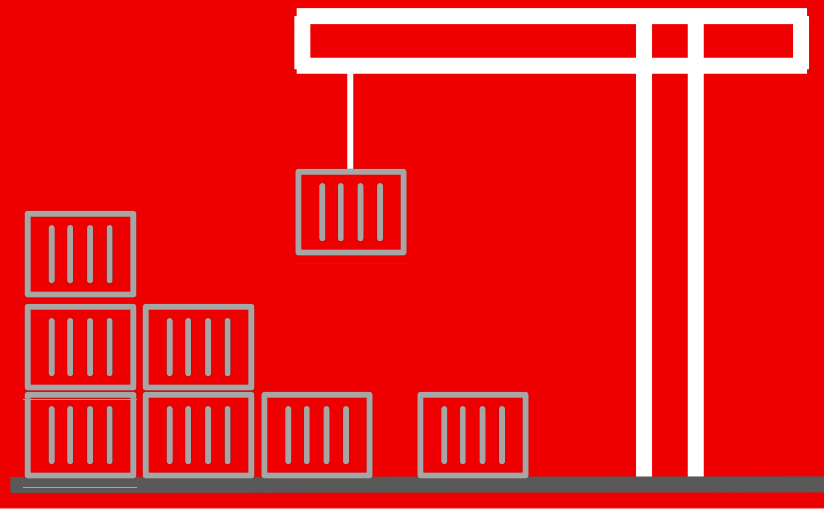
Strategic Migration Guidance, Hands-On ROSA Workshop Red Hat OpenShift on AWS

Ingolstadt, 24th of June, 2026



Yury Titov,
EMEA Senior Sp. Solution Architect,
managed Cloud Services

Red Hat

Agenda

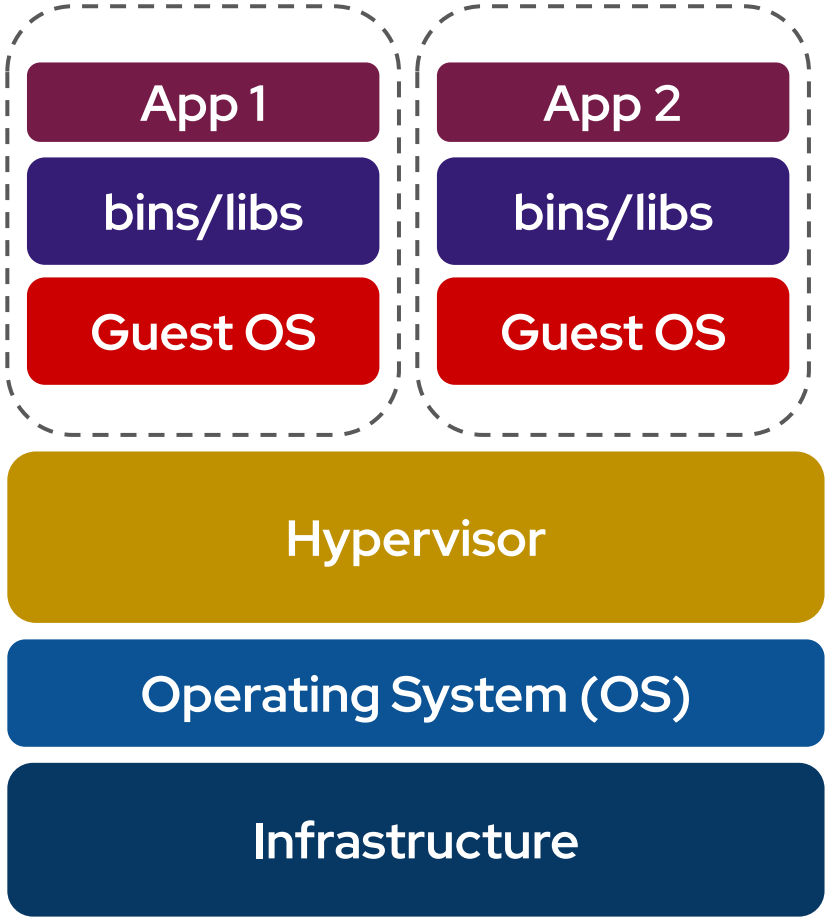
- Intro: Red Hat OpenShift Cloud Services
- Migration Guidance
- Workshop

Red Hat OpenShift Cloud Services

Red Hat

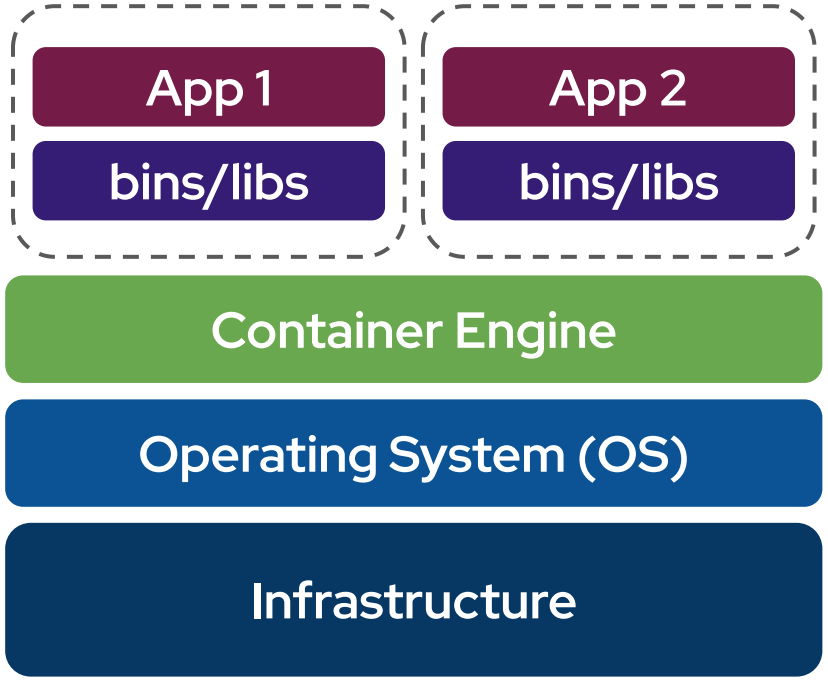
Virtual Machines vs Containers

Virtual Machine Model



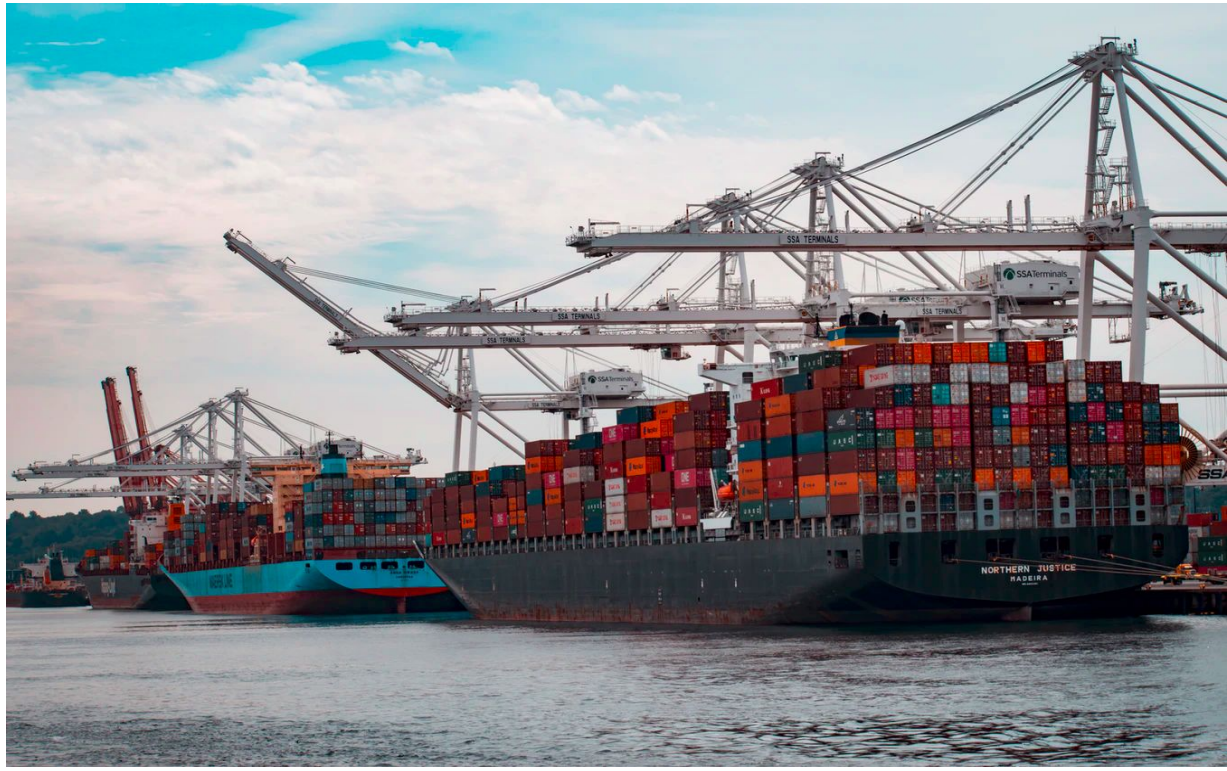
Containerized Model

- Process isolation
- Filesystem isolation
- Network isolation
- Resource limits



Red Hat

Virtual Machines vs Containers



Red Hat

Containers must be managed in the right way

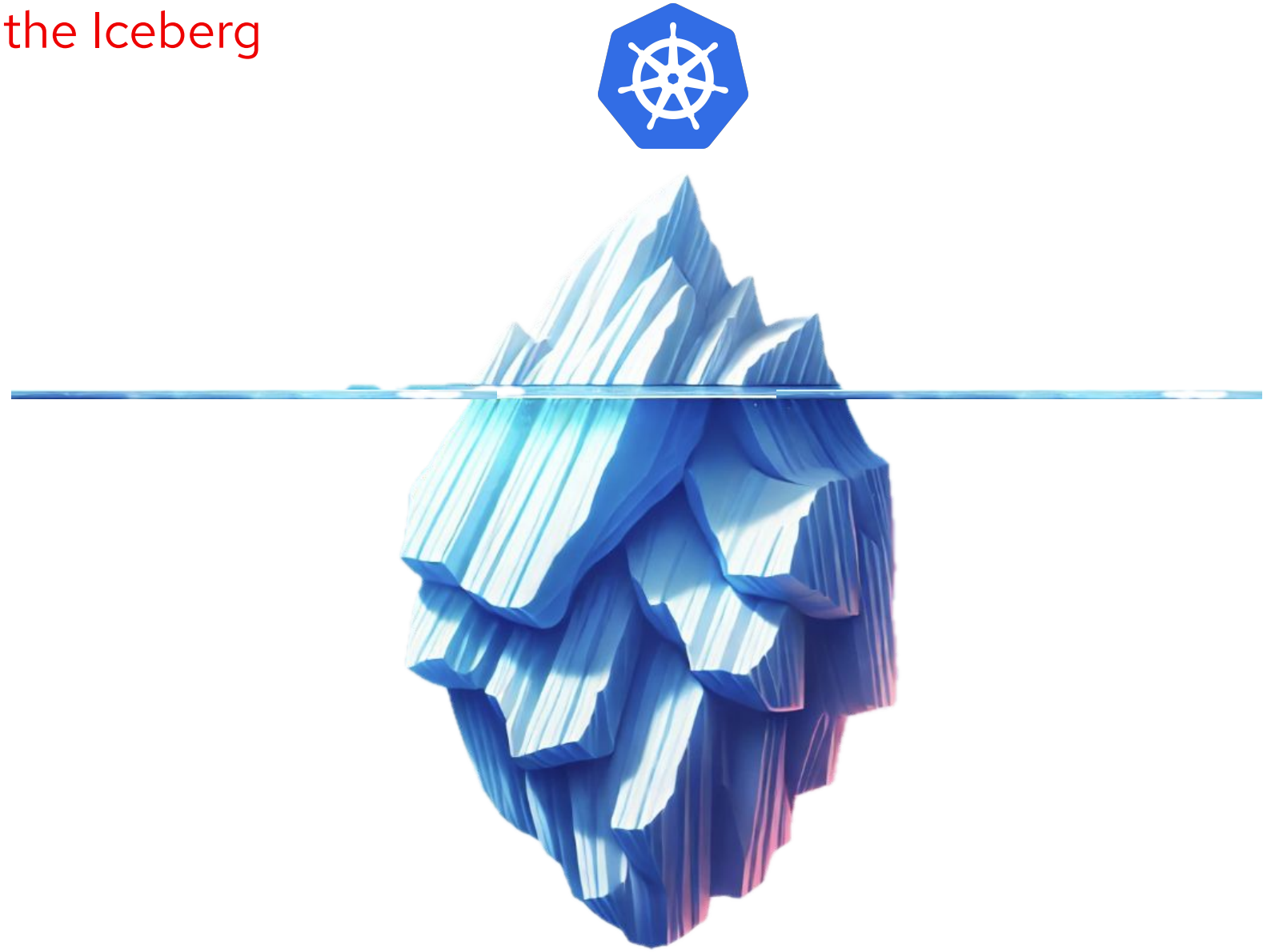
- Increased moving parts
- Interdependencies
- Infrastructure

