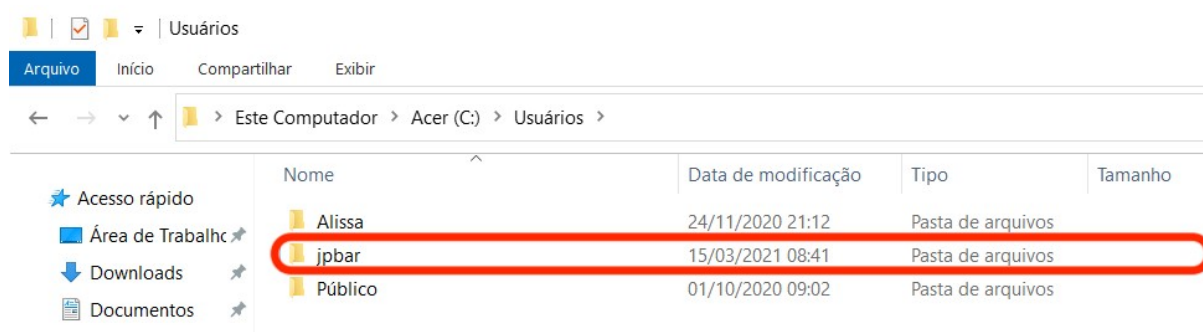


Instruções de Configuração – MapReduce/Hadoop

O Hadoop é uma tecnologia que pode ser executada em diferentes sistemas operacionais. Contudo, sua instalação no Windows requer etapas extra que não são demandadas em outros sistemas operacionais. Apesar de simples, qualquer detalhe de configuração mal realizado inviabilizará a implementação e execução de rotinas MapReduce. Ao ler esse tutorial, realize as etapas em seu computador seguindo, **sem exceções** as tratativas sugeridas. **Importante:** Caso seu sistema operacional seja baseado em UNIX (Linux/macOS), você deve iniciar este passo a passo com a Etapa 4 (download do projeto) e seguir diretamente para a Etapa 7.

Etapa 1: Verificação da pasta do usuário

- Abra o “Meu Computador” e acesse seu drive “C:/Usuários/” e encontre o seu usuário.
- Verifique se a pasta do seu usuário possui espaço.



- No exemplo acima, note que o usuário “jpbar” **não** possui nenhum espaço, e portanto, **funcionará** sem problemas.
- Caso seu usuário e o nome da pasta **possuam espaço(s)**, você **deverá** criar um novo usuário em seu computador sem espaços.
- Minha sugestão é criar um novo usuário chamado “hadoop” com permissão de Administrador para que você consiga executar tudo sem problemas.

Etapa 2: Instalação do Java 8















- Acesse o link a seguir e realize o download do Java JDK 8 de acordo com a arquitetura do seu computador.

Link: <https://www.oracle.com/br/java/technologies/javase/javase-jdk8-downloads.html>

- Atualmente, boa parte dos computadores e sistemas operacionais são 64 bits, portanto, prefira a versão “Windows x64”, apresentada em vermelho abaixo.

Java SE Development Kit 8u281

This software is licensed under the [Oracle Technology Network License Agreement for Oracle Java SE](#)

Product / File Description	File Size	Download
Linux ARM 64 RPM Package	59.1 MB	 jdk-8u281-linux-aarch64.rpm
Linux ARM 64 Compressed Archive	70.77 MB	 jdk-8u281-linux-aarch64.tar.gz
Linux ARM 32 Hard Float ABI	73.47 MB	 jdk-8u281-linux-arm32-vfp-hflt.tar.gz
Linux x86 RPM Package	108.46 MB	 jdk-8u281-linux-i586.rpm
Linux x86 Compressed Archive	136.95 MB	 jdk-8u281-linux-i586.tar.gz
Linux x64 RPM Package	108.06 MB	 jdk-8u281-linux-x64.rpm
Linux x64 Compressed Archive	137.06 MB	 jdk-8u281-linux-x64.tar.gz
macOS x64	205.26 MB	 jdk-8u281-macosx-x64.dmg
Solaris SPARC 64-bit (SVR4 package)	125.96 MB	 jdk-8u281-solaris-sparcv9.tar.Z
Solaris SPARC 64-bit	88.77 MB	 jdk-8u281-solaris-sparcv9.tar.gz
Solaris x64 (SVR4 package)	134.68 MB	 jdk-8u281-solaris-x64.tar.Z
Solaris x64	92.66 MB	 jdk-8u281-solaris-x64.tar.gz
Windows x86	154.69 MB	 jdk-8u281-windows-i586.exe
Windows x64	166.97 MB	 jdk-8u281-windows-x64.exe

Após o download, realize a instalação do Java.

Etapa 3: Instalação da biblioteca C++ 2020

- Acesse o link a seguir e realize o download do pacote C++ 2010.

Link: <https://www.microsoft.com/en-us/download/details.aspx?id=5555>

- Após o download, realize a instalação.

Etapa 4: Instalação do IntelliJ IDEA

- Acesse a página do IntelliJ IDEA e realize o download.

Link: <https://www.jetbrains.com/pt-br/idea/>

Atenção: ambas as versões Community ou Ultimate são suficientes. A versão Ultimate requer licença. Para obter a licença, é necessário realizar cadastro na JetBrains com sua conta institucional.

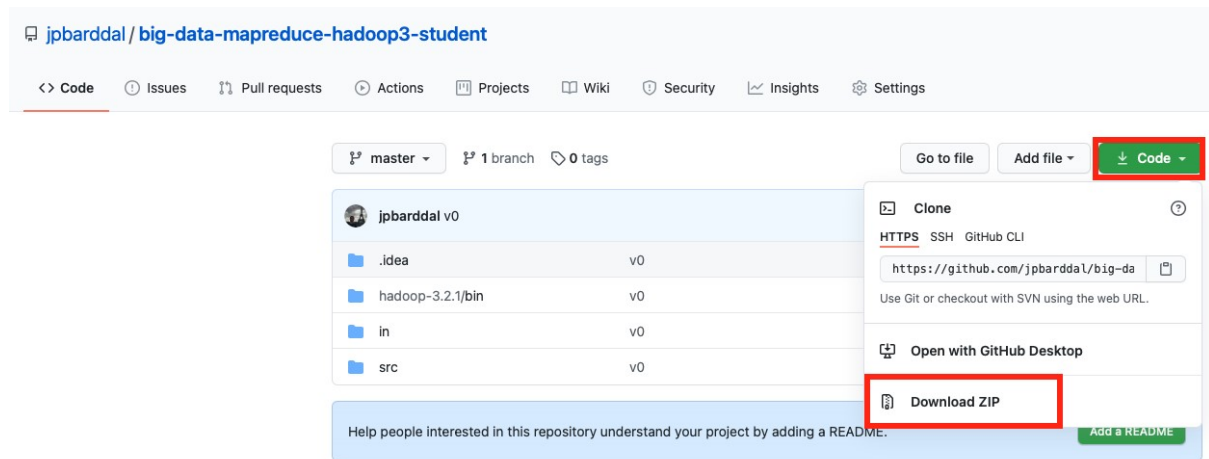
- Após o download, realize a instalação do IntelliJ IDEA.

Etapa 5: Download do projeto de MapReduce

- Realize o download do projeto de MapReduce no link abaixo.

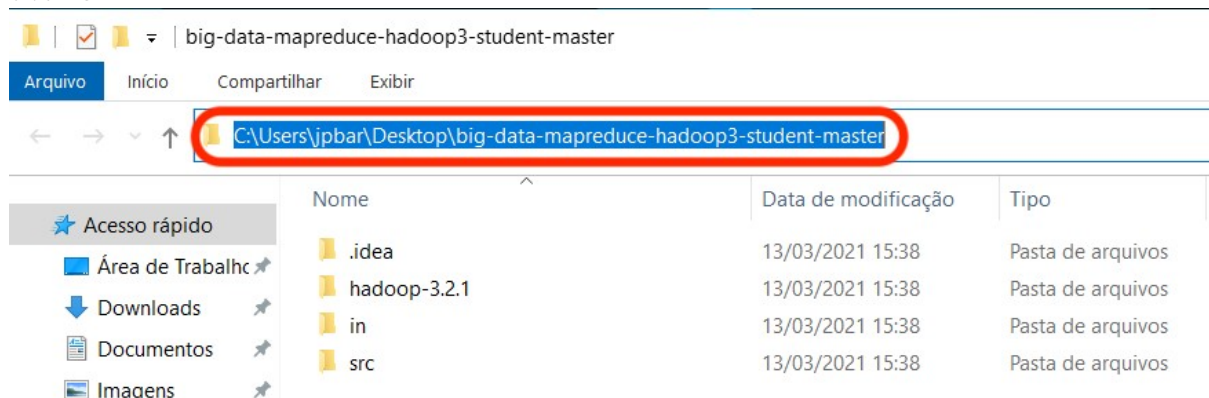
Link: <https://github.com/jpbarddal/big-data-mapreduce-hadoop3-student>

Para realizar o download, sugere-se selecionar a opção “Code > Download ZIP”, conforme representado na imagem abaixo. Caso prefira, sinta-se a vontade para realizar um fork.



