LAB MANUAL
PART A
(PART A: TO BE REFERRED BY STUDENTS)

Experiment No-02

A.1 Aim:
   To install Hadoop

A.2 Prerequisite
   Knowledge of Java, Python and VMware software pack.

A.3 OutCome
   Students will be able to To acquire fundamental enabling techniques and scalable algorithms like Hadoop, Map Reduce and NO SQL in big data analytics.

A.4 Theory:
   To install Hadoop core clusters needed are:

   - Install java on the system

   - Install VMware

   - Download VM file of Hadoop

   - Load it into VMware and start

   Installation of Hadoop using IBM Infosphere Biginsight.

   Steps are:

   1. **Check for VTX mode required configuration.**

      8 GB RAM for better performance and minimum i3 processor with 80 GB space.

      Vtx mode setting is in-> BIOS under-> Security option.

   2. **Open fie into VMware.**

      The file is IBM Infospherebiginsight image

      - Open user Biadmin

      - It will start your OS, Redhat Linux.

      - It contains python, Java, IBMInfosphere and Eclipse IDE.

   3. **Start Hadoop with all its Components**:

      Cd   /opt/ibm/biginsight/bin

         ./start-all.sh

      Components of all Hadoop gets started.

      Hadoop starts successfully.

1

./status-all.sh

Status of all Hadoop components get displayed

PART B
(PART B: TO BE COMPLETED BY STUDENTS)

(Students must submit the soft copy as per the following segments within two hours of the practical. The soft copy must be uploaded on the Blackboard or emailed to the concerned lab in charge faculties at the end of the practical in case there is no Blackboard access available)

| Roll. No. 50	Name: AMEY MAHENDRA THAKUR
Class: COMPS-BE-B	Batch: B3
Date of Experiment: 28-07-2021	Date of Submission: 28-07-2021
Grade:	

B.1. Explain how to install Hadoop?
*(Paste your Search material completed during the 2 hours of practical in the lab here)*
- Hortoworks Sandbox
- IBM InfoSphereBiginsight

The steps are as follows:

1. Download IBM Infosphere BigInsights Quick Start Edition VMware Image.
2. Extract the downloaded file.
3. Open the 'RHEL55-64.vmx' file using VMware Workstation.
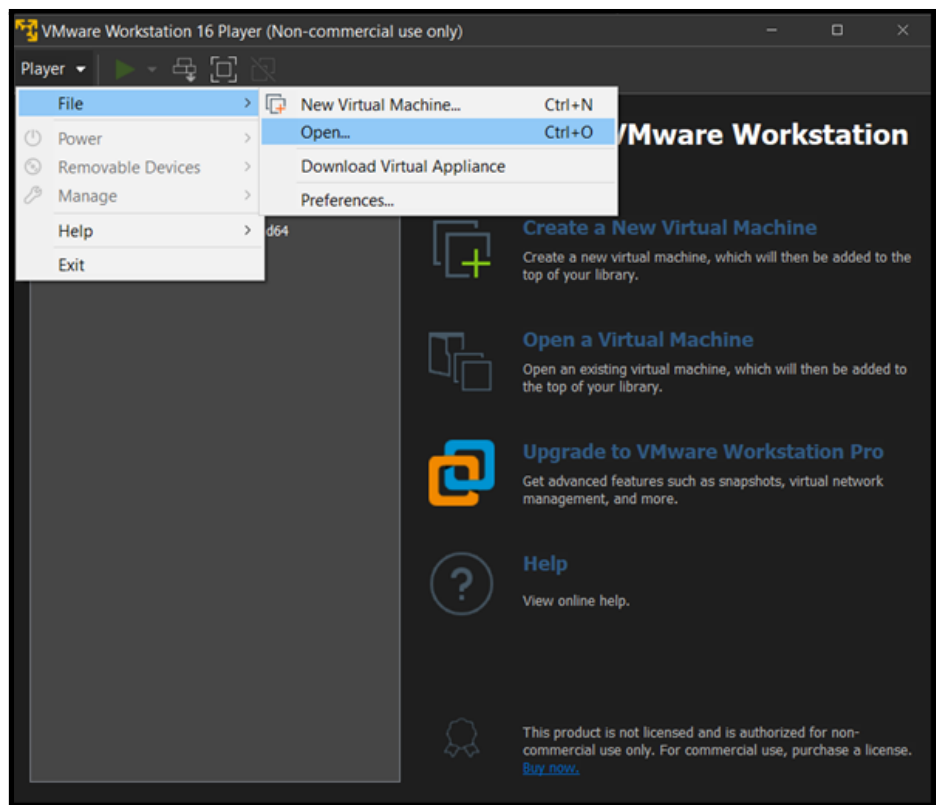
4. Start the VMware image by clicking the Play virtual machine button in the VMware Player if it is not already on.
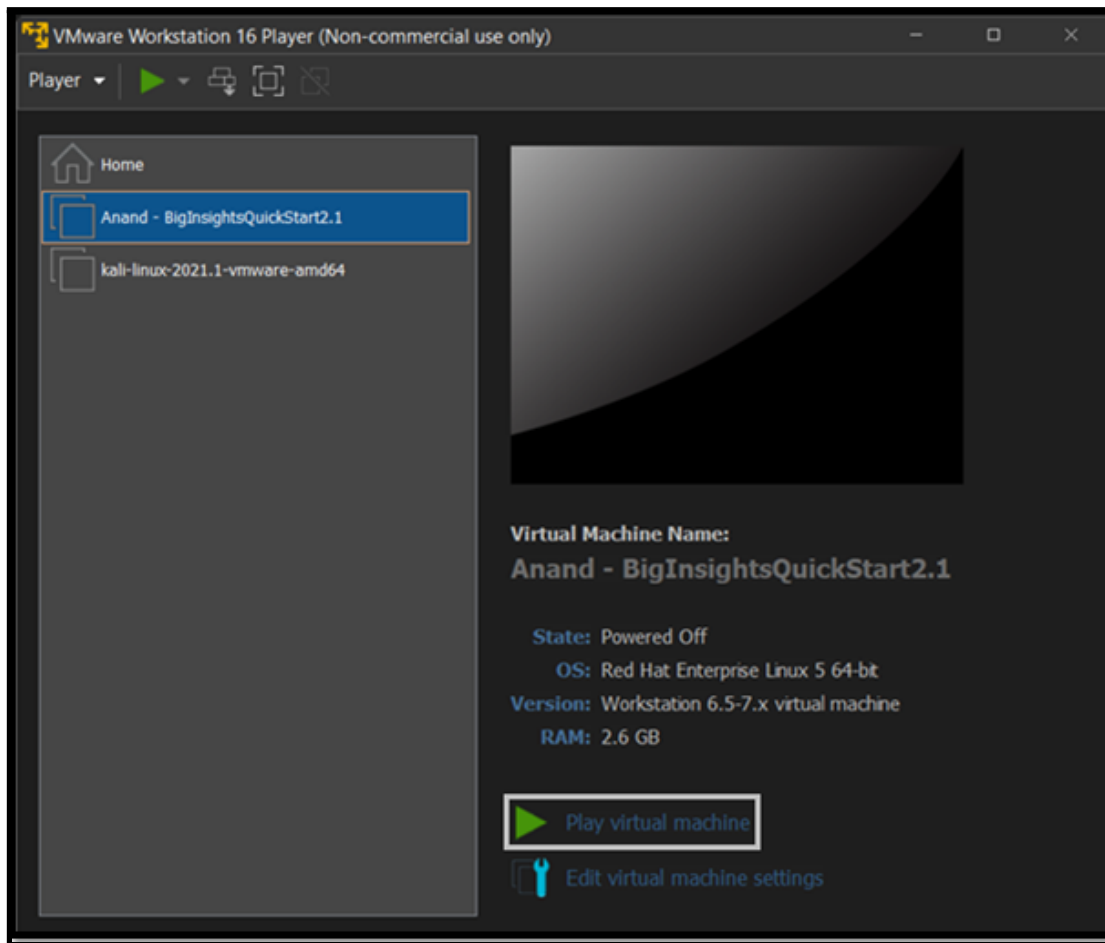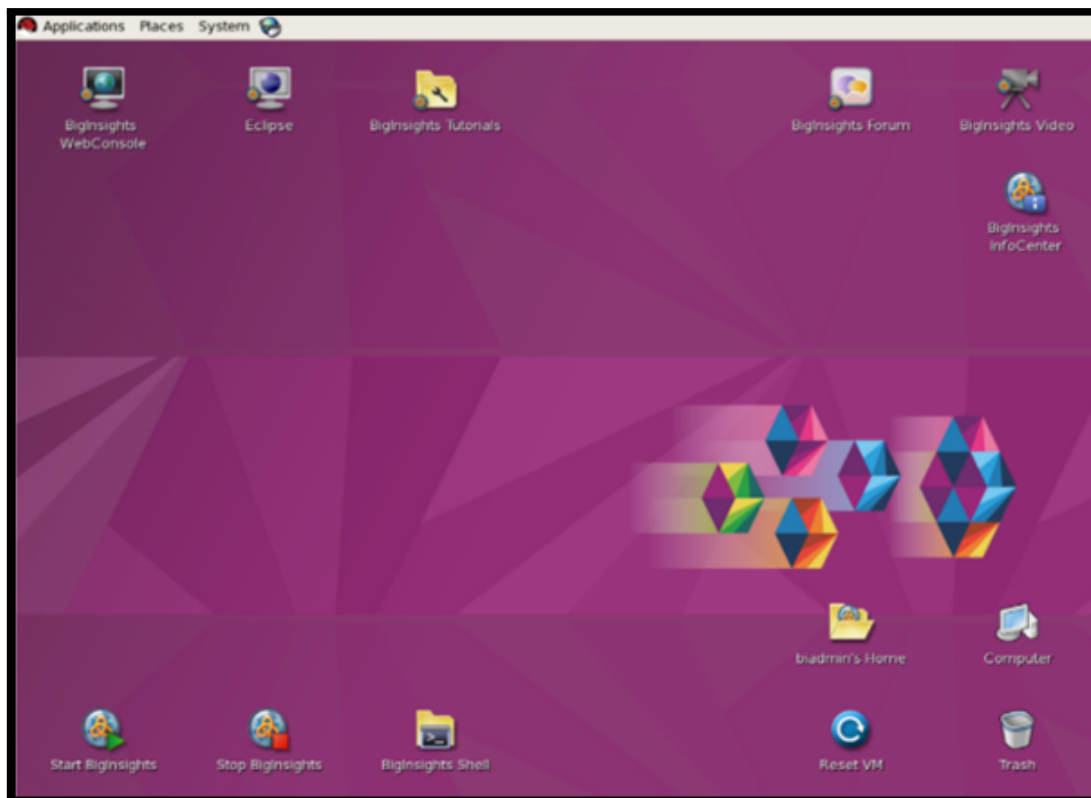

