

Installation of Hadoop on Windows

A.K. Chela^{1*}, H.S. Sidhu²

¹Computer Science & Engineering, Desh Bhagat University, Mandi Gobindgarh, India

²Computer Science, Desh Bhagat University, Mandi Gobindgarh, India

Corresponding Author: preetk117@gmail.com,

Available online at: www.ijcseonline.org

Accepted: 15/Aug/2018, Published: 31/Aug/2018

Abstract— Technology brings a rapid growth of data. It becomes a challenge to process data with previous methods and applications. The unstructured, scalable, complex data is available through distinct sources like social sites, news, applications and so on. Hadoop is a most demanding development tool for big data because it processes the data in the structured and unstructured form. When you think about Hadoop usually it comes into mind regarding Linux. It is not mandatory for all to learn Linux first and then install Hadoop on a system and then use it. There is an advantage of Linux that it provides basic production and development platform for Hadoop. This paper reflects the light on the variety of platforms, open source systems available for Hadoop installation. Now Hadoop is not only limited to the Linux platform, you can put it on Windows too. Windows is user-friendly and you can install Hadoop without Linux on another hand there are enormous resources by which you can easily go into the Linux environment and place Hadoop. Java is a prerequisite of Hadoop.

Keywords—Powershell, Hyper-V, Dual Booting, Virtual Machine, Cygwin

I. INTRODUCTION

The Twenty-First century is an age of social media. There is no limitation on time, amount of data to access and to represent your views in the form of comments, likes/dislikes or through emoji on different applications like Facebook, Instagram and so on. Hadoop is a framework written in Java for running applications on large clusters. It provides high process rate to applications data. Whenever we talk about Hadoop, it means we are going to discuss something about Linux. Presently, it is not limited to Linux only. We can install it on a variety of operating systems such as Windows, Mac, Linux. [2] Normally, Linux developers like to run windows operating system on Linux as well as Windows developers like to do vice versa.

The aim of this paper to describe the ways through which you can setup single node cluster on windows. Cygwin, VirtualBox, Hyper- V, Powershell. Installation of Hadoop is easy on Linux and Mac but for Windows it's a delicate process.

II. METHODOLOGY

1. Cygwin [2]

Cygwin is a windows library that provides a graphical environment and a command line interface. Cygwin gave us a Linux like environment. It allows users to have

access on various standard UNIX utilities. They can be used bash from one of the provided shells. It is used by both Windows friendly and Unix friendly users.

Run the setup.exe file to install Cygwin and follow the set of instructions for default settings, otherwise, you should customize it and select the package of the particular tool. Windows users are able to write console with the use of text editor such as nano.

2. PowerShell

PowerShell is a task-based command-line shell and scripting language; it is designed specifically for system administrators and power-users, to rapidly automate the administration of multiple operating systems (Linux, macOS, Unix, and Windows) and the processes related to the applications that run on those operating systems.

3. Virtual-Machine [2]

You can run any operating system on the Windows operating system with the help of VirtualBox or VMware Player. When we like to boot Linux than firstly boot windows because Ubuntu has installed in VirtualBox. You can even run Linux applications on your desktop by trying VMware's unity mode and seamless mode for Virtual Box.

4. Hyper-V

To create a Virtual Machine with Unix/ Fedora with Windows 10, you can use Hyper-V. If you are working on Windows 10, then enable Hyper-V feature to create a

Virtual Machine to install a Linux distribution package. Following are the footsteps to install Linux distributions on virtual machines by deploying Hyper-V virtualization.

Basic Requirements:

- An operating system with support for Hyper-V.
- An ISO file to install your proposed Linux distribution.

Hyper-V support

Hyper-V is a new feature of Windows 10, there is a minimum hardware requirement a computer must need to enable it.

- Second Level Address Translation (SLAT) with 64-bit CPU.
- VM Monitor Mode Extension (VT-c on Intel chips) must be supported by the processor.
- RAM requires- Minimum 4GB.

Check your system BIOS to confirm the Virtualization Technology and Hardware Enforced Data Execution Prevention options are enabled.

Quickly justify your computer has the required support for Hyper-V.

1. Click **Start**.
2. Open the **Command Prompt**.
3. Enter the following command and press **Enter**:
systeminfo.exe (figure 1)
4. Under **Hyper-V Requirements**, if the result is **Yes**, then you are able to run Hyper-V.

Each of four Hyper-V Requirements displays yes means that VM Monitor mode Extension is there, Virtualization is enabled in Firmware, Address Translation. Data Execution Prevention is available. Hardware supports all the feature, and there is no need to enable them by your own.