- Extraia a pasta completa em uma pasta de sua preferência.

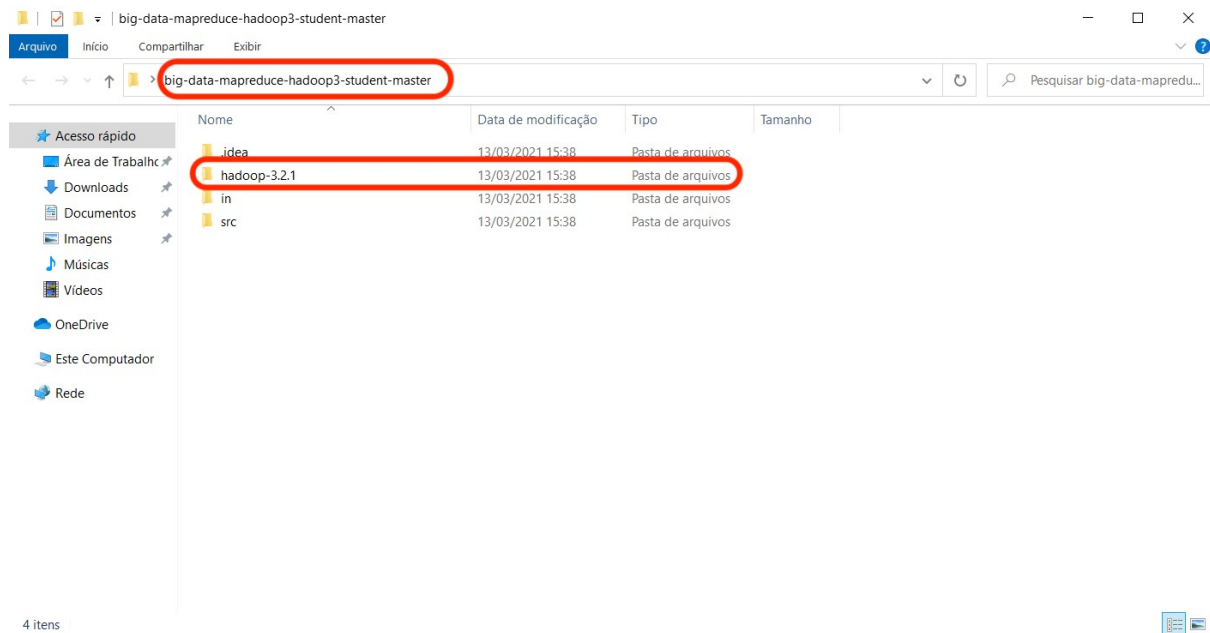
- Neste tutorial, a pasta foi colocada na “Área de Trabalho” (Desktop), conforme imagem abaixo.



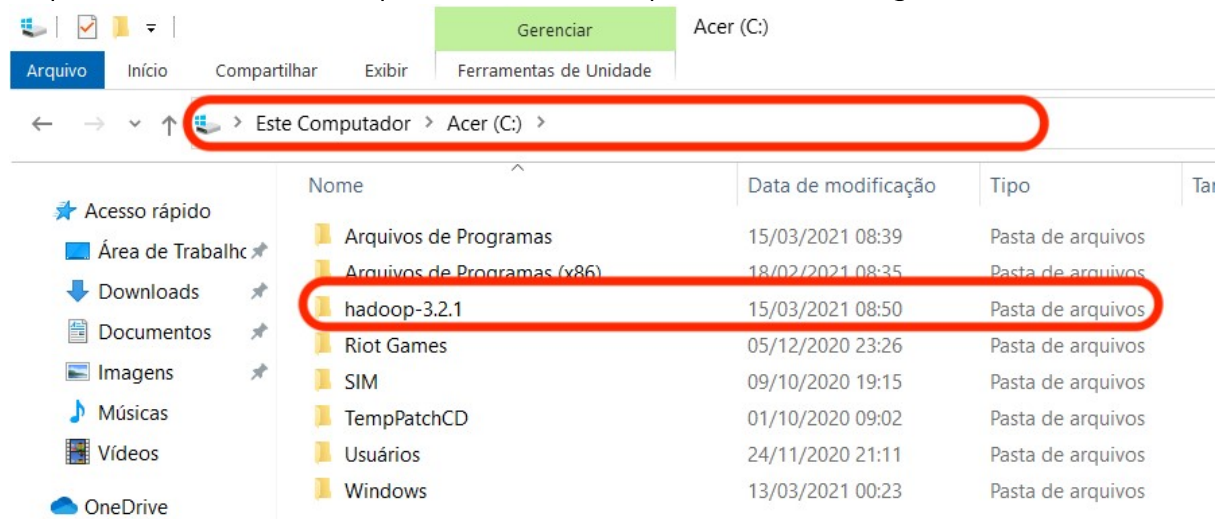
Nome	Data de modificação	Tipo
.idea	13/03/2021 15:38	Pasta de arquivos
hadoop-3.2.1	13/03/2021 15:38	Pasta de arquivos
in	13/03/2021 15:38	Pasta de arquivos
src	13/03/2021 15:38	Pasta de arquivos

Etapa 6: Configuração do Hadoop e Paths de Sistema

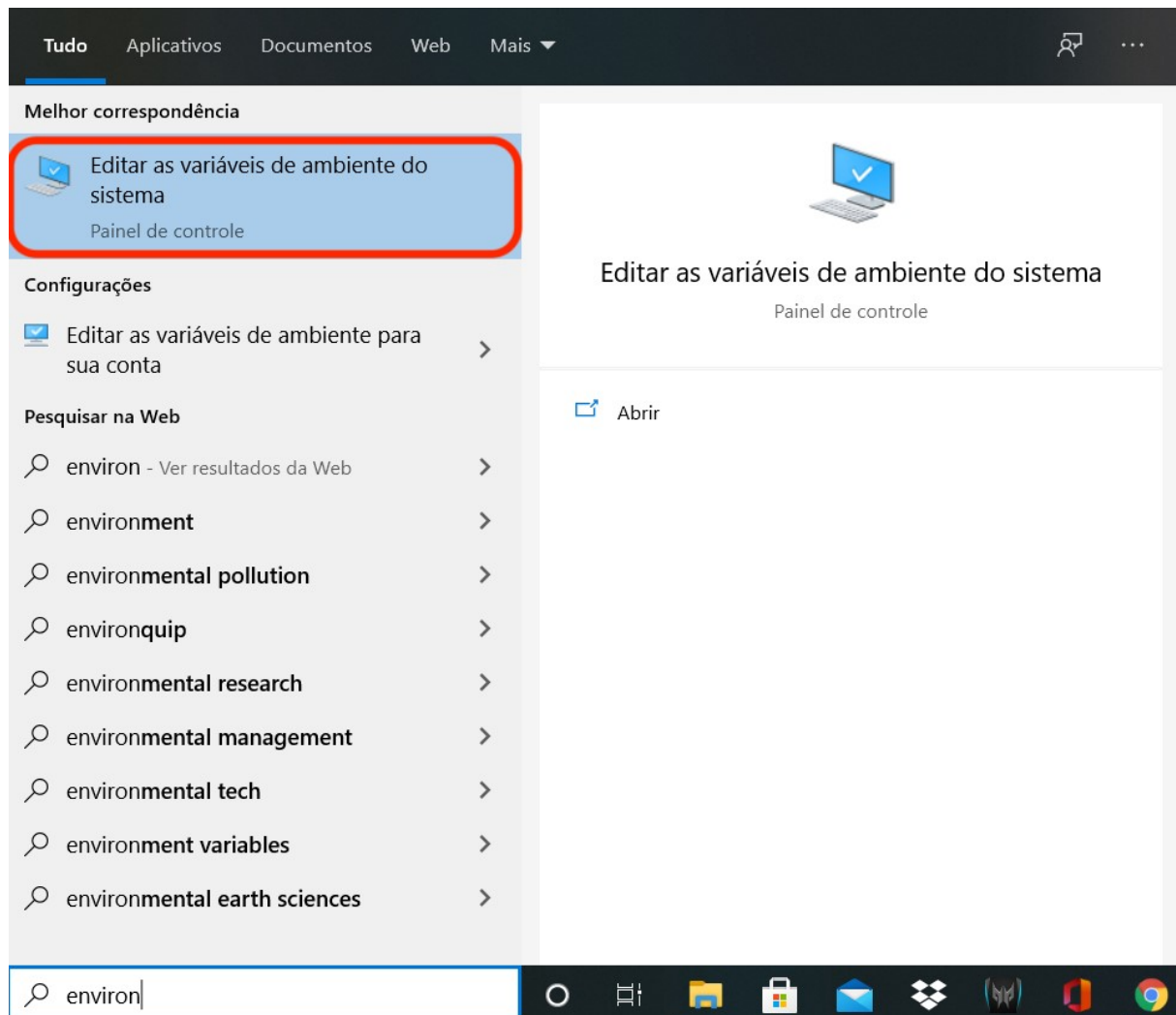
- Abra a pasta do projeto que foi extraída na etapa anterior. Haverá uma pasta com nome no formato “hadoop-X.Y.Z”. No momento da confecção deste tutorial, o nome exato da pasta é “hadoop-3.2.1”, contudo, isso pode mudar ao longo do tempo e conforme o Hadoop e o projeto são atualizados. A imagem abaixo representa a estrutura da pasta e enaltece a pasta “hadoop-3.2.1”.



