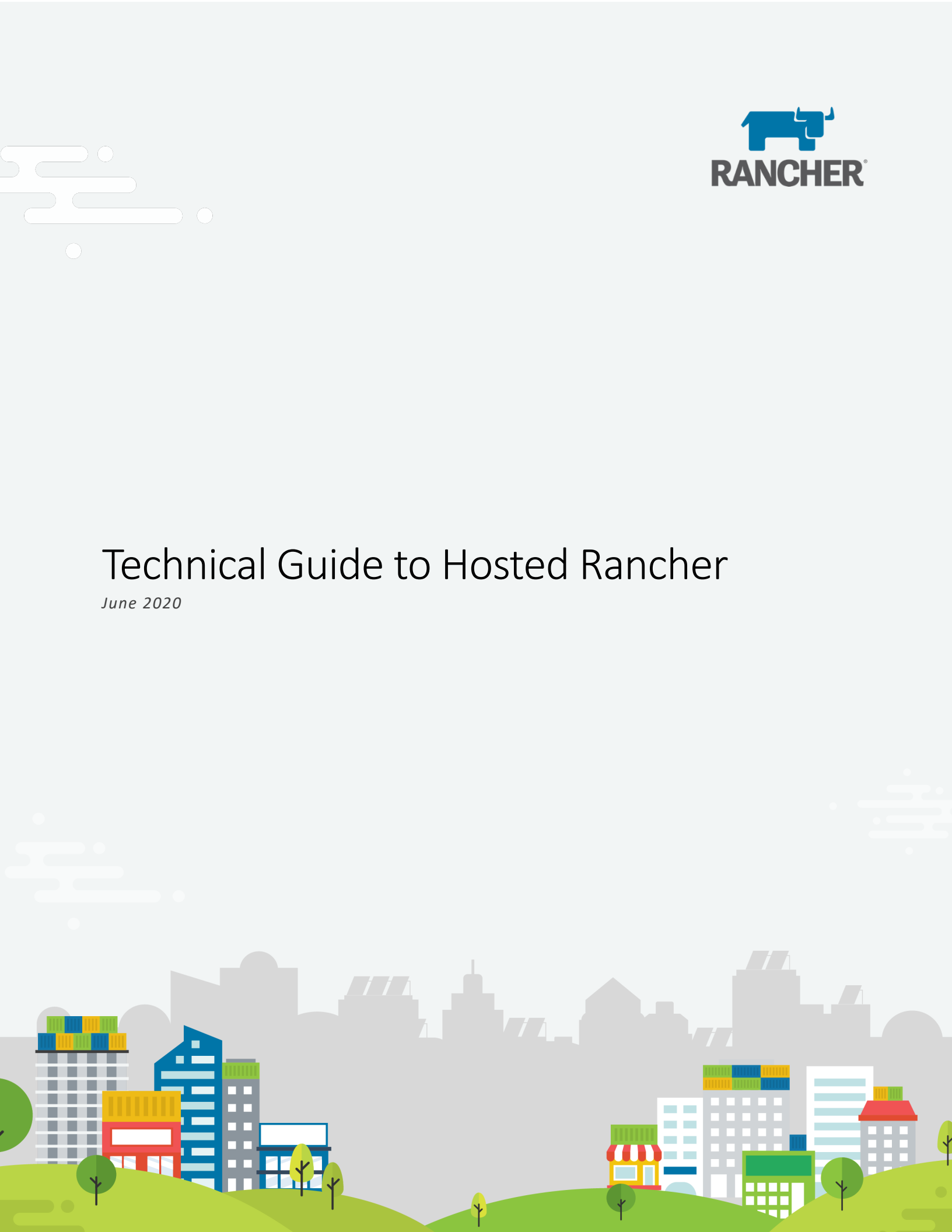


Technical Guide to Hosted Rancher

June 2020

Contents

1 Overview

Hosted Rancher is a new service offering from Rancher Labs. It is a fully managed cloud-hosted service of Rancher, Rancher Lab's popular open source Kubernetes management platform.

