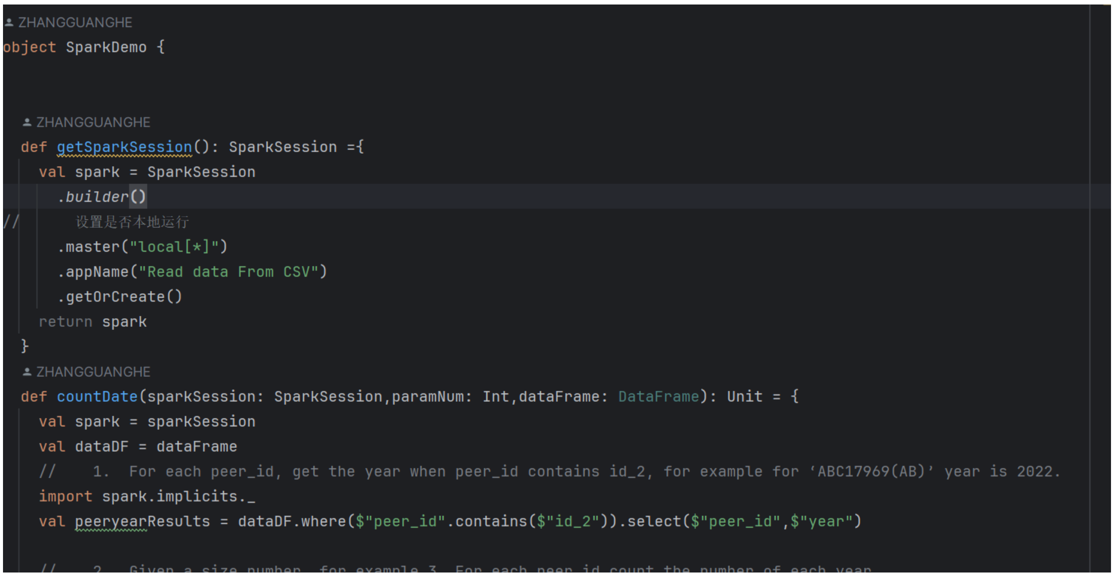
1、 主类：包含主要代码逻辑，测试数据是需求文挡的 tip 中数据，配合传入 5 和 7 参数验证。



```
ZHANGGUANGHE
object SparkDemo {


  ZHANGGUANGHE
  def getSparkSession(): SparkSession ={
    val spark = SparkSession
      .builder()
//      设置是否本地运行
      .master("local[*]")
      .appName("Read data From CSV")
      .getOrCreate()
    return spark
  }
  ZHANGGUANGHE
  def countDate(sparkSession: SparkSession,paramNum: Int,dataFrame: DataFrame): Unit = {
    val spark = sparkSession
    val dataDF = dataFrame
    //    1.  For each peer_id, get the year when peer_id contains id_2, for example for 'ABC17969(AB)' year is 2022.
    import spark.implicits._
    val peeryearResults = dataDF.where($"peer_id".contains($"id_2")).select($"peer_id",$"year")

    //    2.  Given a size number, for example 3. For each peer_id count the number of each year
```

