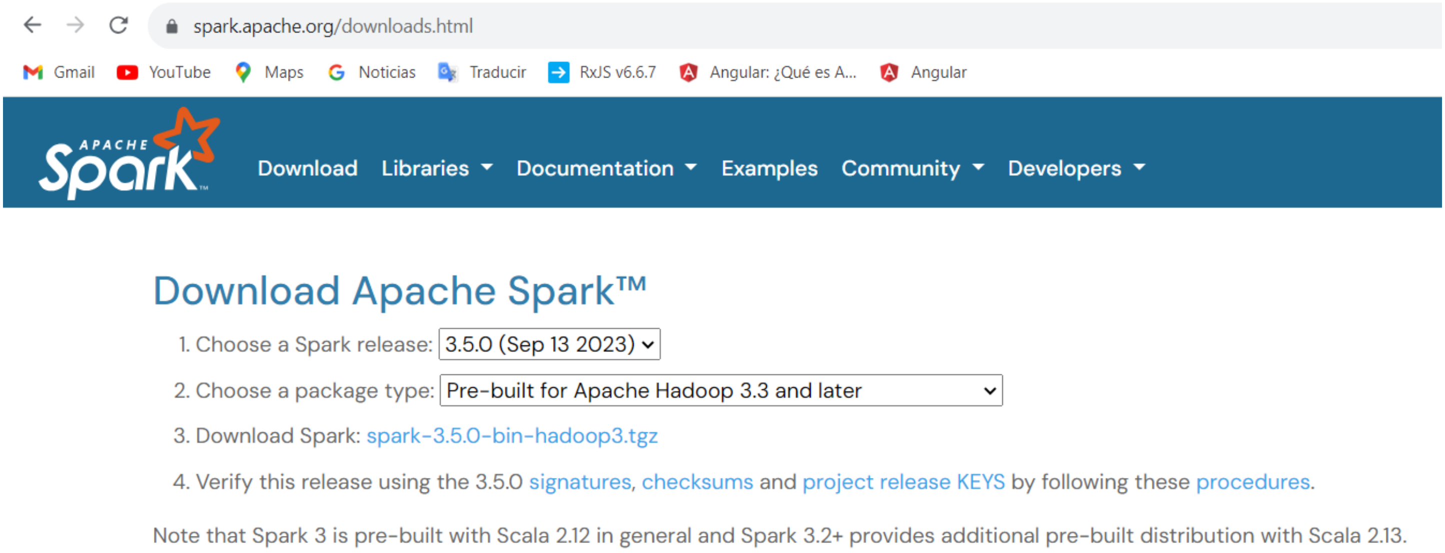
Spark with IntelliJ Community

https://www.youtube.com/@hackprotech/videos

1. Download Spark

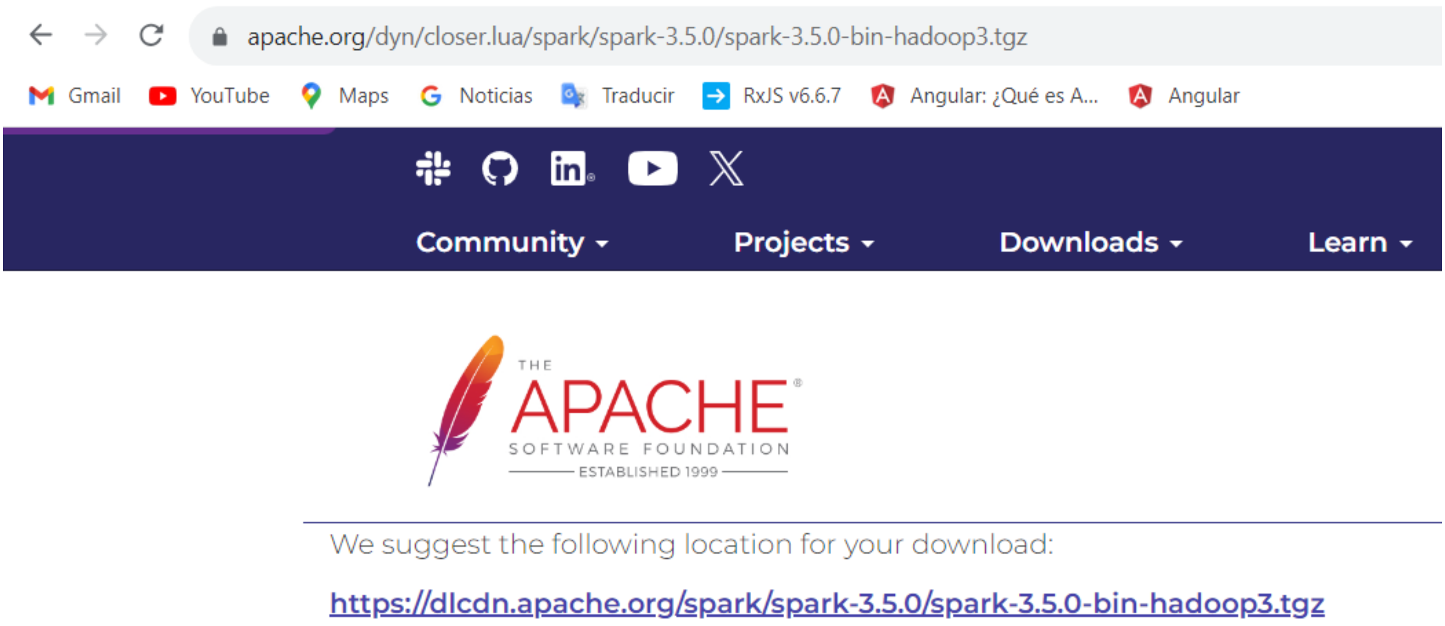


Download Apache Spark™

1. Choose a Spark release: 3.5.0 (Sep 13 2023)
2. Choose a package type: Pre-built for Apache Hadoop 3.3 and later
3. Download Spark: spark-3.5.0-bin-hadoop3.tgz
4. Verify this release using the 3.5.0 signatures, checksums and project release KEYS by following these procedures.

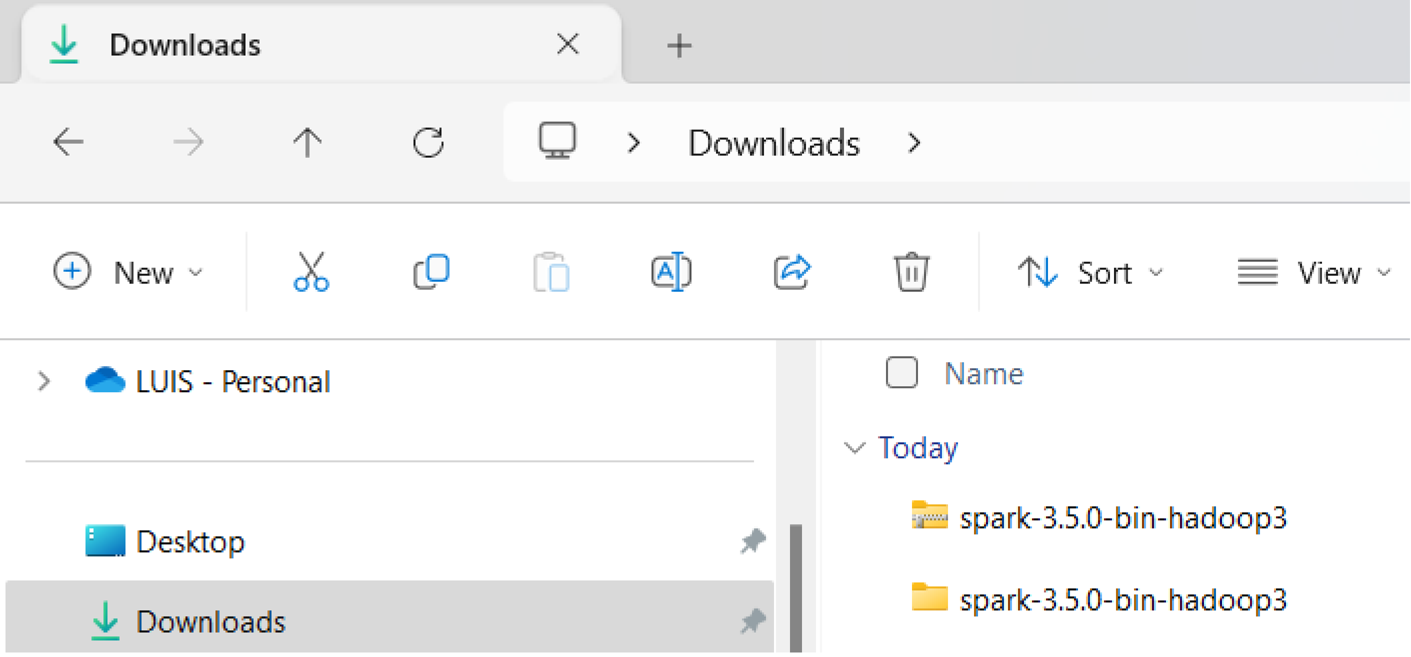
Note that Spark 3 is pre-built with Scala 2.12 in general and Spark 3.2+ provides additional pre-built distribution with Scala 2.13.

Unzip the "spark-3.5.0-bin-hadoop3.tgz"



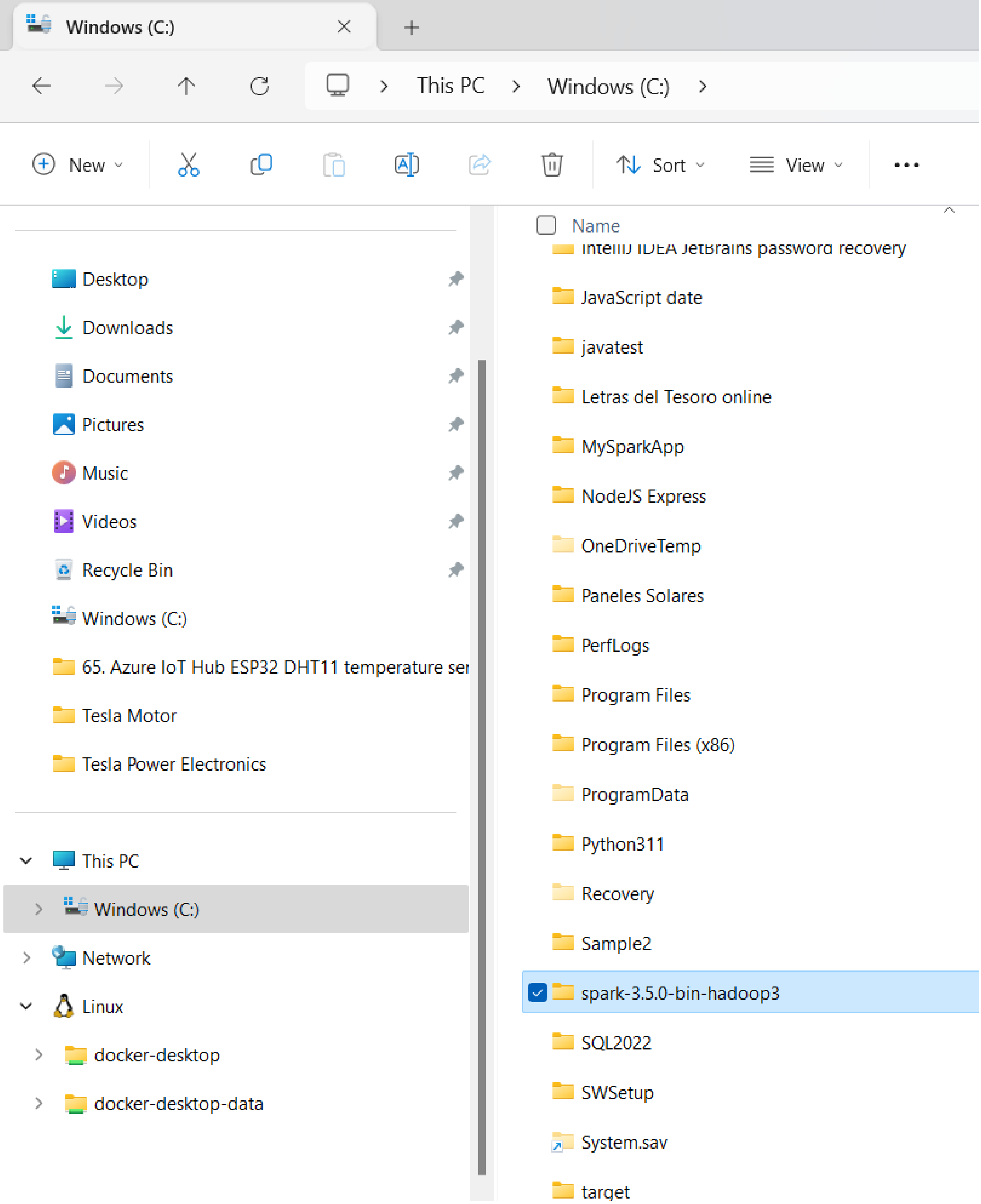
We suggest the following location for your download:

https://dlcdn.apache.org/spark/spark-3.5.0/spark-3.5.0-bin-hadoop3.tgz



spark-3.5.0-bin-hadoop3

spark-3.5.0-bin-hadoop3

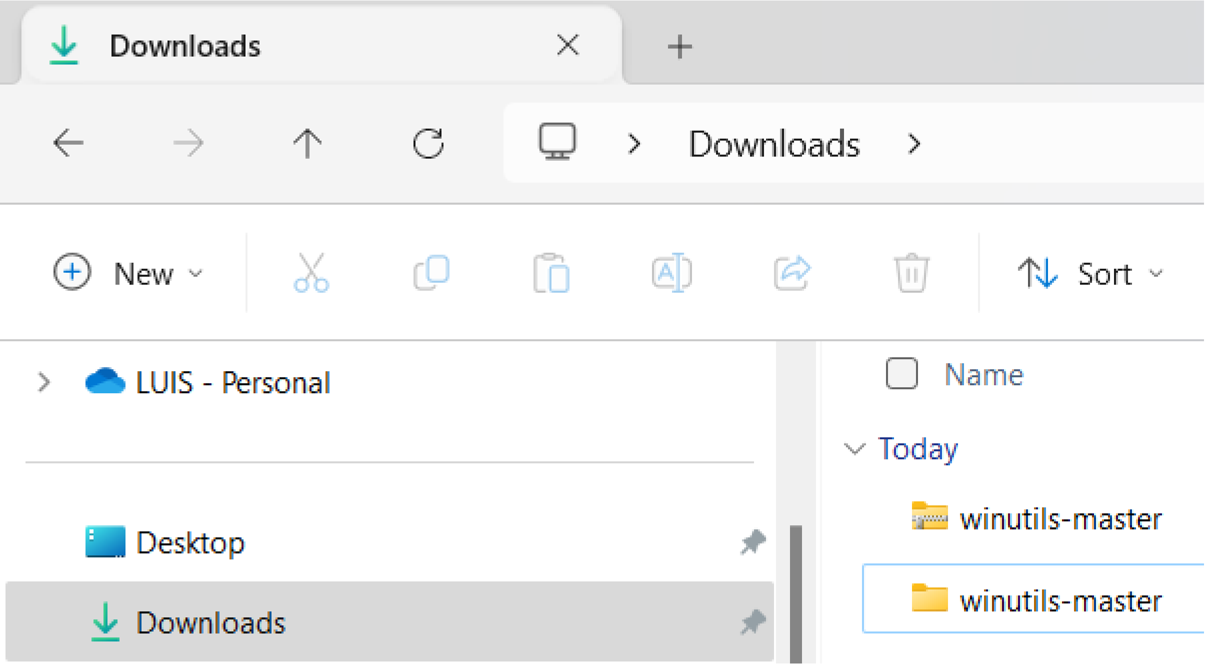


1. Download winutils for Hadoop

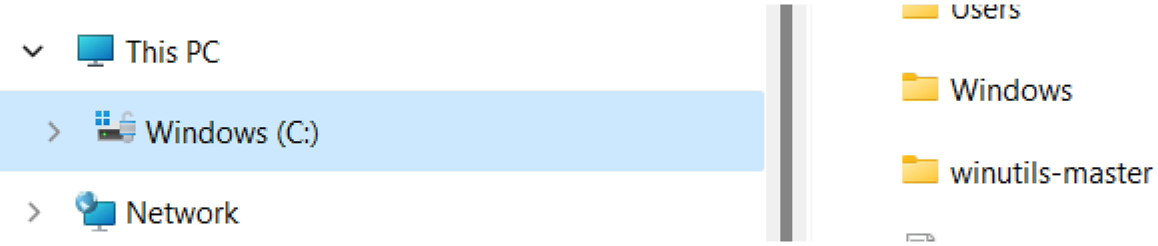
Go to this repository and download it

https://github.com/kontext-tech/winutils

Once downloaded the winutils repository, Unzip the file

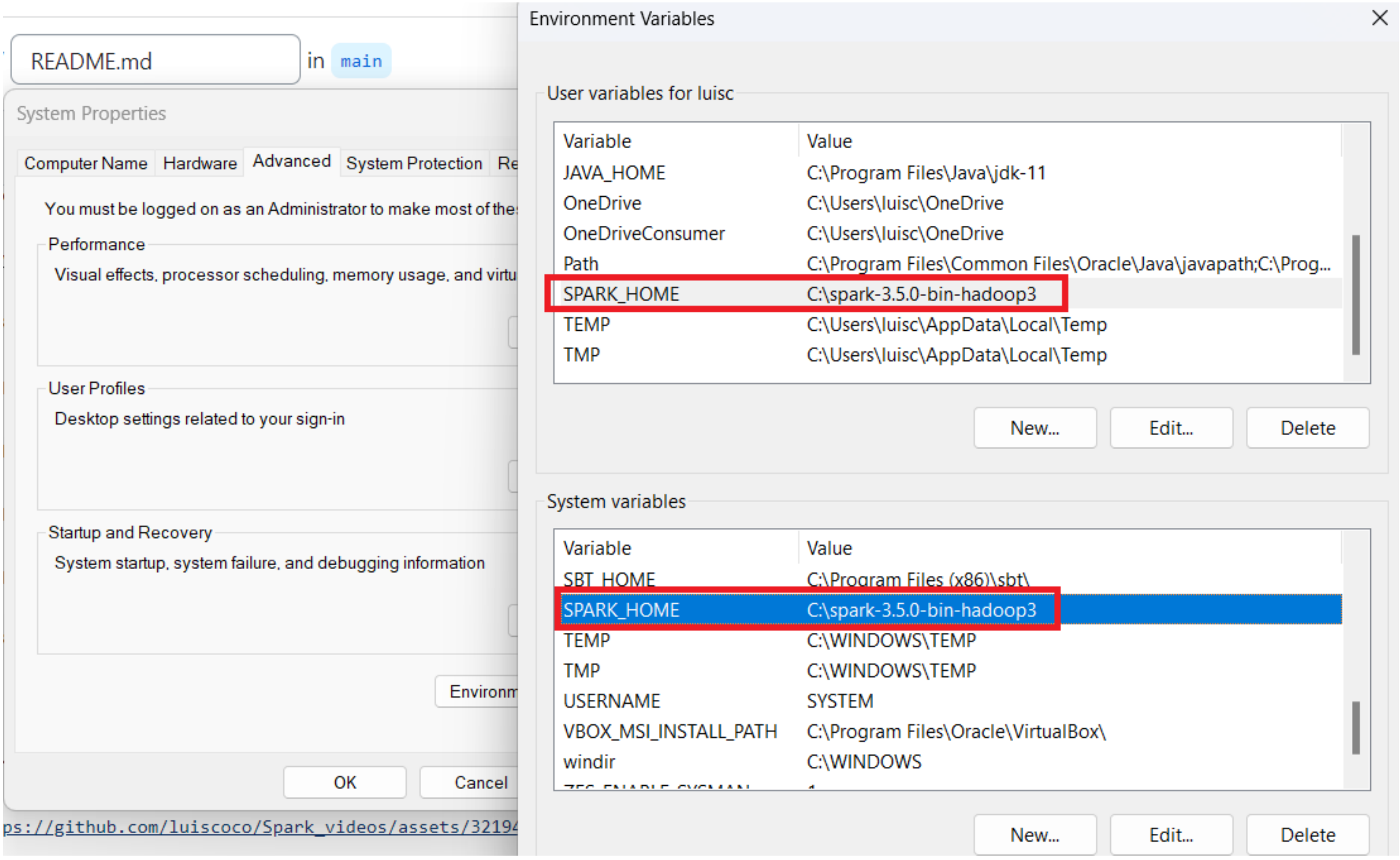


Now we cut from the Download folder and we copy it to the C:/ root



In the following section we have to set the HADOOP_HOME environmental variable

2. Set Spark environmental variable

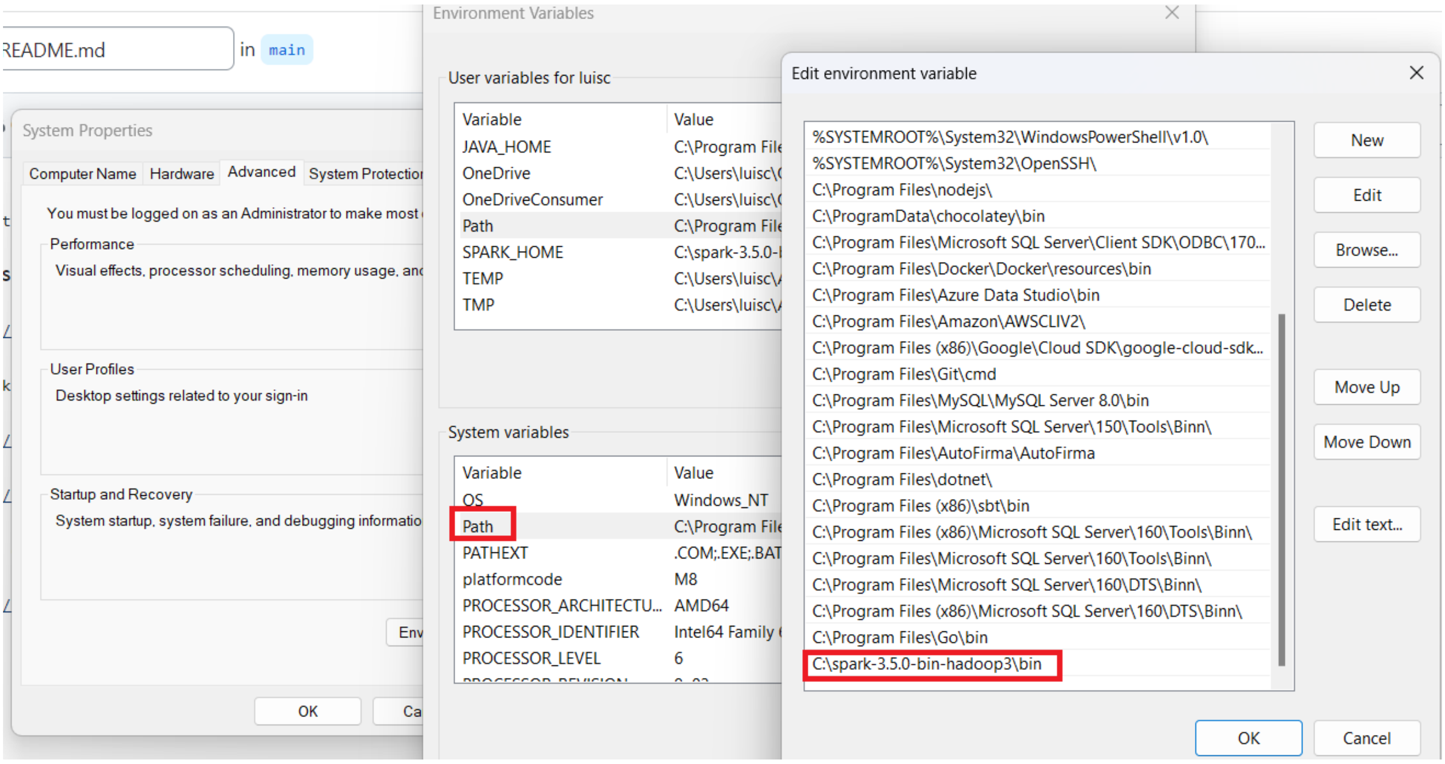

README.md
in main
System Properties
Computer Name
Hardware
Advanced
System Protection
Re
You must be logged on as an Administrator to make most of the
Performance
Visual effects, processor scheduling, memory usage, and virtu
User Profiles
Desktop settings related to your sign-in
Startup and Recovery
System startup, system failure, and debugging information
Environm
OK
Cancel
ps://github.com/luiscoco/Spark_videos/assets/32194
Environment Variables
User variables for luisc
New...
Edit...
Delete
System variables
New...
Edit...
Delete

User variables for luisc

Variable	Value
JAVA_HOME	C:\Program Files\Java\jdk-11
OneDrive	C:\Users\luisc\OneDrive
OneDriveConsumer	C:\Users\luisc\OneDrive
Path	C:\Program Files\Common Files\Oracle\Java\javapath;C:\Prog...
SPARK_HOME	C:\spark-3.5.0-bin-hadoop3
TEMP	C:\Users\luisc\AppData\Local\Temp
TMP	C:\Users\luisc\AppData\Local\Temp

System variables

Variable	Value
SBT_HOME	C:\Program Files (x86)\sbt\
SPARK_HOME	C:\spark-3.5.0-bin-hadoop3
TEMP	C:\WINDOWS\TEMP
TMP	C:\WINDOWS\TEMP
USERNAME	SYSTEM
VBOX_MSI_INSTALL_PATH	C:\Program Files\Oracle\VirtualBox\
windir	C:\WINDOWS



User variables for luisc

Variable	Value
JAVA_HOME	C:\Program File
OneDrive	C:\Users\luisc\
OneDriveConsumer	C:\Users\luisc\
Path	C:\Program File
SPARK_HOME	C:\spark-3.5.0-l
TEMP	C:\Users\luisc\
TMP	C:\Users\luisc\

System variables

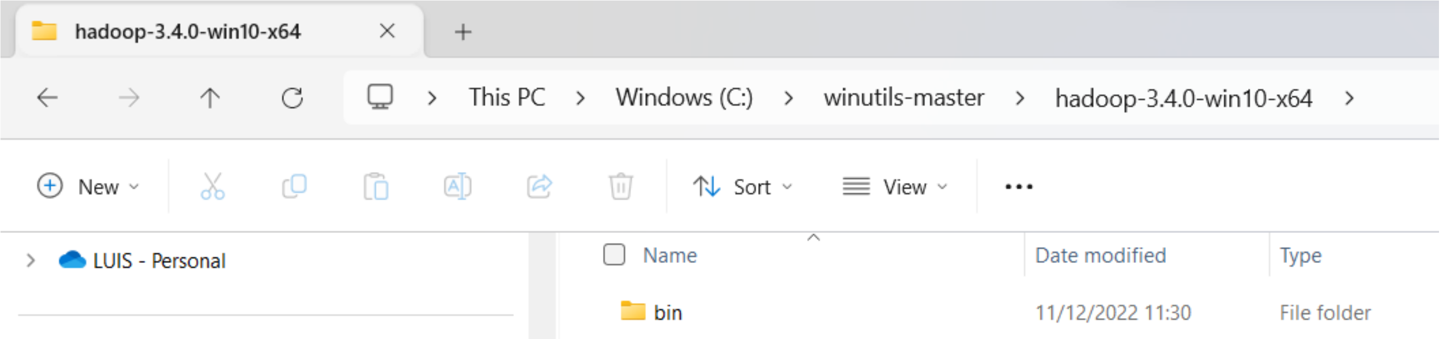
Variable	Value
OS	Windows_NT
Path	C:\Program File
PATHEXT	.COM;.EXE;.BAT
platformcode	M8
PROCESSOR_ARCHITECTU...	AMD64
PROCESSOR_IDENTIFIER	Intel64 Family
PROCESSOR_LEVEL	6

Edit environment variable

```
%SYSTEMROOT%\System32\WindowsPowerShell\v1.0\
%SYSTEMROOT%\System32\OpenSSH\
C:\Program Files\nodejs\
C:\ProgramData\chocolatey\bin
C:\Program Files\Microsoft SQL Server\Client SDK\ODBC\170...
C:\Program Files\Docker\Docker\resources\bin
C:\Program Files\Azure Data Studio\bin
C:\Program Files\Amazon\AWSCLIV2\
C:\Program Files (x86)\Google\Cloud SDK\google-cloud-sdk...
C:\Program Files\Git\cmd
C:\Program Files\MySQL\MySQL Server 8.0\bin
C:\Program Files\Microsoft SQL Server\150\Tools\Binn\
C:\Program Files\AutoFirma\AutoFirma
C:\Program Files\dotnet\
C:\Program Files (x86)\sbt\bin
C:\Program Files (x86)\Microsoft SQL Server\160\Tools\Binn\
C:\Program Files\Microsoft SQL Server\160\Tools\Binn\
C:\Program Files\Microsoft SQL Server\160\DTS\Binn\
C:\Program Files (x86)\Microsoft SQL Server\160\DTS\Binn\
C:\Program Files\Go\bin
C:\spark-3.5.0-bin-hadoop3\bin
```

2. Set Hadoop winutils environmental variable

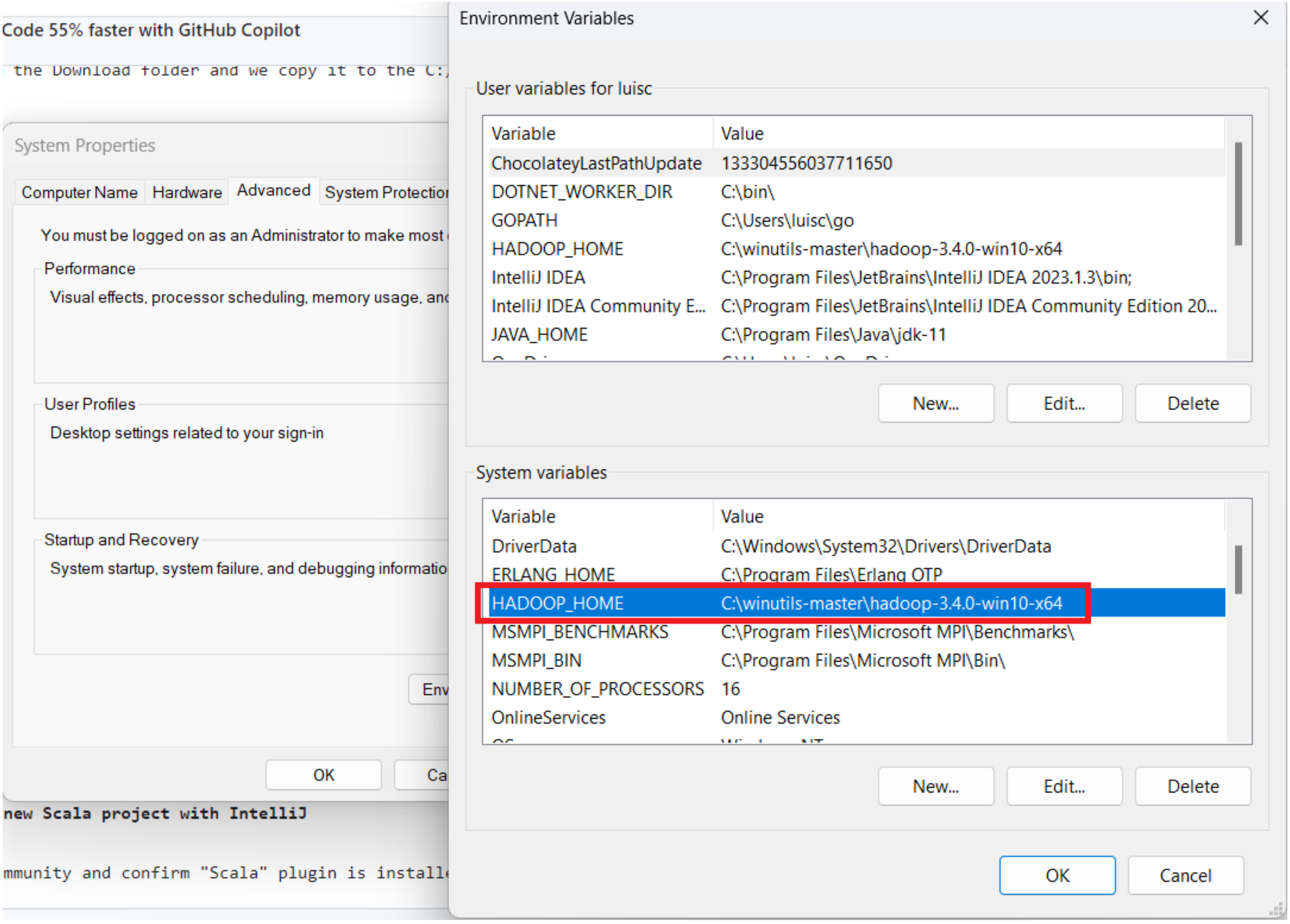
We navigate to the bin folder where is located the winutils.exe file. This path will be copied later in the HADOOP_HOME environmental variable.



hadoop-3.4.0-win10-x64
This PC > Windows (C:) > winutils-master > hadoop-3.4.0-win10-x64 >
New
Sort
View
LUIS - Personal

Name	Date modified	Type
bin	11/12/2022 11:30	File folder

We create a new environmental variable called HADOOP_HOME and we set the value to the path where is located the bin folder, as explained above