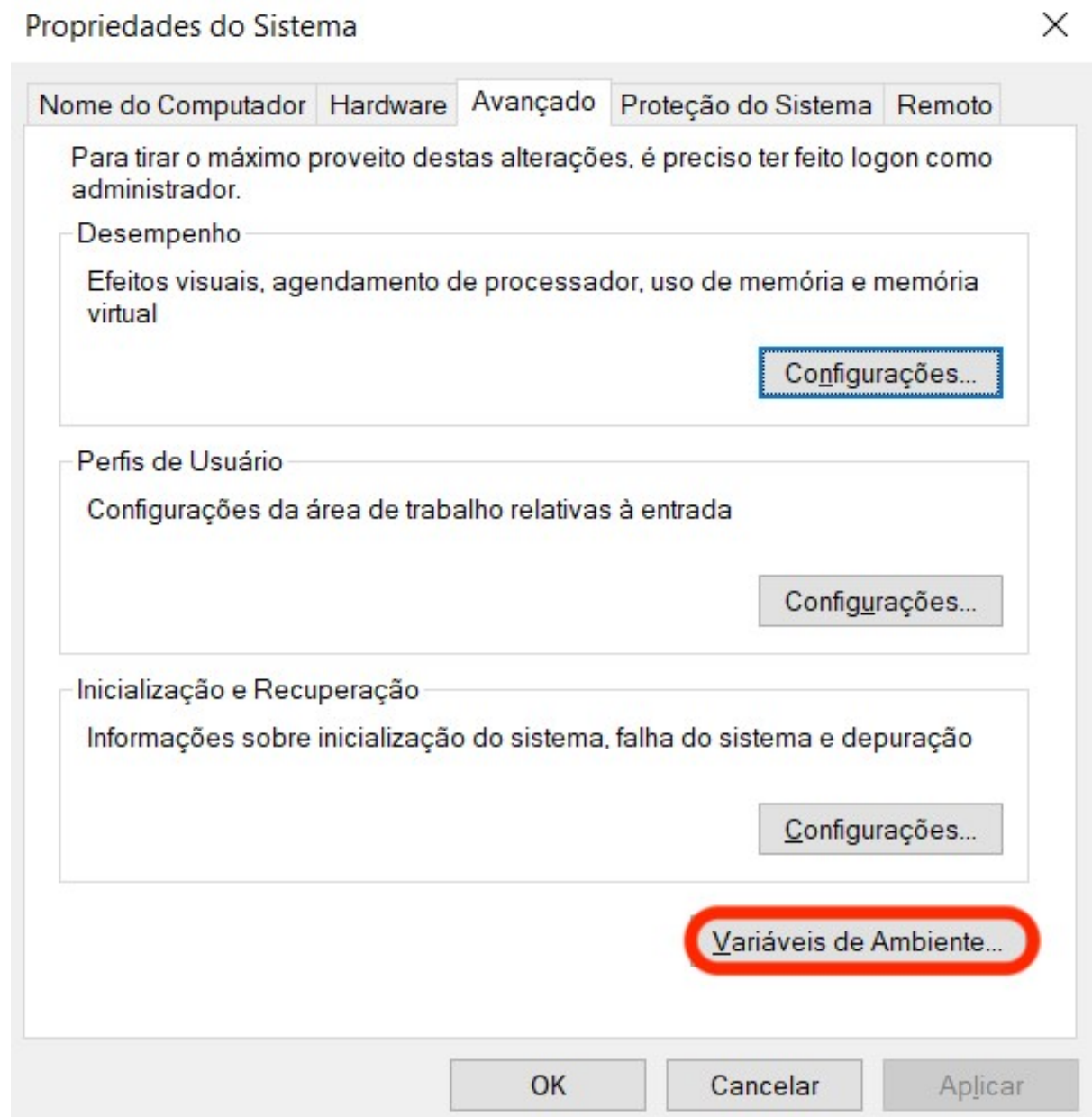
- Copie a pasta “hadoop-3.2.1” para o drive C:/ do seu computador. Desta forma, o caminho da pasta deve ser “C:/hadoop-3.2.1”, conforme apresentado na imagem abaixo:



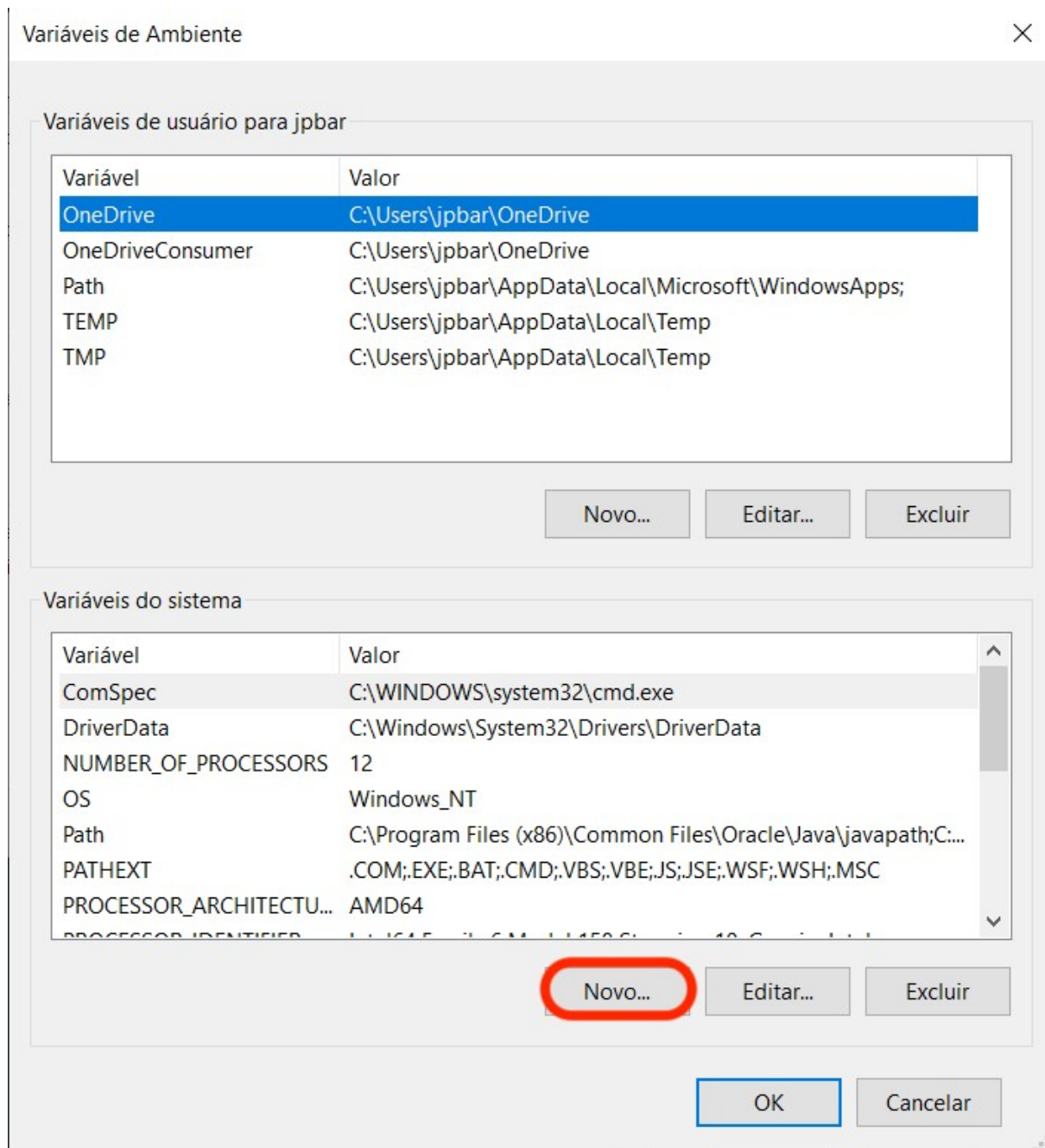
- Abra o menu “Iniciar” e digite “environ”. Você deverá encontrar e clicar na opção apresentada em vermelho na imagem abaixo.



