

Package Explorer

- SalesReportGenerated
- spark-streaming-examples
 - Scala Library container [2.11.11]
 - JRE System Library [JavaSE-1.8]
 - src
 - org.bibek.training.sparkstreaming
 - BasicExample.scala
 - Maven Dependencies
 - bin
 - target
 - pom.xml
- SparkSqlExamples
 - Scala Library container [2.11.11]
 - JRE System Library [JavaSE-1.8]
 - src
 - org.bibek.training.sparksq
 - HiveExample1.scala
 - RddToDFExample.scala
 - WithColumnExample.scala
 - org.bibek.training.sparksq
 - bean
 - Maven Dependencies
 - bin
 - spark-warehouse
 - target
 - pom.xml

```
BasicExample.scala  spark-streaming-examples/pom.xml  RddToDFExample.scala
package org.bibek.training.sparkstreaming

import org.apache.spark.SparkConf
import org.apache.spark.streaming.{Seconds, StreamingContext}

object BasicExample {

  def main(args: Array[String]) {
    val conf = (new SparkConf).setAppName("Spark Streaming Basic Example")
    val streamingContext = new StreamingContext(conf, Seconds(5))

    val dStream = streamingContext.socketTextStream("localhost", 21)

    dStream.print()

    streamingContext.start()
    streamingContext.awaitTermination()
  }
}
```

Tasks Console

```
BasicExample$ [Scala Application] C:\Program Files\Java\jre1.8.0_311\bin\javaw.exe (Dec 9, 2021, 2:21/12/09 15:09:10 INFO BlockManager: Removing RDD 234
21/12/09 15:09:10 INFO SocketInputDStream: Removing blocks of RDD BlockRDD[235] at socketTextStream at BasicExample.scala:12 of time 1639080555000 ms
21/12/09 15:09:10 INFO ReceivedBlockTracker: Deleting batches: 1639080545000 ms
21/12/09 15:09:10 INFO InputInfoTracker: remove old batch metadata: 1639080545000 ms
21/12/09 15:09:15 INFO JobScheduler: Added jobs for time 1639080555000 ms
-----
Time: 1639080555000 ms
-----
```

```
21/12/09 15:09:15 INFO JobScheduler: Starting job streaming job 1639080555000 ms
21/12/09 15:09:15 INFO JobScheduler: Finished job streaming job 1639080555000 ms
21/12/09 15:09:15 INFO BlockRDD: Removing RDD 235 from persistence list
21/12/09 15:09:15 INFO JobScheduler: Total delay: 0.001 s for time 1639080555000 ms (execution: 0.000 s)
21/12/09 15:09:15 INFO BlockManager: Removing RDD 235
21/12/09 15:09:15 INFO SocketInputDStream: Removing blocks of RDD BlockRDD[235] at socketTextStream at BasicExample.scala:12 of time 1639080555000 ms
21/12/09 15:09:15 INFO ReceivedBlockTracker: Deleting batches: 1639080545000 ms
21/12/09 15:09:15 INFO InputInfoTracker: remove old batch metadata: 1639080545000 ms
```

```
C:\Users\Lenovo\Downloads\nc111nt (1)>
```

```
listening on [any] 21 ...
connect to [127.0.0.1] from Bibek-Shah-Shankhar [127.0.0.1] 53434
Hello
How are you
```

Package Explorer

- SalesReportGenerated
- spark-streaming-examples
 - Scala Library container [2.11.11]
 - JRE System Library [JavaSE-1.8]
 - src
 - org.bibek.training.sparkstreaming
 - BasicExample.scala
 - Maven Dependencies
 - bin
 - target
 - pom.xml
- SparkSqlExamples
 - Scala Library container [2.11.11]
 - JRE System Library [JavaSE-1.8]
 - src
 - org.bibek.training.sparksq
 - HiveExample1.scala
 - RddToDFExample.scala
 - WithColumnExample.scala
 - org.bibek.training.sparksq.bean
 - Maven Dependencies
 - bin
 - spark-warehouse
 - target
 - pom.xml

```
BasicExample.scala  spark-streaming-examples/pom.xml  RddToDFExample.scala  SparkSqlExamples/pom.xml

package org.bibek.training.sparkstreaming

import org.apache.spark.SparkConf
import org.apache.spark.streaming.{Seconds, StreamingContext}

object BasicExample {

  def main(args: Array[String]) {
    val conf = (new SparkConf).setAppName("Spark Streaming Basic Example").setMaster("local[*]")
    val streamingContext = new StreamingContext(conf, Seconds(5))

    val dStream = streamingContext.socketTextStream("localhost", 21)

    dStream.print()

    streamingContext.start()
    streamingContext.awaitTermination()
  }
}
```