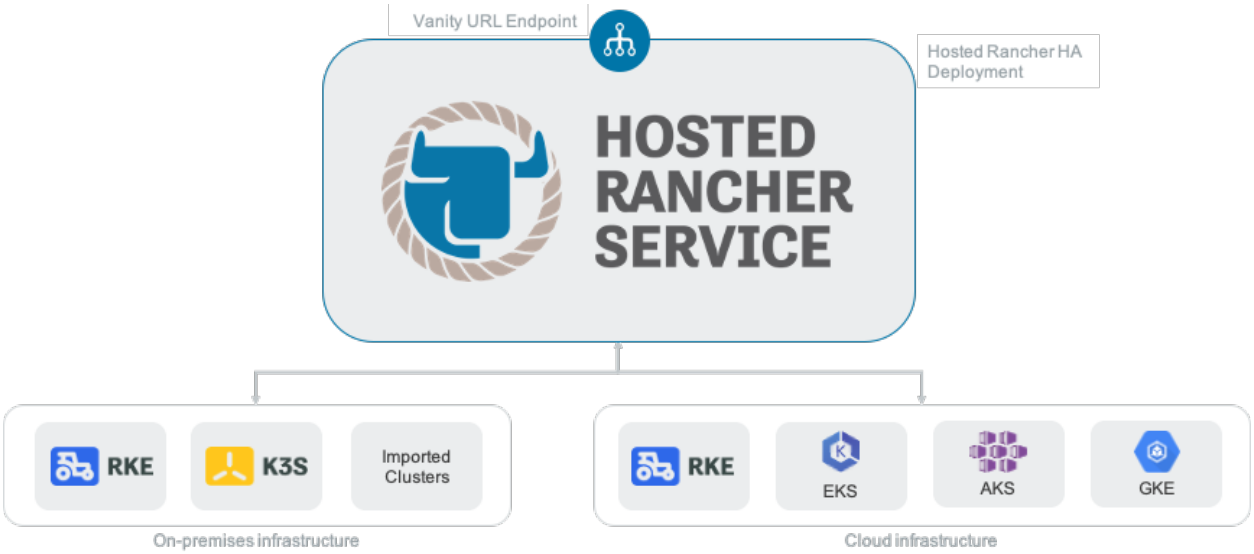
Hosted Rancher alleviates DevOps teams from the burden of installing and operating Rancher and allows them to focus on managing their downstream clusters and application workloads.

Rancher Labs manages all aspects of the service, including uptime, monitoring, logging, security, upgrades, patches, backups, restores and anything else associated with operating a cloud-based service. This reduces total cost of ownership and improves business continuity for customers. We fully leverage the Rancher open source project and Hosted Rancher does not include any proprietary or closed-source components in our service offering.

2 Architecture

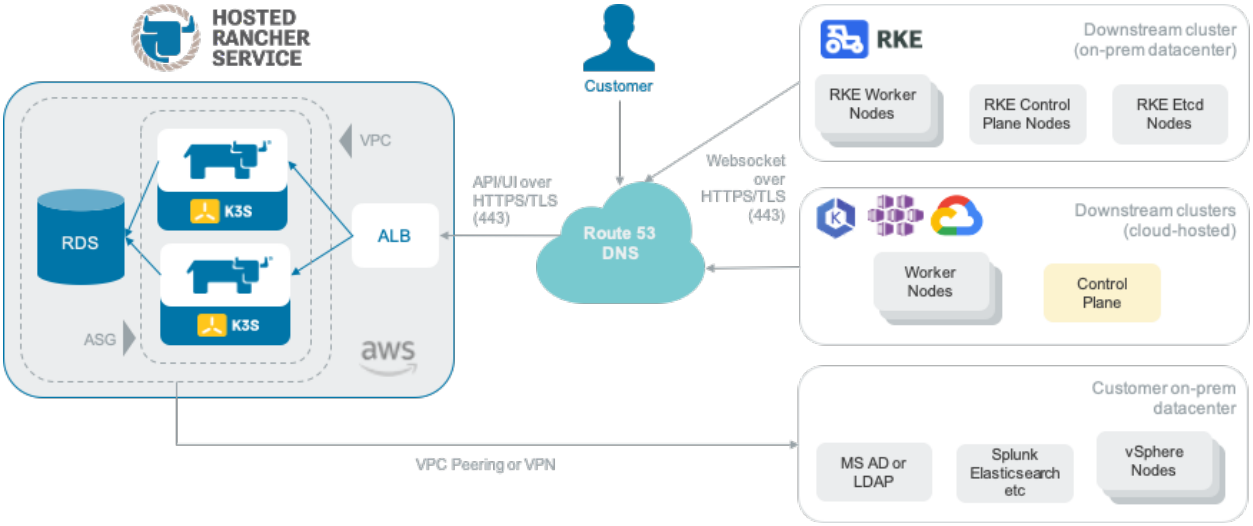
Hosted Rancher was architected from the ground up with security and high availability in mind. At a high level, the Hosted Rancher service is designed to manage any CNCF-certified Kubernetes distribution including RKE, K3s, EKS, AKS, and GKE to build clusters in the cloud or in on-premise datacenters.

