PYSPARK

ASSIGNMENT 21-12-2023

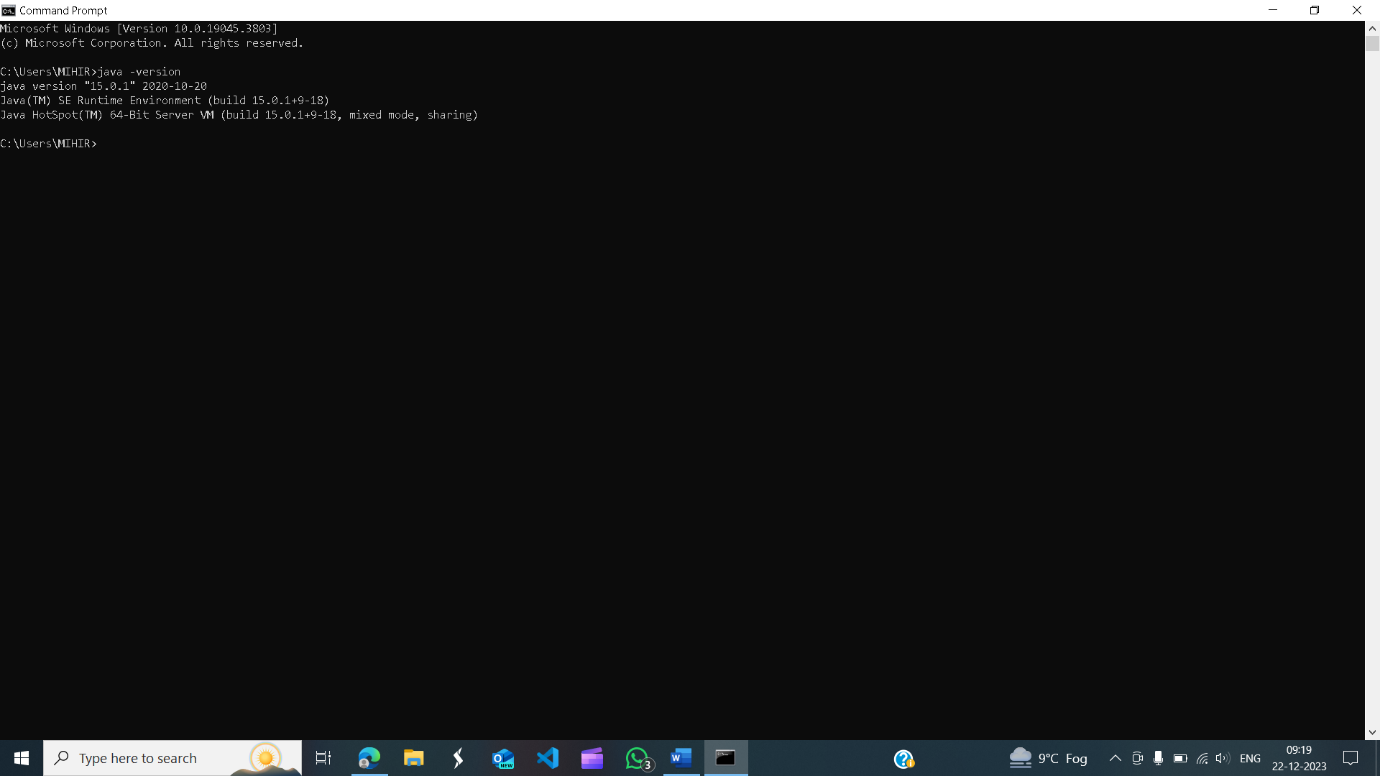
PYSPARK

**Pyspark = Python + Apache Spark**

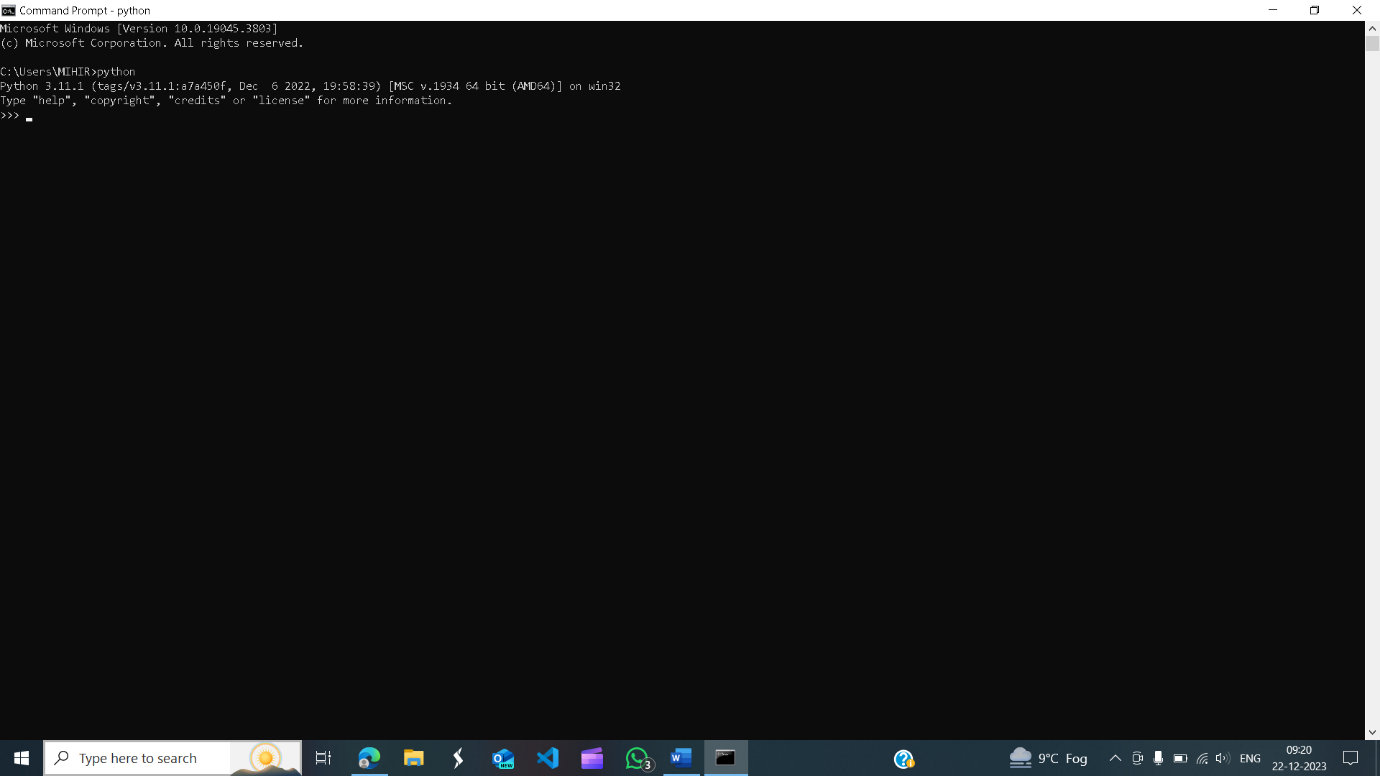
Apache Spark is a new and open-source framework used in the big data industry for real-time processing and batch processing.