Tasks Console

BasicExample\$ [Scala Application] C:\Program Files\Java\jre1.8.0_311\bin\javaw.exe (Dec 9, 2021, 2:49:34 PM)

```
21/12/09 15:09:45 INFO BlockManager: Removing RDD 241
21/12/09 15:09:45 INFO SocketInputDStream: Removing blocks of RDD BlockRDD[241] at
21/12/09 15:09:45 INFO ReceivedBlockTracker: Deleting batches: 1639080575000 ms
21/12/09 15:09:45 INFO InputInfoTracker: remove old batch metadata: 1639080575000
21/12/09 15:09:50 INFO JobScheduler: Added jobs for time 1639080590000 ms
```

Time: 1639080590000 ms

```
21/12/09 15:09:50 INFO JobScheduler: Starting job streaming job 1639080590000 ms.0 from job set of time 1639080590000 ms
21/12/09 15:09:50 INFO JobScheduler: Finished job streaming job 1639080590000 ms.0 from job set of time 1639080590000 ms
21/12/09 15:09:50 INFO BlockRDD: Removing RDD 242 from persistence list
21/12/09 15:09:50 INFO JobScheduler: Total delay: 0.001 s for time 1639080590000 ms (execution: 0.000 s)
21/12/09 15:09:50 INFO BlockManager: Removing RDD 242
21/12/09 15:09:50 INFO SocketInputDStream: Removing blocks of RDD BlockRDD[242] at socketTextStream at BasicExample.scala:12 of time 1639080590000 ms
21/12/09 15:09:50 INFO ReceivedBlockTracker: Deleting batches: 1639080580000 ms
21/12/09 15:09:50 INFO InputInfoTracker: remove old batch metadata: 1639080580000 ms
```

C:\Windows\System32\cmd.exe -vv -l -p 21

```
C:\Users\Lenovo\Downloads\nc111nt (1)>
C:\Users\Lenovo\Downloads\nc111nt (1)>
C:\Users\Lenovo\Downloads\nc111nt (1)>
C:\Users\Lenovo\Downloads\nc111nt (1)>
C:\Users\Lenovo\Downloads\nc111nt (1)>
C:\Users\Lenovo\Downloads\nc111nt (1)>
C:\Users\Lenovo\Downloads\nc111nt (1)>nc.exe -vv -l -p 21
listening on [any] 21 ...
connect to [127.0.0.1] from Bibek-Shah-Shankhar [127.0.0.1] 53434
Hello
How are you
```

Bibek Shah
Shankhar

workspace - spark-streaming-examples/src/org/bibek/training/sparkstreaming/BasicExamples.scala - Scala IDE

File Edit Refactor Navigate Search Project Scala Run Window Help

Quick Access

Package Explorer

BasicExamples.scala

```
package org.bibek.training.sparkstreaming

import org.apache.spark._
import org.apache.spark.streaming._
import org.apache.spark.SparkConf
import org.apache.spark.SparkContext

object BasicExamples {

    def main(args:Array[String]) {
        val conf = (new SparkConf).setAppName("Spark Streaming Basic Example").setMaster("local[1]")

        val streamingContext = new StreamingContext(conf, Seconds(5))
        val dStream = streamingContext.socketTextStream("localhost", 21)

        dStream.print()

        streamingContext.start()

        streamingContext.awaitTermination()
    }
}
```

