

Spark by {Examples} (https://sparkbyexamples.com/)

PySpark Tutorial

[PySpark Tutorial For Beginners \(https://sparkbyexamples.com/pyspark-tutorial/\)](https://sparkbyexamples.com/pyspark-tutorial/)

[PySpark – Features \(https://sparkbyexamples.com/pyspark-tutorial/#features\)](https://sparkbyexamples.com/pyspark-tutorial/#features)

[PySpark – Advantages \(https://sparkbyexamples.com/pyspark-tutorial/#advantages\)](https://sparkbyexamples.com/pyspark-tutorial/#advantages)

[PySpark – Modules & Packages \(https://sparkbyexamples.com/pyspark-tutorial/#modules-packages\)](https://sparkbyexamples.com/pyspark-tutorial/#modules-packages)

[PySpark – Cluster Managers \(https://sparkbyexamples.com/pyspark-tutorial/#cluster-manager\)](https://sparkbyexamples.com/pyspark-tutorial/#cluster-manager)

[PySpark – Install on Windows \(https://sparkbyexamples.com/pyspark-tutorial/#pyspark-installation\)](https://sparkbyexamples.com/pyspark-tutorial/#pyspark-installation)

[PySpark – Web/Application UI \(https://sparkbyexamples.com/spark/spark-web-ui-understanding/\)](https://sparkbyexamples.com/spark/spark-web-ui-understanding/)

[PySpark – SparkSession \(https://sparkbyexamples.com/pyspark/pyspark-what-is-sparksession/\)](https://sparkbyexamples.com/pyspark/pyspark-what-is-sparksession/)

[PySpark – RDD \(https://sparkbyexamples.com/pyspark-rdd\)](https://sparkbyexamples.com/pyspark-rdd)

[PySpark – Parallelize \(https://sparkbyexamples.com/pyspark/pyspark-parallelize-create-rdd/\)](https://sparkbyexamples.com/pyspark/pyspark-parallelize-create-rdd/)

[PySpark – repartition\(\) vs coalesce\(\) \(https://sparkbyexamples.com/pyspark/pyspark-repartition-vs-coalesce/\)](https://sparkbyexamples.com/pyspark/pyspark-repartition-vs-coalesce/)

[PySpark – Broadcast Variables \(https://sparkbyexamples.com/pyspark/pyspark-broadcast-variables/\)](https://sparkbyexamples.com/pyspark/pyspark-broadcast-variables/)

[PySpark \(https://sparkbyexamples.com/pyspark-tutorial/\)](https://sparkbyexamples.com/pyspark-tutorial/)

[Hive \(https://sparkbyexamples.com/apache-hive-tutorial/\)](https://sparkbyexamples.com/apache-hive-tutorial/)

[HBase \(https://sparkbyexamples.com/apache-hbase-tutorial/\)](https://sparkbyexamples.com/apache-hbase-tutorial/)

[Kafka \(https://sparkbyexamples.com/apache-kafka-tutorials-with-examples/\)](https://sparkbyexamples.com/apache-kafka-tutorials-with-examples/)

[FAQ's \(https://sparkbyexamples.com/spark-faq-questions/\)](https://sparkbyexamples.com/spark-faq-questions/)

[More \(https://sparkbyexamples.com/more/\)](https://sparkbyexamples.com/more/)

 [NNK \(https://sparkbyexamples.com/author/admin/\)](https://sparkbyexamples.com/author/admin/)

 [PySpark \(https://sparkbyexamples.com/category/pyspark/\)](https://sparkbyexamples.com/category/pyspark/)

In PySpark, `toDF()` function of the RDD is used to convert RDD to DataFrame. We would need to convert RDD to DataFrame as DataFrame provides more advantages over RDD. For instance, DataFrame is a distributed collection of data organized into named columns similar to Database tables and provides optimization and performance improvements.



- [Create PySpark RDD](#)
- [Convert PySpark RDD to DataFrame](#)
 - [using toDF\(\)](#) with it.
 - [using createDataFrame\(\)](#)
 - [using RDD row type & schema](#)



[PySpark – Accumulator
\(https://sparkbyexamples.com/pyspark/pyspark-accumulator-with-example/\)](https://sparkbyexamples.com/pyspark/pyspark-accumulator-with-example/)

PySpark DataFrame

[PySpark – Create a DataFrame
\(https://sparkbyexamples.com/pyspark/different-ways-to-create-dataframe-in-pyspark/\)](https://sparkbyexamples.com/pyspark/different-ways-to-create-dataframe-in-pyspark/)

[PySpark – Create an empty DataFrame
\(https://sparkbyexamples.com/pyspark/pyspark-create-an-empty-dataframe/\)](https://sparkbyexamples.com/pyspark/pyspark-create-an-empty-dataframe/)

[PySpark – Convert RDD to DataFrame
\(https://sparkbyexamples.com/pyspark/convert-pyspark-rdd-to-dataframe/\)](https://sparkbyexamples.com/pyspark/convert-pyspark-rdd-to-dataframe/)

