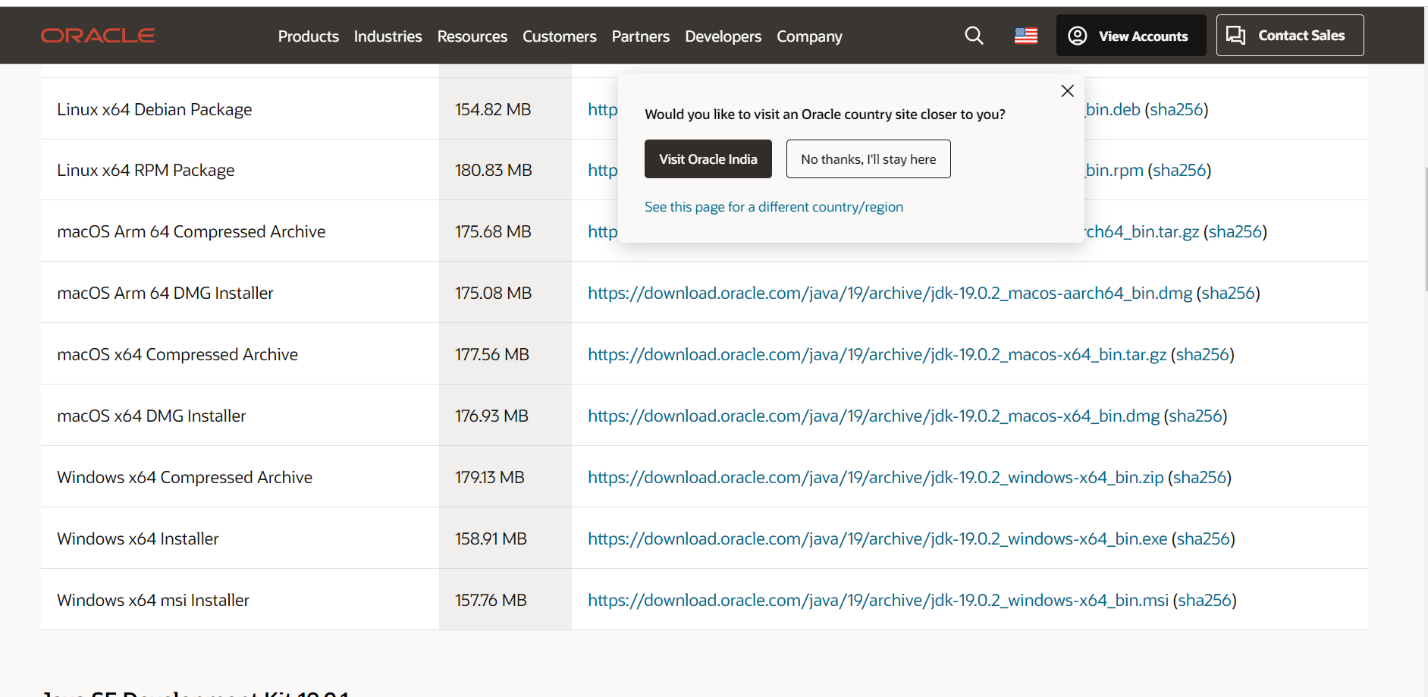
Installation of pyspark:

Pyspark = Python + Apache Spark

Apache Spark is a new and open-source framework used in the big data industry for real-time processing and batch processing. It supports different languages, like Python, Scala, Java, and R.

Apache Spark is initially written in a Java Virtual Machine(JVM) language called Scala, whereas Pyspark is like a Python API which contains a library called Py4J. This allows dynamic interaction with JVM objects.

