

SPONSORED SEARCHES

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](#)
- ✦ [Configure Spark Application](#)
- ✦ [Spark Cluster Managers](#)

Spark RDD

- ✦ [Spark RDD](#)
- ✦ [Spark RDD - Print Contents of RDD](#)

Spark RDD flatMap()



Contents

1 [Spark RDD flatMap\(\)](#)

1.1 [Syntax](#)

1.2 [Spark RDD](#)

[flatMap Java](#)

[Example –](#)

[String to Words](#)

1.3 [Spark RDD](#)

[flatMap Python](#)

[Example –](#)

[String to Words](#)

Search the site

- ★ [Interview Questions](#)
- ★ [Careers at TutorialKart](#)
- ★ [Contact Us](#)
- ★ [Blog / Guest Post](#)
- ★ [Feedback](#)

Spark RDD flatMap()

In this [Spark Tutorial](#), we shall learn to **flatMap** one RDD to another. Flat-Mapping is transforming each RDD element using a function that could return multiple elements to new RDD. Simple example would be applying a flatMap to Strings and using split function to return words to new RDD.

- ◆ Spark RDD - foreach
- ◆ Spark RDD - Create RDD
- ◆ Spark Parallelize
- ◆ Spark RDD - Read Text File to RDD
- ◆ Spark RDD - Read Multiple Text Files to Single RDD
- ◆ Spark RDD - Read JSON File to RDD
- ◆ Spark RDD - Containing Custom Class Objects
- ◆ Spark RDD - Map
- ◆ Spark RDD - FlatMap
- ◆ Spark RDD - Filter
- ◆ Spark RDD - Distinct
- ◆ Spark RDD - Reduce



Spark RDD flatMap



Syntax

```
RDD.flatMap(<function>)
```

SPONSORED SEARCHES



Data Storage Examples

How Can I Learn Python

where<function> is the transformation function that could return multiple elements to new RDD for each of the element of source RDD.

Spark Dataset

Spark RDD flatMap Java Example – String to Words

RDDflatMapExample.java

◆ 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

◆ Decision Tree Classification

◆ Random Forest Classification

◆ Naive Bayes Classification

◆ Logistic Regression Classification

◆ Topic Modelling

Spark SQL

◆ Spark SQL Tutorial

◆ Spark SQL - Load JSON file and execute SQL Query

Spark Others

◆ Spark Interview Questions

```
import java.util.Arrays;
```

```
import org.apache.spark.SparkConf;
import org.apache.spark.api.java.JavaRDD;
import org.apache.spark.api.java.JavaSparkContext;
```

```
public class RDDflatMapExample {
```

```
    public static void main(String[] args) {
        // configure spark
        SparkConf sparkConf = new SparkConf().setAppName("Read Text to RDD")
            .setMaster("local[2]").set("spark.executor.memory", "2g");

        // start a spark context
        JavaSparkContext sc = new JavaSparkContext(sparkConf);
```

```
        // provide path to input text file
        String path = "data/rdd/input/sample.txt";
```

```
        // read text file to RDD
        JavaRDD<String> lines = sc.textFile(path);
```

```
        // flatMap each line to words in the line
        JavaRDD<String> words = lines.flatMap(s -> Arrays.asList(s.split(" ")).iterator());
```

```
        // collect RDD for printing
        for(String word:words.collect()){
            System.out.println(word);
        }
    }
}
```

sample.txt

Welcome to TutorialKart
Learn Apache Spark
Learn to work with RDD

Java Example Output

17/11/29 12:33:59 INFO DAGScheduler: ResultStage 0 (collect at RDDflatMapExample.java:26) finished in 0.513 s
17/11/29 12:33:59 INFO DAGScheduler: Job 0 finished: collect at RDDflatMapExample.java:26, took 0.793858 s

Welcome

to

TutorialKart

Learn

Apache

SPONSORED SEARCHES

spark learning

spark machine

how to use apache spark

machine learning examples

python programming examples

Spark
Learn
to
work
with
RDD

17/11/29 12:33:59 INFO SparkContext: Invoking stop() from shutdown hook

Spark RDD flatMap Python Example – String to Words

spark-rdd-flatmap-example.py
import sys

from pyspark import SparkContext, SparkConf

if __name__ == "__main__":

```
# create Spark context with Spark configuration
conf = SparkConf().setAppName("Read Text to RDD - Python")
sc = SparkContext(conf=conf)
```

```
# read input text file to RDD
lines = sc.textFile("/home/tutorialkart/workspace/spark/sample.txt")
```

```
# flatMap each line to words
words = lines.flatMap(lambda line: line.split(" "))
```

```
# collect the RDD to a list
l1ist = words.collect()
```

```
# print the list
for line in l1ist:
    print line
```

Run the above python program using spark-submit

```
~$ spark-submit spark-rdd-flatmap-example.py
```

Python Program Output

```
17/11/29 15:15:30 INFO DAGScheduler: ResultStage 0 (collect
at /home/tutorialkart/workspace/spark/spark-rdd-flatmap-example.py:18) finished in 1.127 s
17/11/29 15:15:30 INFO DAGScheduler: Job 0 finished: collect
at /home/tutorialkart/workspace/spark/spark-rdd-flatmap-example.py:18, took 1.299076 s
```

Welcome
to
TutorialKart
Learn
Apache
Spark
Learn
to
work
with
RDD

17/11/29 15:15:30 INFO SparkContext: Invoking stop() from shutdown hook

[◀ Previous](#) [Next ▶](#)

[Download Spark RDD flatMap\(\) 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](#)