# Running version control systems and continuous integration to fulfill CERN and LHC experiments needs

Ismael Posada Trobo, Konstantinos Evangelou, Subhashis Suara

## Introduction

At CERN we host an on-premises instance of GItLab to provide our Version Control System for our scientific community. GitLab is a self-service code hosting application based on Git that provides collaboration and code review features for users and CERN service providers. It is one of the most critical infrastructures running at CERN, being widely used daily by more than 19,000 users, and hosting more than 120,000 projects, fitting the needs of the entire community at CERN. Furthermore, we provide several Kubernetes clusters to host the GitLab Runners that are needed for the execution of CI/CD pipelines. Each cluster provides different capabilities based on the needs of our users.

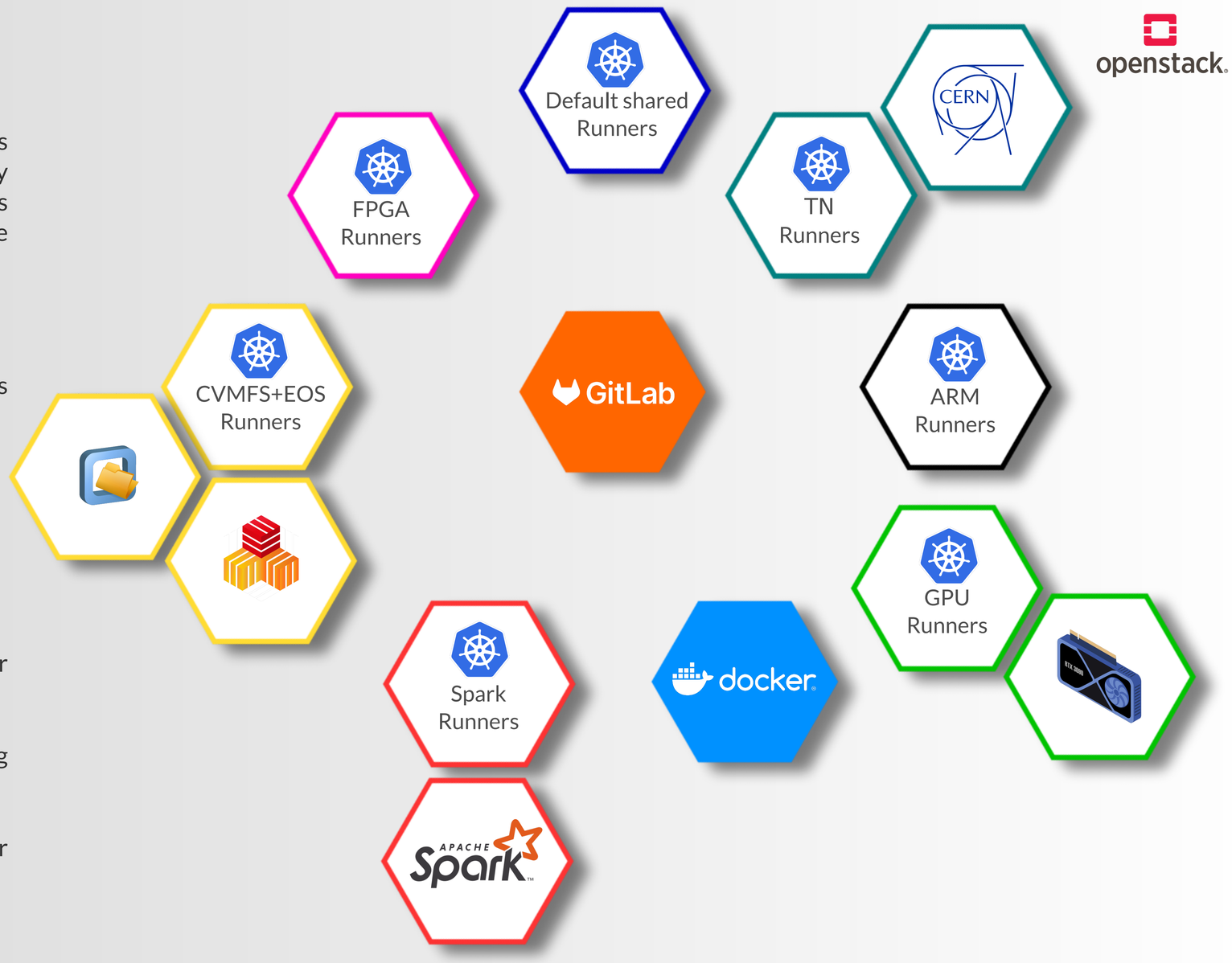
## CERN's Kubernetes Runner Solution

Kubernetes runners offer the benefits of reliability, scalability, and availability inherent in Kubernetes, providing a robust infrastructure that supports our users' diverse activities. This new system is better suited to the needs of CERN, which spans research, IT infrastructure, and development. Kubernetes runners can scale from 0 to 100 pods based on demand and handle 200 jobs concurrently when needed. With multiple clusters, we tailor capabilities for our wide user base, including physicists, machine learning developers, and hardware designers. Using OpenTofu, we create and manage clusters seamlessly, ensuring flexibility and rapid recovery from any issues.

## One Platform, Many Clusters

With multiple clusters, we are able to provide a variety of capabilities to our users based on their needs. Having an extended community means that there is not a silver bullet for shared runners. Hence, it is GitLab's great responsibility to provide multiple instances to facilitate users' activities. Those instances are:

- **Default cluster:** Generic runners for the majority of use cases.
- **CVMFS cluster:** CVMFS stands for CernVM File System. This cluster mounts CVMFS and EOS volumes per job.
- **ARM cluster:** For building ARM executables.
- **FPGA cluster:** Supports FPGA-based systems for tailored circuits.
- **Apache Spark cluster:** For Data-Analysis workloads.
