

# Cloud Application Development

## Week -8

### WEEK-8: DATA INTENSIVE PROGRAMMING Install Hadoop single node cluster and run simple applications like word count

You must have got a theoretical idea about Hadoop, HDFS and its architecture. But to get Hadoop Certified you need good hands-on knowledge. I hope you would have liked our previous blog on [HDFS Architecture](#), now I will take you through the practical knowledge about Hadoop and HDFS. The first step forward is to install Hadoop.

There are two ways to install Hadoop, i.e. **Single node** and **Multi-node**.

**A single node cluster** means only one DataNode running and setting up all the NameNode, DataNode, ResourceManager, and NodeManager on a single machine. This is used for studying and testing purposes. For example, let us consider a sample data set inside the healthcare industry. So, for testing whether the Oozie jobs have scheduled all the processes like collecting, aggregating, storing, and processing the data in a proper sequence, we use a single node cluster. It can easily and efficiently test the sequential workflow in a smaller environment as compared to large environments which contain terabytes of data distributed across hundreds of machines.

While in a **Multi-node cluster**, there are more than one DataNode running and each DataNode is running on different machines. The multi-node cluster is practically used in organizations for analyzing Big Data. Considering the above example, in real-time when we deal with petabytes of data, it needs to be distributed across hundreds of machines to be processed. Thus, here we use a multi-node cluster.

### Prerequisites

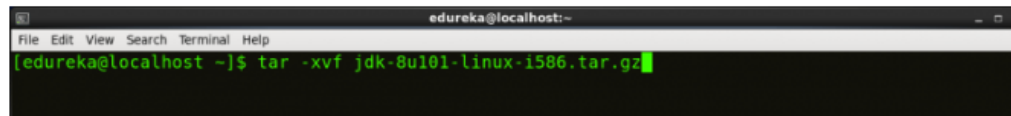
- *VIRTUAL BOX*: it is used for installing the operating system on it.
- *OPERATING SYSTEM*: You can install Hadoop on Linux-based operating systems. Ubuntu and CentOS are very commonly used. In this tutorial, we are using CentOS.
- *JAVA*: You need to install the Java 8 package on your system.
- *HADOOP*: You require Hadoop 2.7.3 package.

## Install Hadoop

**Step 1:** [Click here](#) to download the Java 8 Package. Save this file in your home directory.

**Step 2:** Extract the Java Tar File.

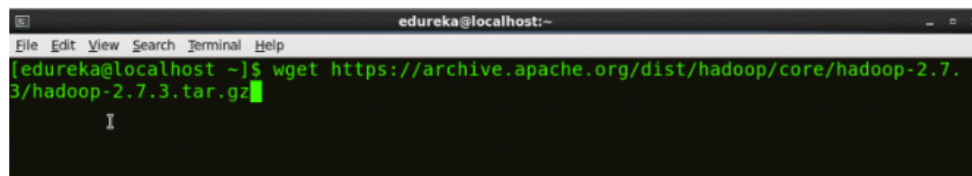
**Command:** `tar -xvf jdk-8u101-linux-i586.tar.gz`

A terminal window titled 'edureka@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The command 'tar -xvf jdk-8u101-linux-i586.tar.gz' is entered and executed, with a green cursor at the end of the line.

*Fig: Hadoop Installation – Extracting Java Files*

**Step 3:** Download the Hadoop 2.7.3 Package.

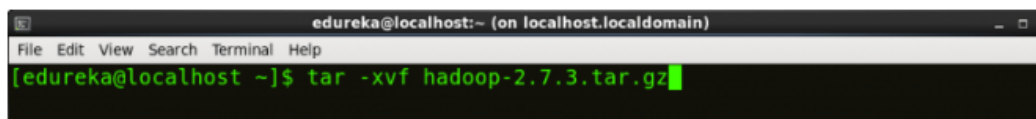
**Command:** `wget https://archive.apache.org/dist/hadoop/core/hadoop-2.7.3/hadoop-2.7.3.tar.gz`

A terminal window titled 'edureka@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The command 'wget https://archive.apache.org/dist/hadoop/core/hadoop-2.7.3/hadoop-2.7.3.tar.gz' is entered and executed, with a green cursor at the end of the line.