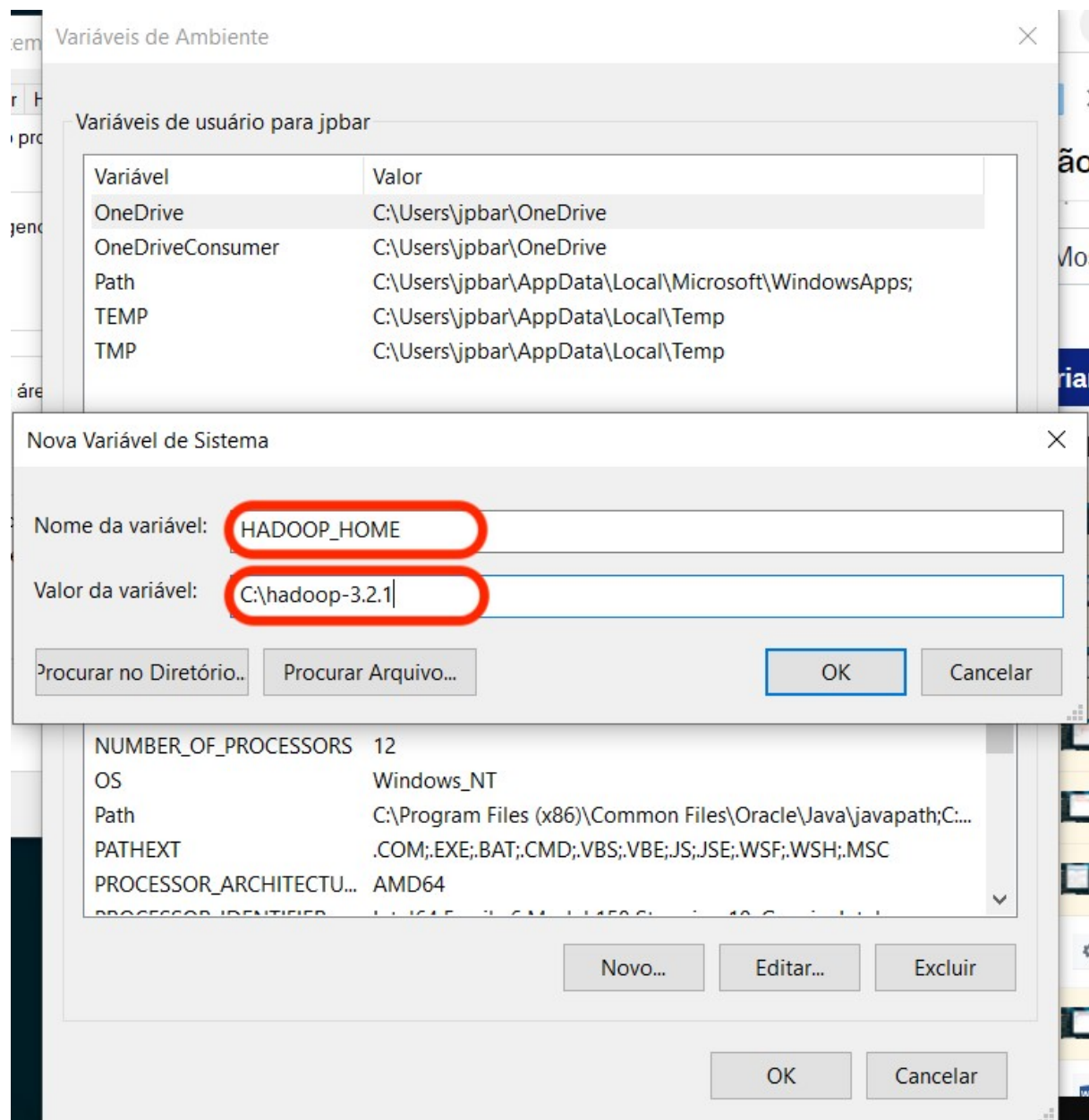
- Uma nova janela abrirá, e você deve clicar em “Variáveis de Ambiente...”.



- Posteriormente, na tela descrita na imagem abaixo, clique em “Novo...” na parte de “Variáveis do sistema”.

