

LAB 01 – Cómo ejecutar un programa MapReduce

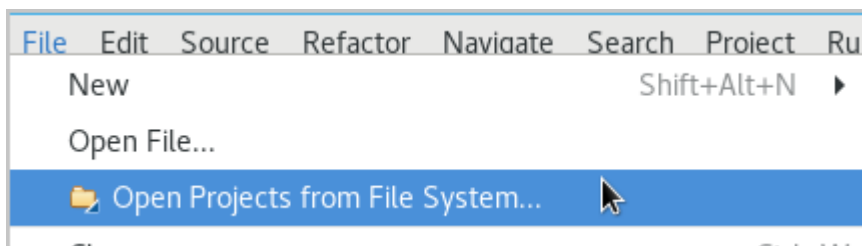
Paso 1: antes de comenzar

Asegúrese de tener Eclipse y Java 1.8 instalados. Si aún no lo ha hecho, está disponible en el área de Descargas.

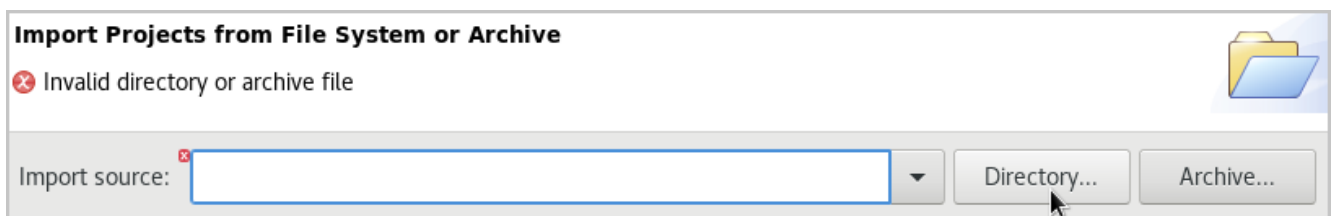
Paso 2: configure un proyecto de Eclipse con las bibliotecas de Hadoop

Para comenzar con MapReduce, deberá configurar un proyecto Eclipse con las Bibliotecas Hadoop. Para este curso, le proporcionamos una plantilla de proyecto Eclipse que incluye estas bibliotecas listas para usar.

1. Obtenga el archivo MapReduce.zip (55Mb), el cual contiene los archivos del proyecto para Eclipse
2. Extraiga el archivo .ZIP en la raíz del Disco, Esto debería extraerse como una sola carpeta llamada MapReduce.
3. Abra su espacio de trabajo de Eclipse
4. Utilice el menú **Abrir proyecto desde el Sistema de Archivos** desde el menú **Archivo**

