

## TASK 1

1) Sum of all numbers

```
import org.apache.spark.{SparkConf, SparkContext}
object SampleSparkDemo {
    def main(args: Array[String]): Unit = {
        val conf = new SparkConf().setAppName("SparkSampleTest").setMaster("local[*]")
        val sc = new SparkContext(conf)
        val sampleList = List(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
        val parallelized_list = sc.parallelize(sampleList)
        println("Sum of given list elements is " + parallelized_list.reduce(_ + _))
    }
}
```

SparkSampleModule [C:\Users\Raj\IdeaProjects\SparkSampleModule] - ...src\main\scala-2.11\org\spark\_samples\SampleSparkDemo.scala [sparksamplemodule]

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

SparkSampleModule > src > main > scala-2.11 > org > spark\_samples > SampleSparkDemo.scala

SampleSparkDemo.scala

```
2
3 import org.apache.spark.{SparkConf, SparkContext}
4
5 object SampleSparkDemo {
6
7     def main(args: Array[String]): Unit = {
8
9         val conf = new SparkConf().setAppName("SparkSampleTest").setMaster("local[*]")
10        val sc = new SparkContext(conf)
11        val sampleList = List(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
12        val parallelized_list = sc.parallelize(sampleList)
13        println("Sum of given list elements is " + parallelized_list.reduce(_ + _))
14    }
15}
16
```

Run SampleSparkDemo

```
18/02/17 18:56:02 INFO TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
18/02/17 18:56:02 INFO DAGScheduler: ResultStage 0 (reduce at SampleSparkDemo.scala:13) finished in 0.528 s
18/02/17 18:56:02 INFO DAGScheduler: Job 0 finished: reduce at SampleSparkDemo.scala:13, took 1.451050 s
Sum of given list elements is 55
18/02/17 18:56:02 INFO SparkContext: Invoking stop() from shutdown hook
18/02/17 18:56:02 INFO SparkUI: Stopped Spark web UI at http://192.168.43.50:4040
18/02/17 18:56:02 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/02/17 18:56:02 INFO MemoryStore: MemoryStore cleared
18/02/17 18:56:02 INFO BlockManager: BlockManager stopped
18/02/17 18:56:02 INFO BlockManagerMaster: BlockManagerMaster stopped
18/02/17 18:56:02 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
18/02/17 18:56:02 INFO SparkContext: Successfully stopped SparkContext
18/02/17 18:56:02 INFO ShutdownHookManager: Shutdown hook called
18/02/17 18:56:02 INFO ShutdownHookManager: Deleting directory C:\Users\Raj\AppData\Local\Temp\spark-922eb20c-3c7e-4

Process finished with exit code 0
```

