<u>Hadoop - Installation</u>
❗❗ ONLY_FOR_WINDOWS ❗❗

👉 System Requirements

    ✅ Windows 10 (64 bit)

    ✅ Minimum 8GB RAM

    ✅ Internet connection (approx. 2GB)

    ✅ Virtualization should be enabled 👌

🚫If these conditions are not met exactly then there might be issues in installation🚫

1. Download & install Oracle VirtualBox

   🌍[https://download.virtualbox.org/virtualbox/6.1.6/VirtualBox-6.1.6-137129-Win.exe]

2. Download Bitnami Hadoop Stack

   🌍[https://bitnami.com/redirect/to/995396/bitnami-hadoop-3.2.1-2-linux-debian-9-x86_64.ova]

3. Download & install Putty

   🌍[https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html]

4. Download & install WinSCP

   🌍[https://winscp.net/download/WinSCP-5.17.3-Setup.exe]

5. Import the VM

   a. Launch VirtualBox

   b. Select "File➡ Import Appliance" menu option in VirtualBox and select the *.ova* file downloaded from the Bitnami website. Click "Continue" to proceed.

   c. Click "Import" to proceed.

       ❗ Ensure that the memory is set to at least 6GB ❗

   d. Settings➡Network adapter 1:NAT

   e. ➡Network adapter 2:host-only

6. Start VM

7. Login to the VM using username as *bitnami* and password as *bitnami*

   ❗ *password is not visible while typing don't be afraid it's there* ❗

8. Enable SSH - enter the commands below
   a. `sudo rm -f /etc/ssh/sshd_not_to_be_run`
   b. `sudo systemctl enable ssh`
   c. `sudo systemctl start ssh`
   d. `sudo apt install openssh-server`
   e. `sudo systemctl status ssh`
      ❗ *After this status should be* ✅ *active* ❗

```
lo: flags=73<UP,LOOPBACK,RUNNING>  mtu 65536
        inet 127.0.0.1  netmask 255.0.0.0
        inet6 ::1  prefixlen 128  scopeid 0x10<host>
        loop  txqueuelen 1  (Local Loopback)
        RX packets 5246  bytes 889539 (868.6 KiB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 5246  bytes 889539 (868.6 KiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

bitnami@debian:~$ sudo systemctl status ssh
 ssh.service - OpenBSD Secure Shell server
   Loaded: loaded (/lib/systemd/system/ssh.service; enabled; vendor preset: enab
   Active: active (running) since Thu 2020-04-16 20:54:43 UTC; 5h 3min left
  Process: 375 ExecStartPre=/usr/sbin/sshd -t (code=exited, status=0/SUCCESS)
 Main PID: 388 (sshd)
    Tasks: 1 (limit: 4915)
   CGroup: /system.slice/ssh.service
           └─388 /usr/sbin/sshd -D

Apr 16 20:54:43 debian systemd[1]: Starting OpenBSD Secure Shell server...
Apr 16 20:54:43 debian sshd[388]: Server listening on 0.0.0.0 port 22.
Apr 16 20:54:43 debian sshd[388]: Server listening on :: port 22.
Apr 16 20:54:43 debian systemd[1]: Started OpenBSD Secure Shell server.
lines 1-13/13 (END)
```

   f. `sudo ufw allow ssh`

🌐 Refer this site for further info [https://linuxize.com/post/how-to-enable-ssh-on-ubuntu-18-04/]

9. Run the following command to check the IP address of the VM

    `sudo ifconfig`

    ❗ Note down the **inet** IP. It will be something like **192.168.x.x** ❗

```
      RX errors 0  dropped 0  overruns 0  frame 0
      TX packets 88  bytes 9200 (8.9 KiB)
      TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0
      device interrupt 19  base 0xd000

enp0s8: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
      inet 192.168.56.101  netmask 255.255.255.0  broadcast 192.168.56.255
      inet6 fe80::a00:27ff:fe81:2fd9  prefixlen 64  scopeid 0x20<link>
      ether 08:00:27:81:2f:d9  txqueuelen 1000  (Ethernet)
      RX packets 19  bytes 4261 (4.1 KiB)
      RX errors 0  dropped 0  overruns 0  frame 0
      TX packets 30  bytes 3864 (3.7 KiB)
      TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0
      device interrupt 16  base 0xd240

lo: flags=73<UP,LOOPBACK,RUNNING>  mtu 65536
      inet 127.0.0.1  netmask 255.0.0.0
      inet6 ::1  prefixlen 128  scopeid 0x10<host>
      loop  txqueuelen 1  (Local Loopback)
      RX packets 5246  bytes 889539 (868.6 KiB)
      RX errors 0  dropped 0  overruns 0  frame 0
      TX packets 5246  bytes 889539 (868.6 KiB)
      TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

bitnami@debian:~$
```

    ❗❗❗ If mouse pointer is stuck inside VM press **right ctrl** ❗❗❗

10. Open Chrome
    a. Type the IP➡Hit enter
    b. You'll be prompted to type Username & Password
    c. Go to the VM and type **cat ./bitnami_credentials**
    d. Note 'em and type them in the chrome tab [Yes, I know the pass is irritating 😡]
    e. You'll see a cute 🐘 (Yellow Jumbo) which implies you're good to proceed to the next step.
11. Create a **.txt** file on your PC ➡name it accordingly➡put a ton of 💩 inside it➡save it.
12. Add the file to the VM
    a. Open WinSCP
    b. Select protocol as **SFTP**
    c. Enter the inet IP address noted above in the Host Name
    d. Enter username and password as **bitnami**
    e. Click Login
    f. The left pane is the VM's file directory and right one is your file directory
    g. In the VM's directory navigate to **/home/bitnami** and paste the .txt you created in the above step
    h. Close WinSCP (you can even uninstall it if you want 😂, JK don't uninstall it)

13. SSH to the VM
    a. Open PuTTY
    b. Enter the inet IP address noted above in the Host Name
    c. Connection type: SSH
    d. Click Open
14. Finally 🏆 run these commands in Putty's terminal
    a. `hadoop fs -rmr /tmp/hdfs-example-output`
    b. `hadoop fs -put -f /home/bitnami/input_file.txt /tmp/hdfs-example-input`
    c. `hadoop jar /opt/bitnami/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar grep /tmp/hdfs-example-input /tmp/hdfs-example-output '[a-zA-Z0-9]+'`
    d. `hadoop fs -cat /tmp/hdfs-example-output/part-r-00000`
15. Take 📸 of the output. Create doc and submit.

🔥**This might look easy for you but it had a lot of hard work and time(almost 2 days & nights)involved,So...cheer up and work your ass off.**🔥

#StaySafe!!!...#Go Corona, Corona Go