

```
[1] # Instalar SDK Java 8
lapt-get install openjdk-8-jdk-headless -qq > /dev/null

[2] # Descargar Spark 3.2.2
wget -q https://archive.apache.org/dist/spark/spark-3.2.3/spark-3.2.3-bin-hadoop3.2.tgz

[3] # Descomprimir el archivo descargado de Spark
tar xf spark-3.2.3-bin-hadoop3.2.tgz

[4] # Establecer las variables de entorno
import os
os.environ["JAVA_HOME"] = "/usr/lib/jvm/java-8-openjdk-amd64"
os.environ["SPARK_HOME"] = "/content/spark-3.2.3-bin-hadoop3.2"

[5] # Instalar la librería findspark
pip install -q findspark
```

```
[6] # Instalar pyspark

!pip install -q pyspark

[7] ### verificar la instalación ###
import findspark
findspark.init()

[9] from pyspark.sql import SparkSession

spark = SparkSession.builder.master("local[*]").getOrCreate()

[10] # Crear un RDD desde una lista
data = [1, 2, 3, 4, 5]
rdd = spark.sparkContext.parallelize(data)

# Mostrar los datos del RDD
print("Contenido del RDD:", rdd.collect())

Contenido del RDD: [1, 2, 3, 4, 5]
```