Install Hadoop 3.3.0 on Windows 10 Step by Step Guide

Raymond 36,228 29 2020-08-01 2 years ago

This detailed step-by-step guide shows you how to install the latest Hadoop v3.3.0 on Windows 10. It leverages Hadoop 3.3.0 winutils tool. WLS (Windows Subsystem for Linux) is not required. This version was released on July 14 2020. It is the first release of Apache Hadoop 3.3 line. There are significant changes compared with Hadoop 3.2.0, such as Java 11 runtime support, protobuf upgrade to 3.7.1, scheduling of opportunistic containers, non-volatile SCM support in HDFS cache directives, etc.

Please follow all the instructions carefully. Once you complete the steps, you will have a shiny pseudo-distributed single node Hadoop to work with.

References

Refer to the following articles if you prefer to install other versions of Hadoop or if you want to configure a multi-node cluster or using WSL.

- Install Hadoop 3.3.0 on Windows 10 using WSL (Windows Subsystems for Linux is requried)
- Install Hadoop 3.0.0 on Windows (Single Node)
- Configure Hadoop 3.1.0 in a Multi Node Cluster
- Install Hadoop 3.2.0 on Windows 10 using Windows Subsystem for Linux (WSL)

Required tools

Before you start, make sure you have these following tools enabled in Windows 10.

Tool	Comments

PowerShell	We will use this tool to download package.
	
	In my system, PowerShell version table is listed below:
	
	`$PSversionTable`
	
	`Name Value`
	`---- -----`
	`PSVersion 5.1.19041.1`
	`PSEdition Desktop`
	`PSCompatibleVersions {1.0, 2.0, 3.0, 4.0...}`
	`BuildVersion 10.0.19041.1`
	`CLRVersion 4.0.30319.42000`
	`WSManStackVersion 3.0`
	`PSRemotingProtocolVersion 2.3`
	`SerializationVersion 1.1.0.1`

Git Bash or 7 Zip	We will use Git Bash or 7 Zip to unzip Hadoop binary package. You can choose to install either tool or any other tool as long as it can unzip *.tar.gz files on Windows.
Command Prompt	We will use it to start Hadoop daemons and run some commands as part of the installation process.
Java JDK	JDK is required to run Hadoop as the framework is built using Java. In my system, my JDK version is jdk1.8.0_161. Check out the supported JDK version on the following page. https://cwiki.apache.org/confluence/display/HADOOP/Hadoop+Java+Versions From Hadoop 3.3.0, Java 11 runtime is now supported.

Now we will start the installation process.

Step 1 - Download Hadoop binary package

Select download mirror link

Go to download page of the official website:

Apache Download Mirrors - Hadoop 3.3.0

And then choose one of the mirror link. The page lists the mirrors closest to you based on your location. For me, I am choosing the following mirror link:

http://apache.mirror.amaze.com.au/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz

> ⓘ In the following sections, this URL will be used to download the package. Your URL might be different from mine and you can replace the link accordingly.

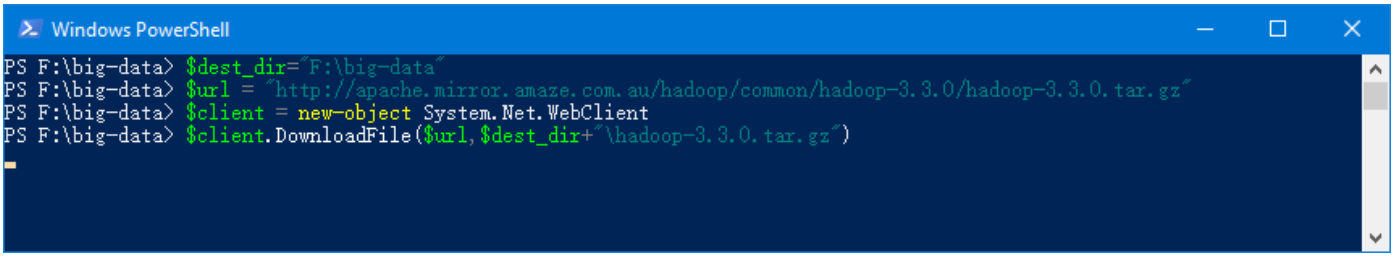
Download the package

> ⓘ In this guide, I am installing Hadoop in folder big-data of my F drive (F:\big-data). If you prefer to install on another drive, please remember to change the path accordingly in the following command lines. This directory is also called destination directory in the following sections.

Open PowerShell and then run the following command lines one by one:

```
$dest_dir="F:\big-data"
$url = "http://apache.mirror.amaze.com.au/hadoop/common/hadoop-3.3.0/hadoop-
3.3.0.tar.gz"
$client = new-object System.Net.WebClient
$client.DownloadFile($url,$dest_dir+"\hadoop-3.3.0.tar.gz")
```

It may take a few minutes to download.



Once the download completes, you can verify it:

```
PS F:\big-data> cd $dest_dir
PS F:\big-data> ls


    Directory: F:\big-data


Mode                 LastWriteTime         Length Name
----                 -------------         ------ ----
-a----         2/08/2020   1:55 AM      500749234 hadoop-3.3.0.tar.gz
```

You can also directly download the package through your web browser and save it to the destination directory.

> ⚠ Please keep this PowerShell window open as we will use some variables in this session in the following steps. If you already closed it, it is okay, just remember to reinitialise the above variables: $client, $dest_dir.

Step 2 - Unpack the package

Now we need to unpack the downloaded package using GUI tool (like 7 Zip) or command line. For me, I will use git bash to unpack it.

Open git bash and change the directory to the destination folder:

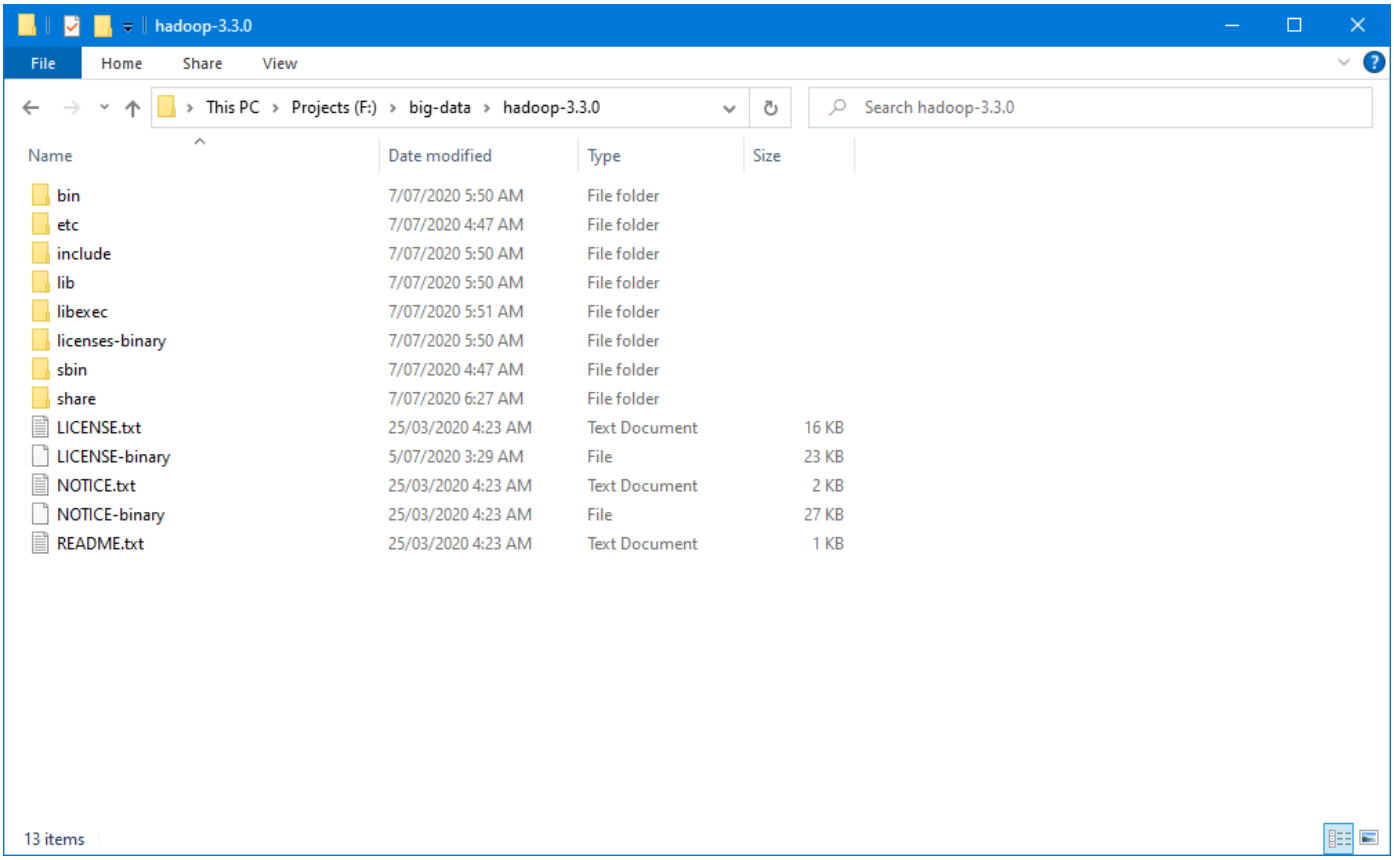
```
cd $dest_dir
```

And then run the following command to unzip:

```
tar -xvzf  hadoop-3.3.0.tar.gz
```

The command will take quite a few minutes as there are numerous files included and the latest version introduced many new features.

After the unzip command is completed, a new folder hadoop-3.3.0 is created under the destination folder.



When running the command you may experience errors like the following:

```
tar: Exiting with failure status due to previous errors
```

Please ignore it for now.

Step 3 - Install Hadoop native IO binary

Hadoop on Linux includes optional Native IO support. However Native IO is mandatory on Windows and without it you will not be able to get your installation working. The Windows native IO libraries are not included as part of Apache Hadoop release. Thus we need to build and install it.

The following repository already pre-built Hadoop Windows native libraries:
https://github.com/kontext-tech/winutils

These libraries are not signed and there is no guarantee that it is 100% safe. We use it purely for test&learn purpose.

Download all the files in the following location and save them to the bin folder under Hadoop folder. For my environment, the full path is: F:\big-data\hadoop-3.3.0\bin. Remember to change it

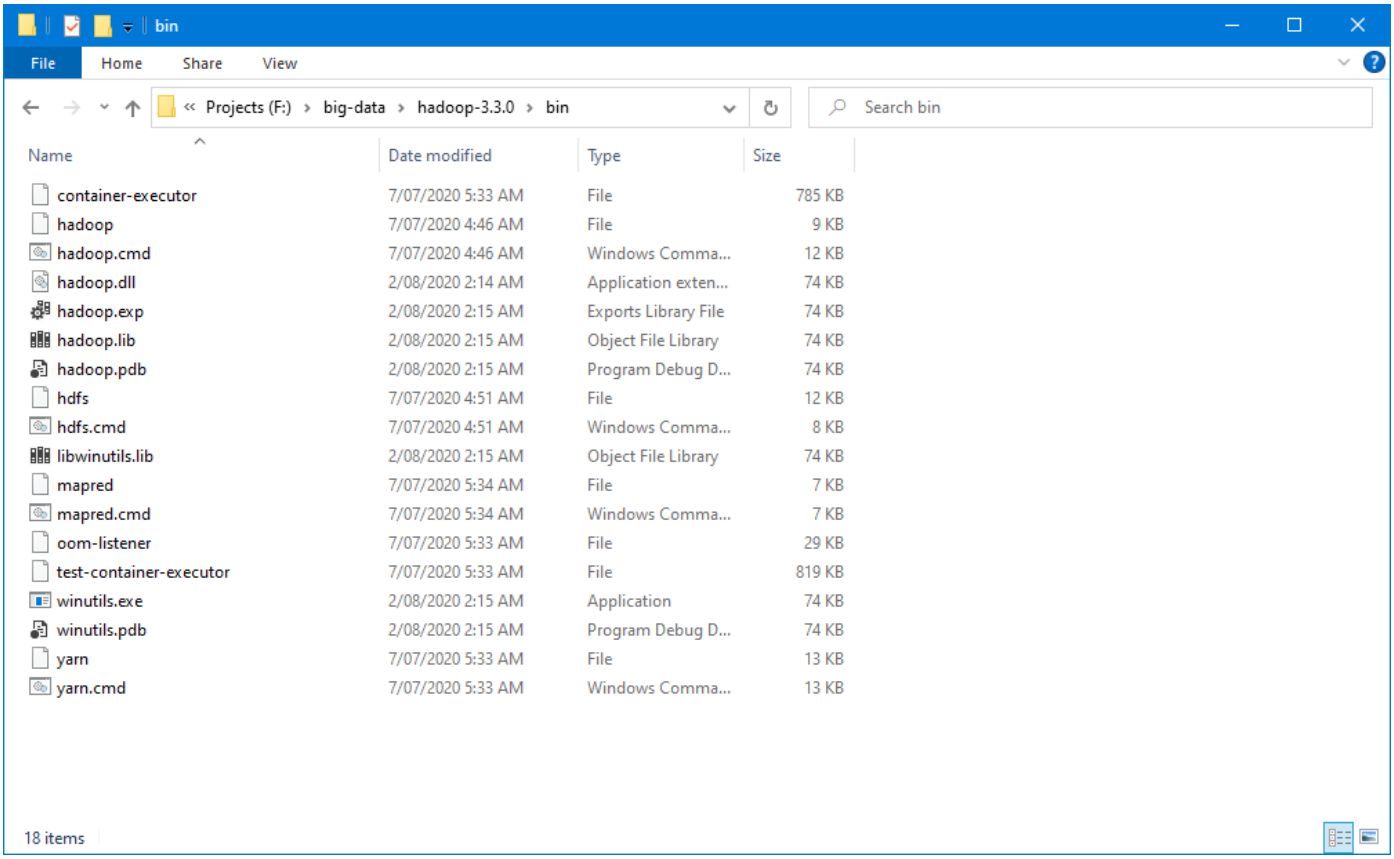
to your own path accordingly.

https://github.com/kontext-tech/winutils/tree/master/hadoop-3.3.0/bin

Alternatively, you can run the following commands in the previous PowerShell window to download:

```
$client.DownloadFile("https://github.com/kontext-tech/winutils/tree/master/hadoop-3.3.0/bin/hadoop.dll",$dest_dir+"\hadoop-3.3.0\bin\"+"hadoop.dll")
$client.DownloadFile("https://github.com/kontext-tech/winutils/tree/master/hadoop-3.3.0/bin/hadoop.exp",$dest_dir+"\hadoop-3.3.0\bin\"+"hadoop.exp")
$client.DownloadFile("https://github.com/kontext-tech/winutils/tree/master/hadoop-3.3.0/bin/hadoop.lib",$dest_dir+"\hadoop-3.3.0\bin\"+"hadoop.lib")
$client.DownloadFile("https://github.com/kontext-tech/winutils/tree/master/hadoop-3.3.0/bin/hadoop.pdb",$dest_dir+"\hadoop-3.3.0\bin\"+"hadoop.pdb")
$client.DownloadFile("https://github.com/kontext-tech/winutils/tree/master/hadoop-3.3.0/bin/libwinutils.lib",$dest_dir+"\hadoop-3.3.0\bin\"+"libwinutils.lib")
$client.DownloadFile("https://github.com/kontext-tech/winutils/tree/master/hadoop-3.3.0/bin/winutils.exe",$dest_dir+"\hadoop-3.3.0\bin\"+"winutils.exe")
$client.DownloadFile("https://github.com/kontext-tech/winutils/tree/master/hadoop-3.3.0/bin/winutils.pdb",$dest_dir+"\hadoop-3.3.0\bin\"+"winutils.pdb")
```

After this, the bin folder looks like the following:



Step 4 - (Optional) Java JDK installation

Java JDK is required to run Hadoop. If you have not installed Java JDK, please install it.

