

LAB-07

- Jainil Trivedi (CE166)

➤ **AIM:** SSH, User Equivalence, Clusters / Cloud / Horizontal Scaling

1. We can check host name using command **hostname** in the terminal.
2. We can see IP address using command **ipconfig** in the terminal.
3. Now we connect to ssh using command
ssh<ip address> (ex: 192.168.32.35)
4. It would then ask us for password to connect. We can also configure to connect without having to give password as input.

- **Application of clustering:**

- market research, like which products are bought together.
- Finding patterns, similar to market research, we can find out and suggest people what they should buy, for example, if a customer buys laptop, we can suggest them external mouse.
- It can also be used in medical sector.
- Credit card fraud detection: If we have credit card from India and while visiting USA if we make a payment, we will be alerted.

- **Purpose of clustering:**

- Enables us to store and perform analytics on large data like 11 Tb, with many computers working as a worker.
- In lab the master node was also made worker

- **Set Up:**

- Stop all processing using
stop-*.sh
-
- In core-site.xml, add tag to define master node,

```
<property>
```

```
    <name>fs.defaultFS</name>
```

```
    <value>hdfs://hadoop-master:9000/</value>
```

```
</property>
```

- In hdfs-site.xml, we define list of worker nodes that we want to add. For the same purpose, we need to make a file with the list of worker nodes, IP addresses

```
<property>
```

```
    <name>dfs.datanode.data.dir</name>
```

```
    <value>/opt/hadoop/tmp/dfs/data</value>
```

```
</property>
```

- Now, start the distributed file system and yarn. Using

```
start dfs.sh
```

```
yarm.sh
```