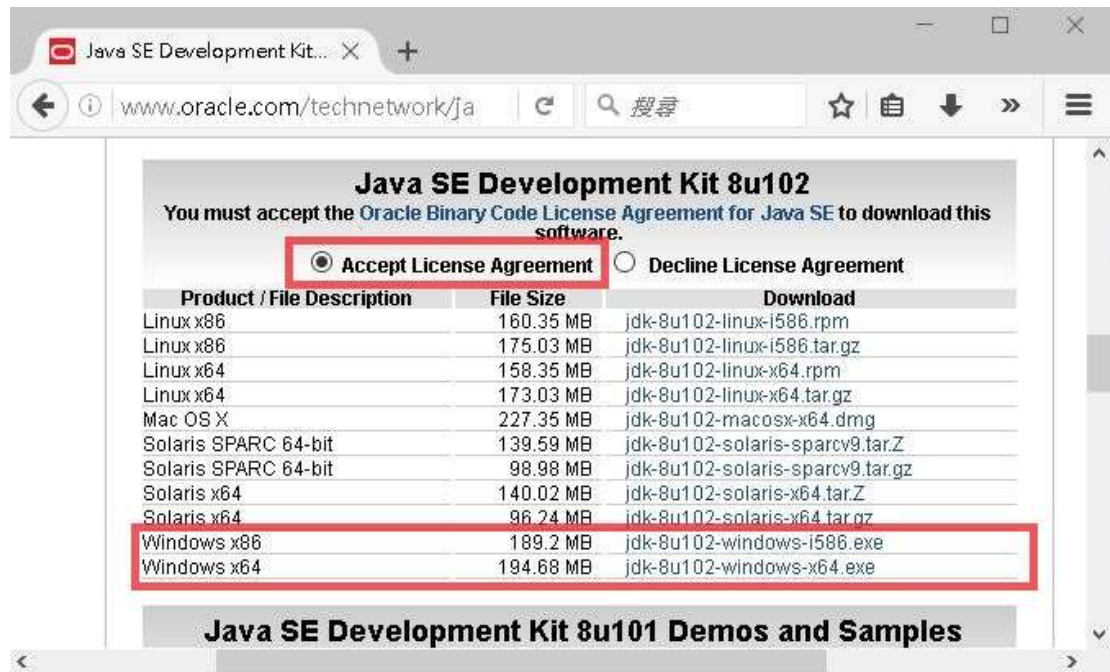


本文件將說明如何安裝 Spark 本機開發環境，以下畫面皆為 Windows 環境截圖，Linux 環境步驟亦相同

1. download and install Java JDK8 ([網址](#))

