Spark by {ExampleSpathtypshttpspt/splatkbyexamptels.com/)m/)

Spark Tutorial

PySpark (https://sparkbyexamples.com/pyspark-

Spark - Installation on

Windows

(https://sparkbyexamples.com/tutorial/)

spark/apache-spark-

installation-on-windows/)

Spark - Installation on Linux | Hive (https://sparkbyexamples.com/apache-hive-

<u>Ubuntu</u>

(https://sparkbyexamples.com/

spark/spark-installation-on-

linux-ubuntu/)

tutorial/)

Get Certified in Data Science

Get the data science skills most companies are looking for. Invest in vourself today!

Spark - Cluster Setup with

Hadoop Yarn

(https://sparkbyexamples.com/

spark/spark-setup-on-hadoop-

yarn/)

HBase Shttps://sparkbyexamples.com/apache-

nested array to

hbase-tutorial/) single array column

Spark - Web/Application UI

(https://sparkbyexamples.com/

spark/spark-web-ui-

understanding/)

scala-intellij/)

Kafka (https://sparkbyexamples.com/apache-kafka-

(https://sparkbyexamples.com/author/admin/) -

Spark - Setup with Scala and

IntelliJ

Apache Spark

tutorials-with-examples/amples.com/category/spark/) /

(https://sparkbyexamples.com/ Spark SQL Functions spark/spark-setup-run-with-

(https://sparkbyexamples.com/category/spark/spark-

FAQ's sqhftmstilogsarkbyexamples.com/spark-

Spark - How to Run Examples

From this Site on IntelliJ IDEA

spark/how-to-run-spark-

examples-from-intellij/)

(https://sparkbyexamples.com/questions/) Problem: How to flatten the Array of Array or Nested Array DataFrame column into a single array column using

More Sparktys://sparkbyexamples.com/)

Spark - SparkSession

(https://sparkbyexamples.com/ spark/sparksession-explained-

with-examples/)

Spark - SparkContext

(https://sparkbyexamples.com/

spark/spark-sparkcontext/)

Master Python, Machine Learning Data Analysis, AI, Tableau with No

Intellipaat

Spark RDD Tutorial

Spark RDD - Parallelize (https://sparkbyexamples.com/ apache-spark-rdd/how-tocreate-an-rdd-usingparallelize/)

<u>Spark RDD – Read text file</u> (<u>https://sparkbyexamples.com/apache-spark-rdd/spark-read-multiple-text-files-into-a-single-rdd/</u>)</u>

<u>Spark RDD – Read CSV</u> (https://sparkbyexamples.com/ <u>apache-spark-rdd/spark-load-csv-file-into-rdd/)</u>

<u>Spark RDD – Create RDD</u> (https://sparkbyexamples.com/ apache-spark-rdd/differentways-to-create-spark-rdd/)</u>

<u>Spark RDD – Create Empty</u> RDD

(https://sparkbyexamples.com/ apache-spark-rdd/spark-howto-create-an-empty-rdd/)

<u>Spark RDD – Transformations</u> (<u>https://sparkbyexamples.com/apache-spark-rdd/spark-rdd-transformations/</u>)

<u>Spark RDD – Actions</u> (<u>https://sparkbyexamples.com/apache-spark-rdd/spark-rdd-actions/)</u>

<u>Spark RDD – Pair Functions</u> (https://sparkbyexamples.com/ apache-spark-rdd/spark-pairrdd-functions/)

<u>Spark RDD – Repartition and</u>
<u>Coalesce</u>
(https://sparkbyexamples.com/
<u>spark/spark-repartition-vs-coalesce/)</u>

<u>Spark RDD – Shuffle Partitions</u> (https://sparkbyexamples.com/ spark/spark-shuffle-partitions/)

<u>Spark RDD – Cache vs Persist</u> (https://sparkbyexamples.com/ <u>spark/spark-difference-</u> <u>between-cache-and-persist/)</u>

<u>Spark RDD – Persistance</u>
<u>Storage Levels</u>
(https://sparkbyexamples.com/spark/spark-persistence-storage-levels/)

Solution: Spark SQL provides flatten function to convert an Array of Array column (nested Array)

ArrayType(ArrayType(StringType)) to single array column on Spark

DataFrame using scala example.

