Consul Enterprise Production Readiness

August 2022

HashiCorp

The HashiCorp Customer Success Team has created this document to augment the content presented in the Enterprise Onboarding Program. This guide is a starting point for reviewing your installation and environment and creation of a Consul runbook. This checklist is a collection of suggestions and best practices documents, please use it as a foundation to review your environment and adjust to your business needs, requirements and operating conditions. This document comes with no warranty.

Architecture and Infrastructure
Are all cluster architectures aligned with the reference architecture and best practices? • Consul Reference Architecture • Consul & Kubernetes Reference Architecture • Is autopilot configured with redundancy zones properly configured? • Are voter nodes configured across multiple AZs / failure zones to provide resilience in accordance with your uptime requirements? • Are the nodes sized appropriately (RAM and CPU)?
Is LAN gossip in all data centers within the 8ms latency budget for all nodes and clients?
Consul in Kubernetes specific considerations • Is Consul installed into a dedicated Kubernetes namespace? • Have persistent volumes (PV) been configured for Consul storage • Is Consul agent deployed on every Kubernetes worker node
Is Consul agent deployed onto every Consul server node in all data centers?
Are Consul clusters being used for Vault storage backends dedicated to only this task?
Is the primary data center running as an administrative only instance (does not host services)?
Do any data centers have more than 5,000 clients?
Can secondary data centers federated via Mesh Gateways call directly to the primary data center at start up?
Have network segments been implemented where appropriate? Is there a documented standard naming scheme for network segments?

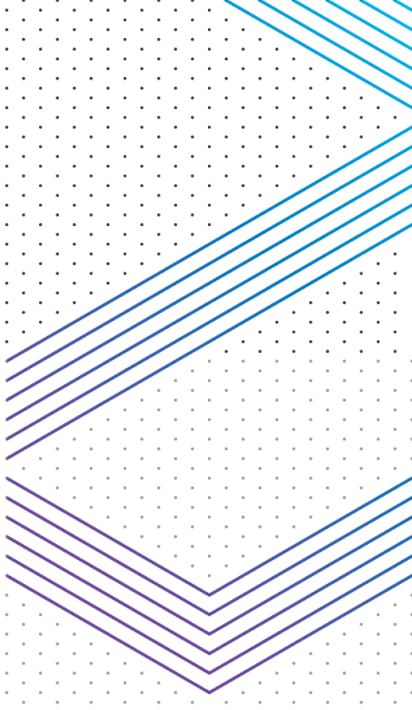
Consul Operations & Business Continuity
Has telemetry and monitoring been enabled for every Consul data center?
Have alerts been created on Consul server node health? • Disk space and file handles • RAM utilization • CPU utilization • Network activity & utilization
Have Consul key metrics been reviewed and baseline and alerting thresholds been established for the following? • Leadership changes • Autopilot health • Performance indicators • Raft timing • Stale queries • Transaction timing • DNS queries
Has audit logging been enabled on all servers and agents? • Have alerts been created for log outages/stoppage? • If a server is destroyed/lost are logs and events available post-mortem?
Are Envoy metrics and logs being collected on key endpoints and gateways?
Is NTP synchronized across clusters and Consul data centers?
Is the Consul snapshot agent configured for each datacenter? • Do snapshots have a logical naming scheme to ensure operators know which data center they are associated with? • Are Consul snapshots accessible and secured at rest? • Is there a documented retention policy in place around snapshots? • Is alerting in place around snapshot failure?
Are Consul disaster recovery (DR) operations documented and tested?
Have the 4 essential secrets been backed outside snapshots up for Consul on Kubernetes instances?
Have common DR failures been tested and recovery procedures documented • Node failure, availability zone failure, both singular and multiple

• Complete of primary data center • Loss of quorum in the primary datacenter • Loss of secondary data centers (leveraging prepared queries for resilience) • TLS certificate distribution process • ACL down policy
Are DR operations tested at least semi-annually?
Have runbooks been created around common operations? • Major and minor version upgrades • Key rotation • Adding / removing federated data center(s) • User and Operator onboarding and offboarding
Have your operations and security teams subscribed to the release notifications group and/or the security and vulnerability announcements list?

Security
Has the Consul security model been reviewed by the operations team?
Has a "*default deny*" policy been implemented for intentions
Has mTLS been configured for RPC and gossip (LAN) traffic both incoming and outgoing for every data center?
Has a supported CA been configured and agent certificates distributed to all Consul Agents?
Is there automation in place for agent certificate distribution and rotation?
Is there an automated process in place to rotate the Service Mesh Certificate?
Have Sentinel policies been implemented for KV writes
Have ACLs been enabled and tokens created for all agents and services?
Has an ACL token rotation strategy been created, tested, and implemented? Is the process scalable and as automated as possible?
Have global management tokens been created and stored in a secure location with the ACL bootstrap global token deleted?
Are admin users minting short lived tokens for admin activities instead of using global management tokens?

Has the Consul UI been secured with mTLS, HTTP write restrictions, and ACLs?
Is Consul running with a user account with proper permissions on all clients and servers (should not be running as root anywhere)?

Performance
Has the Consul Server Performance Guide been reviewed by the operations team?
Has the `allow_stale` configuration been reviewed and set for clusters that service a large number of DNS queries?
Have read replicas been implemented on clusters with heavy read traffic (a large volume of DNS traffic and/or RPC calls)
Has rate limiting for RPC requests been implemented where appropriate?
Do clusters that service Service Mesh / Consul Connect have adequate CPU provisioned to service spikes associated with certificate signing?
Have clusters been tuned for disk I/O, and available RAM following the tuning guide?
Has the `raft_multiplier` value been properly set on all servers?



USA Headquarters

101 Second St., Suite 700, San Francisco, CA, 94105
www.hashicorp.com