You can install JDK 8 from the following page:

https://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html

Once you complete the installation, please run the following command in PowerShell or Git Bash to verify:

```
$ java -version
java version "1.8.0_161"
Java(TM) SE Runtime Environment (build 1.8.0_161-b12)
Java HotSpot(TM) 64-Bit Server VM (build 25.161-b12, mixed mode)
```

If you got error about 'cannot find java command or executable'. Don't worry we will resolve this in the following step.

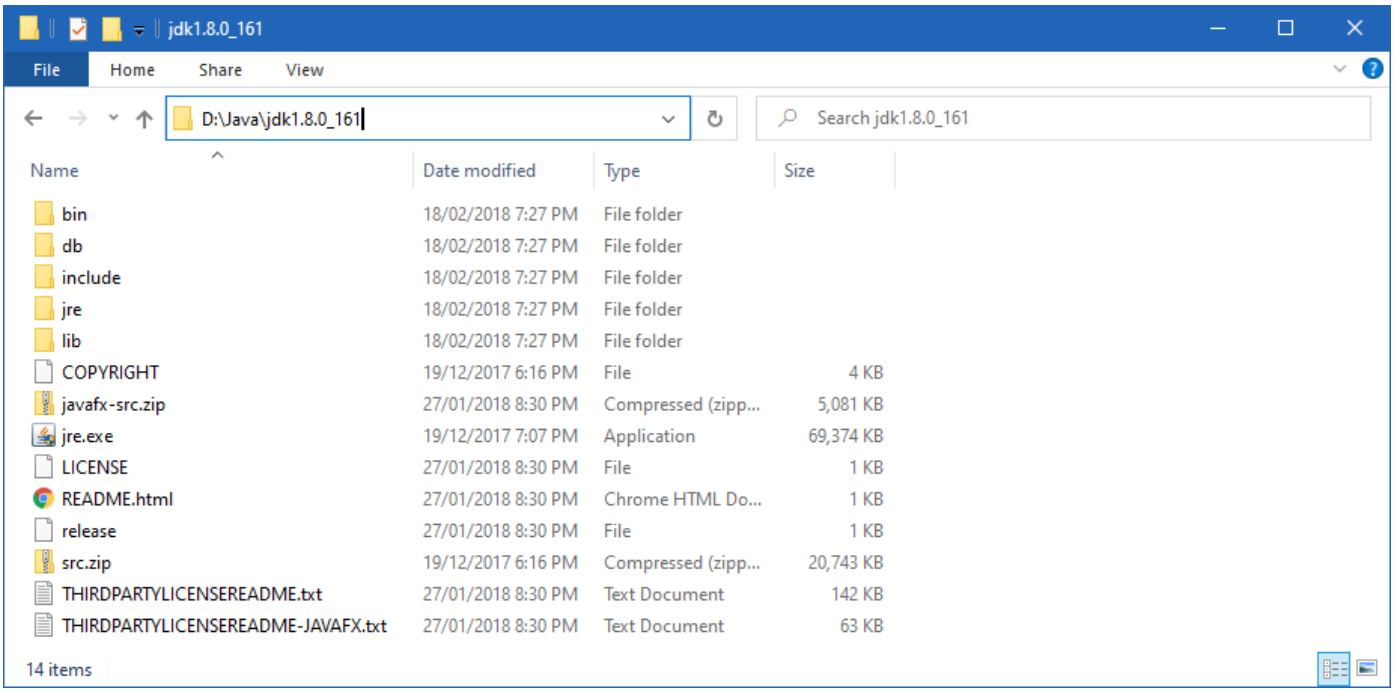
Step 5 - Configure environment variables

Now we've downloaded and unpacked all the artefacts we need to configure two important environment variables.

Configure JAVA_HOME environment variable

As mentioned earlier, Hadoop requires Java and we need to configure JAVA_HOME environment variable (though it is not mandatory but I recommend it).

First, we need to find out the location of Java SDK. In my system, the path is: D:\Java\jdk1.8.0_161.



Your location can be different depends on where you install your JDK.

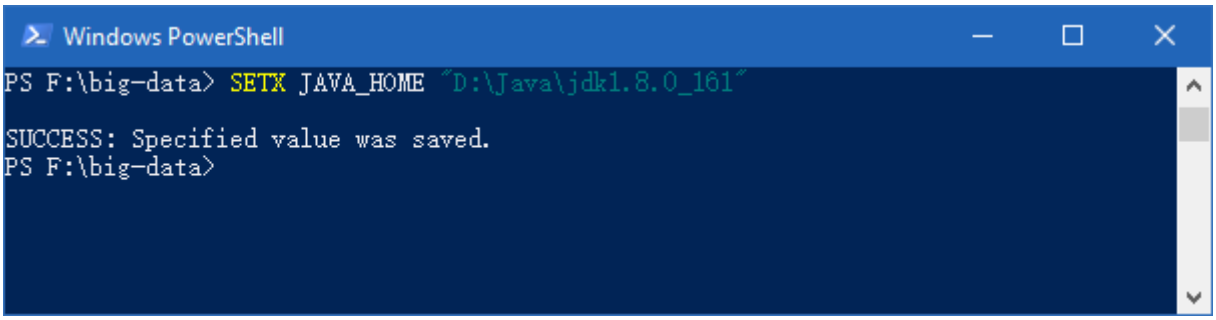
And then run the following command in the previous PowerShell window:

```
SETX JAVA_HOME "D:\Java\jdk1.8.0_161"
```

Remember to quote the path especially if you have spaces in your JDK path.

You can setup environment variable at system level by adding option /M however just in case you don't have access to change system variables, you can just set it up at user level.

The output looks like the following:



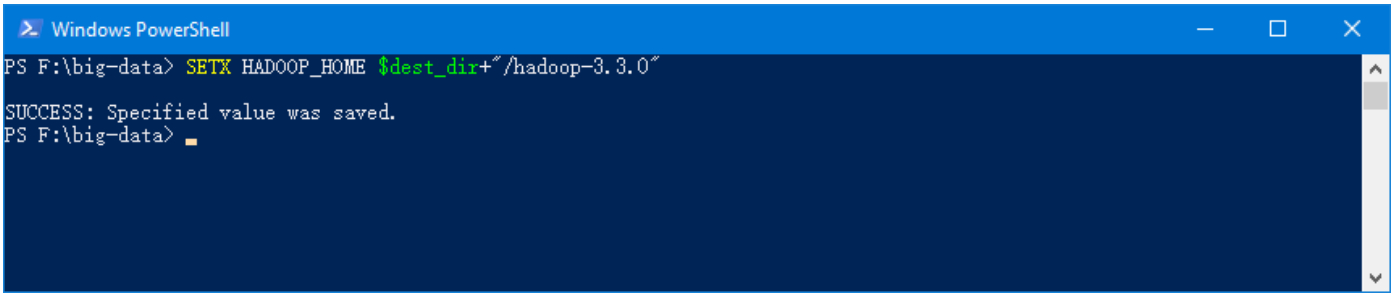
Configure HADOOP_HOME environment variable

Similarly we need to create a new environment variable for HADOOP_HOME using the following command. The path should be your extracted Hadoop folder. For my environment it is: F:\big-data\hadoop-3.3.0.

