Vault

Implementation Foundations

Module: Vault Deployment Guidelines

What You Will Learn

- Vault Production Deployment Best Practices
- Vault Deployment Considerations
- Vault Storage Model
- Vault Deployment Security Model

Production Best Practices

Things to Consider

Location	Infrastructure	Security
• Public vs. Private • Availability Zone • Redundancy	• Physical vs. Virtual • Platform Support • Sizing Requirements • Network Requirements	• Risk Assesement • Security Model • Production Hardening

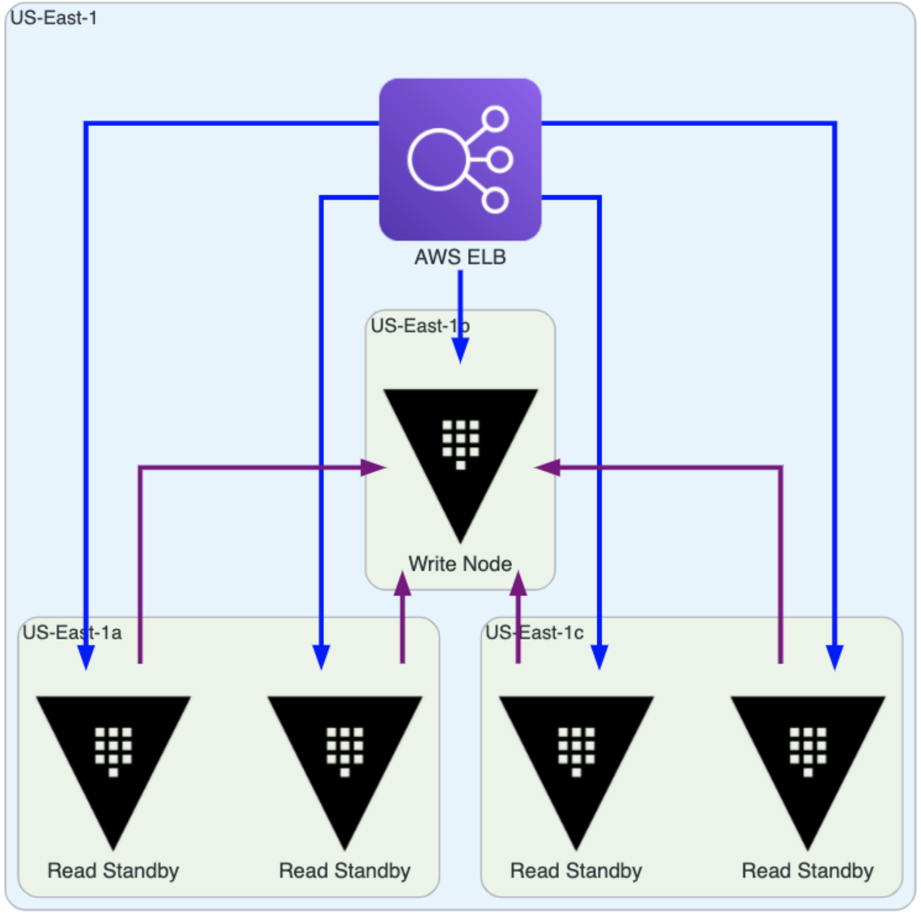
Public vs Private Considerations

Private

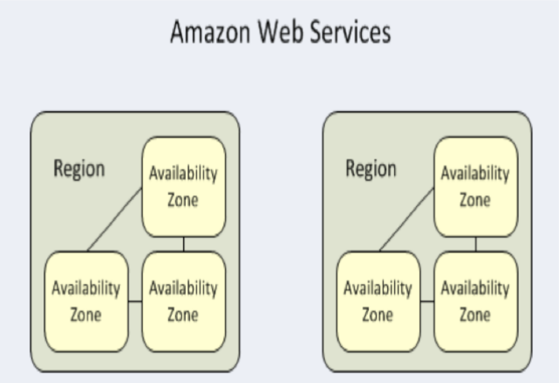
- vSphere Clustering

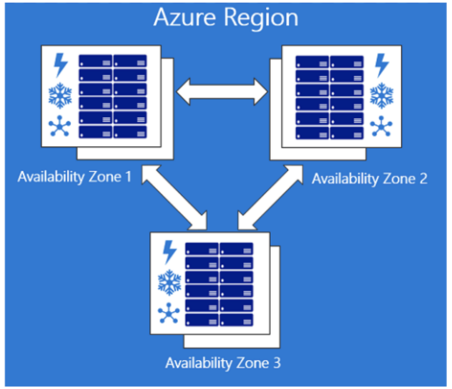
Public

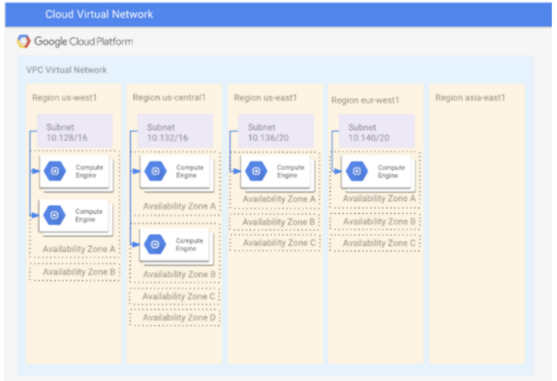
- Availability Zones
- Cross Region Connectivity
- Failure Topology