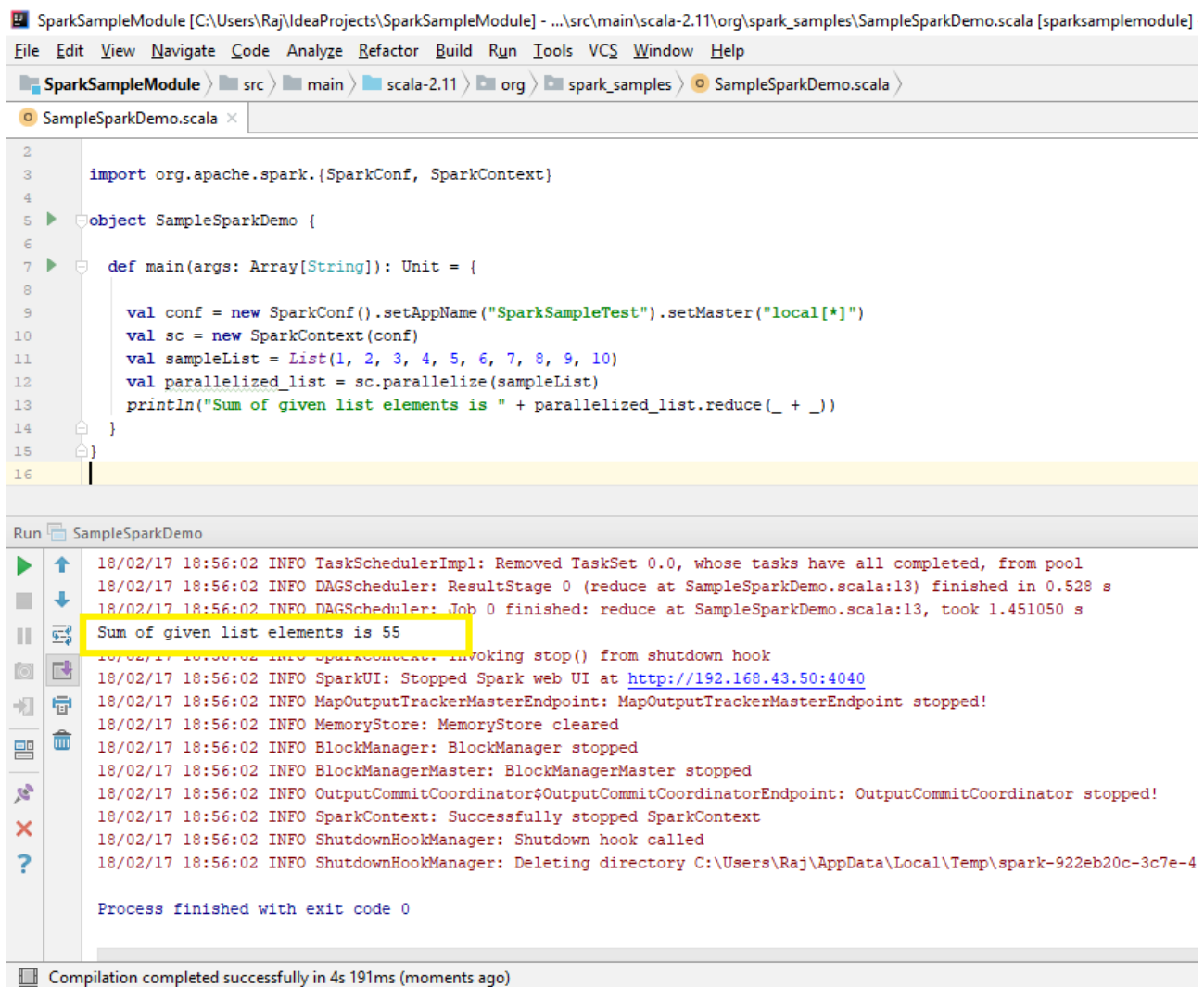
Compilation completed successfully in 4s 191ms (moments ago)

### Output:

Total number of elements in the list is 55

## 2) Total elements in the list

```
import org.apache.spark.{SparkConf,SparkContext}
object SampleSparkDemo {
    def main(args: Array[String]): Unit = {
        val conf = new SparkConf().setAppName("SparkSampleTest").setMaster("local[*]")
        val sc = new SparkContext(conf)
        val sampleList = List(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
        val parallelized_list = sc.parallelize(sampleList)
        println("Total number of elements in the list is " + parallelized_list.count())
    }
}
```



The screenshot displays an IDE window for a project named 'SparkSampleModule'. The editor shows the Scala code for 'SampleSparkDemo.scala'. The code imports 'org.apache.spark.{SparkConf, SparkContext}', defines an object 'SampleSparkDemo', and a 'main' method. Inside 'main', it creates a 'SparkConf' with app name 'SparkSampleTest' and master 'local[\*]', a 'SparkContext' from the conf, a 'List' of numbers 1 to 10, and a 'parallelized\_list' by calling 'sc.parallelize(sampleList)'. It then prints the sum of the list elements using 'parallelized\_list.reduce(\_ + \_)'. The output console shows the execution of 'SampleSparkDemo', with logs from 'TaskSchedulerImpl', 'DAGScheduler', and 'SparkContext'. A yellow box highlights the output line: 'Sum of given list elements is 55'. The console also shows the shutdown of the Spark web UI and other components. At the bottom, a status bar indicates 'Compilation completed successfully in 4s 191ms (moments ago)'.

```
import org.apache.spark.{SparkConf, SparkContext}
object SampleSparkDemo {
    def main(args: Array[String]): Unit = {
        val conf = new SparkConf().setAppName("SparkSampleTest").setMaster("local[*]")
        val sc = new SparkContext(conf)
        val sampleList = List(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
        val parallelized_list = sc.parallelize(sampleList)
        println("Sum of given list elements is " + parallelized_list.reduce(_ + _))
    }
}
```

