|  |
| --- |
| Máster en Big Data |
| Tecnologías de Almacenamiento |
| 1. Hands-On: Desarrollo MapReduce |

Índice

[1. Introducción 3](#_Toc506891313)

[2. Entorno de desarrollo 3](#_Toc506891314)

[3. IpCount 5](#_Toc506891315)

[4. Avarage Length 5](#_Toc506891316)

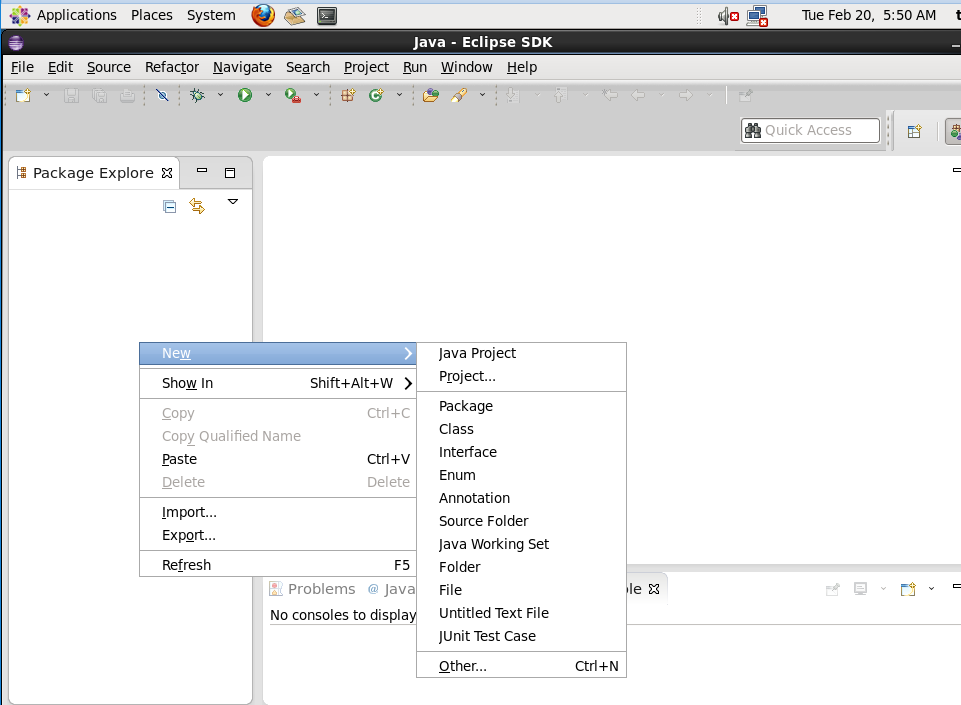
# Introducción

El objetivo de este Hands-On es realizar el desarrollo, en Java, de dos Jobs sencillos de MapReduce y ejecutarlos.

