```
In [1]:  !apt-get update # Update apt-get repository
```

Hit:1 http://security.ubuntu.com/ubuntu jammy-security InRelease
Hit:2 http://archive.ubuntu.com/ubuntu jammy InRelease
Hit:3 http://archive.ubuntu.com/ubuntu jammy-updates InRelease
Hit:4 http://archive.ubuntu.com/ubuntu jammy-backports InRelease
Hit:5 https://cloud.r-project.org/bin/linux/ubuntu jammy-cran40/ InRelease
Hit:6 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64 InRelease
Hit:7 https://r2u.stat.illinois.edu/ubuntu jammy InRelease
Hit:8 https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu jammy InRelease
Hit:9 https://ppa.launchpadcontent.net/graphics-drivers/ppa/ubuntu jammy InRelease
Hit:10 https://ppa.launchpadcontent.net/ubuntugis/ppa/ubuntu jammy InRelease
Reading package lists... Done
W: Skipping acquire of configured file 'main/source/Sources' as repository 'https://r2u.stat.illinois.edu/ubuntu jammy
InRelease' does not seem to provide it (sources.list entry misspelt?)

```
In [2]:  !apt-get install openjdk-8-jdk-headless -qq > /dev/null # Install Java.
```

```
In [9]:  !wget -q https://archive.apache.org/dist/spark/spark-4.0.0/spark-4.0.0-bin-hadoop3.tgz # Download Apache Sparks.
```

```
In [3]:  !tar xf spark-4.0.0-bin-hadoop3.tgz # Unzip the tgz file.
```

```
In [4]:  !pip install -q findspark # Install findspark. Adds PySpark to the System path during runtime.
```

```
In [5]:  # Set environment variables
         import os
```

```
In [6]:  # Initialize findspark
         import findspark
         findspark.init()
         findspark.find()
```

```
Out[6]:  '/usr/local/lib/python3.11/dist-packages/pyspark'
```

```
In [7]:  # Create a PySpark session
         from pyspark.sql import SparkSession
         spark = SparkSession.builder.appName("PySpark Bussiness Analysis").getOrCreate()
         spark
```

Out[7]: **SparkSession - in-memory**

SparkContext

[Spark UI](#)

Version	`v3.5.1`
Master	`local[*]`
AppName	`PySpark Bussiness Analysis`