Planning for High Availability

Ned Bellavance
Founder, Ned in the Cloud LLC

@ned1313 | nedinthecloud.com

Overview

Vault clustering options

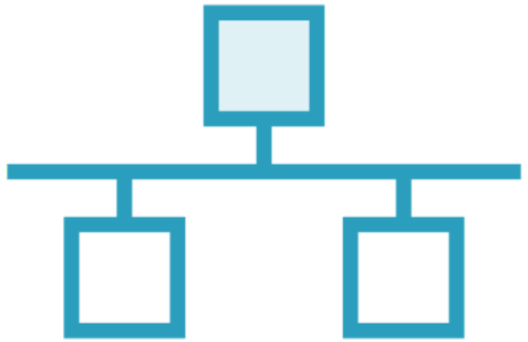
Cluster communications

Replication options

Vault Server Clustering

High Availability Components

Compute



Network

Storage

Compute

Active and standby

Forward or redirect

Read-only for Enterprise

Lock based in datastore

Different storage for HA

Network Components

Listener
cluster_address

Node
cluster_addr

Node
api_addr

Direct access

Load balancer

Cluster Communications

```
listener "tcp" {
  address = 10.1.1.1:8200
  cluster_address = 10.1.1.1:8201
}
```

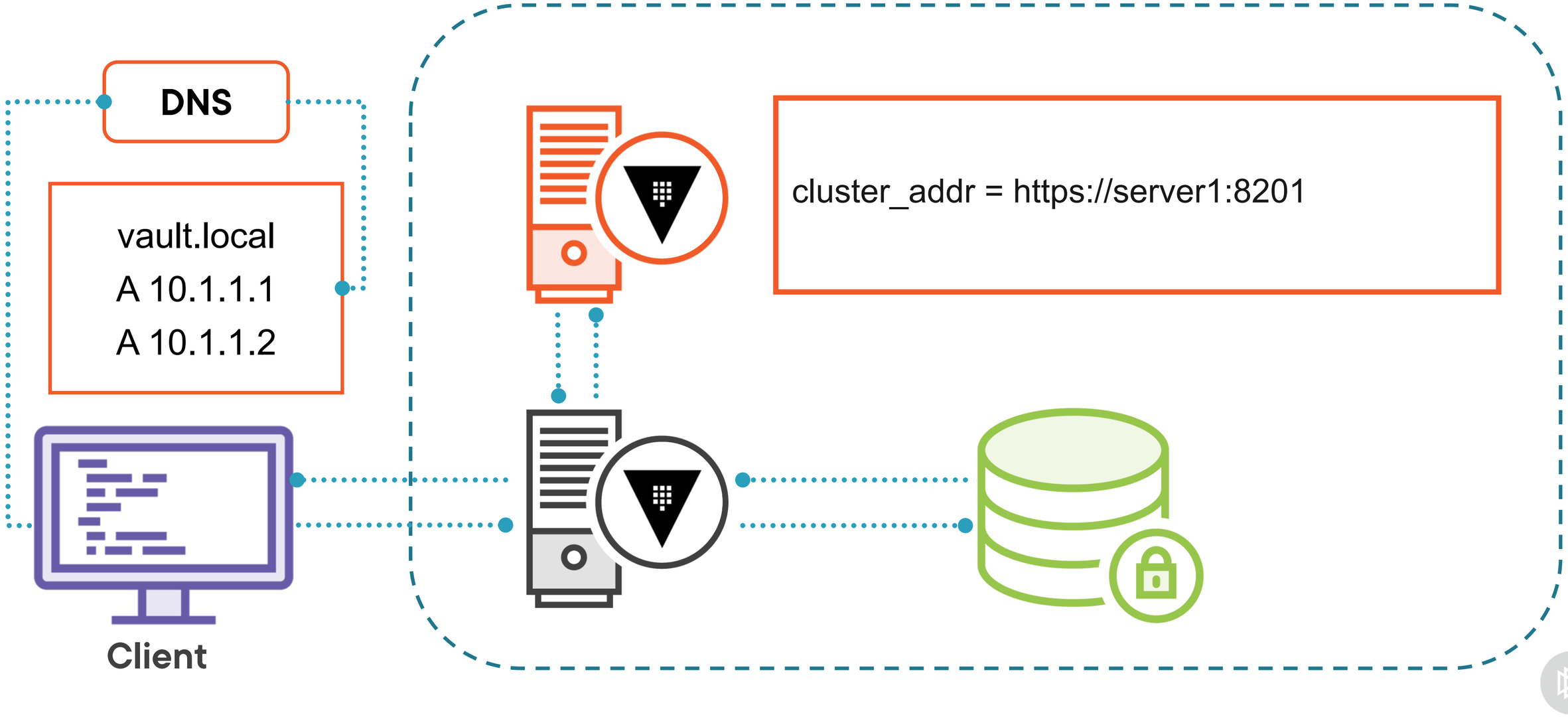
```
listener "tcp" {
  address = 10.1.1.2:8200
  cluster_address = 10.1.1.2:8201
}
```

