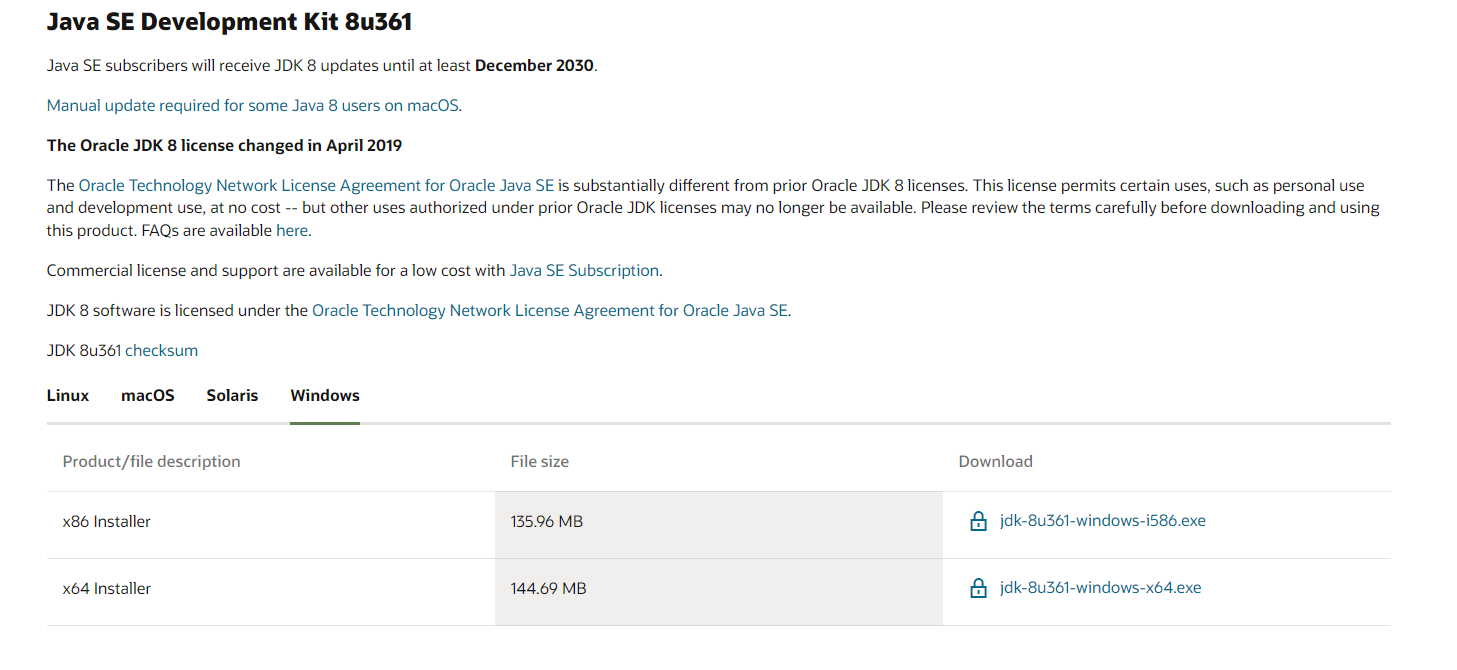
# Hadoop for Windows

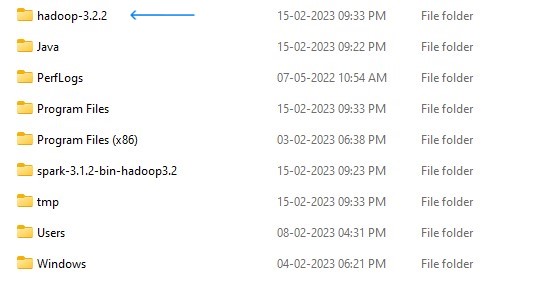
1. Install Java:
   * Download the Java Development Kit (JDK) from the Oracle website: <https://www.oracle.com/java/technologies/downloads/#java8-windows>
   * Choose the appropriate version based on your system specifications (32-bit or 64-bit).
   * Once the download is complete, run the installation file and follow the prompts to complete the installation.



1. Download Hadoop:
   * Download the Hadoop distribution for Windows from the Apache website:

<https://archive.apache.org/dist/hadoop/core/hadoop-3.2.2/hadoop-3.2.2.tar.gz>

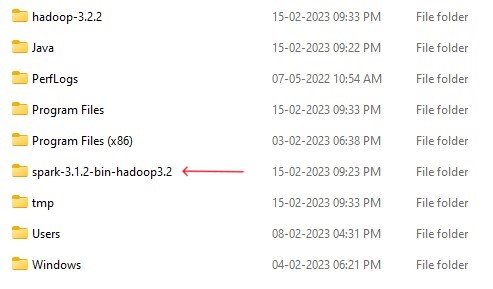
* + Once the download is complete, extract the contents of the file to a folder on your local machine.



