How to Set Up Hadoop on Windows: A Step-by-Step Guide

- Hadoop is a well-known big data processing system for storing and analysing enormous volumes of data.
- It's an open-source project that you can use for free.
- If you're new to Hadoop, you may find the installation process difficult.

Steps of installing Hadoop on Windows.

Step 1: Download and install Java

Hadoop is built on Java, so you must have Java installed on your PC. You can get the most recent version of Java from the official website. After downloading, follow the installation wizard to install Java on your system.
JDK: https://www.oracle.com/java/technologies/javase-downloads.html

Step 2: Download Hadoop

Hadoop can be downloaded from the Apache Hadoop website. Make sure to have the latest stable release of Hadoop. Once downloaded, extract the

contents to a convenient location.

Hadoop: https://hadoop.apache.org/releases.html

Step 3: Set Environment Variables

You must configure environment variables after downloading and unpacking Hadoop. Launch the Start menu, type "Edit the system environment variables," and select the result. This will launch the System Properties dialogue box. Click on "Environment Variables" button to open.
Click "New" under System Variables to add a new variable. Enter the variable name "HADOOP_HOME" and the path to the Hadoop folder as the variable value. Then press "OK."
Then, under System Variables, locate the "Path" variable and click "Edit." Click "New" in the Edit Environment Variable window and enter "% HADOOP_HOME%bin" as the variable value. To close all the windows, use the "OK" button.

Step 4: Setup Hadoop

You must configure Hadoop in this phase by modifying several configuration files. Navigate to the "etc/hadoop" folder in the Hadoop folder. You must make changes to three files:

- core-site.xml
- hdfs-site.xml
- mapred-site.xml

Open each file in a text editor and edit the following properties:
In core-site.xml
`<configuration>`

`<property>`

```xml
<name>fs.default.name</name>

<value>hdfs://localhost:9000</value>

</property>

</configuration>
```

In hdfs-site.xml

```xml
<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>file:/hadoop-3.3.1/data/namenode</value>
```

```
</property>


<property>


<name>dfs.datanode.data.dir</name>


<value>file:/hadoop-3.3.1/data/datanode</value>


</property>


</configuration>
```

In mapred-site.xml

```
<configuration>


<property>


<name>mapred.job.tracker</name>


<value>localhost:54311</value>


</property>


</configuration>
```

Save the changes in each file.

Step 5: Format Hadoop NameNode

You must format the NameNode before you can start Hadoop. Navigate to

the Hadoop bin folder using a command prompt. Execute this command:

`hadoop namenode -format`

Step 6: Start Hadoop

To start Hadoop, open a command prompt and navigate to the Hadoop bin folder.

Run the following command:

`start-all.cmd`

This command will start all the required Hadoop services, including the NameNode, DataNode, and JobTracker. Wait for a few minutes until all the services are started.

Step 7: Verify Hadoop Installation

To ensure that Hadoop is properly installed, open a web browser and go to http://localhost:50070/. This will launch the web interface for the Hadoop NameNode.

➢ Wrapping Up

Remember to get the most recent stable version of Hadoop, install Java, configure Hadoop, format the NameNode, and start Hadoop services. Finally, check the NameNode web interface to ensure that Hadoop is properly installed.