主类：SparkDemo

函数：countDate

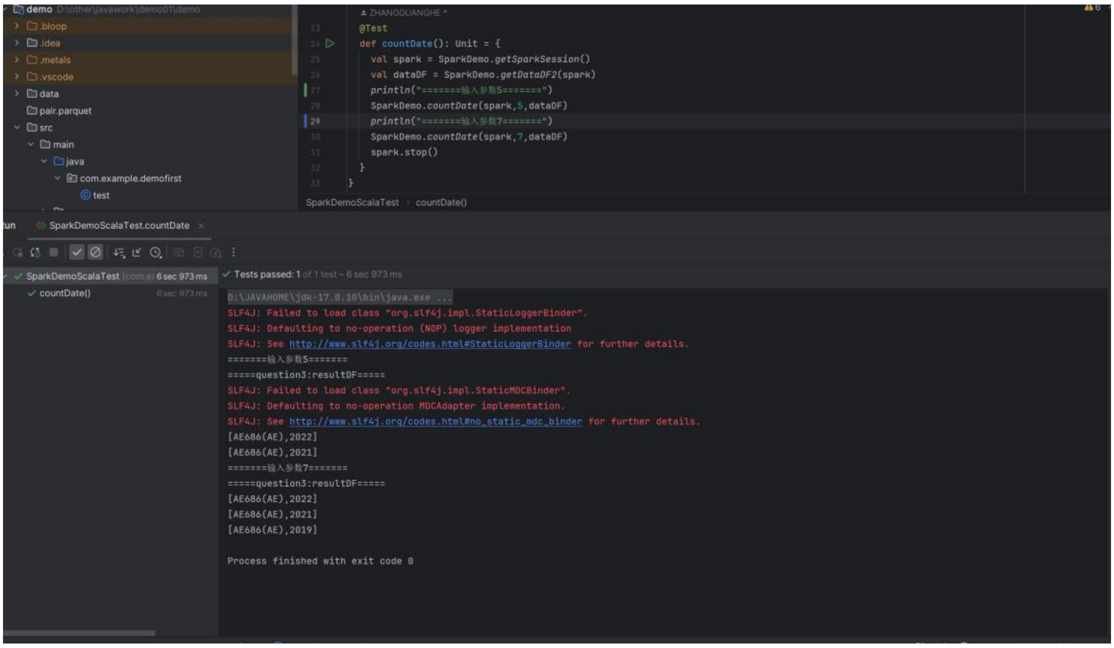
输入参数：sparkSession: SparkSession

paramNum: Int

dataFrame: DataFrame

2、 unit test 类 SparkDemoScalaTest.scala，分别测试传入参数 5 和 7，结果如下

测试结果截图，传入参数 5 和 7 的效果与 tip 中一致

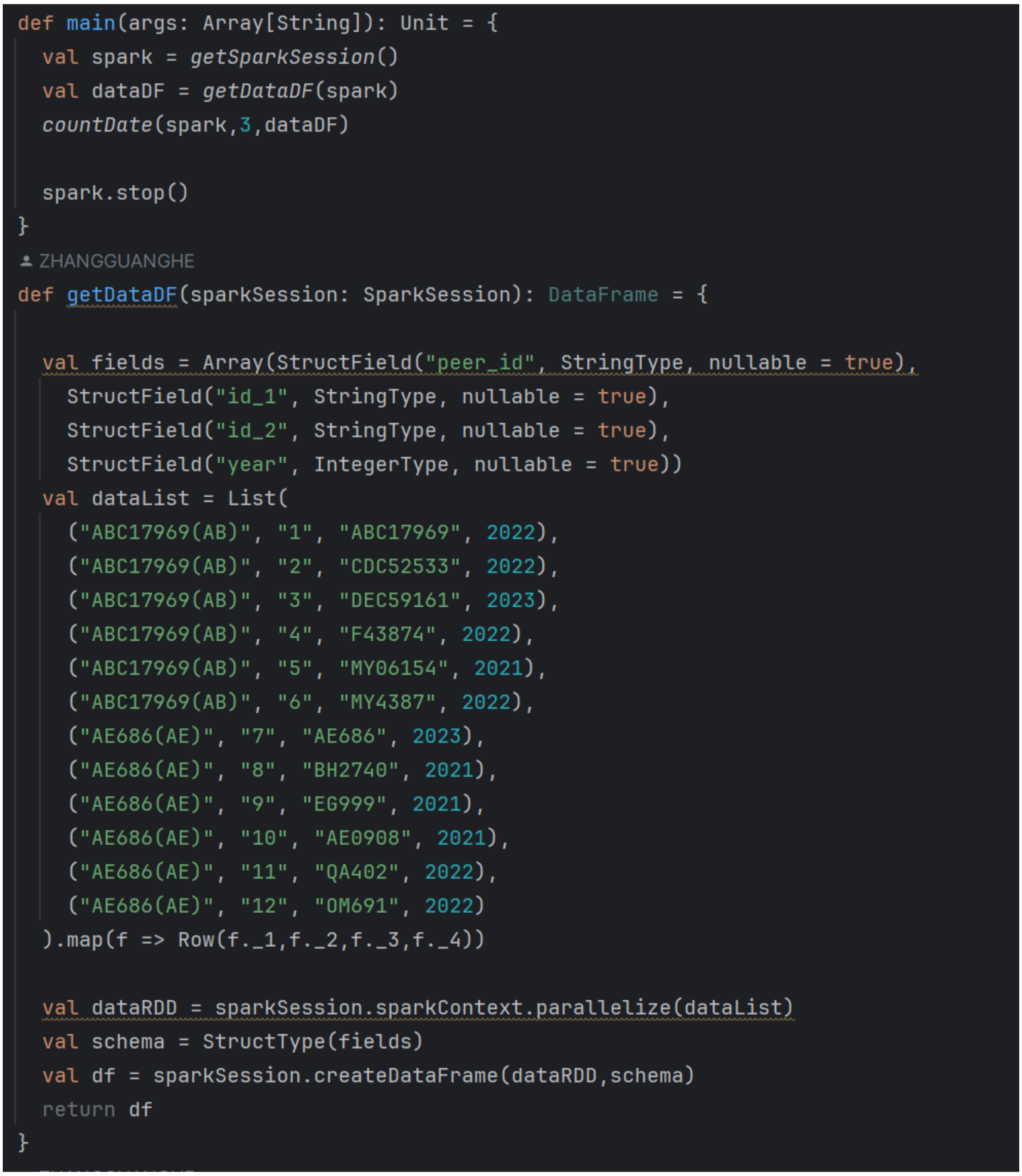

demo D:\other\javawork\demo01\demo
.bloop
.idea
.metals
.vscode
data
pair.parquet
src
main
java
com.example.demofirst
test
SparkDemoScalaTest
countDate()
Run
SparkDemoScalaTest.countDate
SparkDemoScalaTest (com.e) 6 sec 973 ms
countDate() 6 sec 973 ms
Tests passed: 1 of 1 test – 6 sec 973 ms

```
  ZHANGGUANGHE *
  @Test
  def countDate(): Unit = {
    val spark = SparkDemo.getSparkSession()
    val dataDF = SparkDemo.getDataDF2(spark)
    println("=======输入参数5=======")
    SparkDemo.countDate(spark,5,dataDF)
    println("=======输入参数7=======")
    SparkDemo.countDate(spark,7,dataDF)
    spark.stop()
  }
}
```

```
D:\JAVAHOME\jdk-17.0.10\bin\java.exe ...
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
=======输入参数5=======
=====question3:resultDF=====
SLF4J: Failed to load class "org.slf4j.impl.StaticMDCBinder".
SLF4J: Defaulting to no-operation MDCAdapter implementation.
SLF4J: See http://www.slf4j.org/codes.html#no_static_mdc_binder for further details.
[AE686(AE),2022]
[AE686(AE),2021]
=======输入参数7=======
=====question3:resultDF=====
[AE686(AE),2022]
[AE686(AE),2021]
[AE686(AE),2019]

Process finished with exit code 0
```