Enable Hyper-V

Following are the steps to enable Hyper-V on Windows 10:

1. Type control panel in *type here to search*, Open Control Panel.
2. Click on Programs.
3. Select Turn Windows features on or off. As shown in (figure 2)
4. Confirm the Hyper-V option making sure: Hyper-V Management Tools and Hyper-V Platform are also checked.
5. Click OK to start the process. (figure 3)
6. After the installation completes, click **Restart now** to apply the changes. (figure4)

Configure a virtual switch on Hyper-V

1) Create a virtual switch

Method to enable Hyper-V on Windows 10

Hyper-V allows you to create one or more virtual machines to install and run distinct Operating Systems on the same hardware. This virtualization technology is available from Microsoft on Windows 10 Pro, Enterprise, and Education.

Create a virtual switch on Hyper-V by using the following measures:

- a) Click **Start**.
- b) Search for **Hyper-V Manager** and click the top result.
- c) Click on the **Action** menu.
- d) Select **New** and click on **Virtual Switch Manager**.
- e) On the left pane, select **New virtual network switch**.
- f) On the right pane, select **External**.
- g) Click on **Create Virtual Switch** button.
- h) Enter a new desired name for the switch.
- i) In connection type, select **External network** under your network adapter.
- j) Click **Apply**.
- k) Click **OK**.

2) Virtual machine

Next step is to create a new virtual machine after creating the virtual switch, you can move forward to:

- a) Click **Start**.
- b) Search for **Hyper-V Manager** and open it.
- c) Click on the **Action** menu.
- d) Select **New** and click on **Virtual Machine**.
- e) Click the **Next** button.
- f) Type the desired name for your virtual machine.
- g) Use the default location to store your virtual machine or check the **Store the virtual machine in a different location** option to select a different path.
- h) Click **Next** to continue.
- i) Leave the selected default option **Generation 1** or you can select option **Generation 2** if you require a UEFI-based firmware.
- j) Click **Next**.
- k) For Ubuntu, you require a minimum amount of 2GB memory. Select the amount of RAM to allocate virtual machine.
- l) You can allocate as much memory as you want, but this setting will always depend on the physical memory available on your computer.
- m) Click **Next**.
- n) Select the virtual switch from the drop-down menu you created earlier.
- o) Click **Next**.
- p) Stay all the default settings as it is to create a virtual hard drive, allocate at least 25 GB storage, which is the lowest requirement to run Ubuntu.
- q) Click **Next** to continue.

- r) From Installation options, select **Install an operating system from a bootable CD/DVD-ROM** option.
- s) Choose the **Image file (.iso)** option.
- t) Choose the path for the ISO file with the Ubuntu installation files.
- u) Click **Next**.
- v) Click **Finish**.
- w) The last step is to build the virtual machine and install the Linux distribution you require.

Installation of Hadoop on Windows without Linux

The basic requirements are:

5. Latest JDK
6. Latest Hadoop package

```

Network Card(s):
[05]: K84338819
4 NIC(s) Installed.
[01]: Intel(R) Dual Band Wireless-AC 3160
Connection Name: Wi-Fi
DHCP Enabled: Yes
DHCP Server: 192.168.1.1
IP address(es)
[01]: 192.168.1.79
[02]: fe80::a835:1f6c:39f:a35a
[02]: Realtek PCIe GBE Family Controller
Connection Name: Ethernet
Status: Media disconnected
[03]: Bluetooth Device (Personal Area Network)
Connection Name: Bluetooth Network Connection
Status: Media disconnected
[04]: VirtualBox Host-Only Ethernet Adapter
Connection Name: VirtualBox Host-Only Network
DHCP Enabled: No
IP address(es)
[01]: 192.168.56.1
[02]: fe80::bc:9566:7c91:7020

Hyper-V Requirements:
VM Monitor Mode Extensions: Yes
Virtualization Enabled In Firmware: Yes
Second Level Address Translation: Yes
Data Execution Prevention Available: Yes

