

High availability Kubernetes cluster using Octavia Ingress Controller (Remote Presentation)

Tuesday, 26 March 2024 15:00 (30 minutes)

With the widespread adoption of containers by various organizations and companies, Kubernetes (K8s), an open-source software dedicated to container management, has in recent years become de-facto the standard for the deployment and operation of applications focused on this technological solution. K8s offers several advantages: workload balancing, dynamic resource allocation, automated rollout and rollback, storage orchestration, management of sensitive information, self-healing, etc. Such a resilient infrastructure allows us, for example, to deploy authentication systems used in the WLCG world. Furthermore, the same infrastructure will be used in the future to provide services to the developer group. Obviously K8s has some limitations that can be supplanted by the easy integration with many external software.

Thanks to its flexibility and scalability features, K8s can be integrated with cloud native solutions such as Openstack, a modular cloud operating system capable to offer computing and storage management services according to the Infrastructure as a Service (IaaS) paradigm, deployed at INFN CNAF.

The inner complementary relationship between K8s and Openstack has pushed us to widely use this solution in our Cloud infrastructure. An interesting aspect of the integration between the two software, which we are investigating, is the possibility of exposing K8s services externally via a Load Balancer (LB), by making use of the Octavia service. Octavia is an open source, operator-scale load balancing solution designed to work with OpenStack. Octavia achieves load balancing services by overseeing a fleet of virtual machines, containers, or bare metal servers, referred to collectively as amphorae. These amphorae are dynamically deployed as needed, distinguishing Octavia from alternative load balancing solutions and making it particularly well-suited for cloud environments.

Using Octavia to publish a service externally will enable a sort of service resilience given by the adoption of a Master-Backup HAProxy. In the event of a master downtime, the backup will assume the role of master, keeping external communication alive. The advantage of using the Octavia service is to limit the resources (e.g. FloatingIPs) used. Furthermore, delegating the creation of the LB, and therefore the VMs, to Octavia may allow us to speed up the process and avoid any human errors resulting from manual installation.

In the present work the deployment of the functionality is shown and performance of the whole system will be provided to prove the solidity of the adopted technological solutions.

Primary author: SINISI, Francesco (INFN-CNAF)

Co-authors: ALKANSA, Ahmad (INFN CNAF); COSTANTINI, Alessandro (INFN-CNAF); Dr MICHELOTTO, Diego (INFN CNAF)

Presenter: SINISI, Francesco (INFN-CNAF)

Session Classification: Infrastructure Clouds & Virtualisation

Track Classification: Track 8: Infrastructure Clouds and Virtualizations