→ Distributed Computing
→ Private Cloud (datacenters)
→ Hyperscalers specifics

Red Hat

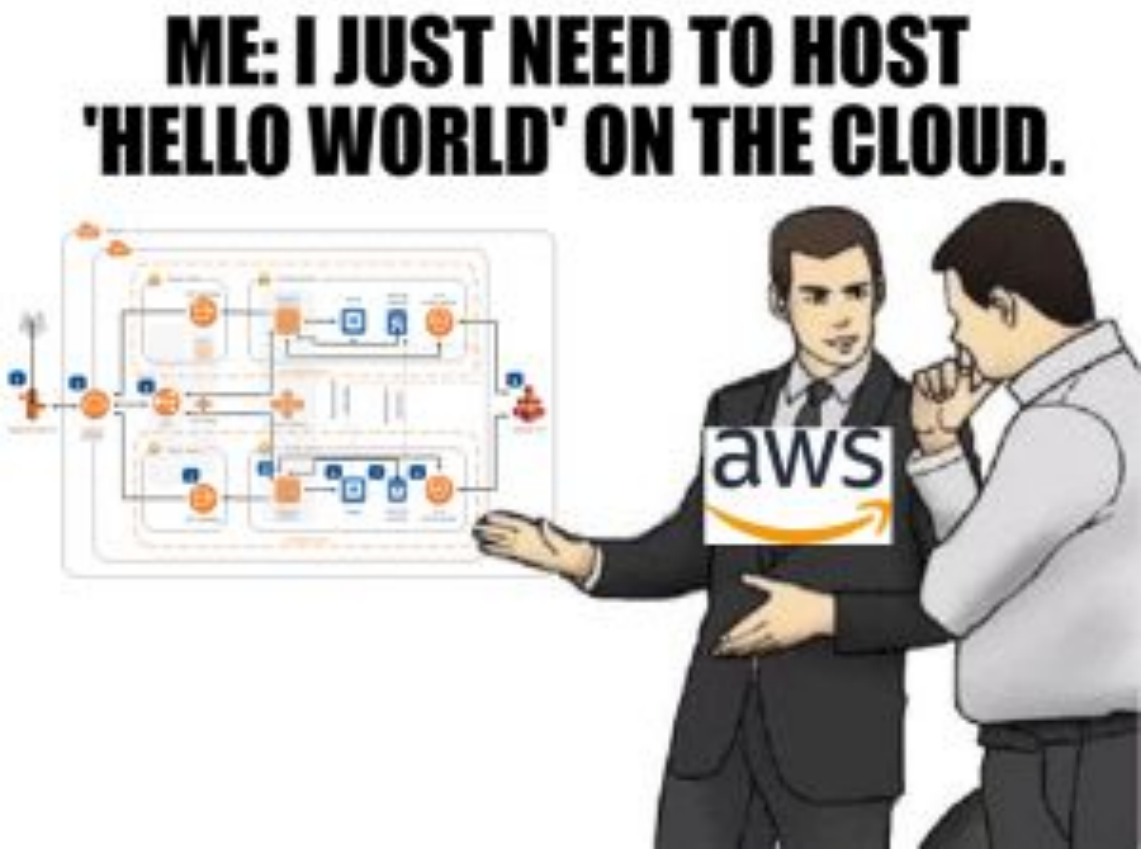
Kubernetes

Tip of the Iceberg



*AI generated image (Iceberg)

Red Hat

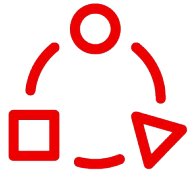


Red Hat

What's in a Modern Application Platform?

Unified platform for Dev, Sec and Ops

Transparent to developers



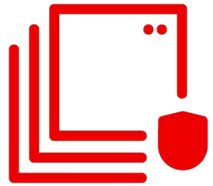
Extensible - works with what you have

Observability, management and monitoring

Runs on any infrastructure or cloud

Security configuration management and enforcement

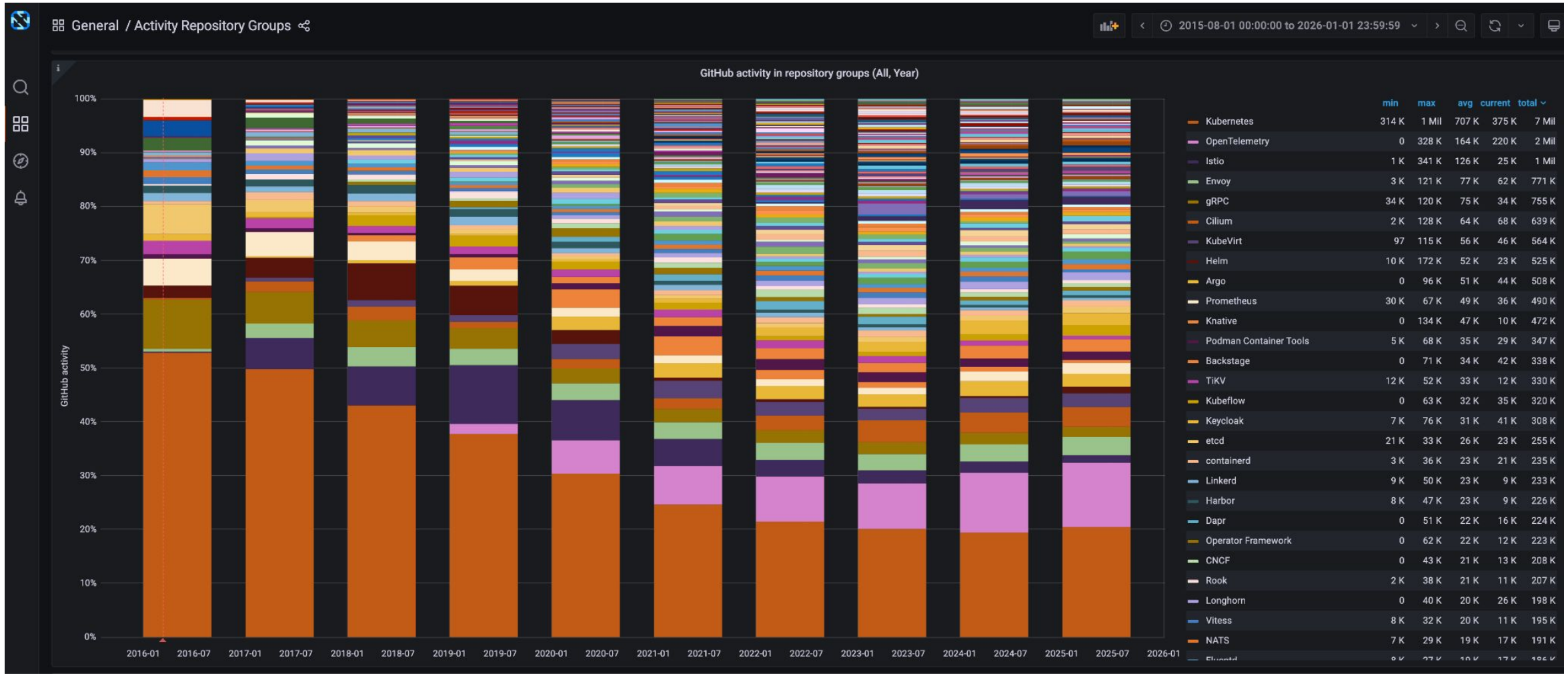
Consistent data management



Vulnerability scanning and secure image management

What is an Application Platform?

K8S and Innovation Focus on the Surrounding Areas

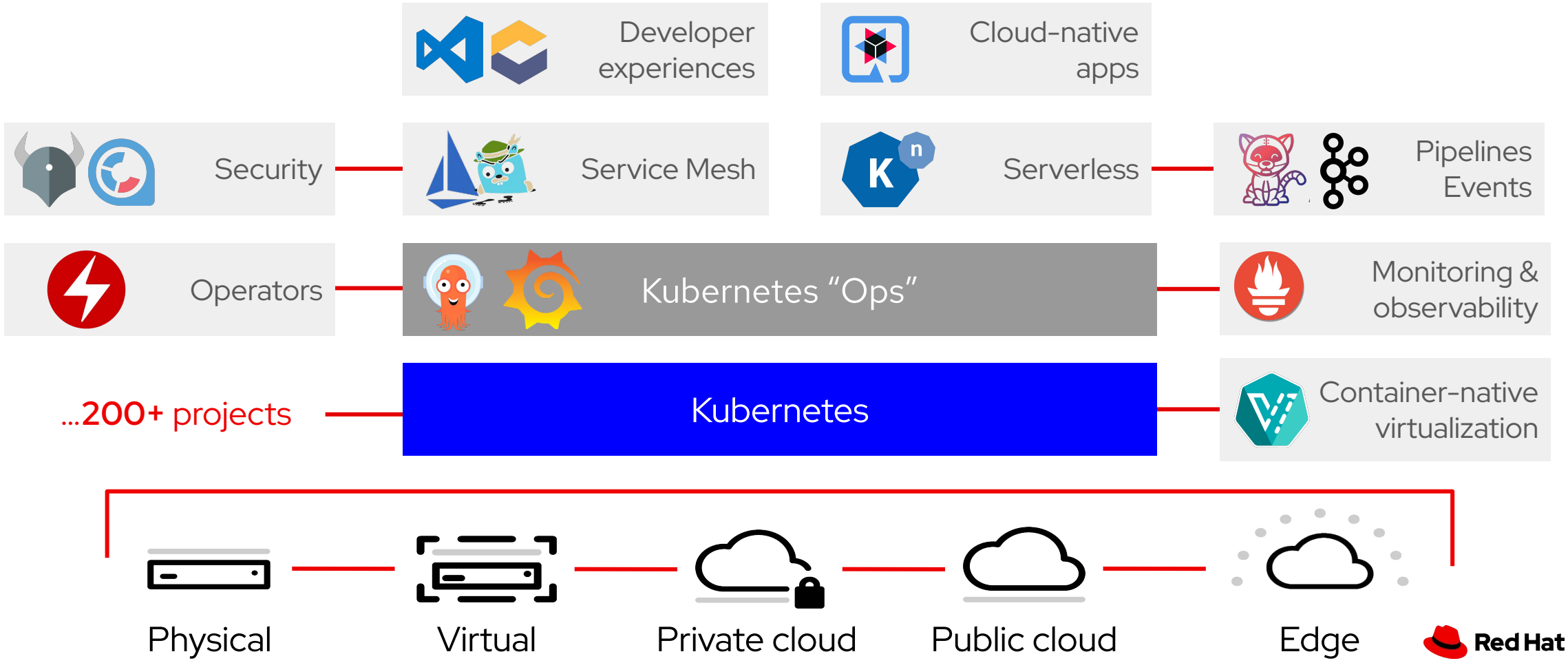


https://all.devstats.cncf.io/d/1/activity-repository-groups?orgId=1&var-period=y&var-repogroups=All&from=1438380000000&to=1767308399000

Red Hat

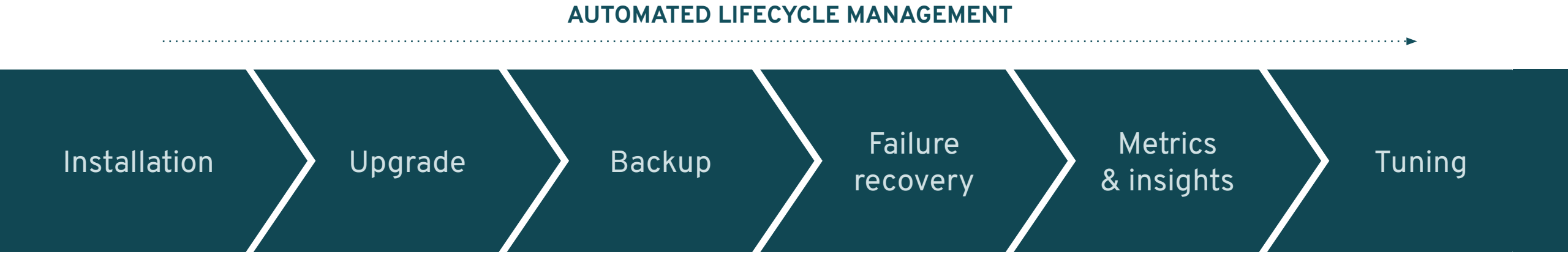
Applications require more than a Runtime!