Code 55% faster with GitHub Copilot
the Download folder and we copy it to the C:/
System Properties
Computer Name
Hardware
Advanced
System Protection
You must be logged on as an Administrator to make most
Performance
Visual effects, processor scheduling, memory usage, and
User Profiles
Desktop settings related to your sign-in
Startup and Recovery
System startup, system failure, and debugging informatio
Env
OK
Ca
new Scala project with IntelliJ
mmunity and confirm "Scala" plugin is install
Environment Variables

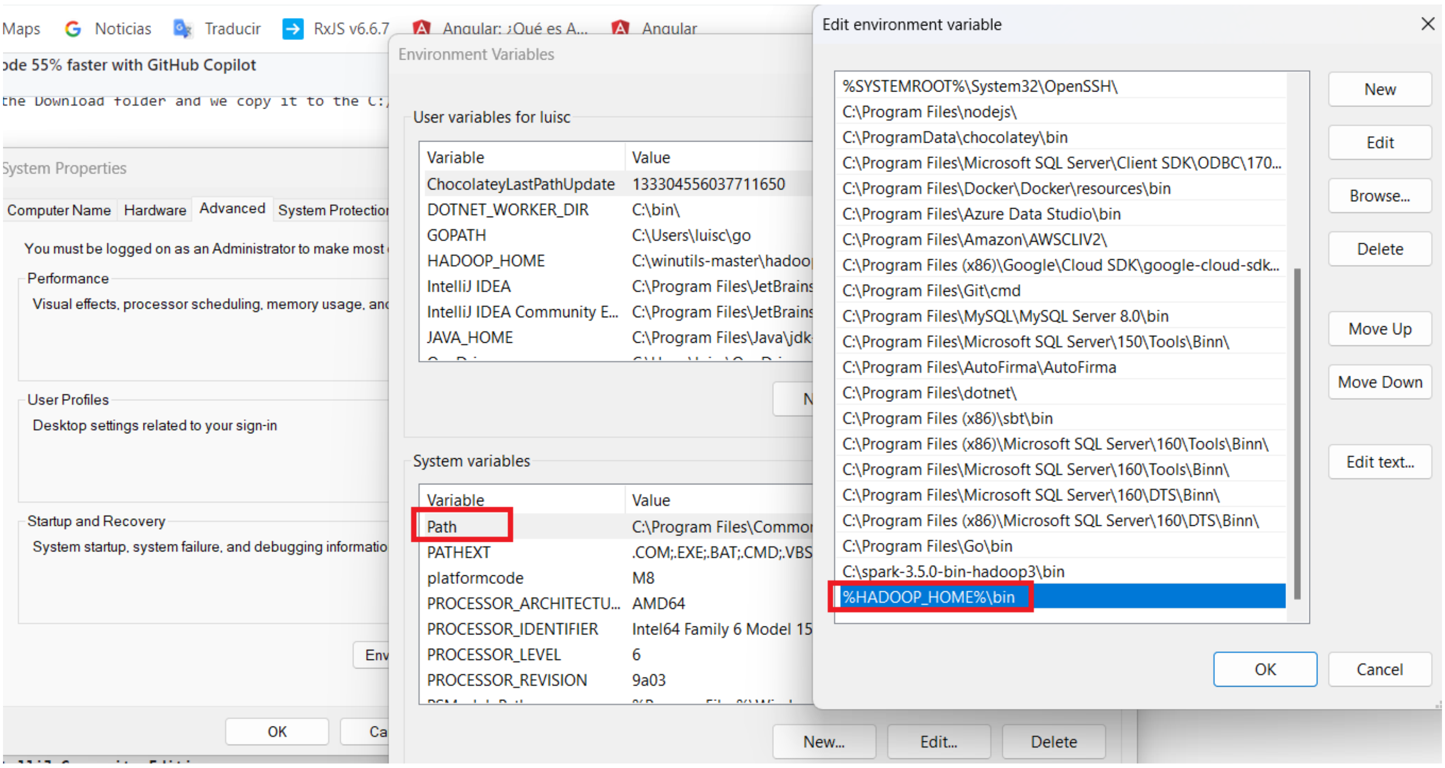
User variables for luisc

Variable	Value
ChocolateyLastPathUpdate	133304556037711650
DOTNET_WORKER_DIR	C:\bin\
GOPATH	C:\Users\luisc\go
HADOOP_HOME	C:\winutils-master\hadoop-3.4.0-win10-x64
IntelliJ IDEA	C:\Program Files\JetBrains\IntelliJ IDEA 2023.1.3\bin;
IntelliJ IDEA Community E...	C:\Program Files\JetBrains\IntelliJ IDEA Community Edition 20...
JAVA_HOME	C:\Program Files\Java\jdk-11

System variables

Variable	Value
DriverData	C:\Windows\System32\Drivers\DriverData
ERLANG_HOME	C:\Program Files\Erlang OTP
HADOOP_HOME	C:\winutils-master\hadoop-3.4.0-win10-x64
MSMPI_BENCHMARKS	C:\Program Files\Microsoft MPI\Benchmarks\
MSMPI_BIN	C:\Program Files\Microsoft MPI\Bin\
NUMBER_OF_PROCESSORS	16
OnlineServices	Online Services

We add the path to the winutils.exe file in the PATH environmental variable


Maps
Noticias
Traducir
RxJS v6.6.7
Angular: ¿Qué es A...
Angular
ode 55% faster with GitHub Copilot
the Download folder and we copy it to the C:/
System Properties
Computer Name
Hardware
Advanced
System Protectio
You must be logged on as an Administrator to make most
Performance
Visual effects, processor scheduling, memory usage, and
User Profiles
Desktop settings related to your sign-in
Startup and Recovery
System startup, system failure, and debugging informatio
Env
OK
Ca
Environment Variables
N
New...
Edit...
Delete
Edit environment variable
New
Edit
Browse...
Delete
Move Up
Move Down
Edit text...
OK
Cancel

User variables for luisc

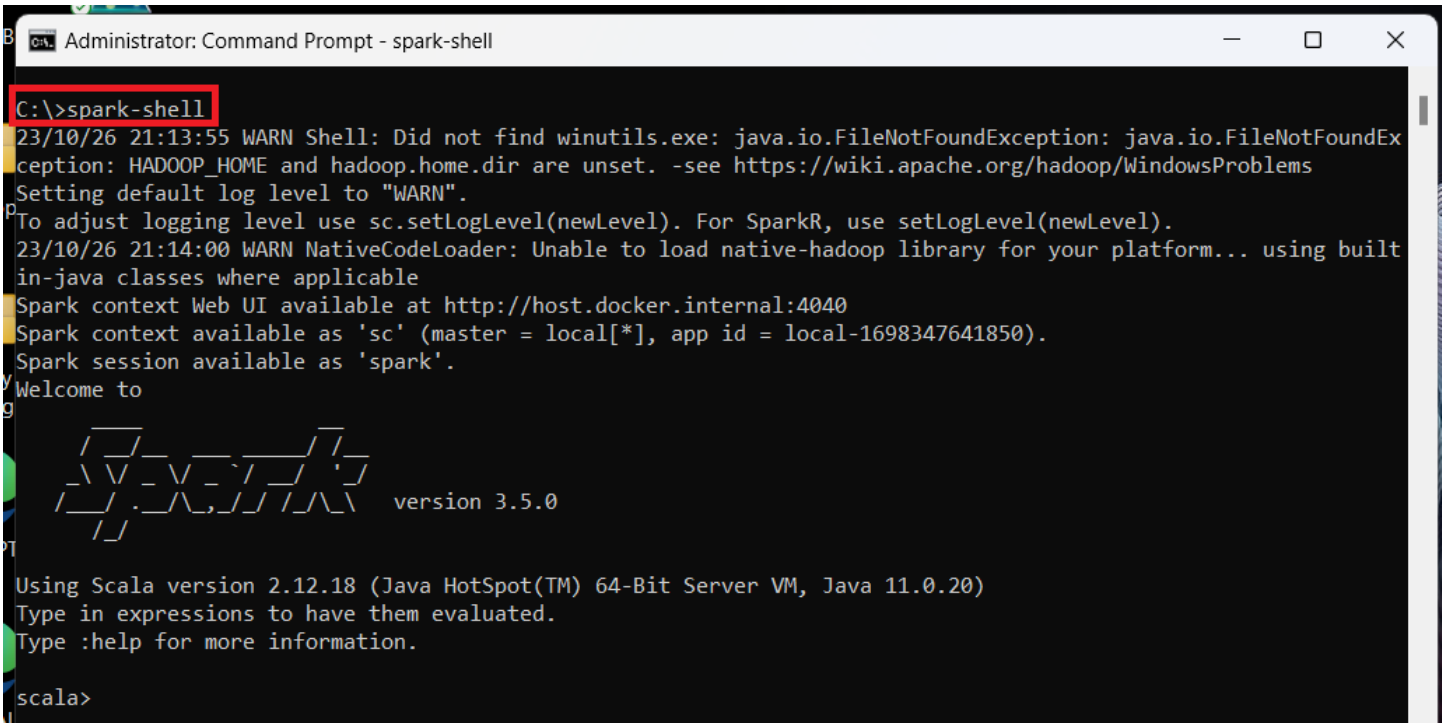
Variable	Value
ChocolateyLastPathUpdate	133304556037711650
DOTNET_WORKER_DIR	C:\bin\
GOPATH	C:\Users\luisc\go
HADOOP_HOME	C:\winutils-master\hadoo
IntelliJ IDEA	C:\Program Files\JetBrains
IntelliJ IDEA Community E...	C:\Program Files\JetBrains
JAVA_HOME	C:\Program Files\Java\jdk

System variables

Variable	Value
Path	C:\Program Files\Commor
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS
platformcode	M8
PROCESSOR_ARCHITECTU...	AMD64
PROCESSOR_IDENTIFIER	Intel64 Family 6 Model 15
PROCESSOR_LEVEL	6
PROCESSOR_REVISION	9a03

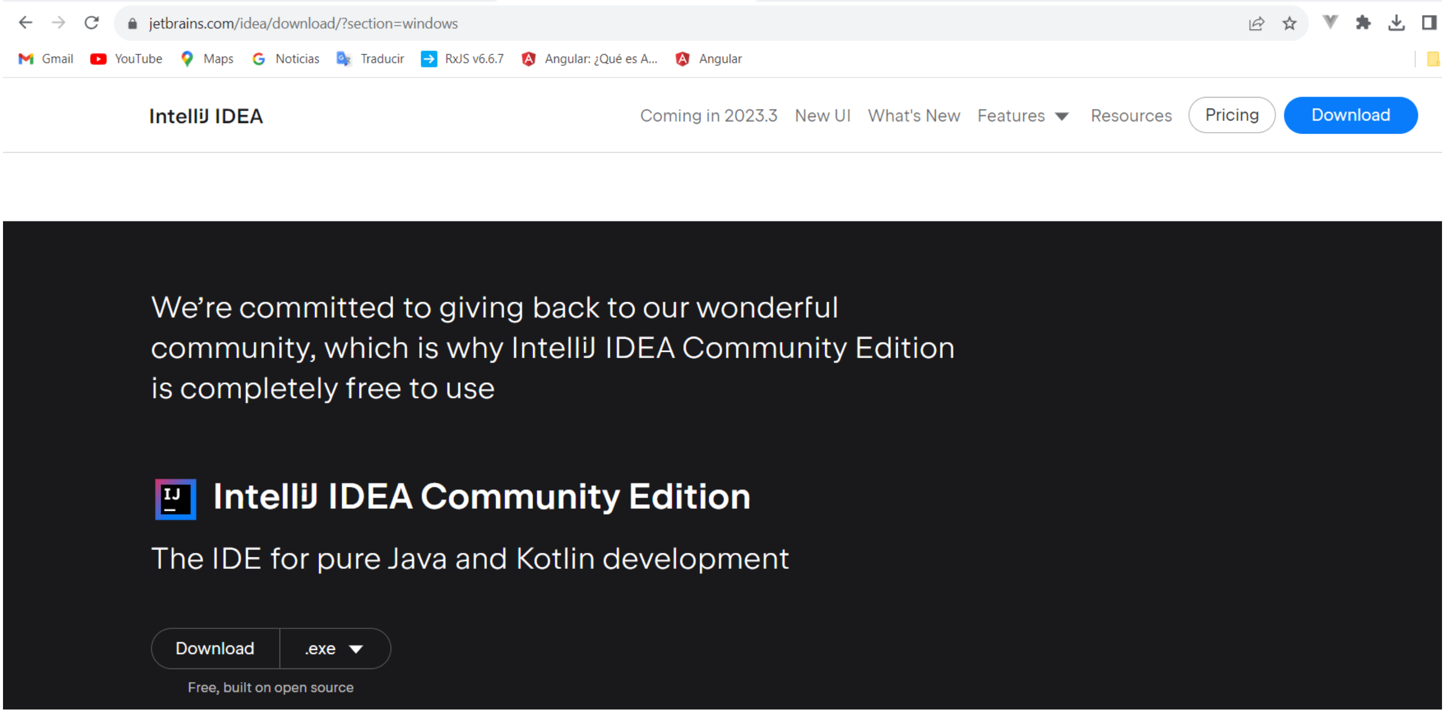
```
%SYSTEMROOT%\System32\OpenSSH\
C:\Program Files\nodejs\
C:\ProgramData\chocolatey\bin
C:\Program Files\Microsoft SQL Server\Client SDK\ODBC\170...
C:\Program Files\Docker\Docker\resources\bin
C:\Program Files\Azure Data Studio\bin
C:\Program Files\Amazon\AWSCLIV2\
C:\Program Files (x86)\Google\Cloud SDK\google-cloud-sdk...
C:\Program Files\Git\cmd
C:\Program Files\MySQL\MySQL Server 8.0\bin
C:\Program Files\Microsoft SQL Server\150\Tools\Binn\
C:\Program Files\AutoFirma\AutoFirma
C:\Program Files\dotnet\
C:\Program Files (x86)\sbt\bin
C:\Program Files (x86)\Microsoft SQL Server\160\Tools\Binn\
C:\Program Files\Microsoft SQL Server\160\Tools\Binn\
C:\Program Files\Microsoft SQL Server\160\DTS\Binn\
C:\Program Files (x86)\Microsoft SQL Server\160\DTS\Binn\
C:\Program Files\Go\bin
C:\spark-3.5.0-bin-hadoop3\bin
%HADOOP_HOME%\bin
```

3. Check the spark-shell is installed



```
C:\>spark-shell
23/10/26 21:13:55 WARN Shell: Did not find winutils.exe: java.io.FileNotFoundException: java.io.FileNotFoundEx
ception: HADOOP_HOME and hadoop.home.dir are unset. -see https://wiki.apache.org/hadoop/WindowsProblems
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
23/10/26 21:14:00 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using built
in-java classes where applicable
Spark context Web UI available at http://host.docker.internal:4040
Spark context available as 'sc' (master = local[*], app id = local-1698347641850).
Spark session available as 'spark'.
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /___/ .__/\_,_/_/ /_/\_\   version 3.5.0
      /_/

Using Scala version 2.12.18 (Java HotSpot(TM) 64-Bit Server VM, Java 11.0.20)
Type in expressions to have them evaluated.
Type :help for more information.

scala>
```

4. Install IntelliJ Community Edition



We're committed to giving back to our wonderful community, which is why IntelliJ IDEA Community Edition is completely free to use