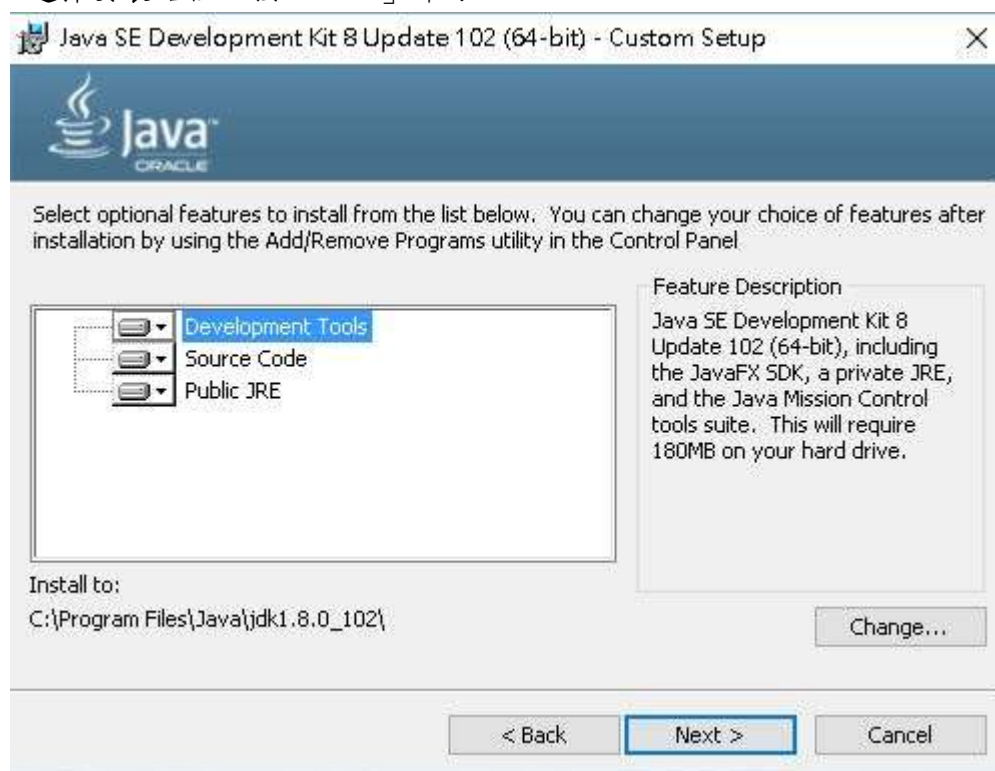
-到 Java 網站下載 JDK8u102



-執行安裝，只需一直「Next」即可完成



-選擇安裝路徑，按「Next」即可



-選擇安裝路徑，按「Next」即可




-Java 安裝中



-安裝完成，按「Close」結束程式



-安裝完成後，可至 cmd，輸入 `java -version` 檢視是否可正常運行

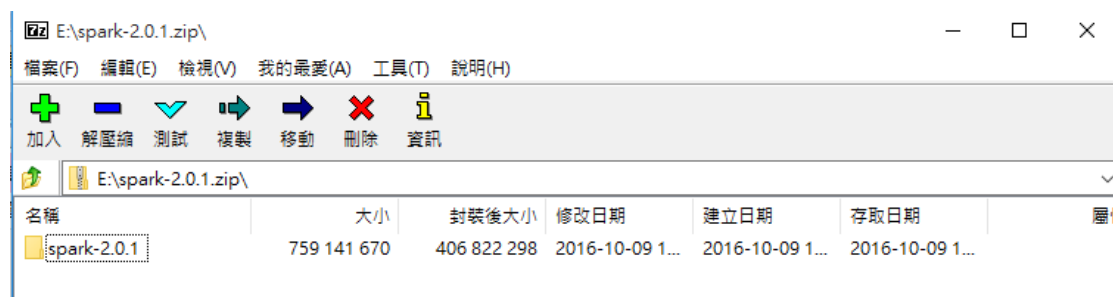


```
命令提示字元
Microsoft Windows [版本 10.0.10586]
(c) 2015 Microsoft Corporation. 著作權所有，並保留一切權利。

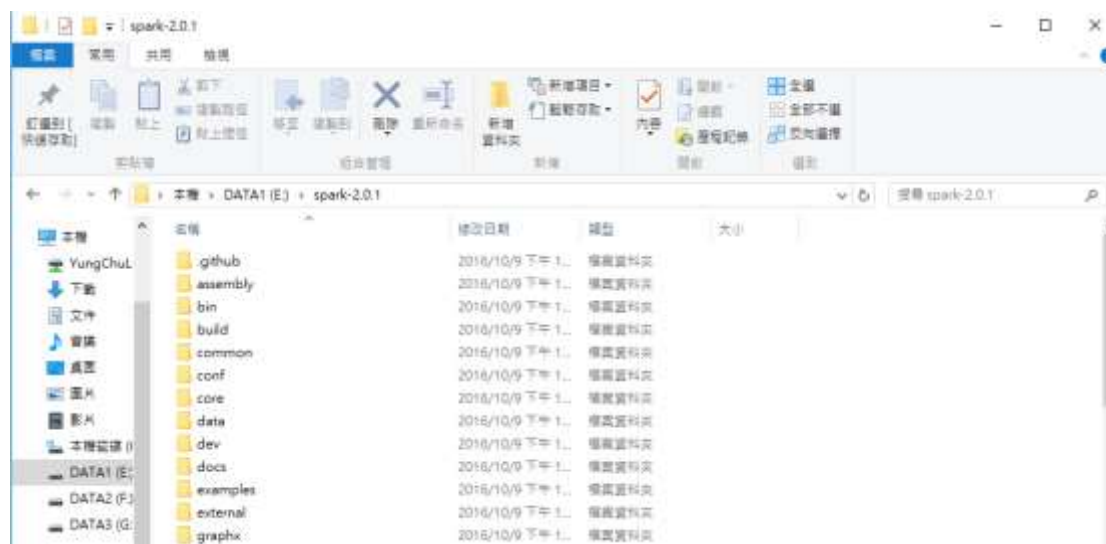
C:\Users\YungChul>java -version
java version "1.8.0_102"
Java(TM) SE Runtime Environment (build 1.8.0_102-b14)
Java HotSpot(TM) 64-Bit Server VM (build 25.102-b14, mixed mode)
```