Selecting Your Deployment Region







AWS

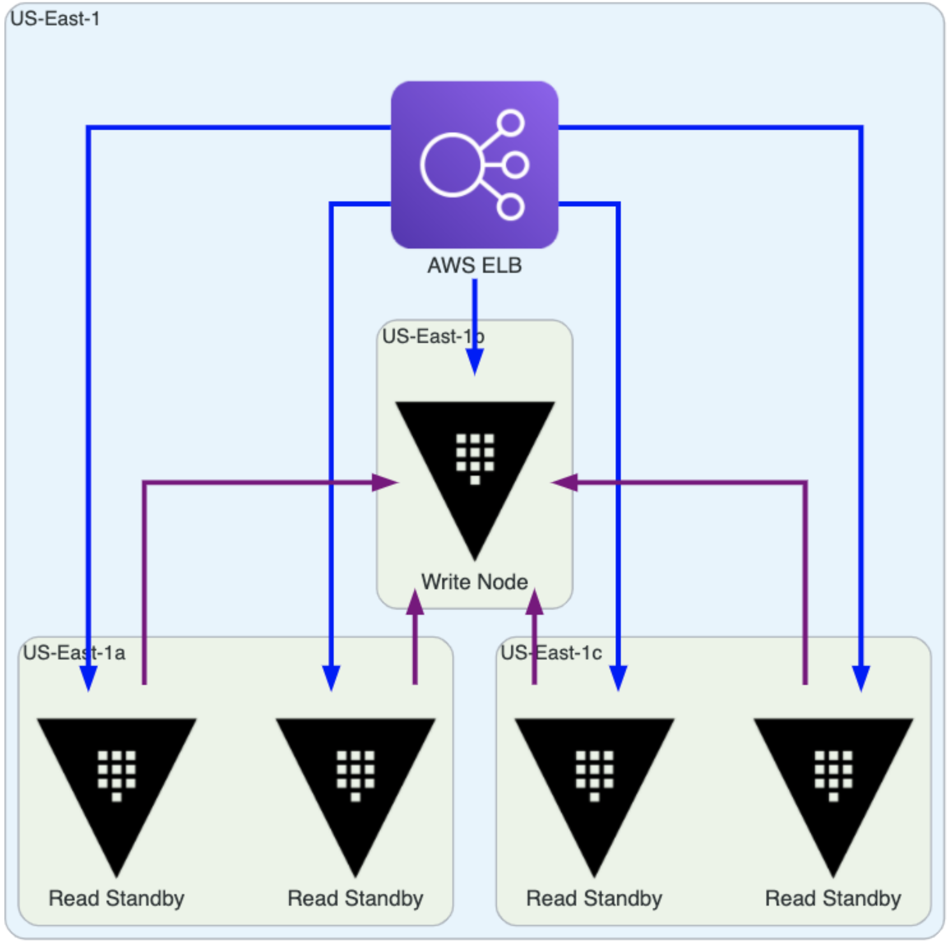
Azure

GCP

Redundant Deployment - Consensus

Servers	Quorum Size	Failure Tolerance
1	1	0
2	2	0
3	2	0
4	3	1
5	3	2
6	4	2
7	4	3

Multi-AZ Deployment - Integrated Raft



Hardware vs Virtual vs Container

Hardware	Virtual	Container
• Best level of security • Limits to on premise resources or expensive cloud options	• Universal Standard • Cost Optimization • Supported Automation Methods	• Service Management • Supported Automation Methods

Vault Storage Model

Vault Backend Storage Model

- For enterprise customers, HashiCorp provides official support for Consul and Vault's Integrated Storage as storage backends.

- Many other options for storage are available with community support

Integrated vs External

	Integrated Storage	External Storage
HashiCorp Supported	Yes	Limited support
Operation	simpler - no additional software installation required	Must install and configure the external storage environment outside of Vault. For high availability, the external storage should be clustered.
Networking	One less network hop	Extra network hop between Vault and the external storage system (e.g., Consul cluster).
Troubleshooting and monitoring	Vault is the only system you need to monitor and troubleshoot	The source of failure could be the external storage; therefore, you need to check the health of both Vault and the external storage.
Data location	The encrypted Vault data is stored on the same host where the Vault server process runs	The encrypted Vault data is stored where the external storage is located

Raft vs Consul

	Integrated Storage	Consul
Deployment	Vault cluster is all you need	Vault cluster & Consul cluster Use a dedicated Consul cluster for Vault storage, and it should not be used for other purposes (e.g., service discovery, service mesh)