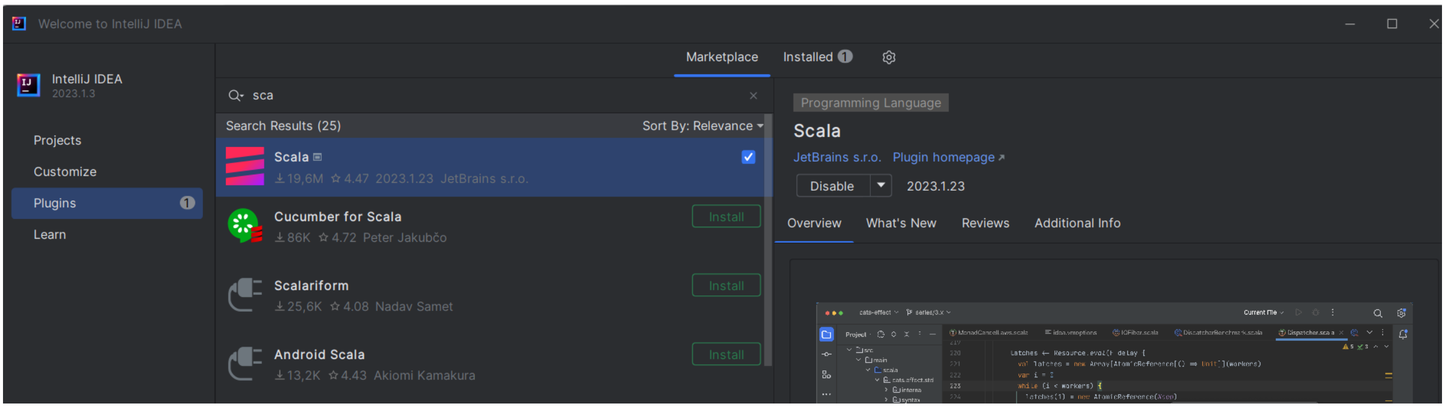
IntelliJ IDEA Community Edition

The IDE for pure Java and Kotlin development

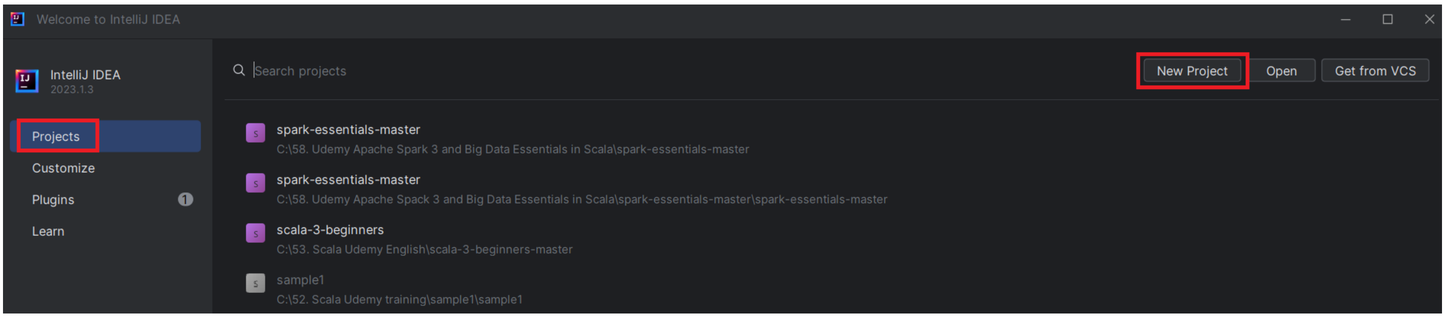
Free, built on open source

5. Create a new Scala project with IntelliJ

Run IntelliJ Community and confirm "Scala" plugin is installed



Select "Projects" in the left menu, and then click on the "New Project" button



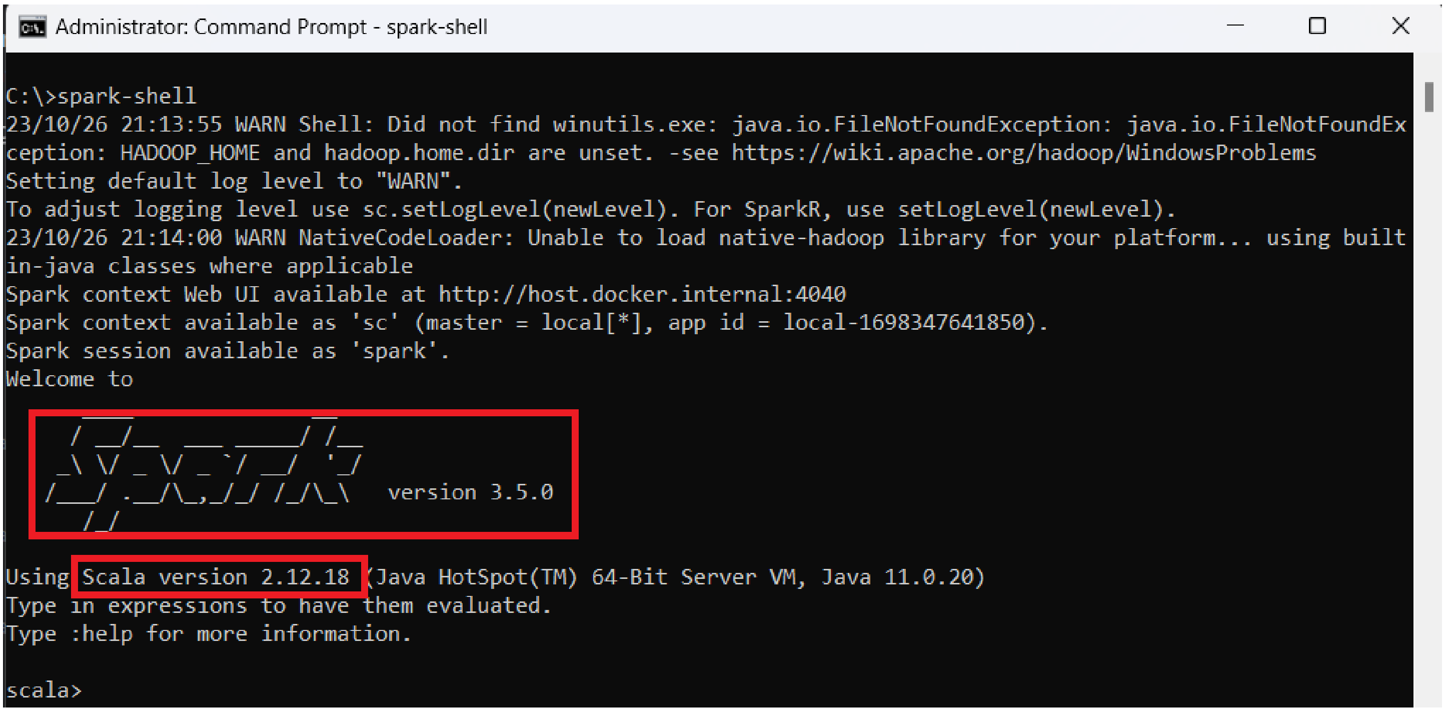
spark-essentials-master
C:\58. Udemy Apache Spark 3 and Big Data Essentials in Scala\spark-essentials-master

spark-essentials-master
C:\58. Udemy Apache Spack 3 and Big Data Essentials in Scala\spark-essentials-master\spark-essentials-master

scala-3-beginners
C:\53. Scala Udemy English\scala-3-beginners-master

sample1
C:\52. Scala Udemy training\sample1\sample1

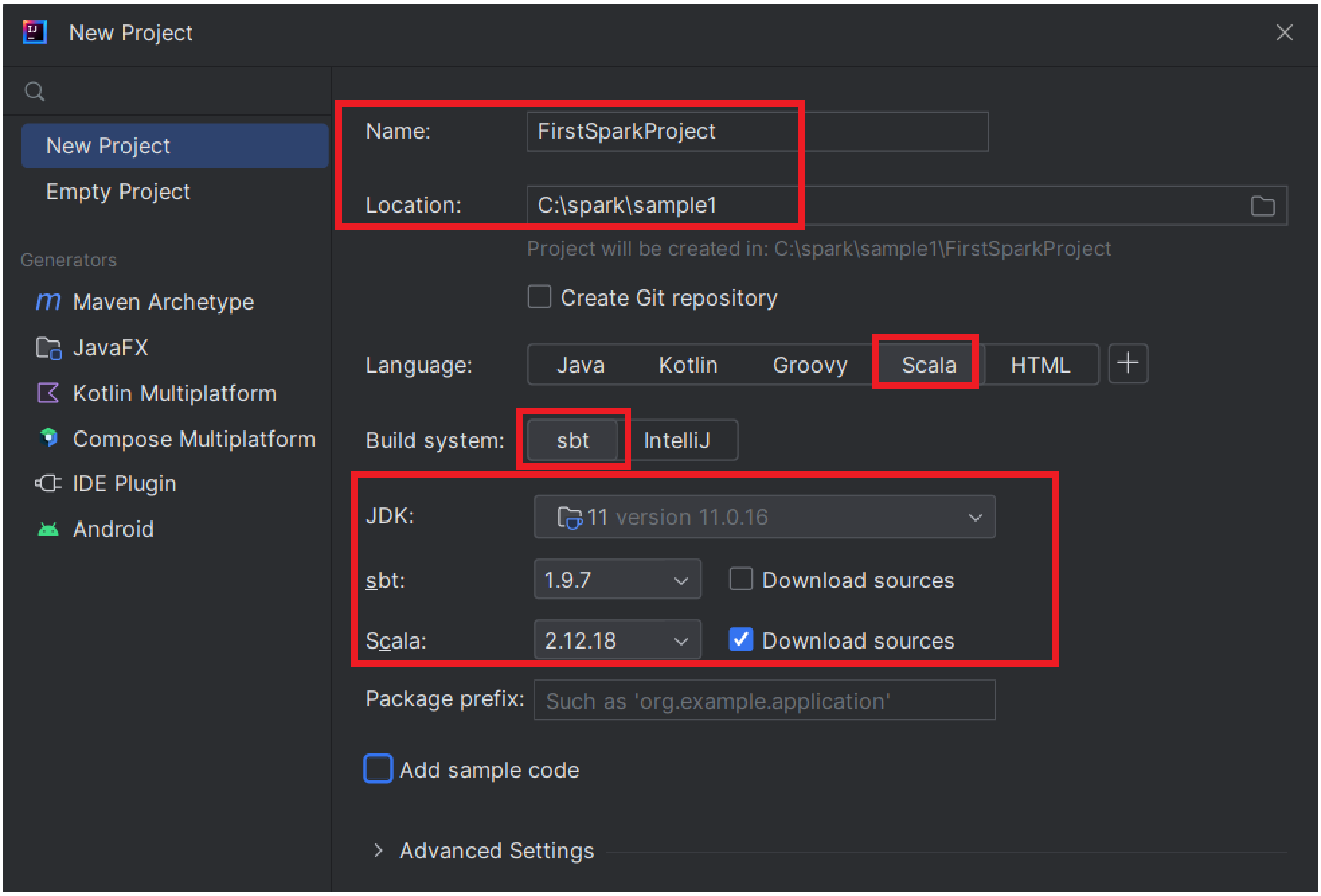
Before we input the new project data, copy Spark version and the Scala Version



Administrator: Command Prompt - spark-shell

```
C:\>spark-shell
23/10/26 21:13:55 WARN Shell: Did not find winutils.exe: java.io.FileNotFoundException: java.io.FileNotFoundEx
ception: HADOOP_HOME and hadoop.home.dir are unset. -see https://wiki.apache.org/hadoop/WindowsProblems
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
23/10/26 21:14:00 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using built
in-java classes where applicable
Spark context Web UI available at http://host.docker.internal:4040
Spark context available as 'sc' (master = local[*], app id = local-1698347641850).
Spark session available as 'spark'.
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /___/ .__/\_,_/_/ /_/\_\   version 3.5.0
      /_/

Using Scala version 2.12.18 (Java HotSpot(TM) 64-Bit Server VM, Java 11.0.20)
Type in expressions to have them evaluated.
Type :help for more information.

scala>
```

Then we input the new project data



New Project

Name: FirstSparkProject

Location: C:\spark\sample1

Project will be created in: C:\spark\sample1\FirstSparkProject

□ Create Git repository

Language: Java Kotlin Groovy **Scala** HTML +

Build system: **sbt** IntelliJ

JDK: 11 version 11.0.16

sbt: 1.9.7 □ Download sources

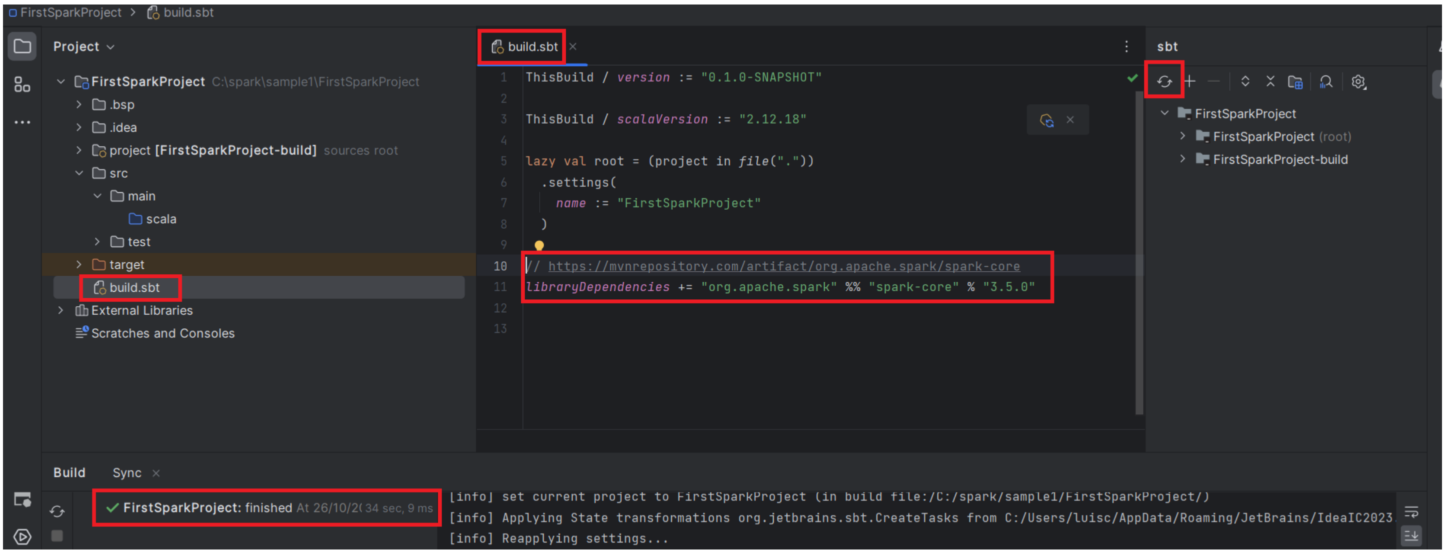
Scala: 2.12.18 ☑ Download sources

Package prefix: Such as 'org.example.application'

□ Add sample code

6. Set the build.sbt file

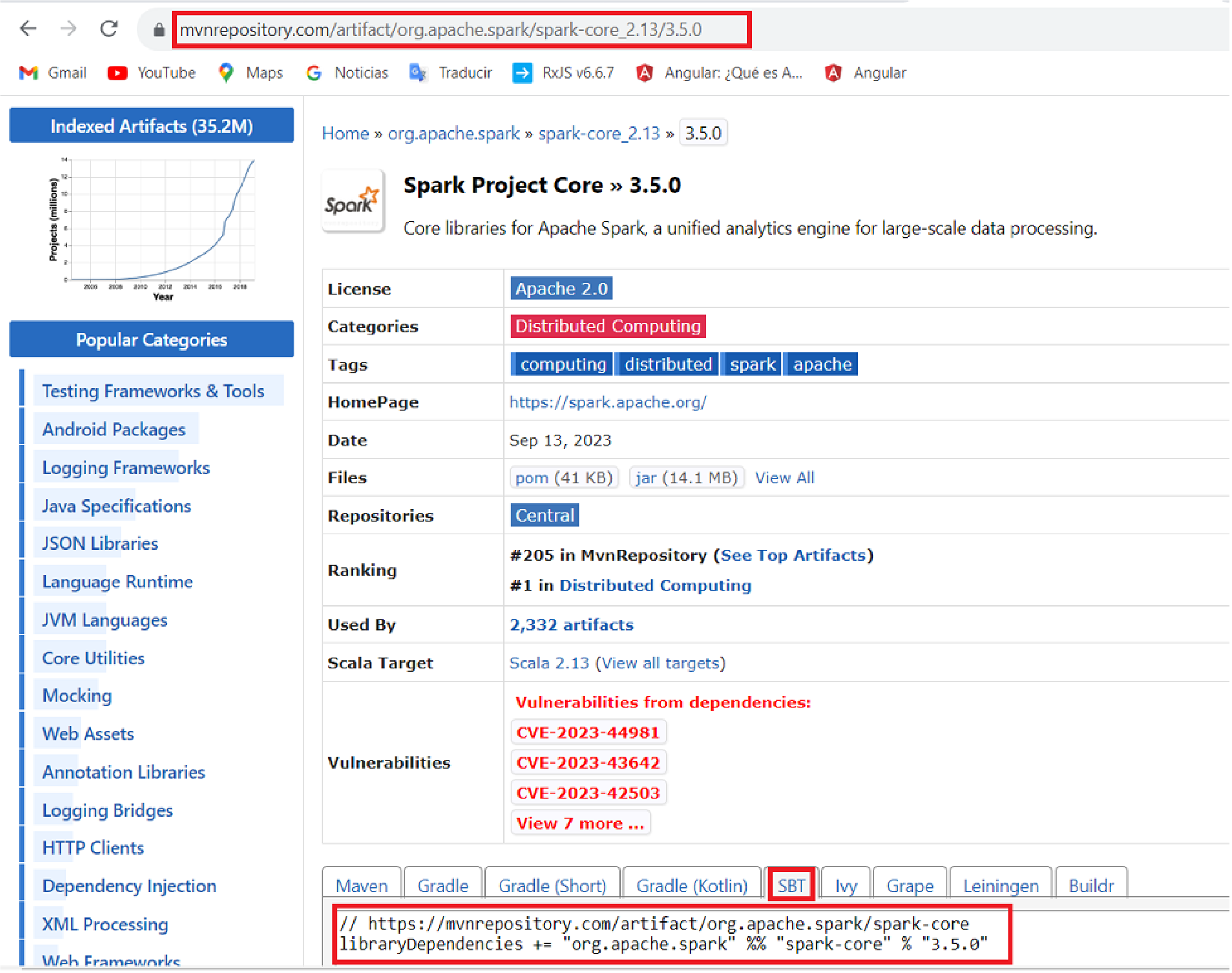
```
ThisBuild / version := "0.1.0-SNAPSHOT"
ThisBuild / scalaVersion := "2.12.18"

lazy val root = (project in file("."))
  .settings(
    name := "FirstSparkProject"
  )

// https://mvnrepository.com/artifact/org.apache.spark/spark-core
libraryDependencies += "org.apache.spark" %% "spark-core" % "3.5.0"
```



```
ThisBuild / version := "0.1.0-SNAPSHOT"

ThisBuild / scalaVersion := "2.12.18"

lazy val root = (project in file("."))
  .settings(
    name := "FirstSparkProject"
  )

// https://mvnrepository.com/artifact/org.apache.spark/spark-core
libraryDependencies += "org.apache.spark" %% "spark-core" % "3.5.0"
```

