**SreeVidya 23-12-23**

**Assessment 18**

**Spark Shell Commands:**

* **Launch Spark Shell (Scala):**

bash

*spark-shell*

* **Launch Spark Shell with Python (PySpark):**

bash

*pyspark*

* **Read a Text File:**

scala

*Val textFile = sc.textFile("path/to/textfile.txt")*

* **Count the Number of Lines:**

scala

*val count = textFile.count()*

* **Display First Few Lines:**

scala

*textFile.take(5). foreach(println)*

* **Filter Data:**

scala

*val filteredData = textFile.filter(line => line.contains("keyword"))*

* **Spark Submit Commands:**

When submitting a Spark application, you use the spark-submit command:

bash

*spark-submit --class YourMainClass --master local [2] YourApp.jar arg1 arg2*

**--class YourMainClass**: Specifies the main class of your Spark application.

**--master local [2]:** Specifies that Spark should run in local mode with 2 threads. Adjust the number of threads as needed.

**YourApp.jar**: Path to the compiled JAR file of your Spark application.

**arg1 arg2:** Command-line arguments that your Spark application might expect.

**DataFrame Operations:**

* **Read a CSV File into DataFrame:**

scala

*val df = spark.read.csv("path/to/csvfile.csv")*

* **Show DataFrame Schema:**

scala

*df. printSchema()*

* **Show First N Rows:**

scala

*df. show (5)*

* **Filter DataFrame:**

scala

*val filteredDF = df.filter(col("columnName") > 10)*

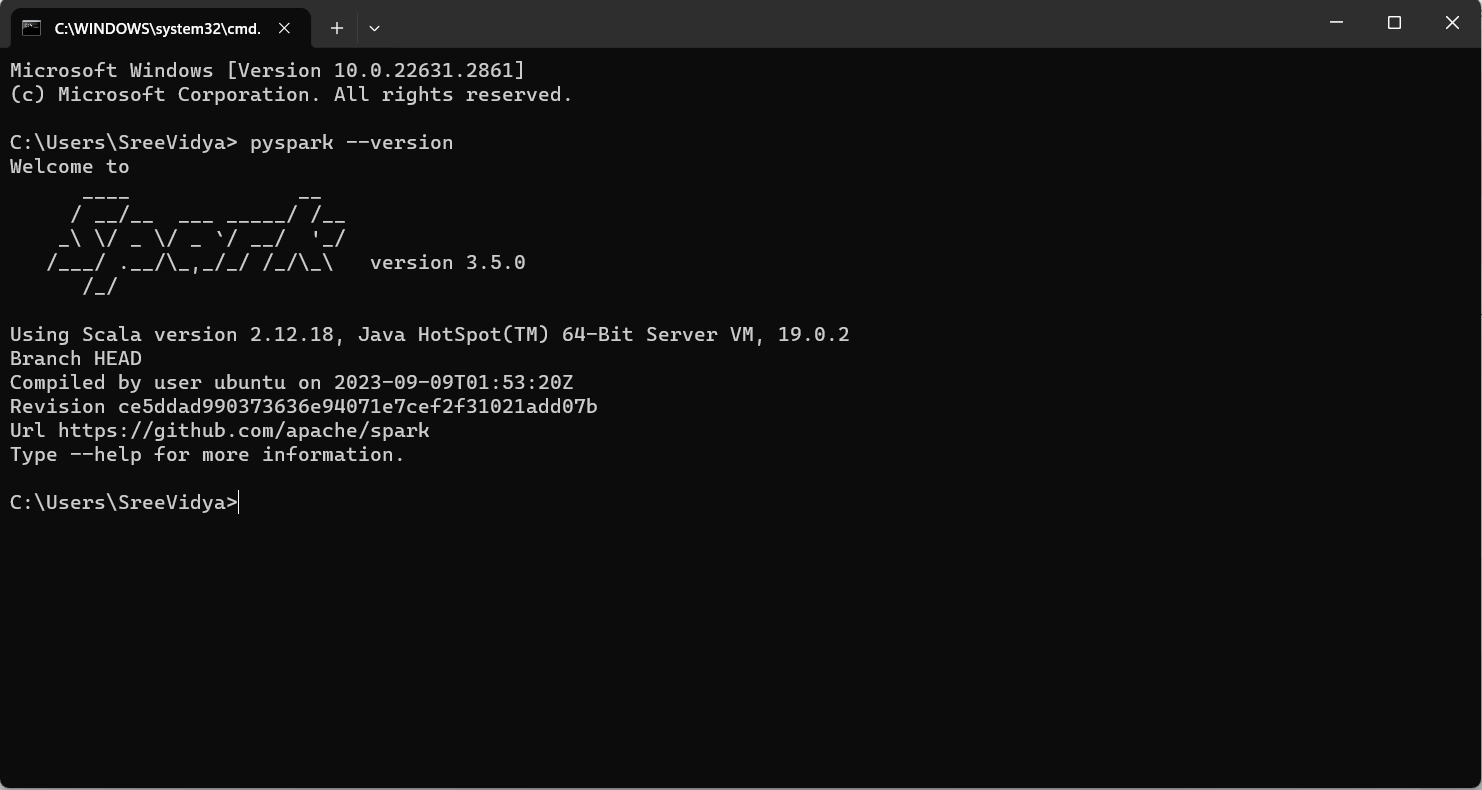
* **Group By and Aggregate:**

scala

*val result = df.groupBy("columnName"). agg(avg("otherColumn"))*

These commands provide a basic overview of Spark functionality. Depending on your specific use case and the complexity of your Spark application

* How to open Spark in command prompt
* Type pyspark or pyspark --version



* To run commands for Apache spark we need to install jupyter notebook, for that we need to type 🡪Python -m pip install jupyter notebook
* To run commands, we have another command to type in command prompt named 🡪 Python -m pip install findspark
* To open Jupyter notebook type jupyter notebook or jupyter lab in command prompt, there we can execute codes.