2.1 Typical Rancher Deployment



The following diagram shows more detail on the Hosted Rancher inner workings along with how it can interoperate with downstream clusters and other infrastructure.

2.2 How Rancher Works



Hosted Rancher is built on top of K3s, containerd and RDS using Aurora. Our service fully leverages virtual private clouds (VPC), auto scale groups (ASG), application load balancers (ALB), security groups, virtual private networks (VPN) and Route 53 within the AWS Cloud to ensure security and high availability. In a standard setup, downstream clusters only require outbound HTTPS/TLS over port TCP/443 to communicate with Hosted Rancher. Once the node agents and cluster agents running on the downstream clusters have established the TLS websocket connection, Hosted Rancher can fully manage the Kubernetes clusters and nodes. See Rancher's Architecture Overview for more details on the Rancher architecture.

Active Directory integration is a fairly common use case for Rancher authentication. For customers set up on Azure AD, there are no additional steps that need to be followed beyond the integration guide in the Rancher docs. Integration with Active Directory within a corporate network will require an AD endpoint is made available to Hosted Rancher. This can be done by opening firewall ports on your corporate network, establishing a VPN connection between Hosted Rancher and your corporate network, or through VPC peering.

For VPC peering and VPN configurations, more details and diagrams are provided in the Security section of this architecture whitepaper.

3 Uptime, Support, and Service Level Agreement

The Hosted Rancher operations team strives to keep the service up and running 24 hours a day, 7 days a week, 365 days of the year. We use best-of-breed services such as Datadog and Pingdom to monitor the uptime and performance of the service. Both services take measurements every minute from multiple datacenters spread across the globe. Both services will immediately alert our operations staff if the Hosted Rancher service fails to respond to an HTTPS-based health check request within five seconds or less.

Hosted Rancher customers are provided with the same support offered to our platinum level customers. This entitles customers to file support tickets for problems, questions or RFEs for both Hosted Rancher

and the downstream clusters that it manages. Our operations staff may be able to diagnose and resolve issues with Hosted Rancher with little or no involvement from the customer's operations staff. Downstream clusters are likely to require more involvement from the customer's operations staff to resolve.

Hosted Rancher offers a comprehensive 99.9% ("three nines") service level agreement (SLA). This is a real SLA, meaning it's a legal document that is agreed upon when a customer signs a master services agreement (MSA) with Rancher Labs. All the details regarding how the uptime is calculated, exclusions, and service credits are outlined in the MSA.

4 Fault Tolerance / High Availability and Scalability

Hosted Rancher was built with fault tolerance, high availability and scalability in mind to ensure the best uptime possible. Various constructs within AWS have been used to achieve this. All Hosted Rancher environments run on a minimum of two virtual machines, placed in separate availability zones. If required, these virtual machines can be scaled up by shifting to an instance with more vCPUs and memory. They can also be scaled out by adding additional instances that join the Kubernetes cluster. VMs are closely monitored to determine if scaling is necessary. Scaling can also be done in real-time without requiring a service outage.

