**Group A**

**Lab 4a**

**Spark Streaming**

**Participants:**

**1. Aadarsha chapagain**

**2.Jyoti shukla**

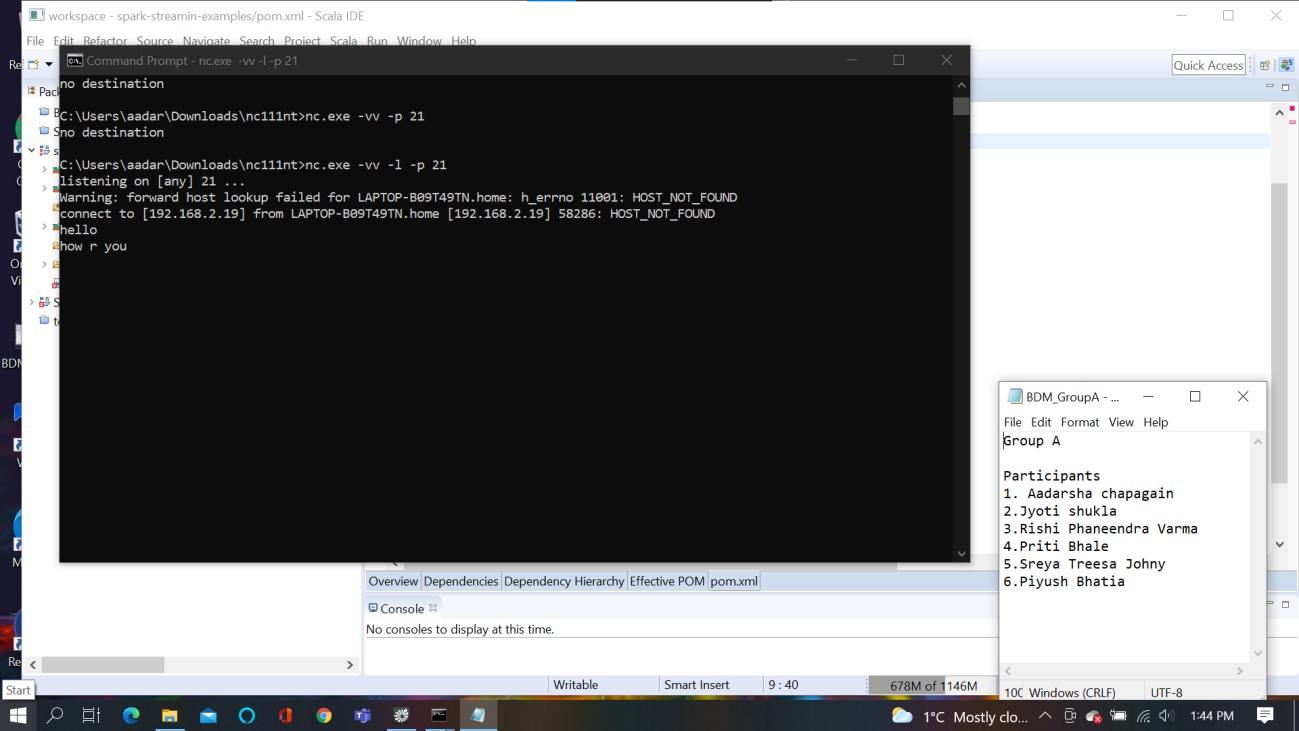
**3.Rishi Phaneendra Varma**

**4.Priti Bhale**

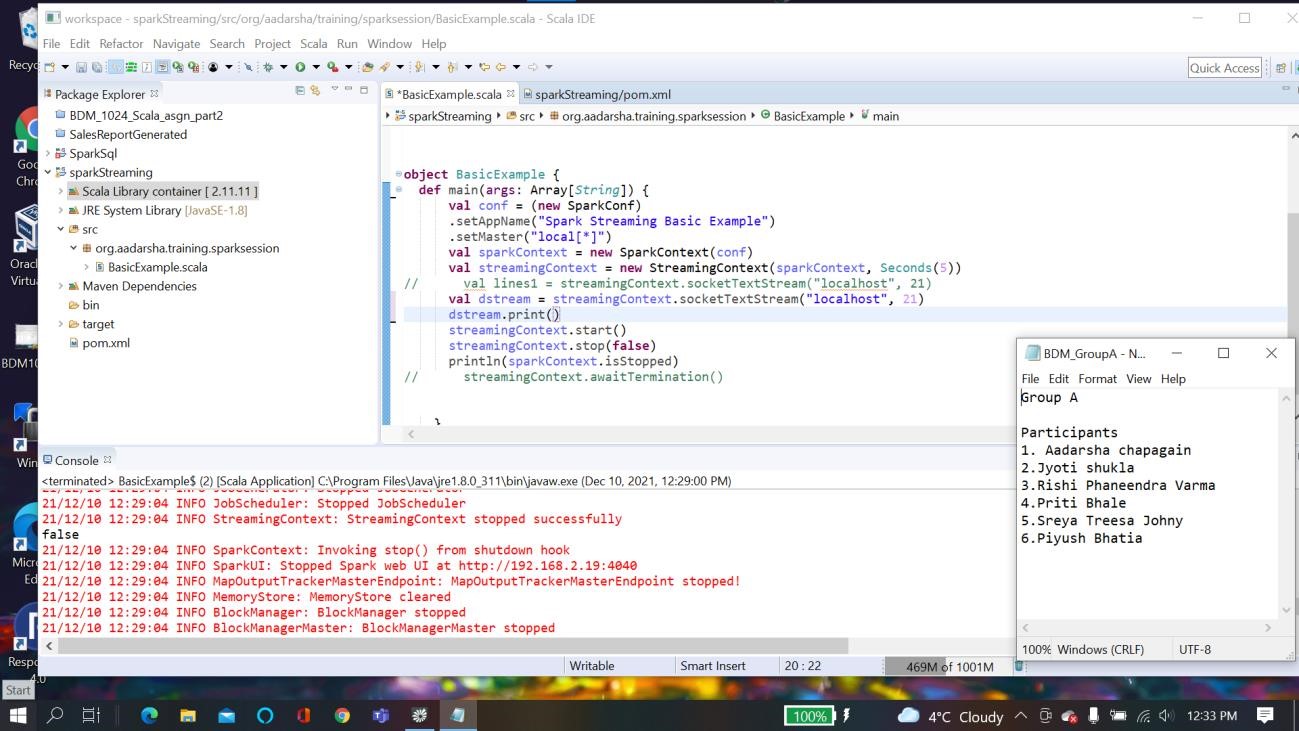
**5.Sreya Treesa Johny**

**6.Piyush Bhatia**

**Slide 9:**



**Slide 20:**



Slide 42:

**package** org.aadarsha.training.sparksession

**import** org.apache.spark.SparkConf **import** org.apache.spark.streaming.Seconds **import** org.apache.spark.streaming.StreamingContext **import** org.apache.spark.SparkContext //import org.apache.spark.sql.SparkSession

**object** BasicExample { **def** main(args: Array[*String*]) { **val** conf = (**new** SparkConf)

.setAppName("Spark Streaming Basic Example")

.setMaster("local[\*]")

**val** sparkContext = **new** SparkContext(conf)

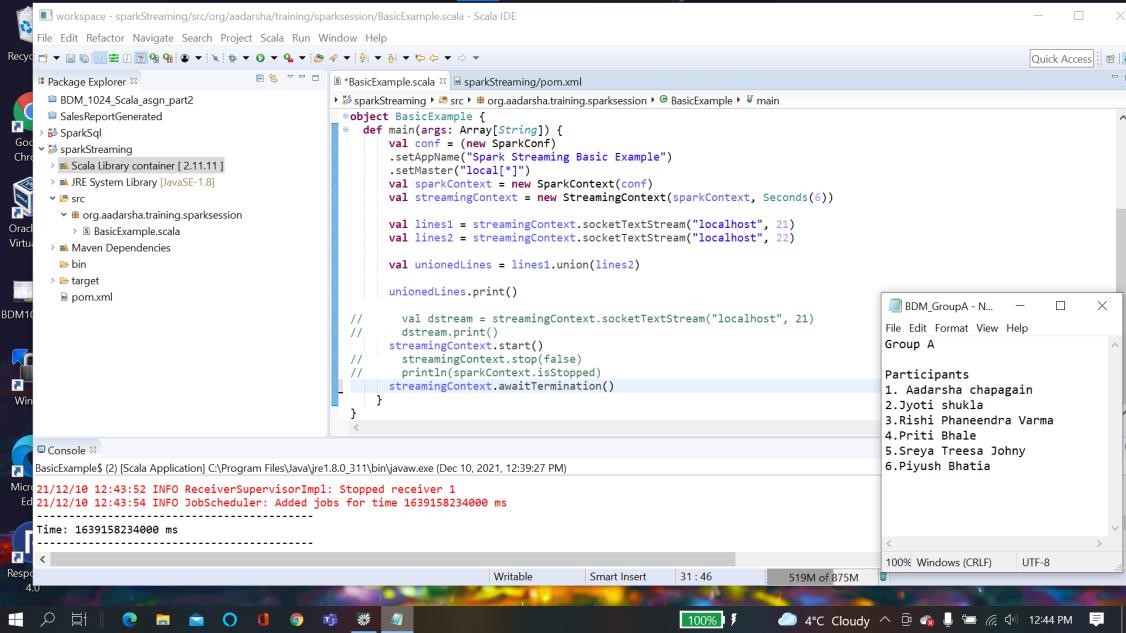
**val** streamingContext = **new** StreamingContext(sparkContext, Seconds(6))