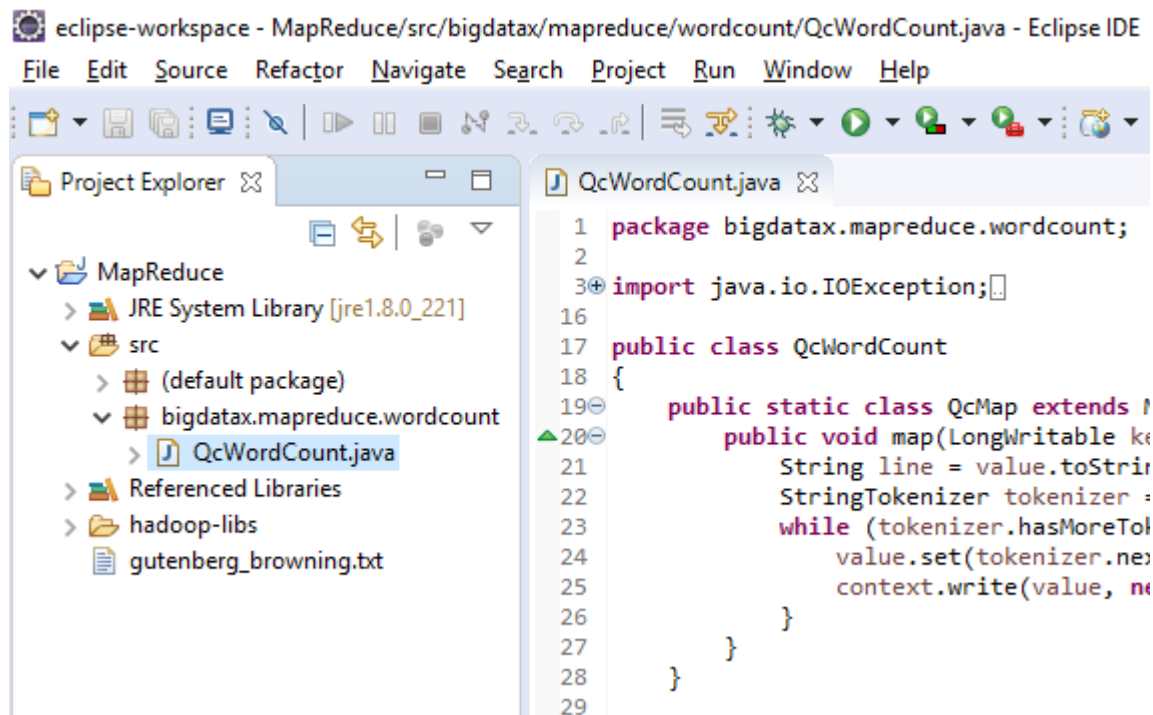


5. En el cuadro de dialogo que aparece, de click en el botón Directorio



6. En el cuadro de dialogo que aparece, navegue y seleccione la carpeta generada en el paso #2 y presione Finalizar.

Este proyecto de plantilla incluye varios archivos Stub que puede usar al hacer sus propios programas MapReduce.

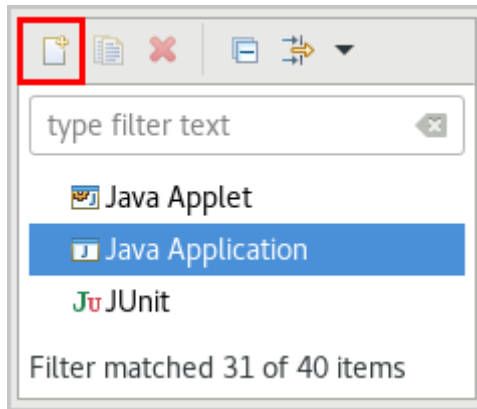


Paso 3: Configure Eclipse para ejecutar el ejemplo QcWordCount

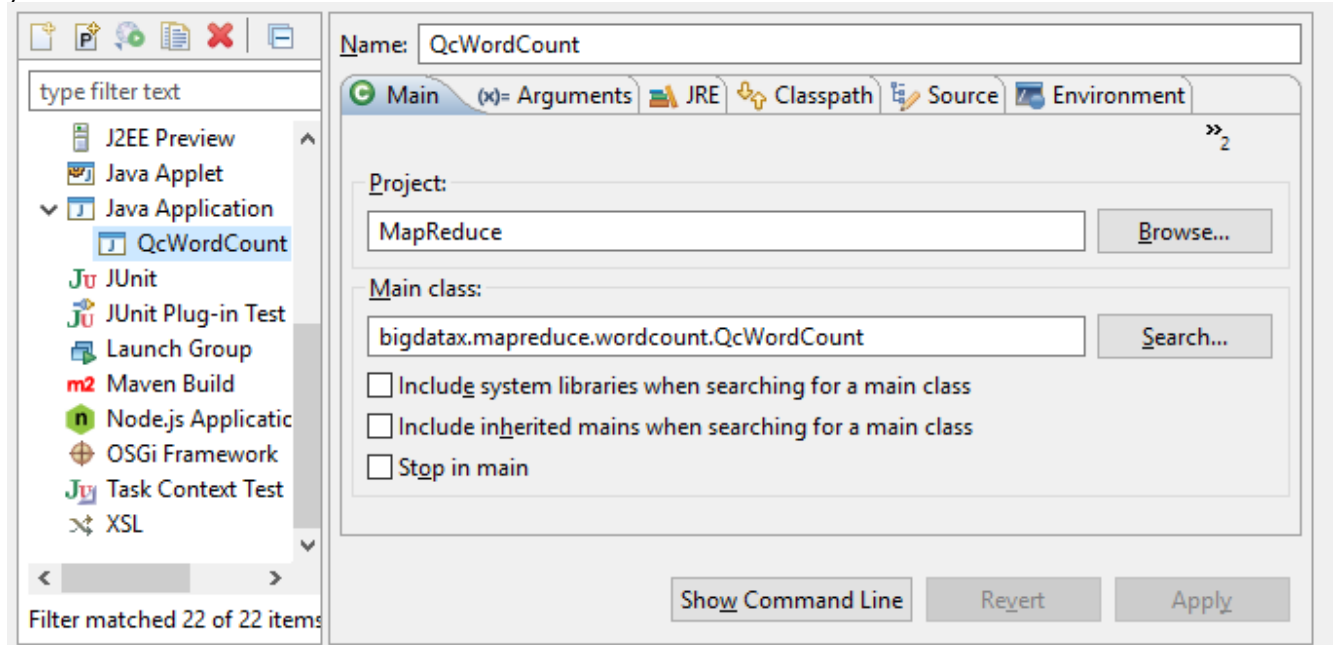
El siguiente paso es configurar Eclipse para ejecutar el ejemplo QcWordCount. Para hacer esto:

1. Vaya a **src/bigdatax.mapreduce.wordcount** y de click derecho en el archivo **QcWordCount.java**, y en el menú contextual que aparece, seleccione Run As -> Run Configurations.
2. En el cuadro de dialogo que aparece, seleccione **Java Application** en el panel izquierdo y de click en el botón **New**.