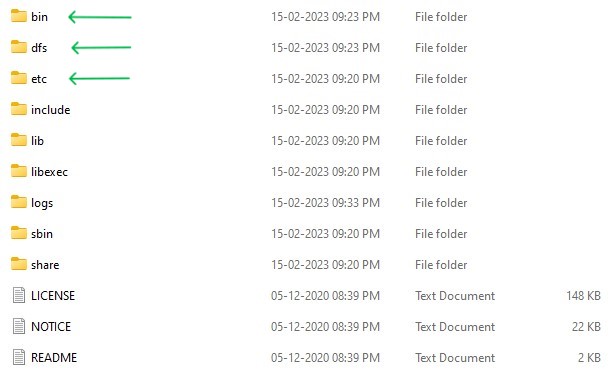
1. Download Spark:
   * Download the Spark distribution for Windows from the Apache website:

<https://archive.apache.org/dist/spark/spark-3.1.2/spark-3.1.2-bin-hadoop3.2.tgz>

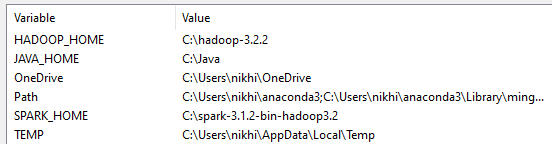
* + Once the download is complete, extract the contents of the file to a folder on your local machine.



1. Download Hadoop Dependencies:
   * Download the Hadoop dependencies file from this link: https://github.com/NikhilRaj-24/Hadoop-Necessaries/blob/main/hadoop-dependencies-3.2.2.zip
   * Extract the contents of the zip file to a folder on your local machine in the Hadoop folder in C drive.



1. Set Environment Variables:
   * Open the Windows Control Panel and go to System and Security -> System -> Advanced system settings.
   * Click on the "Environment Variables" button at the bottom of the window.
   * Under the "System Variables" section, click the "New" button and add the following variables:
     + Variable name: JAVA\_HOME Variable value: C:\Java
     + Variable name: HADOOP\_HOME Variable value: C:\hadoop-3.2.2
     + Variable name: SPARK\_HOME Variable value: C:\spark-3.1.2-bin-hadoop3.2
   * Edit the "Path" variable and add the following paths to the end of the existing value:
     + C:\Java\bin
     + C:\hadoop-3.2.2\bin
     + C:\hadoop-3.2.2\sbin
     + C:\spark-3.1.2-bin-hadoop3.2\bin
     + C:\spark-3.1.2-bin-hadoop3.2\sbin



Graphical user interface, text, application

Description automatically generated

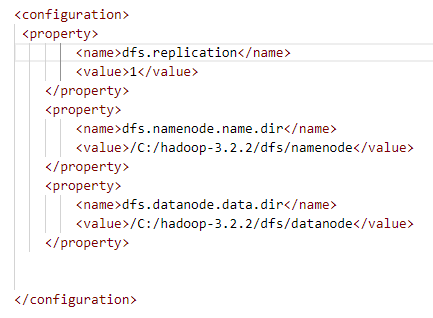
1. Configure etc folder:

* Open Hadoop folder and then open core-site.xml
* Edit the configuration (edit the value according to your path)

Text

Description automatically generated with medium confidence

* + Open hdfs-site.xml
  + Edit the configuration (edit the value according to your path)