**val** lines1 = streamingContext.socketTextStream("localhost", 21) **val** lines2 = streamingContext.socketTextStream("localhost", 22)

**val** unionedLines = lines1.union(lines2) unionedLines.print() streamingContext.start()

streamingContext.awaitTermination() }

}



**Slide 52:**

**package** org.aadarsha.training.sparksession

**import** org.apache.spark.SparkConf **import** org.apache.spark.streaming.Seconds **import** org.apache.spark.streaming.StreamingContext **import** org.apache.spark.SparkContext

**object** BasicExample { **def** main(args: Array[*String*]) { **val** conf = (**new** SparkConf)

.setAppName("Spark Streaming Basic Example")

.setMaster("local[\*]")

**val** sparkContext = **new** SparkContext(conf)

**val** streamingContext = **new** StreamingContext(sparkContext, Seconds(6)) **val** constantsRdd = createConstants(sparkContext)

**val** lines = streamingContext.socketTextStream("localhost", 21)

**val** resultDstream = lines.transform(rdd => rdd.union(constantsRdd)) resultDstream.print()

streamingContext.start()

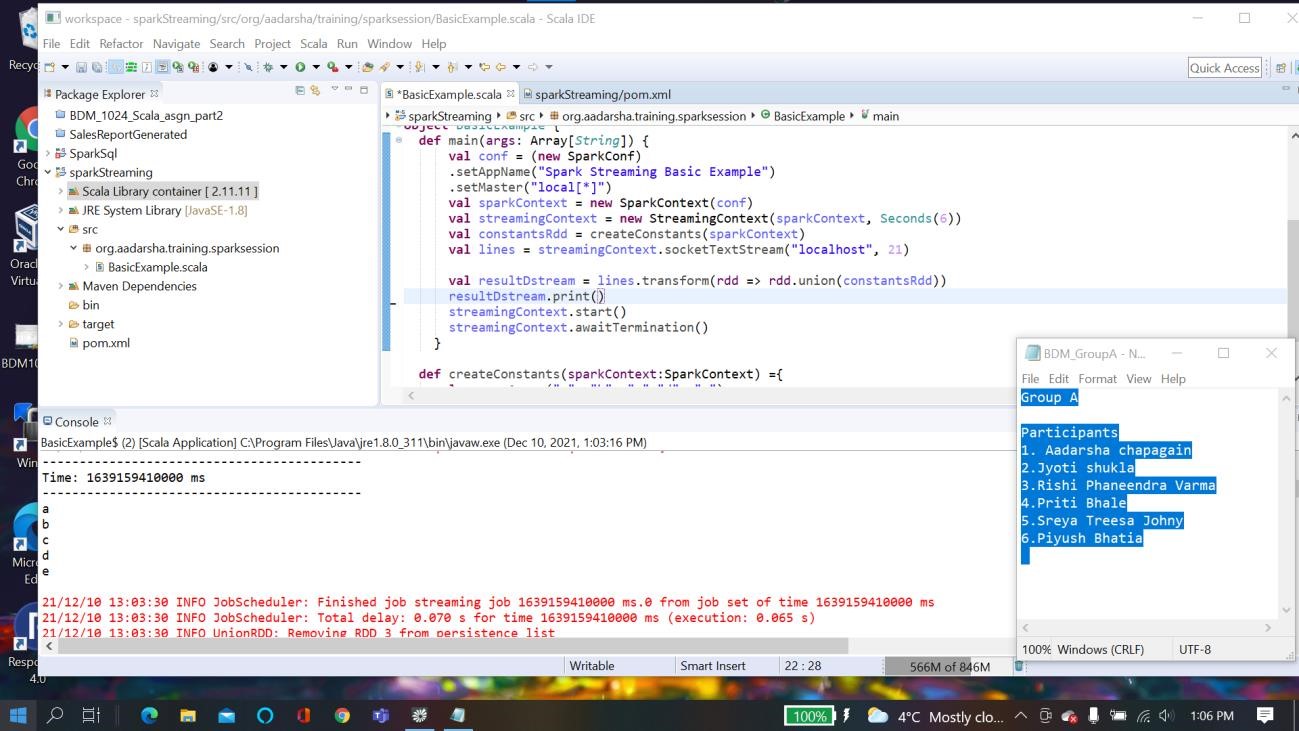
streamingContext.awaitTermination()