Console

```
BasicExamples$ [Scala Application] C:\Program Files\Java\jre1.8.0_311\bin\javaw.exe (Dec 9, 2021, 3:47:26 PM)
21/12/09 15:48:15 INFO JobScheduler: total delay: 0.016 s for time 1639082895000 ms (executed)
21/12/09 15:48:15 INFO BlockRDD: Removing RDD 9 from persistence list
21/12/09 15:48:15 INFO BlockManager: Removing RDD 9
21/12/09 15:48:15 INFO SocketInputDStream: Removing blocks of RDD BlockRDD[9] at socketText
21/12/09 15:48:15 INFO ReceivedBlockTracker: Deleting batches: 1639082885000 ms
21/12/09 15:48:15 INFO InputInfoTracker: remove old batch metadata: 1639082885000 ms
How are you

21/12/09 15:48:15 INFO BlockManagerInfo: Removed input-0-1639082887600 on 192.168.239.1:5470
21/12/09 15:48:20 INFO JobScheduler: Added jobs for time 1639082900000 ms
-----
Time: 1639082900000 ms
```

*Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

C:\Windows\System32\cmd.exe - vv -l -p 21

```
C:\Users\Lenovo\Downloads\nc111nt (1)>
listening on [any] 21 ...
connect to [127.0.0.1] from Bibek-Shah-Shankhar [127.0.0.1] 5470
Hello
How are you
```

3:48 PM
09-Dec-21

Bibek Shah
Shankhar

workspace - spark-streaming-examples/src/org/bibek/training/sparkstreaming/BasicExamples.scala - Scala IDE

File Edit Refactor Navigate Search Project Scala Run Window Help

Quick Access

Package Explorer

BasicExamples.scala

```
package org.bibek.training.sparkstreaming

import org.apache.spark._
import org.apache.spark.streaming._
import org.apache.spark.SparkConf
import org.apache.spark.SparkContext

object BasicExamples {

    def main(args:Array[String]) {
        val conf = (new SparkConf).setAppName("Spark Streaming Basic Example").setMaster("local[1]")
        val sparkContext = new SparkContext(conf)

        val streamingContext = new StreamingContext(sparkContext, Seconds(5))
        val dStream = streamingContext.socketTextStream("localhost", 21)

        dStream.print()

        streamingContext.start()
        //streamingContext.awaitTermination()
        streamingContext.stop(false)
        println(sparkContext.isStopped)
    }
}
```

Console

```
<terminated> BasicExamples$ [Scala Application] C:\Program Files\Java\jre1.8.0_311\bin\javaw.exe (Dec 9, 2021, 3:50:10 PM)
false
21/12/09 15:50:12 INFO SparkContext: Invoking stop() from shutdown hook
21/12/09 15:50:12 INFO SparkUI: Stopped Spark web UI at http://192.168.239.1:4041
21/12/09 15:50:12 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
21/12/09 15:50:12 INFO MemoryStore: MemoryStore cleared
21/12/09 15:50:12 INFO BlockManager: BlockManager stopped
21/12/09 15:50:12 INFO BlockManagerMaster: BlockManagerMaster stopped
21/12/09 15:50:12 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoord
21/12/09 15:50:12 INFO SparkContext: Successfully stopped SparkContext
21/12/09 15:50:12 INFO ShutdownHookManager: Shutdown hook called
21/12/09 15:50:12 INFO ShutdownHookManager: Deleting directory C:\Users\Lenovo\AppData\Local\Temp
```

Writable Smart Insert 19 : 29 598M of 1178M

μTorrent

VMS Workstation

JRE System Library

src

Maven Dependencies

bin

target

pom.xml

SparkSQLExamples

Scala Library cont.

