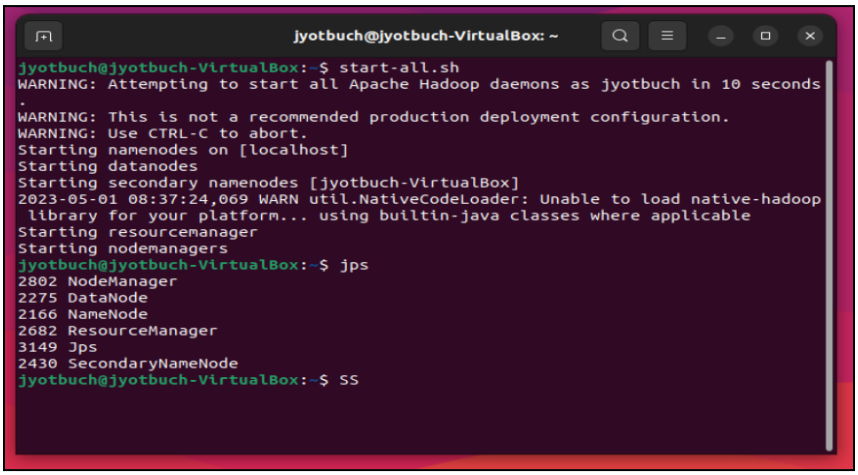
Vishwakarma University, Pune
University Grants Commission (UGC) Approved State Private University

Assignment 1	
Name - Jyot Buch	Roll No. - 09
Date of Submission - 01/05/2023	Class - A (2023 Batch)

1.	List the steps for installing Ubuntu Operating System. Give step-by-step procedure of configuring Apache Hadoop on Ubuntu OS.
	Part 1: Installing Ubuntu Operating System 1. Download the Ubuntu Operating System from the official website https://ubuntu.com/download 2. Create a bootable USB drive with the Ubuntu ISO file. You can use tools like Rufus or Etcher to create a bootable USB drive. 3. Insert the bootable USB drive into your computer and restart the computer. 4. Enter the boot menu of your computer by pressing the appropriate key. The key varies depending on your computer manufacturer. Commonly used keys are F2, F12, or ESC. 5. Select the USB drive from the boot menu and press enter. 6. The Ubuntu installer will start. Select the language and click on the "Install Ubuntu" button. 7. Follow the installation wizard instructions. You will be asked to select the installation type, partition the hard drive, and create a username and password. 8. Once the installation is complete, reboot your computer. Part 2: Configuring Apache Hadoop on Ubuntu OS 1. Install Java on your Ubuntu system by running the following command: 2. sudo apt-get install default-jdk 3. Download Apache Hadoop from the official website https://hadoop.apache.org/releases.html 4. Extract the downloaded Apache Hadoop tarball by running the following command: 5. tar -xvf hadoop-3.3.1.tar.gz 6. Move the extracted Hadoop directory to the /usr/local directory by running the following command: 7. sudo mv hadoop-3.3.1 /usr/local/hadoop 8. Set up the environment variables by adding the following lines to the ~/.bashrc file: export HADOOP_HOME=/usr/local/hadoop export PATH=$PATH:$HADOOP_HOME/bin export PATH=$PATH:$HADOOP_HOME/sbin

	9. Save the changes to the ~/.bashrc file and run the following command to apply the changes: source /etc/profile 10. Edit the core-site.xml file located in the /usr/local/hadoop/etc/hadoop directory and add the following lines: <property> <name>fs.default.name</name> <value>hdfs://localhost:54310</value> </property> 11. Edit the hdfs-site.xml file located in the /usr/local/hadoop/etc/hadoop directory and add the following lines: <property> <name>dfs.replication</name> <value>1</value> </property> 12. Format the Hadoop filesystem by running the following command: hdfs namenode -format 13. Start the Hadoop daemons by running the following command: start-all.sh 14. Verify that Hadoop is running by accessing the Hadoop web interface at http://localhost:50070 
2.	Why is there a need for ssh in Apache Hadoop? How it supports communication between master and slaves of configuration.
	In Apache Hadoop, Secure Shell (SSH) is necessary for secure communication between the master and slave nodes in a distributed computing model. SSH provides secure communication through encryption, secure authentication using public-key cryptography and password authentication, secure file transfer, and remote administration capabilities. The master node uses SSH to log in to the slave nodes and execute the assigned tasks, and the slave nodes use SSH to communicate back to the master node about the task execution status. In summary, SSH plays a crucial role in ensuring a Hadoop cluster's secure and efficient operation.