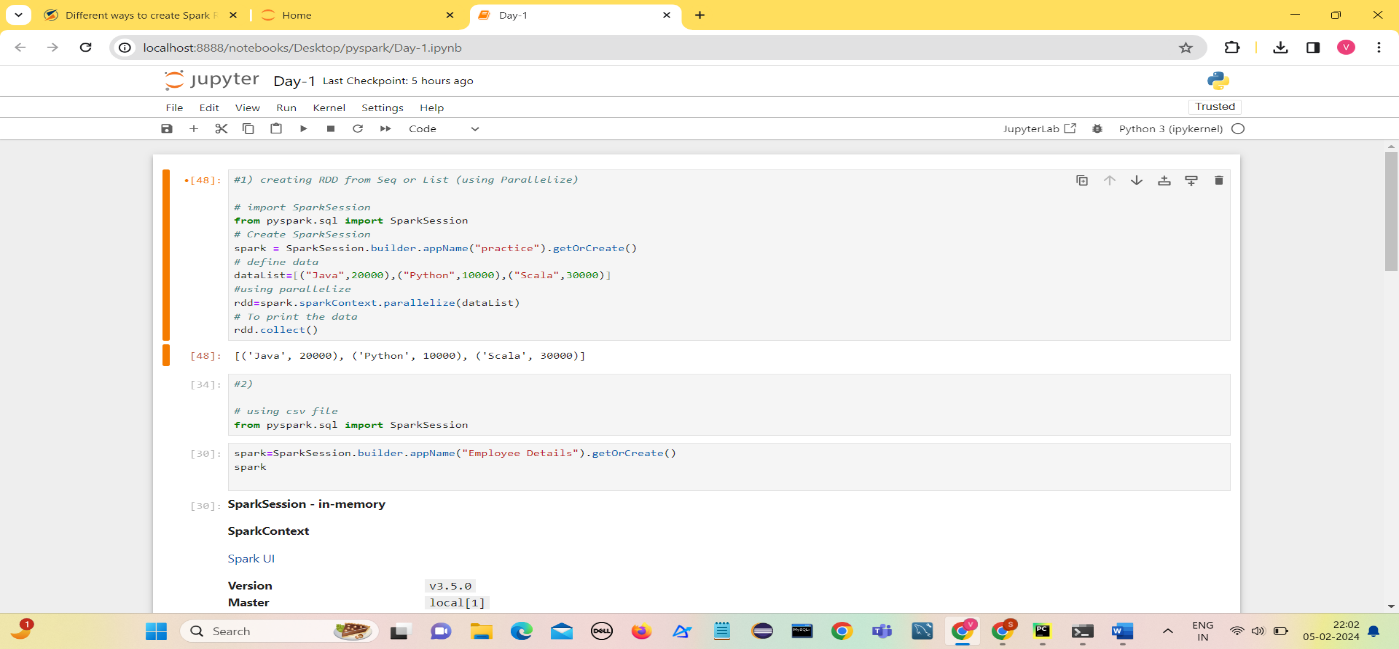
**Pyspark Day-2**

**Creating RDD:**

Spark RDD can be created in several ways, for example, It can be created by using sparkContext.parallelize(), from text file, from another RDD, DataFrame, and Dataset.

1. **Spark Create RDD from Seq or List (using Parallelize)**

RDD’s are generally created by parallelized collection i.e. by taking an existing collection from driver program (scala, python etc) and passing it to SparkContext’s  parallelize() method.

## Create an RDD from a text file

## Spark Read CSV file into DataFrame

## Apache Spark provides a DataFrame API that allows an easy and efficient way to read a CSV file into DataFrame. DataFrames are distributed collections of data organized into named columns. spark.read.csv("path") is used from the API to read a CSV file.

## 

## 

## 