JRE System Library

src

org.bibek.training

HiveExample

RddToDFExample

WithColumn

org.bibek.training

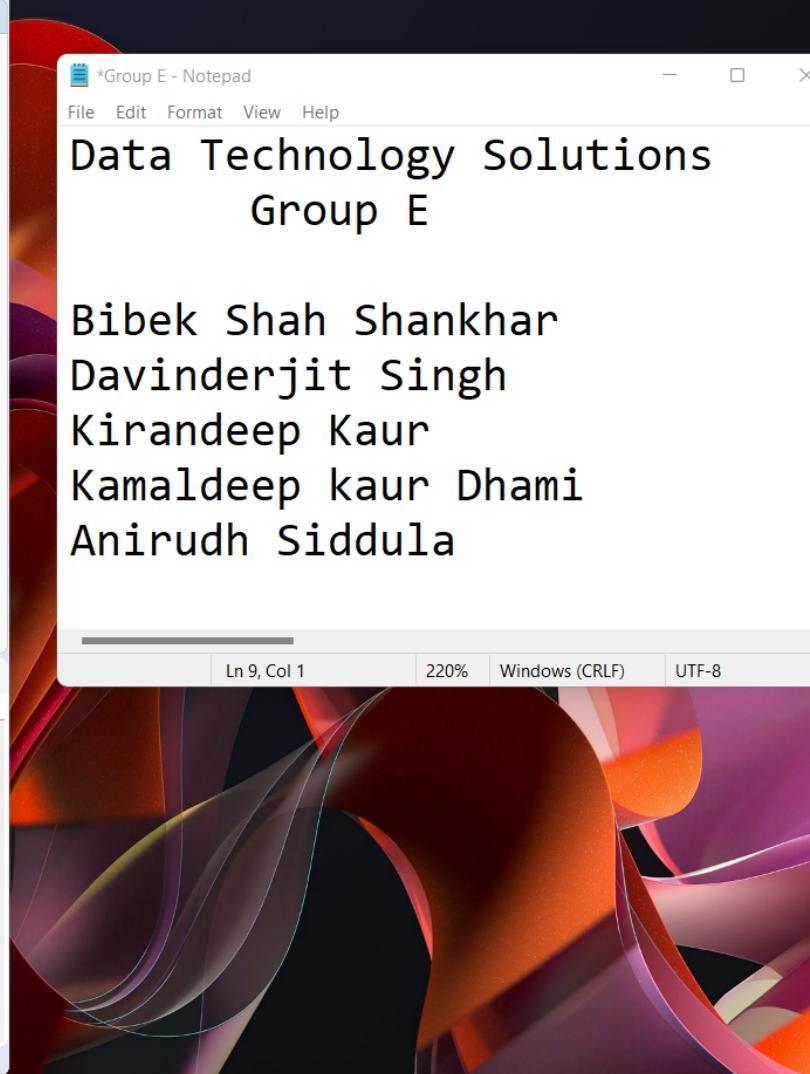
Maven Dependencies

bin

spark-warehouse

target

pom.xml



workspace - spark-streaming-examples/src/org/bibek/training/sparkstreaming/BasicExamples.scala - Scala IDE

File Edit Refactor Navigate Search Project Scala Run Window Help

Quick Access

Packa... BasicExamples.scala spark-streaming-examples/pom.xml

```
import org.apache.spark._  
import org.apache.spark.streaming._  
import org.apache.spark.SparkConf  
import org.apache.spark.SparkContext  
  
object BasicExamples {  
  
    def main(args: Array[String]) {  
        val conf = (new SparkConf).setAppName("Spark Streaming Basic Example")  
        val sparkContext = new SparkContext(conf)  
  
        val streamingContext = new StreamingContext(sparkContext, Seconds(2))  
        val lines1 = streamingContext.socketTextStream("localhost", 21)  
        val lines2 = streamingContext.socketTextStream("localhost", 22)  
  
        val unionedLines = lines1.union(lines2)  
  
        unionedLines.print()  
  
        streamingContext.start()  
  
        streamingContext.awaitTermination()  
    }  
}
```

