

Crear un proyecto en Apache Spark:

1. Configurar el Entorno de Desarrollo

- **Instalar Java:** Asegúrate de tener Java instalado.
- **Instalar Apache Spark:** Descarga e instala Spark como se mencionó anteriormente.
- **Instalar un IDE:** Utiliza un entorno de desarrollo integrado (IDE) como IntelliJ IDEA o Eclipse.

2. Crear un Proyecto en Scala

Si prefieres trabajar con Scala, sigue estos pasos:

a. Crear un Proyecto con SBT

1. **Instalar SBT:** Descarga e instala SBT desde sbt download.
2. **Crear la Estructura del Proyecto:**

```
mkdir MiProyectoSpark
```

```
cd MiProyectoSpark
```

```
mkdir -p src/main/scala
```

3. Crear el Archivo build.sbt:

```
name := "MiProyectoSpark"
```

```
version := "0.1"
```

```
scalaVersion := "2.12.10"
```

```
libraryDependencies += "org.apache.spark" %% "spark-core" % "3.1.2"
```

```
libraryDependencies += "org.apache.spark" %% "spark-sql" % "3.1.2"
```

4. Escribir el Código Scala: Crea un archivo src/main/scala/Main.scala:

```
import org.apache.spark.sql.Session
```

```
object Main {
```

```
  def main(args: Array[String]): Unit = {
```

```
    val spark = Session.builder
```

```

    .appName("MiProyectoSpark")

    .master("local[*]")

    .getOrCreate()

val data = Seq((1, "uno"), (2, "dos"), (3, "tres"))

val df = spark.createDataFrame(data).toDF("id", "nombre")

df.show()

spark.stop()

}

}

```

5. **Compilar y Ejecutar el Proyecto:**

```
sbt package
```

```
sbt run
```

3. **Crear un Proyecto en Python**

Si prefieres trabajar con Python, sigue estos pasos:

1. **Instalar PySpark:**

```
pip install pyspark
```

2. **Crear el Archivo del Proyecto:** Crea un archivo `mi_proyecto_spark.py`:

```

from pyspark.sql import SparkSession

if __name__ == "__main__":

    spark = SparkSession.builder \

        .appName("MiProyectoSpark") \

        .master("local[*]") \

        .getOrCreate()

    data = [(1, "uno"), (2, "dos"), (3, "tres")]

```

```
df = spark.createDataFrame(data, ["id", "nombre"])
```

```
df.show()
```

```
spark.stop()
```

3. Ejecutar el Proyecto:

```
python mi_proyecto_spark.py
```

4. Ejecutar el Proyecto en Spark

- Usar spark-submit:

```
spark-submit --master local[*] mi_proyecto_spark.py
```