*Fig: Hadoop Installation – Downloading Hadoop*

**Step 4:** Extract the Hadoop tar File.

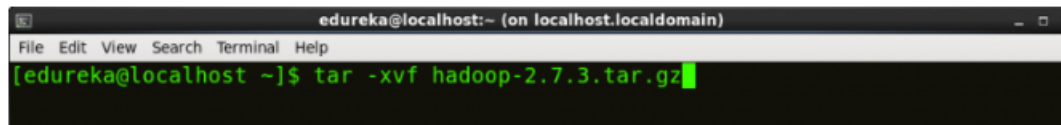
**Command:** `tar -xvf hadoop-2.7.3.tar.gz`

A terminal window titled 'edureka@localhost:~ (on localhost.localdomain)' with a menu bar (File, Edit, View, Search, Terminal, Help). The command 'tar -xvf hadoop-2.7.3.tar.gz' is entered and executed, with a green cursor at the end of the line.

*Fig: Hadoop Installation – Extracting Hadoop Files*

**Step 4:** Extract the Hadoop tar File.

**Command:** tar -xvf hadoop-2.7.3.tar.gz

A terminal window titled 'edureka@localhost:~ (on localhost.localdomain)' with a menu bar (File, Edit, View, Search, Terminal, Help). The command '[edureka@localhost ~]\$ tar -xvf hadoop-2.7.3.tar.gz' is entered and executed, with a green cursor at the end of the line.

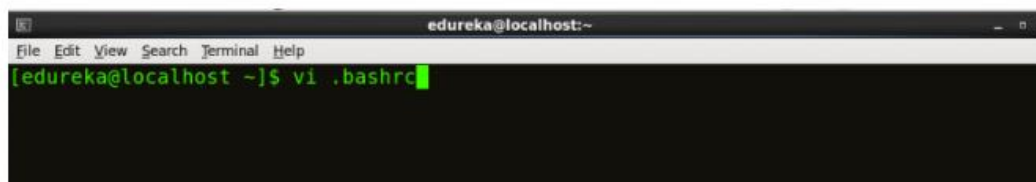
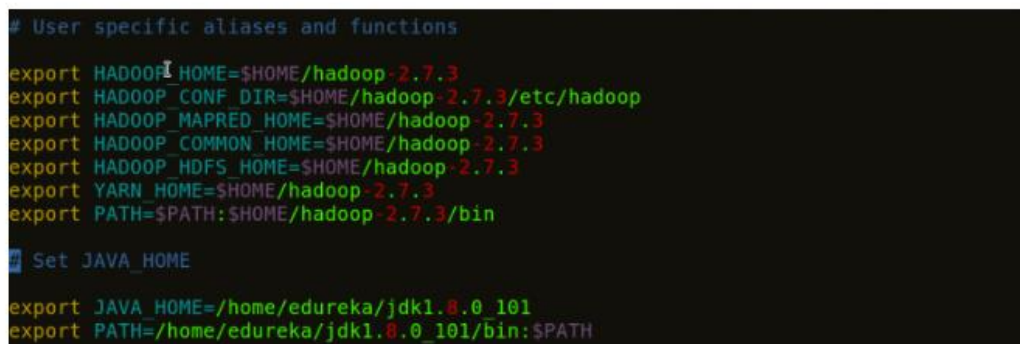
*Fig: Hadoop Installation – Extracting Hadoop Files*

**Step 5:** Add the Hadoop and Java paths in the bash file (.bashrc).

Open. **bashrc** file. Now, add Hadoop and Java Path as shown below.

Learn more about the Hadoop Ecosystem and its tools with the [Hadoop Certification](#).

**Command:** vi .bashrc

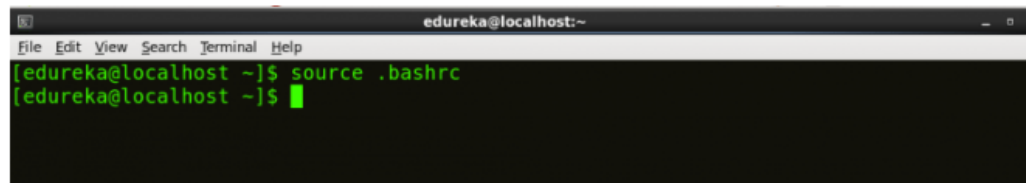
A terminal window titled 'edureka@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The command '[edureka@localhost ~]\$ vi .bashrc' is entered and executed, with a green cursor at the end of the line.A screenshot of the .bashrc file content. It shows a section for user-specific aliases and functions, followed by export statements for Hadoop environment variables (HADOOP\_HOME, HADOOP\_CONF\_DIR, HADOOP\_MAPRED\_HOME, HADOOP\_COMMON\_HOME, HADOOP\_HDFS\_HOME, YARN\_HOME, and PATH) and a section for setting JAVA\_HOME (export JAVA\_HOME=/home/edureka/jdk1.8.0\_101 and export PATH=/home/edureka/jdk1.8.0\_101/bin:\$PATH).