Console

```
BasicExamples$ [Scala Application] C:\Program Files\Java\jre1.8.0_311\bin\javaw.exe (Dec 9, 2021, 21/12/09 16:22:05 INFO UnionRDD: Removing RDD 934 from persistence list  
21/12/09 16:22:05 INFO JobScheduler: Total delay: 0.001 s for time 1639084925000  
21/12/09 16:22:05 INFO BlockManager: Removing RDD 934  
21/12/09 16:22:05 INFO BlockRDD: Removing RDD 932 from persistence list  
21/12/09 16:22:05 INFO BlockManager: Removing RDD 932  
21/12/09 16:22:05 INFO SocketInputDStream: Removing blocks of RDD BlockRDD[932]  
21/12/09 16:22:05 INFO BlockRDD: Removing RDD 933 from persistence list  
21/12/09 16:22:05 INFO BlockManager: Removing RDD 933  
21/12/09 16:22:05 INFO SocketInputDStream: Removing blocks of RDD BlockRDD[933]  
21/12/09 16:22:05 INFO ReceivedBlockTracker: Deleting batches: 1639084915000 ms  
21/12/09 16:22:05 INFO InputInfoTracker: remove old batch metadata: 1639084915000 ms
```

C:\Windows\System32\cmd.exe - nc.exe -vv -l -p 21

```
Microsoft Windows [Version 10.0.22000.318]  
(c) Microsoft Corporation. All rights reserved.  
  
C:\Users\Lenovo\Downloads\nc111nt (1)>nc.exe -vv -l -p 21  
nc.exe: invalid option -- 1  
nc -h for help  
  
C:\Users\Lenovo\Downloads\nc111nt (1)>nc.exe -vv -l -p 21  
listening on [any] 21 ...  
connect to [127.0.0.1] from Bibek-Shah-Shankhar [127.0.0.1] 55266  
21  
21  
21
```

C:\Windows\System32\cmd.exe - nc.exe -vv -l -p 22

```
Microsoft Windows [Version 10.0.22000.318]  
(c) Microsoft Corporation. All rights reserved.  
  
C:\Users\Lenovo\Downloads\nc111nt (1)>nc.exe -vv -l -p 22  
listening on [any] 22 ...  
connect to [127.0.0.1] from Bibek-Shah-Shankhar [127.0.0.1] 57673  
22
```

*Group E - Notepad

File Edit Format View Help

Data Technology Solutions

Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 9, Col 1 220% Windows (CRLF)

Writable Smart Insert 25 657M of 1182M

4:22 PM 09-Dec-21

workspace - spark-streaming-examples/src/org/bibek/training/sparkstreaming/BasicExamples.scala - Scala IDE

File Edit Refactor Navigate Search Project Scala Run Window Help

Quick Access

Packa... BasicExamples.scala spark-streaming-examples/pom.xml

```
package org.bibek.training.sparkstreaming

import org.apache.spark._
import org.apache.spark.streaming._
import org.apache.spark.SparkConf
import org.apache.spark.SparkContext

object BasicExamples {

  def main(args: Array[String]) {
    val conf = (new SparkConf).setAppName("Spark Streaming Basic Example").setMaster("local[2]")
    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)
  }
}
```

Console

```
<terminated> BasicExamples$ [Scala Application] C:\Program Files\Java\jre1.8.0_311\bin\javaw.exe (Dec 9, 2021, 4:38:21) 21/12/09 16:41:55 INFO BlockManager: Removing RDD 111
21/12/09 16:41:55 INFO SocketInputDStream: Removing blocks of RDD BlockRDD[111] at sc
21/12/09 16:41:55 INFO ReceivedBlockTracker: Deleting batches: 1639086105000 ms
21/12/09 16:41:55 INFO BlockManagerInfo: Removed input-0-1639086107000 on 192.168.239
21/12/09 16:41:55 INFO InputInfoTracker: remove old batch metadata: 1639086105000 ms
d
e

21/12/09 16:41:55 INFO BlockManagerInfo: Removed input-0-1639086109200 on 192.168.239
```