Data location	Data is on disk.	All data is in memory.
Snapshots	Normal data backup strategy of your organization.	More frequent snapshots are necessary since data is in memory.
Max message size	1 MiB (Configurable using the [max_entry_size]() parameter)	512 KiB (Configurable using the [kv_max_value_size]() parameter)

Raft Configuration

- Using Vault Integrated Storage requires configuring the Raft storage backend
- Raft peers may be initialized:
    - manually with hard-coded configuration values
    - via the cloud auto-join feature on supported cloud providers

Manual Configuration Example

```
storage "raft" {
  path    = "/opt/vault/data"
  node_id = "<UNIQUE_ID_FOR_THIS_HOST>"

  retry_join {
    leader_tls_servername   = "<VALID_TLS_SERVER_NAME>"
    leader_api_addr         = "https://<ADDRESS_OF_PEER_1>:8200"
    leader_ca_cert_file     = "/opt/vault/tls/vault-ca.pem"
    leader_client_cert_file = "/opt/vault/tls/vault-cert.pem"
    leader_client_key_file  = "/opt/vault/tls/vault-key.pem"
  }
}
```

Auto-join Configuration Example

```
storage "raft" {
  path    = "/opt/vault/data"
  node_id = "<UNIQUE_ID_FOR_THIS_HOST>"

  retry_join {
    auto_join               = "provider=aws region=<AWS_REGION> tag_key=
    auto_join_scheme        = "https"
    leader_tls_servername   = "<VALID_TLS_SERVER_NAME>"
    leader_ca_cert_file     = "/opt/vault/tls/vault-ca.pem"
    leader_client_cert_file = "/opt/vault/tls/vault-cert.pem"
    leader_client_key_file  = "/opt/vault/tls/vault-key.pem"
  }
}
```

Integrated Storage - Hardware Sizing

Size	CPU	Memory	Disk Capacity	Disk IO	Disk Throughput
Small	2-4 core	8-16 GB RAM	100+ GB	3000+ IOPS	75+ MB/s
Large	4-8 core	32-64 GB RAM	200+ GB	10000+ IOPS	250+ MB/s

Small clusters would be appropriate for most initial production deployments or for development and testing environments.