Hosted Rancher also uses AWS's Auto Scale Group (ASG) feature. Both of the virtual machines running Rancher are contained in the ASG. The VMs inside the ASG are monitored by AWS using the Rancher health check endpoint. If the health check endpoint fails, the ASG will automatically terminate the virtual machine and launch a new one. Hosted Rancher VMs are configured to automatically join the K3s cluster at boot time. To summarize, the ASG configuration allows Hosted Rancher clusters to self-heal if a single Rancher instance stops responding.

Elastic Load Balancing (ELB) is also used in combination with the ASG to ensure high availability. The load balancer will direct traffic to all heathy Hosted Rancher instances, which in most deployments is two VMs. In the scenario where Rancher stops responding, the ELB will stop directing traffic to the unhealthy instance and only direct traffic to the healthy instance. The ELB's health check is integrated with ASG, which will trigger the termination of the unhealthy VM and launch of the replacement VM.

5 Maintenance

One of the significant advantages to using Hosted Rancher is the ongoing maintenance of your Rancher environment. This starts with the initial spin-up of your environment in the cloud. Rancher Labs has done extensive work to automate all the initial setup of your Rancher cluster. This includes the base Ubuntu operating system with the latest patches, the latest version of K3s and latest version of Rancher. We also take care of provisioning the networking such as the VPC, subnets, security groups, and load balancer. An HTTPS/TLS certificate signed by a public certificate authority is also created based on the ‘rancher.cloud’ vanity hostname you have selected. All of this can be provisioned in less than an hour. If your environment needs to be brought down entirely and relaunched, that can also be done quickly through automation.

Once you are up and running with Hosted Rancher, we also take care of the ongoing maintenance of your cluster. Operating system updates and patches are done transparently by rolling image upgrades. The underlying K3s cluster that powers Rancher can also be upgraded with no downtime required. For Rancher upgrades, we coordinate with each customer on the timing and upgrade version. We generally try to keep all customers on releases no older than three months. In the future, we plan to allow self-service Rancher upgrades through the user interface.

6 Monitoring

Monitoring is a critical component and a significant value proposition for Hosted Rancher. We highly leverage the built-in monitoring and alerting features in Rancher. This includes Prometheus and Grafana that alerts our staff using PagerDuty and Slack. We have a 24/7/365 schedule rotation with escalations to three levels up our organization chart. We monitor metrics such as CPU utilization, CPU load, memory utilization, and disk space very closely. We are continuously adjusting our monitoring and alert thresholds that are too sensitive or not sensitive enough.

In addition to Rancher's built-in monitoring, Datadog and Pingdom are used to monitor uptime and are also integrated with our PagerDuty on-call support schedule. AWS CloudWatch is used to monitor health metrics outside Rancher such as RDS, ASG, and ELB components.

7 Logging

Hosted Rancher uses various logs for auditing and troubleshooting purposes. The Rancher deployment has Rancher audit logs turned on at level 2. These audit logs contain the API requests, including UI JSON calls, against the Rancher server. Level 2 includes log event metadata and request body but does not include response metadata and response body. Rancher APIs typically do not contain personally identifiable information (PII). One exception is the user API which can contain a user's full name. More details on Rancher's API audit logging can be found in the Rancher Documentation. Audit logs are stored locally on each virtual machine and retained for up to 7 days. Each virtual machine resides in the region/country you have chosen for your Hosted Rancher deployment. Only the Rancher Labs operations team has access to these logs for troubleshooting purposes. Hosted Rancher customers can also get access to their audit logs upon request.