Run SampleSparkDemo

```
18/02/17 18:56:02 INFO TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
18/02/17 18:56:02 INFO DAGScheduler: ResultStage 0 (reduce at SampleSparkDemo.scala:13) finished in 0.528 s
18/02/17 18:56:02 INFO DAGScheduler: Job 0 finished: reduce at SampleSparkDemo.scala:13, took 1.451050 s
Sum of given list elements is 55
18/02/17 18:56:02 INFO SparkContext: Invoking stop() from shutdown hook
18/02/17 18:56:02 INFO SparkUI: Stopped Spark web UI at http://192.168.43.50:4040
18/02/17 18:56:02 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/02/17 18:56:02 INFO MemoryStore: MemoryStore cleared
18/02/17 18:56:02 INFO BlockManager: BlockManager stopped
18/02/17 18:56:02 INFO BlockManagerMaster: BlockManagerMaster stopped
18/02/17 18:56:02 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
18/02/17 18:56:02 INFO SparkContext: Successfully stopped SparkContext
18/02/17 18:56:02 INFO ShutdownHookManager: Shutdown hook called
18/02/17 18:56:02 INFO ShutdownHookManager: Deleting directory C:\Users\Raj\AppData\Local\Temp\spark-922eb20c-3c7e-4

Process finished with exit code 0
```

Compilation completed successfully in 4s 191ms (moments ago)

## Output:

Total number of elements in the list is 10

### 3) Average of the numbers in the list

```
import org.apache.spark.{SparkConf, SparkContext}

object SampleSparkDemo {
    def main(args: Array[String]): Unit = {
        val conf = new SparkConf().setAppName("SparkSampleTest").setMaster("local[*]")
        val sc = new SparkContext(conf)
        val sampleList = List(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
        val parallelized_list = sc.parallelize(sampleList)
        println("Average of numbers in the list is " + parallelized_list.reduce(_ + _).toFloat /
parallelized_list.count())
    }
}
```

```
SparkSampleModule [C:\Users\Raj\IdeaProjects\SparkSampleModule] - ...src\main\scala-2.11\org\spark_samples\SampleSparkDemo.scala [sparksamplemodule] - Int
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
SparkSampleModule > src > main > scala-2.11 > org > spark_samples > SampleSparkDemo.scala >
SampleSparkDemo.scala x
4
5 object SampleSparkDemo {
6
7 def main(args: Array[String]): Unit = {
8
9     val conf = new SparkConf().setAppName("SparkSampleTest").setMaster("local[*]")
10    val sc = new SparkContext(conf)
11    val sampleList = List(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
12    val parallelized_list = sc.parallelize(sampleList).cache()
13    println("Sum of given list elements is " + parallelized_list.reduce(_ + _))
14    println("Total number of elements in the list is " + parallelized_list.count())
15    println("Average of numbers in the list is " + parallelized_list.reduce(_ + _).toFloat / parallelized_list.count())
16  }
17 }
18

Run SampleSparkDemo
18/02/17 19:38:30 INFO TaskSetManager: Finished task 0.0 in stage 3.0 (TID 12) in 20 ms on localhost (1/4)
18/02/17 19:38:30 INFO TaskSetManager: Finished task 1.0 in stage 3.0 (TID 13) in 20 ms on localhost (2/4)
18/02/17 19:38:30 INFO Executor: Finished task 2.0 in stage 3.0 (TID 14). 954 bytes result sent to driver
18/02/17 19:38:30 INFO TaskSetManager: Finished task 3.0 in stage 3.0 (TID 15) in 18 ms on localhost (3/4)
18/02/17 19:38:30 INFO TaskSetManager: Finished task 2.0 in stage 3.0 (TID 14) in 20 ms on localhost (4/4)
18/02/17 19:38:30 INFO TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
18/02/17 19:38:30 INFO DAGScheduler: ResultStage 3 (count at SampleSparkDemo.scala:15) finished in 0.023 s
Average of numbers in the list is 5.5
18/02/17 19:38:30 INFO DAGScheduler: Job 3 finished: count at SampleSparkDemo.scala:15, took 0.040473 s
18/02/17 19:38:30 INFO SparkContext: Invoking stop() from shutdown hook
18/02/17 19:38:30 INFO SparkUI: Stopped Spark web UI at http://192.168.43.50:4040
18/02/17 19:38:30 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/02/17 19:38:30 INFO MemoryStore: MemoryStore cleared
18/02/17 19:38:30 INFO BlockManager: BlockManager stopped
18/02/17 19:38:30 INFO BlockManagerMaster: BlockManagerMaster stopped
18/02/17 19:38:30 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
18/02/17 19:38:30 INFO SparkContext: Successfully stopped SparkContext
18/02/17 19:38:30 INFO ShutdownHookManager: Shutdown hook called

Compilation completed successfully in 3s 577ms (moments ago)
```