Large clusters are production environments with a consistently high workload

Autopilot

Vault 1.7 introduced autopilot to simplify and automate the cluster management for Integrated Storage. The autopilot includes:

- Cluster node health check
- Server stabilization: prevent disruption to raft quorum due to an unstable new node
  - Monitor newly added node health for a period and decide promotion to voter status
- Dead server cleanup - periodic, automatic clean-up of failed servers

Vault Security Model

Vault Security Model

Due to the nature of Vault and the confidentiality of data it is managing, the Vault security model is very critical. The overall goal of Vault's security model is to provide:

- confidentiality
- integrity
- availability
- accountability
- authentication

Vault Production Hardening

- Disable swap
- Single tenancy
- Enable `mlock`
- End-to-End TLS
- Firewall Traffic
- Disable SSH / RDP
- Don't run as root
- Immutable upgrades
- Turn core dumps off
- Root Token management

Platform Optimization

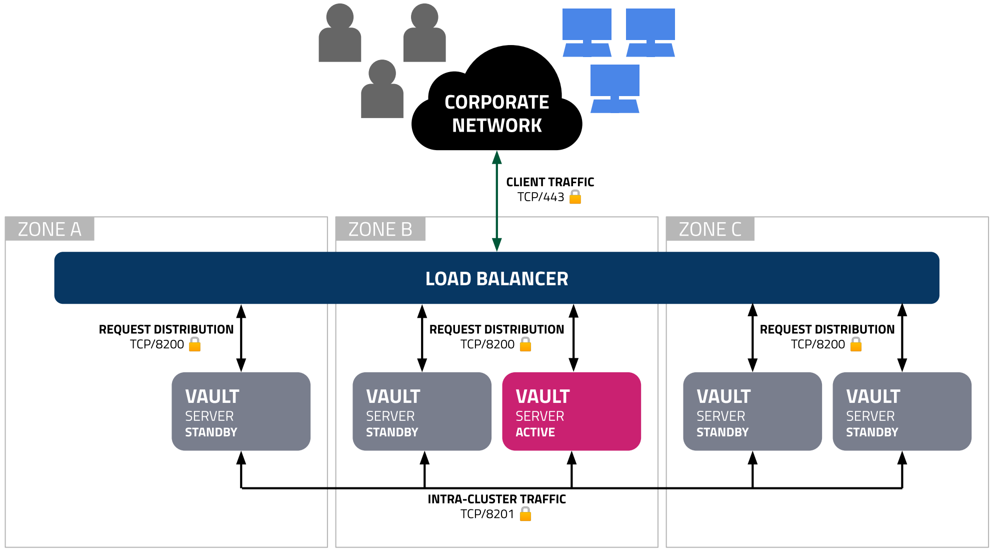
Officially Supported Platforms

- AWS Marketplace
- Terraform Provider
- Docker container
- Helm Chart

Community Supported Platforms

- Chef/Puppet/Ansible/Salt
- Openshift/Openstack

External Load Balancing



- Poll the sys/health endpoint to detect active node
- Prefer L4 over L7 load balancing
- If L7 required, must terminate TLS on Vault

Chapter Summary

- Deployment Location
- Security Considerations
- Load Balancer Management
- Hardware, Virtual, or Container
- Platform's Native Support Capabilities

Reference Links

- Vault Security Models
- Vault Architecture
- Vault Reference Architecture
- Vault Deployment Guide