```

Figure 1

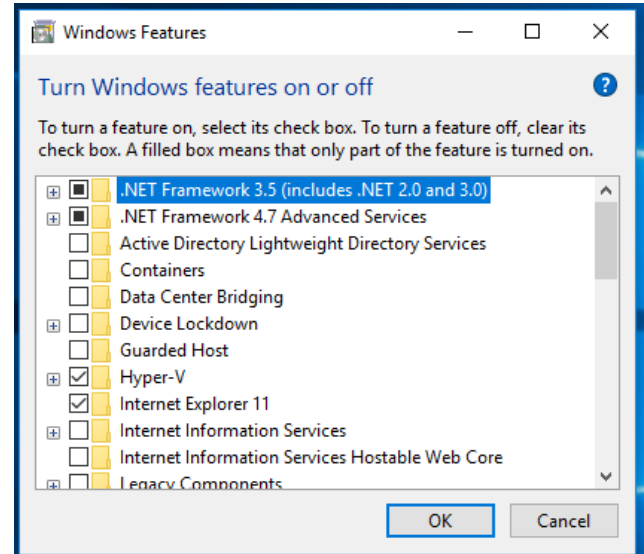


Figure 3

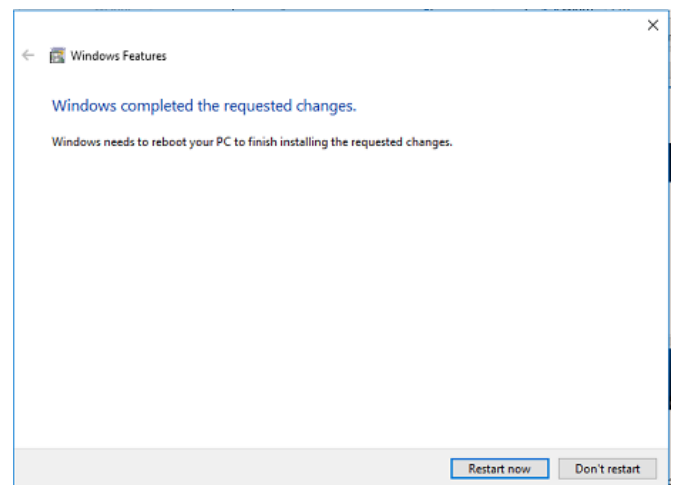


Figure 4

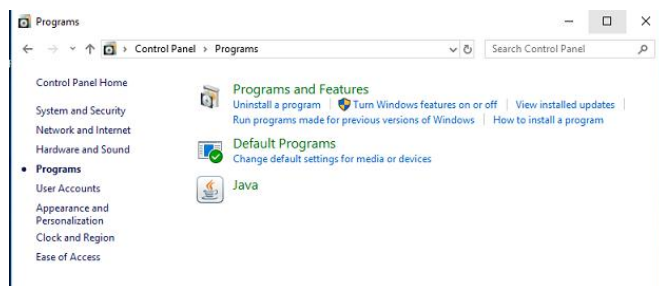


Figure 2

Download the JDK from Oracle website. It depends on the version of your system, either it is 86-bit or 64 bit and operating system (Linux, Windows, Mac, Solaris).

For Windows you will download 64-bit version (depends on your system). Accept the license agreement and download it into your system.

Second download Hadoop software from hadoop.apache.org

Click on releases and select latest version from various available or select which you require. But make sure select the binary package instead of the source package. Choose binary package because it has pre-built libraries you don't

need to compile all which takes a lot of time. Extract the Hadoop folder.

Start installation.

- 1) Install JDK package. Install it into default location or you can change the location.
- 2) Make sure change the destination folder location to save it in a different folder.
- 3) Just add *jre* folder under *jdk* folder in Java
- 4) Installation of Java is complete

Add Java home to the environment variables

1. Open Windows Settings.
2. Go to environment variables.
3. Add a new user variable called *Java_home* as variable name and set variable value as a location where Java stored.
4. Set the path of environment variable up to *java bin*.

Open the command Prompt

To check java installed properly use a command

```
java -version
```

Install Hadoop

1. Copy the extracted folder of Hadoop into the directory where Java has installed.
2. Copy the path. Open the windows settings and then environment variables.
3. Add a user variable named *Hadoop_home* as variable name and paste the Hadoop path as variable value.
4. Set the path of environment variable up to Hadoop bin.
5. Add one more location in path for *sbin* folder, It contains all the executables for Linux and command files for Windows.
6. Close the Windows settings.