If you used PowerShell to download and if the window is still open, you can simply run the following command:

```
SETX HADOOP_HOME $dest_dir+"/hadoop-3.3.0"
```

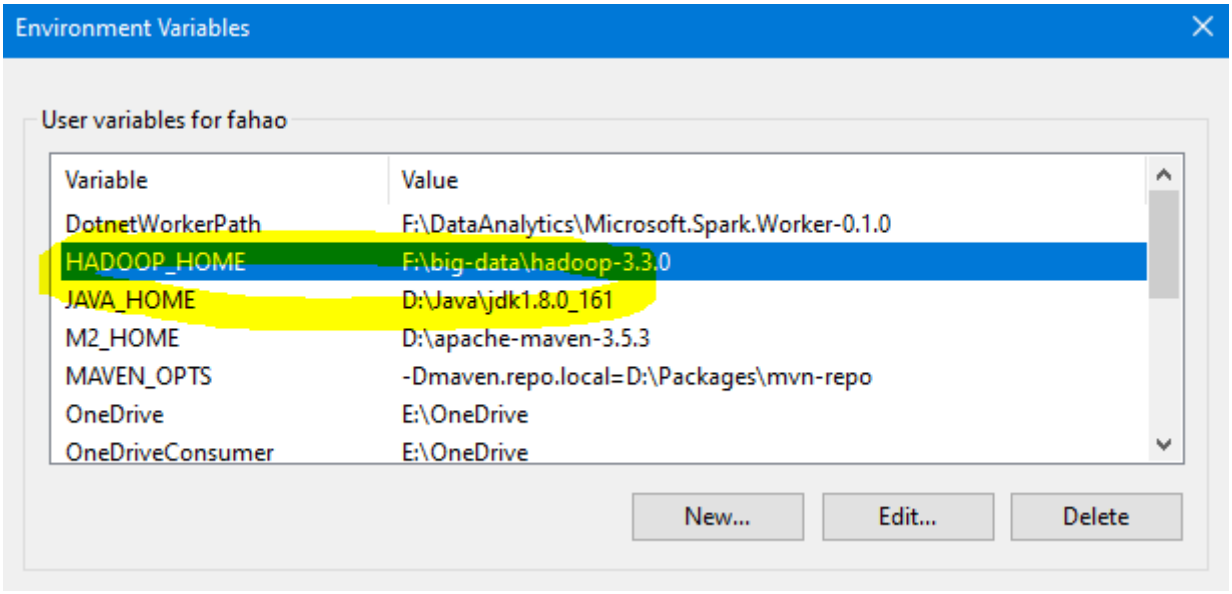
The output looks like the following screenshot:



Alternatively, you can specify the full path:

```
SETX HADOOP_HOME "F:\big-data\hadoop-3.3.0"
```

Now you can also verify the two environment variables in the system:



Configure PATH environment variable

Once we finish setting up the above two environment variables, we need to add the bin folders to the PATH environment variable.

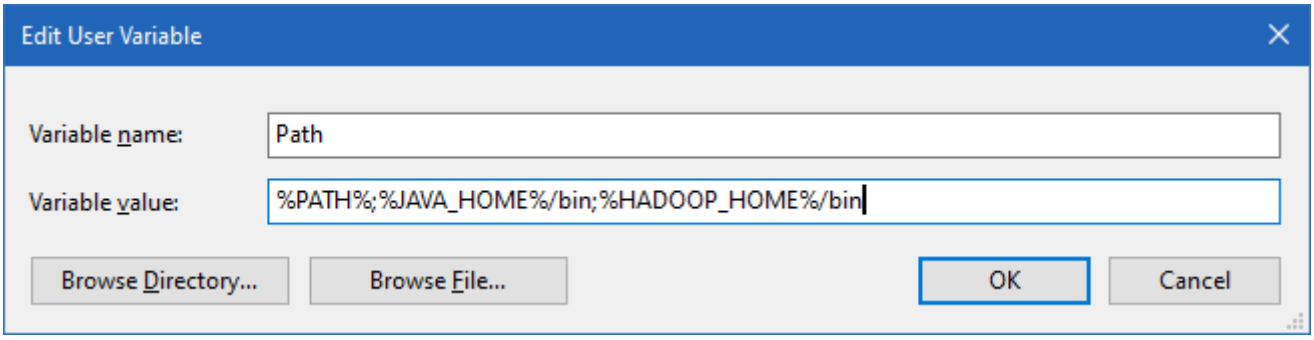
If PATH environment exists in your system, you can also manually add the following two paths to it:

- %JAVA_HOME%/bin
- %HADOOP_HOME%/bin

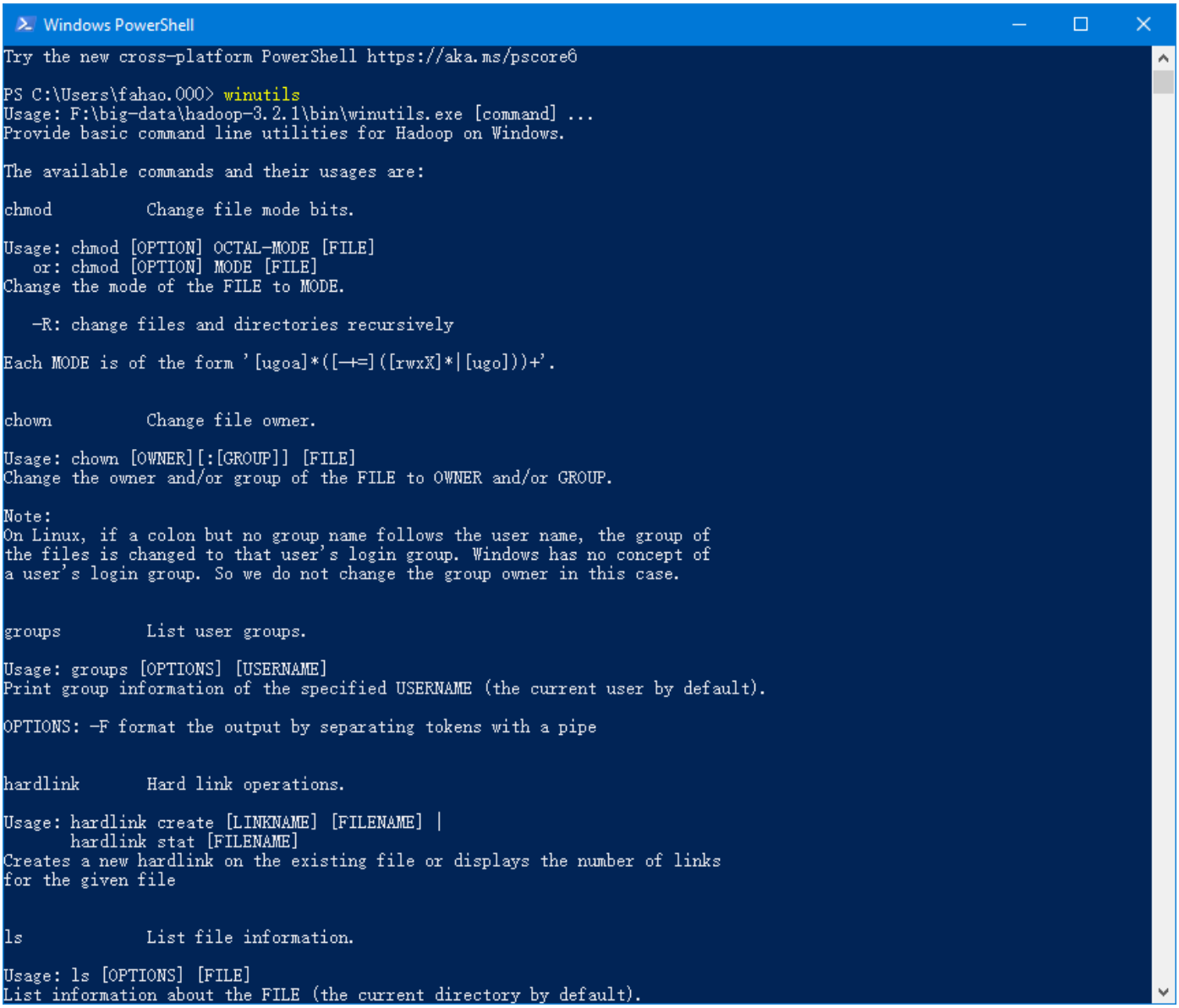
Alternatively, you can run the following command to add them:

```
setx PATH "$env:PATH;$env:JAVA_HOME/bin;$env:HADOOP_HOME/bin"
```

If you don't have other user variables setup in the system, you can also directly add a Path environment variable that references others to make it short:



Close PowerShell window and open a new one and type winutils.exe directly to verify that our above steps are completed successfully:



You should also be able to run the following command:

```
hadoop -version
java version "1.8.0_161"
Java(TM) SE Runtime Environment (build 1.8.0_161-b12)
Java HotSpot(TM) 64-Bit Server VM (build 25.161-b12, mixed mode)
```

Step 6 - Configure Hadoop

Now we are ready to configure the most important part - Hadoop configurations which involves Core, YARN, MapReduce, HDFS configurations.

