

BIG DATA ANALYTICS
Installation of pyspark on windows

Step 1: Install Java

java -version

OUTPUT:

java version "22.0.2" 2024-07-16
Java(TM) SE Runtime Environment (build 22.0.2+9-70)
Java HotSpot(TM) 64-Bit Server VM (build 22.0.2+9-70, mixed mode, sharing)

Step 2: Install Python


python --version

OUTPUT:

Python 3.13.5


Step 3: Install Spark

OUTPUT:

 spark-4.0.0-bin-hadoop3.tgz	17-08-2025 13:39	Compressed Archi...	5,30,708 KB
---	------------------	---------------------	-------------

Step 4: Install Hadoop Winutils

OUTPUT:

 winutils.exe	17-08-2025 13:25	Application	117 KB
--	------------------	-------------	--------

Step 5: Install PySpark

pip install pyspark

OUTPUT:

Defaulting to user installation because normal site-packages is not writeable Requirement already satisfied: pyspark in c:\users\prab\appdata\roaming\python\python313\site-packages

(4.0.0) Requirement already satisfied: py4j==0.10.9.9 in
c:\users\prab\appdata\roaming\python\python313\site-packages (from pyspark) (0.10.9.9)

[notice] A new release of pip is available: 25.0.1 -> 25.2

[notice] To update, run: python.exe -m pip install --upgrade pip

Step 6: Verify the Installation

```
import pyspark

from pyspark.sql import SparkSession

spark = SparkSession.builder.appName("StudentExample").getOrCreate()

print(spark.version)
```

OUTPUT:

3.5.1

Step 7: Running a example pyspark code to check the working of successful installation

```
data = [

    ("Sudharsan", "Maths", 95),

    ("Priya", "Physics", 88),

    ("Arun", "Chemistry", 92),

    ("Meena", "Biology", 85),

]

df = spark.createDataFrame(data, ["Name", "Subject", "Marks"])

df.show()
```

OUTPUT:

Name	Subject	Marks
Sudharsan	Maths	95
Priya	Physics	88
Arun	Chemistry	92
Meena	Biology	85

-Prabhu chennimalai K D

23ALR066