

Optimizing Spot Instance Reliability and Security Using Cloud-Native Data and Tools

Muhammad Saqib*, Shubham Malhotra[†], Dipkumar Mehta[‡], Jagdish Jangid[§], Fnu Yashu[¶], and Sachin Dixit^{||}

^{*}Texas Tech University, Dept. of Computer Science saqibraopk@hotmail.com

[†]Rochester Institute of Technology, Dept. of Software Engineering
shubham.malhotra28@gmail.com

[‡]C.K.Pithawalla College of Engineering and Technology
dipkumar.mehta@gmail.com

[§]Infinera Corp

jangid.jagdish@gmail.com

[¶]Stony Brook University, Department of Computer Science
yyashu@cs.stonybrook.edu

Stripe Inc

spdixit@gmail.com

ARTICLE INFO

ABSTRACT

Received: 15 Nov 2024

Revised: 18 Jan 2025

Accepted: 26 Jan 2025

This paper presents "Cloudlab," a comprehensive, cloud-native laboratory designed to support network security research and training. Built on Google Cloud and adhering to GitOps methodologies, Cloudlab facilitates the creation, testing, and deployment of secure, containerized workloads using Kubernetes and serverless architectures. The lab integrates tools like Palo Alto Networks firewalls, Bridgecrew for "Security as Code," and automated GitHub workflows to establish a robust Continuous Integration/Continuous Machine Learning pipeline. By providing an adaptive and scalable environment, Cloudlab supports advanced security concepts such as role-based access control, Policy as Code, and container security. This initiative enables data scientists and engineers to explore cutting-edge practices in a dynamic cloud-native ecosystem, fostering innovation and improving operational resilience in modern IT infrastructures.

Keywords: Cloud-Native, Continuous Integration (CI), Continuous Machine Learning (CML), GitOps, Kubernetes Security, Policy as Code, Security as Code, Container Security, Serverless Computing, Infrastructure as Code (IaC), DevSecOps, Google Cloud (GCP), Container Orchestration, Observability and Monitoring.

I. Introduction

The ever-existing nature of cloud-native technologies has brought a shift in the way organizations build, deploy, and manage modern applications. As businesses switch to cloud environments—spanning public, private, and hybrid models—they gain access to advanced scalability and flexibility. However, this transition introduces a host of complexities, particularly in ensuring the reliability, security, and adaptability of cloud-based solutions. Organizations must steer their way through an ever-expanding ecosystem of tools, frameworks, and methodologies to address challenges such as securing containerized workloads, managing role-based access, implementing continuous integration, and adhering to Policy as Code principles. These challenges are further compounded by the dynamic nature of cloud-native environments, where traditional security paradigms often fall short.

To address these pressing concerns, this paper introduces Cloudlab, a dedicated cloud-native laboratory meticulously designed to support advanced security research, testing, and training. Built on Google Cloud's infrastructure and adhering to GitOps methodologies, Cloudlab serves as a comprehensive platform for exploring and validating cutting-edge security concepts. It incorporates advanced technologies such as Kubernetes, serverless architectures, and automated workflows to create a scalable and adaptive environment. The lab integrates tools like Palo Alto Networks CN-Series firewalls and Bridgecrew's "Security as Code" framework, enabling

researchers and engineers to adopt best practices in cloud security.

The foundation of Cloudlab lies in its dual pipelines: Continuous Integration (CI) and Continuous Machine Learning (CML). The CI pipeline is responsible for generating Docker images, managing version control, and automating the testing and deployment of secure workloads. On the other hand, the CML pipeline facilitates experimentation with machine learning models and their secure deployment. Together, these pipelines foster an ecosystem of automation, innovation, and operational resilience, aligning with the needs of modern IT infrastructures.

Through Cloudlab, this paper seeks to explore a range of topics critical to cloud-native environments. These include provisioning secure infrastructure, implementing serverless cloud functions, automating security workflows, and ensuring the security of containerized applications. The lab's design not only promotes the adoption of advanced security practices but also bridges the gap between theoretical research and real-world application.

By delving into the nuances of API endpoints, container security, role-based access control, and Policy as Code, Cloudlab provides a hands-on platform for engineers and security practitioners to enhance their expertise. Furthermore, the lab underscores the importance of automating security processes, leveraging tools like Terratest and Kyverno to validate and enforce security policies. These capabilities demonstrate the potential of cloud-native approaches to revolutionize the field of network security, offering robust solutions to the challenges posed by modern cloud ecosystems.

This paper aims to provide a detailed account of Cloudlab's architecture, capabilities, and applications, showcasing how it addresses key challenges in cloud security and reliability. By presenting a scalable, adaptable, and secure laboratory environment, this study highlights the transformative potential of cloud-native tools and methodologies. Ultimately, Cloudlab serves as a testament to the critical role of innovation in advancing security practices and supporting the next generation of cloud-native technologies.[1].

The Cloudlab is a cutting-edge, private laboratory environment meticulously designed to advance security research, testing, and training in cloud-native ecosystems. It represents a paradigm shift in how modern security practices are approached, leveraging the principles of cloud-native technologies to ensure scalability, adaptability, and operational efficiency. What sets Cloudlab apart is its adherence to GitOps practices, a revolutionary methodology that redefines infrastructure management. By treating infrastructure as code (IaC), GitOps facilitates the seamless storage, review, and maintenance of complex cloud configurations, ensuring a robust foundation for experimentation and deployment.

