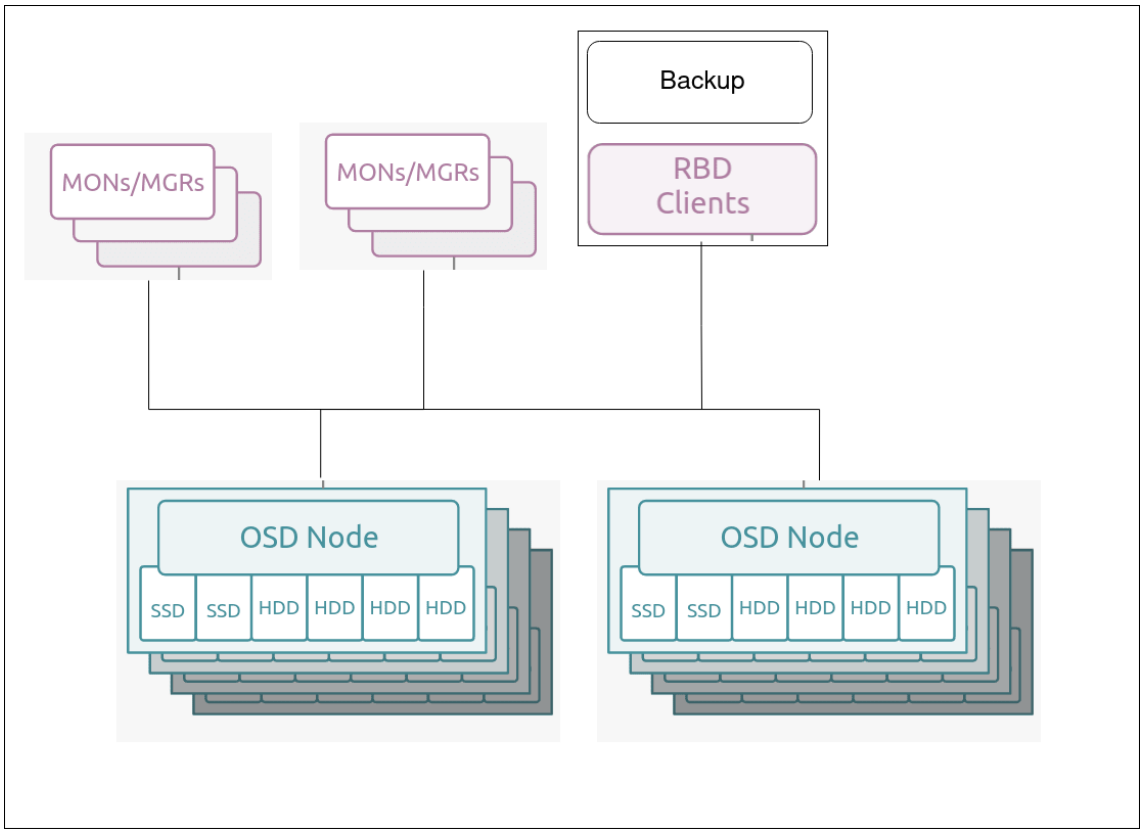
University of Porto, Faculty of Science

Ceph Cluster Using Virtual Machines and Terraform with Backup Strategy

System Administration

October 3, 2023



1 Project Description

In this systems administration project, you will have the opportunity to design, deploy, and manage a Ceph cluster using virtual machines (VMs) and optionally, Terraform. Ceph is

a distributed storage system widely used in enterprise environments for its scalability and reliability. Your task is to create a functional Ceph cluster with the following components:

- Object Storage Devices (OSDs): These are responsible for storing data within the Ceph cluster.
- Monitor Nodes (MONs): MONs are essential for maintaining cluster metadata and monitoring cluster health.
- Manager Nodes (MGRs): MGRs provide management and monitoring services for the Ceph cluster.

In addition to setting up the Ceph cluster and implementing a backup solution, your project will also involve deploying an RBD client to interact with the Ceph cluster.
All the project will be done using GCP (Google Cloud Platform)

2 Architecture

2.1 Distributed Storage Cluster

The distributed storage cluster forms the core of the architecture, consisting of OSDs, MONs, and MGRs. It requires a total of four virtual machines (VMs) provisioned for the primary Ceph cluster:

- **2x OSD VMs**
- **1x MON VM**
- **1x MGR VM**

2.2 Client Integration

Client integration is a crucial component of the architecture, and it is implemented with the backup node to prevent an additional layer of complexity. PostgreSQL will be deployed over RBD (Rados Block Device) for data storage. This adds complexity to the client environment while showcasing the versatility of the solution.

- PostgreSQL Deployment: PostgreSQL, a powerful open-source relational database management system, will be deployed within the client environment. It will be configured to store and manage data using RBD for efficient and scalable storage.

This architecture choice provides a seamless and reliable database solution within the client environment, utilizing Ceph's RBD for data storage and replication. It enhances data availability and ensures that client applications can leverage PostgreSQL's features and scalability for their data storage needs.

2.3 Backup Solution with Rsync

The backup solution is implemented using ***rsync***, a reliable and efficient tool for file synchronization. It requires an additional VM:

- **1x VM for Backup:** This VM is dedicated to serving as the backup server, utilizing rsync for regular backups, and storing backup copies securely. The backup server will synchronize data from the Ceph cluster to ensure data protection.

The combination of rsync for data synchronization and a centralized backup repository ensures comprehensive data protection within the project.

3 Project Tasks

- Use Terraform to provision and configure the necessary VM instances for OSDs, MONs, MGRs, backup components, and the RBD client within Google Cloud. (optional, but will be valued)
- Install and configure the Ceph software on the provisioned VMs as previously outlined.
- Setup the Ceph cluster, specifying roles for each VM and configuring networking and storage within the Google Cloud environment.
- Deploy and configure an RBD client on the designated Google Cloud VM to interact with the Ceph cluster for data storage and retrieval.
- Deploy a backup solution that is compatible with the Ceph cluster and configure it to protect cluster data.
- Test the Ceph cluster and the RBD client to ensure data storage, replication, and retrieval functionality within Google Cloud.
- Perform backup and recovery tests to validate the effectiveness of the backup solution in the cloud environment.
- Monitor and evaluate the health and performance of the Ceph cluster, backup system, and RBD client within Google Cloud.
- Document your entire project, including configurations, backup strategies, RBD client setup, and troubleshooting steps.
- Prepare a project presentation to showcase your cluster working. (recommend installing **pgAdmin 4** to showcase connection to the DB)

4 Evaluation Criteria

You will be assessed based on your project documentation, presentation, the functionality and reliability of your Ceph cluster within Google Cloud, the effectiveness of your backup solution, and the successful deployment and configuration of the RBD client in the cloud environment.