Driven by open source projects



KUBERNETES OPERATOR FRAMEWORK

Operator Framework is an open source toolkit to manage application instances on Kubernetes in an effective, automated and scalable way.



Red Hat

An opinionated platform for building, deploying and running applications

Red Hat OpenShift

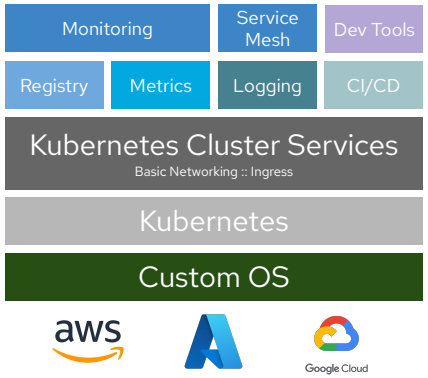
Service Mesh	App-Services	AI
CI/CD	DNS	Authentication
Monitoring	**Kubernetes**	Automation
Logging	Registry	Security
Compute	Virtualisation	Network

- Fully integrated and supported components
- Expert SRE and Customer Success support by hyperscalers
- Abstracts away technical details
- Consistent experience across clouds

Red Hat

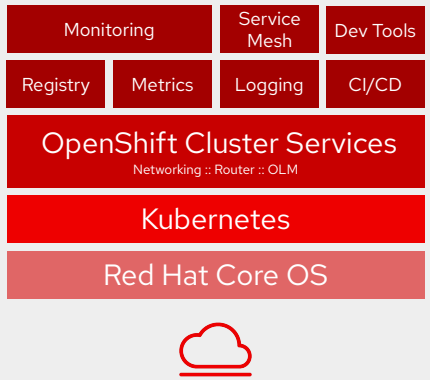
Build and run a platform *versus* using a turnkey cloud service

The Parts



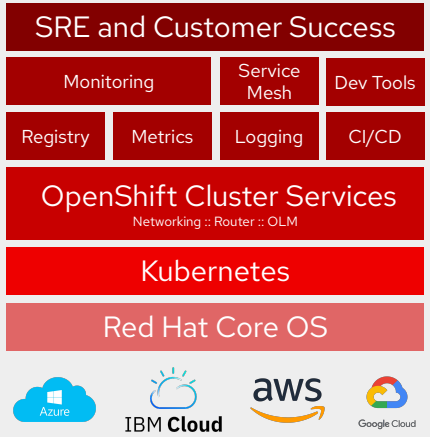
xKS + 'native' services

The Assembled Car



- Application Platform - Self-managed Red Hat OpenShift

The Car & Pit Crew



- Turnkey Application Platform - Red Hat OpenShift cloud services

"Batteries Included"

... but swappable

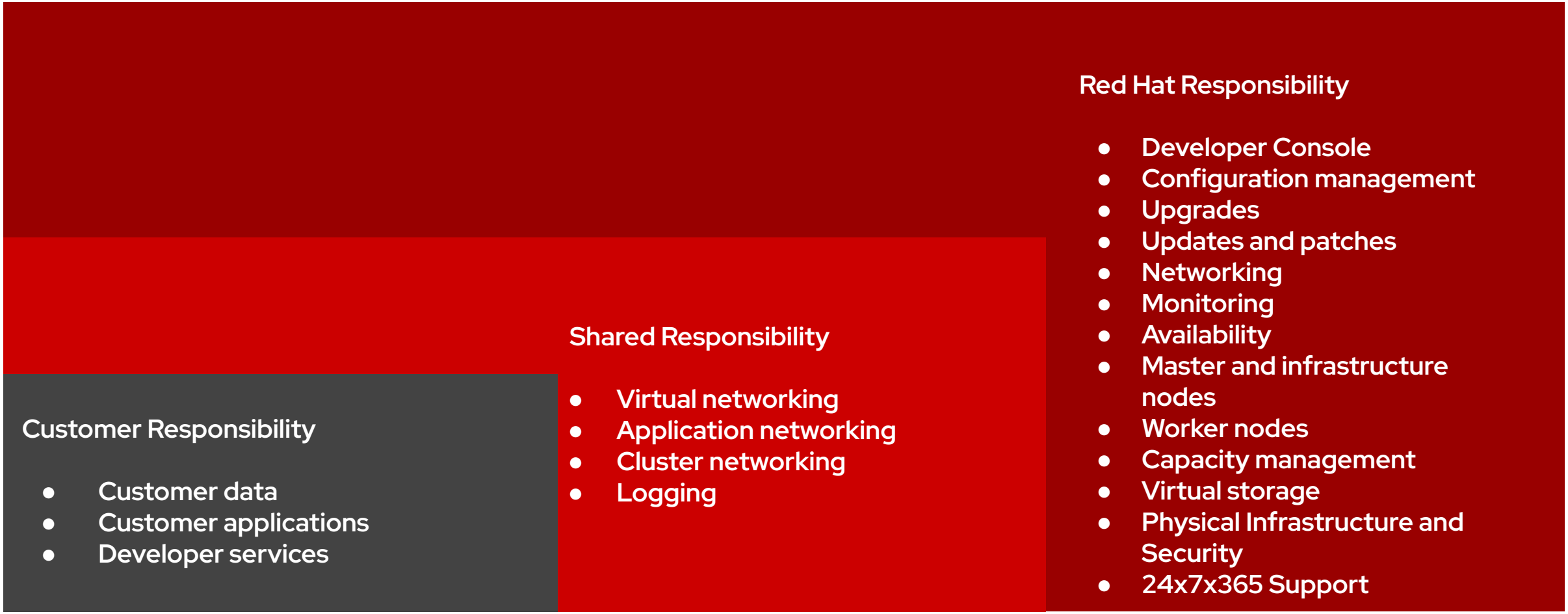
Individual components can be swapped out

Eg.

- Using AWS CloudWatch for logging on AWS
- Use specific cloud services or ISV offerings

Red Hat

Managed OpenShift Simplified Responsibility Model 99.95% SLA



Why Red Hat Cloud Services?

Jointly engineered, operated, and supported by Red Hat and Hyperscaler

Complete application development platform

Consistent OpenShift experience across the hybrid cloud

Fully managed, from infrastructure to daily operations

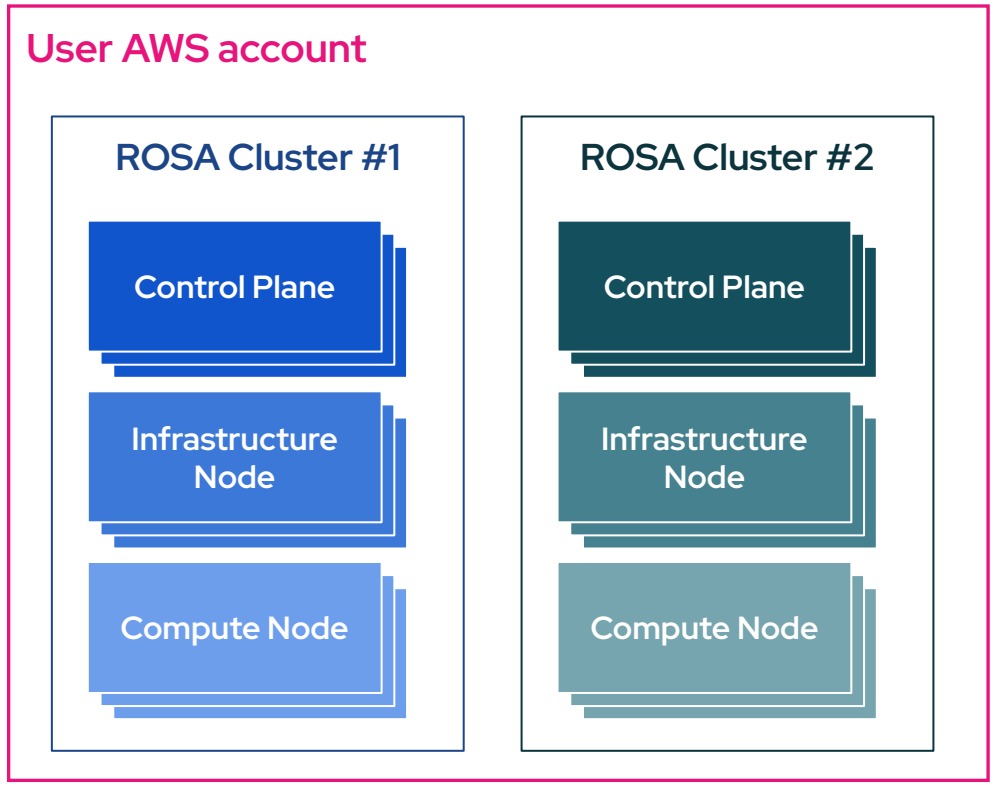
Red Hat

Migration Guidance

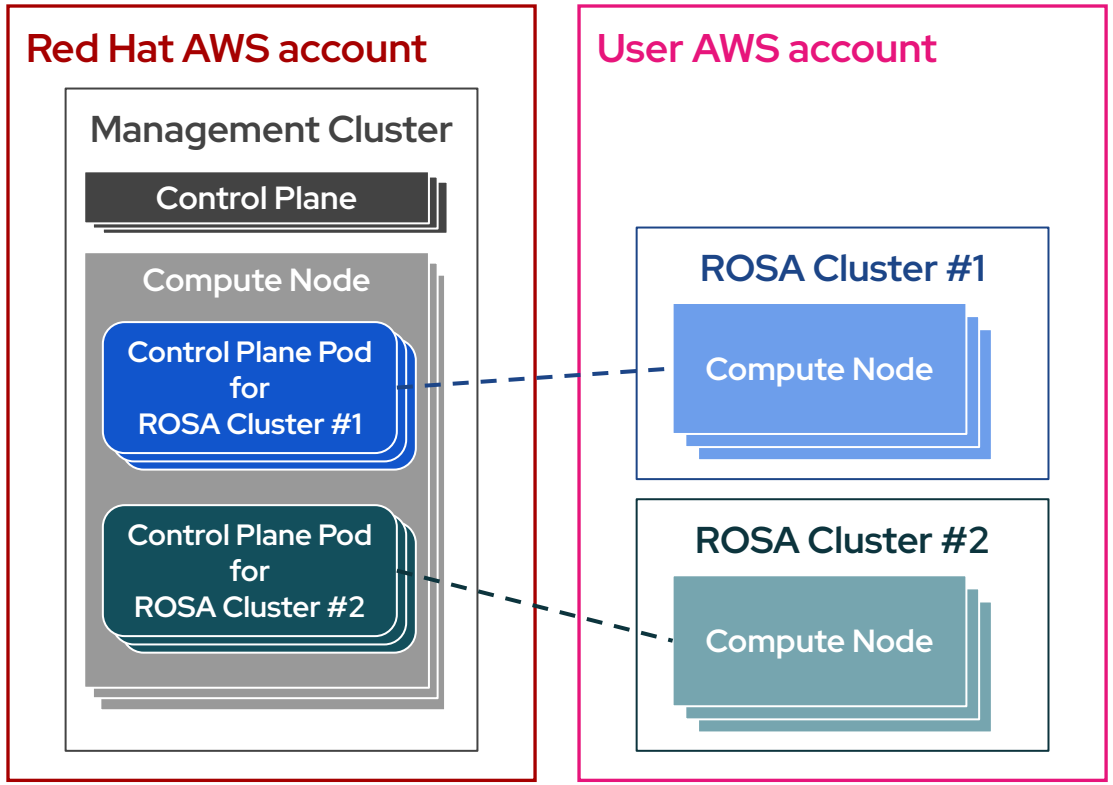
ROSA Variants

Generally Available

ROSA Dedicated/Classic



ROSA with Hosted Control Plane (HCP)



Red Hat

ROSA Dedicated vs. ROSA Hosted Control Planes

	Fully-managed Service (AWS + Red Hat joint team)	3d party Managed Service (Red Hat team only)
Architectural Dimension	**ROSA HCP (The Modern Standard)**	**ROSA Dedicated/Classic**
Control Plane Location	Hosted in Red Hat-managed AWS Account	Deployed in Customer AWS Account
Minimum Node Footprint	Only 2 Worker Nodes required	7-9 Nodes (3 Master, 3 Worker, 2 Infra)
Provisioning Speed	Rapid (~10 minutes)	Coordinated (~40 minutes)
EC2 Cost Overhead	Highly Reduced (No Control Plane Instance fees)	Full Customer Account EC2 resource rates
Upgrade Lifecycle	Decoupled (Control plane upgraded independently)	Fully Coordinated Cluster upgrade cycle

Migration Strategy & Framework

Strategic Guidance

1. Start with **business objectives**
2. Assess **application readiness**
3. Build a ROSA landing zone (**Kubika team helps!**)
4. **Pilot** with low-risk **workloads**
5. Scale through **migration waves** **(using migration toolkits)**
6. **Modernize** where business value exists

Framework:

01 / ASSESS

Catalog existing workloads, map persistent volume dependencies, and establish targeted TCO benchmarks.