Related:

- How to flatten nested Struct column (https://sparkbyexamples.com/spark/spark-flatten-nested-struct-column/)
- How to explode Array & Map columns to rows (https://sparkbyexamples.com/spark/e xplode-spark-array-and-mapdataframe-column/)

First, let's create a DataFrame (https://sparkbyexamples.com/spark/diffe rent-ways-to-create-a-spark-dataframe/) with an array column within another array column, from below example column "subjects" is an array of ArraType which holds all subjects learned.

```
val arrayArrayData = Seq(
   Row("James",List(List("Java"
   Row("Michael",List(List("Spa
   Row("Robert",List(List("CSha
))

val arrayArraySchema = new Str
   .add("subjects",ArrayType(Ar

val df = spark.createDataFrame
   spark.sparkContext.parallel
   df.printSchema()
   df.show()
```

df.printSchema() and df.show() returns the following schema and table.

<u>Spark RDD – Broadcast</u>
<u>Variables</u>
(https://sparkbyexamples.com/
<u>spark/spark-broadcast-</u>
variables/)

<u>Spark RDD – Accumulator</u>
<u>Variables</u>
(https://sparkbyexamples.com/spark/spark-accumulators/)

<u>Spark RDD – Convert RDD to</u>
<u>DataFrame</u>
(https://sparkbyexamples.com/apache-spark-rdd/convert-spark-rdd-to-dataframe-dataset/)

Spark SQL Tutorial

Spark SQL - Create

DataFrame
(https://sparkbyexamples.com/
spark/different-ways-to-createa-spark-dataframe/)

<u>Spark SQL – Select Columns</u> (<u>https://sparkbyexamples.com/spark/spark-select-columns-from-dataframe/</u>)

<u>Spark SQL – Add and Update</u>
<u>Column (withColumn)</u>
(https://sparkbyexamples.com/
spark/spark-dataframewithcolumn/)

<u>Spark SQL – Rename Nested</u> <u>Column</u> (https://sparkbyexamples.com/

(https://sparkbyexamples.com/ spark/rename-a-column-onspark-dataframes/)

<u>Spark SQL – Drop column</u> (<u>https://sparkbyexamples.com/spark/spark-drop-column-from-dataframe-dataset/)</u>

<u>Spark SQL – Where | Filter</u> (<u>https://sparkbyexamples.com/spark/spark-dataframe-where-filter/</u>)

<u>Spark SQL – When Otherwise</u> (https://sparkbyexamples.com/ <u>spark/spark-case-when-otherwise-example/)</u> Master Python, Machine Learning Data Analysis, Al, Tableau with No

Intellipaat

Flatten - Nested array to single array

Flatten – Creates a single array from an array of arrays (nested array). If a structure of nested arrays is deeper than two levels then only one level of nesting is removed. below snippet convert "subjects" column to a single array.

```
Syntax : flatten(e: Column): Col

df.select($"name",flatten($"subj
```

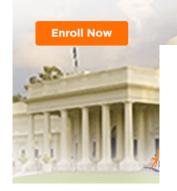




Advanced Certification in

Cloud Computing & DevOps

- Learn from IIT Roorkee
 Faculty & Industry Experts
- · Live Classes 7 Month Program



<u>Spark SQL – Collect data to</u>
<u>Driver</u>
(https://sparkbyexamples.com/spark/spark-dataframe-collect/)

<u>Spark SQL – Distinct</u> (https://sparkbyexamples.com/ <u>spark/spark-remove-duplicate-rows/)</u>

<u>Spark SQL- Pivot Table</u>
<u>DataFrame</u>
(https://sparkbyexamples.com/
spark/how-to-pivot-table-andunpivot-a-spark-dataframe/)

<u>Spark SQL – Data Types</u> (https://sparkbyexamples.com/ <u>spark/spark-sql-dataframe-data-types/</u>)

<u>Spark SQL – StructType |</u>
<u>StructField</u>
(https://sparkbyexamples.com/
<u>spark/spark-sql-structtype-on-</u>
dataframe/)

<u>Spark SQL – Schema</u> (https://sparkbyexamples.com/ spark/spark-schema-explainedwith-examples/)

<u>Spark SQL – Groupby</u> (https://sparkbyexamples.com/ <u>spark/using-groupby-on-</u> dataframe/)

