

The image shows a Windows desktop environment with two open windows. On the left is a Scala IDE window titled "workspace - SparkSqlExamples/src/org/bibek/training/sparksq/RddToDFExample.scala - Scala IDE". The code editor displays a Scala script named RddToDFExample.scala, which reads a CSV file and creates a DataFrame. The package is org.bibek.training.sparksq. The code uses imports from org.apache.spark and org.bibek.training.sparksq.bean. On the right is a Notepad window titled "Group E - Notepad" with the title "Data Technology Solutions Group E". It contains a list of names: Bibek Shah, Shankhar, Davinderjit Singh, Kirandeep Kaur, Kamaldeep kaur Dhami, and Anirudh Siddula. The desktop background is a blue-toned landscape image. The taskbar at the bottom shows various pinned icons and the system tray with the date and time.

File Edit Refactor Navigate Search Project Scala Run Window Help

Quick Access

RddToDFExample.scala WithColumnExample.scala Sales.scala SparkSqlExam...

package org.bibek.training.sparksq

```
import org.apache.spark.SparkConf
import org.apache.spark.SparkContext
import org.apache.spark.sql.SparkSession
import org.bibek.training.sparksq.bean.Sales
```

object RddToDFExample {

```
def main(args: Array[String]) {
    val conf = (new SparkConf)
        .setAppName("RDD to DF Example")
        .setMaster("local[*]")

    val context = new SparkContext(conf)

    val sparkSession = SparkSession.builder()
        .appName("RDD to DF Example")
        .master("local[*]")
        .getOrCreate()

    val rdd = context
        .textFile("C:/Users/Lenovo/Documents/SalesJan2009.csv")
        .map(line => {
            val fields = line.split(",")
            new Sales(fields(0).trim(), fields(1).trim().toInt, fields(2).trim())
        })
}

import sparkSession.implicits._

val df = rdd.toDF()
//df.printSchema()
df.show(10)
```

Tasks

0 items

Description Resource Path

Writable Smart Insert 14 545M of 1061M

File Edit Format View Help

Data Technology Solutions  
Group E

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

Ln 8, Col 16 170% Windows (CRLF) UTF-8

8:38 PM 01-Dec-21

The screenshot shows the IntelliJ IDEA interface with a Scala project named "SparkSQLExamples". The code editor displays the file `RddToDFExample.scala` which contains Scala code for reading data from Hive and creating a DataFrame. The terminal window below shows the execution results, including the generated code and the resulting DataFrame output.

```
package org.bibek.training.sparksql

import org.apache.spark.SparkConf
import org.apache.spark.SparkContext
import org.apache.spark.sql.SparkSession
import org.bibek.training.sparksql.bean.Sales

object RddToDFExample {

  def main(args: Array[String]) {
    val conf = (new SparkConf)
      .setAppName("RDD to DF Example")
      .setMaster("local[*]")

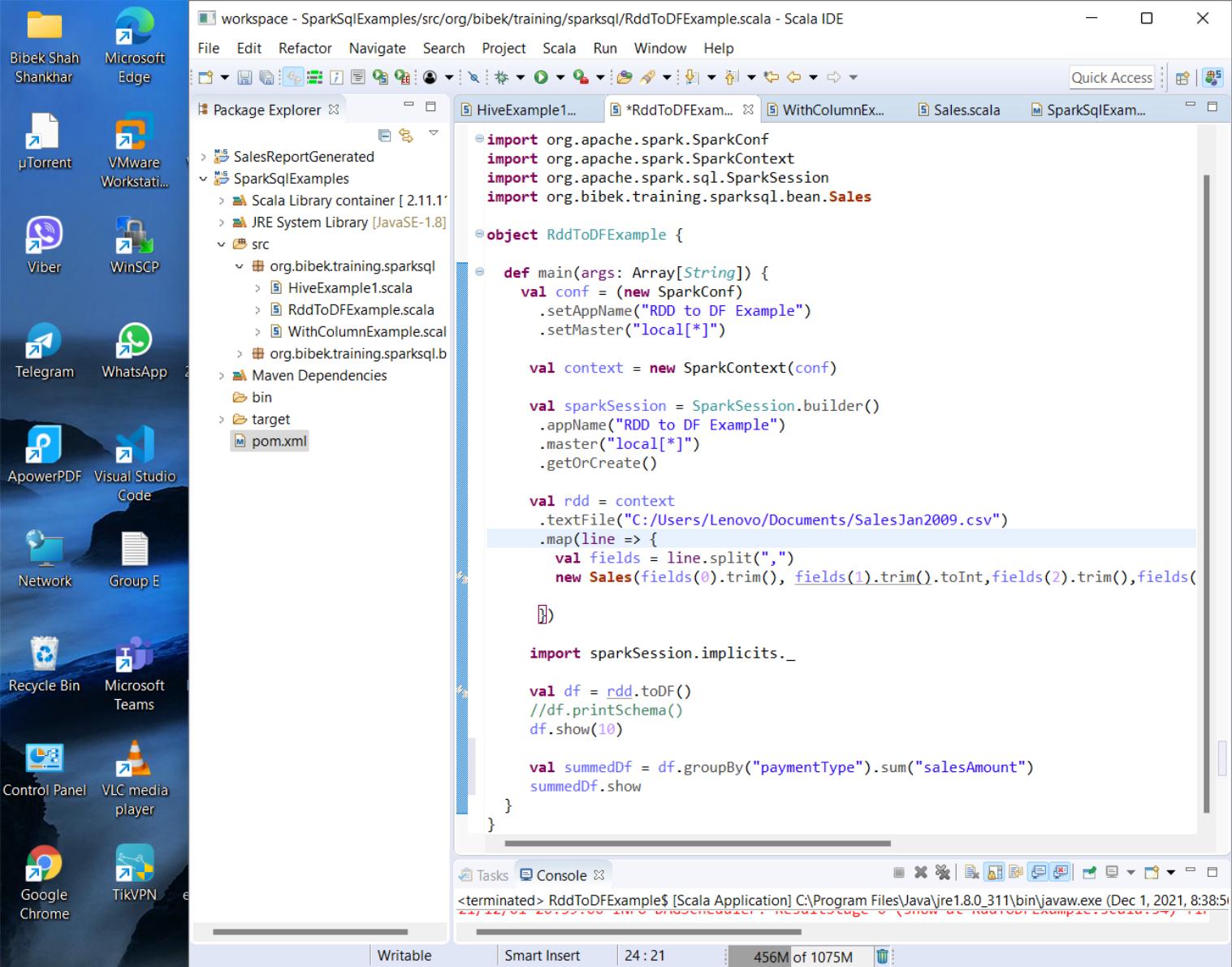
    val context = new SparkContext(conf)

    val sparkSession = SparkSession.builder()
      .appName("RDD to DF Example")
      .master("local[*]")
      .getOrCreate()
  }
}
```

