

Establishing a network between the nodes

The first thing to do is to connect the networks physically using a router or a hub. Here, I will be connecting multiple virtual machines using a virtual router. Then, we need to make each node discoverable to the other. Finally, we need to share the master's public key with the slaves. That concludes the network required to be setup for running Hadoop and HDFS.

1. Name the computer > Open `/etc/hostname` with `sudo` permissions and save the file with the computer-name.
> > **NOTE:** Do not use underscores in the name because they are restricted and will make your computer undiscoverable or not work in the expected manner on the network. (This applies possibly for other special characters as well).
2. Setup a static IP for each node (Ubuntu 18.04 and Ubuntu 20.04)
 - Choose the 'Wired Network Settings' under Network.
 - Go to the IPv4 tab, and choose to manually configure the network.
 - Assign your static IP and netmask for your gateway under 'Address'
 - Set your DNS (192.168.43.1)
 - Copy your 'Address' into your 'Route'
 - Repeat this process for each node
3. Configure `/etc/hostname`
 - For each node, make an with its IP address and name.
4. (Not sure if this step is required) Configure `/etc/hosts.allow`
 - make an entry `ALL:ALL` to allow each node to access the other nodes.

SOURCES: 1. `/etc/hostname` naming restrictions