

Hadoop installations in Windows

- What is hadoop?

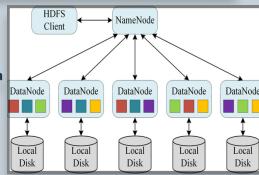
An open source framework that allows to store & process large data sets in parallel & distributed manner.

- What are the main components of hadoop?

There are two main components of hadoop.

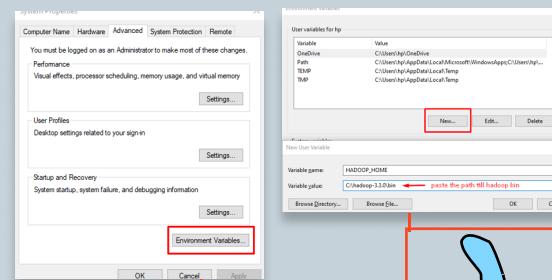
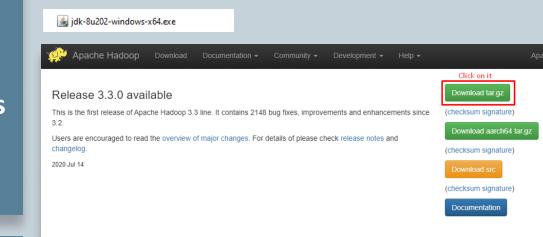
- Hadoop Distributed File System(HDFS) is the primary data storage system used by Hadoop applications.
- MapReduce is the processing unit of Hadoop.

With the help of Name node HDFS client can read and write the file. Name node stores the data in distributed manner. The DataNode perform the low level read and write requests from the file system's client.



1. Edit file C:/Hadoop-3.3.0/etc/hadoop/core-site.xml, paste the xml code in folder and save

```
<configuration> paste this code between configuration
<property>
  <name>fs.defaultFS</name>
  <value>hdfs://localhost:9000</value>
</property>
</configuration>
```



Download & Installation

- Java
- Hadoop

Environment Variables and PATH

- Java Home
- Hadoop Home

Configurations

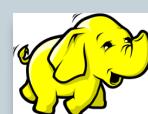
- Edit 5 files
- Configure Hadoop

Testing

Testing if installation was successful.

3. Create folder "data" under "C:\Hadoop-3.3.0"

- Create folder "datanode" under "C:\Hadoop-3.3.0\data"
- Create folder "namenode" under "C:\Hadoop-3.3.0\data"



hdfs-site.xml

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

```
mapred-site.xml
```

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

4. Edit file C:/Hadoop-3.3.0/etc/hadoop/hadoop-env.cmd

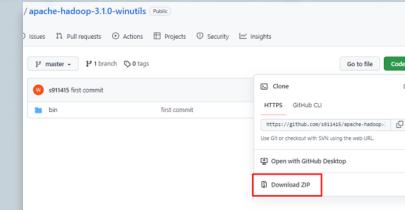
Open this file and paste the your java folder path till jdk here.
rem set JAVA_HOME in this file, so that it is correctly
rem remote nodes.
rem The java implementation to use. Required.
+ JAVA_HOME=%JAVA_HOME%

rem remote nodes.
rem The java implementation to use. Required.
set JAVA_HOME=C:\Java\jdk1.8.0_202

yarn-site.xml

```
<configuration>
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
<property>
  <name>dfs.namenode.name.dir</name>
  <value>/hadoop-3.3.0/data/namenode</value>
</property>
<property>
  <name>dfs.datanode.data.dir</name>
  <value>/hadoop-3.3.0/data/datanode</value>
</property>
</configuration>
```

5. After downloading unzip this folder and copy bin folder and replace existing bin folder in C:\Hadoop-3.3.0\bin



- Where hadoop is used?

It is used at financial sectors, Security and Law Enforcement, Companies use Hadoop for understanding customers requirements, Retail industry, Government sectors, Advertisements Targeting Platforms, Health care sectors, Optimizing machine performance

6. Open cmd and type command "hdfs namenode -format"

```
Microsoft Windows [Version 10.0.19041.572]
(c) 2020 Microsoft Corporation. All rights reserved.

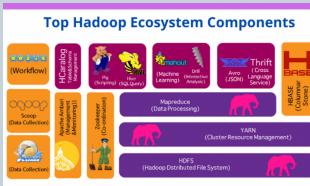
C:\hadoop-3.3.0\bin>hdfs namenode -format
```

7. Open cmd and change directory to C:\Hadoop-3.3.0\sbin
- type start-all.cmd

```
Microsoft Windows [Version 10.0.19041.572]
(c) 2020 Microsoft Corporation. All rights reserved.

C:\hadoop-3.3.0\sbin>start-all.cmd
```

7. Open browser and type localhost:8088



Casestudy done by
20bsit019 & 20bsit079
given by
Dr. Dharmendra Patel