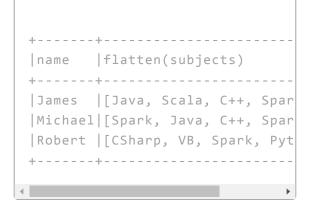
<u>Spark SQL – Sort DataFrame</u> (https://sparkbyexamples.com/ <u>spark/spark-how-to-sort-dataframe-column-explained/)</u>

<u>Spark SQL – Join Types</u> (<u>https://sparkbyexamples.com/spark/spark-sql-dataframe-join/)</u>

<u>Spark SQL – Union and</u>
<u>UnionAll</u>
(https://sparkbyexamples.com/
spark/spark-dataframe-unionand-union-all/)

<u>Spark SQL - map() vs</u> <u>mapPartitions()</u> (<u>https://sparkbyexamples.com/spark/spark-map-vs-mappartitions-transformation/)</u>

Outputs:



Complete Spark Flatten
Nested Array Example
(flatten function)

IIT Madras Data Science & Al

Learn from IIT
Madras Professors
& Industry
Practitioners.
Advance Your
Career Now!

Intellipaat

Open

<u>Spark SQL - foreach() vs</u> <u>foreachPartition()</u> (<u>https://sparkbyexamples.com/spark/spark-foreachpartition-vs-foreach-explained/)</u>

Spark SQL - map() vs flatMap() (https://sparkbyexamples.com/ spark/spark-map-vs-flatmapwith-examples/)

<u>Spark SQL – Persist and</u>
<u>Cache</u>
(https://sparkbyexamples.com/spark/spark-dataframe-cache-and-persist-explained/)

<u>Spark SQL – UDF (User Defined Functions)</u>
(https://sparkbyexamples.com/spark/spark-sql-udf/)

<u>Spark SQL – Array (ArrayType)</u>
<u>Column</u>
(https://sparkbyexamples.com/spark/spark-array-arraytype-dataframe-column/)

<u>Spark SQL – Map (MapType)</u> <u>column</u> (https://sparkbyexamples.com/ <u>spark/spark-dataframe-map-maptype-column/)</u>

<u>Spark SQL – Flatten Nested</u>
<u>Struct Column</u>
(https://sparkbyexamples.com/spark/spark-flatten-nested-struct-column/)

Spark SQL – Flatten Nested
Array Column
(https://sparkbyexamples.com/
spark/spark-flatten-nestedarray-column-to-singlecolumn/)

<u>Spark SQL – Explode Array & Map Columns</u>
(https://sparkbyexamples.com/spark/explode-spark-array-and-map-dataframe-column/)

<u>Spark SQL - Sampling</u> (https://sparkbyexamples.com/ <u>spark/spark-sampling-with-examples/)</u>

```
package com.sparkbyexamples.spar
import org.apache.spark.sql.{Row
import org.apache.spark.sql.func
import org.apache.spark.sql.type
object ArrayOfArrayType extends
 val spark = SparkSession.build
    .master("local[1]")
    .getOrCreate()
 val arrayArrayData = Seq(
    Row("James", List(List("Java"
    Row("Michael",List(List("Spa
    Row("Robert", List(List("CSha
 val arrayArraySchema = new Str
    .add("subjects", ArrayType(Ar
 val df = spark.createDataFrame
    spark.sparkContext.paralleli
 df.printSchema()
  df.show(false)
 //Convert Array of Array into
 df.select($"name",flatten($"su
```

Conclusion

In this article, you have learned how to defined nested array using StructType and how to flatten the nested array to a single array using Spark Flatten function and Scala example.

Happy Learning !!

Share this:

Advance Certification Program

Learn from IIT
Madras Professors
& Industry
Practitioners.
Advance Your
Career Now!

