ASSESSMENT 15

Apache Spark is a new and open-source framework used in the big data industry for real-time processing and batch processing.

The steps to install pyspark:

Install java jdk

Go to Command Prompt and type java -version

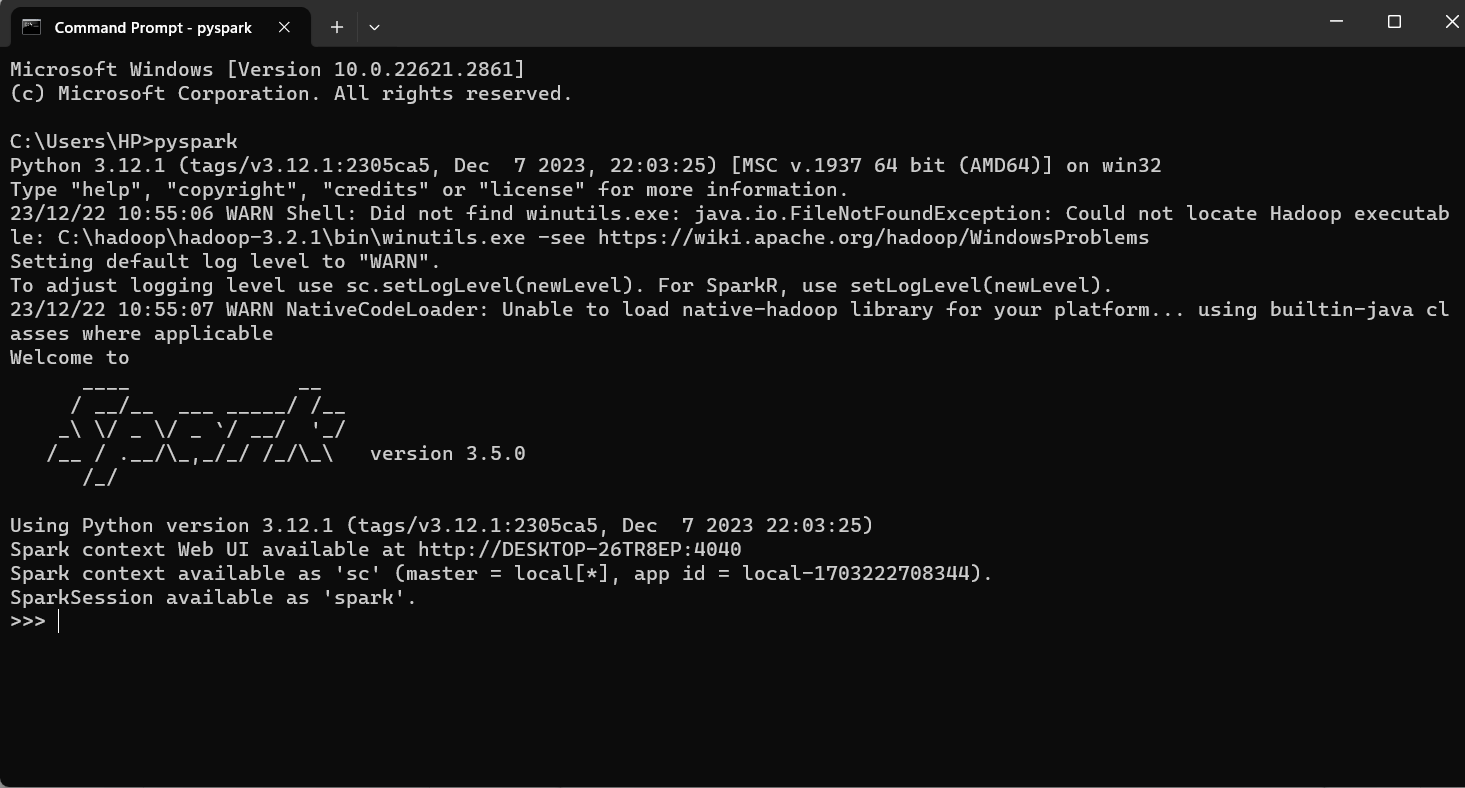
In the Environment variables, ad the java path

Then we have to download the spark file

Download the Hadoop file and extract that and save it in the c folder

Create a path for Hadoop and spark in the Environment variables

Open command prompt and type pyspark



Then in the browser type http://DESKTOP-26TR8EP:4040