At its core, Cloudlab is engineered to support researchers, engineers, and practitioners in tackling the multifaceted challenges of securing modern IT infrastructures. It does so through two primary, purpose-driven pipelines. The first pipeline is dedicated to Continuous Integration (CI), streamlining the development process by automating the testing, building, and deployment of containerized workloads. This pipeline ensures that new code integrations are rigorously tested and securely deployed, fostering a culture of rapid innovation without compromising stability.

The second pipeline, Continuous Machine Learning, or CML[2], is equally transformative. As artificial intelligence (AI) and machine learning (ML) become integral to cloud-native systems, the CML pipeline facilitates the development, testing, and secure deployment of ML models. By automating the end-to-end lifecycle of machine learning workflows, the CML pipeline enables researchers to iterate rapidly, test at scale, and deploy models with a high degree of reliability. This capability is particularly valuable for exploring novel AI-driven security solutions, enhancing the lab's capacity for cutting-edge research.

Together, these pipelines form the backbone of Cloudlab, creating a dynamic and adaptive environment capable of addressing the demands of modern security challenges. They empower users to experiment with advanced concepts like role-based access control (RBAC), Policy as Code, and container security while integrating state-of-the-art tools and methodologies. Cloudlab's adherence to GitOps practices ensures that every change is version-controlled and reproducible, providing a transparent and collaborative platform for innovation.

This holistic approach makes Cloudlab an invaluable resource for security practitioners aiming to stay ahead in the rapidly evolving world of cloud-native technologies. By combining rigorous methodologies with advanced tools, the lab not only supports groundbreaking research but also bridges the gap between theoretical knowledge and

practical application. In doing so, it paves the way for a more secure and resilient future in cloud computing.

As security practitioners, it is important for us to understand the infrastructure and applications being built in the public clouds as a first step to making things more secure.

II. Project Goals

There are several goals associated with the development and operation of this project, each of which plays a pivotal role in achieving a comprehensive understanding of cloud-native environments. These goals are interconnected, emphasizing the integration of advanced tools, methodologies, and security concepts. The overarching objective is to explore, evaluate, and enhance the composition, management, and resilience of various components, as outlined below:

- 1) Provisioning Palo Alto Networks CN-Series firewall products, integrating them with Calico and protecting containerized workloads (Kubernetes “pods”).
- 2) Researching containerized deployments, workloads, and security.
- 3) Integration of Bridgecrew “Security as Code” tooling with GitHub repositories.
- 4) Developing and demonstrating expertise in “serverless cloud functions” (GCP Cloud Functions in this case).
- 5) Developing and utilizing a cloud-native Continuous Integration build pipeline. The output of this pipeline is a Docker image that is stored in gcr.io. These images include a fully contained set of tools, documentation, and Terraform code for customer deployments.
- 6) Demonstrating Policy as Code concepts using Terratest and Kyverno.

III. Learning Objectives

The primary aim of building and operating this lab is to foster learning and experimentation in key areas of cloud-native security. Engineers and researchers can benefit from engaging with various topics that are critical to modern cloud operations. The following list highlights the main learning objectives:

- 1) API Endpoints.
- 2) Deployment and operation of CN-Series firewalls.
- 3) Kubernetes role-based access control (RBAC) and security.
- 4) Developing serverless functions in Python.
- 5) ML/AI pipelines, containerized workloads, and their security.
- 6) Containers and container security.
- 7) Automation.
- 8) Testing perspectives, including Policy as Code and Security as Code.
- 9) CI/CD Pipelines and security.

Training materials derived from the lab’s construction and operation are readily available, enabling engineers to build proficiency in these areas and apply their knowledge to real-world scenarios.

A. API Endpoints

API endpoints are foundational to the interaction of applications across environments, including serverless systems and Kubernetes clusters. Their construction and operation provide an ideal space for exploring potential vulnerabilities and security challenges within cloud-native ecosystems. Understanding and securing API endpoints is critical for maintaining the reliability and integrity of applications, as they often serve as gateways for data exchange and system integration.

B. Containers and Container Security

Containerized environments, while offering immense flexibility and scalability, introduce unique security

challenges. Addressing these challenges involves understanding container architecture, identifying vulnerabilities, and implementing best practices for container security. This includes securing Kubernetes pods, employing role-based access controls, and integrating automated tools to monitor and protect containerized workloads.

C. Automation

Automation is a cornerstone of cloud-native operations, enabling organizations to scale efficiently and maintain operational consistency. By automating processes such as infrastructure provisioning, security scans, and policy enforcement, organizations can significantly reduce manual errors and improve overall system reliability. The lab emphasizes the use of tools and methodologies to automate complex workflows, ensuring that security and operational excellence remain integral to cloud-native environments.

