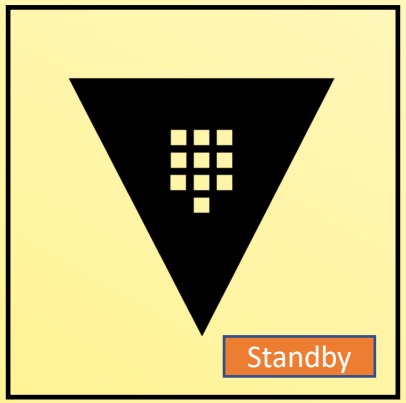
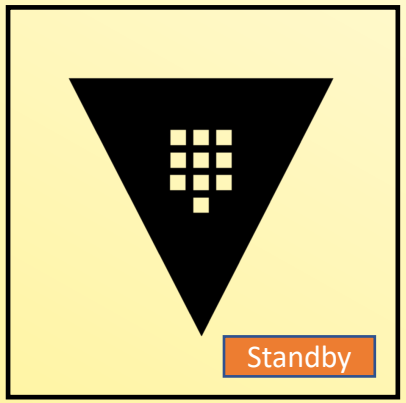
Describe the Use Cases of Performance Standby Nodes

Vault Clustering - OSS

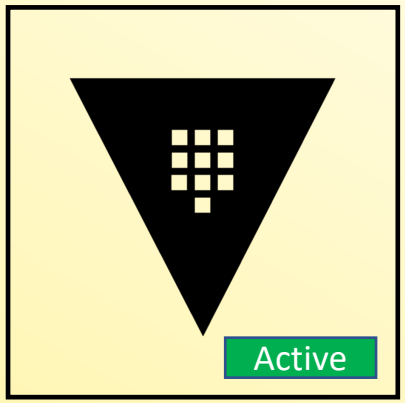
Vault Node A



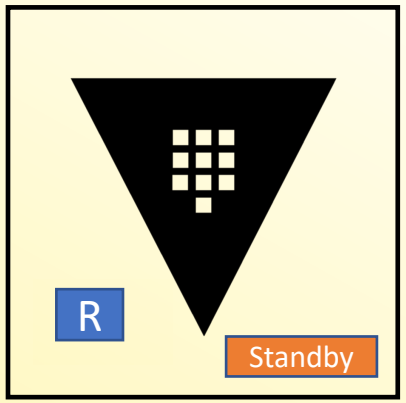
Vault Node B



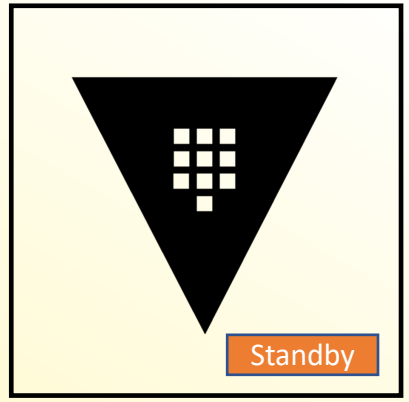
Vault Node C



Vault Node D



Vault Node E



√ read
√ write

X read
X write

Developer

~~Credential Request~~

Credential Request

Vault OSS is a scale UP application

Vault Clustering - Enterprise

Vault Node A

Performance Standby

√ read
X write

Vault Node B

Performance Standby

√ read
X write

Vault Node C

Active

√ read
√ write

Only Forward Write Request

Vault Node D

R

Performance Standby

√ read
X write

Read Response

Vault Node E

Performance Standby

√ read
X write

√ read
√ write

Developer

~~Credential Request~~

Credential Request

Vault Enterprise is a scale OUT application

What is a Read?

Any operation that does **NOT** result in a storage write is considered a **READ**

- Not necessarily limited to **HTTP GET** or `vault read` operations

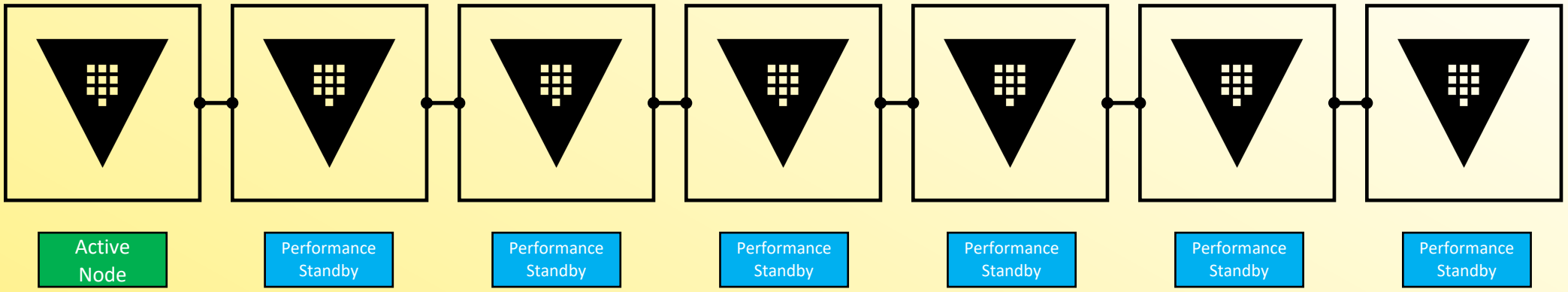
- Common read-only actions performed by applications may include:
    1. Reading secrets stored in the Key/Value engine
    2. Transit Secrets Engine – Encrypt or Decrypt operations
    3. Sign SSH client keys

