Canary Deployments

Document Level Classication

100

Introduction

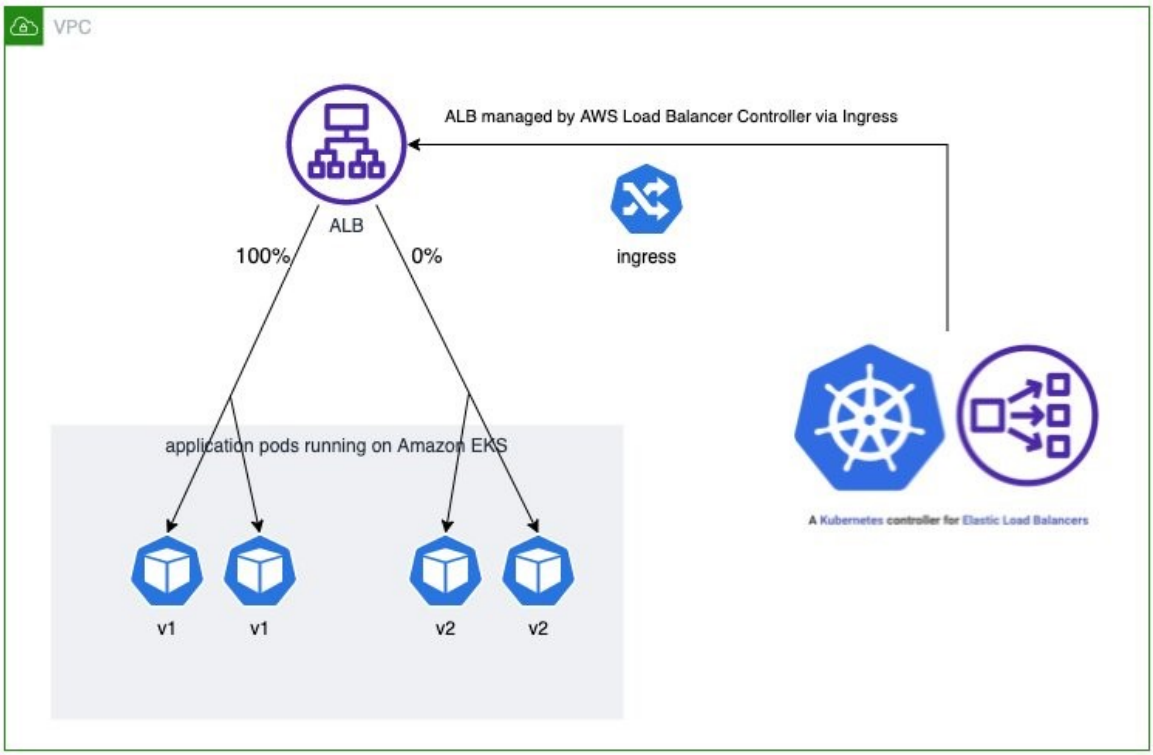
The following document describes the Canary deployment strategy at a high level. This deployment strategy falls under the Reliability Pillar of the AWS Well-Architected Framework. Specifically, REL08-BP04 Deploy using immutable infrastructure.

What is a Canary Deployment?

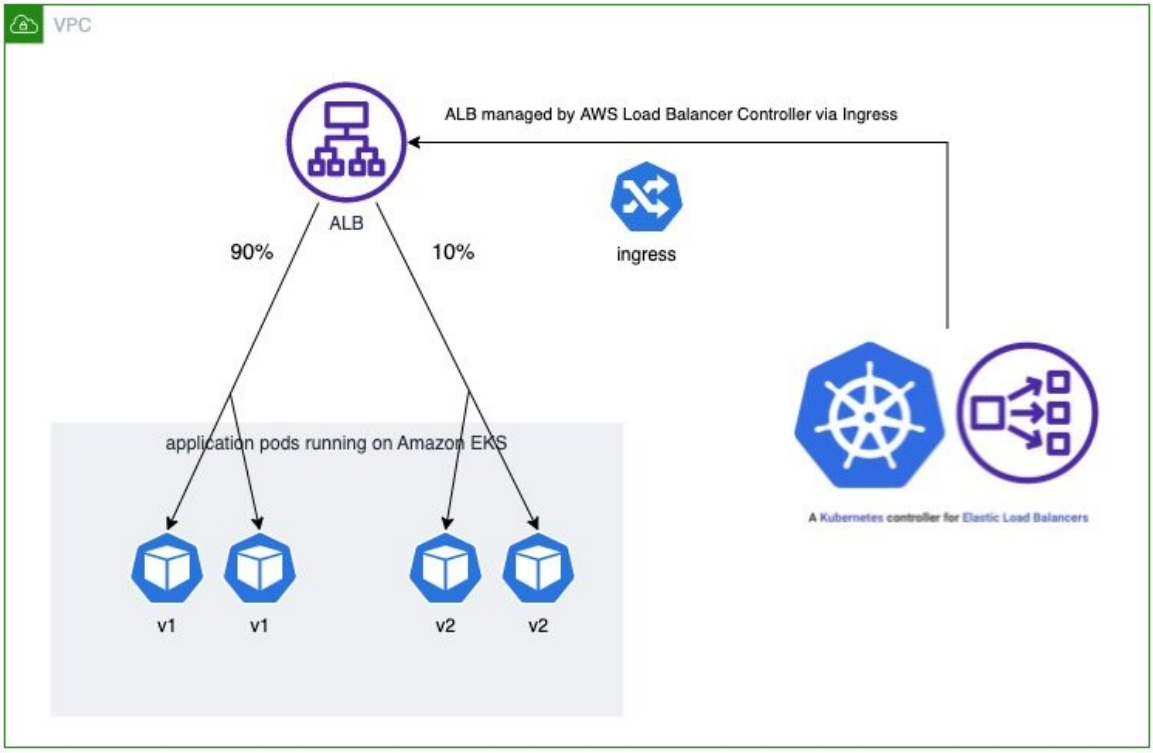
A canary deployment is a method of deploying software that exposes a new feature to an early sub-segment of users. The goal is to test new functionality on a subset of customers before releasing it to the entire user base.

