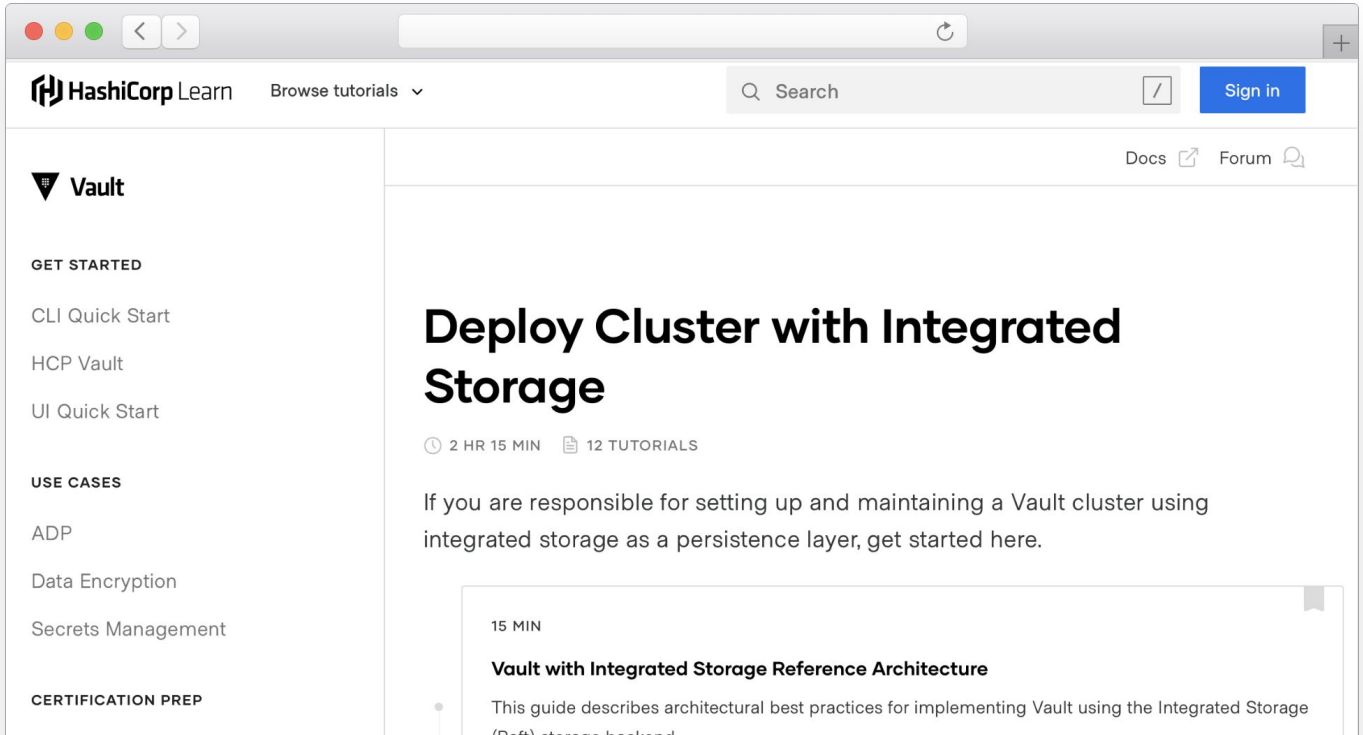
Discuss

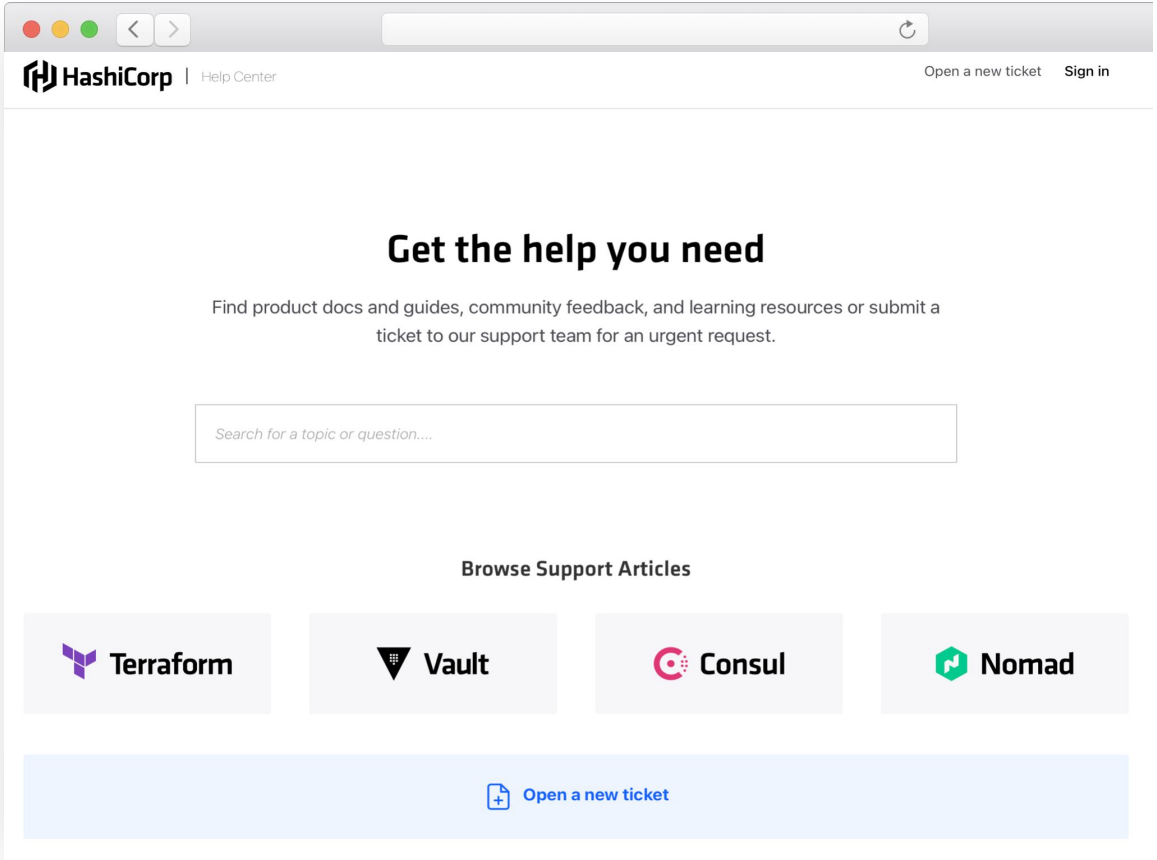
Engage with the HashiCorp Cloud community including HashiCorp Architects and Engineers.

discuss.hashicorp.com

Learn

Step-by-step guides to accelerate deployment of Vault





Support

https://support.hashicorp.com

Need Additional Help?

Customer Success

Contact our Customer Success Management team with any questions. We will help coordinate the right resources for you to get your questions answered.

customer.success@hashicorp.com

Technical Support

Something not working quite right? Engage with HashiCorp Technical Support by opening a new ticket for your issue at support.hashicorp.com.

Q & A

Thank You

hello@hashicorp.com
www.hashicorp.com