Vault Enterprise with Performance Standby

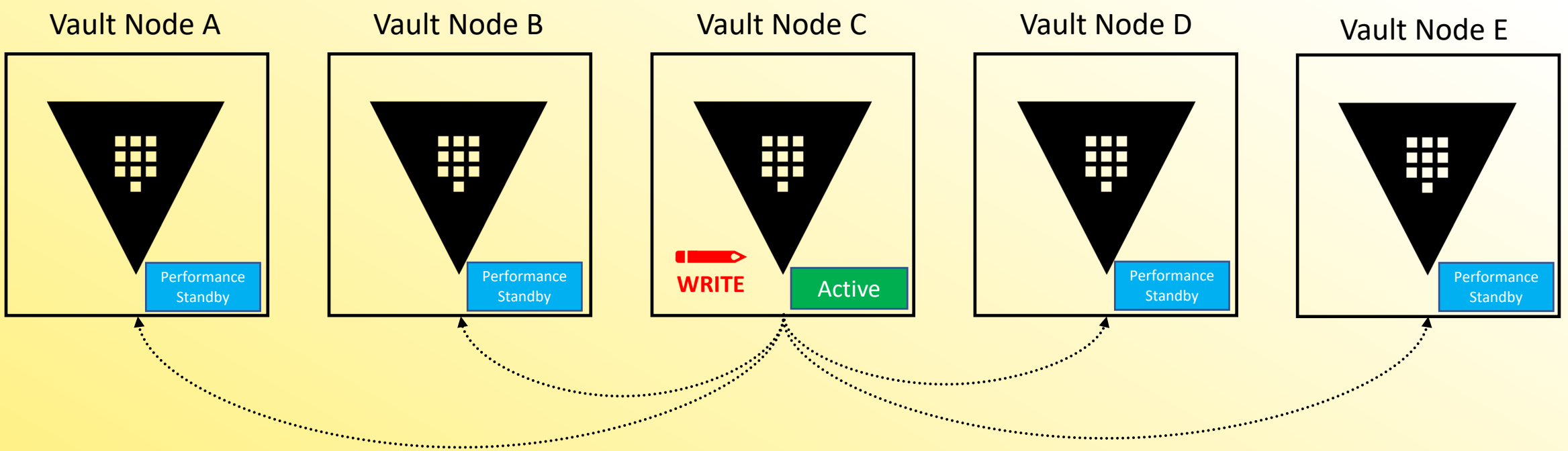
- To scale a Vault Enterprise cluster, performance standby nodes can respond to read requests from clients rather than sending the request to the Active node
- Applications known to require reads can be directed to performance standby nodes
    - this will help offload traffic from the Active node and allow you to scale **OUT** your cluster
- Performance Standby nodes can still take over as an Active node to continue providing high-availability within the local cluster

💡 **Reminder:** Performance Standby functionality is a Vault Enterprise feature

Scaling Out with Performance Secondaries

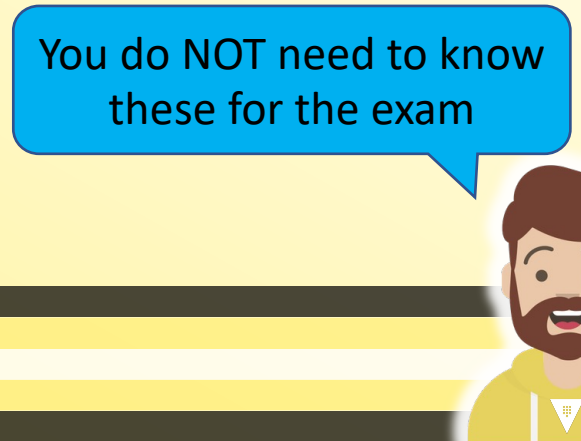


Eventual Consistency



How Do I Target a Performance Standby?

- Vault provides health information via the `/sys/health` endpoint
- Load Balancers can target specific return codes to determine an Active node vs. a Performance Standby node
- The default status codes include:
  - **`200`** – initialized, unsealed, and active node
  - **`429`** – unsealed but standby node
  - **`472`** – DR replication secondary and active node
  - **`473`** – Performance Standby
  - **`501`** – Not Initialized
  - **`503`** – Sealed node



How Do I Enable Performance Standby

It's enabled by default for Vault Enterprise - if licensed

You can disable it if you want by adding the following flag:

```
disable_performance_standby=true
```