✓ FirstSparkProject: finished At 26/10/2(34 sec, 9 ms

```
[info] set current project to FirstSparkProject (in build file:/C:/spark/sample1/FirstSparkProject/)
[info] Applying State transformations org.jetbrains.sbt.CreateTasks from C:/Users/luisc/AppData/Roaming/JetBrains/IdeaIC2023
[info] Reapplying settings...
```



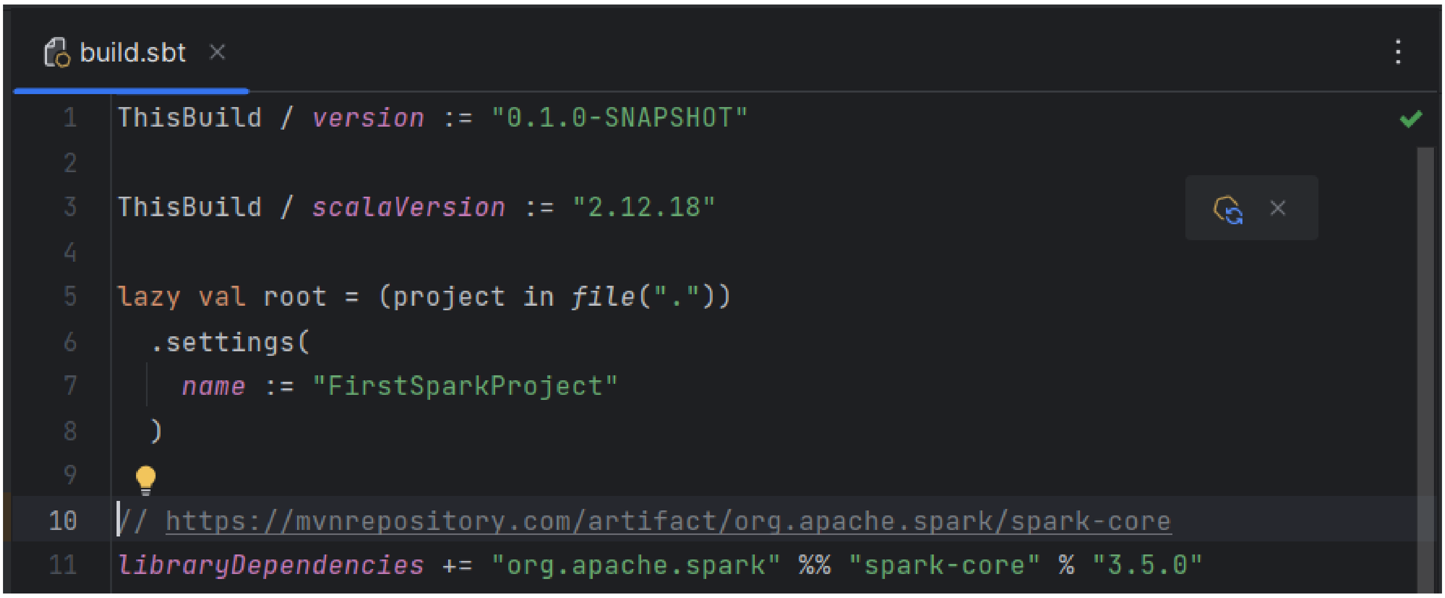
Home » org.apache.spark » spark-core_2.13 » 3.5.0

Spark Project Core » 3.5.0

Core libraries for Apache Spark, a unified analytics engine for large-scale data processing.

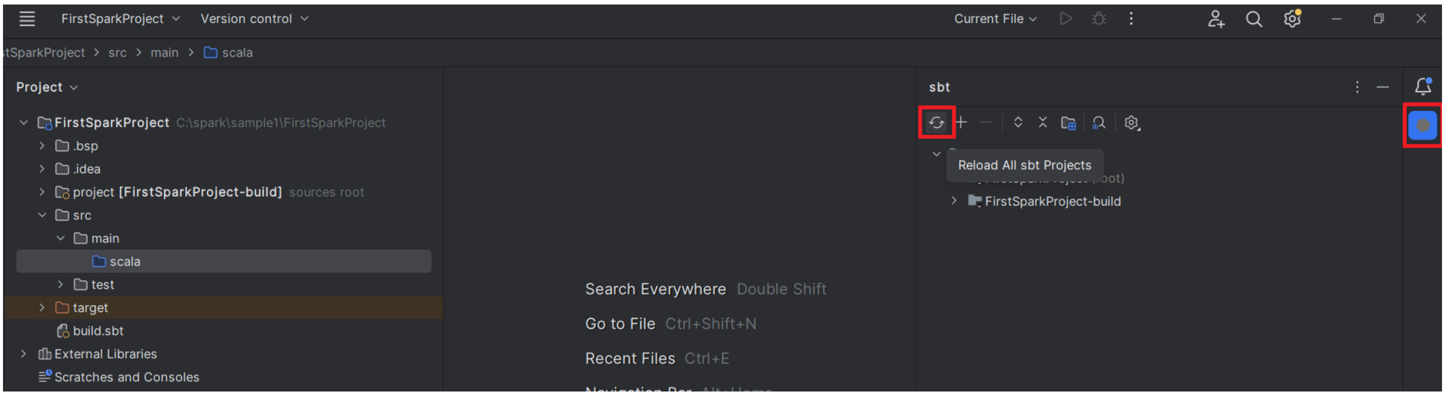
License	Apache 2.0
Categories	Distributed Computing
Tags	computing distributed spark apache
HomePage	https://spark.apache.org/
Date	Sep 13, 2023
Files	pom (41 KB) jar (14.1 MB) View All
Repositories	Central
Ranking	#205 in MvnRepository (See Top Artifacts) #1 in Distributed Computing
Used By	2,332 artifacts
Scala Target	Scala 2.13 (View all targets)
Vulnerabilities	Vulnerabilities from dependencies: CVE-2023-44981 CVE-2023-43642 CVE-2023-42503 View 7 more ...

```
// https://mvnrepository.com/artifact/org.apache.spark/spark-core
libraryDependencies += "org.apache.spark" %% "spark-core" % "3.5.0"
```

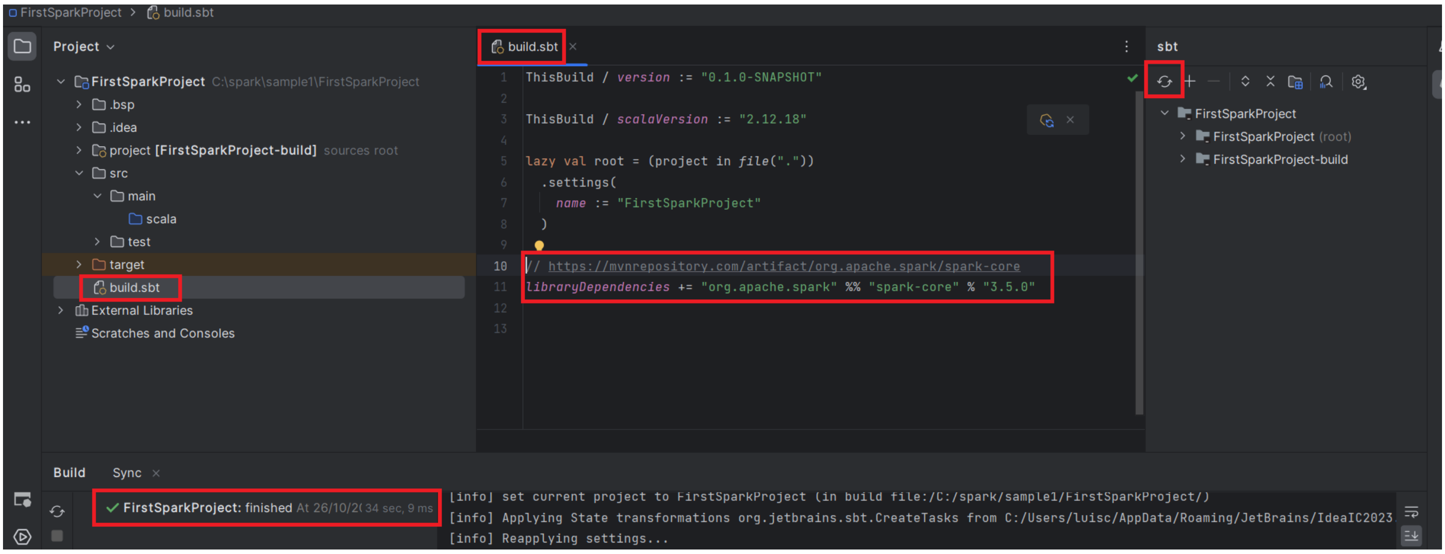


```
ThisBuild / version := "0.1.0-SNAPSHOT"

ThisBuild / scalaVersion := "2.12.18"

lazy val root = (project in file("."))
  .settings(
    name := "FirstSparkProject"
  )

// https://mvnrepository.com/artifact/org.apache.spark/spark-core
libraryDependencies += "org.apache.spark" %% "spark-core" % "3.5.0"
```

Press the "sbt" button and also press the button "Reload sbt projects" to load the spark library dependency



We check the project was reloaded with all the dependencies including the new spark dependency



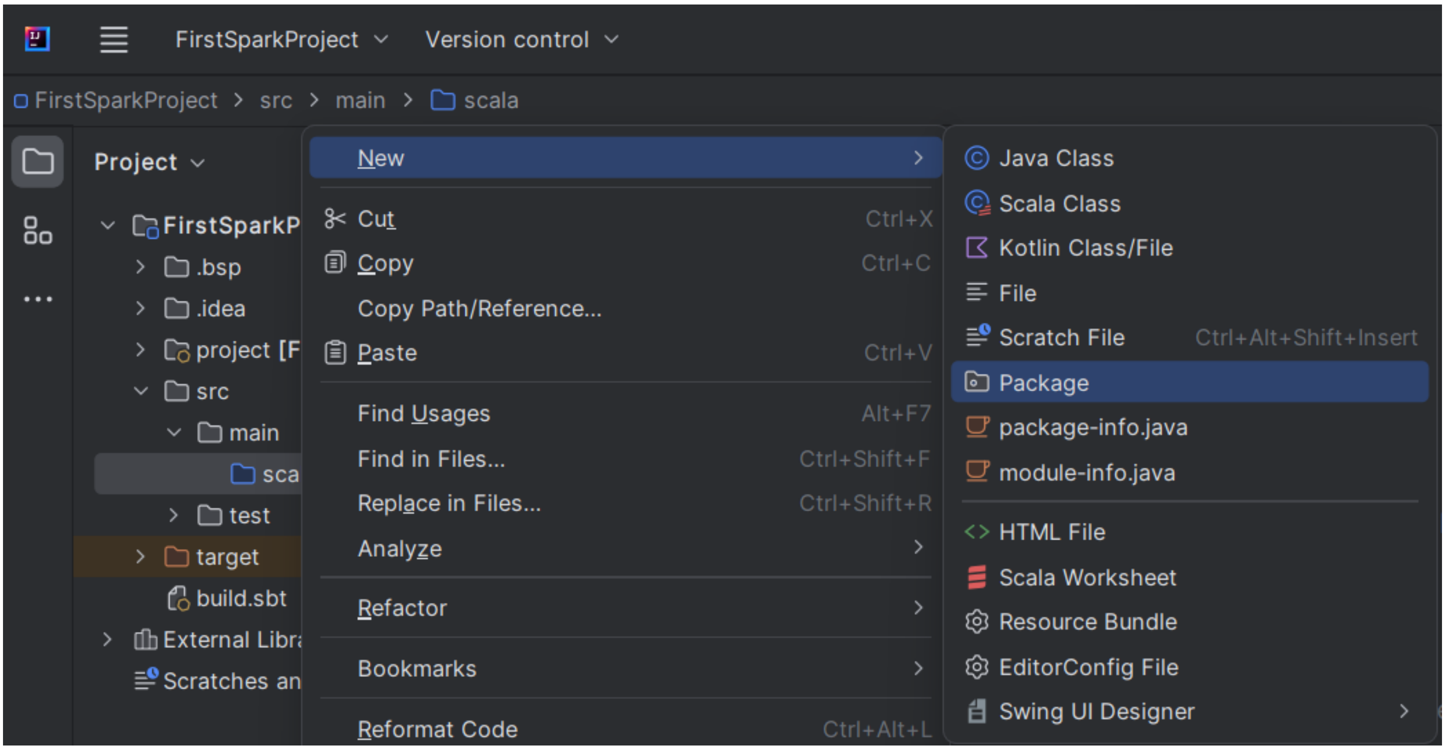
```
ThisBuild / version := "0.1.0-SNAPSHOT"

ThisBuild / scalaVersion := "2.12.18"

lazy val root = (project in file("."))
  .settings(
    name := "FirstSparkProject"
  )

// https://mvnrepository.com/artifact/org.apache.spark/spark-core
libraryDependencies += "org.apache.spark" %% "spark-core" % "3.5.0"
```

FirstSparkProject: finished At 26/10/2(34 sec, 9 ms

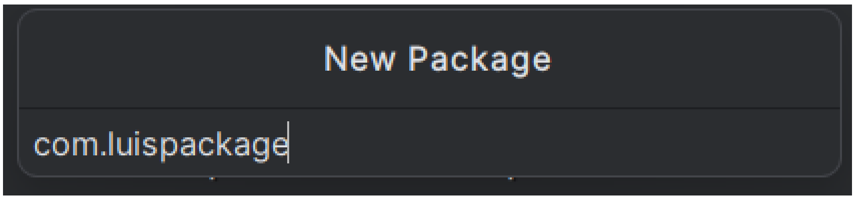
```
[info] set current project to FirstSparkProject (in build file:/C:/spark/sample1/FirstSparkProject/)
[info] Applying State transformations org.jetbrains.sbt.CreateTasks from C:/Users/luisc/AppData/Roaming/JetBrains/IdeaIC2023
[info] Reapplying settings...
```