2. 至 <https://goo.gl/LdlGm8> 下載 spark-2.0.1.zip, and unzip

-zip 檔可用系統內建解壓縮程式或 7zip 開啟(用 windows 內建解壓縮程式會有檔名過長無法建檔的問題)



-將 spark-2.0.1 資料夾解壓縮到硬碟(EX: E:\spark-2.0.1\)



-開啟「命令提示字元」，到 spark\bin 底下，執行 `spark-shell`

```
C:\> 命令提示字元

E:\>cd spark-2.0.1\bin

E:\spark-2.0.1\bin>spark-shell
```

將 spark\bin 加入 Path 環境變數中，可以在任何地方呼叫 spark-shell 指令



-spark 順利執行畫面(出現 scala> 表示已可以執行，可試著輸入簡單程式執行看看)

```
命令提示字元 - spark-shell

E:\>cd spark-2.0.1\bin

E:\spark-2.0.1\bin>spark-shell
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel).
16/10/10 12:39:45 WARN NativeCodeLoader: Unable to load native-adoop library for your platform... using builtin-java classes where applicable
16/10/10 12:39:47 WARN SparkContext: Use an existing SparkContext, some configuration may not take effect.
Spark context Web UI available at http://192.168.1.107:4040
Spark context available as 'sc' (master = local[*], app id = local-1476074386975).
Spark session available as 'spark'.
Welcome to

  ____ _
 / ___ \| | | |
/ /   \| |_| |
\ \___) |  __/
 \____|_|_|_| version 2.0.1

Using Scala version 2.11.8 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_77)
Type in expressions to have them evaluated.
Type :help for more information.

scala> val i=3+5
i: Int = 8

scala>
```

-若執行中遇到 NullPointerException 是因為缺乏在 windows 環境下模擬所需的 winutils.exe，以及 Win10 權限問題，檔案讀寫會出現問題，接下來步驟可以排除此問題(Linux 環境不需安裝 winutils.exe，可直接跳至步驟 6)

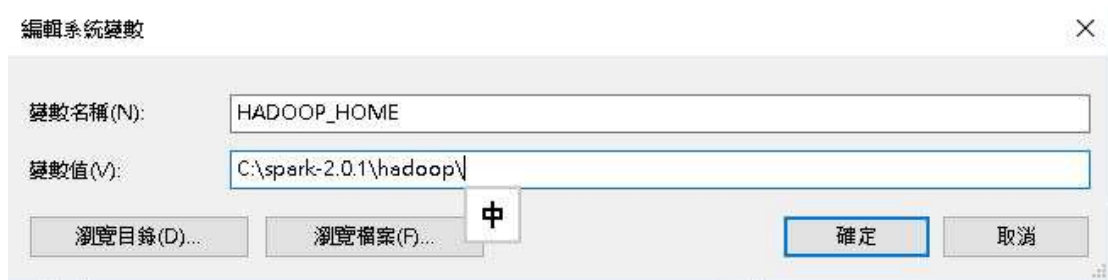
3. 本機下載 winutils.exe([網址](#))，在 Spark 目錄下建立 Hadoop\bin 資料夾，並將 winutils.exe 放入



4. 設定環境變數

-HADOOP_HOME=C:\spark-2.0.1\hadoop\





➔注意 HADOOP_HOME 需設定為放置 winutils.exe 的上一層

➔可執行 `echo %HADOOP_HOME%` 來檢查環境變數是否設定完成

5. 測試以下指令

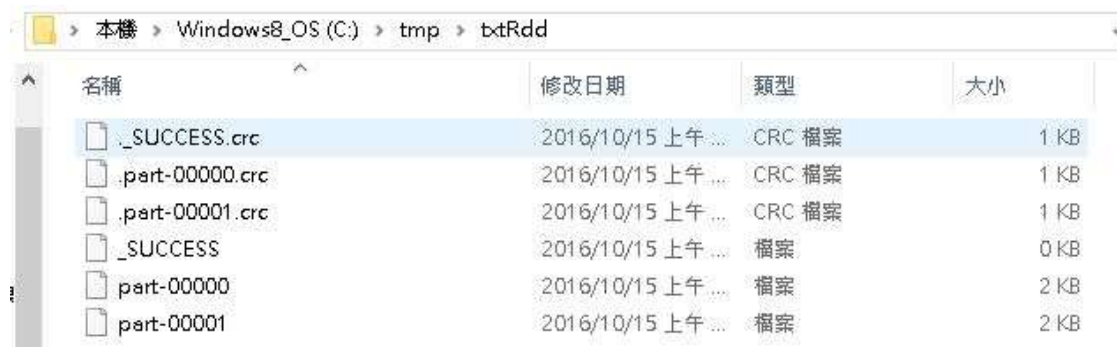
```
val txtRdd=sc.textFile("file:///c:/spark-2.0.1/README.md")
```

```
txtRdd.saveAsTextFile("File:///C:/tmp/txtRdd")
```