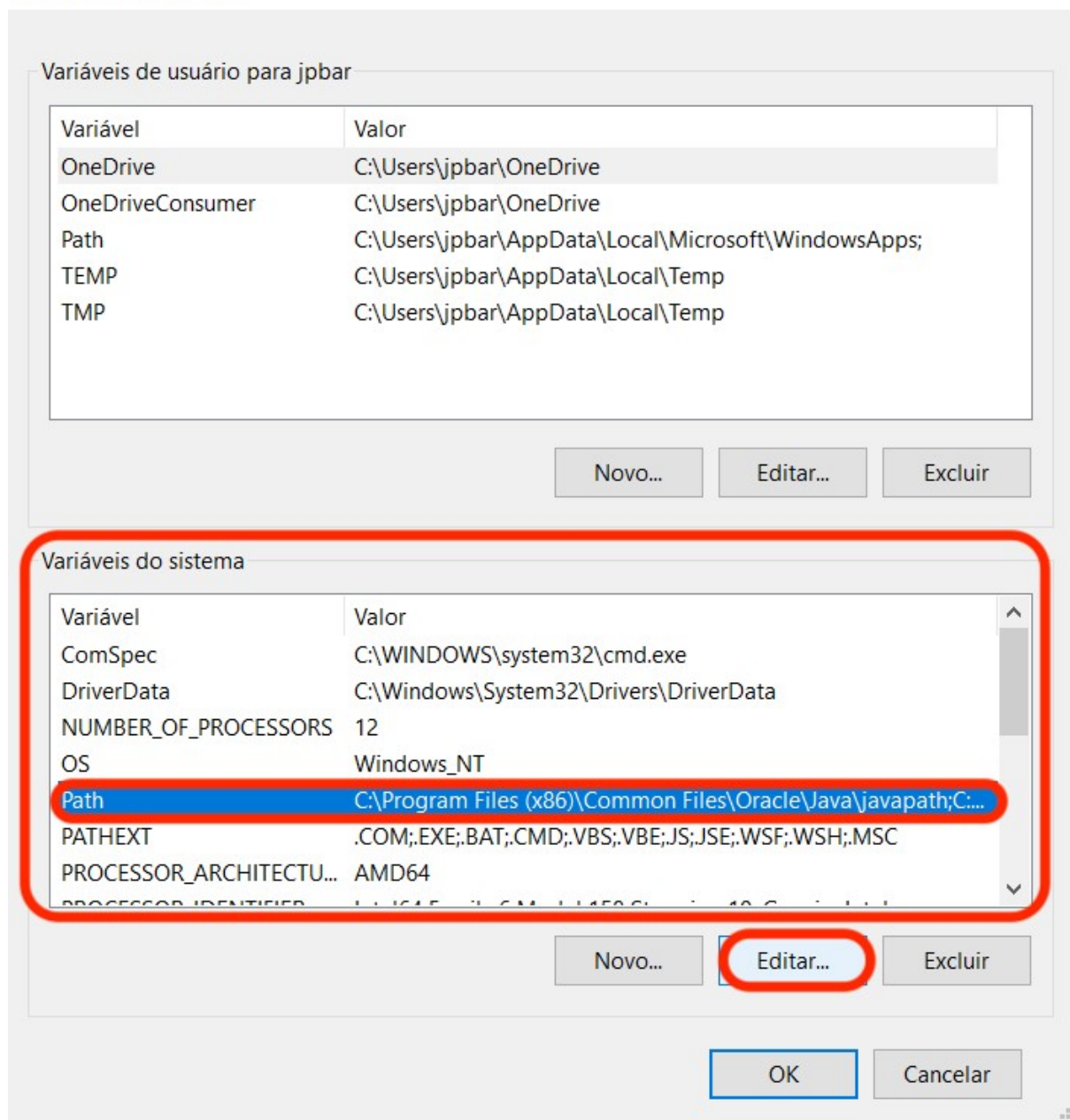


- Você deve inserir uma nova variável com nome “HADOOP_HOME”, e o caminho do diretório é a pasta do Hadoop, definido anteriormente em “C:/hadoop-X.Y.Z”. No exemplo, abaixo, o diretório é “C:/hadoop-3.2.1”.



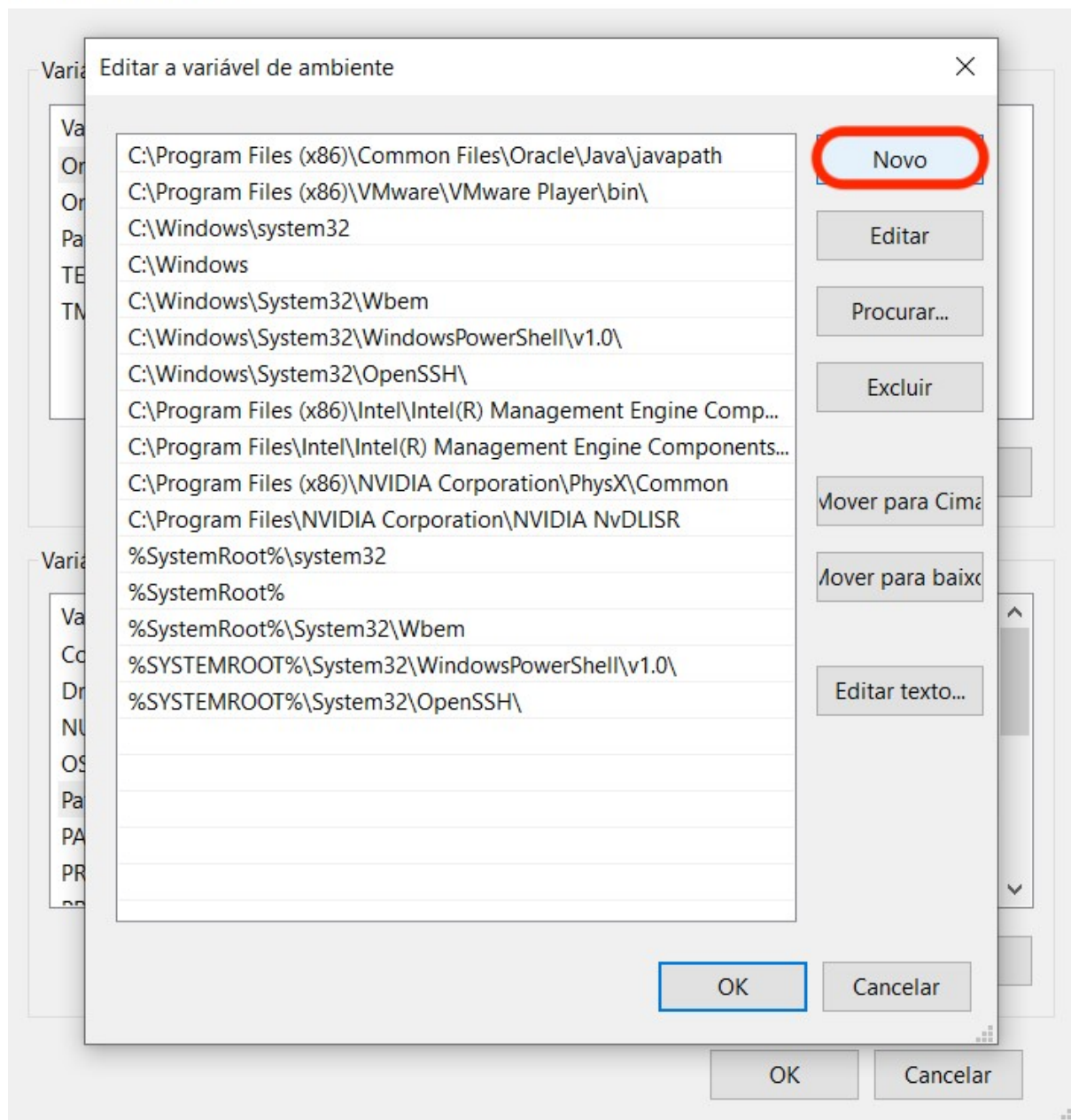
- Na sequência, selecione a variável “Path”, também no contexto de “Variáveis do sistema” e clique em “Editar...”.