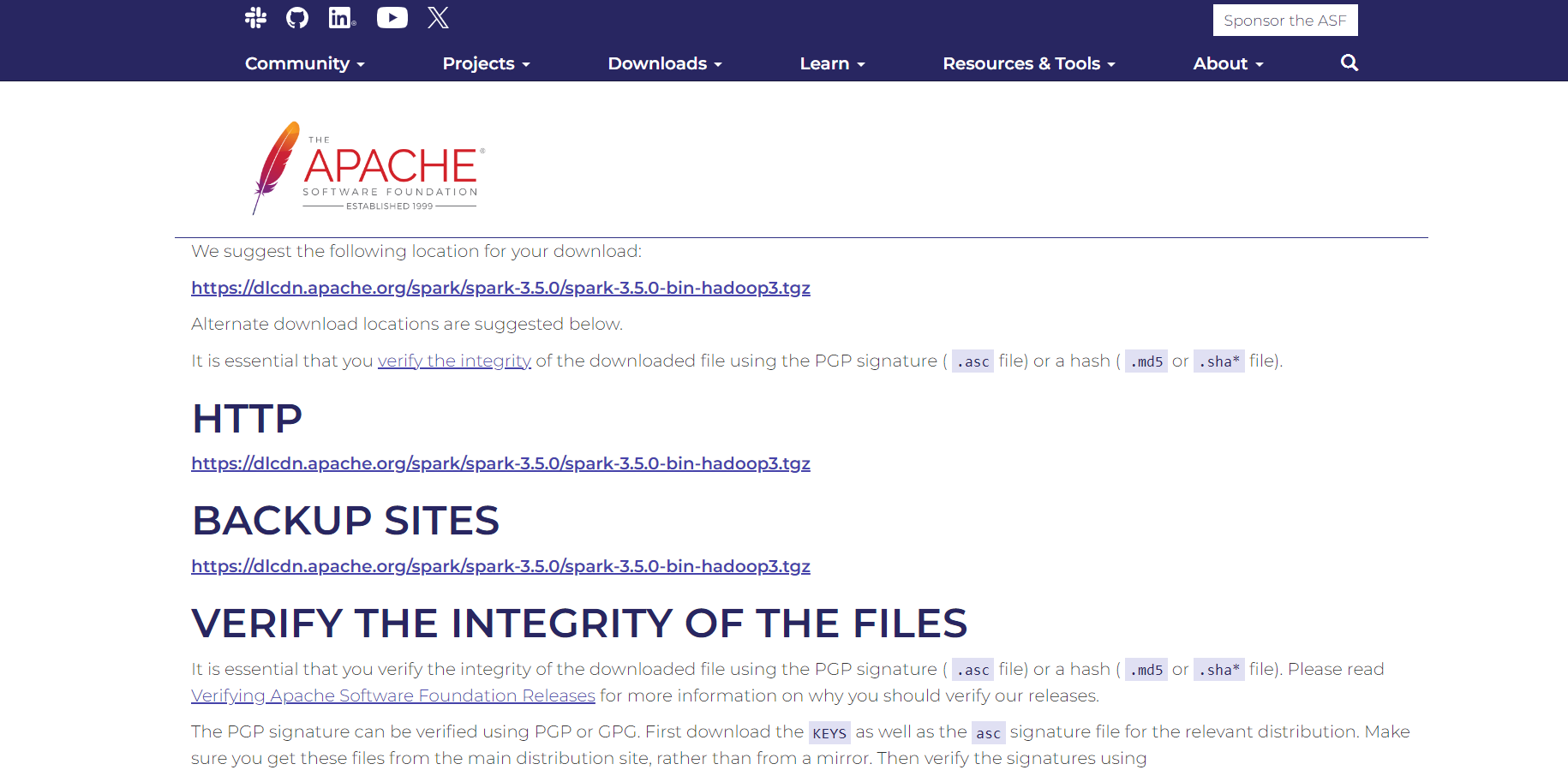
Download javajdk



Add the java path

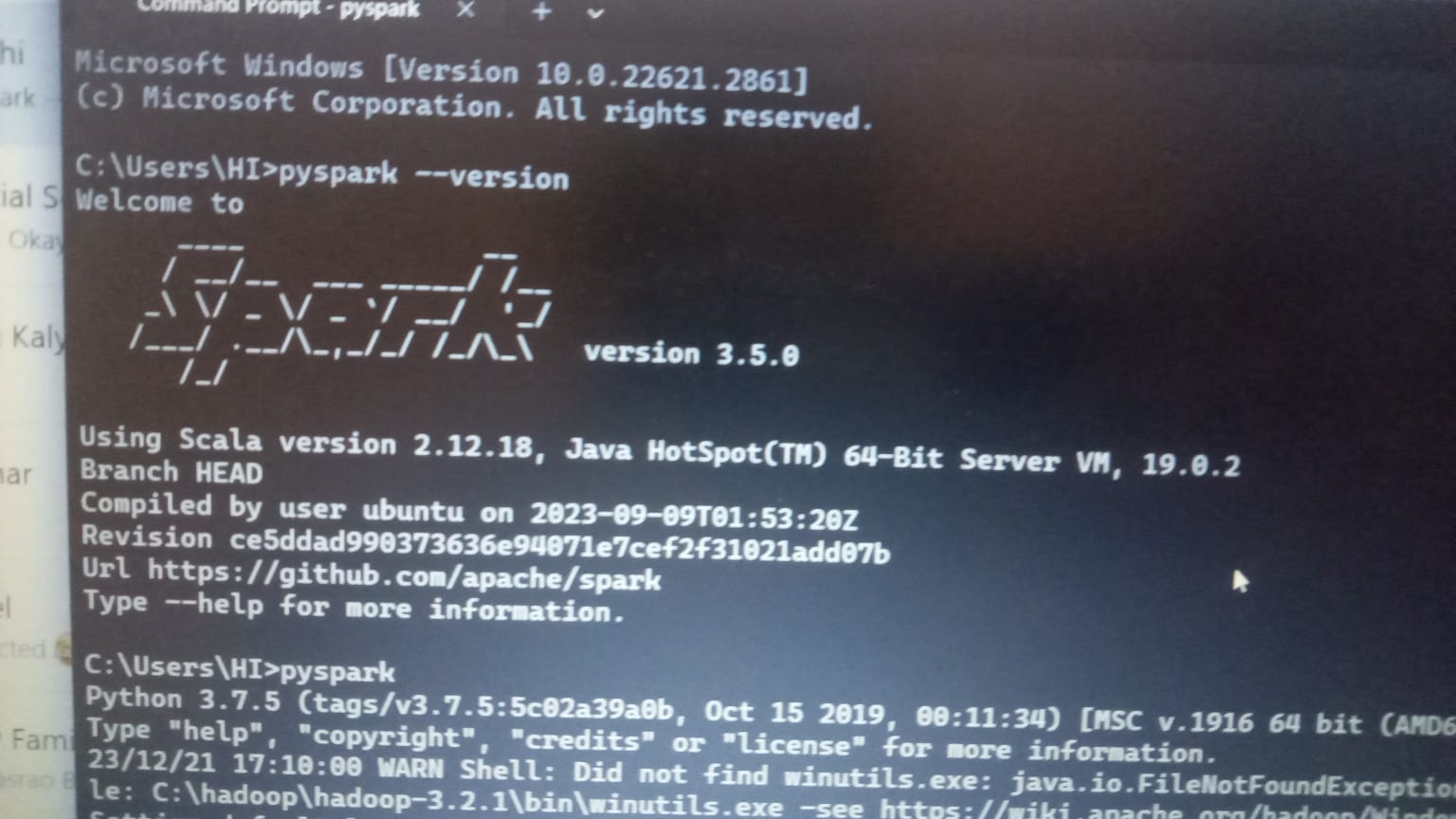
Edit the environment variables

Install Hadoop and spark



Add path and edit the environment varaibales

Open command prompt and type pyspark ot will be installed.

To connect to pyspark ui: http://localhost:4040/jobs

We will get the web ui of pyspark

A screenshot of a computer

Description automatically generated