IV. Lab Configuration Details

The configuration and codebase for the lab are meticulously maintained on GitHub, ensuring transparency, collaboration, and version control. By leveraging GitHub repositories as a centralized storage system, the lab facilitates efficient infrastructure management, version tracking, and seamless collaboration between engineers, researchers, and developers. Hosted on Google Cloud—one of the “big three” public cloud providers—the lab serves a dual purpose: acting as a proof of concept for innovative cloud-native solutions and as a training platform to enhance hands-on expertise in modern security and automation practices.

The lab’s architecture is composed of several core components that work cohesively to create a secure, scalable, and adaptive environment. Each of these components is described below:

- 1) **GitHub Repositories:** The GitHub repositories are the foundation of the lab’s infrastructure as code (IaC) approach. They store all configuration files, pipeline scripts, and containerized application code, ensuring a single source of truth. By maintaining these files in GitHub, the lab enables automated workflows such as pull request reviews, code scanning, and testing, which are crucial for ensuring the reliability and integrity of the infrastructure. These repositories also facilitate collaboration by allowing multiple contributors to work simultaneously, with built-in features for managing changes and resolving conflicts.
- 2) **Google Kubernetes Engine (GKE):** GKE provides the backbone for the lab’s computational and orchestration needs. It hosts the Kubernetes clusters that underpin the Continuous Integration (CI) pipeline and enable machine learning experimentation. The GKE clusters are designed to support containerized workloads efficiently, ensuring seamless scalability and resource optimization. By integrating role-based access control (RBAC) and network policies, the GKE environment ensures that workloads are both secure and compliant with best practices. This infrastructure allows for real-time testing and deployment of workloads while maintaining high availability and fault tolerance.
- 3) **Private Container Repository (GCR):** The Google Container Registry (GCR) serves as a secure storage location for Docker images generated by the CI pipeline. These images encapsulate the tools, libraries, and application code required for deployment, ensuring consistency across environments. By maintaining a private repository, the lab guarantees that sensitive configurations and dependencies are securely stored, reducing the risk of unauthorized access. Additionally, GCR supports automated vulnerability scanning, allowing the lab to identify and address potential security risks in container images.
- 4) **Automation Frameworks:** Automation is a cornerstone of the lab’s design, with frameworks such as GitHub Actions and webhooks playing a critical role in streamlining CI/CD workflows. These automation tools trigger events such as testing, building, and deployment whenever changes are made to the codebase. Webhooks enable seamless integration between GitHub repositories and other services, such as the CI pipeline hosted on GKE. The automation frameworks also incorporate security measures like “Security as Code” tooling, ensuring that every build and deployment adheres to predefined security standards.

The lab’s design underscores the importance of integrating automation, security, and scalability into cloud-native environments. By harmonizing these elements, the lab creates a robust platform for exploring advanced cloud concepts and gaining practical expertise. It empowers researchers and engineers to test innovative solutions in a controlled environment, bridging the gap between theoretical knowledge and real-world application.

A detailed description of the five main blocks seen in Figure 1 is provided below:

- 1) **CI/CD Pipelines:** The lab incorporates a robust CI/CD pipeline hosted on the GKE cluster. This pipeline automates the process of building, testing, and deploying applications. Each code commit triggers a series of automated tasks, including static code analysis, container image creation, and security checks. The pipeline ensures that only tested and verified code is promoted to production environments, significantly reducing deployment risks.