}

**def** createConstants(sparkContext:SparkContext) ={ **val** arr = Array("a", "b", "c","d", "e") sparkContext.parallelize(arr)

}

}



**Slide 57:**

**package** org.aadarsha.training.sparksession

**import** org.apache.spark.SparkConf

**import** org.apache.spark.streaming.StreamingContext **import** org.apache.spark.streaming.Seconds

**object** CountByWindowExample { **def** main(args: Array[*String*]) {

**val** conf = (**new**

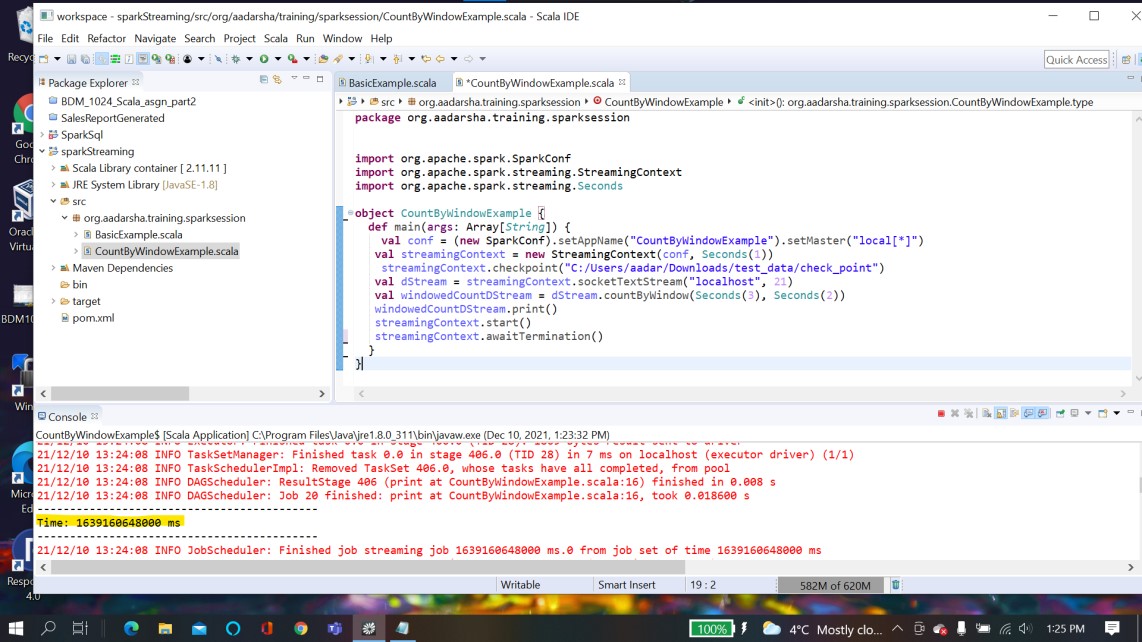
SparkConf).setAppName("CountByWindowExample").setMaster("local[\*]") **val** streamingContext = **new** StreamingContext(conf, Seconds(1))

streamingContext.checkpoint("C:/Users/aadar/Downloads/test\_data/check\_point") **val** dStream = streamingContext.socketTextStream("localhost", 21) **val** windowedCountDStream = dStream.countByWindow(Seconds(3), Seconds(2)) windowedCountDStream.print() streamingContext.start()

streamingContext.awaitTermination()

}

}



**Slide 75:**

**package** org.aadarsha.training.sparksession **import** org.apache.spark.SparkConf

**import** org.apache.spark.streaming.StreamingContext **import** org.apache.spark.streaming.Seconds

**import** org.apache.spark.streaming.dstream.DStream.toPairDStreamFunctions **import** java.time.Instant

**object** ReduceByKeyAndWindowExample { **def** main(args: Array[*String*]) {

**val** conf = (**new**

SparkConf).setAppName("CountByWindowExample").setMaster("local[\*]") **val** streamingContext = **new** StreamingContext(conf, Seconds(2))

streamingContext.checkpoint("C:/Users/aadar/Downloads/test\_data/check\_point") **val** dStream = streamingContext.socketTextStream("localhost", 21) **val** windowedCountDStream = dStream.window(Seconds(10), Seconds(8)) windowedCountDStream.foreachRDD(rdd=>{ println("Hello")

rdd.saveAsTextFile("C:/Users/aadar/Downloads/test\_data/output")

})

streamingContext.start()

streamingContext.awaitTermination()

}

}