639M of 1246M

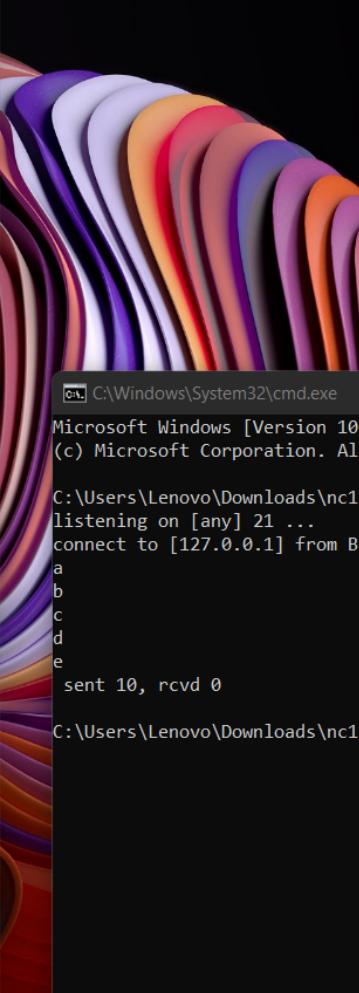
*Group E - Notepad

File Edit Format View Help

Data Technology Solutions

Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula



C:\Windows\System32\cmd.exe

Microsoft Windows [Version 10.0.22000.318]
(c) Microsoft Corporation. All rights reserved.

```
C:\Users\Lenovo\Downloads\nc111nt (1)>nc.exe -vv -l -p 21
listening on [any] 21 ...
connect to [127.0.0.1] from Bibek-Shah-Shankhar [127.0.0.1] 58805
a
b
c
d
e
sent 10, rcvd 0
C:\Users\Lenovo\Downloads\nc111nt (1)>
```

The screenshot shows the IntelliJ IDEA interface with the following details:

- File Structure:** The left sidebar displays the project structure under "workspace - spark-streaming-examples". It includes packages like SalesReportGenerate, spark-streaming-examples, SparkSQLExamples, and org.bibek.training. Each package has its own source code files (e.g., BasicExample.scala) and configuration files (e.g., pom.xml).
- Editor:** The main editor window shows the content of `BasicExamples.scala`. The code defines a `CountByWindowExample` object with a `main` method. This method sets up a `SparkConf`, creates a `StreamingContext`, and defines a `dStream` as a socket text stream from "localhost" on port 21. It then performs a `countByWindow` operation on `dStream` using windows of 3 seconds and sliding every 2 seconds. Finally, it prints the results and starts the streaming context.
- Console:** The bottom window shows the execution output of the Scala application. It includes the command `CountByWindowExample$ [Scala Application] C:\Program Files\Java\jre1.8.0_311\bin\javaw.exe`, log messages from the executor and task manager, and the completion message "Job 20 finished: print at BasicExamples.scala:18, took 0.01s". The time taken for the job is listed as 1639086985000 ms.

workspace - spark-streaming-examples/src/org/bibek/training/sparkstreaming/ReduceByKeyAndWindowExample.scala - Scala IDE

