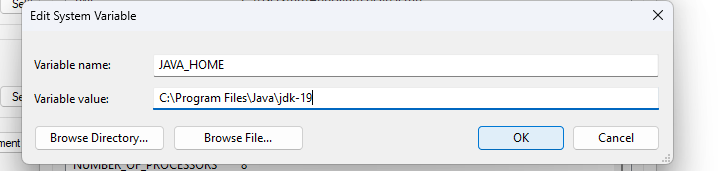
**Alekhya Krishna Balivada Pyspark Assessment 1 (21-12-2023)**

**Pyspark: Python + Apache Spark**

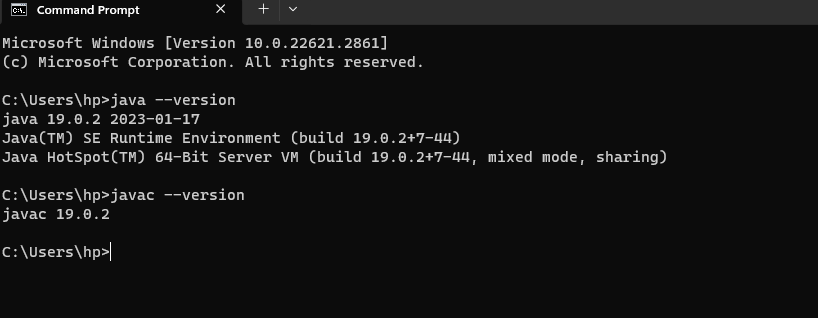
**Apache Spark:** Apache Spark is an open-source, distributed processing system used for big data workloads. It supports different languages like Java, Python, Scala and R.

**Java JDK installation:**

1. Go to google chrome and type Java JDK download.
2. ****Download the link as shown below
3. After downloading install JDK.
4. After installation generate the path for JDK in environment variables.

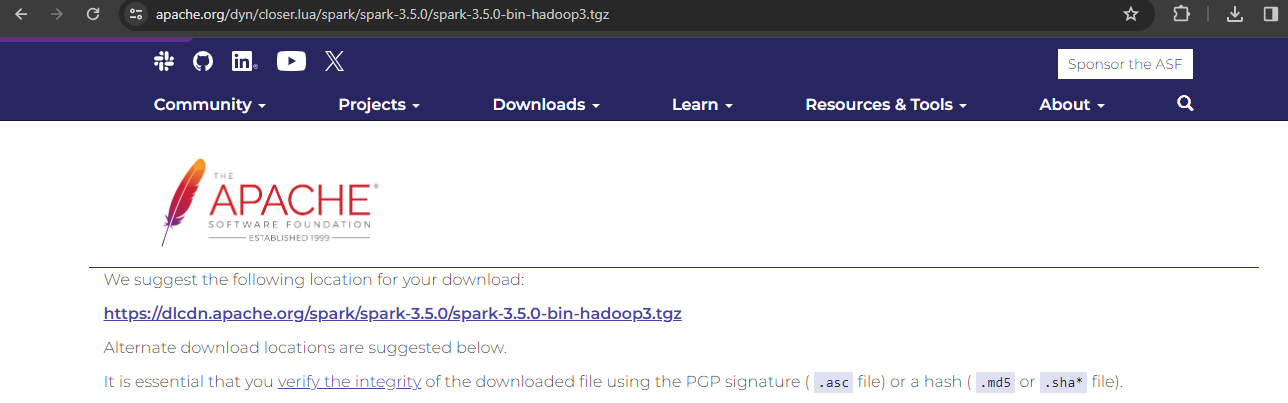
****

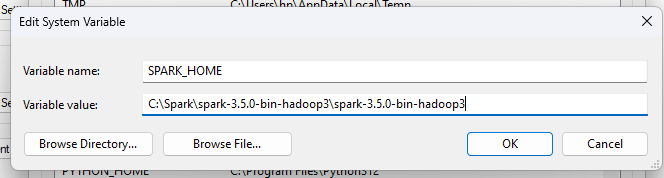
****

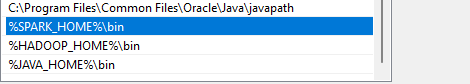
1. ****Now type ‘java –version’ in command prompt.