應該可以正常執行，並在 C:\tmp\ 底下產生 txtRdd 的資料夾

```
scala> val txtRdd=sc.textFile("file:///c:/spark-2.0.1/README.md")
txtRdd: org.apache.spark.rdd.RDD[String] = file:///c:/spark-2.0.1/README.md MapPartitionsRDD[1] at textFile at <console>:24

scala> txtRdd.saveAsTextFile("File:///C:/tmp/txtRdd")
[Stage 0:>                                     (0 + 0) / 2]
[Stage 0:>                                     (0 + 2) / 2]
```



Win10 可能在執行 `saveAsTextFile` 時仍會出現錯誤訊息，但不影響本課進行，可暫時忽略。

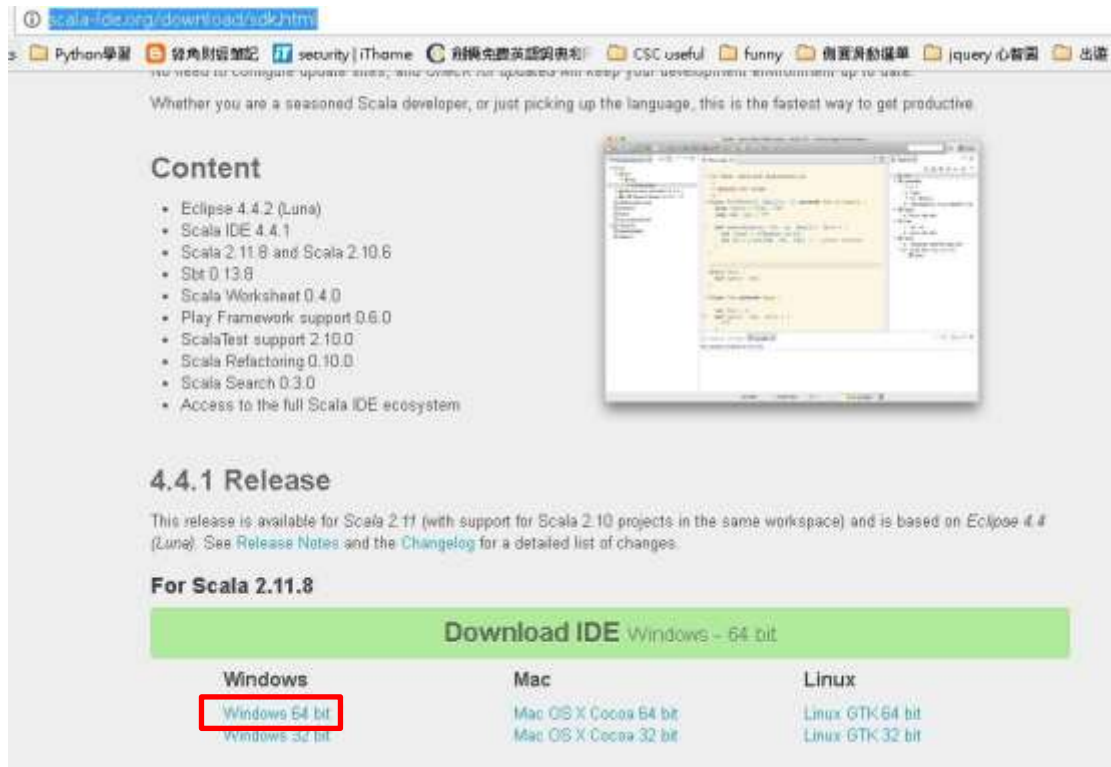
6. 欲結束 Spark，於 cmd 中輸入 `:quit` 或 `CTRL+D`

