

Experiment -1

Installation of Hadoop Single Node Cluster



**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013



What Is Virtualization ?

- Virtualization is a Technology that transforms hardware into software.
- Virtualization allows to run multiple operating systems as virtual machines.
 - Each copy of an operating system is installed in to a virtual machine.

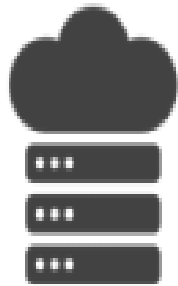


**PRESIDENCY
UNIVERSITY**

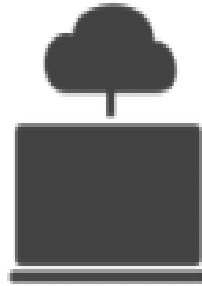
Private University Estd. in Karnataka State by Act No. 41 of 2013



TYPES OF VIRTUALIZATION



Server
Virtualization



Desktop
Virtualization



Application
Virtualization



Network
Virtualization



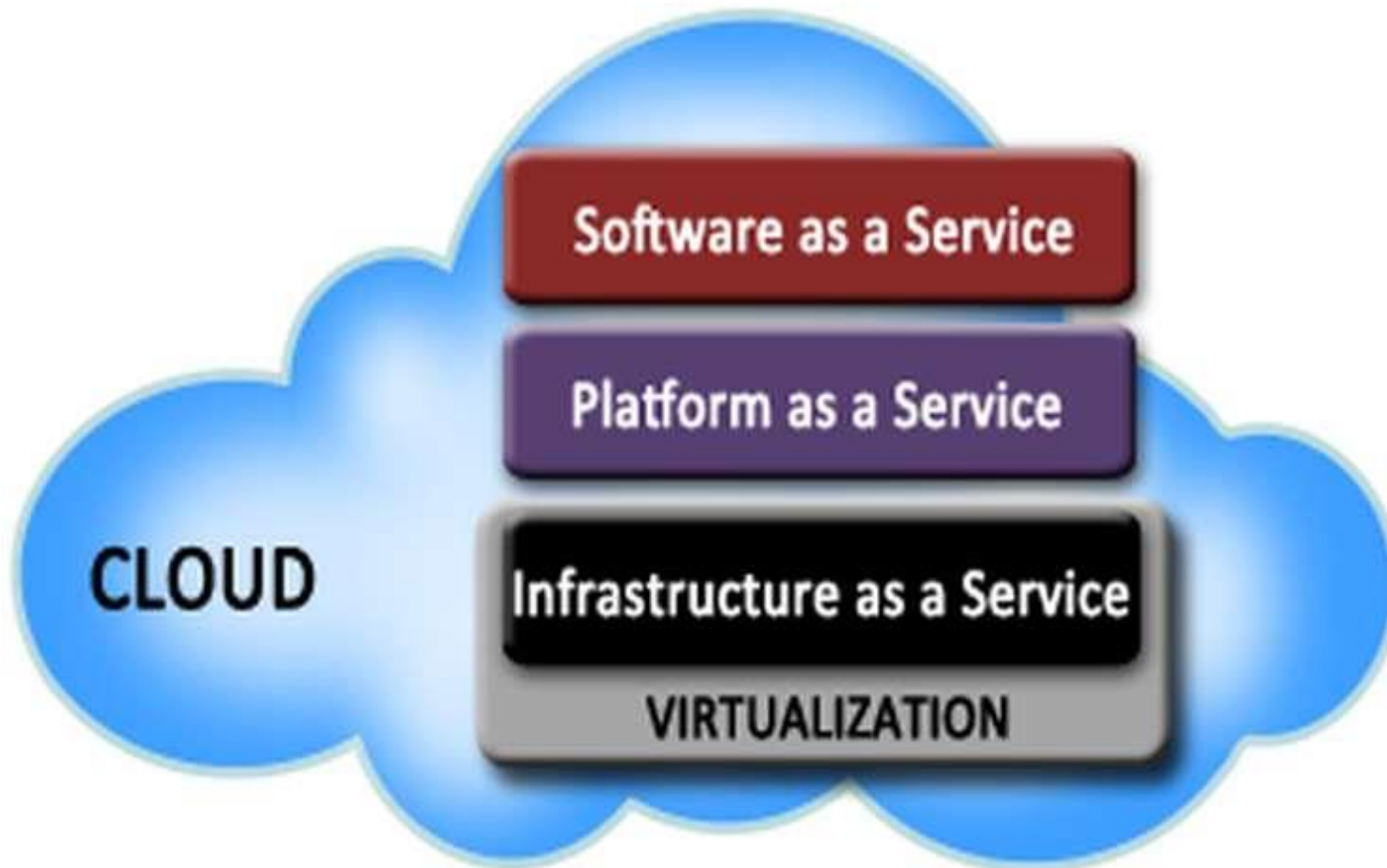
Storage
Virtualization



**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013





**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013



Infrastructure-as-a-service (IaaS)

- With IaaS, you rent IT infrastructure—servers and virtual machines (VMs), storage, networks, operating systems—from a cloud provider on a pay-as-you-go basis.

Platform as a service (PaaS)

- Platform-as-a-service (PaaS) that supply an on-demand environment for developing, testing, delivering and managing software applications.
- PaaS is designed to make it easier for developers to quickly create web or mobile apps, without worrying about setting up or managing the underlying infrastructure of servers, storage, network and databases needed for development.