Tasks Console

```
<terminated> RddToDFExample$ [Scala Application] C:\Program Files\Java\jre1.8.0_311\bin\javaw.exe (Dec 1, 2021, 8:38:56 PM)
21/12/01 20:39:00 INFO DAGScheduler: Job 0 finished: show at RddToDFExample.scala:34, took 0.198466 s
21/12/01 20:39:00 INFO CodeGenerator: Code generated in 11.5433 ms
+-----+-----+-----+-----+-----+-----+-----+
|productId|salesAmount|paymentType|customerName|city|region|country|
+-----+-----+-----+-----+-----+-----+-----+
| Product1| 1200 | Mastercard| carolina| Basildon| England|United Kingdom|
| Product1| 1200 | Visa| Betina| Parkville| MO| United States|
| Product1| 1200 | Mastercard| Federica e Andreia| Astoria| OR| United States|
| Product1| 1200 | Visa| Gouya| Echuca| Victoria| Australia|
| Product2| 3600 | Visa| Gerd W Cahaba Heights| Mickleton| AL| United States|
| Product1| 1200 | Visa| LAURENCE| Martin| NJ| United States|
| Product1| 1200 | Mastercard| Fleur| Peoria| IL| United States|
| Product1| 1200 | Mastercard| adam| Martin| TN| United States|
| Product1| 1200 | Mastercard| Renee Elisabeth| Tel Aviv| Tel Aviv| Israel|
| Product1| 1200 | Visa| Aidan| Chatou| Ile-de-France| France|
+-----+-----+-----+-----+-----+-----+-----+
only showing top 10 rows
```





Bibek Shah  
Shankhar

Microsoft Edge

µTorrent

VMware Workstation

Viber

WinSCP

Telegram

WhatsApp

ApowerPDF

Visual Studio Code

Network

Group E

Recycle Bin

Microsoft Teams

Control Panel

VLC media player

Google Chrome

TikVPN

workspace - SparkSqlExamples/src/org/bibek/training/sparksql/RddToDFExample.scala - Scala IDE