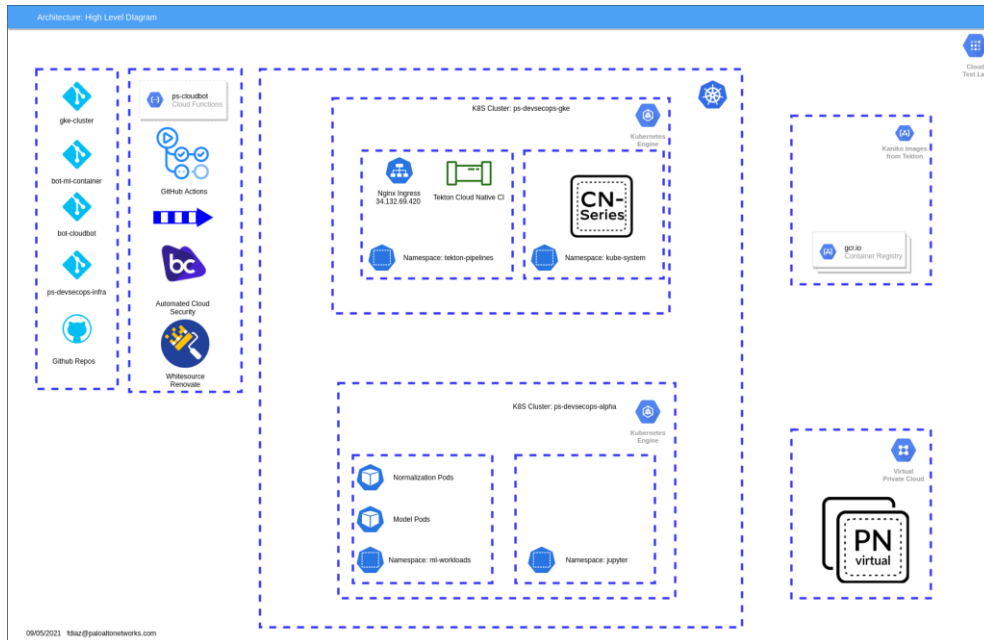


Figure 1. High-Level Lab Design Diagram

- 2) **Kubernetes Clusters:** Two Kubernetes clusters form the core of the lab's infrastructure. The primary cluster hosts the CI pipeline and containerized applications, while the secondary cluster is dedicated to machine learning experiments. These clusters are designed to be flexible and scalable, enabling rapid prototyping and testing of new workloads.
- 3) **Security Tools and Policies:** Security is integral to the lab's architecture. Tools such as Bridgecrew and Kyverno enforce "Security as Code" and "Policy as Code" practices, ensuring that all configurations and deployments adhere to industry standards. Automated security scans are performed on code, configurations, and container images, reducing vulnerabilities across the development lifecycle.
- 4) **Monitoring and Observability:** The lab integrates monitoring and observability tools to provide real-time insights into system performance and security. Metrics and logs are collected from Kubernetes clusters, CI pipelines, and containerized applications, allowing engineers to identify and resolve issues proactively.
- 5) **Training and Experimentation:** As a training platform, the lab is designed to simulate real-world scenarios, providing engineers with hands-on experience in managing cloud-native infrastructures. The inclusion of advanced tools and workflows ensures that users can experiment with cutting-edge technologies while adhering to best practices.

The Cloudlab serves as a versatile platform that aligns theoretical research with practical applications. By integrating the components described above, the lab not only enhances operational resilience but also fosters innovation, making it a critical resource for advancing cloud-native security and scalability.

1 [GitHub Repositories]

The code base for the lab is broken down into several GitHub repositories, more or less around the functional area.

Repo Name	Purpose
bot-cloudbot	A custom Python 3.9 GCP Cloud Function for GitHub pull request task automation.
bot-ml-container	An experimental containerized machine learning model.
gke-cluster	Infra as Code files for GKE cluster, YAML files for Tekton CI pipeline. CN Series firewall nodes.
ps-containerizer	An “invisible shim” with a Docker image for each “public cloud” VM-Series Terraform module development repo. Allows PRs to be ingested into Tekton CI pipeline without any integrations with the source repository.
ps-devsecops-alpha	IaC files for Alpha K8s cluster.
python-project-template	Python project template for writing serverless code in AWS Lambda and GCP cloud functions.

Table I

PROJECT CODEBASE - GITHUB REPOSITORIES

2 [GitHub Actions]

GitHub repositories that have been “on-boarded” to the project have certain “actions” included.

- 1) Bridgecrew is used to scan all commits to all open pull requests.
- 2) Whitesource Renovate is used to track keep project dependencies up to date and secure.
- 3) Another GitHub action defines the parameters of the GCP project and helps the “ps-cloudbot” with pull request maintenance tasks.

Note that there is also a webhook to make the CI pipeline aware of each commit and kick off a test and build cycle.

3 [Kubernetes Clusters] Two clusters are deployed with the Google Kubernetes Engine

(GKE). The “gke” cluster hosts the Tekton CI pipeline. The “alpha” cluster is used for machine learning experimentation.

Pipeline runs can be viewed and managed through a graphical interface that is well suited for development teams. There is also the ability to manage pipeline runs and their requisite tasks using standard command line tooling.

Example pipeline command

```
franklin /gke: tkn pr ls NAME STARTED DURATION STATUS gh-pr-run-ncwrr 1 hour ago 1 minute
Succeeded gh-pr-run-vhrvs 2 hours ago 1 minute Succeeded gh-pr-run-q6szq 3 hours ago 1
minute Succeeded gh-pr-run-8vn22 6 hours ago 1 minute Succeeded gh-pr-run-h7twc 14 hours
ago 17 seconds Failed gh-pr-run-74wm7 21 hours ago 1 minute Succeeded gh-pr-run-tqlfg 1 day
ago 1 minute Succeeded gh-pr-run-xs52f 1 day ago 1 minute Succeeded gh-pr-run-xtdz6 1 day ago
1 minute Succeeded gh-pr-run-w2zn7 1 day ago 1 minute Succeeded
```

4 [GCR Private Container Repository]

To date there are about 15 GitHub repositories that are integrated with the CI pipeline outlined in this paper. Each commit to a pull request causes the CI pipeline to generate a Docker image. These Docker images are stored in a private GCR location.

5 [Panorama Management]

Panorama is a key component in deployment and maintenance of Kubernetes clusters and CN Series firewalls. Currently there are two virtual Panorama devices deployed in a high availability (HA) configuration.

Palo Alto Networks has historically maintained serverless functions, for example in AWS Lambda, for firewall “auto-scaling” tasks. This has since been replaced by a set of official Panorama “plug-ins” that can be downloaded to PanOS devices. Lambda and Cloud functions are still frequently used to augment the capabilities of this new family of plug-ins.

V. Serverless Functions

Google Cloud offers Cloud Functions, among other serverless computing offerings. This is a good set of patterns to learn and comes up often in security infrastructure work since organizations can use it to run jobs at a lower cost. Securing the webhooks and connection between cloud functions and the VPC the GKE cluster is in is an important factor in deployments. Making XML or other sorts of API calls is often used to pass data between management platforms, ticketing systems, custom applications, and so on.