Variáveis de Ambiente

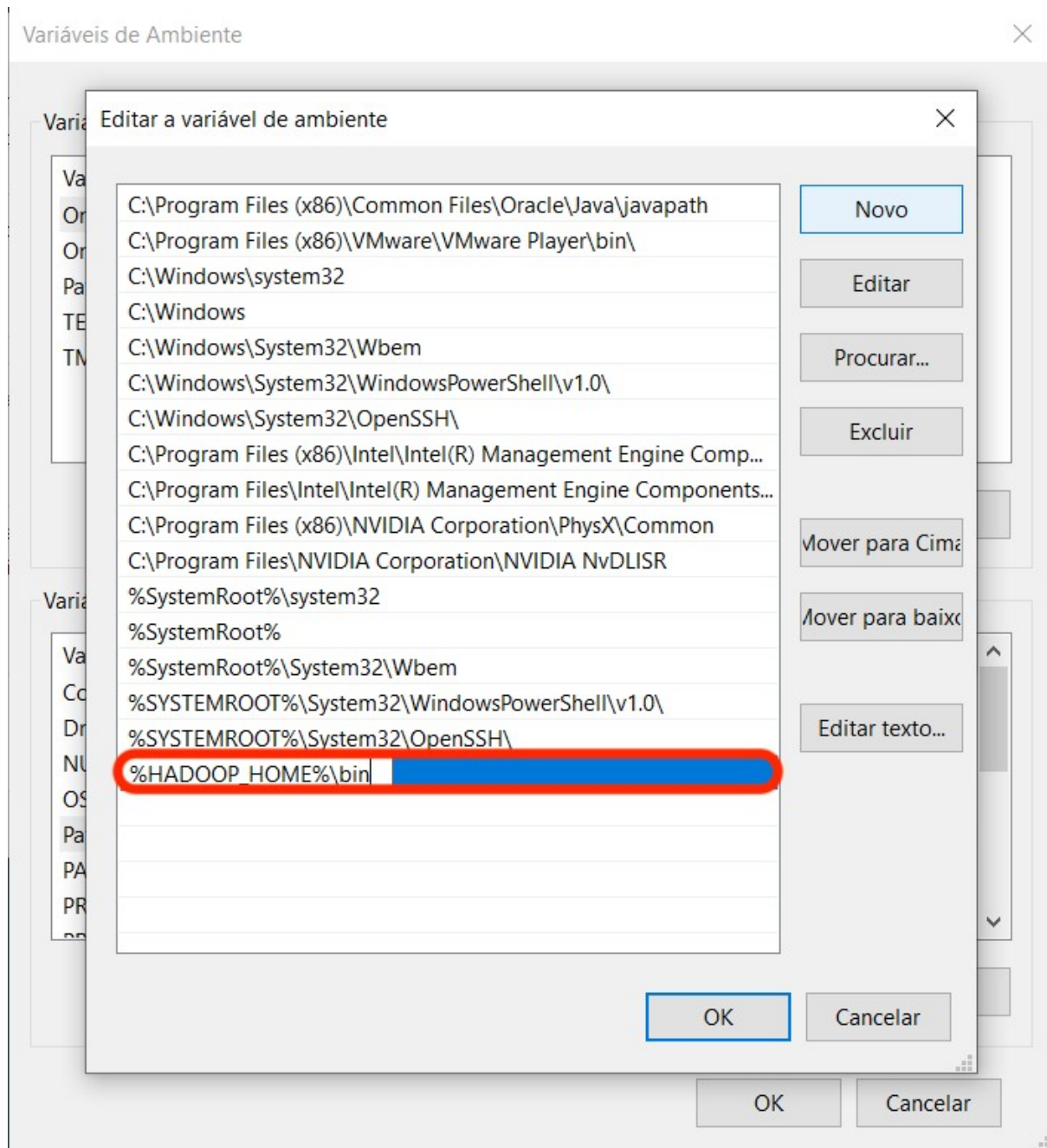


- Na janela que abrirá, clique em “Novo”:

Variáveis de Ambiente



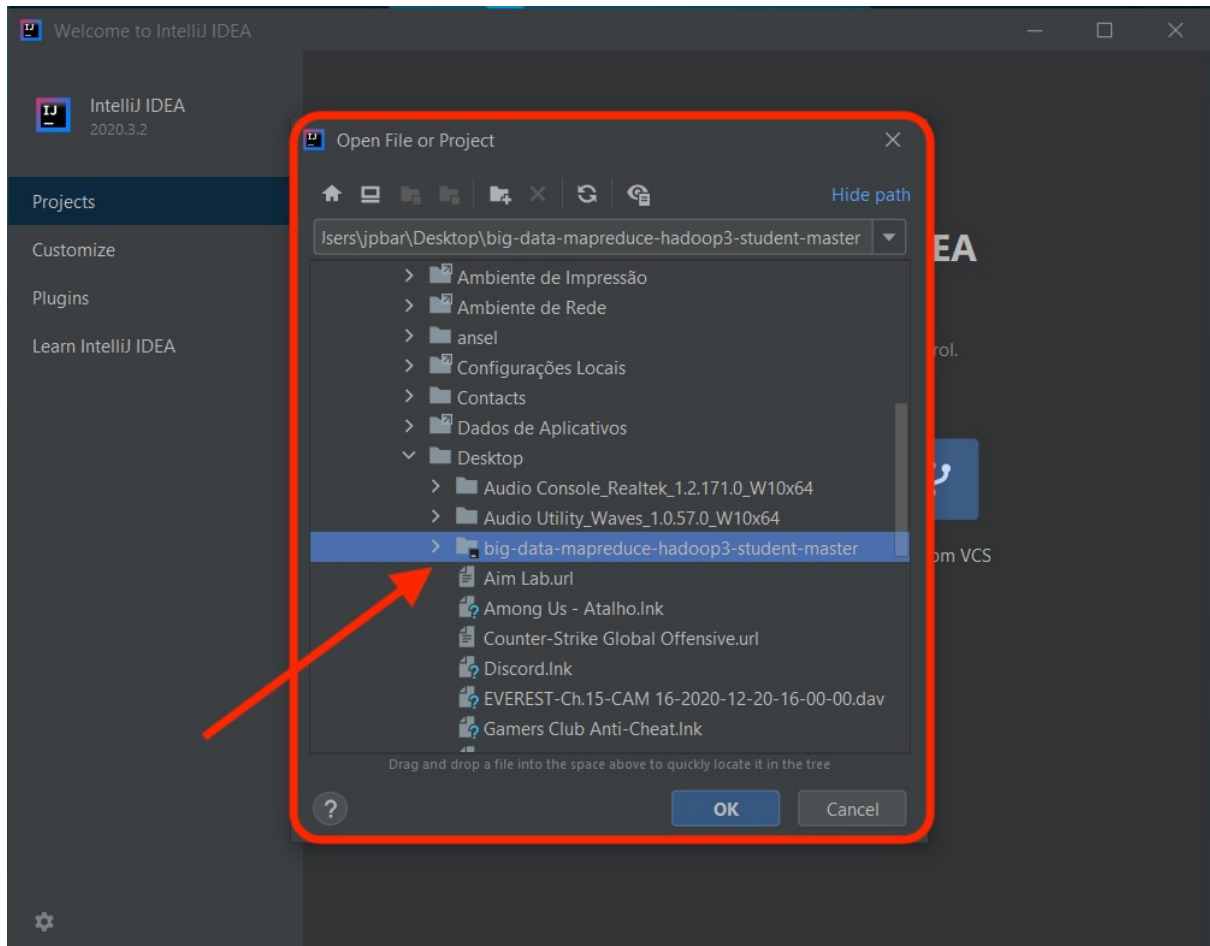
- No campo de edição de texto, adicione: “%HADOOP_HOME%\bin”, conforme apresentado na imagem abaixo.



- Dê “OK” em todas as janelas e reinicie o computador.

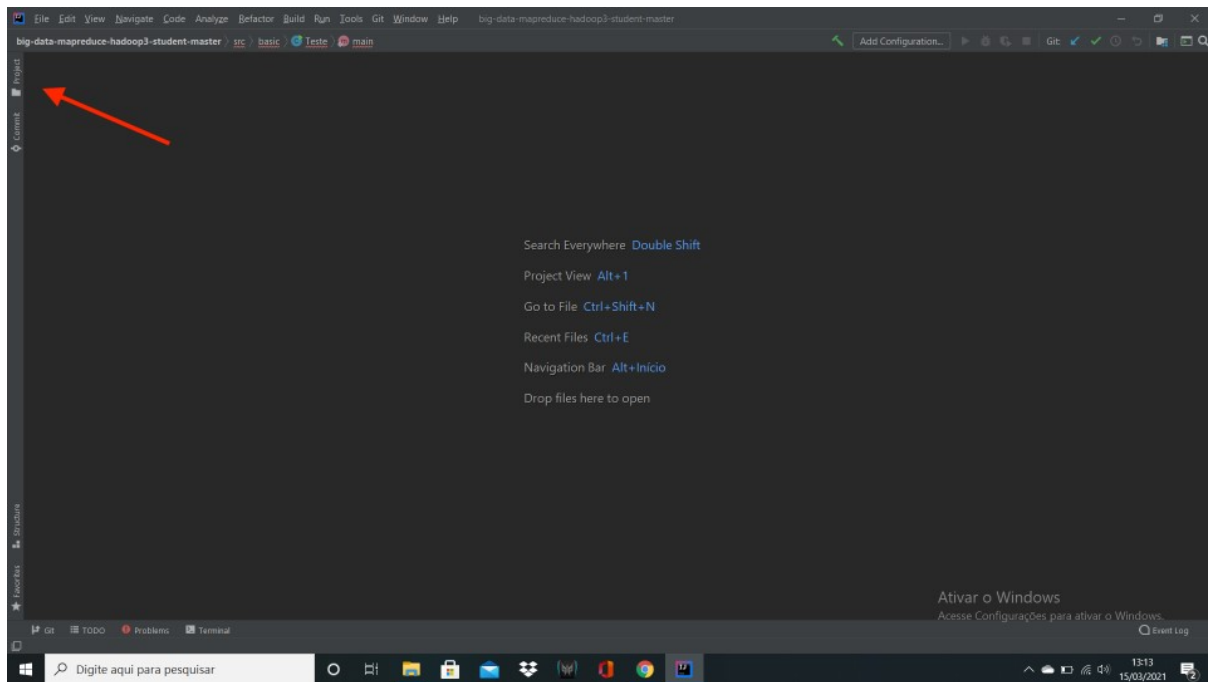
Etapa 7: Teste

- Inicialize o IntelliJ IDEA.
- Na janela abaixo, clique em Open e selecione a pasta do projeto. **Importante:** Note que a pasta do projeto possui um ícone diferenciado, uma vez que há um quadrado preto, diferentemente das demais pastas.