查詢指令則是輸入 `:help`

底下將說明如何使用 Scala IDE for Eclipse 本機開發環境，以下畫面皆為 Windows 環境截圖

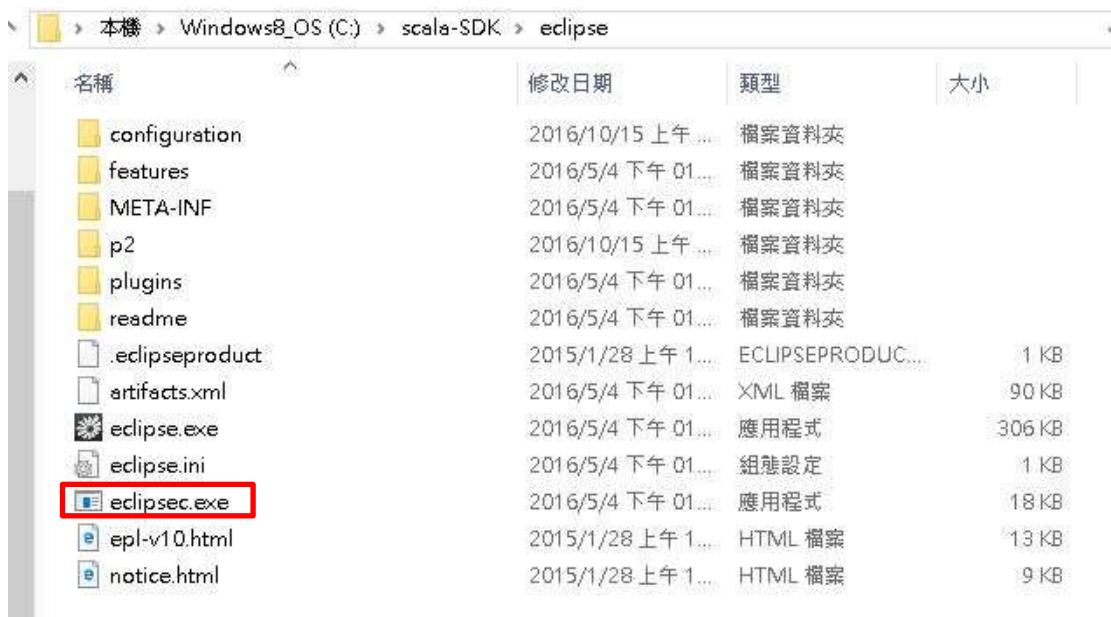
1. download Scala IDE for Eclipse ([網址](#))

-下載 windows 64 bit

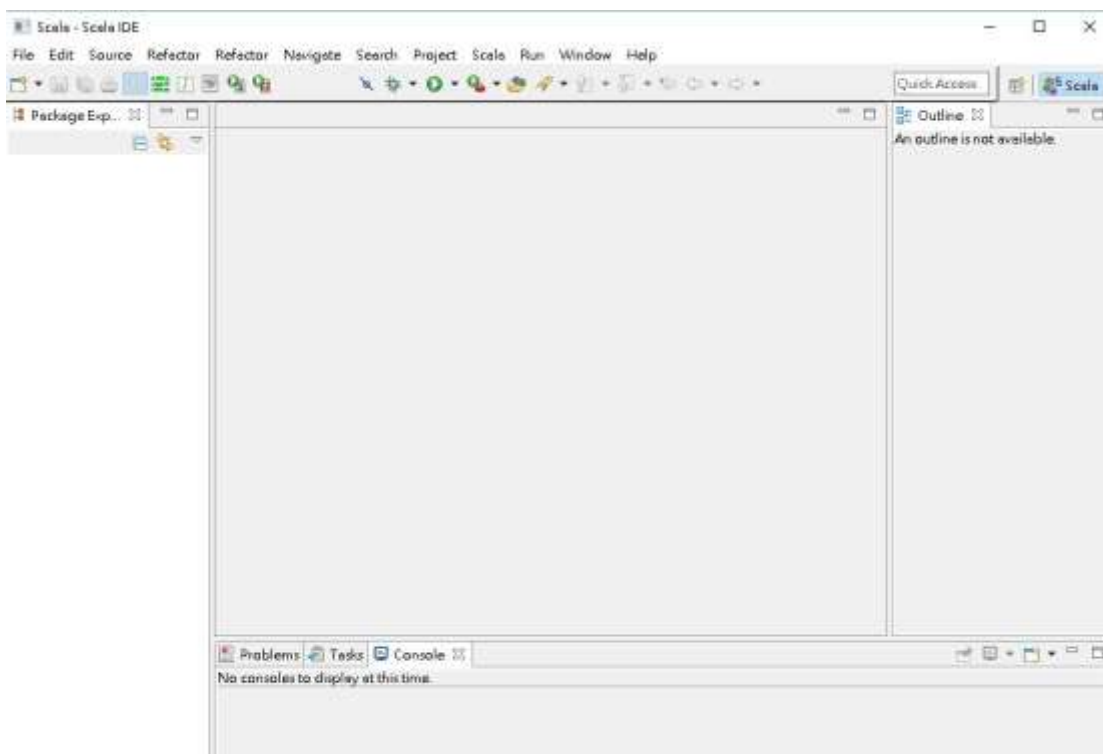


-解壓縮至硬碟(如 C:\scala-SDK)