7. Create a package and the scala source code file

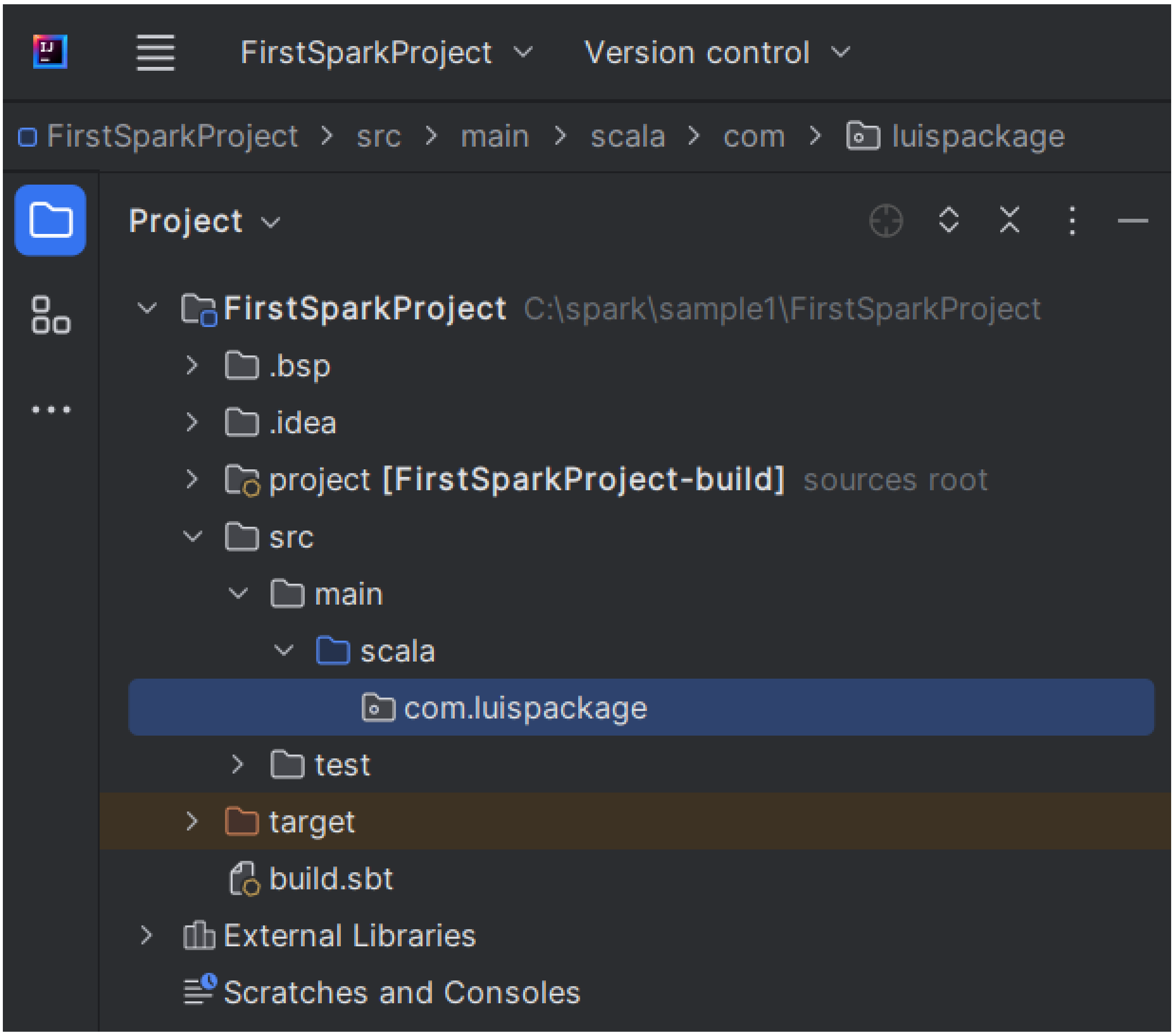
First we create a new package



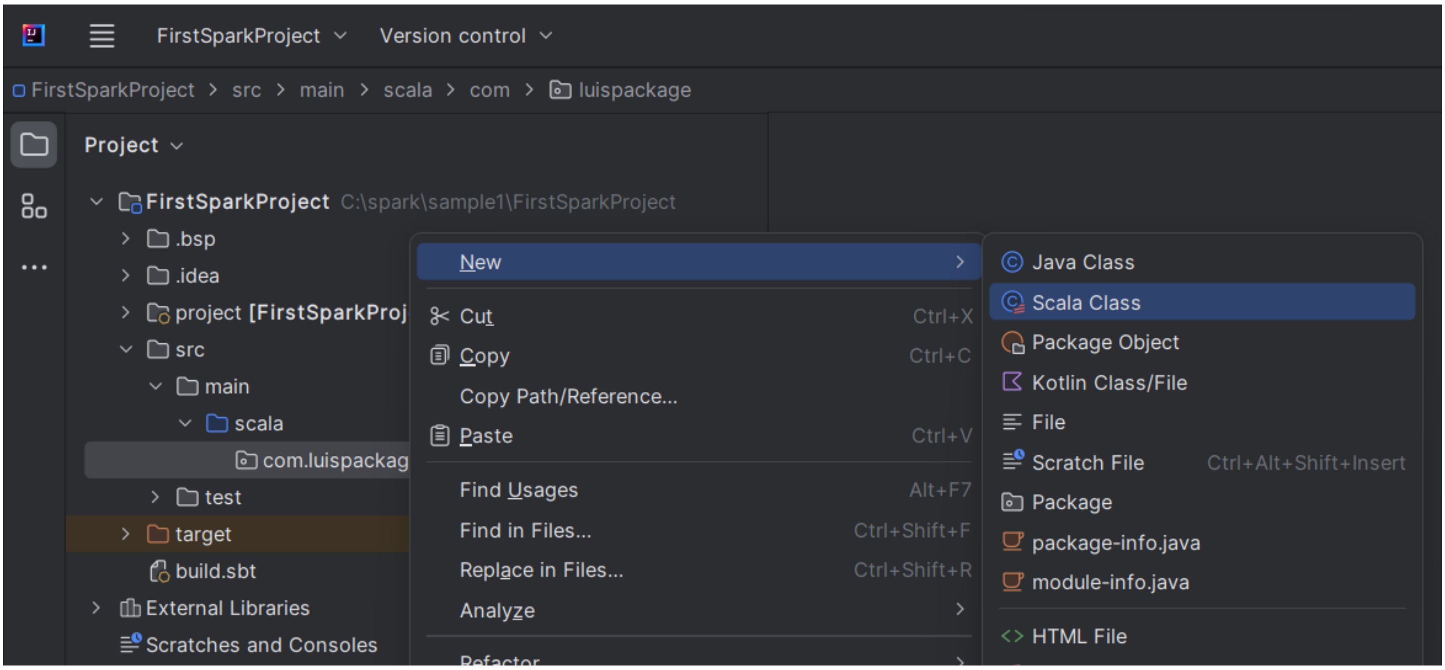
We set the new package name

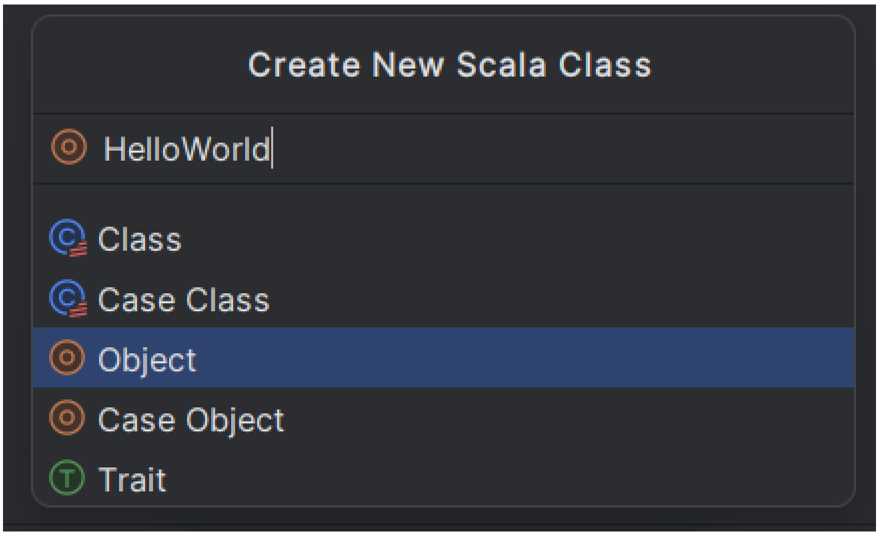


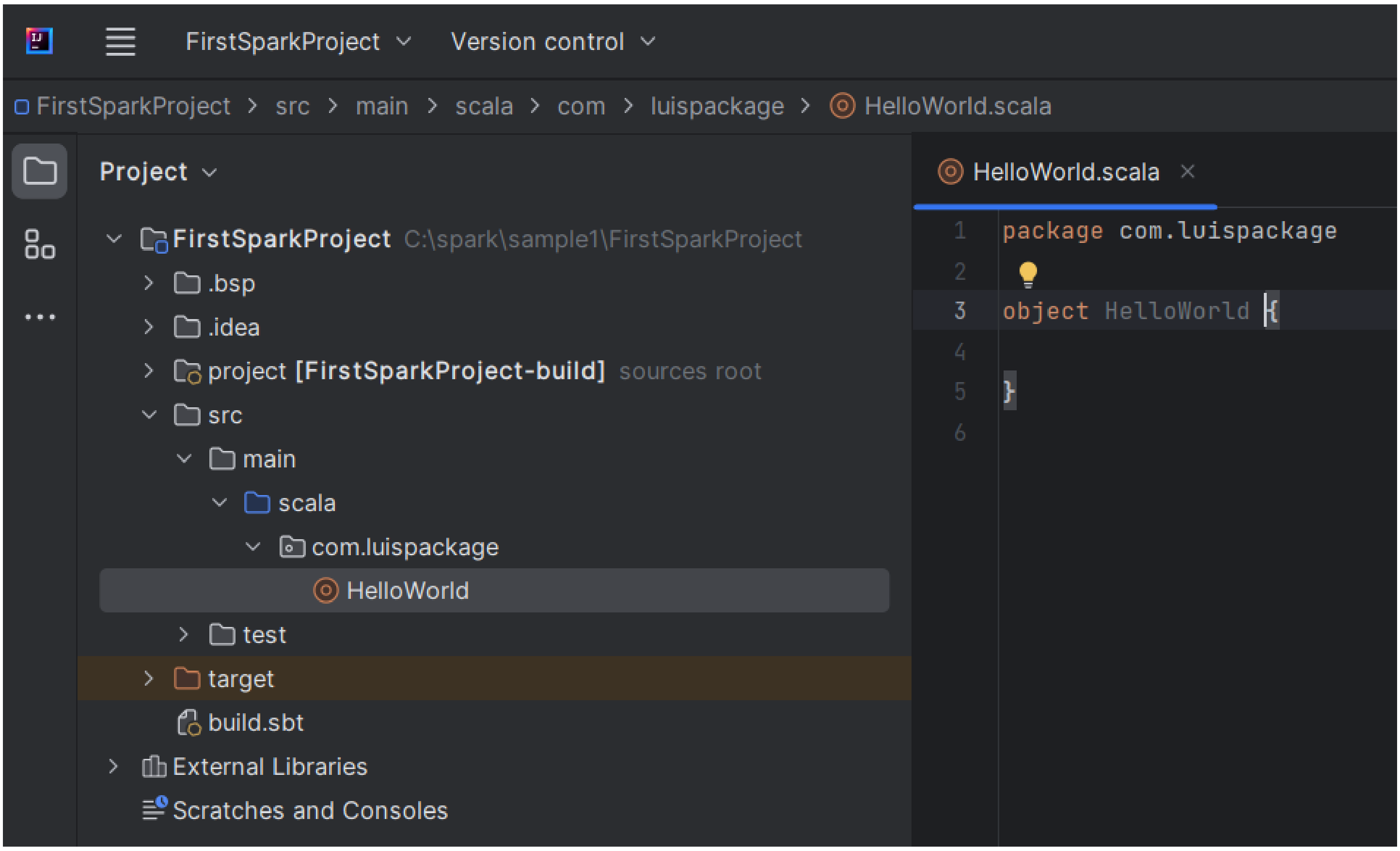
com.luispackage


FirstSparkProject
Version control
FirstSparkProject > src > main > scala > com > luispackage
Project
FirstSparkProject C:\spark\sample1\FirstSparkProject
.bsp
.idea
project [FirstSparkProject-build] sources root
src
main
scala
com.luispackage
test
target
build.sbt
External Libraries
Scratches and Consoles

We create a new scala Object


FirstSparkProject
Version control
FirstSparkProject > src > main > scala > com > luispackage
Project
FirstSparkProject C:\spark\sample1\FirstSparkProject
.bsp
.idea
project [FirstSparkProj
src
main
scala
com.luispackag
test
target
build.sbt
External Libraries
Scratches and Consoles
New
Cut Ctrl+X
Copy Ctrl+C
Copy Path/Reference...
Paste Ctrl+V
Find Usages Alt+F7
Find in Files... Ctrl+Shift+F
Replace in Files... Ctrl+Shift+R
Analyze
Refactor
Java Class
Scala Class
Package Object
Kotlin Class/File
File
Scratch File Ctrl+Alt+Shift+Insert
Package
package-info.java
module-info.java
HTML File

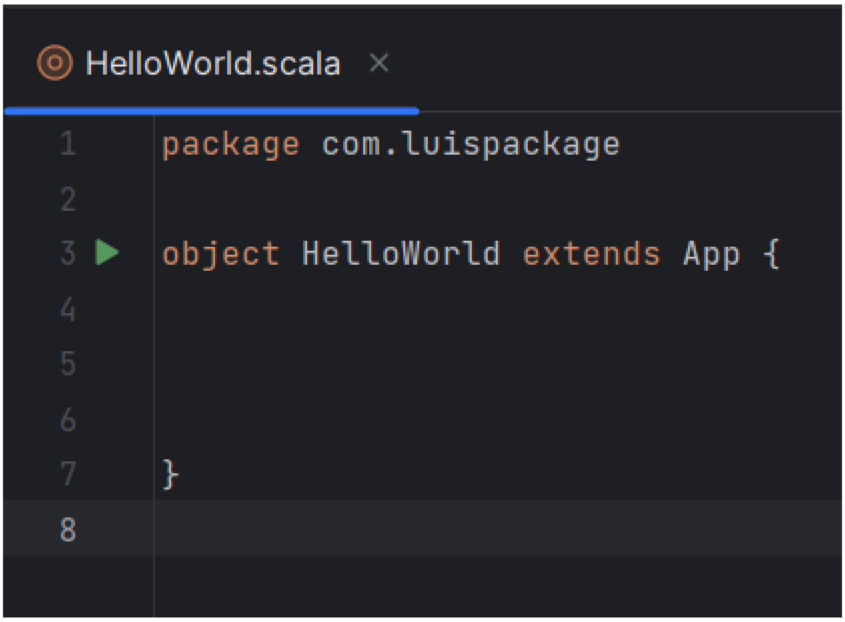

Create New Scala Class
HelloWorld
Class
Case Class
Object
Case Object
Trait


FirstSparkProject
Version control
FirstSparkProject > src > main > scala > com > luispackage > HelloWorld.scala
Project
FirstSparkProject C:\spark\sample1\FirstSparkProject
.bsp
.idea
project [FirstSparkProject-build] sources root
src
main
scala
com.luispackage
HelloWorld
test
target
build.sbt
External Libraries
Scratches and Consoles
HelloWorld.scala

```scala
package com.luispackage

object HelloWorld {

}
```

8. We enter the spark main file code

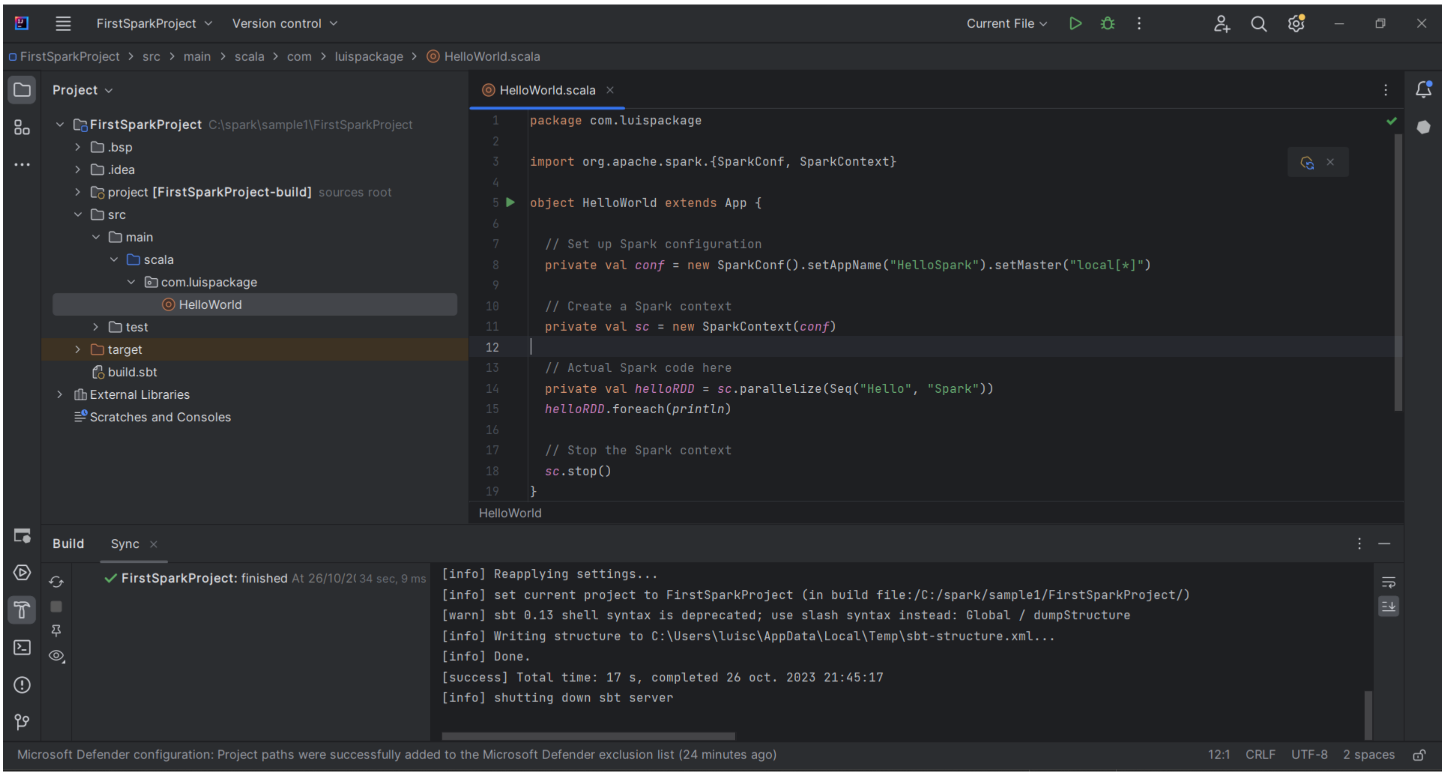
We first extends the object from the App



```scala
package com.luispackage

object HelloWorld extends App {



}
```