Configuration Settings

1. Edit file *G:\hadoop-2.7.6\etc\hadoop\core-site.xml*, write the xml paragraph as shown below and save it.

```
<configuration>
<property>
  <name>fs.defaultFS</name>
  <value>hdfs://localhost:9000</value>
</property>
</configuration>
```

2. Rename "*mapred-site.xml.template*" to "*mapred-site.xml*" and edit this file *G:\hadoop-2.7.6\etc\hadoop\mapred-site.xml*, write the xml paragraph as shown below and save it.

```
<configuration>
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
```

```
</configuration>
```

3. Create "data" folder below *G:\hadoop-2.7.6*
 - Create folder "*datanode*" under "*G:\hadoop-2.7.6\data*"
 - Create folder "*namenode*" under "*G:\hadoop-2.7.6\data*"
4. Edit file *G:\hadoop-2.7.6\etc\hadoop\hdfs-site.xml*, write the xml paragraph as shown below and save it.

```
<configuration>
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>

<property>
  <name>dfs.namenode.name.dir</name>
  <value>G:\hadoop-2.7.6\data\namenode</value>
</property>
```

```
<property>
  <name>dfs.datanode.data.dir</name>
  <value>G:\hadoop-2.7.6\data\datanode</value>
</property>
</configuration>
```

5. Edit file *G:\hadoop-2.7.6\etc\hadoop\yarn-site.xml*, write the xml paragraph as shown below and save it.

```
<configuration>
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>

<property>
  <name>yarn.nodemanager.aux-
services.mapreduce.shuffle.class</name>
  <value>org.apache.hadoop.mapred.ShuffleHandler
</value>
</property>
</configuration>
```

6. Edit file *G:\hadoop-2.7.6\etc\hadoop\hadoop-env.cmd*, by closing the command prompt "Java_Home=%Java_Home%" instead of "set JAVA_HOME=G:\Java\jdk1.8.0_171".

Hadoop Configuration

1. Download file Hadoop Configuration.zip
2. Delete bin file from *G:\hadoop-2.7.6* and replace it with bin file that you just download *Hadoop.configuration.zip*
3. Open command prompt and type command "*hdfs namenode -format*".

Testing

1. Run command prompt as an administrator and change root as "G:\Hadoop-2.7.6\sbin" usually the root is "C:." and type "strat-all.cmd" to run apache.
2. Confirm all applications are running:
 - Hadoop Namenode
 - Hadoop Datanode
 - Yarn Resourcemanager
 - Yarn Nodemanager
3. Open <http://localhost:8088>
It will display all running applications of Hadoop.
4. Open <http://localhost:50070>
It will display the dfshealth.html file. Overview, summary, directory, memory information and so on.

Finally, Hadoop has been installed.

III. CONCLUSION

Linux plays a vital role in Hadoop installation because it provides a development and production platform whereas Windows provides development platform. The Linux terminal is superior to use over the Windows command line for developers. Linux is a best for multi-node clustering. You can easily add a new cluster on Linux. But when you install Linux on Windows it makes the system slow because it puts a double load on memory and effects the speed. The best solution is to make your system dual boot, with this each operating system has its own RAM to run a cluster.

REFERENCES

- [1] <https://www.cio.com/article/3195905/windows/18-things-you-should-know-about-using-linux-tools-in-windows-10.html>
- [2] Shantanu Sharma "Installing Hadoop-2.6.x on Windows 10" University of California, Irvin.
- [3] "Introduction to Hyper-V on Windows 10" Microsoft Documents, United States Feb 05,2016.
- [4] S. Prasad "Hadoop Installation" https://www.it.iitb.ac.in/frg/wiki/images/a/af/03D05011_Sandeep_Prasad_Group_7_Week_3_Report_1_Hadoop_Installation.pdf
- [5] N.S. Hoe, C. Charles "User Guide to using the Linux Desktop", UNDP-APDIP, Kuala Lumpur, Malaysia,2004
- [6] Y. Bassil," Windows and Linux Operating System from a Security Perspective", Journal of Global Research in Computer Science.
- [7] N.Economides, E.Katsamakas,"Linux vs. Windows: A comparison of Applications and platform Innovation Incentives for Open Source and Proprietary Software Platform "Elsevier pp.207-218, 2006.
- [8] K. Parimala G. Rajkumar , A. Ruba , S. Vijayalakshmi , "Challenges and Opportunities with Big Data ", International Journal of Scientific Research in Computer Science and Engineering. pp:16-20, 2017
- [9] A.Tanuja, D. Swetha Swatha, "Processing and Analyzing Big Data using Hadoop" International Journal off Computer Science & Engineering, pp: 91-94, April,2016.

Author's Profile

A.K. Chela is a Research Scholar at Desh Bhagat University, Mandi Gobindgarh. She is a student of Ph.D. in Computer Science & Engineering department at Desh Bhagat University.



H.S. Sidhu is an Assistant Professor at Desh Bhagat University, Mandi Gobindgarh. He has done his Ph.D.in Computer Science from Desh Bhagat University, Mandi Gobindgarh.