A. Github Webhook

The Github webhook can be combined with a set of GCP credentials stored as a “repository secret” in a repository. Available is a powerful and flexible method to combine systems into a more intricate Continuous Integration and testing pipeline.

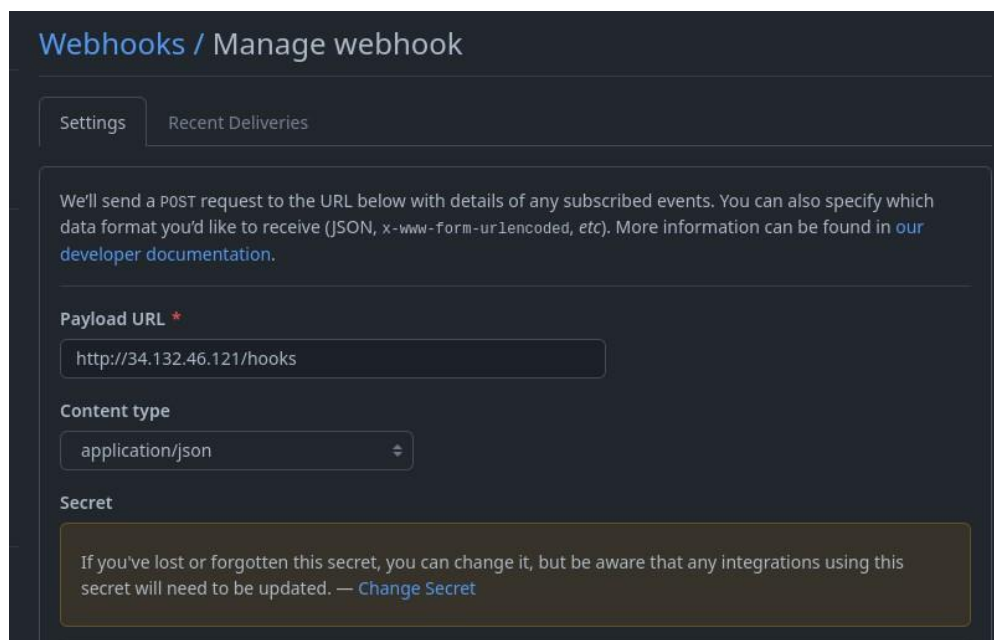
The image shows the 'Webhooks / Manage webhook' interface in a dark-themed Github repository. At the top, there are two tabs: 'Settings' (selected) and 'Recent Deliveries'. Below the tabs, a text block explains that a POST request will be sent to the specified URL with details of subscribed events, and it links to the developer documentation. The 'Payload URL' field is a text input containing 'http://34.132.46.121/hooks'. Below this, the 'Content type' is set to 'application/json' in a dropdown menu. At the bottom, there is a 'Secret' section with a warning message: 'If you've lost or forgotten this secret, you can change it, but be aware that any integrations using this secret will need to be updated.' followed by a 'Change Secret' link.

Figure 2. Webhook configuration settings in Github

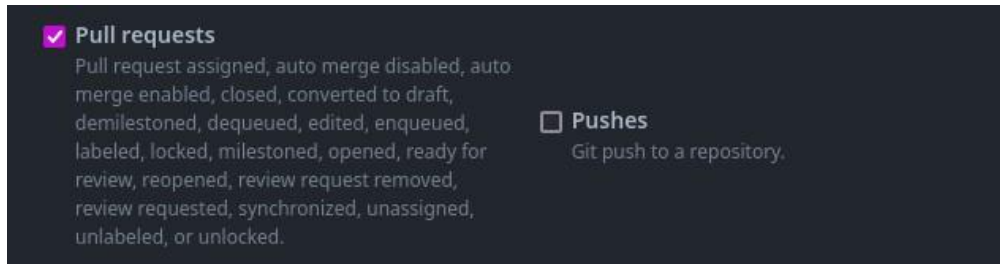


Figure 3. Check only this box in the webhook settings

B. Connecting Serverless to GKE

A network peering is created between the Cloud Function and the VPC that the GKE cluster resides in. This allows the serverless function application to make calls to a containerized deployment, such as a Node application.

Example YAML for a cloudbot service

```
apiVersion: v1 kind: Service metadata: name: cloudbot-service namespace: ci-build annotations:
cloud.google.com/load-balancer-type: "Internal" labels: app: cloudbot-service spec: type:
LoadBalancer selector: app: cloudbot ports: - port: 80 targetPort: 8089 protocol: TCP
```

Example deployment YAML for a cloudbot

```
apiVersion: apps/v1 kind: Deployment metadata: creationTimestamp: null labels: app: cloudbot
name: cloudbot-deployment namespace: ci-build spec: replicas: 3 selector: matchLabels: app:
cloudbot strategy: template: metadata: creationTimestamp: null labels: app: cloudbot spec:
nodeSelector: env: build containers: - name: cloudbot image: gcr.io/gcp-gcs-pso/build-pod
imagePullPolicy: Always volumeMounts:
- name: nfs-volume-1 mountPath: "/data" volumes: - name: nfs-volume-1 persistentVolumeClaim:
claimName: nfs-pvc
```

Example Dockerfile for a cloudbot pod in GKE

```
COPY ["app.js", "package.json", "package-lock.json", "./*"]
RUN npm install --production
COPY . .
ENTRYPOINT ["node", "app.js"]
```

VI. Continuous Integration

Ensuring the security of both internal and external build pipelines, as well as performing comprehensive scans on work products traversing these pipelines, is a critical aspect of modern cloud-native operations. These measures are essential to maintain the integrity, reliability, and security of the development and deployment processes. In alignment with these objectives, a fully operational Cloud Native Continuous Integration (CI) pipeline has been meticulously designed and implemented within the lab environment.