File Edit Refactor Navigate Search Project Scala Run Window Help

Quick Access

Package Explorer

HiveExample1... RddToDFExam... WithColumnEx... Sales.scala SparkSqlExam...

```
val context = new SparkContext(conf)

val sparkSession = SparkSession.builder()
    .appName("RDD to DF Example")
    .master("local[*]")
    .getOrCreate()

val rdd = context
    .textFile("C:/Users/Lenovo/Documents/SalesJan2009.csv")
    .map(line => {
        val fields = line.split(",")
        new Sales(fields(0).trim(), fields(1).trim().toInt, fields(2).trim(), fields(3).trim())
    })

import sparkSession.implicits._

val df = rdd.toDF()
//df.printSchema()
df.show(10)

val summedDF = df.groupBy("paymentType").sum("salesAmount")
summedDF.show
```

Tasks Console

```
<terminated> RddToDFExample$ [Scala Application] C:\Program Files\Java\jre1.8.0_311\bin\javaw.exe (Dec 1, 2021, 8:41:22)  
+-----+  
| paymentType|sum(salesAmount)|  
+-----+  
| Visa| 849350|  
| Diners| 133800|  
| Mastercard| 458450|  
| Amex| 188900|  
+-----+
```

21/12/01 20:41:31 INFO SparkContext: Invoking stop() from shutdown hook  
21/12/01 20:41:31 INFO SparkUI: Stopped Spark web UI at http://192.168.239.1:4040

Writable Smart Insert 31 : 6 537M of 1075M



workspace - SparkSQLExamples/src/org/bibek/training/sparksql/WithColumnExample.scala - Scala IDE

File Edit Refactor Navigate Search Project Scala Run Window Help

Quick Access

Package Explorer

SalesReportGenerated  
SparkSQLExamples  
Scala Library container [ 2.11.1 ]  
JRE System Library [ JavaSE-1.8 ]  
src  
org.bibek.training.sparksql  
HiveExample1.scala  
RddToDFExample.scala  
WithColumnExample.scala  
org.bibek.training.sparksql.t  
Maven Dependencies  
bin  
target  
pom.xml

```
object WithColumnExample {  
  
    def main(args: Array[String]) {  
        val sparkSession = SparkSession.builder()  
            .appName("RDD to DF Example")  
            .master("local[*]")  
            .getOrCreate()  
  
        val df = sparkSession  
            .read  
            .option("header", true)  
            .option("inferSchema", true)  
            .csv("C:/Users/Lenovo/Documents/SalesJan2009.csv")  
            .withColumn("ProcessingDate", lit(getFormattedDate(new Date)))  
  
        //df.withColumnRenamed("Id", "ProductId")  
  
        //df.show(10)  
  
        //save as partitioned data  
        df.write  
            .partitionBy("ProcessingDate")  
            .option("header", true)  
            .csv("C:/Users/Lenovo/Documents/test_data/output")  
    }  
  
    def getFormattedDate(date: Date) = {  
        val dateFormat = new SimpleDateFormat("yyyy-MM-dd")  
        val formattedDate = dateFormat.format(date)  
        formattedDate  
    }  
}
```

Tasks Console

<terminated> WithColumnExample\$ [Scala Application] C:\Program Files\Java\jre1.8.0\_311\bin\javaw.exe (Dec 1, 2021, 8:58 PM)  
in 102 160 220 1.60176 in memory (free: 6.0 KB free: 1007 0 MB)

Writable Smart Insert 30 : 18 616M of 1031M

Group E - Notepad

Data Technology Solutions  
Group E

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

Ln 8, Col 16 170% Windows (CRLF) UTF-8

8:58 PM 01-Dec-21

The image shows a Windows desktop environment with several application icons on the left and two open windows in the center.

**Open Applications:**

- Scala IDE**: The main window displays a Scala code editor for a file named `HiveExample1.scala`. The code implements a `main` method to read a CSV file, process it with a UDF, and save it as a partitioned DataFrame. A taskbar at the bottom shows the Scala IDE is running with 636M of 1011M memory used.
- Notepad**: A window titled "Group E - Notepad" contains the text "Data Technology Solutions Group E" followed by a list of names: Bibek Shah, Shankhar, Davinderjit Singh, Kirandeep Kaur, Kamaldeep kaur Dhami, and Anirudh Siddula.

**Desktop Icons (Left Side):**

- Bibek Shah
- Microsoft Edge
- µTorrent
- VMware Workstation
- Viber
- WinSCP
- Telegram
- WhatsApp
- ApowerPDF
- Visual Studio Code
- Network
- Group E
- Recycle Bin
- Microsoft Teams
- Control Panel
- VLC media player
- Google Chrome
- TikVPN

**Taskbar Icons (Bottom):**

Standard Windows taskbar icons for search, file operations, messaging, email, file explorer, and social media.

**System Status Bar (Bottom Right):**

9:10 PM 01-Dec-21