TutorialKart

SAP   Salesforce   Programming   Kotlin   NodeJS   f   🐦   G   in

Learn Apache Spark

✦ Apache Spark Tutorial

✦ Install Spark on Ubuntu

✦ Install Spark on Mac OS

✦ Scala Spark Shell - Example

✦ Python Spark Shell - PySpark

✦ Setup Java Project with Spark

✦ Spark Scala Application - WordCount Example

✦ Spark Python Application

✦ Spark DAG & Physical Execution Plan

✦ Setup Spark Cluster

✦ Configure Spark Ecosystem

Spark RDD with custom class objects
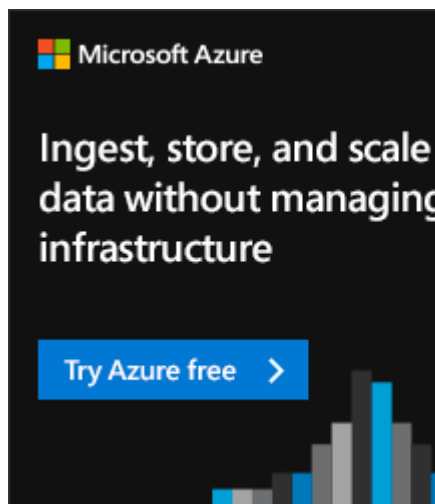
Contents

Spark RDD with custom class objects

To assign Spark RDD with custom class objects, implement the custom class with Serializable interface, create an immutable list of custom class objects, then parallelize the list with SparkContext. Parallelizing returns RDD created with custom class objects as elements.

Java Example

Following example demonstrates the creation of RDD with list of class objects.

```java
import java.io.Serializable;
import java.util.List;
import org.apache.spark.SparkConf;
import org.apache.spark.api.java.JavaRDD;
import org.apache.spark.api.java.JavaSparkContext;
import com.google.common.collect.ImmutableList;

public class CustomObjectsRDD {

    public static void main(String[] args) {
        // configure spark
        SparkConf sparkConf = new SparkConf().setAppName("Pri
                                    .setMaster("local[2]"
        // start a spark context
        JavaSparkContext sc = new JavaSparkContext(sparkConf)

        // prepare list of objects
        List<Person> personList = ImmutableList.of(
                new Person("Arjun", 25),
                new Person("Akhil", 2));

        // parallelize the list using SparkContext
        JavaRDD<Person> perJavaRDD = sc.parallelize(personLis

        for(Person person : perJavaRDD.collect()){
            System.out.println(person.name);
        }

        sc.close();
    }
}
```

```java
class Person implements Serializable{
    private static final long serialVersionUID = -26854442183
    String name;
    int age;
    public Person(String name, int age){
        this.name = name;
        this.age = age;
    }
}
```

Spark Dataseet

✦ Spark - Read JSON file to Dataset

✦ Spark - Write Dataset to JSON file

✦ Spark - Add new Column to Dataset

✦ Spark - Concatenate Datasets

Spark MLlib (Machine Learning Library)

✦ Spark MLlib Tutorial

✦ KMeans Clustering & Classification

```
18/02/10 21:59:10 INFO TaskSchedulerImpl: Removed
TaskSet 0.0, whose tasks have all completed, from pool
18/02/10 21:59:10 INFO DAGScheduler: ResultStage 0 (collect a
CustomObjectsRDD.java:29) finished in 0.223 s
18/02/10 21:59:10 INFO DAGScheduler: Job 0 finished: collect
at CustomObjectsRDD.java:29, took 0.661038 s
Arjun
Akhil
18/02/10 21:59:10 INFO SparkContext: Invoking stop() from
shutdown hook
18/02/10 21:59:10 INFO SparkUI: Stopped Spark web UI at
http://192.168.1.104:4040
18/02/10 21:59:10 INFO
MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoin
stopped!
```

Conclusion :

In this **Spark Tutorial** – **Spark RDD with custom class objects**, we have learnt to initialize RDD from an immutable list of custom objects using SparkContext.parallelize(), with the help of an Example.

‹ Previous   Next ›

📄 Download Spark RDD with custom class objects - Java Example in PDF

Most Read Articles

Apache Kafka Tutorial - Learn Scalable Kafka Messaging System

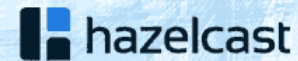Learn to use Spark Machine Learning Library (MLlib)

How to write Spark Application in Python and Submit it to Spark Cluster?

How to Configure Spark Application?

Java Tutorial from Basics with well detailed Examples

Popular Tutorials

- ➤ Salesforce Tutorial
- ➤ SAP Tutorials
- ➤ Kafka Tutorial
- ➤ Kotlin Tutorial

Interview Questions

- ➤ Salesforce Visualforce Interview Questions
- ➤ Salesforce Apex Interview Questions
- ➤ Kotlin Interview Questions

Tutorial Kart

- ➤ About Us
- ➤ Contact Us
- ➤ Careers - Write for us
- ➤ Privacy Policy
- ➤ Terms of Use
- ✎ Write Us Feedback

f   🐦   G   in