Configure core site

Edit file core-site.xml in %HADOOP_HOME%\etc\hadoop folder. For my environment, the actual path is F:\big-data\hadoop-3.3.0\etc\hadoop.

Replace configuration element with the following:

```
<configuration>
   <property>
     <name>fs.default.name</name>
     <value>hdfs://0.0.0.0:19000</value>
   </property>
</configuration>
```

Configure HDFS

Edit file hdfs-site.xml in %HADOOP_HOME%\etc\hadoop folder.

Before editing, please correct two folders in your system: one for namenode directory and another for data directory. For my system, I created the following two sub folders:

- F:\big-data\data\dfs\namespace_logs_330
- F:\big-data\data\dfs\data_330

Replace configuration element with the following (remember to replace the highlighted paths accordingly):

```
<configuration>
   <property>
     <name>dfs.replication</name>
     <value>1</value>
   </property>
   <property>
     <name>dfs.namenode.name.dir</name>
     <value>file:///F:/big-data/data/dfs/namespace_logs_330</value>
   </property>
   <property>
     <name>dfs.datanode.data.dir</name>
     <value>file:///F:/big-data/data/dfs/data_330</value>
   </property>
</configuration>
```

In Hadoop 3, the property names are slightly different from previous version. Refer to the following official documentation to learn more about the configuration properties:

Hadoop 3.3.0 hdfs_default.xml

ⓘ For DFS replication we configure it as one as we are configuring just one single node. By default the value is 3.

The directory configuration are not mandatory and by default it will use Hadoop temporary folder. For our tutorial purpose, I would recommend customise the values.

Configure MapReduce and YARN site

Edit file mapred-site.xml in %HADOOP_HOME%\etc\hadoop folder.

Replace configuration element with the following:

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

<value>%HADOOP_HOME%/share/hadoop/mapreduce/*,%HADOOP_HOME%/share/hadoop/mapreduce/lib/
*,%HADOOP_HOME%/share/hadoop/common/*,%HADOOP_HOME%/share/hadoop/common/lib/*,%HADOOP_H
OME%/share/hadoop/yarn/*,%HADOOP_HOME%/share/hadoop/yarn/lib/*,%HADOOP_HOME%/share/hado
op/hdfs/*,%HADOOP_HOME%/share/hadoop/hdfs/lib/*</value>
    </property>
</configuration>
```

Edit file yarn-site.xml in %HADOOP_HOME%\etc\hadoop folder.

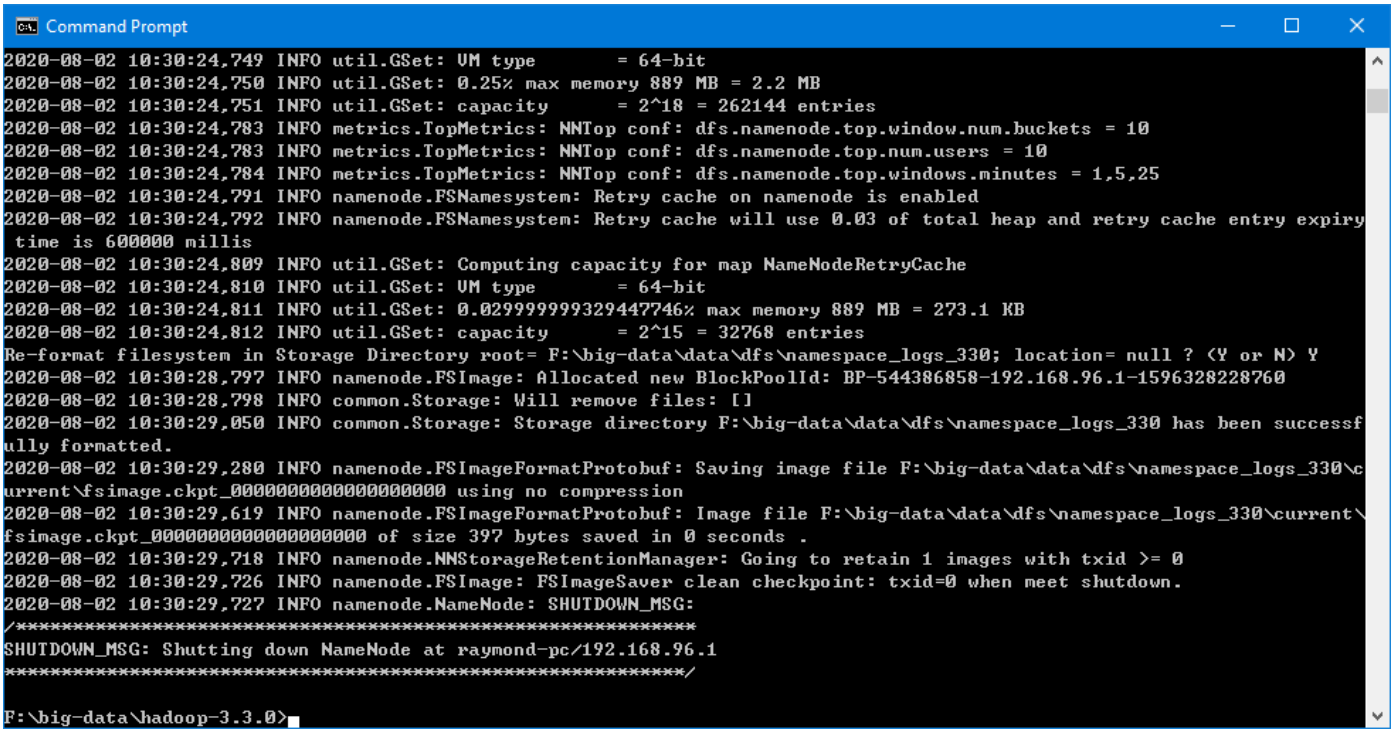
```
<configuration>
    <property>
        <name>yarn.nodemanager.aux-services</name>
        <value>mapreduce_shuffle</value>
    </property>
    <property>
        <name>yarn.nodemanager.env-whitelist</name>

<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PREPEND_
DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
    </property>
</configuration>
```

Step 7 - Initialise HDFS & bug fix

Run the following command in Command Prompt