02 / DESIGN

Establish high-resiliency BYO-VPC topology, configure secure AWS STS, and provision cluster OIDC providers.

03 / MIGRATE

Deploy **Migration Toolkit for containers and virt (MTC/MTV).** Conduct non-disruptive stage replication and finalize cutovers.

04 / RUN

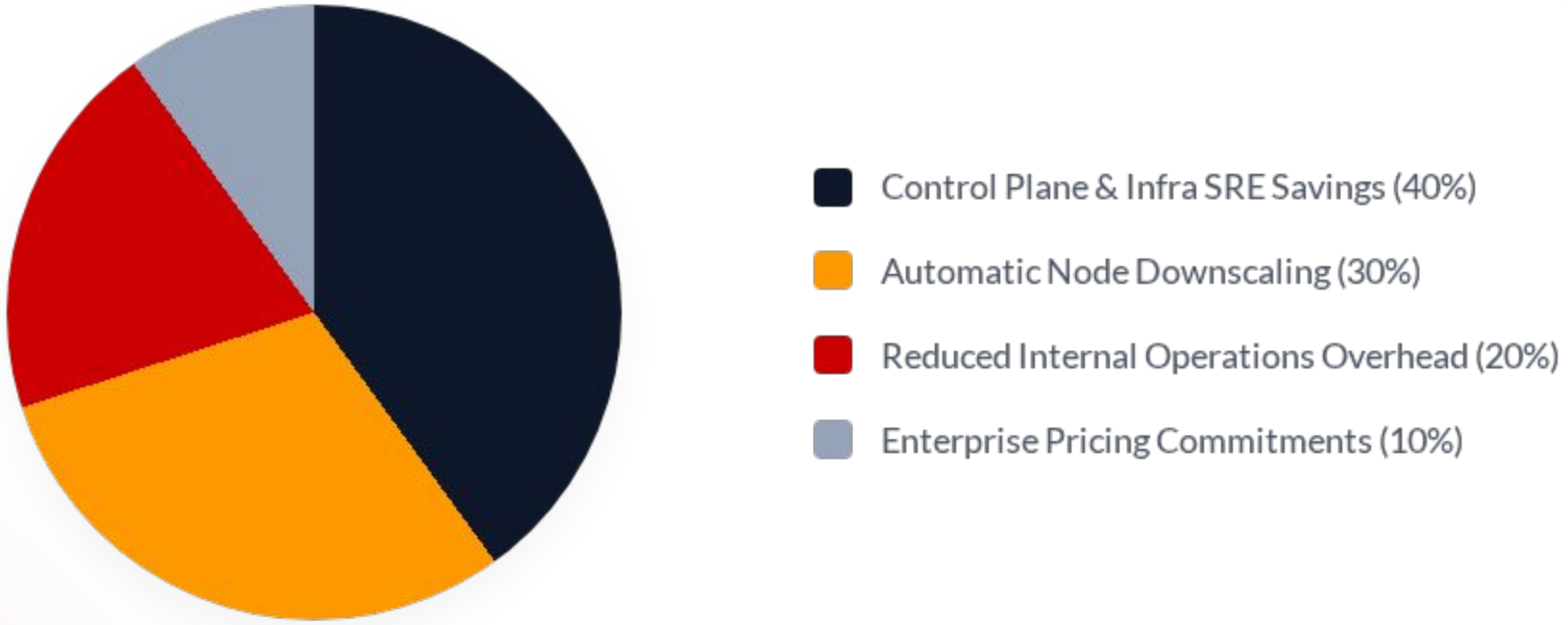
Configure autoscaling via Karpenter/Machine Pools and integrate native AWS services (RDS, S3, IAM).

Migration Best Practices

- **Leverage MTC/MTV Direct Replication** Utilize the Migration Toolkit for Containers (MTC) to coordinate namespace-level transfers from on-premise OpenShift 3/4 or alternative cloud clusters directly into ROSA.
- **Minimize Cutover Windows** Conduct consecutive background "Stage" migrations to replicate core persistent database volumes with zero application downtime, preserving "Cutover" for instantaneous DNS routing.
- **Transition to OVN-Kubernetes CNI** Standardize cluster routing with OVN-Kubernetes. Replace outdated CNI configurations with EgressFirewall, IPsec encryption, and native AWS PrivateLink secure tunnels.

Red Hat

Strategic Cost Allocation



Adopting ROSA with Hosted Control Planes (HCP) allows enterprises to instantly reclaim up to 40% of standard cluster infrastructure fees by offloading control plane provisioning and management to Red Hat.

Red Hat

Pre Optimized Model Garden

Maintained, pre-optimized models ready to deploy to production with vLLM

Broad Collection

Llama
Qwen
Gemma
Mistral
DeepSeek
Phi
Molmo
Granite
Nemotron

Comprehensive Validation

Open LLM Leaderboard evaluation scores

Benchmark	Meta-Llama-3.1-70B-Instruct	Meta-Llama-3.1-70B-Instruct-FP8(this model)	Recovery
MMLU (5-shot)	83.83	83.73	99.88%
MMLU-cot (0-shot)	86.01	85.44	99.34%
ARC Challenge (0-shot)	93.26	92.92	99.64%
GSM-8K-cot (8-shot, strict-match)	94.92	94.54	99.60%
Hellaswag (10-shot)	86.75	86.64	99.87%
Winogrande (5-shot)	85.32	85.95	100.7%
TruthfulQA (0-shot, mc2)	60.68	60.84	100.2%
Average	**84.40**	**84.29**	**99.88%**

Extensive Selection

Formats
- W4/8A16
- W8A8-INT8
- W8A8-FP8
- 2:4 sparse

Algorithms
- GPTQ / AWQ
- SmoothQuant
- SparseGPT
- RTN

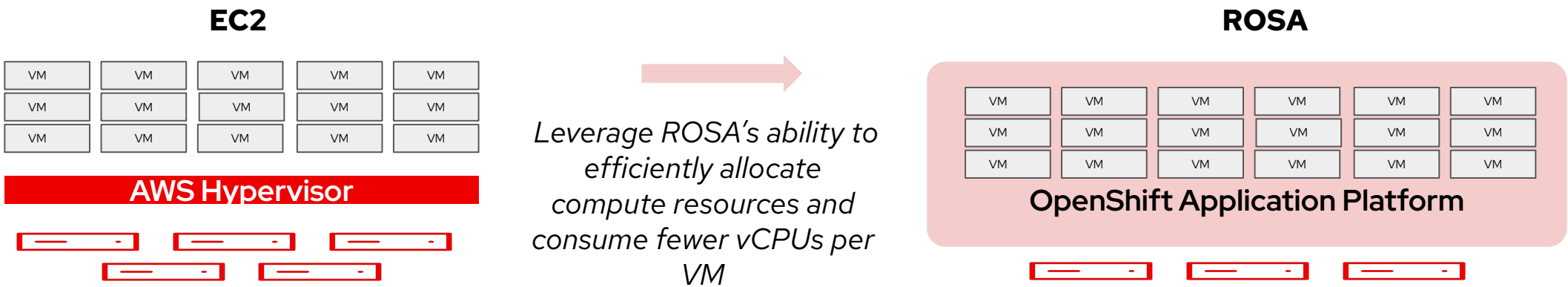
Hardware

GPUs
TPUs
Instinct
CPUs

Cut GPU costs in half with inference optimized models.

VMs: Optimizing Resource Consumption with ROSA

Choose hardware overprovisioning ratios in ROSA can optimize EC2 resources by 30%-60%

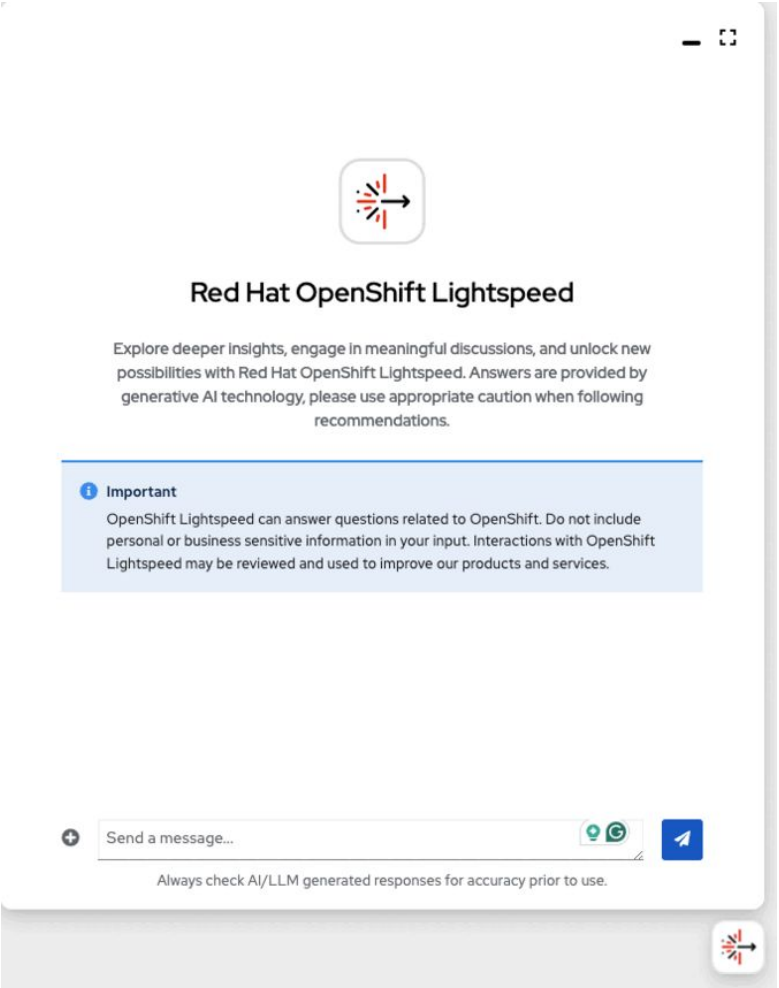


Driving the Achievement of Project Outcomes

- **Digital Transformation & Cloud Modernization**
- **Automation & Operational Efficiency**
- **Ensure Security & Resilience**

Red Hat

AIOps: OpenShift Lightspeed 1.1



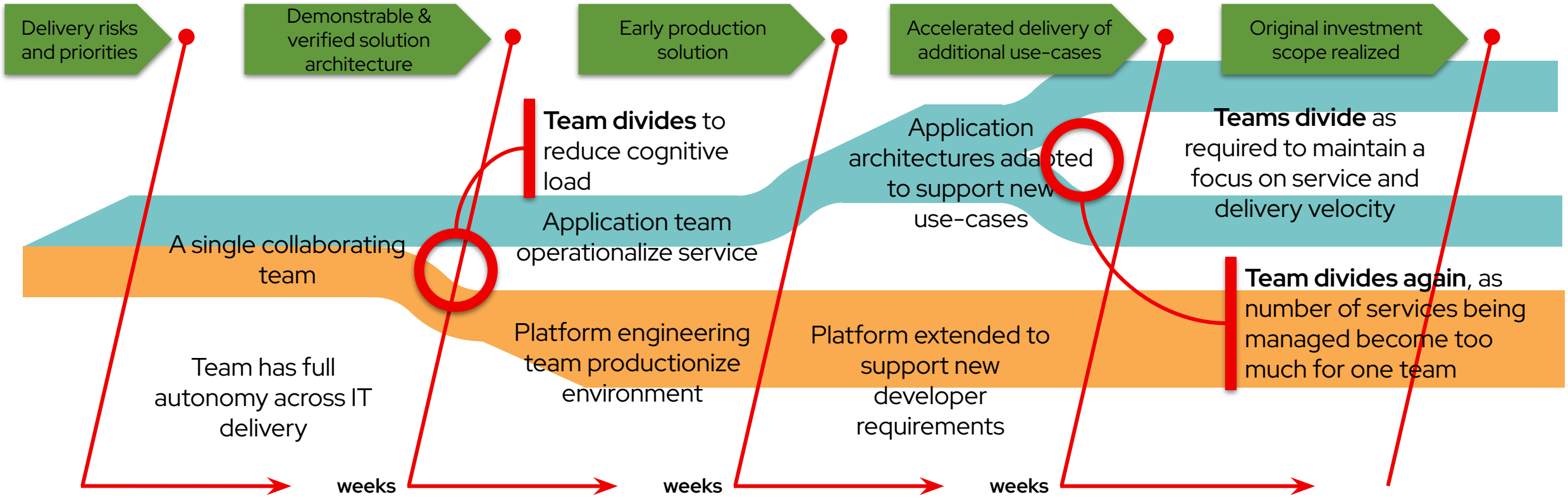
- Works on all supported versions of OpenShift
- Troubleshooting Mode for advanced diagnostics.
- MCP Client/Host Support for connecting your own MCP server using a Static API token, OpenShift user token, or Secret-based header
- Google Gemini and **Anthropic** models support via Google Vertex provider

Product Manager: Gaurav Singh

Red Hat

Red Hat recommends an evolutionary approach to organisational change

Organisational change is seeded through delivery of specific services, and designed to scale as required



Team Topologies: Organizing Business and Technology Teams for Fast Flow, Pias & Skelton
ISBN: 9781942788812

Red Hat's approached are informed by, and align with, Team Topologies

Red Hat

Key Security: AWS STS & IAM

0

Hardcoded IAM Access Keys

Enforcing Least Privilege with STS

ROSA uses the AWS Security Token Service (STS) to completely eliminate long-lived admin credentials and the risk of leaked access keys.