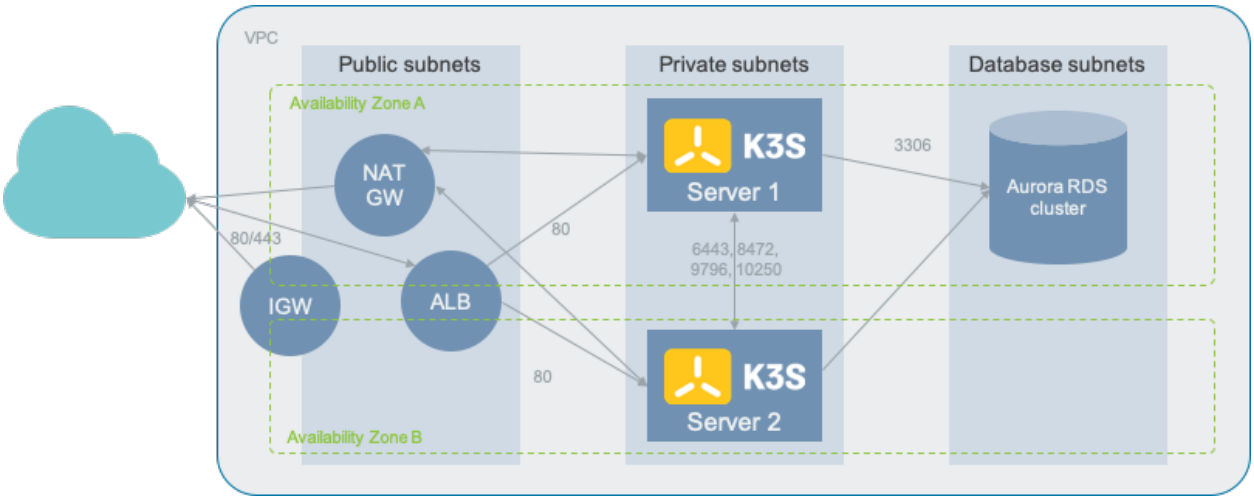
Hosted Rancher also stores application and operating system logs which are generally used for troubleshooting issues and do not contain PII. Some examples of these types of logs are K3s logs, OS messages logs, ssh/console logs, and kernel logs. Full console logging is also enabled using AWS Systems Manager, which gives a full log of who logged into a Hosted Rancher VM, which commands they executed, and the output from those commands.

8 Security

Rancher Labs incorporates security best practices whenever possible and feasible. All endpoint access to Hosted Rancher is done through HTTPS using TLS 1.2 encryption, which is supported by all modern browsers and API tools/clients. This is the primary access point for the Hosted Rancher service. Amazon

is used for our public root certificate authority (CA) and certificates are automatically renewed before expiration.

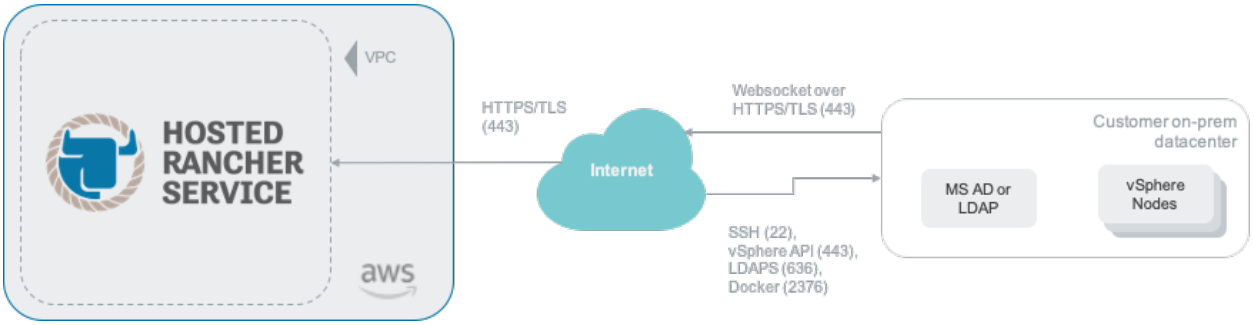
Within the Amazon Web Services cloud infrastructure, several measures are taken to properly secure the environment. First, each tenant is placed in a dedicated virtual private cloud (VPC). Each VPC in Hosted Rancher is completely isolated from each other. There are three groups of subnets in two availability zones for a total of six subnets. There's one subnet group for public-facing services such as the load balancer and NAT gateway. Another subnet group is for the RDS cluster, and the third subnet is for the Rancher K3s cluster. Instances and services in each subnet group are placed in security groups that have all ports blocked by default and only have ports opened that are required. For example, the security group containing the RDS cluster only has port 3306 open to allow database connectivity between K3s and RDS.