Cluster Communications

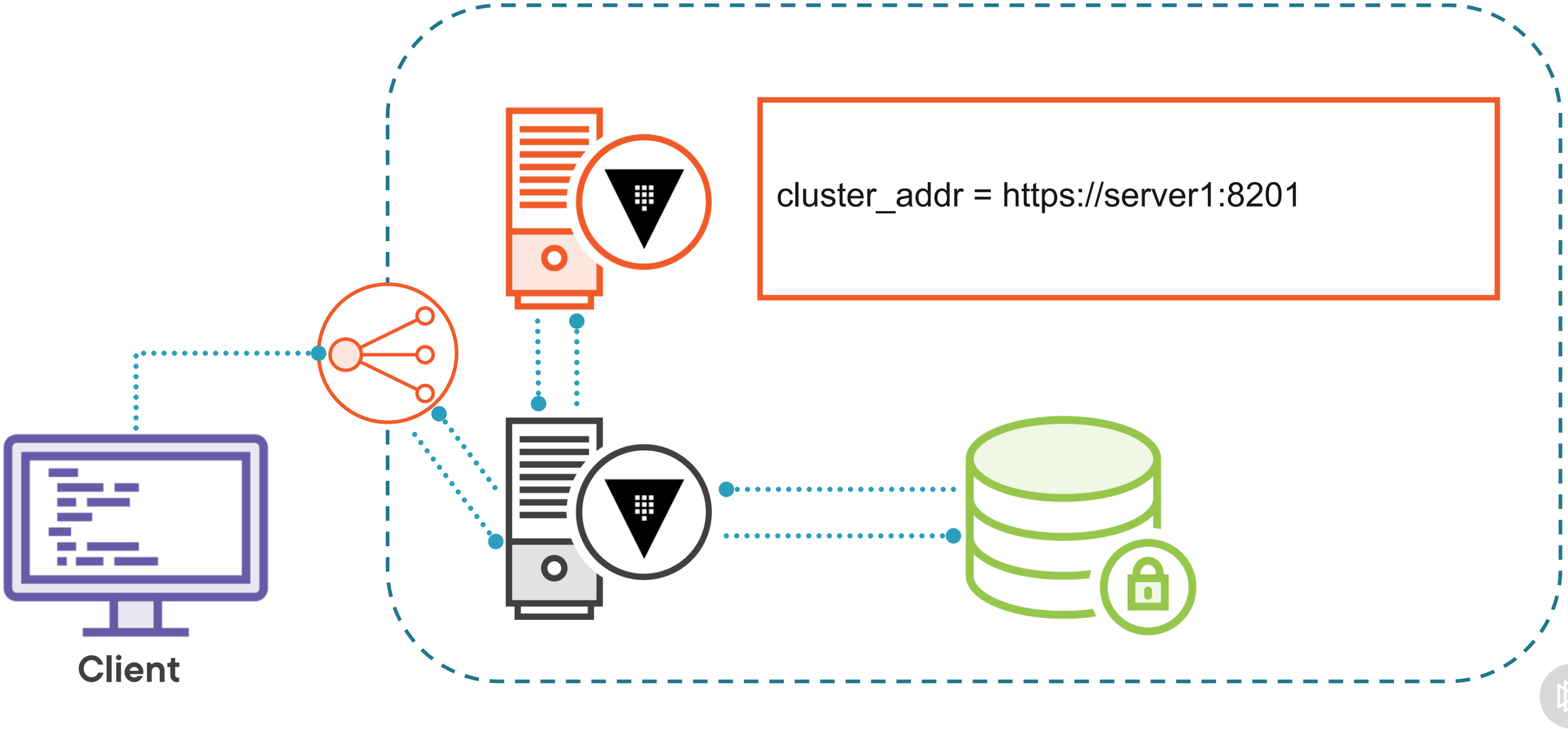
cluster_addr = https://server1:8201
api_addr = https://server1:8200

