Virtual Labs Cluster Automation

1. Introduction

This documentation is about my internship at IIIT Hyderabad on topic automating the Virtual Labs cluster setup using ansible(Configuration management tool).

2. Problem

The problem is bootstrapping steps for cluster creation. For creating VLEAD cluster(Infrastructure for hosting labs), bootstrapping steps takes three to four days for VLEAD Employees(who knows about VLEAD cluster) and for others it may take two to three weeks.

3. Objective

Automate the boostrapping steps and running system-model ansible playbooks. So that eliminate more number of days to setup VLEAD cluster and install cluster in minimal steps in one day.

4. Pre-Requisities and tools used:-

4.1. Version control

The cluster model would involve timely changes, be it in the configuration of the nodes or the architecture of the cluster. Version control provides a mechanism which would keep track of all such changes. Along with tracking changes, it also provides the facility to revert back to a previous state of the system. The version control system used in this model is "git".

4.2. OpenVZ

OpenVZ is a simple open source kernel level virtualization. It is used to create several isolated containers (Virtual Environments) on the same physical resources. Details on why OpenVZ is used can be found.

4.3. Networking

Some basics topics in networking that are required to know are:

- Bridged networks (using OpenVZ)
- Concept of subnets
- Basic knowledge of IP addressing

4.4. Ansible

Ansible is a configuration management tool which is used for configuring the cluster. It provides remote configuration of nodes over SSH.

5. Bootstrapping of a cluster

An independent operating system such as a container or a VM or even a physical machine is referred as a node. The collection of such related and connected nodes forms a cluster. Each node in the cluster has a specific purpose that it serves. To setup an entire cluster initially, each of the component nodes are setup individually in a planned and specific manner. This process of setup of individual nodes in a specific manner to setup a new cluster is called the bootstrapping process. This process is tightly related to the provisioning method used on the corresponding cluster.

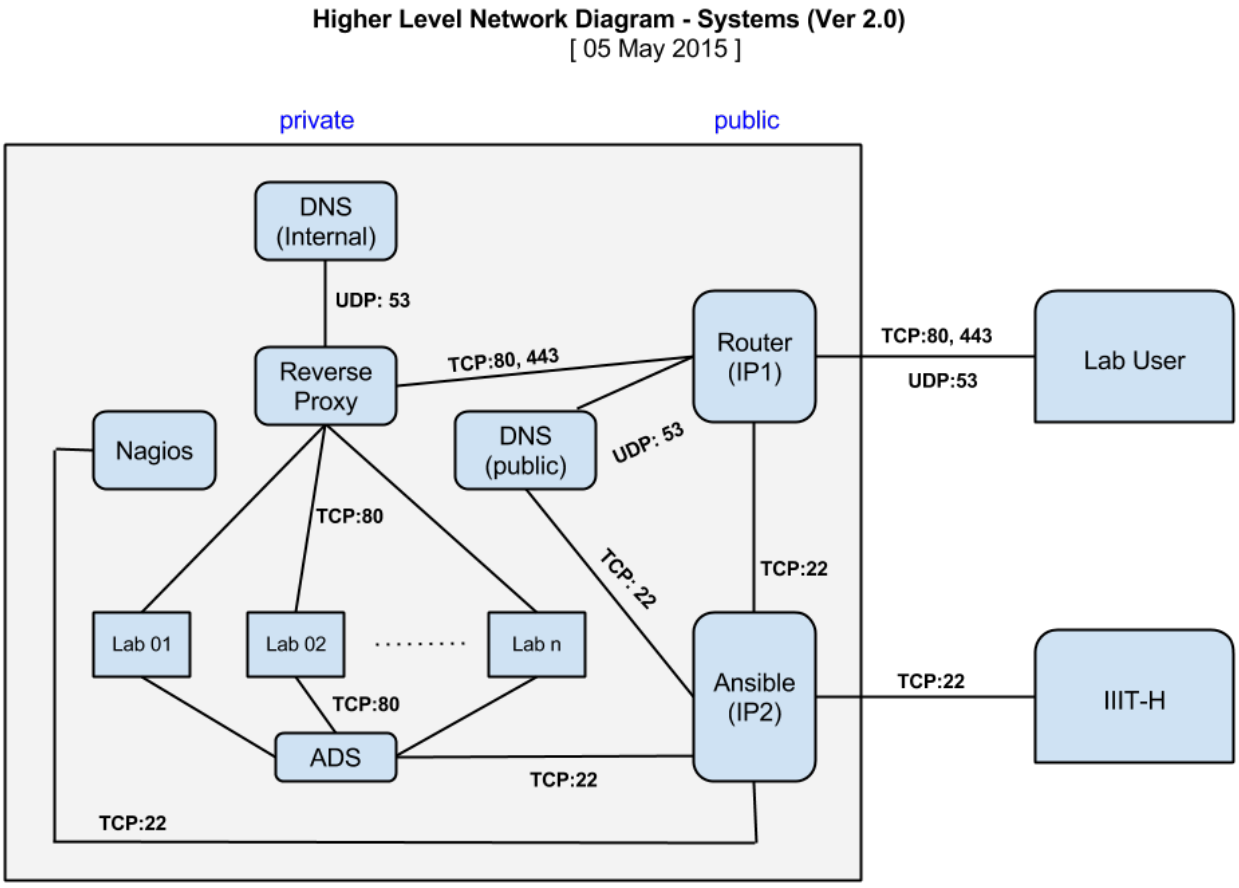
6. Nodes in the cluster

There are several nodes in the cluster which are all common to the various clusters. The list of nodes is as below:

- Rsyslog server
- Private DNS
- Public DNS
- Reverse Proxy
- Nagios server
- Router
- Ansible server
- OSSEC server
- Rsnapshot server
- ADS

7. Design of the cluster model

The diagram below depicts the architecture of the nodes in the cluster model. The diagram shows the network-setup of the cluster and connectivity between a few key nodes. In this diagram IP1 and IP2 refer to the only two public IP's being used by the system. IP1 is the public IP assigned to router and IP2 is the public IP assigned to ansible. Normal virtual-lab users only see and contact router. VLEAD team uses ansible to manage the cluster.This diagram focuses on depicting the important connections in the network. There are several other connections which are not shown. For example, ansible is connected to every node in the cluster. All the connections are not shown in the diagram. Also there are some nodes which are not shown, for example the rsnapshot server, rsyslog server, ossec server are not shown. But these nodes are configured and are being used in the cluster but are not shown to keep the diagram simple.



8. Implementation

Next Contains implementation of each and every node and small snippets of code