Intellipaat

Open

<u>Spark SQL - Partitioning</u> (https://sparkbyexamples.com/ <u>spark/spark-partitioning-understanding/)</u>

Spark SQL Functions

<u>Spark SQL String Functions</u> (https://sparkbyexamples.com/ <u>spark/usage-of-spark-sql-</u> <u>string-functions/)</u>

Spark SQL Date and
Timestamp Functions
(https://sparkbyexamples.com/
spark/spark-sql-date-and-timefunctions/)

<u>Spark SQL Array Functions</u> (https://sparkbyexamples.com/ spark/spark-sql-arrayfunctions/)

<u>Spark SQL Map Functions</u> (<u>https://sparkbyexamples.com/spark/spark-sql-map-functions/</u>)

<u>Spark SQL Sort Functions</u> (<u>https://sparkbyexamples.com/spark/spark-sql-sort-functions/</u>)

Spark SQL Aggregate
Functions
(https://sparkbyexamples.com/
spark/spark-sql-aggregatefunctions/)

<u>Spark SQL Window Functions</u> (<u>https://sparkbyexamples.com/spark/spark-sql-window-functions/)</u>

<u>Spark SQL JSON Functions</u> (<u>https://sparkbyexamples.com/spark/spark-most-used-json-functions-with-examples/</u>)

Spark Data Source API

<u>Spark – Read & Write CSV file</u> (https://sparkbyexamples.com/ <u>spark/spark-read-csv-file-into-dataframe/</u>)

<u>Spark – Read and Write JSON</u> <u>file</u> <u>(https://sparkbyexamples.com/</u>

- (https://sparkbyexamples.com/spark/sparkflatten-nested-array-column-to-single-column/? share=facebook&nb=1)
- (https://sparkbyexamples.com/spark/spark-flatten-nested-array-column-to-single-column/? share=reddit&nb=1)
- $oldsymbol{\phi}$ (https://sparkbyexamples.com/spark/sparkflatten-nested-array-column-to-single-column/? share=pinterest&nb=1)
- t (https://sparkbyexamples.com/spark/spark-flatten-nested-array-column-to-single-column/? share=tumblr&nb=1)
- (https://sparkbyexamples.com/spark/spark-flatten-nested-array-column-to-single-column/?share=pocket&nb=1)
- (https://sparkbyexamples.com/spark/spark-flatten-nested-array-column-to-single-column/?share=linkedin&nb=1)
- (https://sparkbyexamples.com/spark/spark-flatten-nested-array-column-to-single-column/?share=twitter&nb=1)

TAGS: ARRAYTYPE

(HTTPS://SPARKBYEXAMPLES.COM/TAG/ARRAYTYP E/), FLATTEN

(HTTPS://SPARKBYEXAMPLES.COM/TAG/FLATTEN/)

Advance Certification Program

Learn from IIT
Madras Professors
& Industry
Practitioners.
Advance Your
Career Now!

Intellipaat

Open



NNK

(Https://Sparkbyexamples.Com/Author/Admin/)

(https://sp arkbyexa mples.co m/author/ admin/) SparkByExamples.com is a Big Data and Spark examples community page, all examples are simple and easy to understand and well tested in our development environment Read more ..

(https://sparkbyexamples.com/about-sparkbyexamples/)

Leave a Reply

<u>spark/spark-read-and-write-</u> <u>ison-file/)</u>

<u>Spark – Read & Write Parquet</u> <u>file</u> (<u>https://sparkbyexamples.com/spark/spark-read-write-</u>

dataframe-parquet-example/)

<u>Spark – Read & Write XML file</u> (https://sparkbyexamples.com/ spark/spark-read-write-xml/)

<u>Spark – Read & Write Avro</u> <u>files</u> (<u>https://sparkbyexamples.com/spark/read-write-avro-file-spark-dataframe/</u>)

Spark – Read & Write Avro files (Spark version 2.3.x or earlier) (https://sparkbyexamples.com/ spark/using-avro-data-filesfrom-spark-sql-2-3-x/)

<u>Spark – Read & Write HBase</u> <u>using "hbase-spark" Connector</u> (https://sparkbyexamples.com/ <u>spark/spark-read-write-using-hbase-spark-connector/)</u>

<u>Spark – Read & Write from</u>
<u>HBase using Hortonworks</u>
(https://sparkbyexamples.com/spark/create-spark-dataframe-from-hbase-using-hortonworks/)

<u>Spark – Read & Write ORC file</u> (<u>https://sparkbyexamples.com/spark/spark-read-orc-file-into-dataframe/)</u>

<u>Spark – Read Binary File</u> (https://sparkbyexamples.com/ <u>spark/spark-read-binary-file-</u> into-dataframe/)

Spark Streaming & Kafka

<u>Spark Streaming –</u>
<u>OutputModes</u>
(https://sparkbyexamples.com/
<u>spark/spark-streaming-</u>
<u>outputmode/)</u>

Invalid request signature.