Workloads leverage temporary, auto-rotated tokens tied to OpenID Connect (OIDC) identities. Precision IAM policies (such as ROSAInstallerPolicy) strictly restrict control and data plane boundaries, drastically reducing the system's attack surface.

Accessing the Workshop

Who this Workshop is For.

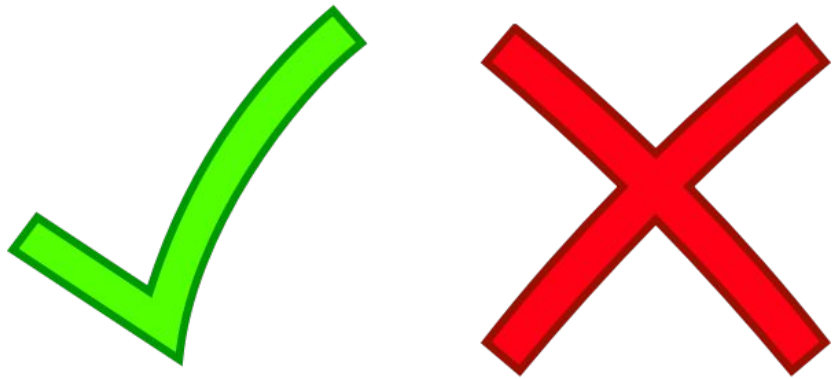
- Platform engineers looking to build an application platform.
- Developers looking to understand foundations of an application platform.
- DevOps looking to Ops with their Dev.

Knowledge Prerequisites

Skills/Knowledge required to be successful in this workshop.

- Basic understanding of OpenShift or Kubernetes concepts.
- Knowledge of running workloads in a Cloud Provider environment.
- Basic CLI/Linux experience.

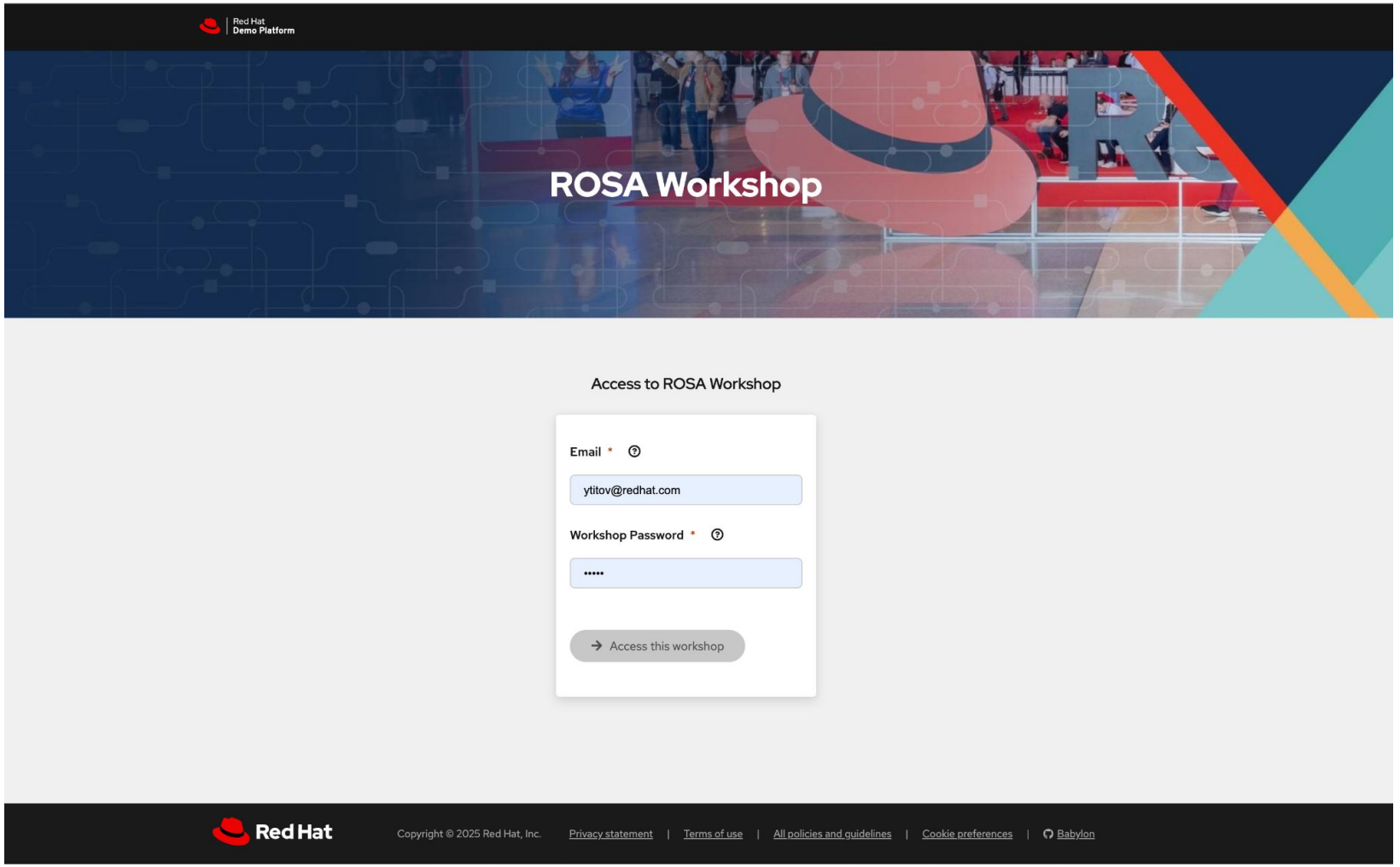
Workshop Guidelines



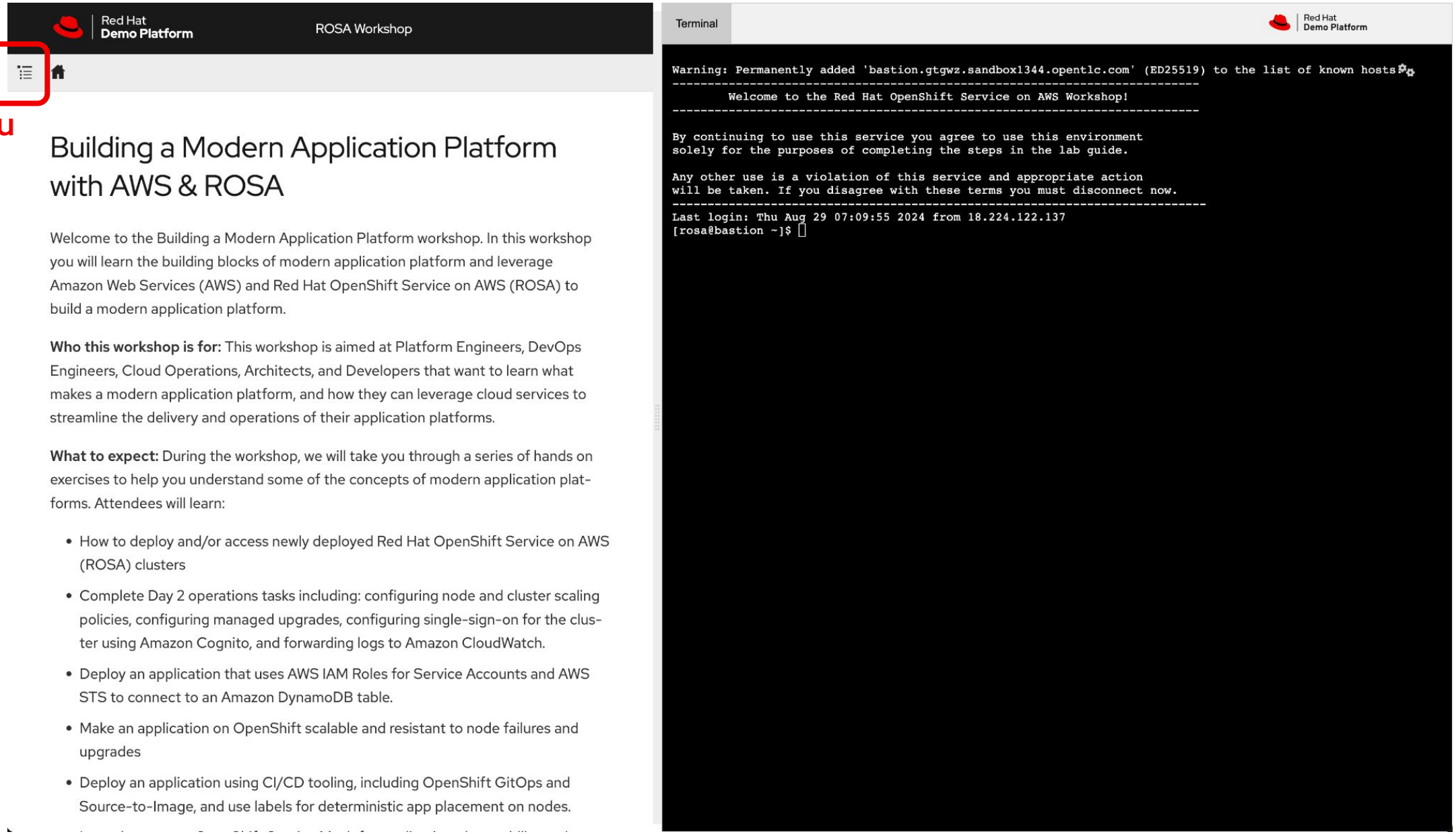
- Be respectful of facilitators, participants, and the compute environments provided.
- Raise your hand or find a facilitator if you need help, have a question, or get stuck.
- Let us know how we did, positive or constructive criticism is welcome!