```
hdfs namenode -format
```

The following is an example when it is formatted successfully:

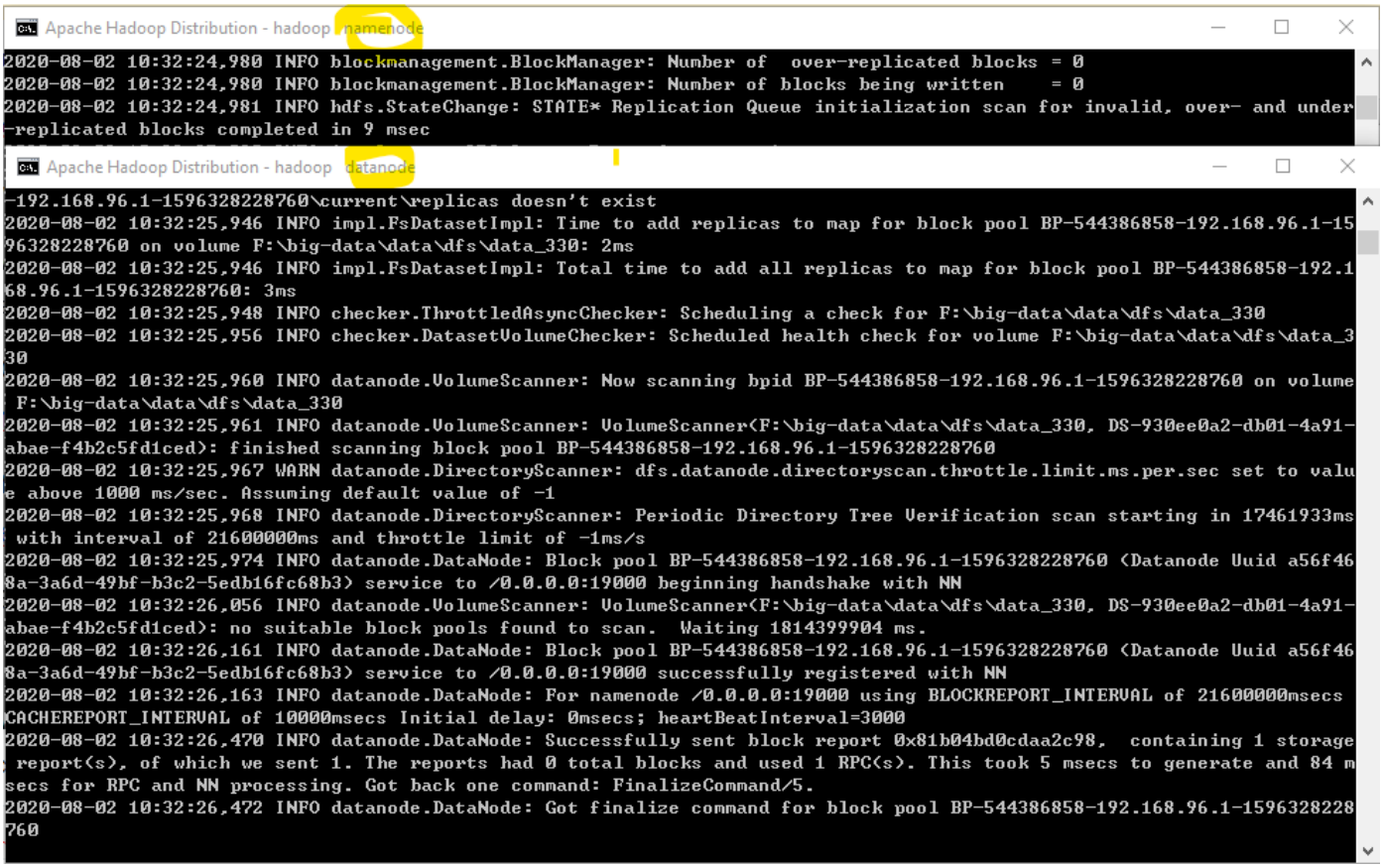


Step 8 - Start HDFS daemons

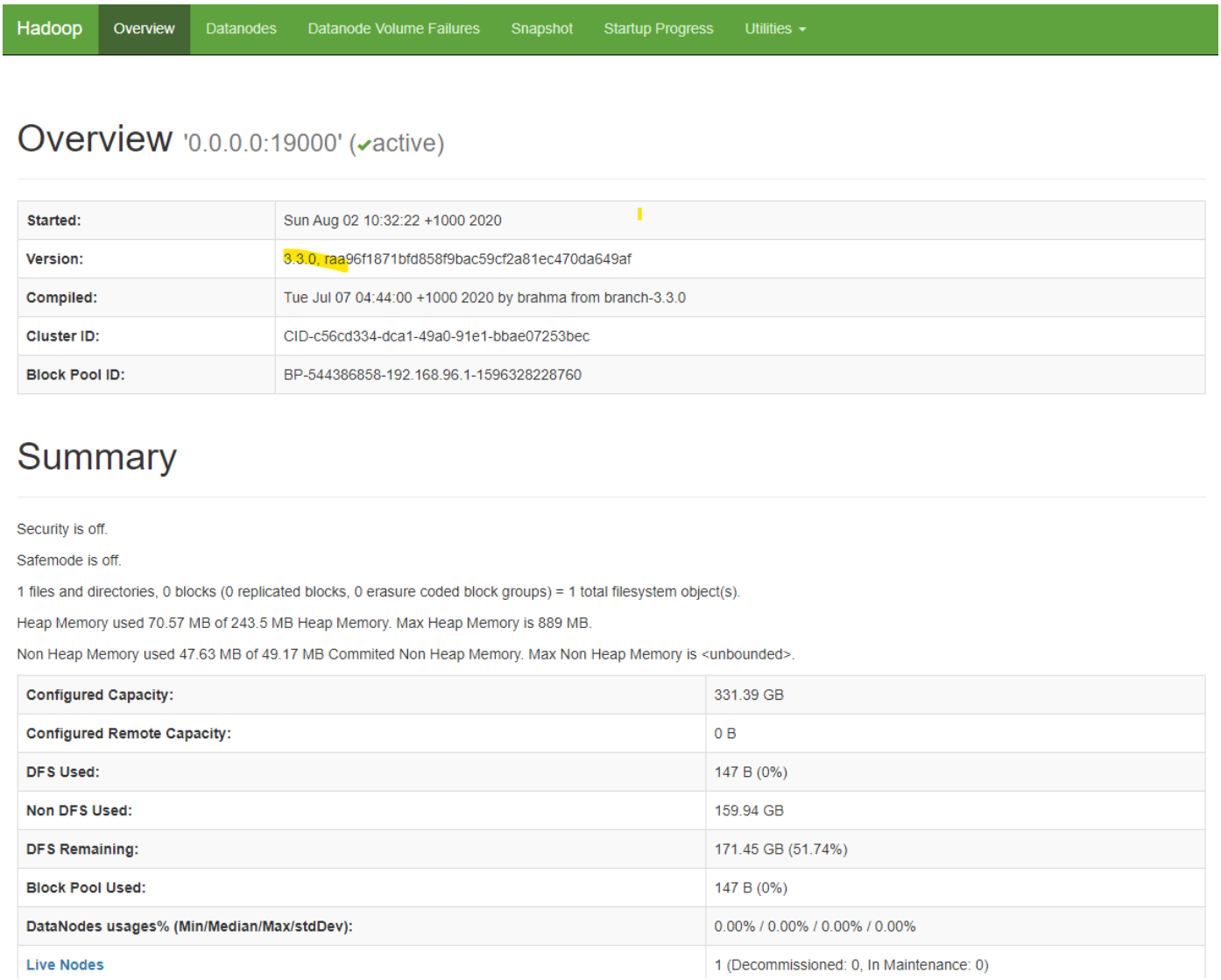
Run the following command to start HDFS daemons in Command Prompt:

```
%HADOOP_HOME%\sbin\start-dfs.cmd
```

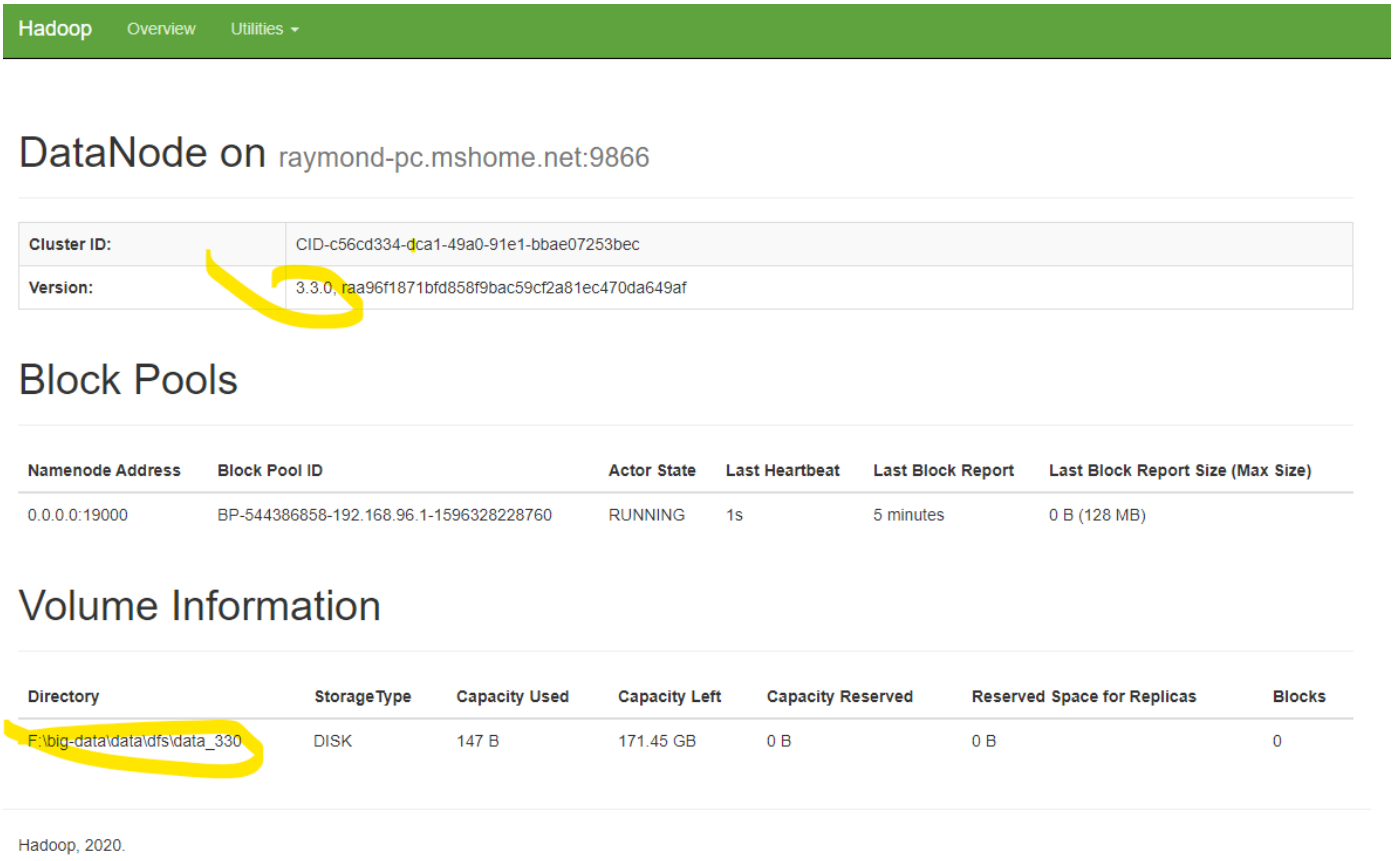
Two Command Prompt windows will open: one for **datanode** and another for **namenode** as the following screenshot shows:



Verify HDFS web portal UI through this link: http://localhost:9870/dfshealth.html#tab-overview.



You can also navigate to a data node UI:



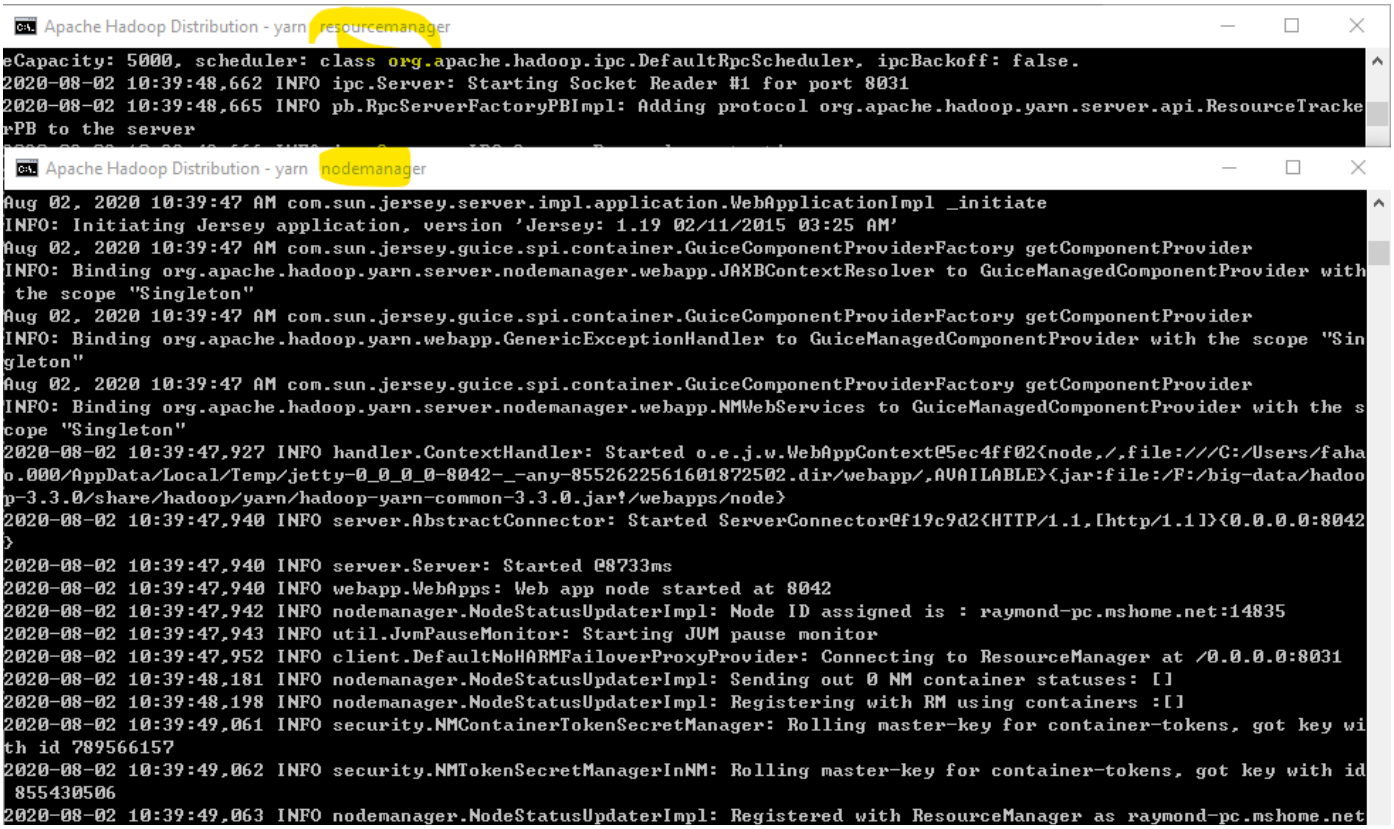
Step 9 - Start YARN daemons

> ⚠ You may encounter permission issues if you start YARN daemons using normal user. To ensure you don't encounter any issues. Please open a Command Prompt window using Run as administrator.
> Alternatively, you can follow this comment on this page which doesn't require Administrator permission using a local Windows account: https://kontext.tech/article/377/latest-hadoop-321-installation-on-windows-10-step-by-step-guide#comment314

Run the following command in an elevated Command Prompt window (Run as administrator) to start YARN daemons:

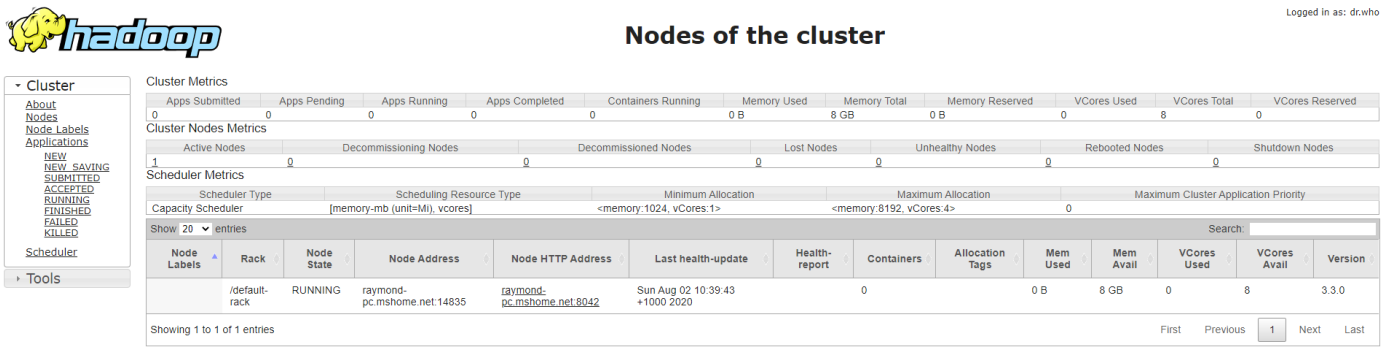
```
%HADOOP_HOME%\sbin\start-yarn.cmd
```

Similarly two Command Prompt windows will open: one for resource manager and another for node manager as the following screenshot shows:



Kontext

> http://localhost:8088



Step 10 - Verify Java processes

Run the following command to verify all running processes:

```
jps
```

The output looks like the following screenshot:

* We can see the process ID of each Java process for HDFS/YARN.

Step 11 - Shutdown YARN & HDFS daemons

You don't need to keep the services running all the time. You can stop them by running the following commands one by one once you finish the test:

```
%HADOOP_HOME%\sbin\stop-yarn.cmd
```

```
%HADOOP_HOME%\sbin\stop-dfs.cmd
```

✓ Congratulations! You've successfully completed the installation of Hadoop 3.3.0 on Windows 10.

Let me know if you encounter any issues. Enjoy with your latest Hadoop on Windows 10.

More from Kontext

- Install Hadoop 3.3.1 on Windows 10 Step by Step Guide
- Install Hadoop 3.2.1 on Windows 10 Step by Step Guide
- Apache Hive 3.0.0 Installation on Windows 10 Step by Step Guide
- Install Hadoop 3.3.0 on macOS
- Install Hadoop 3.3.0 on Linux
- Install Spark 3.3.0 on Linux or WSL

windows10 hadoop yarn hdfs big-data-on-windows-10

Comments

« ‹ 1 2 3 4 › » 1-10 of 36

Raymond

2 years ago

#1499 Re: Install Hadoop 3.3.0 on Windows 10 Step by Step Guide

I need your system information:

```
OS Name:                   Microsoft Windows 10 Pro
OS Version:                10.0.19043 N/A Build 19043
OS Manufacturer:           Microsoft Corporation
OS Configuration:          Standalone Workstation
OS Build Type:             Multiprocessor Free
...
System Type:               x64-based PC
Processor(s):              1 Processor(s) Installed.
                           [01]: Intel64 Family 6 Model 94 Stepping 3
GenuineIntel ~2601 Mhz
```

You can get it from **systeminfo** command.

> Antonio 2 years ago
> Re: Install Hadoop 3.3.0 on Windows 10 Step by Step Guide
> I'm facing the same issue, do you want my powershell version is
> Name Value...
> + Read more

Reply

A Antonio

2 years ago

#1498 Re: Install Hadoop 3.3.0 on Windows 10 Step by Step Guide

I'm facing the same issue, do you want my powershell version is

Name Value...

+ Read more

> Raymond 2 years ago
> Re: Install Hadoop 3.3.0 on Windows 10 Step by Step Guide
> Can you provide your system information here? The one I compiled is for x64 systems.

Reply

Raymond 2 years ago

#1487 Re: Install Hadoop 3.3.0 on Windows 10 Step by Step Guide

Thanks for the feedback. I've update it. :)

> Şahin 2 years ago
> Re: Install Hadoop 3.3.0 on Windows 10 Step by Step Guide
> You could cd $client_dir instead of cd F:/big-data

Reply

Raymond 2 years ago

#1486 Re: Install Hadoop 3.3.0 on Windows 10 Step by Step Guide

Can you provide your system information here? The one I compiled is for x64 systems.

Şahin 2 years ago

Re: Install Hadoop 3.3.0 on Windows 10 Step by Step Guide

After writing winutils.exe, I get the error:

Program 'winutils.exe' failed to run: The specified executable is not a valid application for this OS platform.At line:

1 char:1

+ .\winutils.exe

+ ~~~~~~~~~~~~~~~.

At line:1 char:1

+ .\winutils.exe

+ ~~~~~~~~~~~~~~~

+ CategoryInfo : ResourceUnavailable: (:) [], ApplicationFailedException

+ FullyQualifiedErrorId : NativeCommandFailed

Reply

Ş Şahin

2 years ago

#1485 Re: Install Hadoop 3.3.0 on Windows 10 Step by Step Guide

After writing winutils.exe, I get the error:

Program 'winutils.exe' failed to run: The specified executable is not a valid application for this OS platform.At line:

1 char:1

+ .\winutils.exe

+ ~~~~~~~~~~~~~~~.

At line:1 char:1

+ .\winutils.exe

+ ~~~~~~~~~~~~~~~

+ CategoryInfo : ResourceUnavailable: (:) [], ApplicationFailedException

+ FullyQualifiedErrorId : NativeCommandFailed

Reply

Şahin

2 years ago

#1484 Re: Install Hadoop 3.3.0 on Windows 10 Step by Step Guide

You could cd $client_dir instead of cd F:/big-data

Reply

Raymond

3 years ago

#451 Re: Install Hadoop 3.3.0 on Windows 10 Step by Step Guide

It is confirmed by Naseemuddin that the special build I did for non-English Windows 10 system is working for him:

Re: Install Hadoop 3.3.0 on Windows 10 Step by Step Guide - Kontext

winutils/hadoop-3.3.0-YARN-8246/bin at master · kontext-tech/winutils · GitHub

> R4F 3 years ago
> Re: Install Hadoop 3.3.0 on Windows 10 Step by Step Guide
> HI did you find a solution i have the same issu..

Reply

Raymond

3 years ago

#450 Re: Install Hadoop 3.3.0 on Windows 10 Step by Step Guide

Hi Chenhao,

Is this guide working for you?

-Raymond

> Chenghao 3 years ago
> Re: Install Hadoop 3.3.0 on Windows 10 Step by Step Guide
> Thank u， it doesn’t work after installing it many times, just use yours for a while

Reply

R R4F

3 years ago

#449 Re: Install Hadoop 3.3.0 on Windows 10 Step by Step Guide

HI did you find a solution i have the same issu..

> Raymond 3 years ago
> Re: Install Hadoop 3.3.0 on Windows 10 Step by Step Guide
> Hi, for the winutils I built for Hadoop 3.3.0, I didn't face any issue like that.
> Hadoop 3.3.0 winutils...
> + Read more

Reply

C Chenghao

3 years ago

#448 Re: Install Hadoop 3.3.0 on Windows 10 Step by Step Guide

Thank u， it doesn’t work after installing it many times, just use yours for a while

Reply

« ‹ 1 2 3 4 › » 1-10 of 36

Please log in or register to comment.

Log in Register

Log in with external accounts

Log in with Microsoft account

Log in with Google account

Table of contents

Stats

Big Data Tools on Windows 10

Compile and Build Hadoop 3.2.1 on Windows 10 Guide

Install Hadoop 3.2.1 on Windows 10 Step by Step Guide

Install Hadoop 3.0.0 on Windows (Single Node)

Install and Run Kafka 2.6.0 On Windows 10

Install Apache Spark 3.0.0 on Windows 10

Read more (11)

windows10

hadoop

yarn

hdfs