5. Log in to the VMware virtual machine using the following credentials.

Username: biadmin | Password: biadmin

6. After you log in, your screen should look similar to the one below.

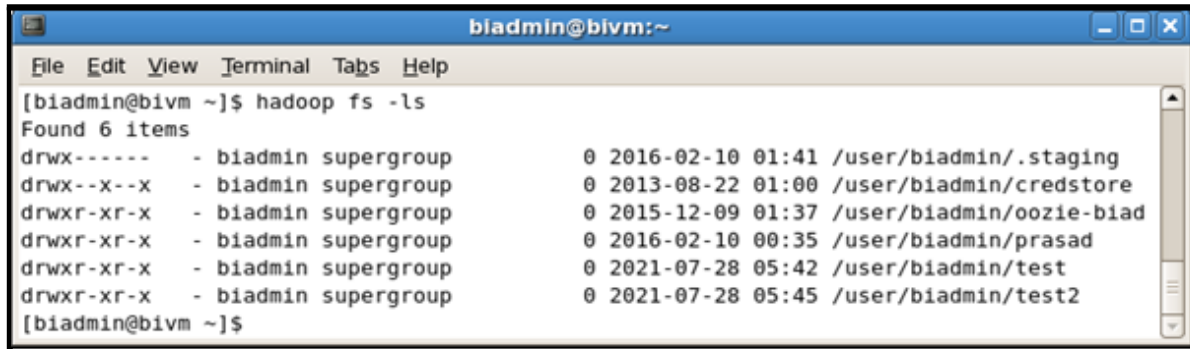7. Double-clicking on the Start BigInsights icon would execute a script that would start all the components.

8. We can stop the components in a similar manner, by double-clicking on the Stop BigInsights icon.

B.2 Input and Output:

Using the command-line interface to explore the Hadoop Distributed File System:

The contents of the root directory are shown.

```
                          biadmin@bivm:~                              _ □ x
 File  Edit  View  Terminal  Tabs  Help
 [biadmin@bivm ~]$ hadoop fs -ls
 Found 6 items
 drwx------    - biadmin supergroup         0 2016-02-10 01:41 /user/biadmin/.staging
 drwx--x--x    - biadmin supergroup         0 2013-08-22 01:00 /user/biadmin/credstore
 drwxr-xr-x    - biadmin supergroup         0 2015-12-09 01:37 /user/biadmin/oozie-biad
 drwxr-xr-x    - biadmin supergroup         0 2016-02-10 00:35 /user/biadmin/prasad
 drwxr-xr-x    - biadmin supergroup         0 2021-07-28 05:42 /user/biadmin/test
 drwxr-xr-x    - biadmin supergroup         0 2021-07-28 05:45 /user/biadmin/test2
 [biadmin@bivm ~]$
```

B.3 Observations and learning:
(Students are expected to comment on the output obtained with clear observations and learning for each task/subpart assigned)

The Hadoop Distributed File System (HDFS) allows users to arrange their data into files and directories. It has a command-line interface known as FS shell that allows users to interact with HDFS data that is available to Hadoop MapReduce applications.

B.4 Conclusion:
(Students must write the conclusion as per the attainment of individual outcome listed above and learning/observation noted in section B.3)

Installed IBM Infosphere BigInsights successfully and explored HDFS on the Hadoop installation using Hadoop a command.

B.5 Question of Curiosity
(To be answered by student based on the practical performed and learning/observations)
Q1: Which are the modes in which Hadoop can run?

Ans:
The three modes in which Hadoop can run are :
1. Standalone mode: This is the default mode. It uses the local FileSystem and a single Java process to run the Hadoop services.
2. Pseudo-distributed mode: This uses a single-node Hadoop deployment to execute all Hadoop services.
3. Fully distributed mode: This uses separate nodes to run Hadoop master and slave services.