Workshop url: https://red.ht/audi-rosa-2026

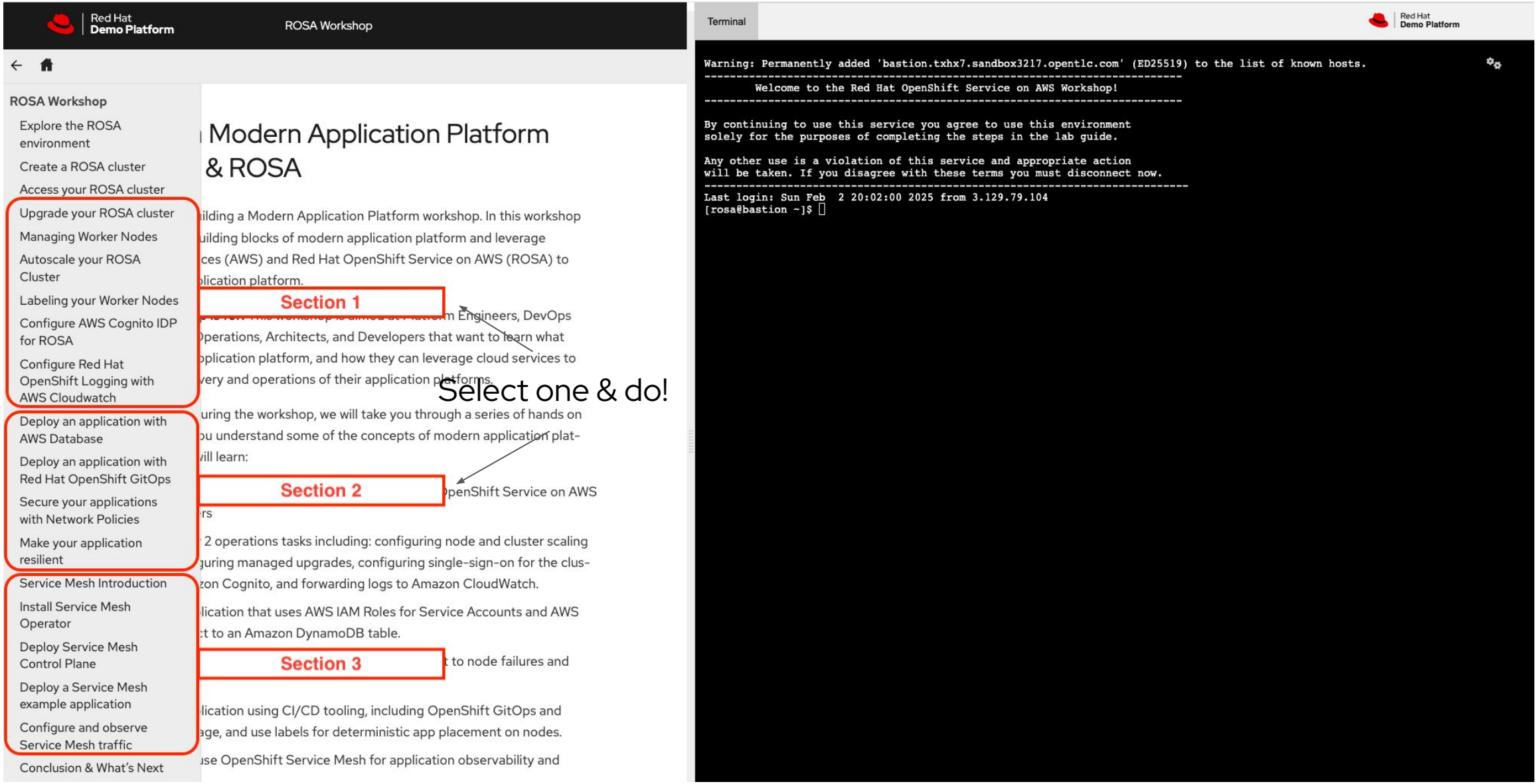
Password: **rosa**



Red Hat


Red Hat
Demo Platform
ROSA Workshop
Menu
Building a Modern Application Platform with AWS & ROSA
Welcome to the Building a Modern Application Platform workshop. In this workshop you will learn the building blocks of modern application platform and leverage Amazon Web Services (AWS) and Red Hat OpenShift Service on AWS (ROSA) to build a modern application platform.
Who this workshop is for: This workshop is aimed at Platform Engineers, DevOps Engineers, Cloud Operations, Architects, and Developers that want to learn what makes a modern application platform, and how they can leverage cloud services to streamline the delivery and operations of their application platforms.
What to expect: During the workshop, we will take you through a series of hands on exercises to help you understand some of the concepts of modern application platforms. Attendees will learn:
How to deploy and/or access newly deployed Red Hat OpenShift Service on AWS (ROSA) clusters
Complete Day 2 operations tasks including: configuring node and cluster scaling policies, configuring managed upgrades, configuring single-sign-on for the cluster using Amazon Cognito, and forwarding logs to Amazon CloudWatch.
Deploy an application that uses AWS IAM Roles for Service Accounts and AWS STS to connect to an Amazon DynamoDB table.
Make an application on OpenShift scalable and resistant to node failures and upgrades
Deploy an application using CI/CD tooling, including OpenShift GitOps and Source-to-Image, and use labels for deterministic app placement on nodes.
Terminal
Red Hat
Demo Platform
Warning: Permanently added 'bastion.gtgwz.sandbox1344.opentlc.com' (ED25519) to the list of known hosts
Welcome to the Red Hat OpenShift Service on AWS Workshop!
By continuing to use this service you agree to use this environment
solely for the purposes of completing the steps in the lab guide.
Any other use is a violation of this service and appropriate action
will be taken. If you disagree with these terms you must disconnect now.
Last login: Thu Aug 29 07:09:55 2024 from 18.224.122.137
[rosa@bastion ~]$

Red Hat

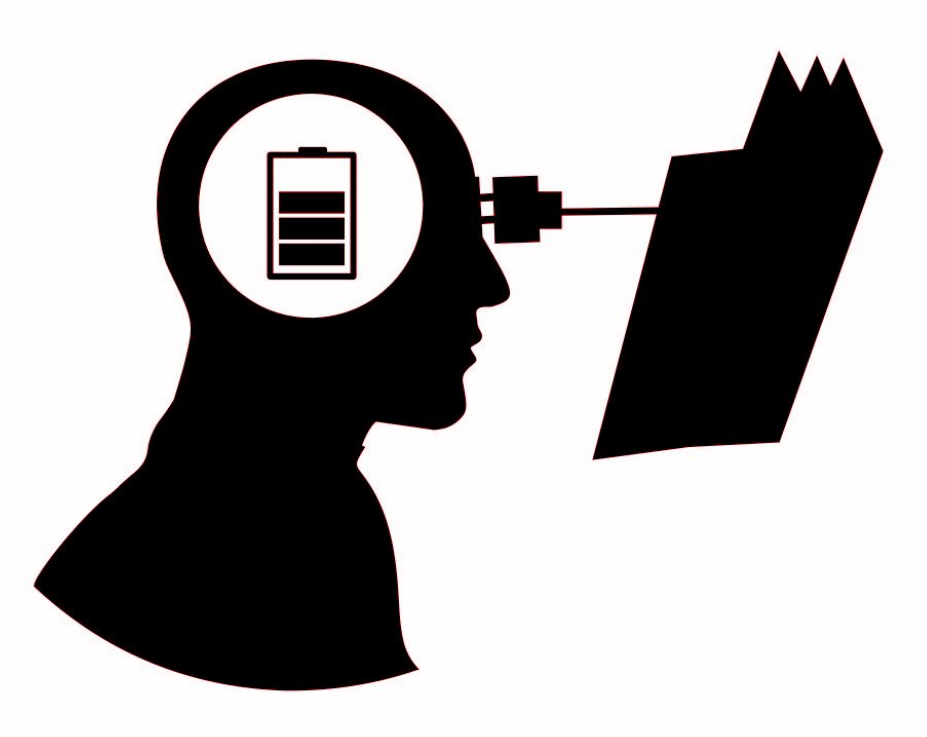

Red Hat
Demo Platform
ROSA Workshop
ROSA Workshop
Explore the ROSA environment
Create a ROSA cluster
Access your ROSA cluster
Upgrade your ROSA cluster
Managing Worker Nodes
Autoscale your ROSA Cluster
Labeling your Worker Nodes
Configure AWS Cognito IDP for ROSA
Configure Red Hat OpenShift Logging with AWS Cloudwatch
Deploy an application with AWS Database
Deploy an application with Red Hat OpenShift GitOps
Secure your applications with Network Policies
Make your application resilient
Service Mesh Introduction
Install Service Mesh Operator
Deploy Service Mesh Control Plane
Deploy a Service Mesh example application
Configure and observe Service Mesh traffic
Conclusion & What's Next
Modern Application Platform
& ROSA
ilding a Modern Application Platform workshop. In this workshop
uilding blocks of modern application platform and leverage
ces (AWS) and Red Hat OpenShift Service on AWS (ROSA) to
plication platform.
Section 1
Engineers, DevOps
Operations, Architects, and Developers that want to learn what
pplication platform, and how they can leverage cloud services to
very and operations of their application platforms.
Select one & do!
uring the workshop, we will take you through a series of hands on
ou understand some of the concepts of modern application plat-
ill learn:
Section 2
penShift Service on AWS
rs
2 operations tasks including: configuring node and cluster scaling
guring managed upgrades, configuring single-sign-on for the clus-
zon Cognito, and forwarding logs to Amazon CloudWatch.
lication that uses AWS IAM Roles for Service Accounts and AWS
t to an Amazon DynamoDB table.
Section 3
t to node failures and
lication using CI/CD tooling, including OpenShift GitOps and
age, and use labels for deterministic app placement on nodes.
se OpenShift Service Mesh for application observability and
Terminal
Red Hat
Demo Platform
Warning: Permanently added 'bastion.txhx7.sandbox3217.opentlc.com' (ED25519) to the list of known hosts.
Welcome to the Red Hat OpenShift Service on AWS Workshop!
By continuing to use this service you agree to use this environment
solely for the purposes of completing the steps in the lab guide.
Any other use is a violation of this service and appropriate action
will be taken. If you disagree with these terms you must disconnect now.
Last login: Sun Feb 2 20:02:00 2025 from 3.129.79.104
[rosa@bastion ~]$

Red Hat

Section 1:
Day Two Operations

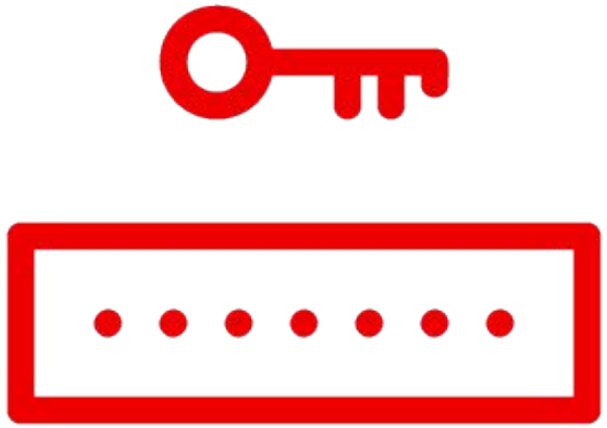
Day Two Operations

What you'll learn today.



- Integrating with Amazon Cognito for IDP
- Managing Cluster Upgrades
- Managing Worker Nodes
- Cluster Autoscaling
- Labeling Nodes
- Logging with AWS CloudWatch

Integrating with IDPs



- ROSA supports a number of Identity Providers:
  - GitHub, GitHub Enterprise, GitLab, Google, LDAP, OpenID Connect.
- In this workshop, we'll use Amazon Cognito via the OpenID Connect integration.

Learn more about Identity Providers here:
https://docs.openshift.com/rosa/rosa_install_access_delete_clusters/rosa-sts-config-identity-providers.html

Cluster Upgrades

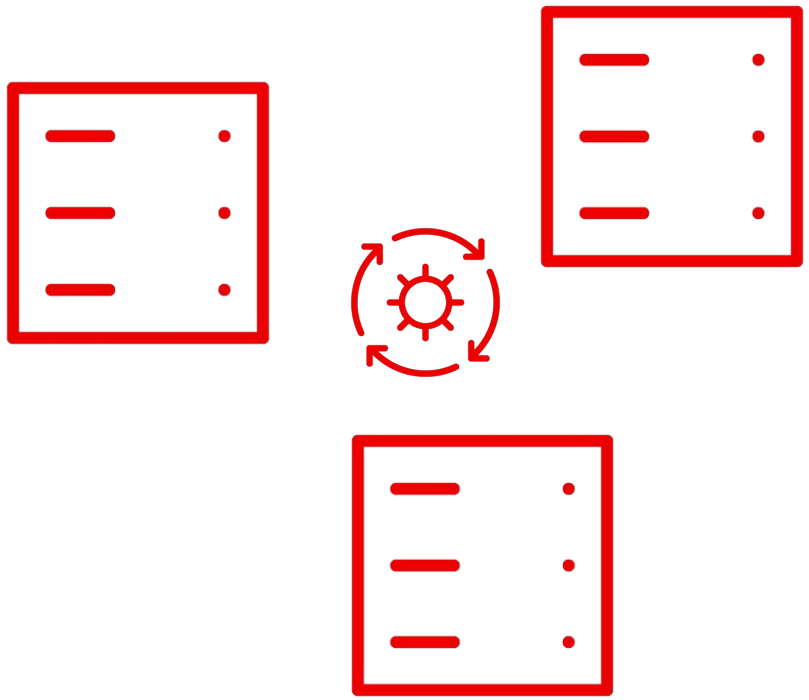
Major.Minor.Patch

- Cluster upgrades can be **manually initiated** or **automatically scheduled**.
- **Critical CVEs** are **automatically patched** within **48 hours** of a Patch release.
  - Impacted Patch releases are deprecated and not supported.
- **Minor versions** are supported for **14 months**.
- **Major versions** are supported for **12 months following** the release of a **new major version**.

Learn more about upgrades here:
https://docs.openshift.com/rosa/rosa_architecture/rosa_policy_service_definition/rosa-life-cycle.html#rosa-life-cycle

Managing Worker Nodes

Providing highly available compute.



- MachinePools allows for worker nodes that span multiple availability zones (AZs).
- MachinePools provide a declarative desired state for worker nodes to ensure consistency across AZs.
- MachinePools can be scaled up or down manually or automatically.