The implemented CI pipeline is designed to handle a wide range of repository types, whether they are public repositories or private ones that require secure authentication credentials. This flexibility ensures that organizations can leverage the pipeline for diverse projects, enabling seamless integration with existing workflows

and infrastructure. The lab uses a specialized repository, referred to as the “containerizer” repository, to streamline the process of collecting files from source repositories, preparing them for deployment, and packaging them into secure and reliable container images.

The “containerizer” repository acts as a critical intermediary, bundling source files with essential components such as command-line tools, test cases, and additional dependencies. These elements are necessary to validate and prepare the application for deployment. Once all components have been processed and verified, the pipeline generates a Docker image and securely stores it in the gcr.io container registry. This repository not only ensures a centralized location for managing container images but also includes features like automated vulnerability scanning, providing an added layer of security to the deployment lifecycle.

A pipeline run is the foundational unit of operation in this system. It represents a sequence of automated tasks, including testing, building, and packaging, that must be completed successfully to produce a deployable artifact. These tasks are orchestrated and executed using Tekton, an open-source framework for creating cloud-native CI/CD pipelines. Tekton provides a scalable and flexible approach to managing these pipelines, allowing engineers to define tasks and workflows as code. This enables greater transparency, reproducibility, and adaptability to evolving project requirements.

The CI pipeline offers several key capabilities and benefits:

- 1) **Automated Testing and Validation:** Each code commit triggers a series of automated tests to validate functionality, security, and compliance with predefined standards. This ensures that only high-quality code progresses through the pipeline.
- 2) **Dynamic Source Code Integration:** The pipeline seamlessly integrates with source repositories, whether public or private, enabling organizations to incorporate multiple contributors and projects without compromising security.
- 3) **Comprehensive Security Scans:** The CI pipeline incorporates tools such as Bridgecrew to perform “Security as Code” scans, identifying and mitigating vulnerabilities in the codebase and container images.
- 4) **Containerization and Image Management:** Leveraging the “containerizer” repository, the pipeline produces Docker images that are ready for deployment. These images include all necessary tools, libraries, and configurations, ensuring consistency across environments.
- 5) **Storage and Accessibility:** All generated Docker images are securely stored in the gcr.io container registry. This centralized **repository** not only simplifies version management but also facilitates the secure distribution of images across environments.
- 6) **Scalability and Modularity:** Tekton allows for the modular definition of tasks and workflows, enabling the CI pipeline to scale with the needs of the project. Engineers can add or modify tasks without disrupting the overall pipeline.
- 7) **Enhanced Observability:** The pipeline includes logging and monitoring tools to provide real-time insights into its **performance**. This allows engineers to quickly identify and resolve issues during pipeline execution.
- 8) **Seamless Deployment Readiness:** By the end of a successful pipeline run, a fully validated and packaged Docker image is produced, ensuring that deployments to production environments are secure, efficient, and reliable.

The role of Tekton in orchestrating the CI pipeline cannot be overstated. Tekton’s task-based framework provides granular control over each step of the pipeline, allowing for extensive customization and optimization. Tasks within a pipeline can include activities such as linting, compiling, testing, security scanning, and packaging. Tekton also supports pipeline-as-code practices, enabling the storage of pipeline definitions in source repositories alongside application code. This ensures that the pipeline is version-controlled, auditable, and easily shareable.

Furthermore, Tekton’s integration with Kubernetes adds a layer of scalability and resilience to the pipeline. By running tasks as Kubernetes pods, the pipeline can dynamically allocate resources based on workload demands. This not only optimizes resource utilization but also ensures that pipeline tasks are isolated and secure.

In summary, the fully operational Cloud Native Continuous Integration pipeline implemented in the lab

exemplifies best practices in modern DevOps. By combining automated testing, secure containerization, and dynamic orchestration through Tekton, the pipeline serves as a powerful tool for accelerating development while maintaining the highest standards of security and reliability. Its design and capabilities demonstrate the transformative potential of cloud-native technologies in streamlining CI/CD processes, making it an indispensable component of the lab's infrastructure.