Then we enter the rest of the code

```scala
package com.luispackage

import org.apache.spark.{SparkConf, SparkContext}

object HelloWorld extends App {
  // Set up Spark configuration
  private val conf = new SparkConf().setAppName("HelloSpark").setMaster("local[*]")

  // Create a Spark context
  private val sc = new SparkContext(conf)

  // Actual Spark code here
  private val helloRDD = sc.parallelize(Seq("Hello", "Spark"))
  helloRDD.foreach(println)

  // Stop the Spark context
  sc.stop()
}
```



```
package com.luispackage

import org.apache.spark.{SparkConf, SparkContext}

object HelloWorld extends App {

  // Set up Spark configuration
  private val conf = new SparkConf().setAppName("HelloSpark").setMaster("local[*]")

  // Create a Spark context
  private val sc = new SparkContext(conf)

  // Actual Spark code here
  private val helloRDD = sc.parallelize(Seq("Hello", "Spark"))
  helloRDD.foreach(println)

  // Stop the Spark context
  sc.stop()
}
```

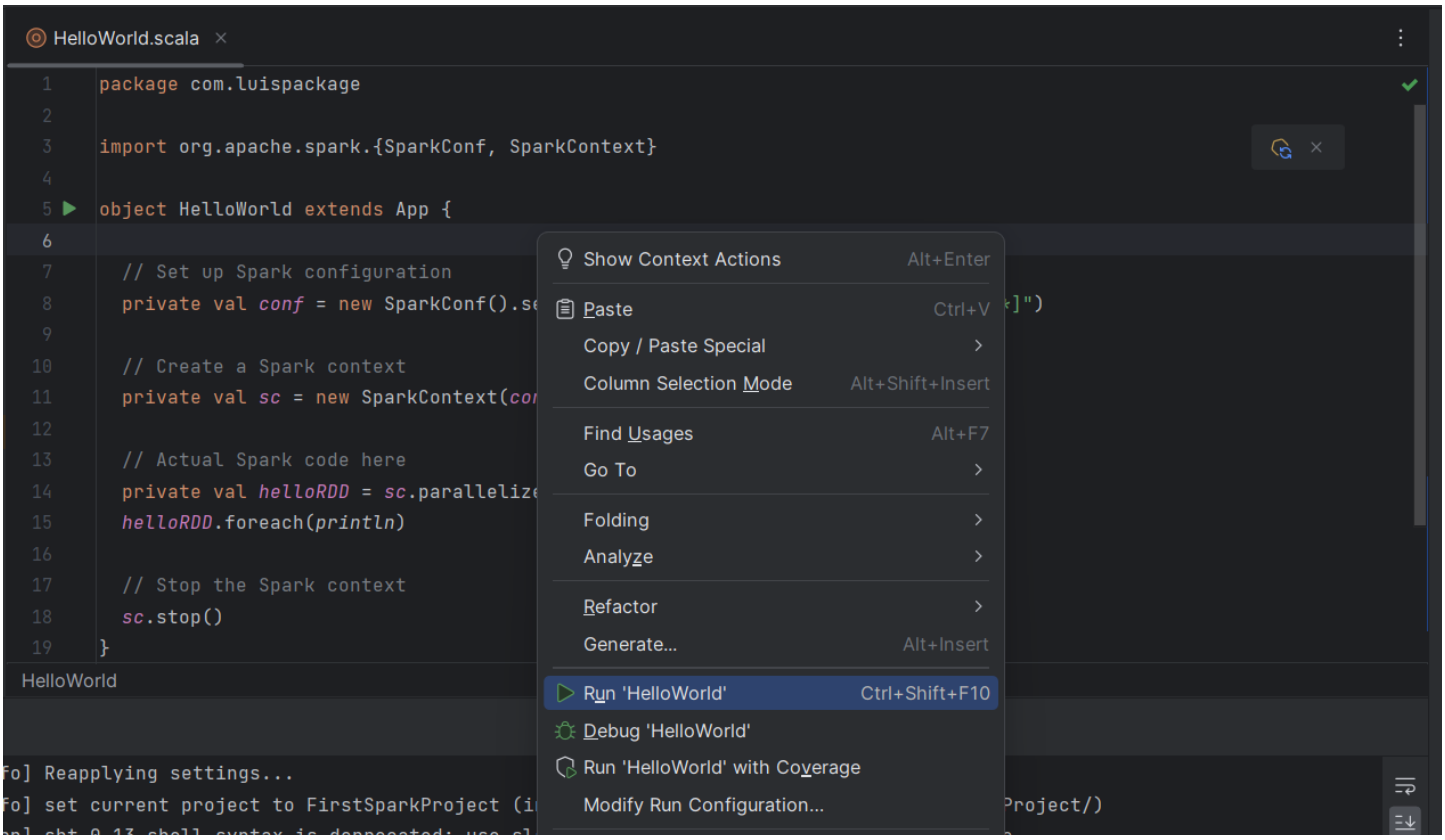
✓ FirstSparkProject: finished At 26/10/2(34 sec, 9 ms

```
[info] Reapplying settings...
[info] set current project to FirstSparkProject (in build file:/C:/spark/sample1/FirstSparkProject/)
[warn] sbt 0.13 shell syntax is deprecated; use slash syntax instead: Global / dumpStructure
[info] Writing structure to C:\Users\luisc\AppData\Local\Temp\sbt-structure.xml...
[info] Done.
[success] Total time: 17 s, completed 26 oct. 2023 21:45:17
[info] shutting down sbt server
```

Microsoft Defender configuration: Project paths were successfully added to the Microsoft Defender exclusion list (24 minutes ago)

9. How to run the spark scala application

We right click on the scala code and select the "Run HelloWorld"

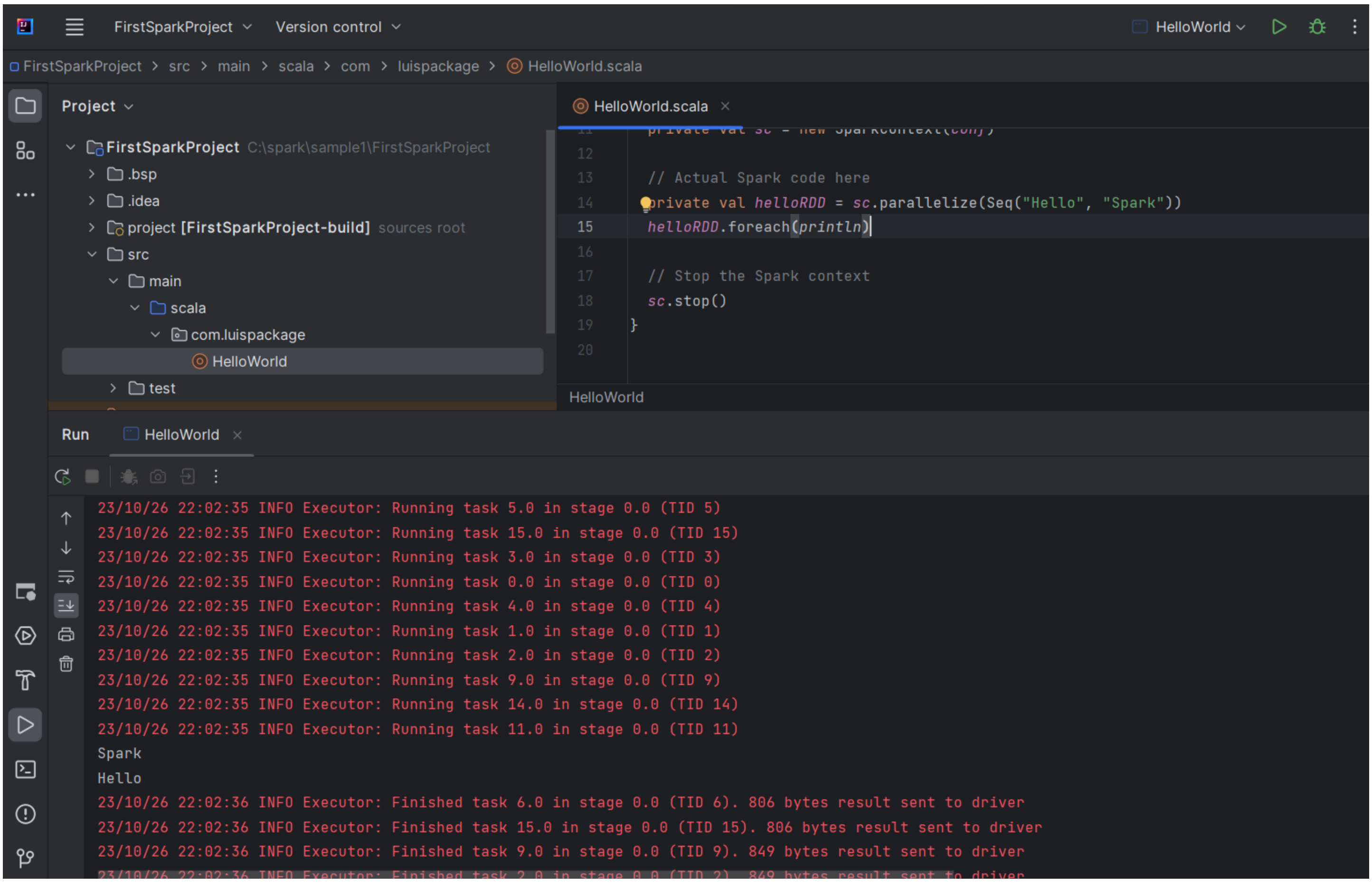


HelloWorld.scala
HelloWorld
Show Context Actions Alt+Enter
Paste Ctrl+V
Copy / Paste Special
Column Selection Mode Alt+Shift+Insert
Find Usages Alt+F7
Go To
Folding
Analyze
Refactor
Generate... Alt+Insert
Run 'HelloWorld' Ctrl+Shift+F10
Debug 'HelloWorld'
Run 'HelloWorld' with Coverage
Modify Run Configuration...

```
package com.luispackage

import org.apache.spark.{SparkConf, SparkContext}

object HelloWorld extends App {

  // Set up Spark configuration
  private val conf = new SparkConf().se                                  ]")

  // Create a Spark context
  private val sc = new SparkContext(co

  // Actual Spark code here
  private val helloRDD = sc.parallelize
  helloRDD.foreach(println)

  // Stop the Spark context
  sc.stop()
}
```

```
fo] Reapplying settings...
fo] set current project to FirstSparkProject (i                          Project/)
```

This is the run output

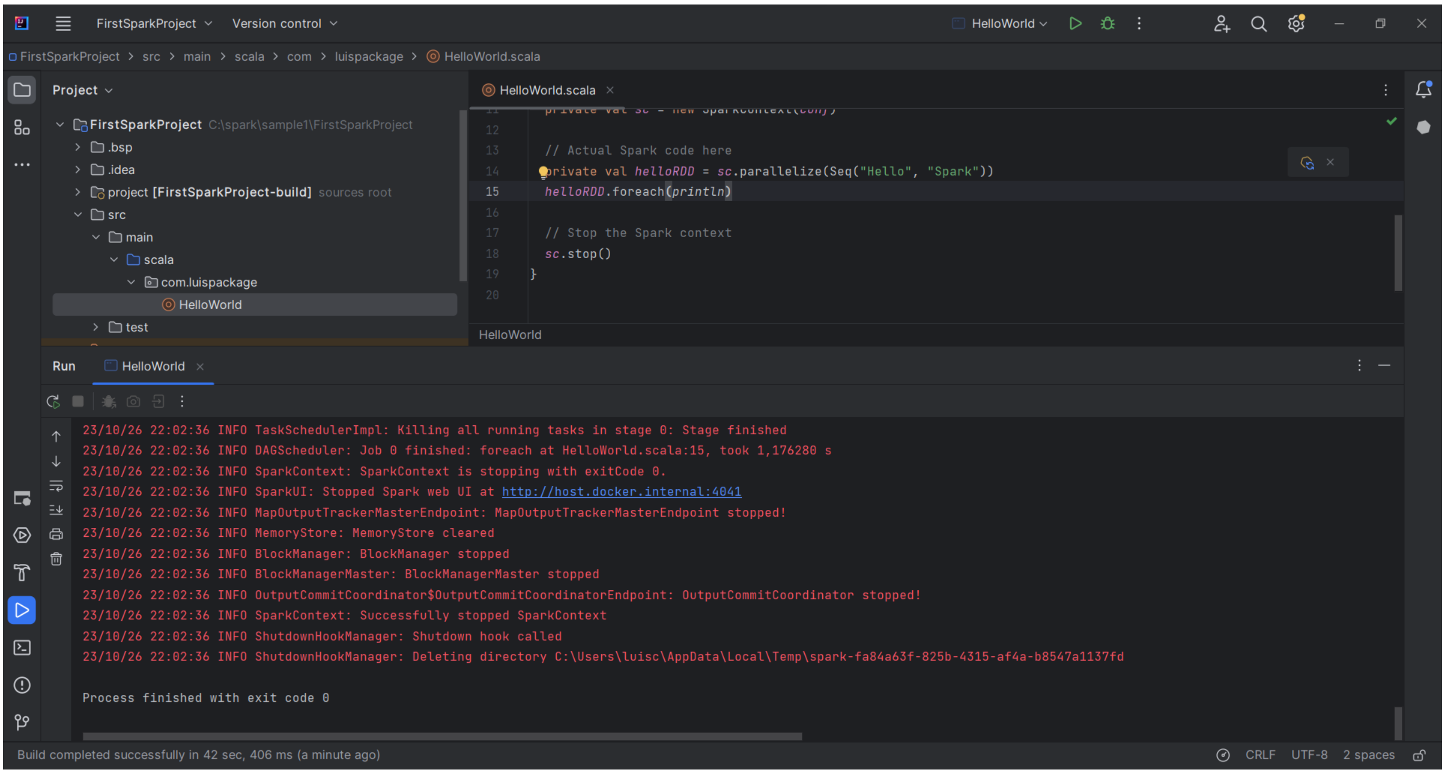


FirstSparkProject
Version control
HelloWorld
FirstSparkProject > src > main > scala > com > luispackage > HelloWorld.scala
Project
FirstSparkProject C:\spark\sample1\FirstSparkProject
.bsp
.idea
project [FirstSparkProject-build] sources root
src
main
scala
com.luispackage
HelloWorld
test
HelloWorld.scala
HelloWorld
Run
HelloWorld

```
  // Actual Spark code here
  private val helloRDD = sc.parallelize(Seq("Hello", "Spark"))
  helloRDD.foreach(println)

  // Stop the Spark context
  sc.stop()
}
```

```
23/10/26 22:02:35 INFO Executor: Running task 5.0 in stage 0.0 (TID 5)
23/10/26 22:02:35 INFO Executor: Running task 15.0 in stage 0.0 (TID 15)
23/10/26 22:02:35 INFO Executor: Running task 3.0 in stage 0.0 (TID 3)
23/10/26 22:02:35 INFO Executor: Running task 0.0 in stage 0.0 (TID 0)
23/10/26 22:02:35 INFO Executor: Running task 4.0 in stage 0.0 (TID 4)
23/10/26 22:02:35 INFO Executor: Running task 1.0 in stage 0.0 (TID 1)
23/10/26 22:02:35 INFO Executor: Running task 2.0 in stage 0.0 (TID 2)
23/10/26 22:02:35 INFO Executor: Running task 9.0 in stage 0.0 (TID 9)
23/10/26 22:02:35 INFO Executor: Running task 14.0 in stage 0.0 (TID 14)
23/10/26 22:02:35 INFO Executor: Running task 11.0 in stage 0.0 (TID 11)
Spark
Hello
23/10/26 22:02:36 INFO Executor: Finished task 6.0 in stage 0.0 (TID 6). 806 bytes result sent to driver
23/10/26 22:02:36 INFO Executor: Finished task 15.0 in stage 0.0 (TID 15). 806 bytes result sent to driver
23/10/26 22:02:36 INFO Executor: Finished task 9.0 in stage 0.0 (TID 9). 849 bytes result sent to driver
23/10/26 22:02:36 INFO Executor: Finished task 2.0 in stage 0.0 (TID 2). 849 bytes result sent to driver
```