Learn more about Managing Worker Nodes here:
https://docs.openshift.com/rosa/rosa_cluster_admin/rosa_nodes/rosa-managing-worker-nodes.html

Machine Pools can span multiple availability zones.

Declarative desired state MachinePool

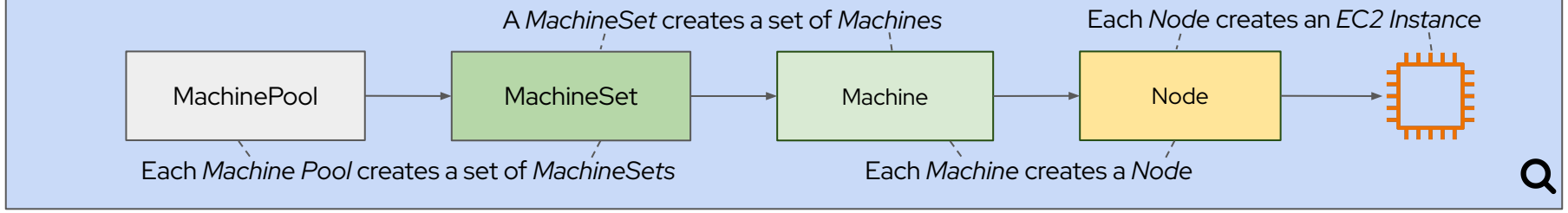
MachineSet → Machines → Nodes

MachineSet → Machines → Nodes

MachineSet → Machines → Nodes

Machine Sets are specific to a single availability zone, which is why there are 3 in this diagram.

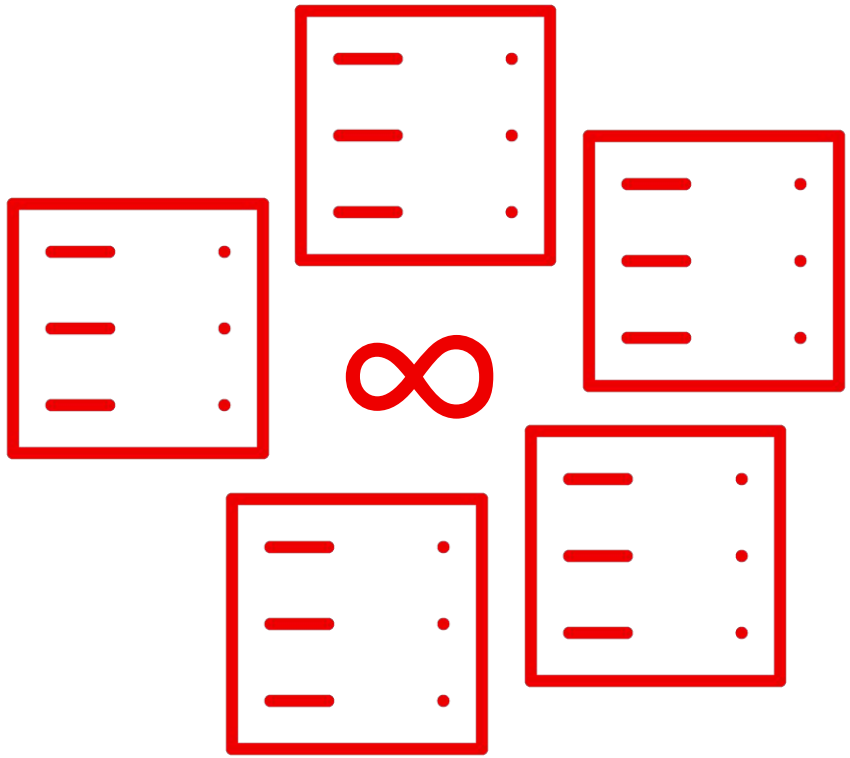
Machine Pools are managed by the *OpenShift Cluster Manager (OCM)*. The rest of the process is managed by the *Machine API Operator*. This operator interacts with the *AWS API* directly to provision *EC2 instances*.



Red Hat

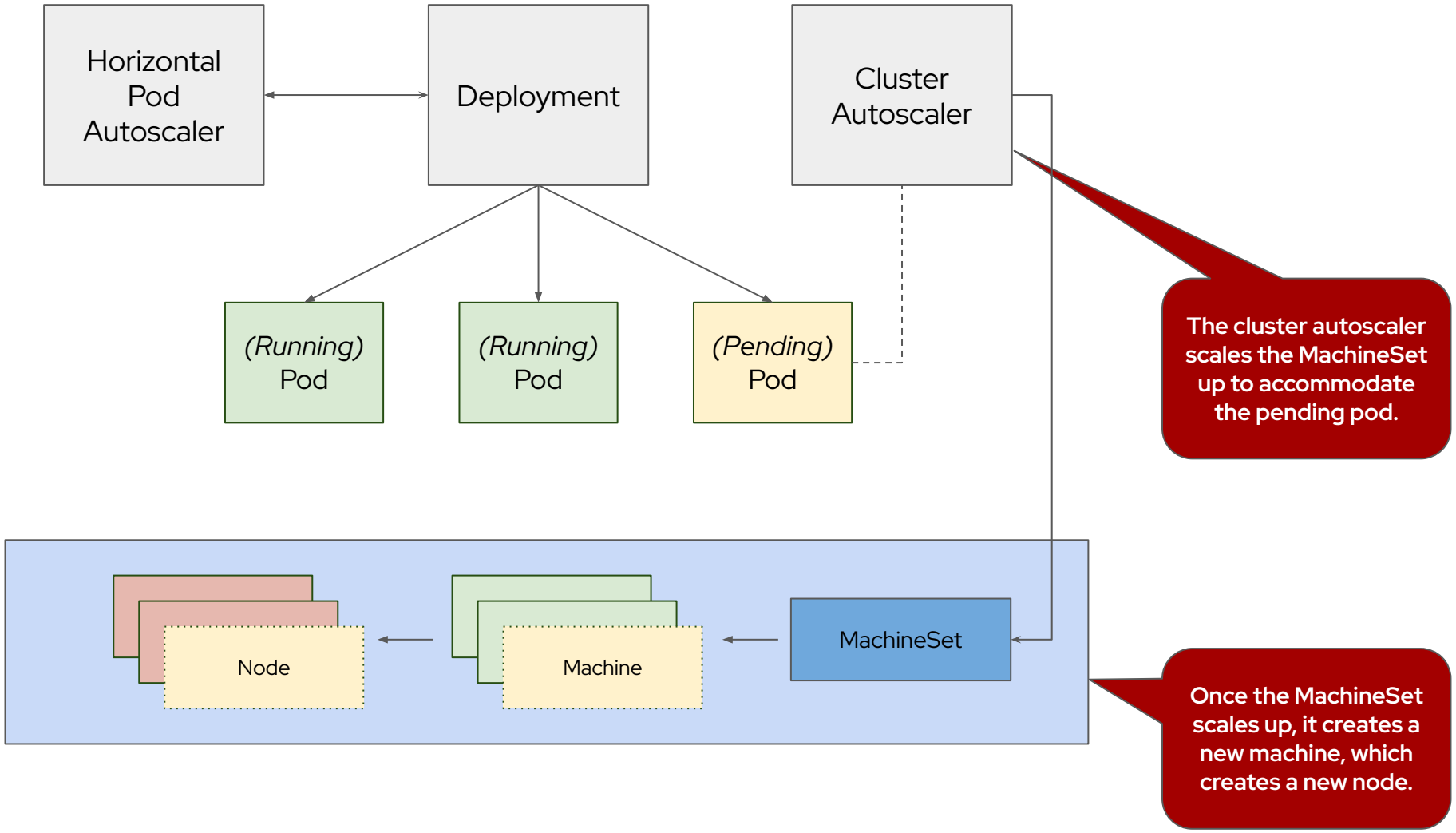
Cluster Autoscaling

Automatically responding to cluster demand.



- MachinePools can be scaled to meet applications demands.
- Cluster AutoScaler will provision additional worker nodes when pods can not be scheduled due to resource constraints.
- Cluster AutoScaler will not scale beyond predefined limits.

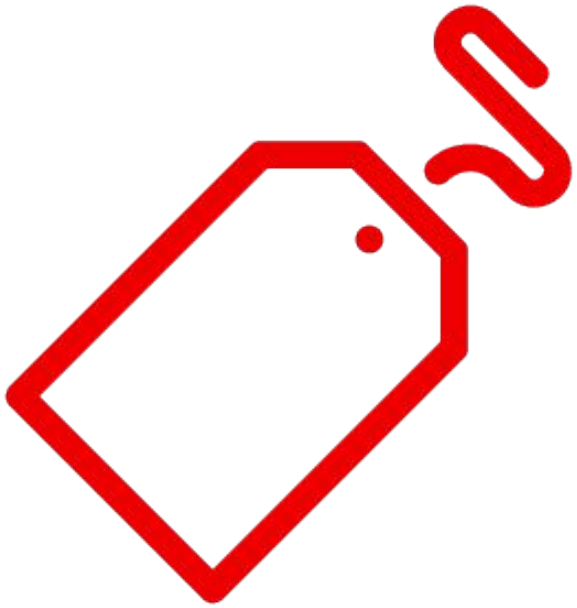
Learn more about Cluster Autoscaling here:
https://docs.openshift.com/rosa/rosa_cluster_admin/rosa_nodes/rosa-nodes-about-autoscaling-nodes.html


Horizontal Pod Autoscaler
Deployment
Cluster Autoscaler
(Running) Pod
(Running) Pod
(Pending) Pod
The cluster autoscaler scales the MachineSet up to accommodate the pending pod.
Node
Machine
MachineSet
Once the MachineSet scales up, it creates a new machine, which creates a new node.

Red Hat

Labeling Nodes

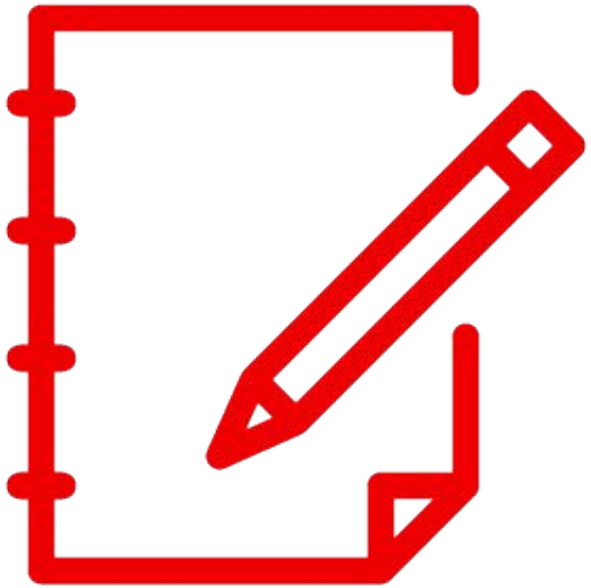
Deploy the right applications to the right compute resources.



- Labels allow application pods to automatically deploy to the correct compute resources.
- Examples include CPU or Memory intensive workloads, or workloads requiring GPU resources.