File Edit Refactor Navigate Search Project Scala Run Window Help

Package Explorer

- SalesReportGenerated
- spark-streaming-examples
- SparkSqlExamples

spark-stream... ReduceByKeyA... BasicExample... HiveExample... >1

```
package org.bibek.training.sparkstreaming

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(1))

    streamingContext.checkpoint("C:/Users/Lenovo/Documents/test_data/checkpoint")

    val dStream = streamingContext.socketTextStream("localhost", 21)
    val windowedCountDStream = dStream.countByWindow(Seconds(3), Seconds(2))

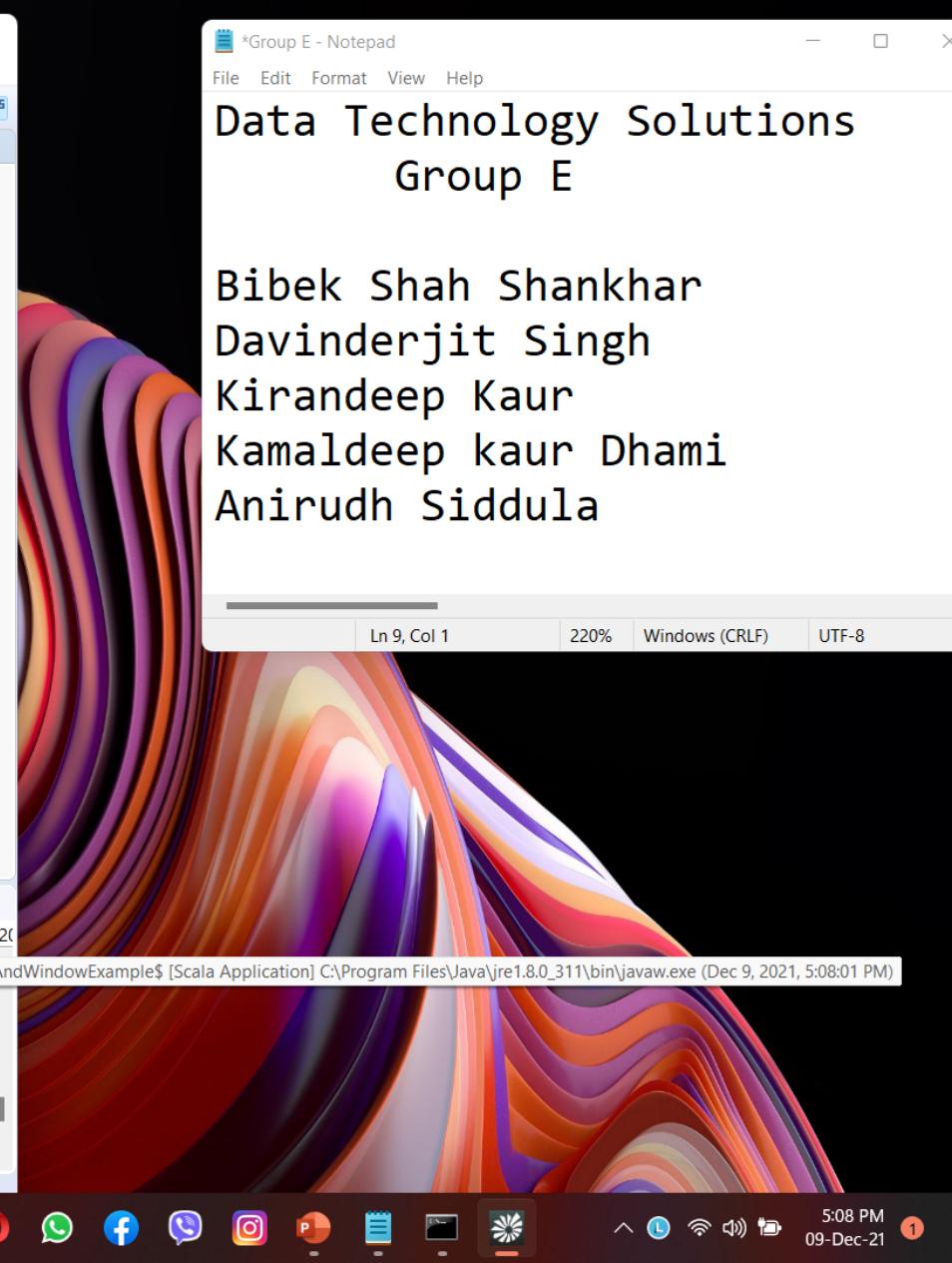
    windowedCountDStream.foreachRDD(rdd => {
      println("Hello")
      rdd.saveAsTextFile("C:/Users/Lenovo/Documents/test_data/output")
    })

    streamingContext.start()
    streamingContext.awaitTermination()
  }
}
```

Console

```
ReduceByKeyAndWindowExample$ [Scala Application] C:\Program Files\Java\jre1.8.0_311\bin\javaw.exe (Dec 9, 2021, 5:08:04 PM)
  at java.lang.Thread.run(Unknown Source)
21/12/09 17:08:04 INFO RecurringTimer: Started timer for BlockGenerator
21/12/09 17:08:04 INFO BlockGenerator: Started BlockGenerator
21/12/09 17:08:04 INFO BlockGenerator: Started block pushing thread
21/12/09 17:08:04 INFO ReceiverTracker: Registered receiver for stream 0 from 192.168.1.11
21/12/09 17:08:04 INFO ReceiverSupervisorImpl: Starting receiver 0
21/12/09 17:08:04 INFO SocketReceiver: Connecting to localhost:21
```

Writable Smart Insert 25 · 30 707M of 1079M



*Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

Ln 9, Col 1 220% Windows (CRLF) UTF-8