3、 主类运行数据 1 时结果如下：

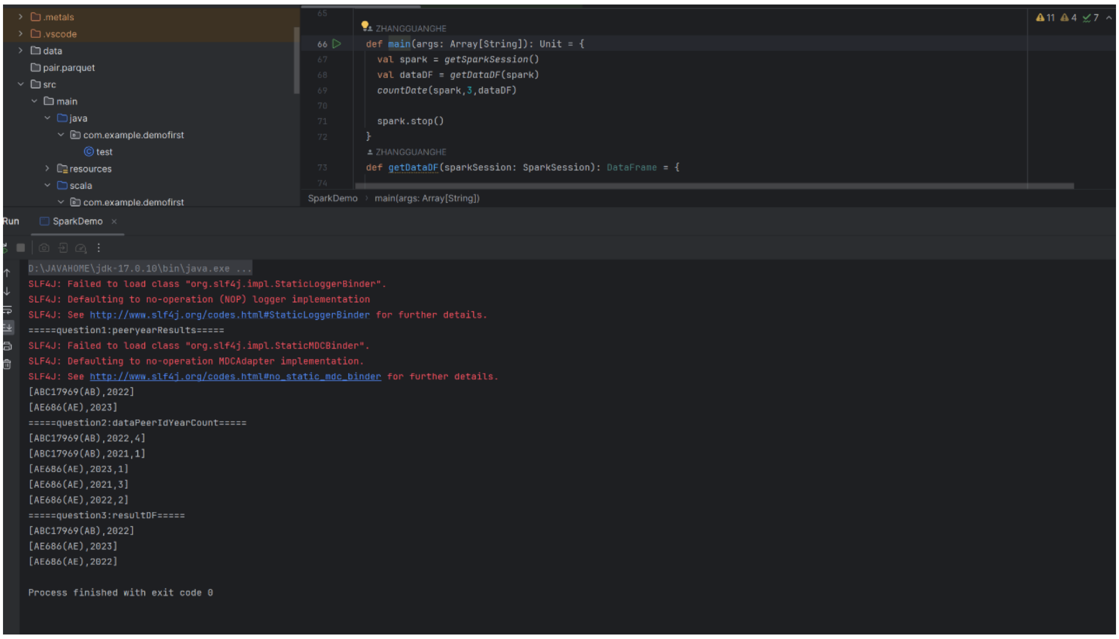


```
def main(args: Array[String]): Unit = {
  val spark = getSparkSession()
  val dataDF = getDataDF(spark)
  countDate(spark,3,dataDF)

  spark.stop()
}
def getDataDF(sparkSession: SparkSession): DataFrame = {

  val fields = Array(StructField("peer_id", StringType, nullable = true),
    StructField("id_1", StringType, nullable = true),
    StructField("id_2", StringType, nullable = true),
    StructField("year", IntegerType, nullable = true))
  val dataList = List(
    ("ABC17969(AB)", "1", "ABC17969", 2022),
    ("ABC17969(AB)", "2", "CDC52533", 2022),
    ("ABC17969(AB)", "3", "DEC59161", 2023),
    ("ABC17969(AB)", "4", "F43874", 2022),
    ("ABC17969(AB)", "5", "MY06154", 2021),
    ("ABC17969(AB)", "6", "MY4387", 2022),
    ("AE686(AE)", "7", "AE686", 2023),
    ("AE686(AE)", "8", "BH2740", 2021),
    ("AE686(AE)", "9", "EG999", 2021),
    ("AE686(AE)", "10", "AE0908", 2021),
    ("AE686(AE)", "11", "QA402", 2022),
    ("AE686(AE)", "12", "OM691", 2022)
  ).map(f => Row(f._1,f._2,f._3,f._4))

  val dataRDD = sparkSession.sparkContext.parallelize(dataList)
  val schema = StructType(fields)
  val df = sparkSession.createDataFrame(dataRDD,schema)
  return df
}
```



```scala
def main(args: Array[String]): Unit = {
  val spark = getSparkSession()
  val dataDF = getDataDF(spark)
  countDate(spark,3,dataDF)

  spark.stop()
}
def getDataDF(sparkSession: SparkSession): DataFrame = {
```

```
D:\JAVAHOME\jdk-17.0.10\bin\java.exe ...
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
=====question1:peeryearResults=====
SLF4J: Failed to load class "org.slf4j.impl.StaticMDCBinder".
SLF4J: Defaulting to no-operation MDCAdapter implementation.
SLF4J: See http://www.slf4j.org/codes.html#no_static_mdc_binder for further details.
[ABC17969(AB),2022]
[AE686(AE),2023]
=====question2:dataPeerIdYearCount=====
[ABC17969(AB),2022,4]
[ABC17969(AB),2021,1]
[AE686(AE),2023,1]
[AE686(AE),2021,3]
[AE686(AE),2022,2]
=====question3:resultDF=====
[ABC17969(AB),2022]
[AE686(AE),2023]
[AE686(AE),2022]

Process finished with exit code 0
```