-進入 scala-SDK\eclipse\ 執行 eclipse.exe，開啟 IDE 畫面

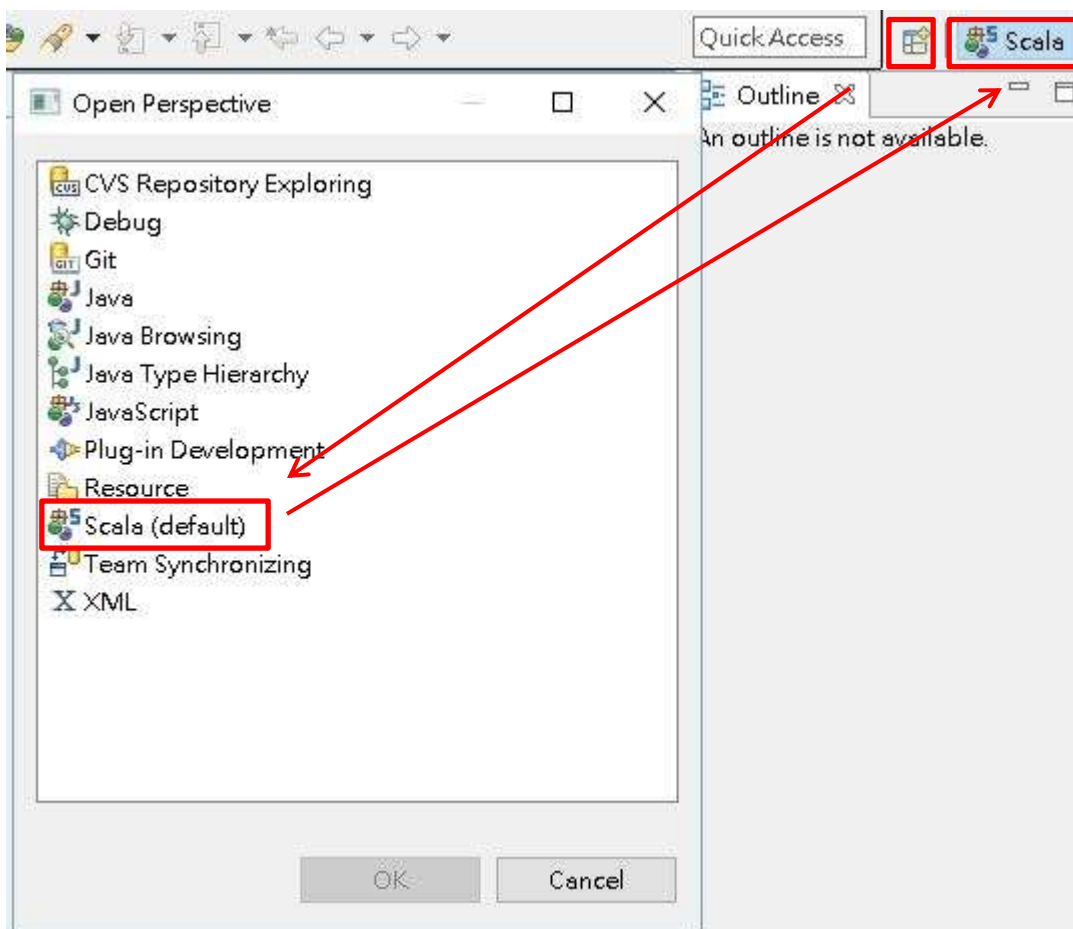


-eclipse 介面

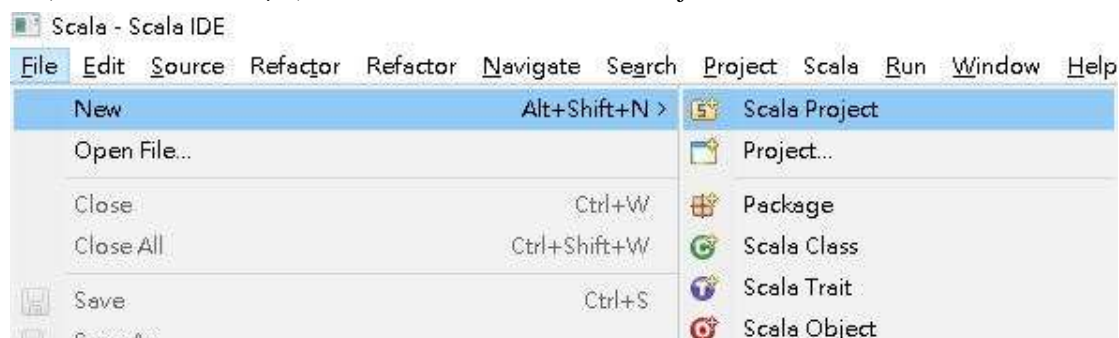


-右上角可以看到 scala 專案圖示(自行於 eclipse 安裝 scala 參考本[視頻](#))