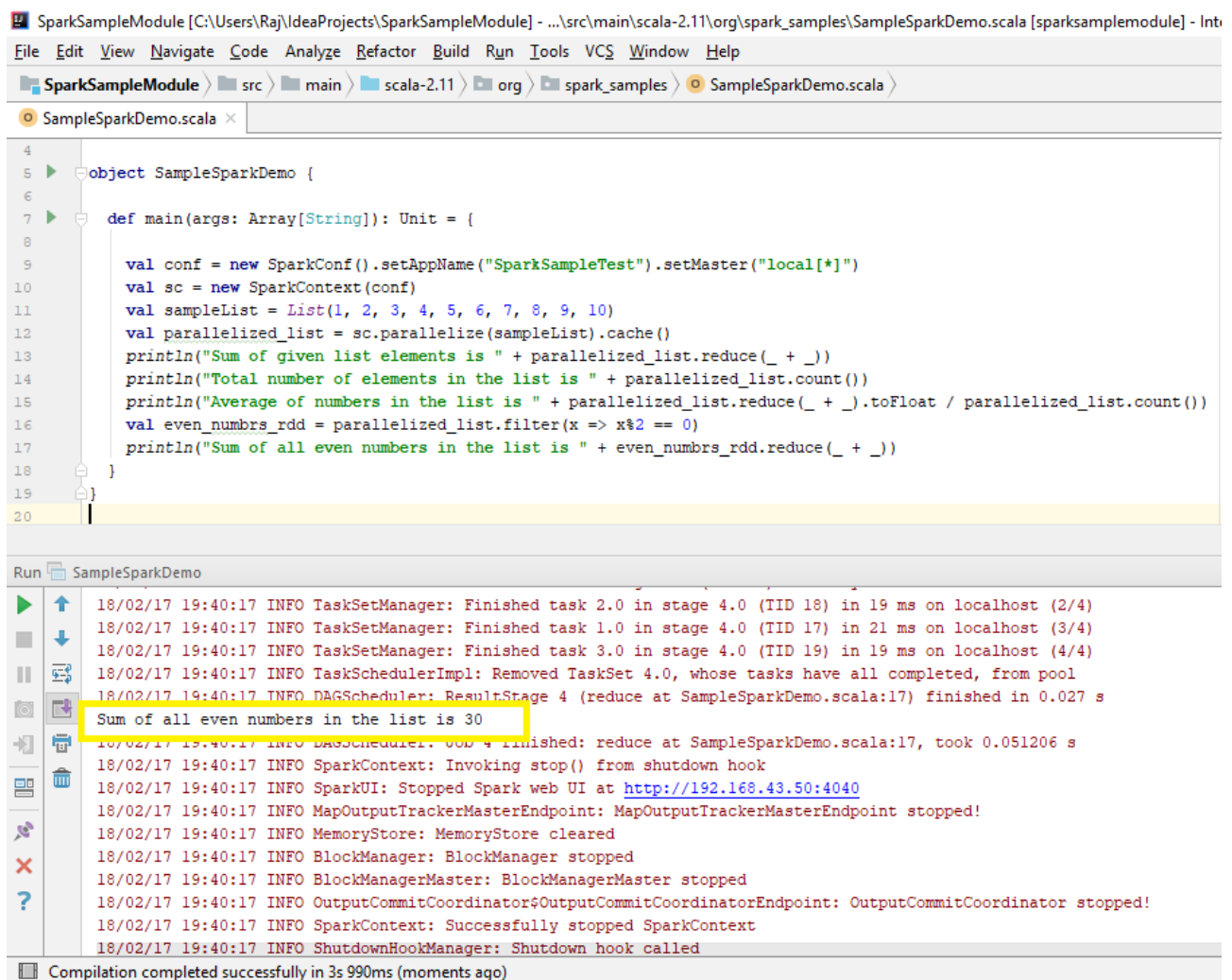
### Output:

Average of numbers in the list is 5.5

4) Sum of all the even numbers in the list

```
import org.apache.spark.{SparkConf, SparkContext}

object SampleSparkDemo {
    def main(args: Array[String]): Unit = {
        val conf = new SparkConf().setAppName("SparkSampleTest").setMaster("local[*]")
        val sc = new SparkContext(conf)
        val sampleList = List(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
        val parallelized_list = sc.parallelize(sampleList)
        val even_numbrs_rdd = parallelized_list.filter(x => x%2 == 0)
        println("Sum of all even numbers in the list is " + even_numbrs_rdd.reduce(_ + _))
    }
}
```



The screenshot shows an IDE window titled "SparkSampleModule [C:\Users\Raj\IdeaProjects\SparkSampleModule] - ...\src\main\scala-2.11\org\spark\_samples\SampleSparkDemo.scala [sparksamplemodule] - Int". The code editor displays the following Scala code:

```
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
object SampleSparkDemo {
    def main(args: Array[String]): Unit = {
        val conf = new SparkConf().setAppName("SparkSampleTest").setMaster("local[*]")
        val sc = new SparkContext(conf)
        val sampleList = List(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
        val parallelized_list = sc.parallelize(sampleList).cache()
        println("Sum of given list elements is " + parallelized_list.reduce(_ + _))
        println("Total number of elements in the list is " + parallelized_list.count())
        println("Average of numbers in the list is " + parallelized_list.reduce(_ + _).toFloat / parallelized_list.count())
        val even_numbrs_rdd = parallelized_list.filter(x => x%2 == 0)
        println("Sum of all even numbers in the list is " + even_numbrs_rdd.reduce(_ + _))
    }
}
```

The Run console shows the following output:

```
18/02/17 19:40:17 INFO TaskSetManager: Finished task 2.0 in stage 4.0 (TID 18) in 19 ms on localhost (2/4)
18/02/17 19:40:17 INFO TaskSetManager: Finished task 1.0 in stage 4.0 (TID 17) in 21 ms on localhost (3/4)
18/02/17 19:40:17 INFO TaskSetManager: Finished task 3.0 in stage 4.0 (TID 19) in 19 ms on localhost (4/4)
18/02/17 19:40:17 INFO TaskSchedulerImpl: Removed TaskSet 4.0, whose tasks have all completed, from pool
18/02/17 19:40:17 INFO DAGScheduler: ResultStage 4 (reduce at SampleSparkDemo.scala:17) finished in 0.027 s
Sum of all even numbers in the list is 30
18/02/17 19:40:17 INFO DAGScheduler: Job 4 finished: reduce at SampleSparkDemo.scala:17, took 0.051206 s
18/02/17 19:40:17 INFO SparkContext: Invoking stop() from shutdown hook
18/02/17 19:40:17 INFO SparkUI: Stopped Spark web UI at http://192.168.43.50:4040
18/02/17 19:40:17 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/02/17 19:40:17 INFO MemoryStore: MemoryStore cleared
18/02/17 19:40:17 INFO BlockManager: BlockManager stopped
18/02/17 19:40:17 INFO BlockManagerMaster: BlockManagerMaster stopped
18/02/17 19:40:17 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
18/02/17 19:40:17 INFO SparkContext: Successfully stopped SparkContext
18/02/17 19:40:17 INFO ShutdownHookManager: Shutdown hook called
```