Also see the Process finished with exit code 0


FirstSparkProject
Version control
HelloWorld
FirstSparkProject > src > main > scala > com > luispackage > HelloWorld.scala
Project
FirstSparkProject C:\spark\sample1\FirstSparkProject
.bsp
.idea
project [FirstSparkProject-build] sources root
src
main
scala
com.luispackage
HelloWorld
test
HelloWorld.scala
Run
HelloWorld
Build completed successfully in 42 sec, 406 ms (a minute ago)
CRLF
UTF-8
2 spaces

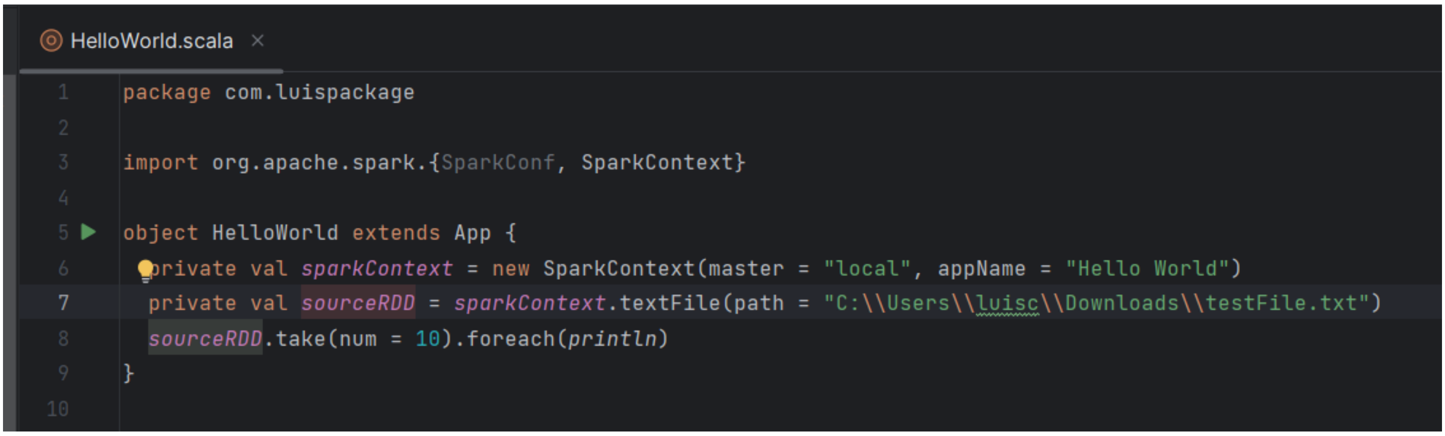
```
  private val sc = new SparkContext(conf)

  // Actual Spark code here
  private val helloRDD = sc.parallelize(Seq("Hello", "Spark"))
  helloRDD.foreach(println)

  // Stop the Spark context
  sc.stop()
}
```

```
23/10/26 22:02:36 INFO TaskSchedulerImpl: Killing all running tasks in stage 0: Stage finished
23/10/26 22:02:36 INFO DAGScheduler: Job 0 finished: foreach at HelloWorld.scala:15, took 1,176280 s
23/10/26 22:02:36 INFO SparkContext: SparkContext is stopping with exitCode 0.
23/10/26 22:02:36 INFO SparkUI: Stopped Spark web UI at http://host.docker.internal:4041
23/10/26 22:02:36 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
23/10/26 22:02:36 INFO MemoryStore: MemoryStore cleared
23/10/26 22:02:36 INFO BlockManager: BlockManager stopped
23/10/26 22:02:36 INFO BlockManagerMaster: BlockManagerMaster stopped
23/10/26 22:02:36 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
23/10/26 22:02:36 INFO SparkContext: Successfully stopped SparkContext
23/10/26 22:02:36 INFO ShutdownHookManager: Shutdown hook called
23/10/26 22:02:36 INFO ShutdownHookManager: Deleting directory C:\Users\luisc\AppData\Local\Temp\spark-fa84a63f-825b-4315-af4a-b8547a1137fd

Process finished with exit code 0
```

10. Another scala spark application sample

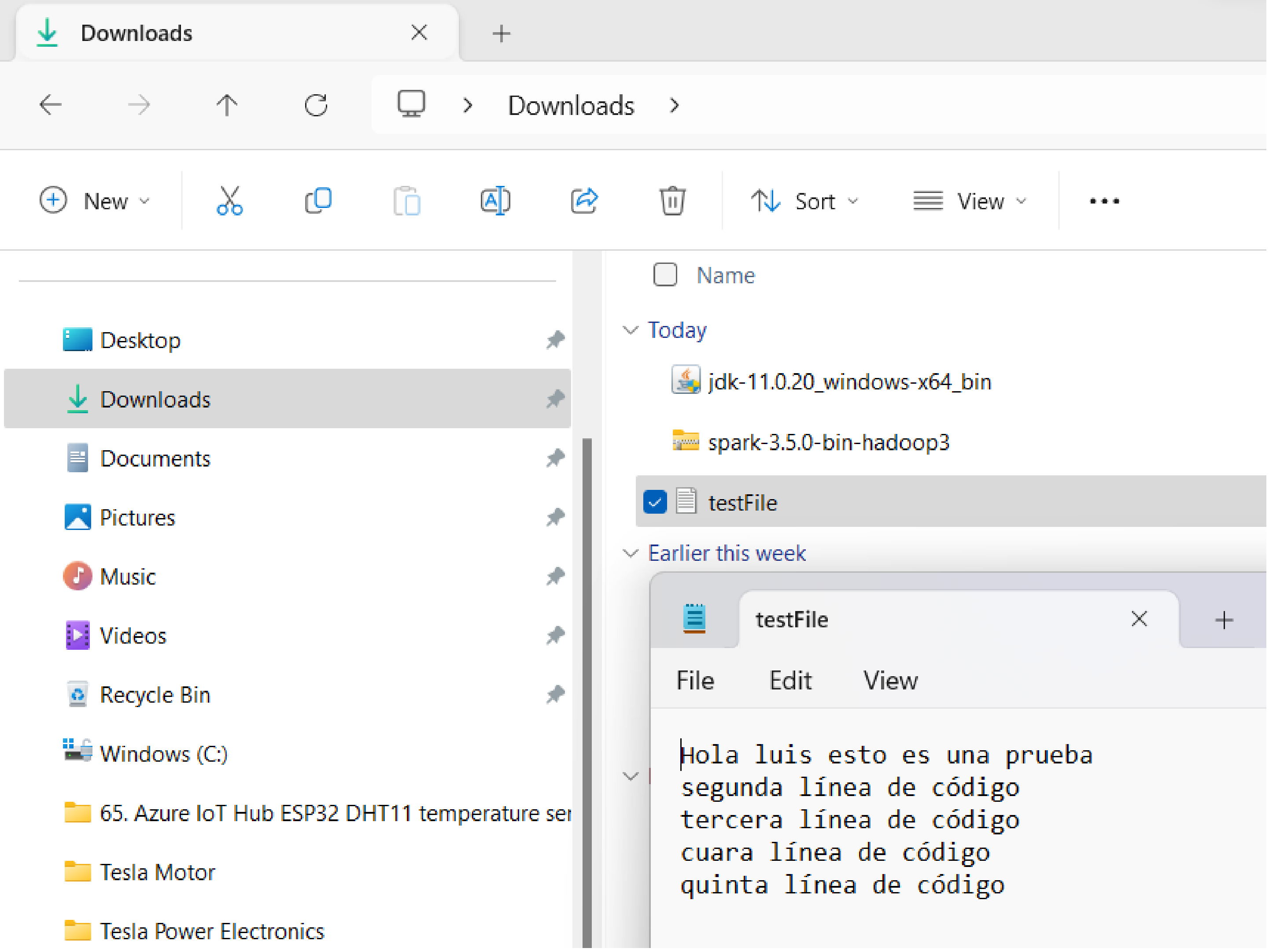
This is the new application source code:

```
package com.luispackage

import org.apache.spark.{SparkConf, SparkContext}

object HelloWorld extends App {
  private val sparkContext = new SparkContext(master = "local", appName = "Hello World")
  private val sourceRDD = sparkContext.textFile(path = "C:\\Users\\luisc\\Downloads\\testFile.
  sourceRDD.take(num = 10).foreach(println)
}
```



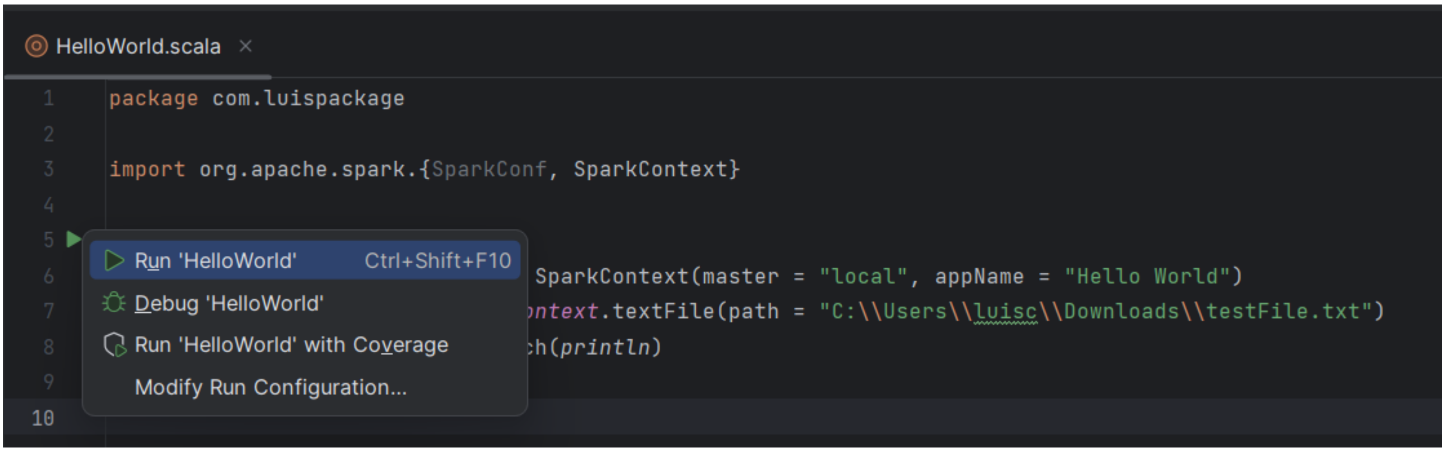
```
package com.luispackage

import org.apache.spark.{SparkConf, SparkContext}

object HelloWorld extends App {
  private val sparkContext = new SparkContext(master = "local", appName = "Hello World")
  private val sourceRDD = sparkContext.textFile(path = "C:\\Users\\luisc\\Downloads\\testFile.txt")
  sourceRDD.take(num = 10).foreach(println)
}
```

We also create the testFile.txt

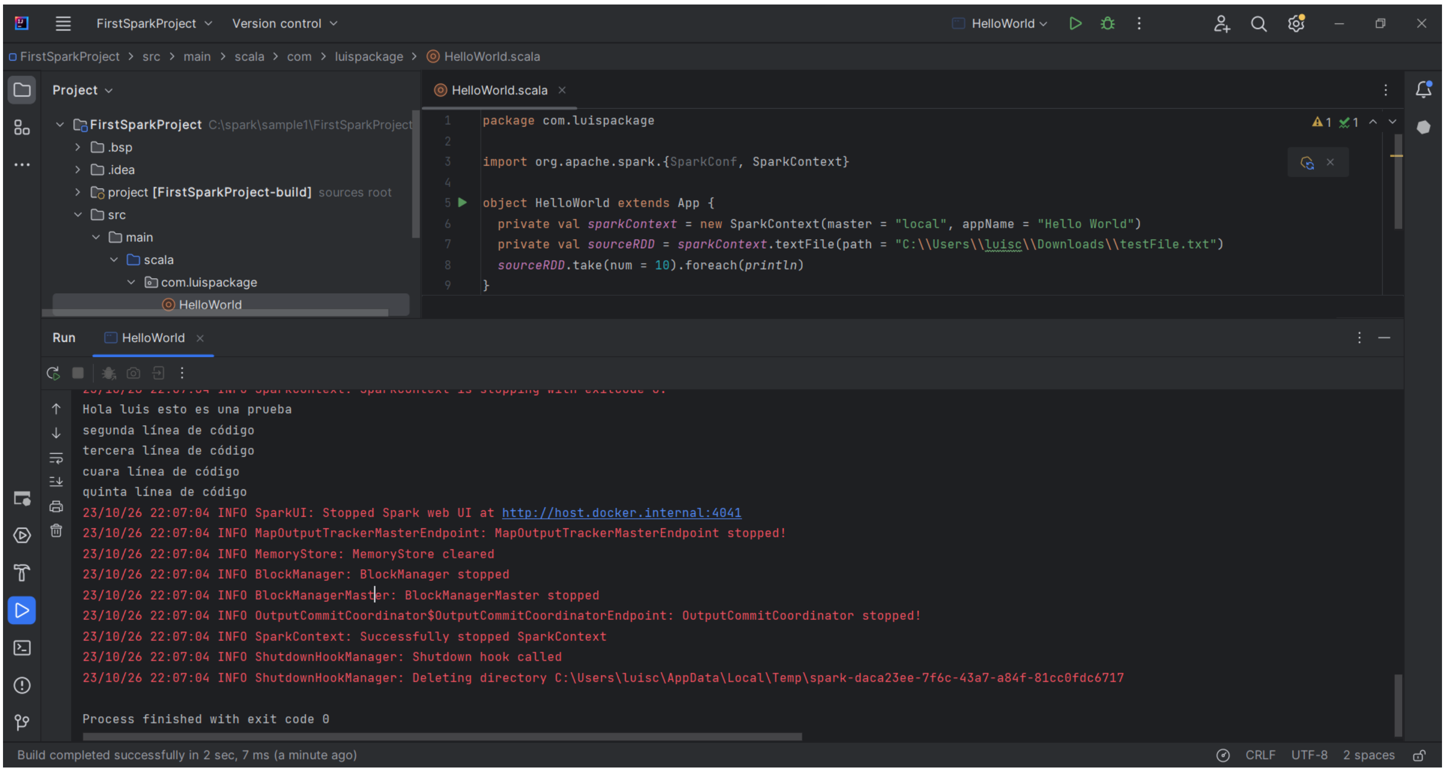


```
Hola luis esto es una prueba
segunda línea de código
tercera línea de código
cuara línea de código
quinta línea de código
```

Now we run the application


HelloWorld.scala
Run 'HelloWorld' Ctrl+Shift+F10
Debug 'HelloWorld'
Run 'HelloWorld' with Coverage
Modify Run Configuration...

```
package com.luispackage

import org.apache.spark.{SparkConf, SparkContext}


SparkContext(master = "local", appName = "Hello World")
ontext.textFile(path = "C:\\Users\\luisc\\Downloads\\testFile.txt")
ch(println)
```


FirstSparkProject
Version control
HelloWorld
FirstSparkProject > src > main > scala > com > luispackage > HelloWorld.scala
Project
FirstSparkProject C:\spark\sample1\FirstSparkProject
.bsp
.idea
project [FirstSparkProject-build] sources root
src
main
scala
com.luispackage
HelloWorld
HelloWorld.scala
Run
HelloWorld
CRLF
UTF-8
2 spaces

```scala
package com.luispackage

import org.apache.spark.{SparkConf, SparkContext}

object HelloWorld extends App {
  private val sparkContext = new SparkContext(master = "local", appName = "Hello World")
  private val sourceRDD = sparkContext.textFile(path = "C:\\Users\\luisc\\Downloads\\testFile.txt")
  sourceRDD.take(num = 10).foreach(println)
}
```

```
23/10/26 22:07:04 INFO SparkContext: SparkContext is stopping with exitcode 0.
Hola luis esto es una prueba
segunda línea de código
tercera línea de código
cuara línea de código
quinta línea de código
23/10/26 22:07:04 INFO SparkUI: Stopped Spark web UI at http://host.docker.internal:4041
23/10/26 22:07:04 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
23/10/26 22:07:04 INFO MemoryStore: MemoryStore cleared
23/10/26 22:07:04 INFO BlockManager: BlockManager stopped
23/10/26 22:07:04 INFO BlockManagerMaster: BlockManagerMaster stopped
23/10/26 22:07:04 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
23/10/26 22:07:04 INFO SparkContext: Successfully stopped SparkContext
23/10/26 22:07:04 INFO ShutdownHookManager: Shutdown hook called
23/10/26 22:07:04 INFO ShutdownHookManager: Deleting directory C:\Users\luisc\AppData\Local\Temp\spark-daca23ee-7f6c-43a7-a84f-81cc0fdc6717

Process finished with exit code 0
```

Build completed successfully in 2 sec, 7 ms (a minute ago)