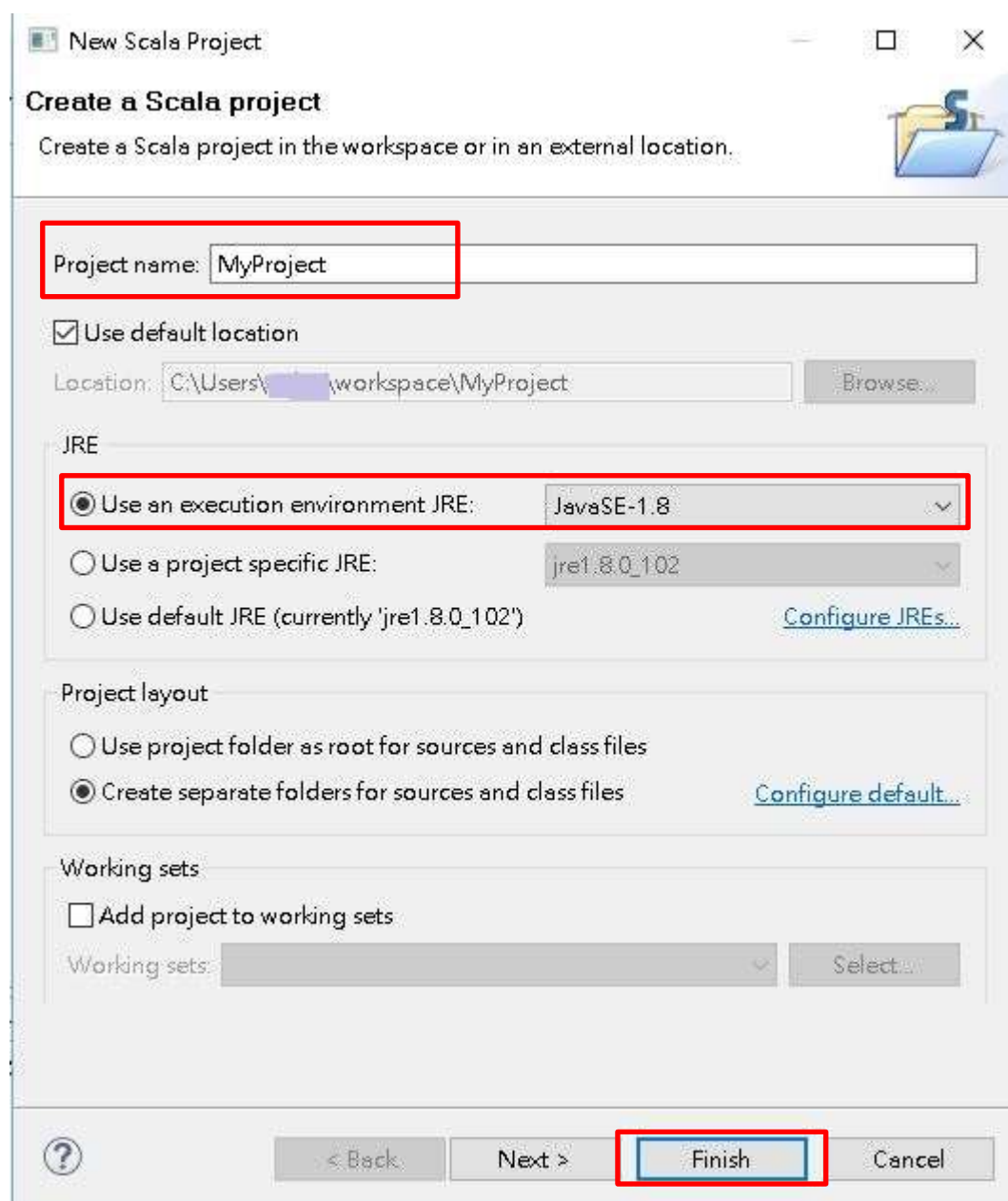
-若沒看到專案圖示，可依下圖順序操作



-開啟一個 scala 專案→File→New→scala Project



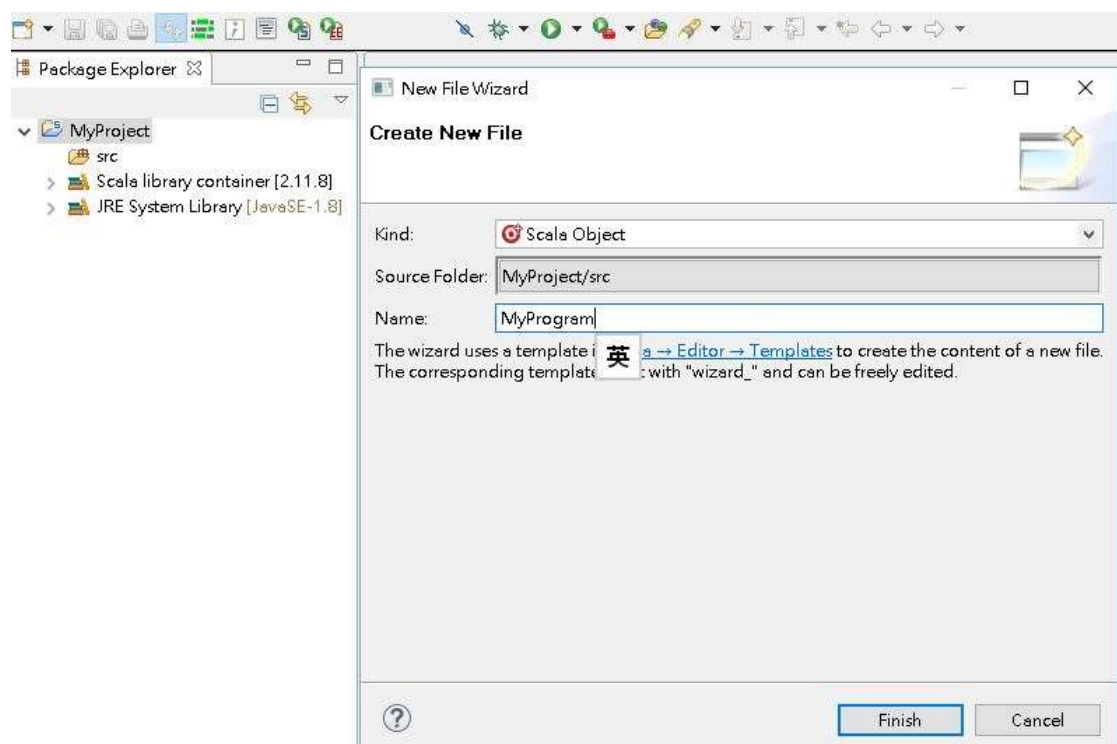
-輸入專案名稱，確認本機安裝的 JDK 版本後，按「Finish」建立專案



- 專案建立後，可於「Package Explorer」看到專案內容
- 建立一個 scala 物件，可用 New → Scala Object 建立