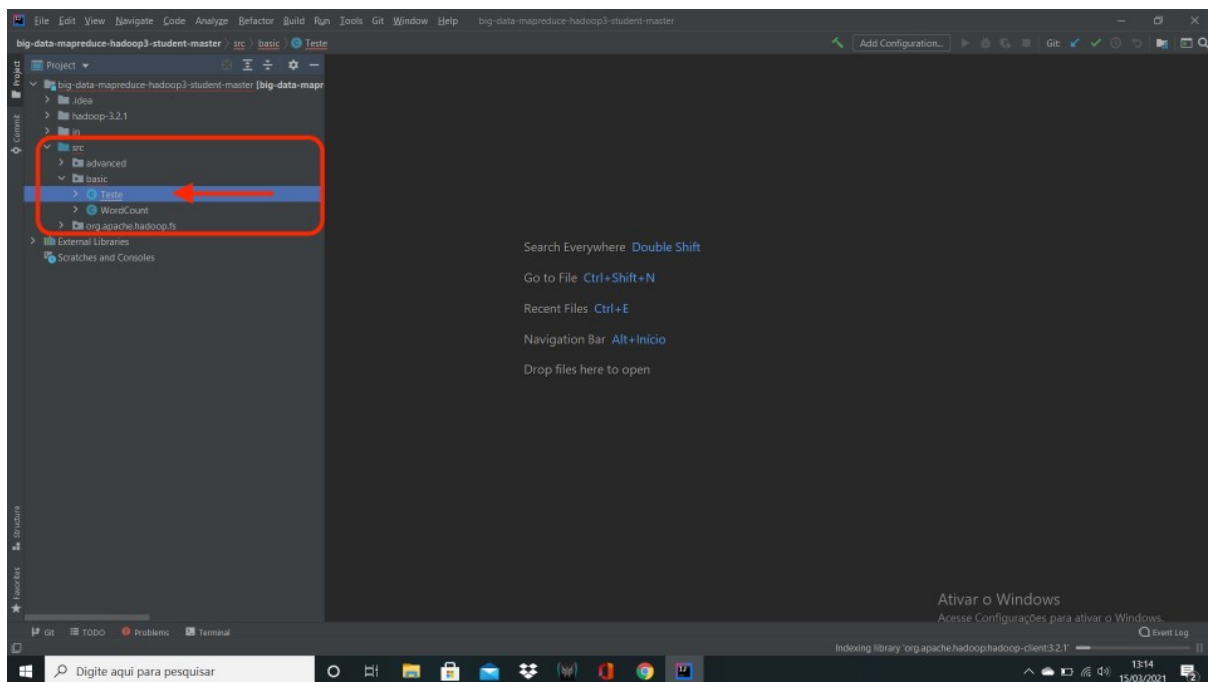


- Aguarde até que a barra de carregamento apresentada no canto inferior direito termine de carregar o Java.

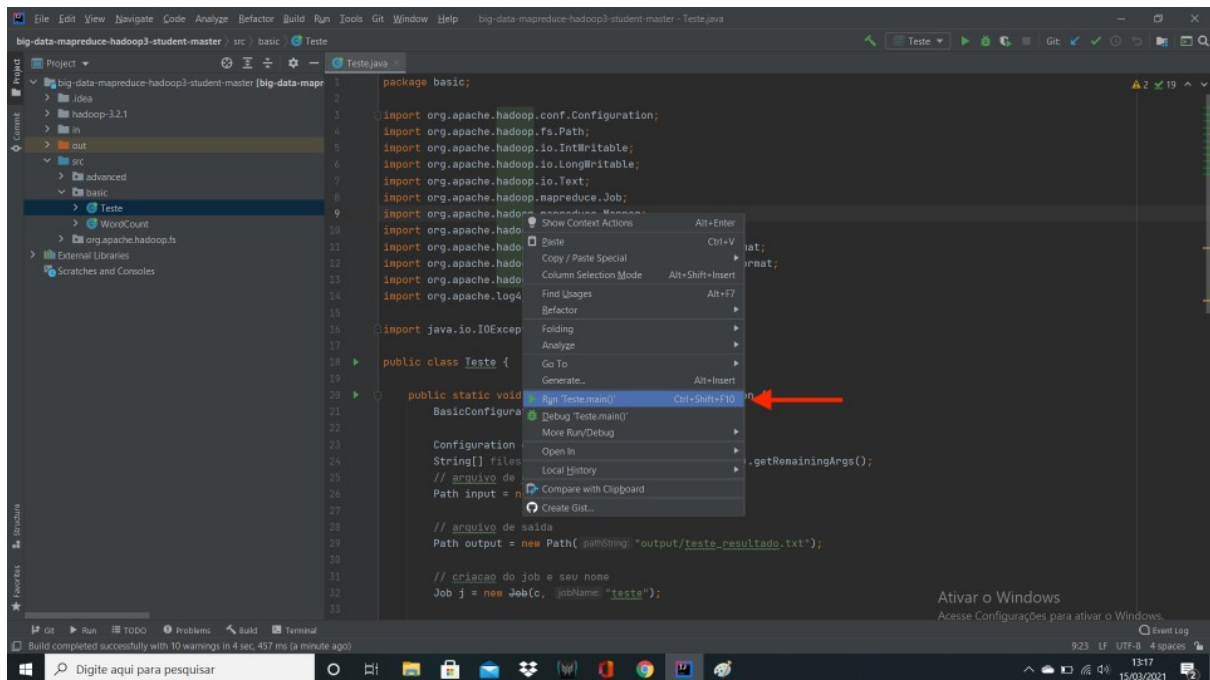
- Abra a aba de Projeto, disponível na parte esquerda da interface.



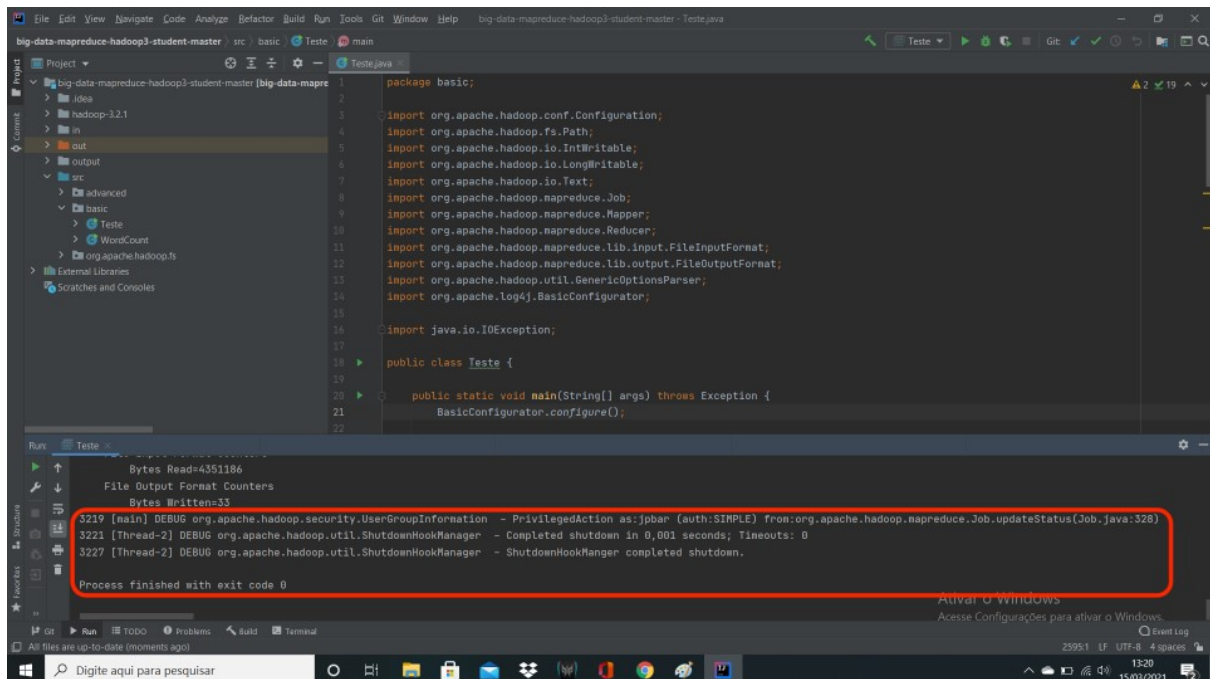
- Abra a classe “Teste.java”, disponível em “src > basic > Teste”.