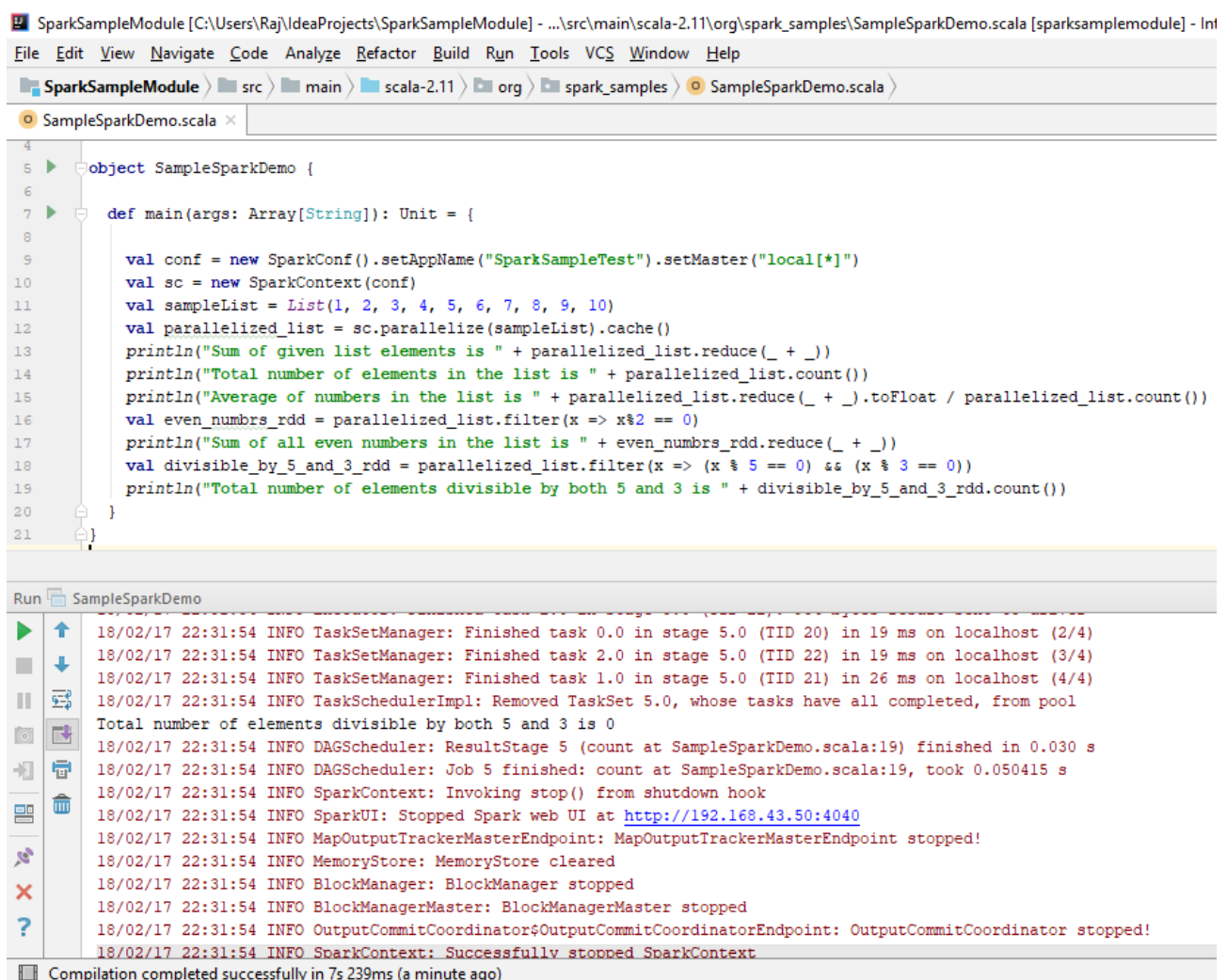
Compilation completed successfully in 3s 990ms (moments ago)