For customers that require Hosted Rancher to have secure access to their private cloud or on-premise infrastructure, we offer several options:

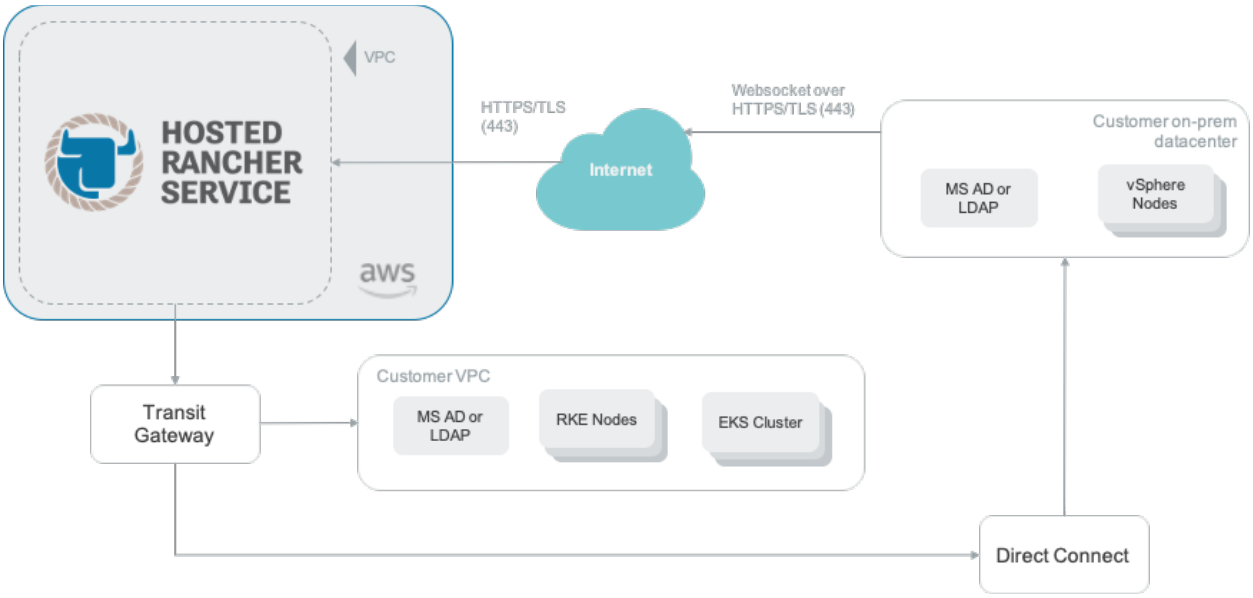
8.1 Option 1 – Port access filtered by IP address

Each customer environment has an Elastic IP address (EIP) that can be used to grant access on a corporate firewall for specific destination ports and IPs. This is the least secure option, but the easiest to set up and maintain.



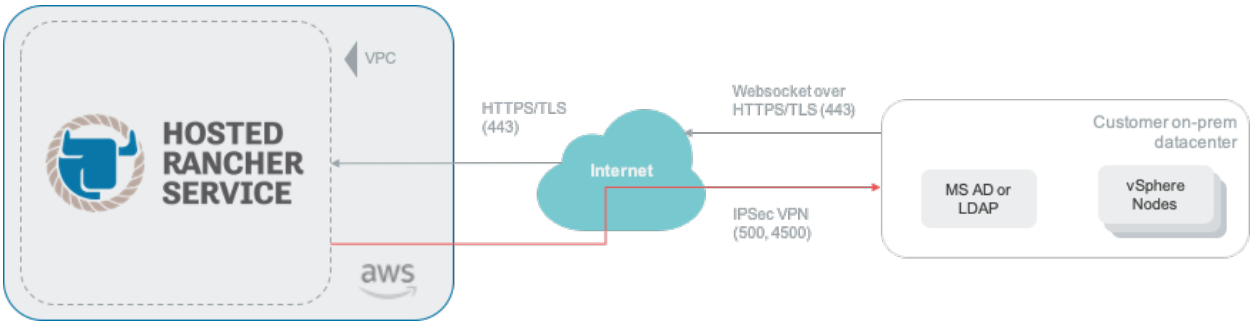
8.2 Option 2 – VPC Peering (Most Common)

For customers that currently have a presence in AWS and host infrastructure directly in AWS or have a Direct Connect established to their on-premise datacenter, this is an attractive option. Hosted Rancher's VPC can be connected to a customer's VPC using a Transit Gateway.