Installation of Pyspark

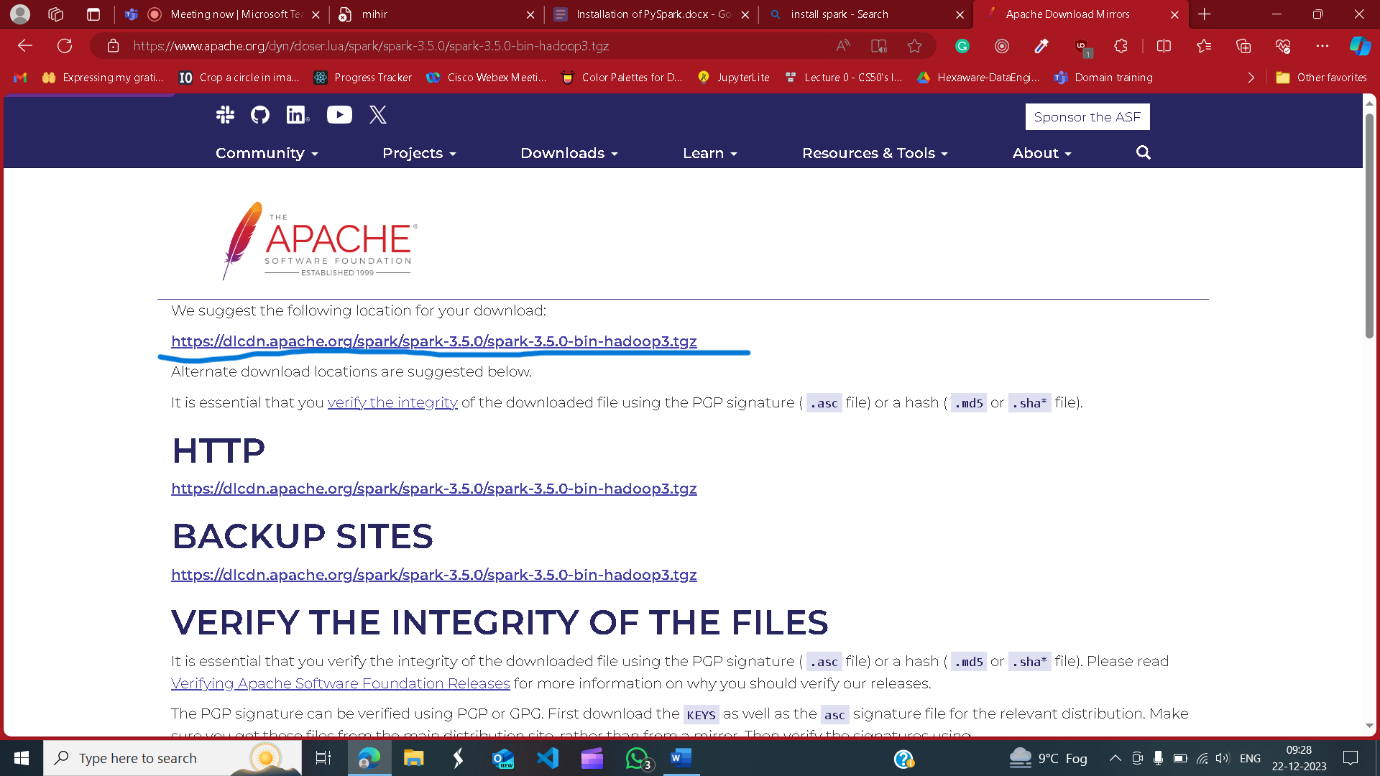
Install jdk



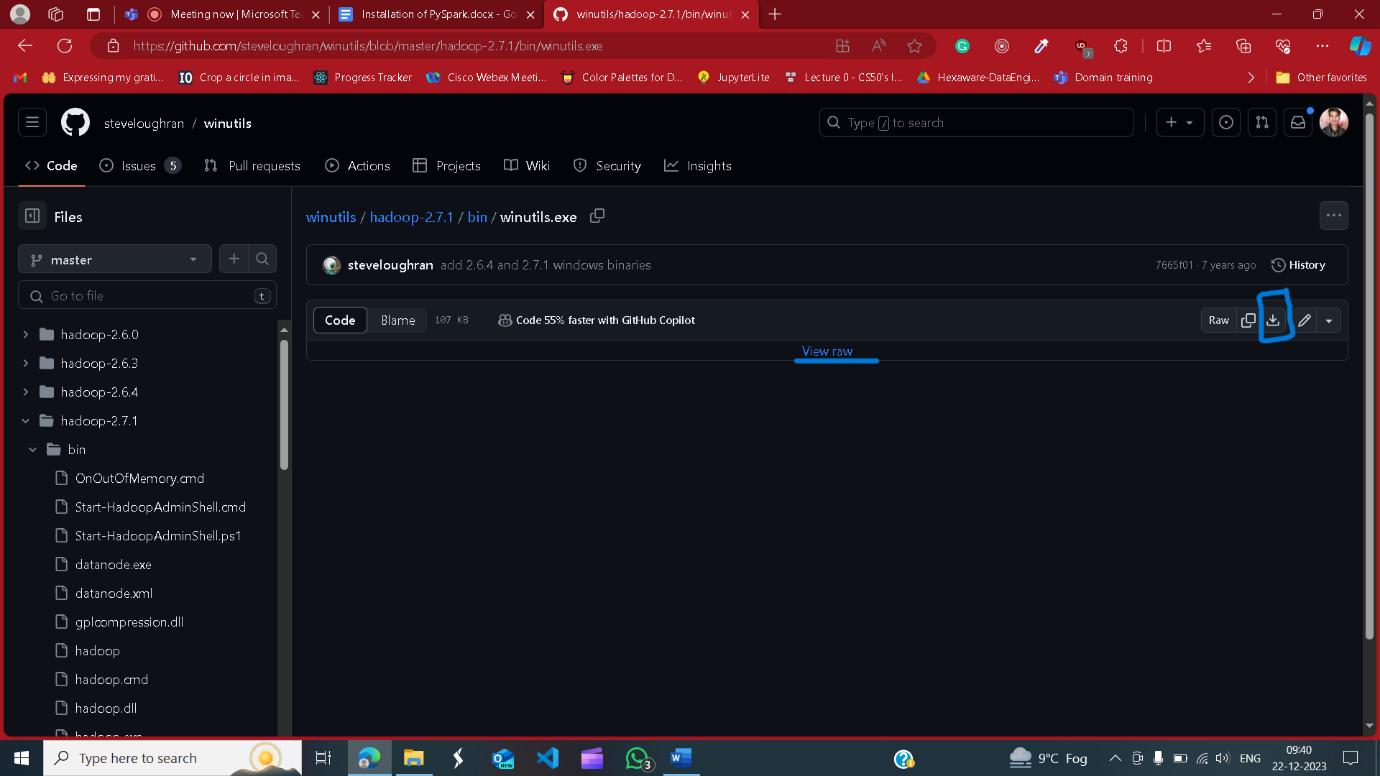
Install python



Install spark



Download hadoop



Along with installing we also have to set up environment variable with this softwares

Final result