3. Algunas nuevas opciones aparecerán en el medio. Los valores predeterminados ya deberían ser correctos:



4. Navegue hasta el tab **Arguments** y en la entrada de texto *program arguments* ingrese **gutenberg_browning.txt output**.

Nuestro programa de ejemplo toma 2 argumentos de línea de comando.

gutenberg_browning.txt es la ruta a un archivo de texto que contiene los trabajos completos de Robert Browning; Las palabras que debe contar nuestro programa. *output* es el nombre de una carpeta que se creará y que contendrá el resultado final de nuestro recuento de palabras.

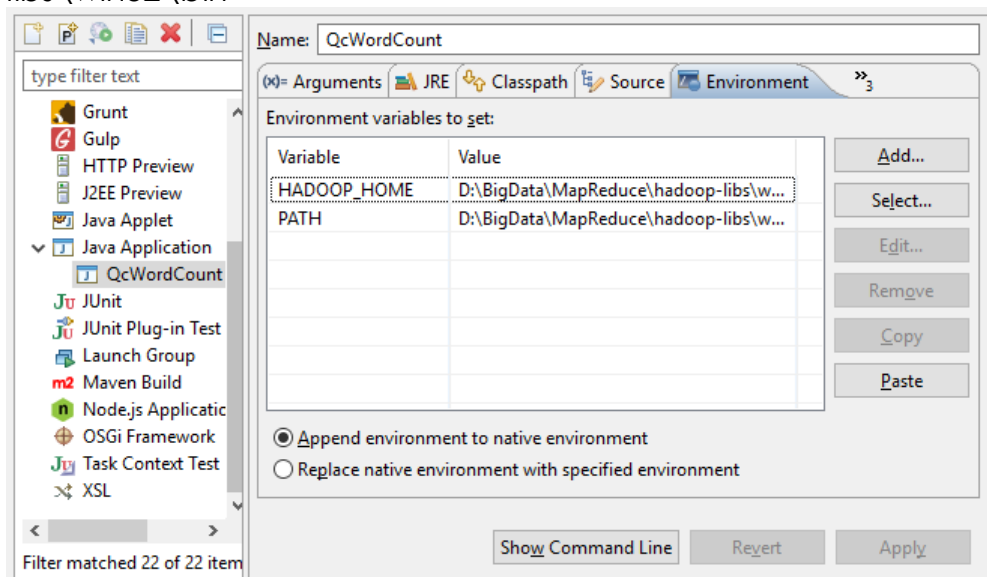
5. Seleccione Apply



Usuarios de Windows

Si tiene Windows instalado, deberá configurar algunas opciones adicionales:

1. Navegue al Tab Environment, Seleccione New.
2. Establezca Nombre en HADOOP_HOME y Valor, la ubicación COMPLETA de las bibliotecas de Windows Hadoop adecuadas para su sistema, luego elija Aceptar:
 - a. Las librerías de Windows Hadoop están ubicadas en la carpeta hadoop-libs.
 - b. Si está ejecutando windows de 64-bit con Java de 64-bit, debe utilizar la carpeta win64.
 - c. Si está ejecutando windows de 32-bit con Java de 32-bit, debe utilizar la carpeta win32.
 - d. Por ejemplo, si esta ejecutando windows de 32-bit y ha descrompimido el Proyecto en C:\Users\YourUsername\eclipse-workspace\MapReduce, entonces el valor que debe ingresar es C:\Users\YourUsername\eclipse-workspace\MapReduce\hadoop-libs\win32
3. Agregue otra variable con el nombre **PATH** y el Valor con el mismo valor que fue establecido para **HADOOP_HOME**, pero con **\bin** al final. Por ejemplo, if C:\Users\YourUsername\eclipse-workspace\MapReduce\hadoop-libs\win32 fue el valor ingresado para HADOOP_HOME, el valor de PATH debe ser C:\Users\YourUsername\eclipse-workspace\MapReduce\hadoop-libs\win32\bin



4. Seleccione la opción 'Append environment to native environment' (como se muestra arriba).
5. Click en Apply.

Paso 4: Ejecute el ejemplo QcWordCount

Ahora está listo para ejecutar el ejemplo QcWordCount. Para hacer esto:

1. Presione el botón **Run** si tiene abierto el cuadro de dialogo **Run Configurations**, de lo contrario click derecho en QcWordCount.java, y desde el menú que aparece seleccione **Run As -> Java Application**.
2. Ahora podrá encontrar un resultado como un archivo de texto llamado part-r-00000 en la carpeta output.
3. Si intenta ejecutar multiples veces, asegurese de borrar la carpeta Output entre cada ejecución del programa, de lo contrario puede recibir un mensaje de error similar al siguiente {PATH_TO_OUPUT_FOLDER} already exists.

