SPARK BASICS SIMPLY EXPLAINED

PART 5: UPDATE TO SPARK 3.5.1 NOW!

Step 1: Update Java

- Install Java v8 latest release: Latest Releases | Adoptium
- Update System Variable for JAVA_HOME = C:\Program Files\Eclipse Adoptium\jdk-8.0.402.6-hotspot

Step 2: Update Spark Installation

- Download Spark 3.5.1: Downloads | Apache Spark and unzip having the path D:\Spark\spark-3.5.1-bin-hadoop3 and then the e.g. „bin“ folder inside
- Update System Variable for SPARK_HOME = D:\Spark\spark-3.5.1-bin-hadoop3

Step 3: Test in CMD

Open CMD

Run *pyspark*

Run *sdf = spark.range(10)*

Run *sdf.count()*

Result should be 10

Step 4: Update conda env

- Open CMD
- Run conda activate pyspark
- Run *pip install pyspark==3.5.1*

Step 5 VSCode Testing

- Clone my repo or download this script: SparkDeltaDatabricksInternals/Lection1 - Architecture/S01E05_SparkSet-Up_3.5.1.ipynb at main · datanikkthegreek/SparkDeltaDatabricksInternals (github.com)
- Open VSCode
- Choose your conda env
- Execute in VS Code