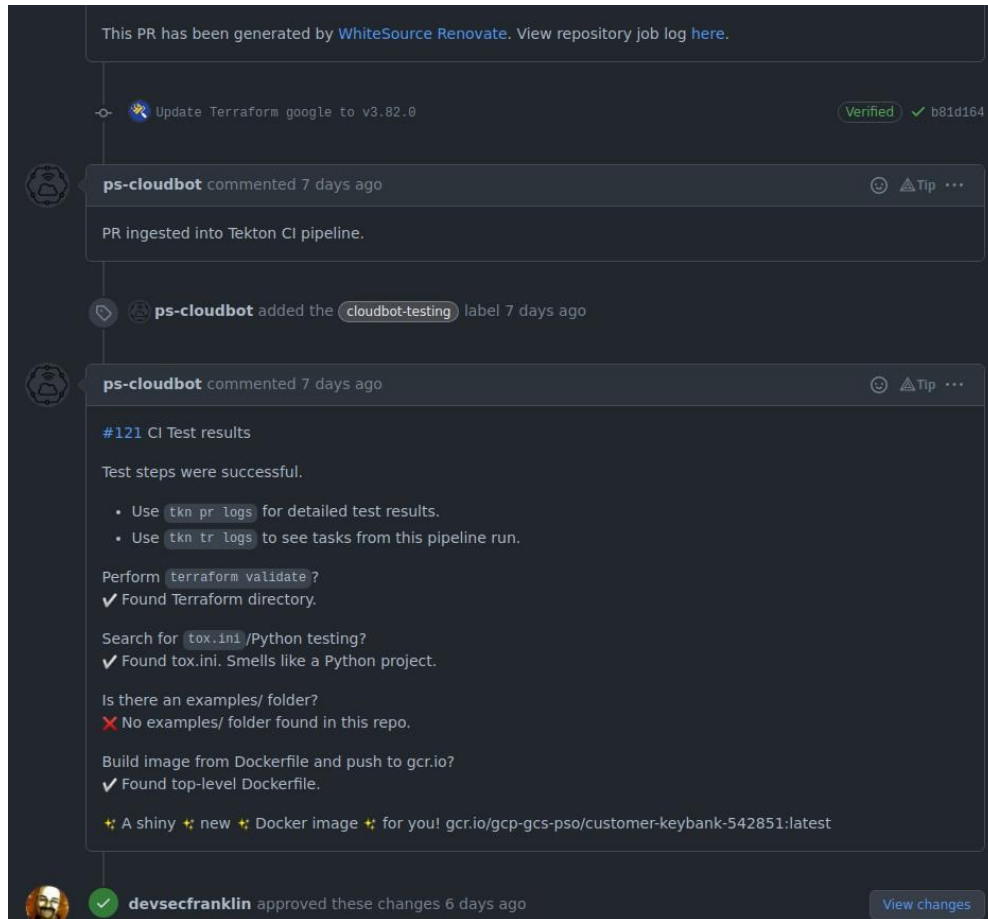


Figure 4. Tekton automated pipeline run results in a GitHub comment

Consider figure 4. The “Renovate” bot detects an out-of-date or insecure dependency. A pull request is opened by the Renovate bot. Next, the “ps-cloudbot” GCP cloud function is notified of the new pull request via webhook. The second bot places a label on the pull request to indicate it is performing administrative actions on the pull request. The cloud function might perform other actions, such as assigning the pull request to a certain user, adding certain users as reviewers, providing documentation, and so on. In this example, the cloud function adds a comment to the pull request to inform project members that it has been accepted by the CI pipeline. A set of test cases based on a certain technology or functional area is executed, and the results are returned to the pull request in a second comment. There is also a link to the container image in the container registry. Although the pull request is merged manually in this instance, the workflow could be modified to approve and merge the pull request with no human interaction whatsoever.

