Bu bölümde Apache Spark ile JSON,CSV,Avro ve Parquet Entegrasyonundan bahsedeceğiz

spark-avro-parquet-csv

JSON Entegrasyonu

Apache Spark ile json formatındaki verileri read().json metodu ile okuyabiliriz

```scala
import org.apache.spark.sql.{Encoders, SaveMode, SparkSession}

import java.util.Properties

case class Person1(name:String, age:Long)

object SparkJson {
 def main(args: Array[String]): Unit = {

    val spark = SparkSession.builder.master("local").
      appName("SparkByExample")
      .getOrCreate()
    import spark.implicits._

    val personDf = spark.read.json("/Users/serkan/Desktop/Course/Spark/data")
    personDf.printSchema()
    personDf.show()

    personDf.foreach(row =>{
      println("Age : " + row.getLong(0) + ", Name : " + row.getString(1))
    })
    println("-----------------------")
    personDf.foreach(row => {
      println("Age : " + row.getAs("age") + ", Name : " + row.getAs("name"))
    })

    val personDs = personDf.as[Person1]
    personDs.show()
    personDs.printSchema()

    personDs.foreach(row => {
      println("DS Age : " + row.age + ", Name : " + row.name)
    })
 }
}
```

Spark ile parquet formatında dosyalar yazabilir ve okuyabilirz

```scala
import org.apache.spark.sql.{Encoders, SaveMode, SparkSession}

import java.util.Properties


object SparkParquet {
 def main(args: Array[String]): Unit = {

   val spark = SparkSession.builder.master("local").
     appName("SparkByExample")
     .getOrCreate()

   val personDf =
spark.read.parquet("/Users/serkan/Desktop/Course/Spark/data/userdata1.parquet")
   personDf.printSchema()
   personDf.show()

   personDf.foreach(row => {
     println("Email : " + row.getAs("email"))
   })

 }
}
```

CSV Entegrasyonu

```scala
import org.apache.spark.sql.SparkSession


object SparkCSV {
 def main(args: Array[String]): Unit = {

   val spark = SparkSession.builder.master("local").
     appName("SparkByExample")
     .getOrCreate()

   val personDf =
spark.read.csv("/Users/serkan/Desktop/Course/Spark/data/people.csv")
   personDf.printSchema()
   personDf.show()

   personDf.foreach(row => {
     println("Name : " + row.get(0))
   })

 }
}
```

Avro entegrasyonu

```scala
import org.apache.spark.sql.SparkSession

object SparkAvro {
 def main(args: Array[String]): Unit = {

    val spark = SparkSession.builder.master("local").
      appName("SparkByExample")
      .getOrCreate()

    val twitterDf =
spark.read.format("avro").load("/Users/serkan/Desktop/Course/Spark/data/twitter.avr
o")
    twitterDf.printSchema()
    twitterDf.show()
 }
}
```