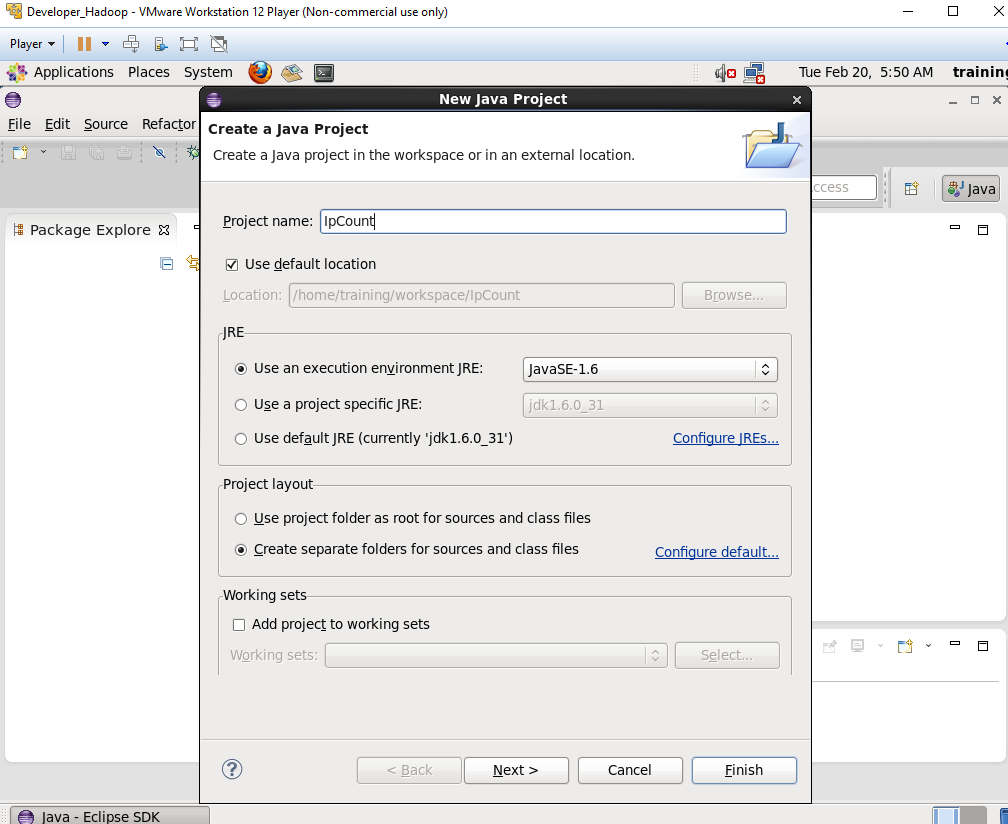
# Entorno de desarrollo

Para realizar el desarrollo lo haremos mediante el IDE Eclipse de la máquina virtual importada en ejercicios anteriores.

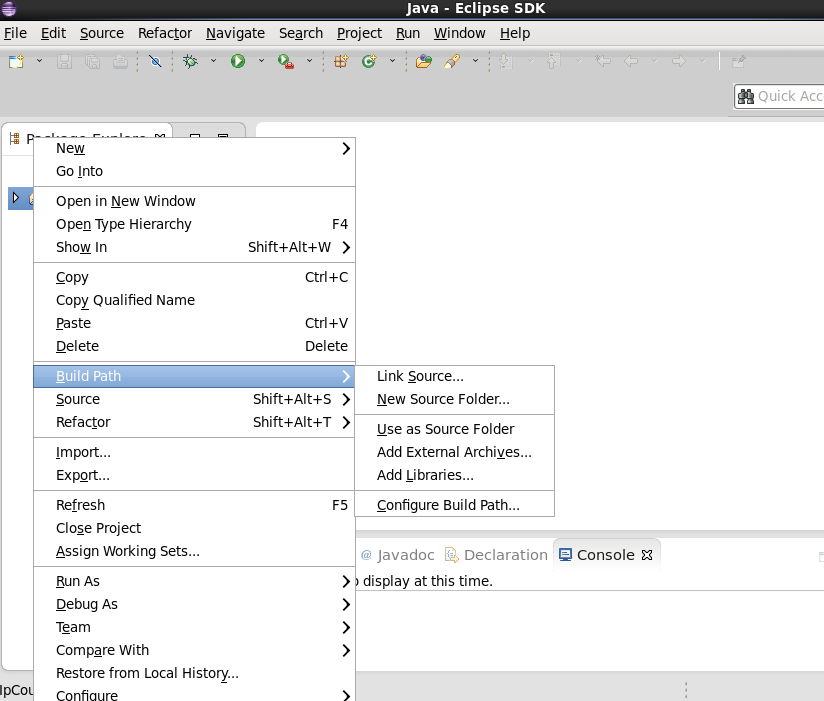
Para crear un nuevo proyecto, haremos click derecho sobre el package explorer New 🡪 Java Project



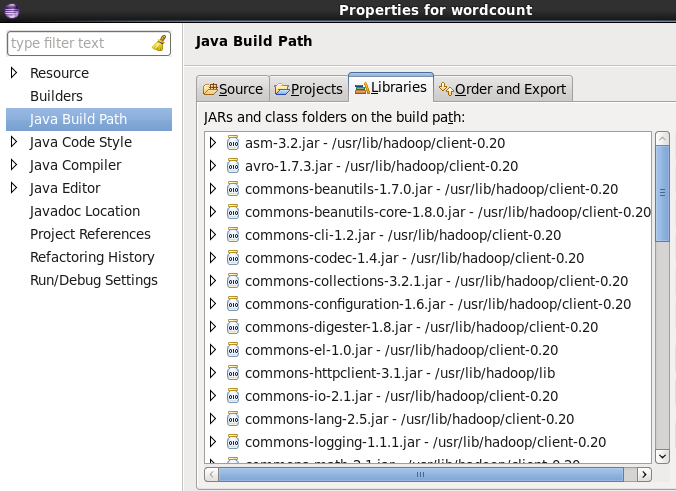
Introducimos el nombre del proyecto y click en Finish



Importamos manualmente las lilbrerías necesarias haciendo click derecho sobre el proyecto que acabamos de crear y seleccionamos Build Path 🡪 Configure Build Path



En la pestaña de libraries, seleccionamos Add Extertnal Jars e importamos todo el contenido de la carpeta /usr/lib/hadoop/client-0.20/



# IpCount

Desarrollar y ejecutar el siguiente MapReduce:

Contar el número de veces que una misma Ip hace una petición en los logs contenidos en /weblog/ importado anteriormente

# Avarage Length

Desarrollar y ejecutar el siguiente MapReduce:

Calcular el promedio de la longitud de las palabras que empiecen por una letra determinada en todas las obras de Shakespeare importadas anteriormente

Ej de salida :

*N 2.0*

*n 3.0*

*d 10.0*

*i 2.0*

*t 3.5*