Architecture

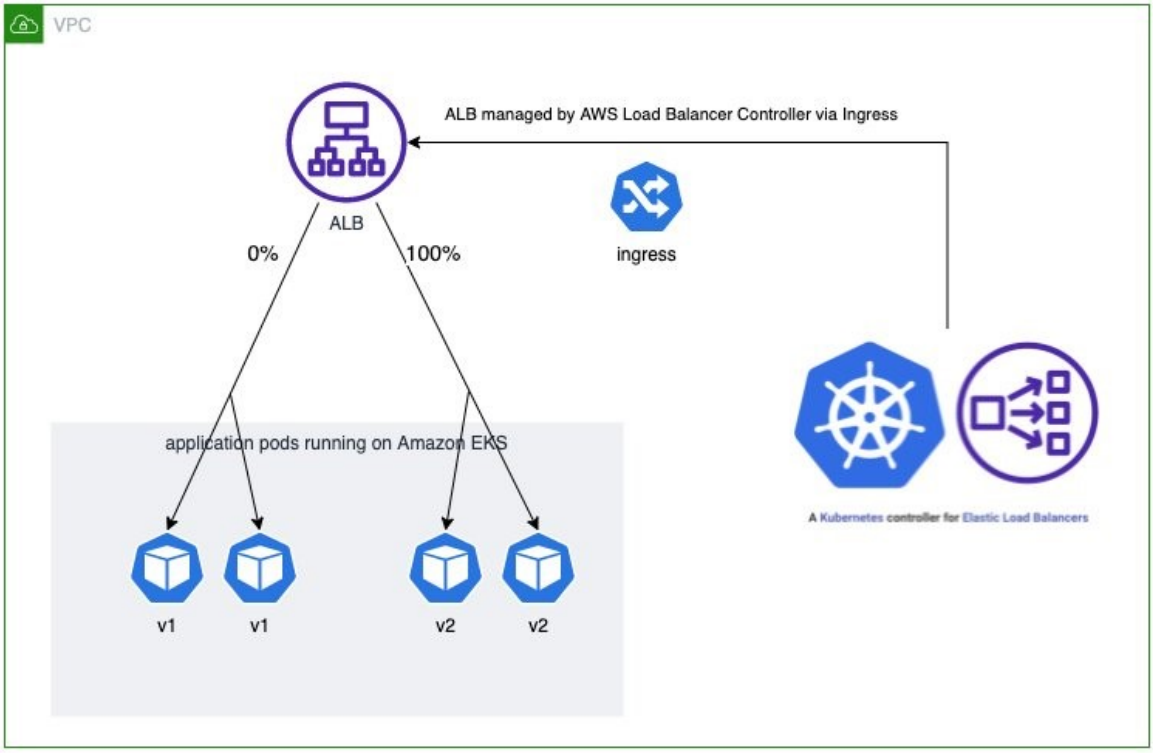
The following diagrams illustrate the Canary deployment strategy. In this example, v1 of the application is taking 100% of the traffic. Traffic is being controlled by the AWS Load Balancer Controller via Kubernetes Ingress.



After v2 of the application has been deployed along side of v2, 10% of the traffic is shifted over to v2 with v1 taking 90%.



If testing is successful, 100% of the traffic can be shifted over to v2 of the application.



Supporting Documentation

- To read more about the Canary deployment strategy using the AWS Load Balancer Controller reference the AWS Blog Post Using AWS Load Balancer Controller for blue/green deployment, canary deployment and A/B testing.
- Codefresh.io also has a in depth guide on Canary deployments in the article What is Canary Deployment?