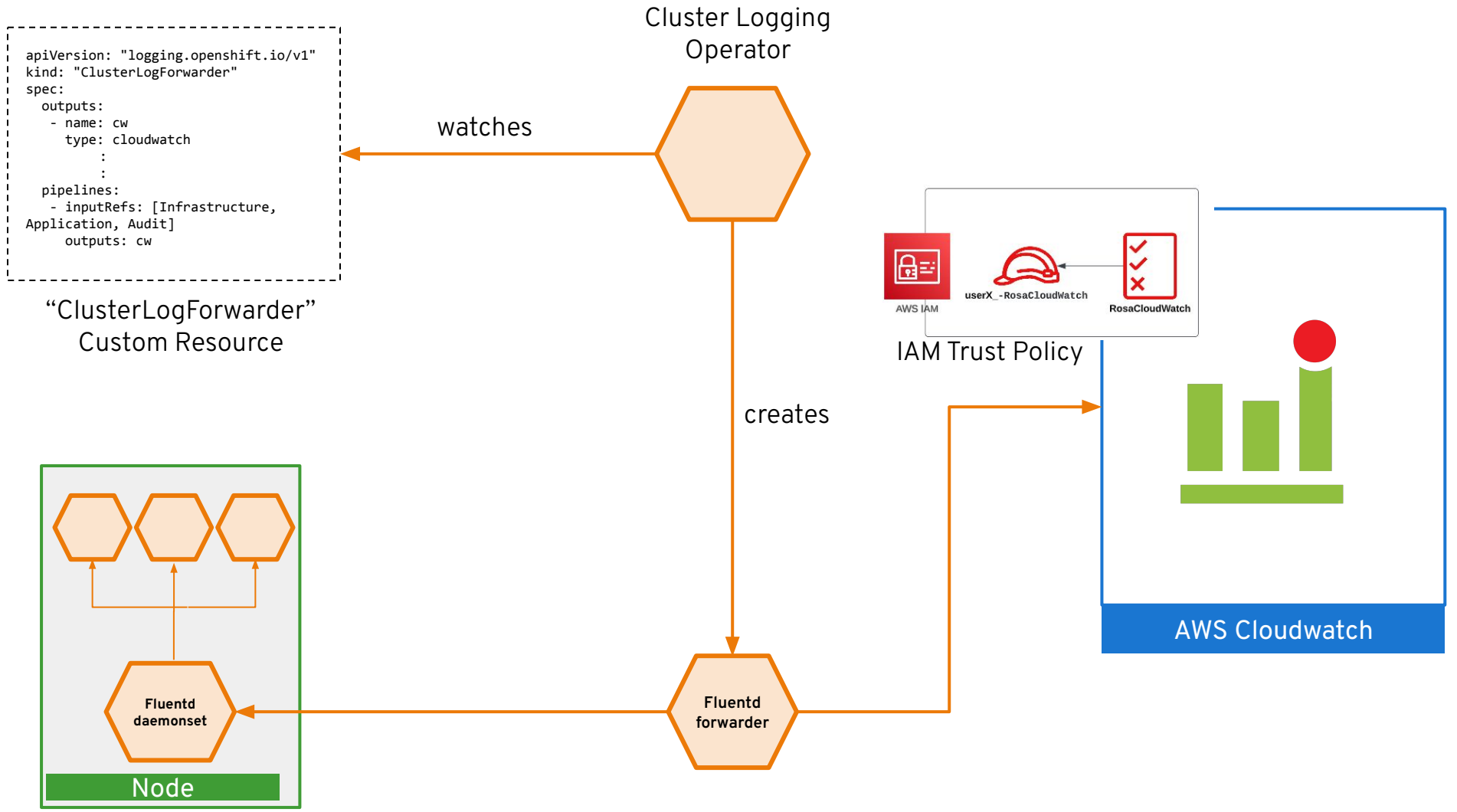
Logging with AWS Cloudwatch

Shipping logs to an enterprise-wide log management system.



- **OpenShift Cluster logs** are **stored in cluster** by **default**.
- **Cluster logs** can be **shipped** to a variety of log management systems such as **FluentD, ElasticSearch, Syslog, AWS CloudWatch, Loki, Kafka, and Splunk.**

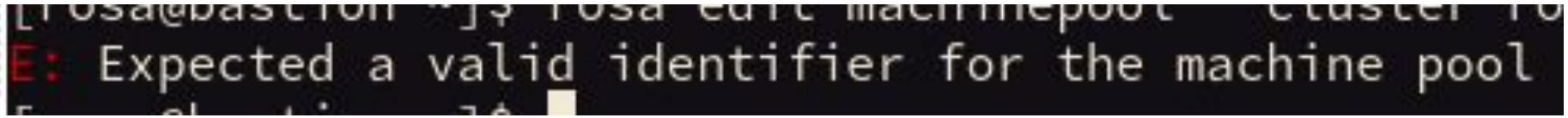
Secure Log Forwarding to Cloudwatch



Red Hat

In the section: "*Enable Autoscaling on the Default MachinePool*"

If you see this error message:



Change the machinepool id from "**Default**" to "**worker**"

Example:
rosa edit machinepool --cluster rosa-${GUID} **Default** --enable-autoscaling --min-replicas=2
--max-replicas=4

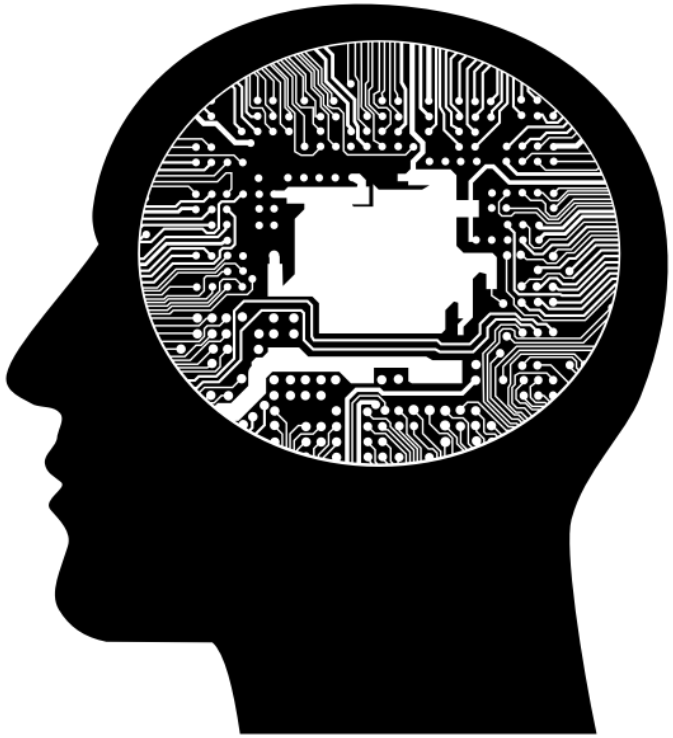
->

rosa edit machinepool --cluster rosa-${GUID} **worker** --enable-autoscaling --min-replicas=2
--max-replicas=4

Section 2: Deploy and Expose an Application

Deploy and Expose an Application

What you'll learn today.



- Deploying Applications
- Restricting Network Access
- Making Applications Resilient

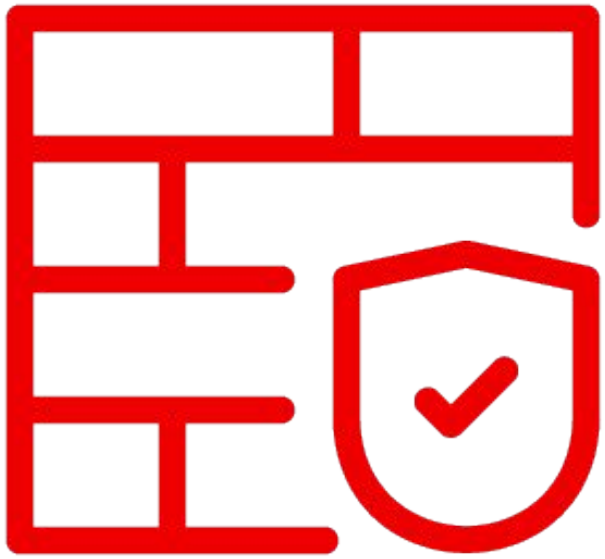
Deploying Applications

Deploy a Java based application using Quarkus and S2I.

- **Source-2-Image** (S2I) takes **application code** and **bundles it into a container** that can be **ran in OpenShift**.
- **Quarkus incorporates S2I** as part of it's build system, and can a**utomatically deploy** an application **to OpenShift** based on the application configuration.
- **Service Accounts** in OpenShift can **map to IAM roles** that **grant access to cloud resources** such as Amazon DynamoDB.

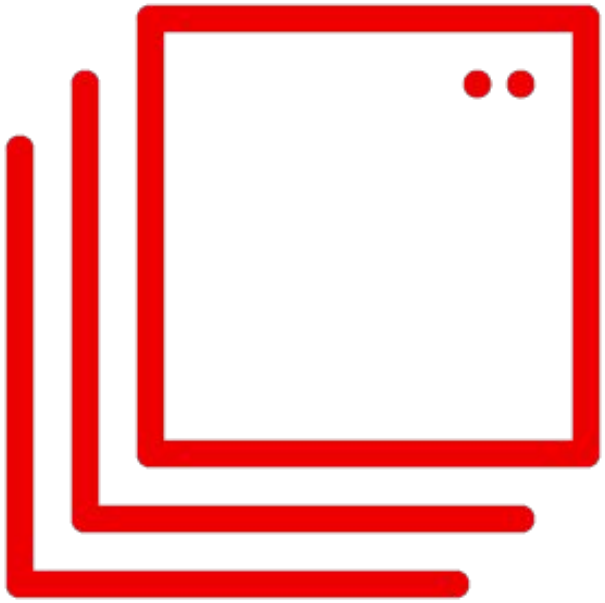
Restricting Network Access

Limit application access using NetworkPolicy.



- **NetworkPolicy allows for** applications to leverage the concepts of **Zero-Trust Networking: Deny by default, explicitly allow ingress/egress.**
- **NetworkPolicy** can **dynamically select** allowed or disallowed **clients** by leveraging **Pod or Namespace labels**.

Making Applications Resilient



- ROSA allows for applications to scale or recover from failure.
- **PodDisruptionBudgets define** the **minAvailable** and **maxUnavailable pods** for a given application (based on labels).
- **HorizontalPodAutoscaler** (HPA) allows for applications to **scale based on resource consumption** such as **CPU or RAM** utilization.

Using OpenShift GitOps

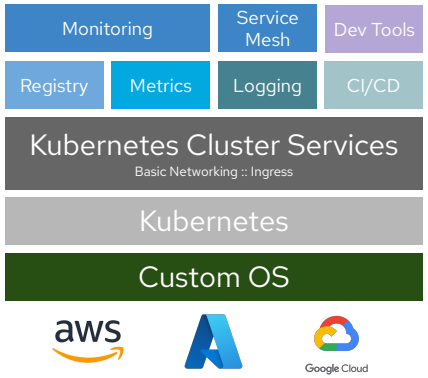
Consistent Code Across Environments

- **Treat everything as code**: Define the state of infrastructure, applications, and configurations with declarative code across environments
- **Single Source of Truth:** Infrastructure and applications are stored and versioned in Git allowing for traceability and visibility into changes that affect their entire state
- **Enhanced security:** Preview changes, detect configuration drifts, and take action
- **Visibility and audit:** Capture and trace any change to clusters through Git history
- **Multi-cluster consistency:** Combine GitOps with Advanced Cluster Manager for Kubernetes to configure multiple clusters and deployments reliably and consistently

Wrapping Up!

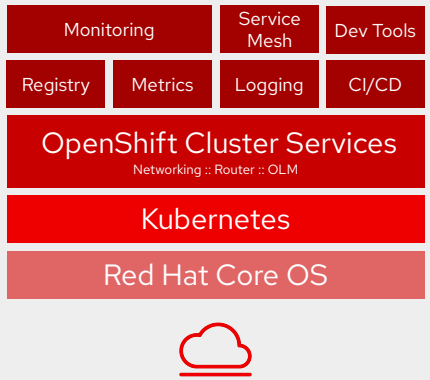
Build and run a platform *versus* using a turnkey cloud service

The Parts



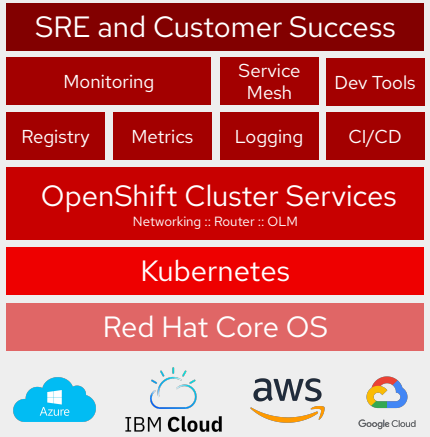
xKS + 'native' services

The Assembled Car



- Application Platform - Self-managed Red Hat OpenShift

The Car & Pit Crew



- Turnkey Application Platform - Red Hat OpenShift cloud services

"Batteries Included"

... but swappable

Individual components can be swapped out

Eg.

- Using AWS CloudWatch for logging on AWS
- Use specific cloud services or ISV offerings

Red Hat

Thank you!


linkedin.com/company/red-hat


youtube.com/user/RedHatVideos


facebook.com/redhatinc


twitter.com/RedHat

Red Hat

Helpful Links

ROSA Documentation

- https://docs.openshift.com/aro/4/welcome/index.html

MOBB.Ninja ROSA Guides

- https://mobb.ninja/#rosa

Introduction to ROSA - Red Hat Training

- https://www.redhat.com/en/services/training/DO120-introduction-to-red-hat-openshift-service-on-aws

ROSA Lightboard Videos

- https://www.redhat.com/en/about/videos/rosa-lightboard

ROSA User Guide - AWS

- https://docs.aws.amazon.com/ROSA/latest/userguide/what-is-rosa.html

Introduction to ROSA - Red Hat Ebook

- https://access.redhat.com/documentation/en-us/red_hat_openshift_service_on_aws/4/pdf/introduction_to_rosa/red_hat_openshift_service_on_aws-4-introduction_to_rosa-en-us.pdf