cluster_addr = https://server2:8201
api_addr = https://server2:8200

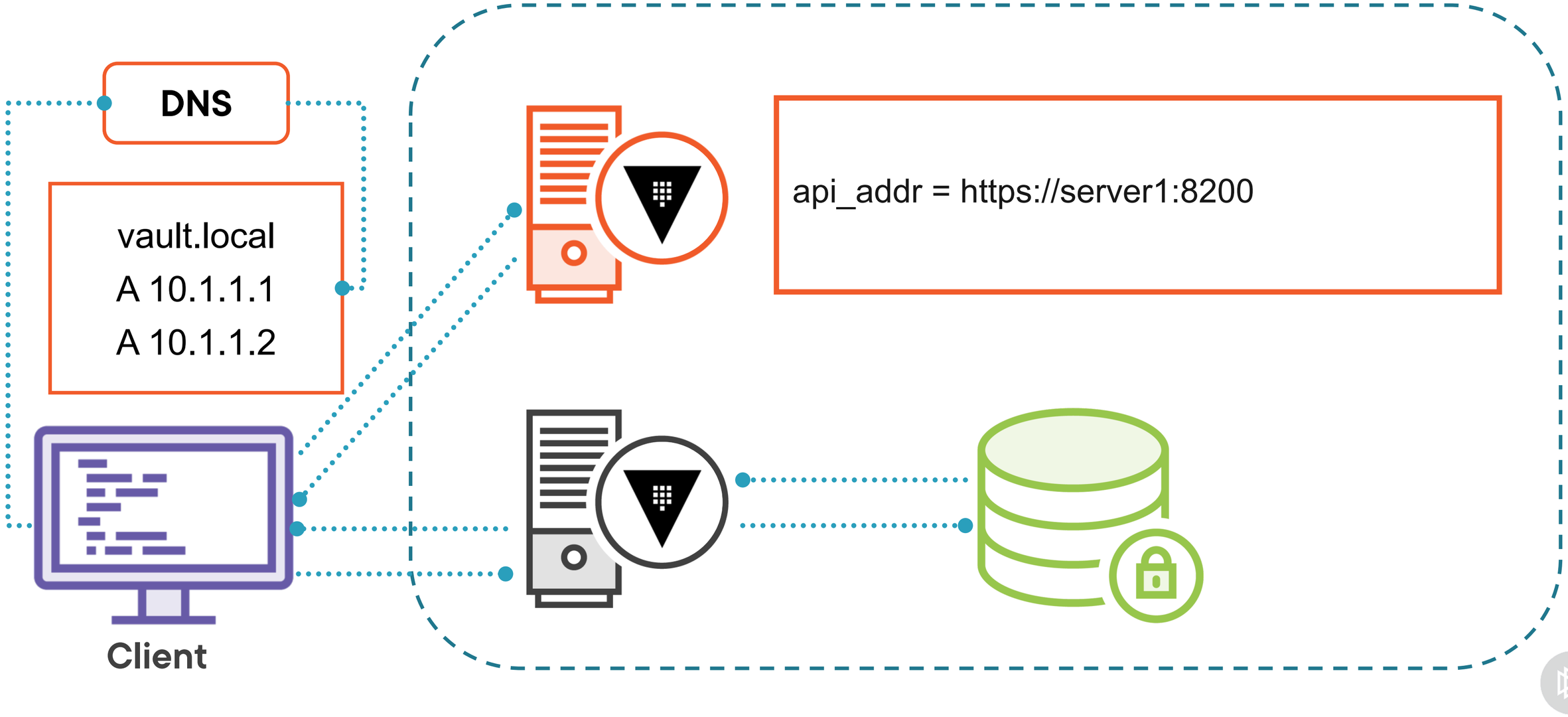
Network Traffic - Request Forwarding



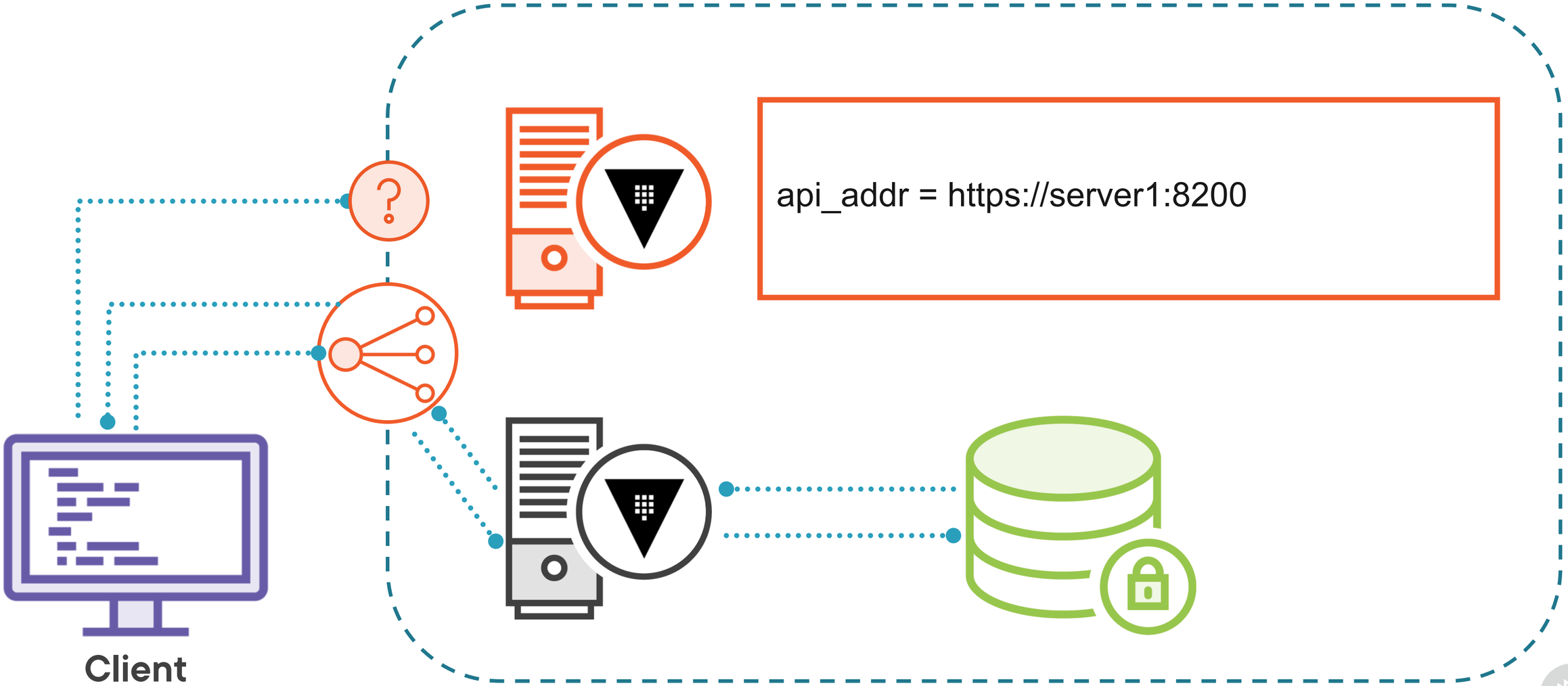
Network Traffic - Request Forwarding



Network Traffic – Client Redirection



Network Traffic – Client Redirection



Globomantics Scenario

Use Case

- Vault services need to be highly available
- Storage backend must be HA capable and HashiCorp supported
- Minimize external dependencies

Solution

- Select Integrated Storage as the backend
- Deploy at least three nodes for HA
- Use DNS round-robin and request forwarding

Vault Replication

Vault Replication

Enterprise only

Cluster is unit of replication

Replication is one-to-many

Replication is asynchronous

Replication Options

Disaster recovery

Replicates tokens and leases

No requests to secondaries

Performance

Replicates data only

Read-only requests allowed

Globomantics Scenario

Use Case

- Vault services must be available within five minutes after an outage
- Current tokens and leases should be honored

Solution

- Create a secondary cluster in another data center
- Purchase an Enterprise license
- Configure disaster recovery replication

Key Takeaways

Vault clusters are composed of one active and one or more standby nodes. Storage for a Vault cluster must support HA for locking.

Client requests can be handled through request forwarding (default) or client redirection.

Vault replication is an Enterprise feature and occurs between a primary cluster and one or more secondary clusters.

Disaster recovery clusters synchronize all data and cannot service requests.

Performance clusters do not synchronize tokens and leases and can service read-only requests.

Up Next: Working with the Identity Engine