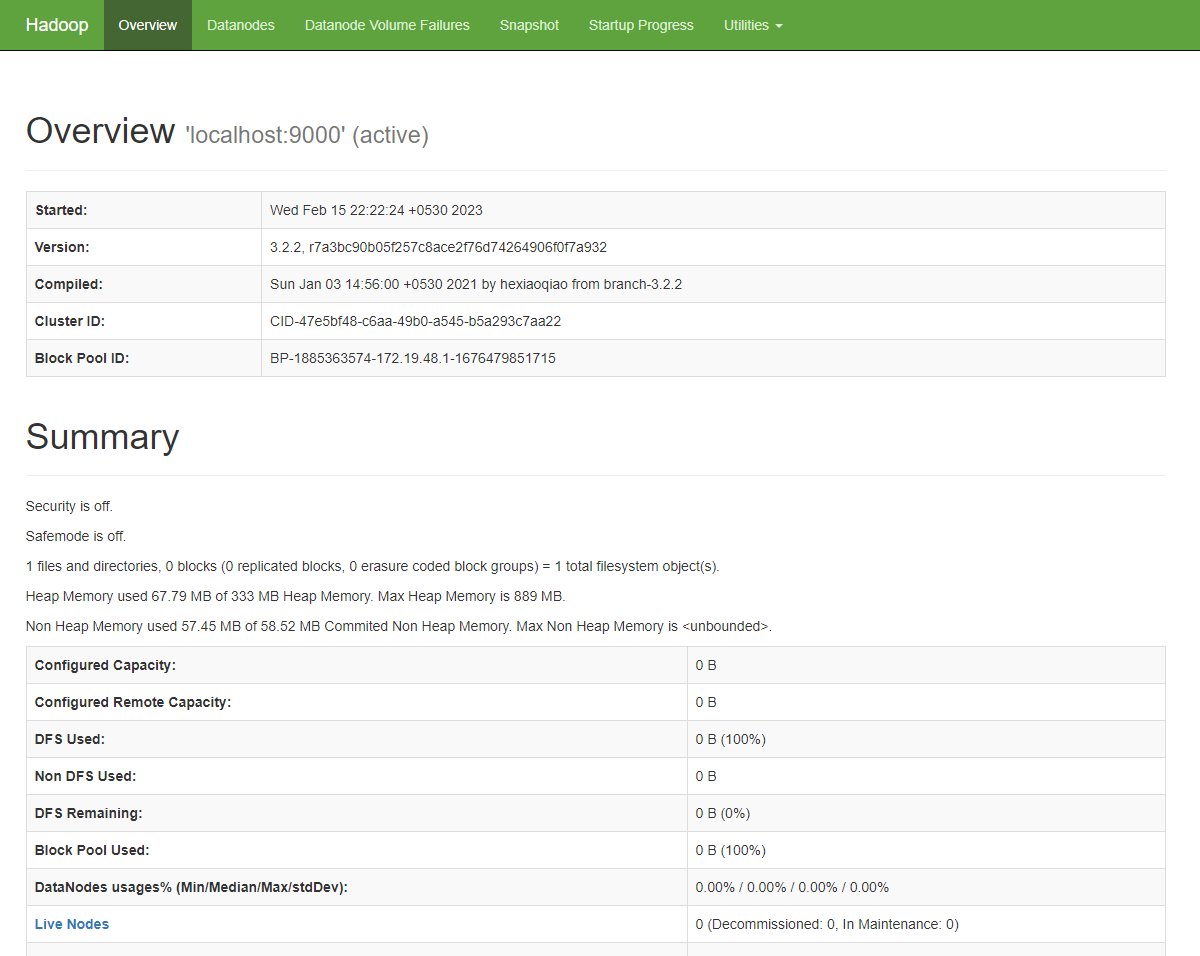
1. Format Hadoop Namenode:
   * Open the Command Prompt and navigate to the Hadoop bin directory (C:\hadoop-3.2.2\bin).
   * Run the following command to format the Hadoop Namenode:
     + hdfs namenode -format



1. Start Hadoop:
   * Open the Command Prompt and navigate to the Hadoop sbin directory (C:\hadoop-3.2.2\sbin – not really necessary, you can directly run the commands.)
   * Run the following command to start Hadoop:
     + start-dfs
   * Run the following command to start YARN:
     + start-yarn

Congratulations, you have successfully installed and started Hadoop on your Windows machine!

To open Hadoop, open chrome and type localhost:9870.



Documentation Team:

1. [@Nikhil Raj P - [CB.SC.I5DAS20023], Amrita Vishwa Vidyapeetham.](mailto:cb.sc.i5das20023@cb.students.amrita.edu)
2. [@Navami Vipith Kumar - [CB.SC.I5DAS20022], Amrita Vishwa Vidyapeetham.](mailto:cb.sc.i5das20022@cb.students.amrita.edu)
3. [@Arya C Prajeesh - [CB.SC.I5DAS20006], Amrita Vishwa Vidyapeetham.](mailto:cb.sc.i5das20006@cb.students.amrita.edu)
4. [@Navaneeth P - [CB.SC.I5DAS20122], Amrita Vishwa Vidyapeetham.](mailto:cb.sc.i5das20122@cb.students.amrita.edu)
5. [@Meera T Ramesh - [CB.SC.I5DAS20119], Amrita Vishwa Vidyapeetham.](mailto:cb.sc.i5das20119@cb.students.amrita.edu)
6. [@Aarsha Nair K S - [CB.SC.I5DAS20001], Amrita Vishwa Vidyapeetham.](mailto:cb.sc.i5das20001@cb.students.amrita.edu)

Facilitator:

1. [@Guru Prakash J - [CSE]](mailto:j_guruprakash@cb.amrita.edu)