8.3 Option 3 – VPN

Hosted Rancher can establish an IPSec based VPN from the AWS VPC directly into a customer on-premise datacenter. Most modern firewall and routing devices support IPSec.



Rancher Labs also takes a few other measures to ensure the highest security possible in the Hosted Rancher environment. The database connection between K3s servers and the Aurora RDS database is over a TLS encrypted connection. All virtual machines running in AWS use encrypted EBS disks.

A limited number of Rancher Lab employees have access to the Hosted Rancher AWS accounts, which provide access to the cloud infrastructure, including the ability to run queries against the RDS database and console access to the Rancher K3s virtual machines. Console access is logged and a full history of commands executed and output returned is included in these logs. Access to these AWS accounts is controlled using single sign-on to the corporate Active Directory. If a Hosted Rancher operations staff

member leaves Rancher Labs, their access is immediately disabled and company assets such as laptops and mobile devices are returned to the company.

As of the writing of this whitepaper, Rancher Labs is currently pursuing SOC 2 type 1 compliance. We have engaged both an auditor and security consulting firm to plan to achieve compliance in Q3 of 2020.

9 Backups and Recovery

Hosted Rancher is operated on top of K3s clusters and utilizes K3s's capabilities to use an external data store such as PostgreSQL or MySQL. RDS's Aurora database is used, which is a high performance and highly available MySQL database. Each Hosted Rancher tenant is provided a dedicated RDS database and configured to do daily backups that are retained for 30 days. In addition to the daily backups, transaction logs are kept for 24 hours, allowing for a point in time recovery to any timestamp within the last 24 hours. Doing a restore to the last 24 hours is almost instant. Restoring from a daily backup can take an hour or longer, depending on the size of the data.

In addition to the database backups, all infrastructure is deployed, configured, and maintained using the Infrastructure as Code (IoC) methodology. All code is source controlled using git and can be used to recreate the infrastructure in the event of a catastrophic failure. Infrastructure can also be redeployed in the event a tenant wants a "hard reset."

10 Terminology

This document, the Rancher website, and other Rancher collateral references the following terms, which are defined below:

"Rancher Labs" is the company that builds, distributes, and licenses products such as Rancher, RKE, Longhorn, and K3s. We have other open-source projects such as K3OS, Rio, Fleet, and Submariner that we develop.

"Rancher" refers to the flagship Kubernetes multi-cluster management open source project. It's also sometimes referred to as "Rancher server" and is the basis for the Hosted Rancher managed service. Rancher v1.x is a legacy product reaching end of life in mid-2020. Rancher v2.x is the current product with v2.4.5 being the latest release as of the writing of this whitepaper.

"Hosted Rancher" is the managed service that hosts Rancher in the cloud.

"Custom Cluster" is a Kubernetes cluster provisioned by Rancher using virtual machines or bare metal servers by running a Docker command.

"RKE Cluster" refers to a cluster provisioned either using the "Custom Cluster" option, a node driver (EC2, Azure, Google Cloud, vSphere, etc.), or the RKE CLI.

"Downstream Cluster" refers to the Kubernetes cluster that is being managed by Hosted Rancher. This could be a custom, imported, hosted Kubernetes (AKS, EKS, GKE, etc.), or RKE on an infrastructure provider (AWS EC2, Azure, vSphere, etc.)

"Control Plane" refers to the node running the Kubernetes management components such as the kube-apiserver, kube-scheduler, and kube-controller.

"Customer" refers to the end-user company or organization that has paid for the Hosted Rancher service or is currently evaluating it. May also be referred to as "Tenant."

11 About Rancher Labs

Rancher Labs delivers open source software that enables organizations to deploy and manage Kubernetes at scale, on any infrastructure across the data center, cloud, branch offices and the network edge. With 35,000+ active users and 100+ million downloads, their flagship product, Rancher, is the industry's most widely adopted Kubernetes management platform.

For additional information, visit www.rancher.com and follow @Rancher_Labs on Twitter.

All product and company names herein may be trademarks of their registered owners.