Преобразования DataFrame

I. RDD в DataFrame

1. Сначала создайте RDD
```
scala> val personRDD = sc.textFile("/opt/module/datas/people1.txt")
```

2. Просмотреть RDD
```
scala> personRDD.collect
res0: Array[String] = Array(zhangsan,1, lisi,10, wangwu,20, zhaoliu,30)
```

3. RDD=>DataFrame
```
scala> personRDD.map{x => val per = x.split(",") ; (per(0),per(1).trim.toInt)}.toDF("name","age")
res1: org.apache.spark.sql.DataFrame = [name: string, age: int]
```

4. Просмотр DF
```
scala> res1.show
+--------+---+
|    name|age|
+--------+---+
|zhangsan|  1|
|    lisi| 10|
|  wangwu| 20|
| zhaoliu| 30|
```

II. DataFrame в RDD

1. Создайте DataFrame из источника данных Spark.
```
scala> val df = spark.read.json("/opt/module/datas/people.json")
df: org.apache.spark.sql.DataFrame = [age: bigint, name: string]
```

2. DF=>RDD
```
scala> val rdd1 = df.rdd
rdd1: org.apache.spark.rdd.RDD[org.apache.spark.sql.Row] = MapPartitionsRDD[12] at rdd at
<console>:28
```

3. Просмотреть RDD
```
scala> rdd1.collect
res4: Array[org.apache.spark.sql.Row] = Array([10,Michael], [30,Andy], [20,Justin])
```

III. Преобразование RDD в DataSet

```
scala> val personRDD = sc.textFile("/opt/module/datas/people1.txt")
```

1. Напишите образец класса "Человек".
```
scala> case class Person(name:String,age:Long)
```

2. RDD=>DS
```
scala> personRDD.map{x => val per = x.split(",") ; Person(per(0),per(1).trim.toInt)}.toDS()
res9: org.apache.spark.sql.Dataset[Person] = [name: string, age: bigint]
```

3. Просмотр DS
```
scala> res9.show
+--------+---+
```

```
|   name|age|
+--------+---+
|zhangsan|  1|
|    lisi| 10|
|  wangwu| 20|
| zhaoliu| 30|
```

DataSet в RDD

1. Напишите образец класса "Человек".
```
scala> case class Person(name:String,age:Long)
```

2. Создайте DataSet.
```
scala> val ds = Seq(Person("Andy",32)).toDS()
ds: org.apache.spark.sql.Dataset[Person] = [name: string, age: bigint]
```

3. DataSet=>RDD
```
scala> val rdd2 = ds.rdd
rdd2: org.apache.spark.rdd.RDD[Person] = MapPartitionsRDD[27] at rdd at <console>:33
```

4. Запросить RDD
```
scala> rdd2.collect
res14: Array[Person] = Array(Person(Andy,32))
```

IV. Преобразование DateFrame в DataSet

1. Создайте DataFrame из источника данных Spark.
```
scala> val df = spark.read.json("/opt/module/datas/people.json")
```

2. Создайте образец класса Person.
```
scala> case class Person(name:String,age:Long)
```

3. DF=>DS
```
scala> val ds = df.as[Person]
ds: org.apache.spark.sql.Dataset[Person] = [age: bigint, name: string]
```

4. Просмотр DS
```
scala> ds.show
+---+-------+
|age|   name|
+---+-------+
| 10|Michael|
| 30|   Andy|
| 20| Justin|
```

DS=>DF
```
scala> val df2 = ds.toDF
df2: org.apache.spark.sql.DataFrame = [age: bigint, name: string]

scala> df2.show
+---+-------+
|age|   name|
+---+-------+
| 10|Michael|
| 30|   Andy|
```

| 20| Justin|
+---+-------+