Advance Certification Program

Learn from IIT
Madras Professors
& Industry
Practitioners.
Advance Your
Career Now!

Intellipaat

Open

IIT Madras Data Science & Al

Learn from IIT Madras Professors & Industry Practitioners. Advance Your Career Now!

<u>Spark Streaming – Reading</u> <u>Files From Directory</u> (https://sparkbyexamples.com/ <u>spark/spark-streaming-read-</u> <u>json-files-from-directory/)</u>

<u>Spark Streaming – Reading</u>
<u>Data From TCP Socket</u>
(https://sparkbyexamples.com/spark/spark-streaming-from-tcp-socket/)

<u>Spark Streaming – Processing</u>
<u>Kafka Messages in JSON</u>
<u>Format</u>
(https://sparkbyexamples.com/spark/spark-streaming-with-kafka/)

Spark Streaming – Processing Kafka messages in AVRO Format (https://sparkbyexamples.com/ spark/spark-streamingconsume-and-produce-kafkamessages-in-avro-format/)

<u>Spark SQL Batch – Consume & Produce Kafka Message</u>
(https://sparkbyexamples.com/
<u>spark/spark-batch-processing-produce-consume-kafka-topic/)</u>

Advance Certification Program

Learn from IIT
Madras Professors
& Industry
Practitioners.
Advance Your
Career Now!

Intellipaat

Open

Advance Certification Program

Learn from IIT
Madras Professors
& Industry
Practitioners.
Advance Your
Career Now!

Intellipaat

Open

Categories

Apache Hadoop (https://sparkbyexamples.com/categor y/hadoop/)

Apache Spark (https://sparkbyexamples.com/categor y/spark/)

Apache Spark Streaming (https://sparkbyexamples.com/category/spark/apache-spark-streaming/)

Apache Kafka (https://sparkbyexamples.com/categor y/kafka/)

Apache HBase (https://sparkbyexamples.com/categor y/hbase/)

Apache Cassandra (https://sparkbyexamples.com/categor y/cassandra/)

Recent Posts

Spark regexp_replace() – Replace String Value

(https://sparkbyexamples.com/spark/spark-regexp_replace-replace-string-value/)

How to Run a PySpark Script from Python?

(https://sparkbyexamples.com/pyspark/ru n-pyspark-script-from-pythonsubprocess/)

Spark SQL like() Using Wildcard Example

(https://sparkbyexamples.com/spark/spark-sql-like-using-wildcard-example/)

Spark isin() & IS NOT IN Operator Example

(https://sparkbyexamples.com/spark/spark-isin-is-not-in-operator-example/)

Spark – Get Size/Length of Array & Map Column

(https://sparkbyexamples.com/spark/spar k-get-size-length-of-array-map-column/) About SparkByExamples.Com

SparkByExamples.com is a Big Data and Spark examples community page, all examples are simple and easy to understand, and well tested in our development environment Read more .. (https://sparkbyexamples.com/about-sparkbyexamples/)

Follow Us

Snowflake Database (https://sparkbyexamples.com/category/snowflake/)

H2O Sparkling Water (https://sparkbyexamples.com/categor y/h2o-sparkling-water/)

PySpark (https://sparkbyexamples.com/categor y/pyspark/) Spark Using Length/Size Of a
DataFrame Column
(https://sparkbyexamples.com/spark/spark-using-length-size-of-a-dataframe-column/)

Spark rlike() Working with Regex Matching Examples (https://sparkbyexamples.com/spark/spar k-rlike-regex-matching-examples/)

Spark Check String Column Has Numeric Values (https://sparkbyexamples.com/spark/spar k-check-string-column-has-numericvalues/)

Spark Check Column Data Type is Integer or String (https://sparkbyexamples.com/spark/spark-check-column-data-type-is-integer-or-string/) (https:

//www. //www.

| (https: | facebo | linkedi | (https: |
|----------|--------|---------|---------|
| //twitte | ok.co | n.com/ | //githu |
| r.com/ | m/spar | in/n- | b.com/ |
| sparkb | kbyex | nk- | spark- |
| yexam | ample | b860a | examp |
| ples) | s/) | 8193/) | les/) |

Copyright sparkbyexamples.com