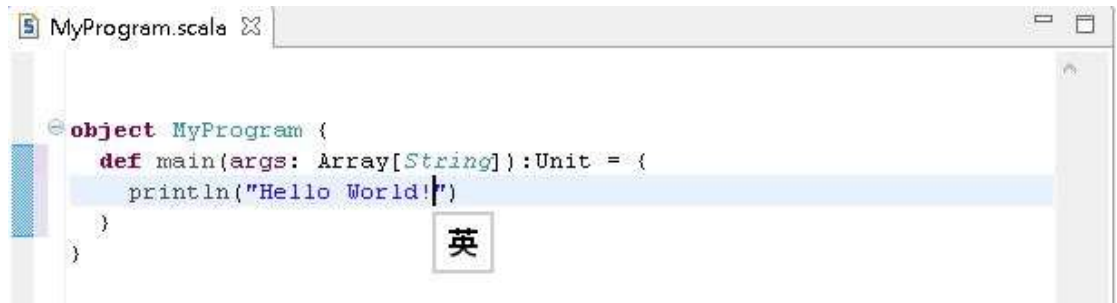


- 建立第一支 scala 程式「MyProgram.scala」



-嘗試在程式中輸入

```
def main(args: Array[String]):Unit = {  
    println("Hello World!")  
}
```



-儲存後，點選「Run」→「Run As」→「Scala Application」來執行程式
-底下「Console」窗格顯示 Hello World，表示程式正常運作