- **GPU cluster:** Run GPU jobs for parallelizable workloads for machine learning jobs.
- **Technical Network cluster:** An air-gapped deployment offering connectivity for accelerator control devices at CERN.
- **Docker Privileged Runners:** Privileged execution of docker containers with root access in one-off hosting VMs.




## Boosting Security

The Kubernetes runners brought significant security improvements by design:

- The GitLab runner manager runs only in the master nodes of the cluster making it inaccessible from the workers.

Future security enhancements include:

- Plans to disable umask 0000 for Kubernetes runner pods, which will align the execution permissions with the user permissions built in the docker image used.
- Adaptation of the "User Namespaces Stateless Pods Support" feature which will enable the separation of pods access to host system, run different pod uids than the host system and execute pods in distinct namespaces.

## System Design Benefits

- **Cluster decoupling:** We can maintain each cluster separately and have specific end-to-end tests without interfering with the GitLab Application.
- **Zero downtime Cluster upgrades:** We can upgrade the runners' clusters to a more recent Kubernetes version with zero downtime by simply creating a new cluster and then registering it as an instance runner with the same tags.
- **Scalability:** We can scale the clusters by adding more nodes which will allow better resource allocation per pod and faster execution of the jobs.
- **Easy Cluster Creation:** We use Infrastructure as Code to create clusters seamlessly for the different types of clusters using OpenTofu.

## The Impact of the Kubernetes Runners

The implementation of GitLab Kubernetes runners has significantly improved the number of concurrent jobs executed, supported diverse workflows, and reduced operational costs associated with our infrastructure. Decoupling the clusters has accelerated our deployment, testing, and provisioning processes, enhancing maintainability. While challenges remain, we are determined to follow GitLab's guidelines and best practices to ensure success. The Git Service is proud to provide our users with exceptional infrastructure that meets the needs of the vast scientific community at CERN.

https://home.cern

Ismael Posada Trobo
Tech Lead & Engineering Manager at CERN
ismael.posada.trobo@cern.ch