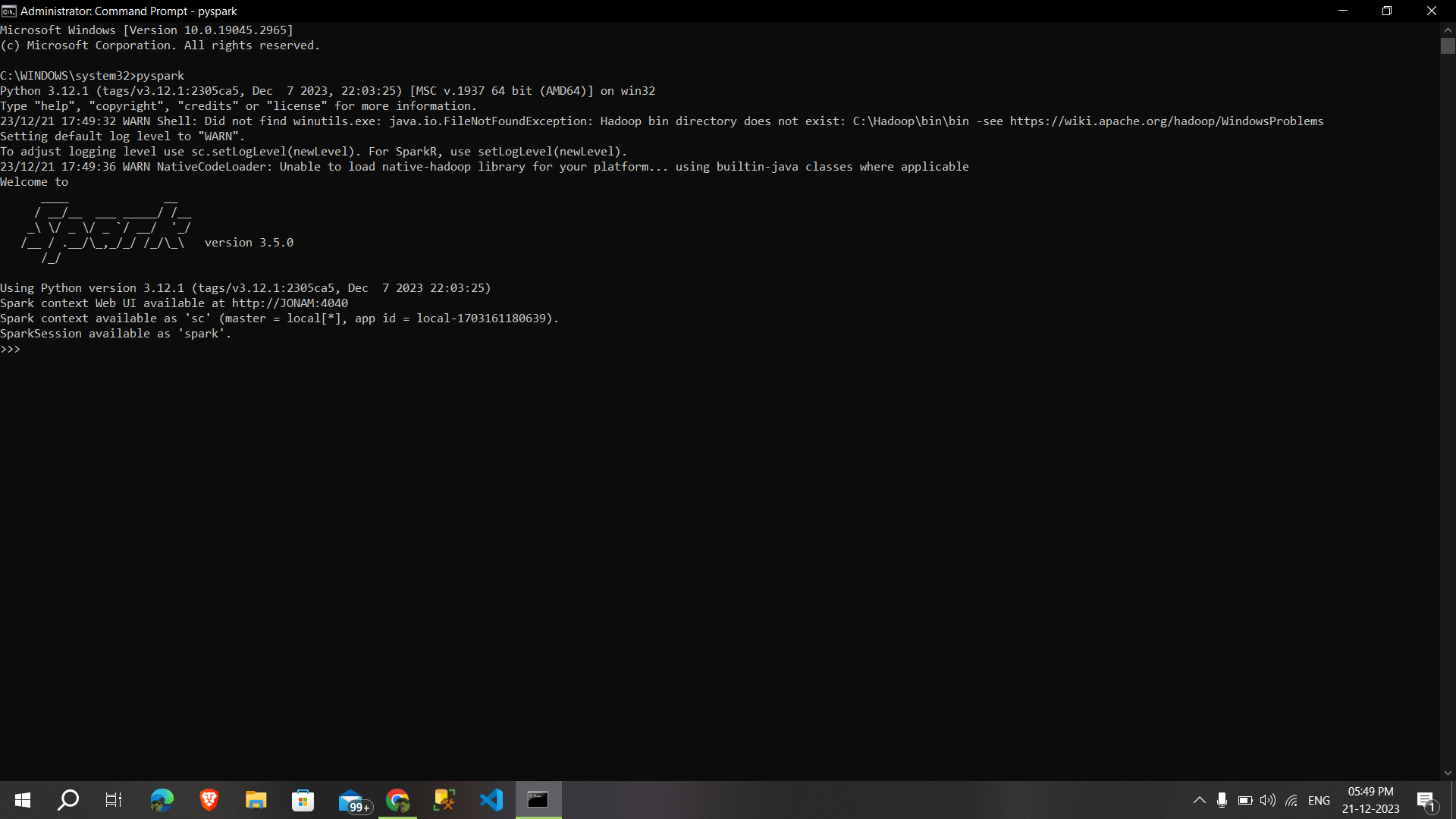
**Installation of PySpark**

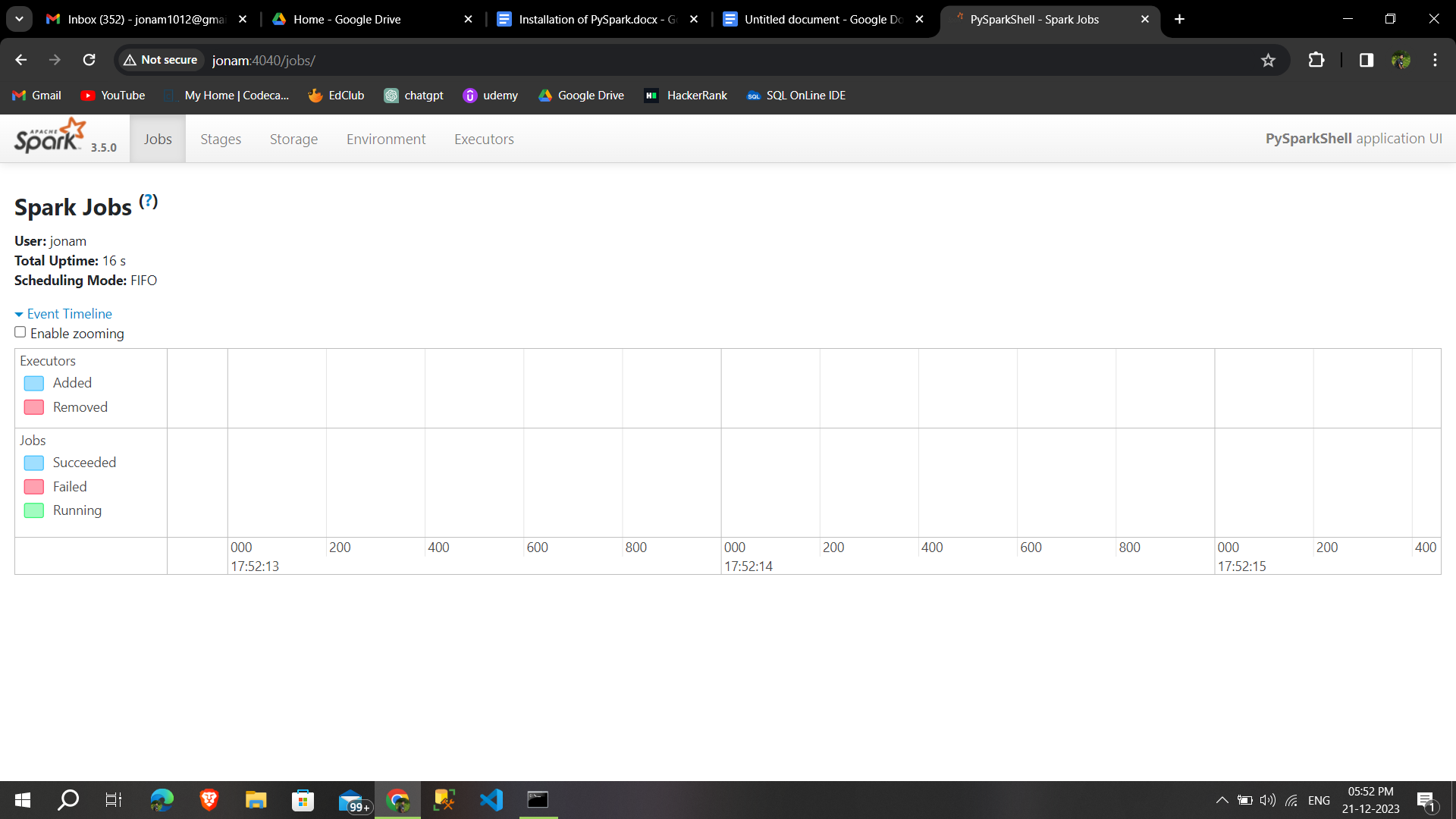
**For installing the spark we need to do follow some steps and those steps are given below,**

1. Install latest version of jdk and set the path in the **environment variables** in your local machine and variable name as **JAVA\_HOME**
2. Install latest version of python and set the path in the **environment variables** in your local machine and variable name as **PYTHON\_HOME**
3. Install the spark and hadoop from the official website and download **same version**
4. After downloading the spark and hadoop version create a folder in your local machine and **extract those files in that folder**
5. Then set a **environment variables** in your local machine and variable name as **SPARK\_HOME,HADOOP\_HOME**

**After successful installation we can see in cmd as below after typing pyspark,**

****

**The web ui of the spark is below,**

****

**Apache Spark:**

* It is an open-source, fast and general-purpose cluster-computing framework for big data processing and analytics.
* it provides APIs for programming in other languages like Python, R, and SQL.
* It has the ability to process data in-memory, which makes it significantly faster than its predecessor, Hadoop MapReduce, for certain use cases.
* Spark supports various data processing tasks, including batch processing, interactive queries, streaming analytics
* PySpark is the Python library for Apache Spark.