*Fig: Hadoop Installation – Setting Environment Variable*

Then, save the bash file and close it.

For applying all these changes to the current Terminal, execute the source command.

**Command:** source .bashrc

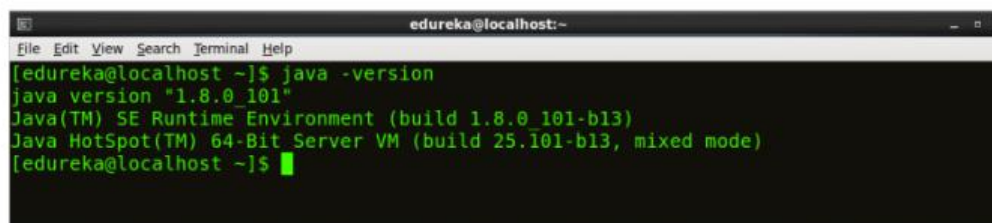


```
edureka@localhost:~  
File Edit View Search Terminal Help  
[edureka@localhost ~]$ source .bashrc  
[edureka@localhost ~]$
```

*Fig: Hadoop Installation – Refreshing environment variables*

To make sure that Java and Hadoop have been properly installed on your system and can be accessed through the Terminal, execute the `java -version` and `hadoop version` commands.

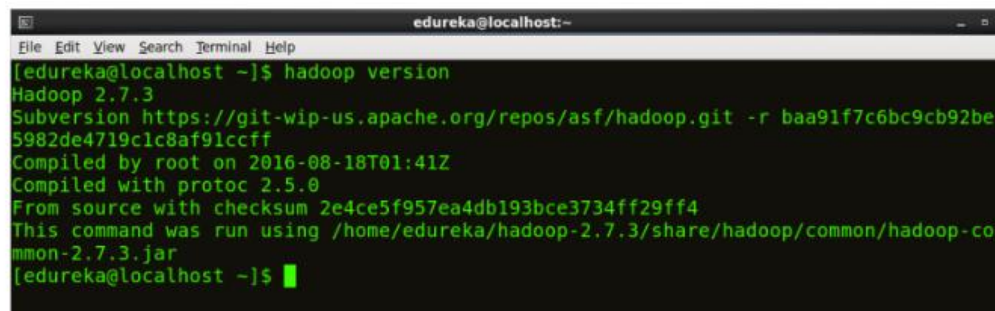
**Command:** `java -version`



```
edureka@localhost:~  
File Edit View Search Terminal Help  
[edureka@localhost ~]$ java -version  
java version "1.8.0_101"  
Java(TM) SE Runtime Environment (build 1.8.0_101-b13)  
Java HotSpot(TM) 64-Bit Server VM (build 25.101-b13, mixed mode)  
[edureka@localhost ~]$
```

*Fig: Hadoop Installation – Checking Java Version*

**Command:** `hadoop version`



```
edureka@localhost:~  
File Edit View Search Terminal Help  
[edureka@localhost ~]$ hadoop version  
Hadoop 2.7.3  
Subversion https://git-wip-us.apache.org/repos/asf/hadoop.git -r baa91f7c6bc9cb92be  
5982de4719c1c8af91ccff  
Compiled by root on 2016-08-18T01:41Z  
Compiled with protoc 2.5.0  
From source with checksum 2e4ce5f957ea4db193bce3734ff29ff4  
This command was run using /home/edureka/hadoop-2.7.3/share/hadoop/common/hadoop-co  
mmon-2.7.3.jar  
[edureka@localhost ~]$
```

*Fig: Hadoop Installation – Checking Hadoop Version*

**Step 6:** Edit the **Hadoop Configuration files**.

**Command:** `cd hadoop-2.7.3/etc/hadoop/`

**Command:** `ls`