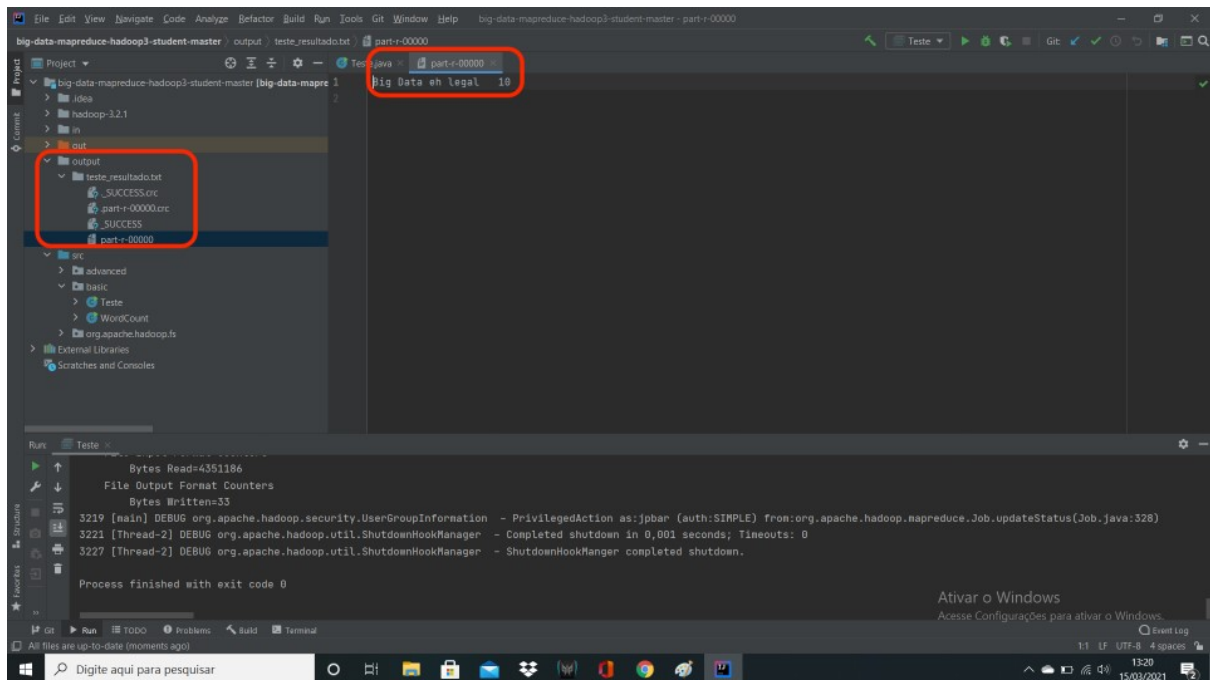
- Com o código-fonte da classe aberta, clique com o botão direito em cima de qualquer parte do código e clique em “Run”. A primeira vez tende a demorar um pouco mais pois o IntelliJ fará o download das bibliotecas necessárias do Hadoop.



- Verifique a saída apresentada no console. No final, você deve encontrar a mensagem apresentada abaixo, terminando com “Process finished with exit code 0” (o que significa que nenhum erro aconteceu):



- Finalmente, volte a aba do projeto e abra a pasta “output > teste_resultado.txt > part-r-0000.txt”, conforme descrito abaixo. Esse arquivo deve possuir o conteúdo “Big Data eh legal 10”.

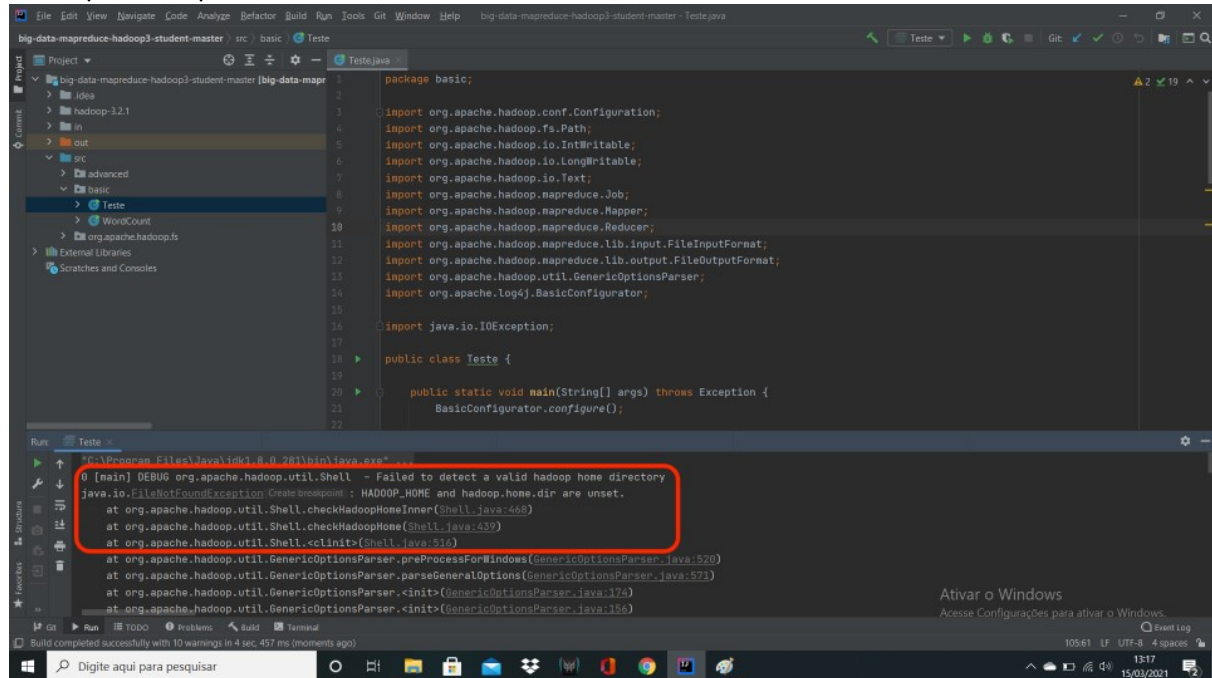


- Se você chegou a este ponto, parabéns, seu computador está configurado para executar MapReduce em modo local. Caso contrário, verifique as etapas anteriores com muita calma. Caso ainda assim você não consiga configurar o ambiente corretamente, verifique os erros comuns apresentados a seguir ou entre em contato com o professor.

Erros comuns

Erro 1: HADOOP_HOME

O erro mais comum é o erro apresentado na imagem abaixo. Caso isso ocorra, por favor, verifique a etapa 6 novamente.



The screenshot shows an IDE window with a Java file named `Teste.java`. The code is a simple test class that imports various Hadoop classes and has a `main` method that calls `BasicConfigurator.configure()`. The error message in the console is:

```
[main] DEBUG org.apache.hadoop.util.Shell - Failed to detect a valid hadoop home directory
java.io.FileNotFoundException: Create breakpoints: HADOOP_HOME and hadoop.home.dir are unset.
at org.apache.hadoop.util.Shell.checkHadoopHomeInner(Shell.java:468)
at org.apache.hadoop.util.Shell.checkHadoopHome(Shell.java:439)
at org.apache.hadoop.util.Shell.<clinit>(Shell.java:516)
at org.apache.hadoop.util.GenericOptionsParser.preProcessForWindows(GenericOptionsParser.java:520)
at org.apache.hadoop.util.GenericOptionsParser.parseGeneralOptions(GenericOptionsParser.java:571)
at org.apache.hadoop.util.GenericOptionsParser.<init>(GenericOptionsParser.java:174)
at org.apache.hadoop.util.GenericOptionsParser.<init>(GenericOptionsParser.java:186)
```

The error is highlighted with a red box. The IDE interface also shows the project structure on the left and the Windows taskbar at the bottom.