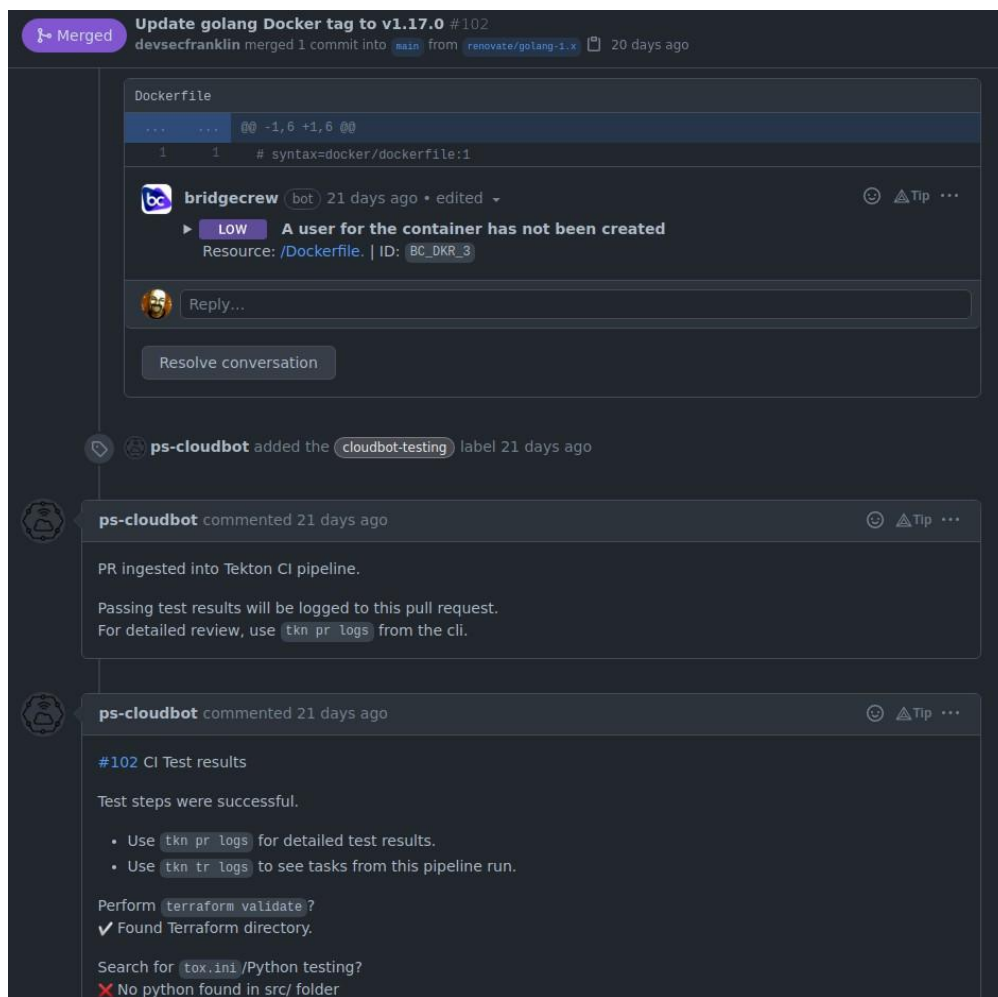


Figure 5. Bridgecrew integration with pull request automation

In addition to the ability to scan for dependencies that are out of date or have vulnerabilities of note, automated security scanning of the code base can be performed as seen in figure 5. As before, the Renovate bot has detected the use of an out-of-date Docker base image for Golang. Because the Dockerfile is updated as part of this pull request, Bridgecrew scans the file and triggers misconfigurations that may lead to security issues. The results of the scans and the issues found are noted in the pull request comments. Automated remediation and merging of these issues may be possible in some cases.

VII. Conclusion

The development and implementation of the Cloudlab project represent a significant step forward in advancing cloud-native security research and training. By combining state-of-the-art tools, methodologies, and infrastructure, the lab provides a robust platform for exploring modern IT challenges, fostering innovation, and equipping engineers with practical expertise. While the lab has already achieved substantial progress in areas such as container security, Policy as Code, and Continuous Integration pipelines, the potential for future enhancements remains vast.

Looking ahead, the lab aims to address several ambitious goals that will further expand its scope and capabilities. These goals include:

- 1) **Orchestration of Multi-Cloud Infrastructure Builds and Deployments:** As organizations increasingly adopt multi-cloud strategies to optimize resource utilization and ensure redundancy, the lab seeks to explore advanced orchestration techniques. This includes leveraging tools like Crossplane, which allows seamless integration and management of resources across multiple cloud providers. By achieving this, the lab

can serve as a testbed for developing scalable, resilient, and interoperable multi-cloud solutions.

- 2) **Expanding Knowledge of Infrastructure as Code (IaC) Languages:** While tools like Terraform (using HCL) have become synonymous with IaC, the lab aims to broaden its horizons by delving into alternative IaC languages and **frameworks**. For instance, Pulumi offers a programming-language-based approach to IaC, enabling the use of familiar languages like Python, JavaScript, and Go for defining and managing infrastructure. Exploring Pulumi will allow the lab to compare and evaluate the strengths and use cases of different IaC paradigms.
- 3) **Conducting Red Team Exercises:** Security remains at the heart of the Cloudlab project, and one of the key future goals **involves** “red teaming” the lab itself. By conducting dynamic application security testing (DAST) exercises, the lab can identify vulnerabilities, validate its defenses, and simulate real-world attack scenarios. These exercises will not only enhance the lab’s security posture but also provide invaluable insights into how cloud-native systems can be hardened against evolving threats.

These future objectives highlight the lab’s commitment to continuous improvement and adaptability in an ever-changing technological landscape. By orchestrating multi-cloud environments, expanding expertise in IaC languages, and proactively addressing security vulnerabilities, the lab will remain at the forefront of cloud-native innovation.

In conclusion, the Cloudlab project exemplifies the transformative potential of cloud-native technologies. It bridges the gap between theoretical knowledge and practical application, empowering researchers and engineers to tackle the complexities of modern IT ecosystems. With its robust infrastructure, innovative methodologies, and ambitious future goals, the lab not only supports groundbreaking research but also fosters a culture of learning and collaboration. As the lab evolves, it is poised to play a pivotal role in shaping the future of secure, scalable, and adaptive cloud-native environments.

References

- [1] C. Pettey, “Cloud shift impacts all it markets,” <https://www.gartner.com/smarterwithgartner/cloud-shift-impacts-all-it-markets/>, 2020, [Online; accessed 2021-09-05].
- [2] E. Alpaydin, *Machine Learning: The New AI*. MIT Press, 2016.
- [3] G. Kim, *The DevOps handbook : how to create world-class agility, reliability, & security in technology organizations*. Portland, OR: IT Revolution Press, LLC, 2016.
- [4] D. Sullivan, *Official Google Professional Cloud Architect : study guide*. Hoboken, NJ: Sybex, a Wiley brand, 2020.
- [5] B. Burns, *Kubernetes best practices : blueprints for building successful applications on Kubernetes*. Sebastopol, CA: O’Reilly Media, 2019.