

Project: Automation of Spark Deployment with Ansible and Terraform

Description

In this project, we will automate the deployment of a Big Data infrastructure based on Apache Spark. To achieve this, we will use Terraform to set up the infrastructure (at least three virtual machines on KVM with libvirt) on minimum 2 physical hosts and Ansible for the deployment and configuration of Spark.

Additionally, we will use WordCount as a test application to validate the proper functioning of the installation.

Project steps.

1. **Build the Infrastructure.** Use Terraform to create and configure at least four virtual machines on a KVM hypervisor and this on two physical machines.
2. **Automate the Deployment of Spark.** Write Ansible playbooks to install and configure Apache Spark on the VMs.
3. **Validation of the Deployment.** Deploy the WordCount application to test the proper functioning of the infrastructure and Spark using Ansible
3. **Write the report.** The report should be at least 3 pages and you should describe your results.
4. **A demo is mandatory.**