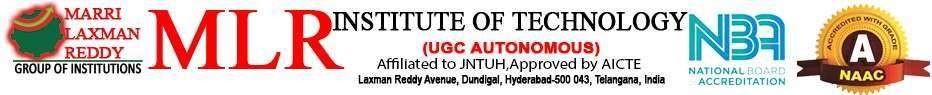
****

**BIG DATA ANALYTICS LAB**

**WEEK-10**

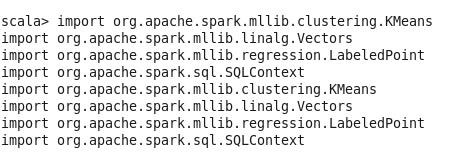
**NAME :** T.Sowmya **BRANCH & SEC :** CSD-A

**ROLLNO :** 22R21A6757 **DATE : 2-12-2024**

**PROBLEM STATEMENT:**

**Implement the k-means algorithm for clustering**

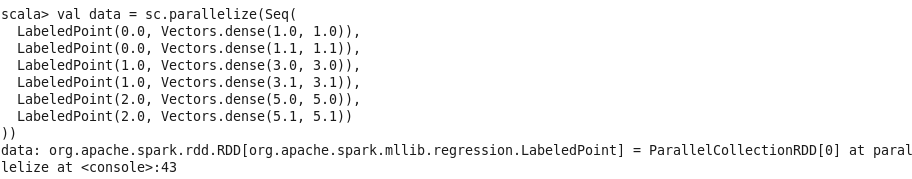
1. **Importing Libraries**

****

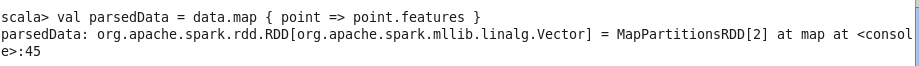
1. **Initialize SQLContext (available in Spark 1.6.0)**

****

1. **Sample dataset: An RDD of LabeledPoint (label, features)**

****

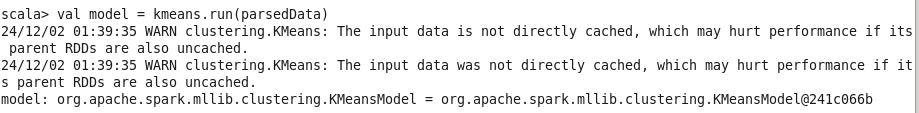
1. **Convert the data to an RDD of Vectors (KMeans expects an RDD of Vectors)**

****

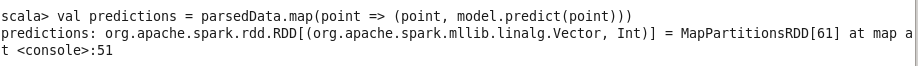
1. **Initialize KMeans clustering**

****

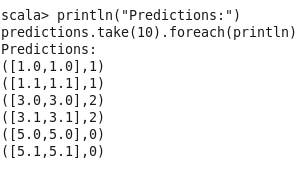
1. **Train the model**

****

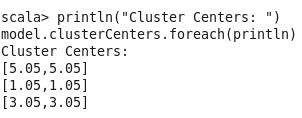
1. **Make predictions (assign each point to a cluster)**

****

1. **Print the predictions**

****

1. **Print Cluster Centers**

****