## Output:

Sum of all even numbers in the list is 30

5) Total number of elements in the list divisible by both 5 and 3

```
import org.apache.spark.{SparkConf,SparkContext}
object SampleSparkDemo {
    def main(args: Array[String]): Unit = {
        val conf = new SparkConf().setAppName("SparkSampleTest").setMaster("local[*]")
        val sc = new SparkContext(conf)
        val sampleList = List(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
        val parallelized_list = sc.parallelize(sampleList)
        val divisible_by_5_and_3_rdd = parallelized_list.filter(x => (x % 5 == 0) && (x % 3 == 0))
        println("Total number of elements divisible by both 5 and 3 is " +
            divisible_by_5_and_3_rdd.count())
    }
}
```



SparkSampleModule [C:\Users\Raj\IdeaProjects\SparkSampleModule] - ...src\main\scala-2.11\org\spark\_samples\SampleSparkDemo.scala [sparksamplemodule] - Int

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

SparkSampleModule > src > main > scala-2.11 > org > spark\_samples > SampleSparkDemo.scala

SampleSparkDemo.scala

```
4
5 object SampleSparkDemo {
6
7 def main(args: Array[String]): Unit = {
8
9     val conf = new SparkConf().setAppName("SparkSampleTest").setMaster("local[*]")
10    val sc = new SparkContext(conf)
11    val sampleList = List(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
12    val parallelized_list = sc.parallelize(sampleList).cache()
13    println("Sum of given list elements is " + parallelized_list.reduce(_ + _))
14    println("Total number of elements in the list is " + parallelized_list.count())
15    println("Average of numbers in the list is " + parallelized_list.reduce(_ + _).toFloat / parallelized_list.count())
16    val even_numbrs_rdd = parallelized_list.filter(x => x%2 == 0)
17    println("Sum of all even numbers in the list is " + even_numbrs_rdd.reduce(_ + _))
18    val divisible_by_5_and_3_rdd = parallelized_list.filter(x => (x % 5 == 0) && (x % 3 == 0))
19    println("Total number of elements divisible by both 5 and 3 is " + divisible_by_5_and_3_rdd.count())
20 }
21 }
```

Run SampleSparkDemo

```
18/02/17 22:31:54 INFO TaskSetManager: Finished task 0.0 in stage 5.0 (TID 20) in 19 ms on localhost (2/4)
18/02/17 22:31:54 INFO TaskSetManager: Finished task 2.0 in stage 5.0 (TID 22) in 19 ms on localhost (3/4)
18/02/17 22:31:54 INFO TaskSetManager: Finished task 1.0 in stage 5.0 (TID 21) in 26 ms on localhost (4/4)
18/02/17 22:31:54 INFO TaskSchedulerImpl: Removed TaskSet 5.0, whose tasks have all completed, from pool
Total number of elements divisible by both 5 and 3 is 0
18/02/17 22:31:54 INFO DAGScheduler: ResultStage 5 (count at SampleSparkDemo.scala:19) finished in 0.030 s
18/02/17 22:31:54 INFO DAGScheduler: Job 5 finished: count at SampleSparkDemo.scala:19, took 0.050415 s
18/02/17 22:31:54 INFO SparkContext: Invoking stop() from shutdown hook
18/02/17 22:31:54 INFO SparkUI: Stopped Spark web UI at http://192.168.43.50:4040
18/02/17 22:31:54 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/02/17 22:31:54 INFO MemoryStore: MemoryStore cleared
18/02/17 22:31:54 INFO BlockManager: BlockManager stopped
18/02/17 22:31:54 INFO BlockManagerMaster: BlockManagerMaster stopped
18/02/17 22:31:54 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
18/02/17 22:31:54 INFO SparkContext: Successfully stopped SparkContext
```

Compilation completed successfully in 7s 239ms (a minute ago)

### Output:

Total number of elements divisible by both 5 and 3 is 0