1. **Transfrom**

Map,flatmap,filter,distinct,sample,union,intersection,subtract,cartesian

1. **Action**

Collect,count,countByValue,take,top,reduce….

1. **Creating an RDD**
2. Val rdd =sc.parallelize(List(1,2,3,4))
3. **val** linesRDD = sc.textFile("input/udemy/spark-scala/ml-100k/u.data")
4. **val** linesRDD=Source.fromFile("path to file”)getLines()//**import** scala.io.Source

**4.**  **var** movieMap: *Map*[Int, *String*] = Map() // creating a Map

**val** lines = Source.fromFile("input/udemy/spark-scala/ml-100k/u.item").getLines() //another way of reading from file without SparkContext/SparkSession

**for** (line <- lines) { **val** fields = line.split('|')

movieMap += (fields(0).toInt -> fields(1)) // initializing Map(key/value) }

**val** result=lines.map(tuple=>(movieMap.value((tuple.\_2).toInt),tuple.\_1))// get map value with key

5.**reduceByKey(),groupBykey(),sortByKey() ,keys(),values()** can only work on a keyValue/pair RDD ,

reduceByKey() function works only on value part

groupBykey() returns a list with the key.

6. Dstream is a stream of RDDs . If the frequency is one second , One RDD gets created every second.

Dstream.foreach /Dstream.transform gets an RDD from a a DStream