**Apache Spark Installation:**

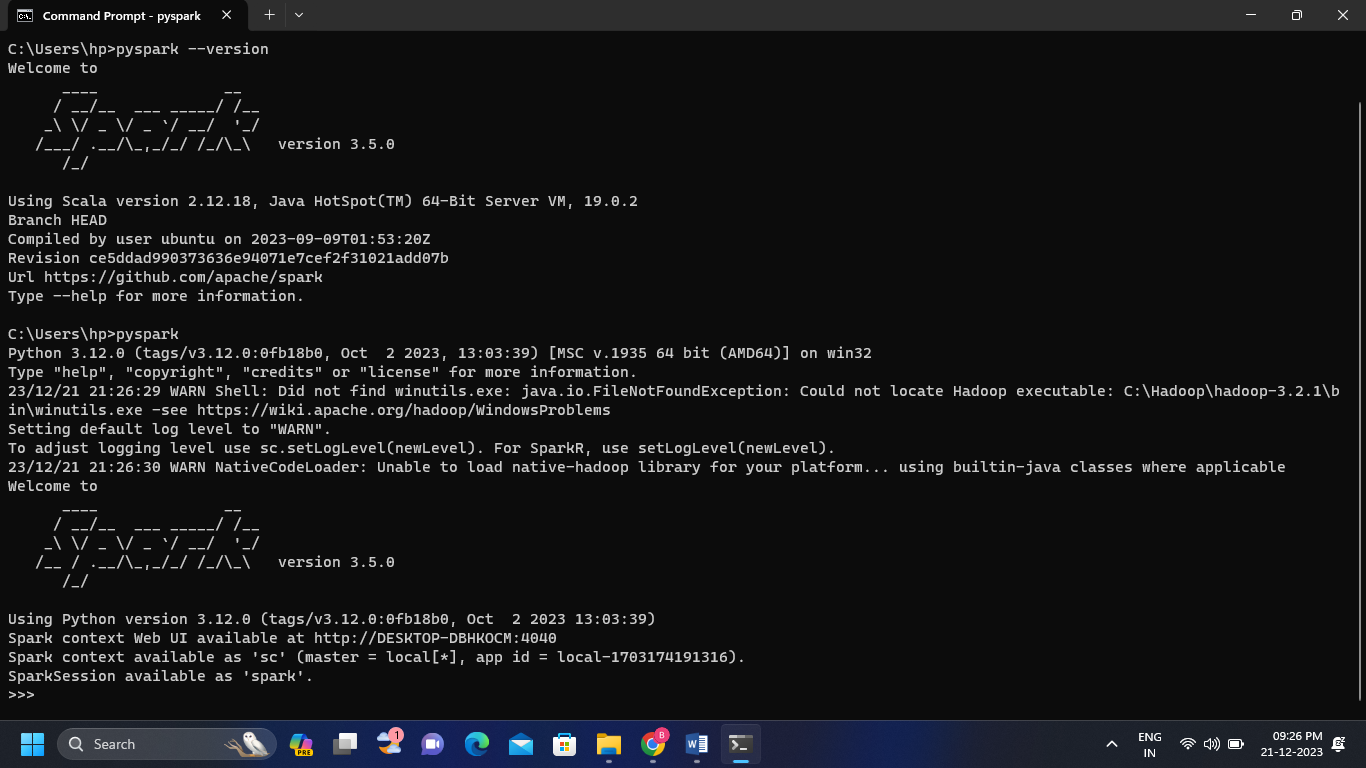
1. Go to google chrome and type ‘Download Apache Spark’ and download the link which is shown as follows –

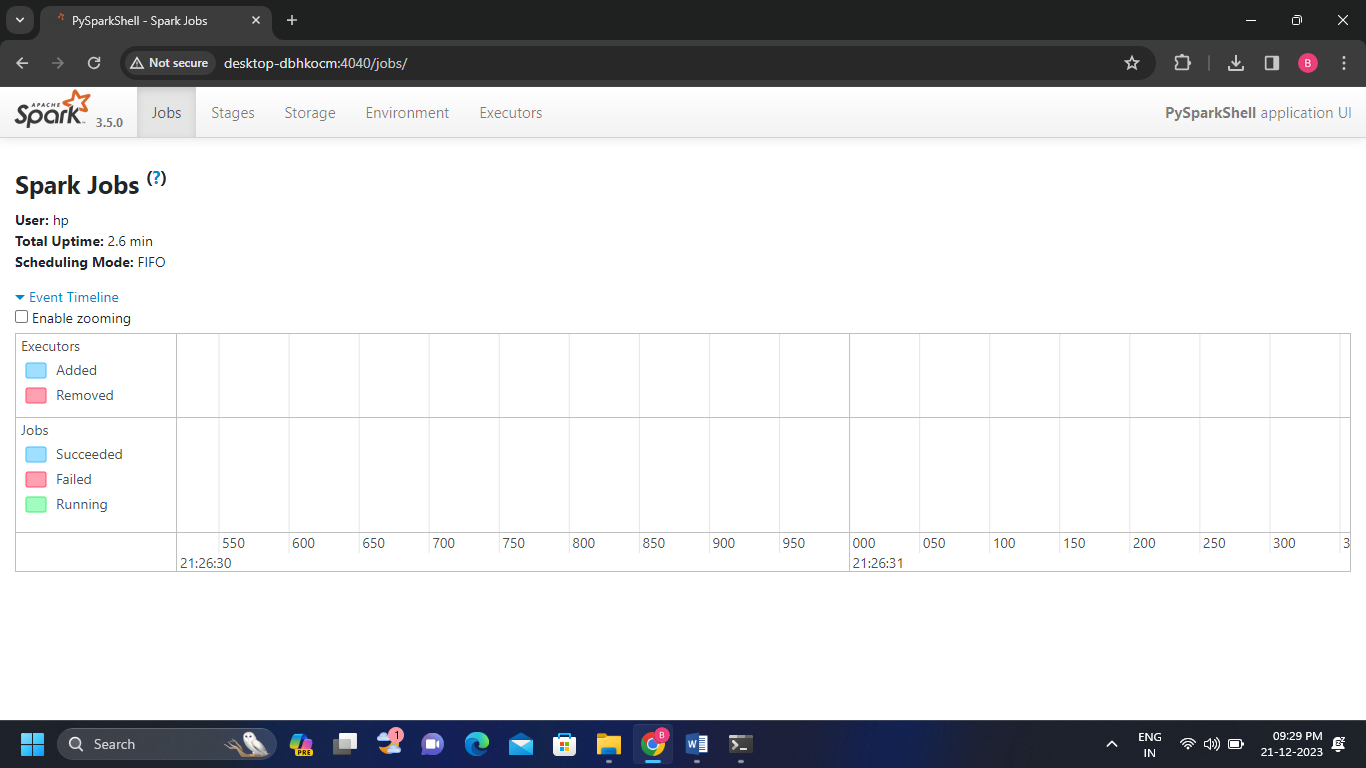


1. After downloading complete the installation process.
2. ****After installing generate the path for spark in environment system variables as -

****

1. After giving path type ‘pyspark’ in command prompt then it will show as follows –

****

1. ****Visit the Spark context web UI and it will be shown as follows-