**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013



Software as a service (SaaS)

- Software-as-a-service (SaaS) is a services for delivering software applications over the Internet, on demand and typically on a subscription basis.
- With SaaS, cloud providers host and manage the software application and underlying infrastructure and handle any maintenance, like software upgrades.



**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013



VMware Workstation

- VMware Workstation is a hosted hypervisor that runs on x64 versions of Windows and Linux operating systems.
- It enables users to set up virtual machines on a single physical machine, and use them simultaneously along with the actual machine.



**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013



REQUIREMENTS

- A Laptop or Desktop with 32 or 64 bit with window (Linux operating system when using VM)
- Minimum of 2 Gb ram
- Minimum of 100 gb hard disk
- Minimum VGA is required



**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013



Installing of Hadoop Single Node Cluster Configuration.

1. Downloading the software Required
2. Untar the software
3. Bashrc configurations
4. Hadoop Configuration File
5. Share public key to localhost
6. Formatting the name node
7. Starting Hadoop Daemons
8. Checking the working hadoop Daemons



**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013



STEP-1

SOFTWARES REQUIRED

- Hadoop 1.2.0-bin.tar.gz
- Jdk 7u67-linux-i586.tar.gz

Link for JDK

<http://www.oracle.com/technetwork/java/javase/downloads/index.html>

Link of Hadoop :

<http://mirror.fibergrid.in/apache/hadoop/common/hadoop-1.2.0/>



**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013



STEP-2

Untar the software

- Open the terminal window
- Type the command : `cd Desktop`
- Type `ls` command
- `tar -zxvf jdk-7u67-linux-i586.tar.gz`
- `tar -zxvf hadoop-1.2.0-bin.tar.gz`



**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013



Step 3

Bashrc configurations

Open the terminal type the command

```
sudo gedit ~/.bashrc
```

```
export JAVA_HOME=/home/user/Desktop/jdk1.7.0_67
```

```
export PATH=$PATH:$JAVA_HOME/bin
```

```
export HADOOP_HOME=/home/user/Desktop/hadoop-1.2.0
```

```
export PATH=$PATH:$HADOOP_HOME/bin
```

Open new terminal and check for java version and Hadoop version.



**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013



Step-4

Hadoop Configuration Files

1. Core-site.xml

(Configuration setting for hadoop core, such as I/O setting that are common to HDFS and Mapreduce)

2.Hdfs-Site.xml

(Configuration setting for HDFS daemons the name node, Secondary name node and data node and Replication factor as well)



**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013



3. Mapred-site.xml

(configuration setting for MapReduce daemons: the job tracker and the task tracker)

4. Hadoop-Env_sh

(Environment variables that are used in the scripts to run Hadoop)



**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013



- Open the new terminal window
- Go to the hadoop folder
- cd Desktop
- cd Hadoop (Press tab) for path
- cd conf
- Type ls
- All the above mentioned files are listed



- Go to terminal and type

sudo gedit core-site.xml

<configuration>

<property>

<name>fs.default.name</name>

<value>hdfs://localhost:9000</value>

</property>

</configuration>

Ip address :127.0.0.0 , 198.1.27.0

Port no: 80,



**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013



- Go to terminal and type

sudo gedit hdfs-site.xml

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

<name>dfs.name.dir</name>

<value>/home/user/Desktop/hadoop-1.2.0/name/data</value>

</property>

</configuration>



**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013



- Go to terminal and type

sudo gedit mapred-site.xml

<configuration>

<property>

<name>mapred.job.tracker</name>

<value>hdfs://localhost:9001</value>

</property>

</configuration>



**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013



- Go to terminal and type

sudo gedit hadoop-env.sh

export JAVA_HOME=/home/user/Desktop/jdk1.7.0_67



**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013



STEP-5

SSh Configuration

- `sudo apt-get install ssh`
- `ssh-keygen -t rsa -P ""`

Sharing the public key with host

- `ssh-copy-id -i ~/.ssh/id_rsa.pub user@ubuntuvvm` and check
- `ssh localhost` (It should not ask password)

STEP 6

Formatting the Name Node

- \$ `hadoop namenode -format`
- format a new distributed file system
- Process creates an empty file system for creating the storage directories



**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013



STEP-7&8

Starting Hadoop Daemons

Open the terminal window and type

```
$ start-all.sh
```

and type

```
$ jps
```



**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013



- **user@ubuntuvirtualmachine:~\$ start-all.sh**
- **Warning: \$HADOOP_HOME is deprecated.**
- **starting namenode, logging to /home/user/Downloads/hadoop-1.2.0/libexec/../logs/hadoop-user-namenode-ubuntuvirtualmachine.out**
- **localhost: datanode running as process 2978.**
- **localhost: secondarynamenode running as process 3123.**
- **jobtracker running as process 3204.**
- **localhost: tasktracker running as process 3342.**
- **user@ubuntuvirtualmachine:~\$ jps**
- **4020 Jps**
- **3342 TaskTracker**
- **3204 JobTracker**
- **3123 SecondaryNameNode**
- **3606 NameNode**
- **2978 DataNode**





**PRESIDENCY
UNIVERSITY**

Private University Estd. in Karnataka State by Act No. 41 of 2013

