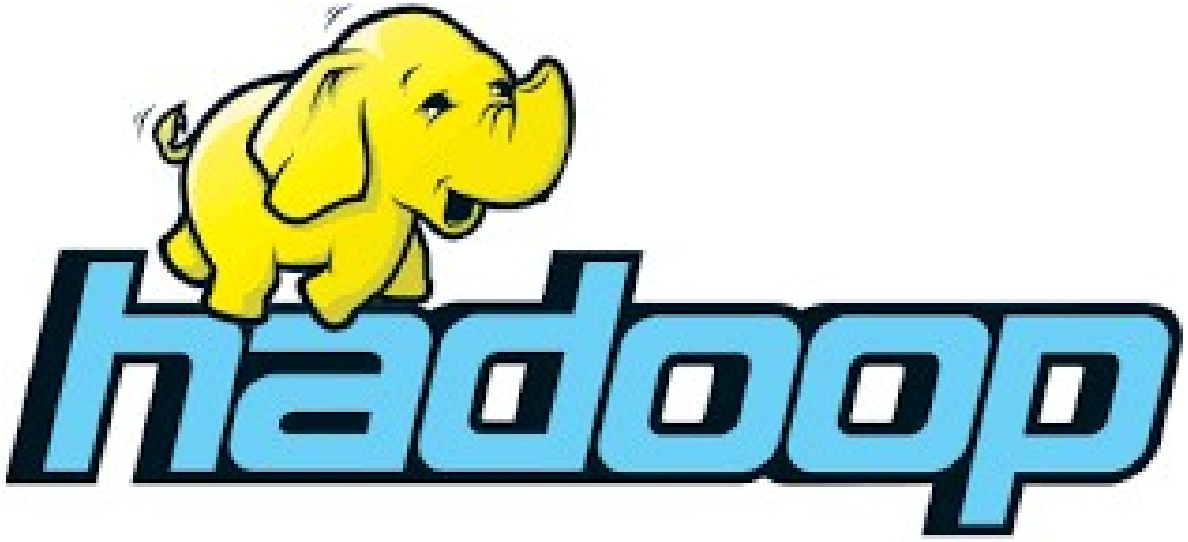
How to Install Hadoop on CentOs 8



CentOS

1. Install Java

- **Run following if you do not have Java installed:**
  - sudo yum update
  - sudo yum install java-1.8.0-openjdk
- **These commands will update the package information on your VPS and then install Java.**
- **Verify that Java has been installed on your system:**
  - java -version

2. Create and Setup SSH Certificates

- **Hadoop uses SSH (to access its nodes) which would normally require the user to enter a password.**
- **This requirement can be eliminated by creating and setting up SSH certificates using the following commands:**
  - ssh-keygen -t rsa -P ''
  - cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys

3. Download and Install Hadoop

- **Download Hadoop from one of the mirrors using the following command:**
  - wget https://downloads.apache.org/hadoop/common/hadoop-3.2.1/hadoop-3.2.1.tar.gz
- **Unzip tar file**
  - tar xfz hadoop-3.2.1.tar.gz
- **Move to your desired location**
  - mv hadoop-3.2.1 /home/prave/hadoop

4. Edit and Setup Configuration Files

- **Setup of Hadoop requires following files to be modified:**
  - ~/.bashrc
  - /home/prave/hadoop/etc/hadoop/hadoop-env.sh
  - /home/prave/hadoop/etc/hadoop/core-site.xml
  - /home/prave/hadoop/etc/hadoop/yarn-site.xml
  - /home/prave/hadoop/etc/hadoop/mapred-site.xml.template
  - /home/prave/hadoop/etc/hadoop/hdfs-site.xml

1) Editing ~/.bashrc

- **Find the path where Java has been installed to set the JAVA_HOME environment variable.**
  - sudo update-alternatives --config java
- **Open ~/.bashrc with vi editor:**
  - vi ~/.bashrc
- **Edit following environment variables to ~/.bashrc**
- **Save changes**
  - **source** ~/.bashrc

1) Editing ~/.bashrc

```
#HADOOP VARIABLES START
export JAVA_HOME=###java-home copied in last step###
export HADOOP_INSTALL=/home/prave/hadoop
export PATH=$PATH:$HADOOP_INSTALL/bin
export PATH=$PATH:$HADOOP_INSTALL/sbin
export HADOOP_MAPRED_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_HOME=$HADOOP_INSTALL
export HADOOP_HDFS_HOME=$HADOOP_INSTALL
export YARN_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_INSTALL/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_INSTALL/lib"
#HADOOP VARIABLES END
```

2) Editing /home/prave/hadoop/etc/hadoop/hadoop-env.sh

- **Open hadoop-env.sh file with vi editor:**
  - vi /home/prave/hadoop/etc/hadoop/hadoop-env.sh
- **Change JAVA_HOME variable:**
  - export JAVA_HOME=###java-home copied in last step###

3) Editing /home/prave/hadoop/etc/hadoop/core-site.xml

- **Open this file with vi using the following command:**
  - vi /home/prave/hadoop/etc/hadoop/core-site.xml
- **In this file, enter the following content in between the <configuration></configuration> tag:**

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

4) Editing /home/prave/hadoop/etc/hadoop/yarn-site.xml

- **Open this file with vi using the following command:**
  - vi /home/prave/hadoop/etc/hadoop/yarn-site.xml
- **In this file, enter the following content in between the <configuration></configuration> tag:**

```
<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>
```

5) Creating and Editing /home/prave/hadoop/etc/hadoop/mapred-site.xml

- **(Below step, Not in all hadoop versions)**
- **If mapred-site.xml does not exist it can to be renamed/copied using template file mapred-site.xml.template**
  - cp /home/prave/hadoop/etc/hadoop/mapred-site.xml.template /home/prave/hadoop/etc/hadoop/mapred-site.xml

5) Creating and Editing /home/prave/hadoop/etc/hadoop/mapred-site.xml

- **open the newly created file with vi using the following command:**
  - vi /home/prave/hadoop/etc/hadoop/mapred-site.xml
- **In this file, enter the following content in between the <configuration></configuration> tag:**

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

6) Editing /home/prave/hadoop/etc/hadoop/hdfs-site.xml

- **Create two directories which will contain the namenode and the datanode for this Hadoop installation:**
  - mkdir -p /home/prave/hadoop_store/hdfs/namenode
  - mkdir -p /home/prave/hadoop_store/hdfs/datanode

6) Editing /home/prave/hadoop/etc/hadoop/hdfs-site.xml

- **open the /home/prave/hadoop/etc/hadoop/hdfs-site.xml file with vi using the following command:**
  - vi /home/prave/hadoop/etc/hadoop/hdfs-site.xml

6) Editing /home/prave/hadoop/etc/hadoop/hdfs-site.xml

- Enter the following content in between the <configuration></configuration> tag:

```
<property>
  <name>dfs.replication</name>
  <value>1</value>
 </property>
 <property>
  <name>dfs.namenode.name.dir</name>
  <value>file:/home/prave/hadoop_store/hdfs/namenode</value>
 </property>
 <property>
  <name>dfs.datanode.data.dir</name>
  <value>file:/home/prave/hadoop_store/hdfs/datanode</value>
 </property>
```

Format and Start

- **Format the New Hadoop Filesystem**
  - hdfs namenode -format
- **Start Hadoop the newly installed single node cluster:**
  - start-dfs.sh
  - start-yarn.sh
- **Verify your hadoop**
  - Jps
  - http://localhost:9870