[PySpark – Convert DataFrame to Pandas
\(https://sparkbyexamples.com/pyspark/convert-pyspark-dataframe-to-pandas/\)](https://sparkbyexamples.com/pyspark/convert-pyspark-dataframe-to-pandas/)

[PySpark – show\(\)
\(https://sparkbyexamples.com/pyspark/pyspark-show-display-dataframe-contents-in-table/\)](https://sparkbyexamples.com/pyspark/pyspark-show-display-dataframe-contents-in-table/)

[PySpark – StructType & StructField
\(https://sparkbyexamples.com/pyspark/pyspark-structtype-and-structfield/\)](https://sparkbyexamples.com/pyspark/pyspark-structtype-and-structfield/)

[PySpark – Row Class
\(https://sparkbyexamples.com/pyspark/pyspark-row-using-rdd-dataframe/\)](https://sparkbyexamples.com/pyspark/pyspark-row-using-rdd-dataframe/)

[PySpark – Column Class
\(https://sparkbyexamples.com/pyspark/pyspark-column-functions/\)](https://sparkbyexamples.com/pyspark/pyspark-column-functions/)

[PySpark – select\(\)
\(https://sparkbyexamples.com/pyspark/select-columns-from-pyspark-dataframe/\)](https://sparkbyexamples.com/pyspark/select-columns-from-pyspark-dataframe/)

[PySpark – collect\(\)
\(https://sparkbyexamples.com/pyspark/pyspark-collect/\)](https://sparkbyexamples.com/pyspark/pyspark-collect/)

[PySpark – withColumn\(\)
\(https://sparkbyexamples.com/pyspark/pyspark-withcolumn/\)](https://sparkbyexamples.com/pyspark/pyspark-withcolumn/)

1. Create PySpark RDD

First, let's create an RDD by passing Python list object to `sparkContext.parallelize()` function. We would need this rdd object for all our examples below.

In PySpark, when you have data in a list meaning you have a collection of data in a PySpark driver memory when you create an RDD, this collection is going to be parallelized (<https://sparkbyexamples.com/pyspark/pyspark-parallelize-create-rdd-from-collection/>).

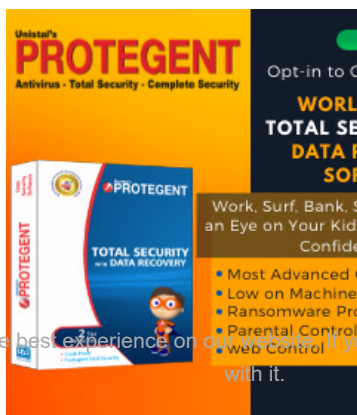
```
from pyspark.sql import SparkSession
spark = SparkSession.builder.appName("Example").getOrCreate()
dept = [("Finance",10),("Marketing",20)]
rdd = spark.sparkContext.parallelize(dept)
```

2. Convert PySpark RDD to DataFrame

Converting PySpark RDD to DataFrame can be done using `toDF()`, `createDataFrame()`. In this section, I will explain these two methods.

2.1 Using `rdd.toDF()` function

PySpark provides `toDF()` function in RDD which can be used to convert RDD into DataFrame



Accept



[PySpark – withColumnRenamed\(\).](#)
(<https://sparkbyexamples.com/pyspark/pyspark-rename-dataframe-column/>).

[PySpark – where\(\) & filter\(\).](#)
(<https://sparkbyexamples.com/pyspark/pyspark-where-filter/>).

[PySpark – drop\(\) & dropDuplicates\(\).](#)
(<https://sparkbyexamples.com/pyspark/pyspark-distinct-to-drop-duplicates/>).

[PySpark – orderBy\(\) and sort\(\).](#)
(<https://sparkbyexamples.com/pyspark/pyspark-orderby-and-sort-explained/>).

[PySpark – groupBy\(\).](#)
(<https://sparkbyexamples.com/pyspark/pyspark-groupby-explained-with-example/>).

[PySpark – join\(\).](#)
(<https://sparkbyexamples.com/pyspark/pyspark-join-explained-with-examples/>).

[PySpark – union\(\) & unionAll\(\).](#)
(<https://sparkbyexamples.com/pyspark/pyspark-union-and-unionall/>).

[PySpark – unionByName\(\).](#)
(<https://sparkbyexamples.com/spark/spark-merge-two-dataframes-with-different-columns/>).

[PySpark – UDF \(User Defined Function\).](#)
(<https://sparkbyexamples.com/pyspark/pyspark-udf-user-defined-function/>).

[PySpark – map\(\).](#)
(<https://sparkbyexamples.com/pyspark/pyspark-map-transformation/>).

[PySpark – flatMap\(\).](#)
(<https://sparkbyexamples.com/pyspark/pyspark-flatmap-transformation/>).

[pyspark – foreach\(\).](#)
(<https://sparkbyexamples.com/pyspark/pyspark-loop-iterate-through-rows-in-dataframe/#use-foreach-loop-through-dataframe>).

[PySpark – sample\(\) vs sampleBy\(\).](#)

```
df = rdd.toDF()
df.printSchema()
df.show(truncate=False)
```

By default, toDF() function creates column names as “_1” and “_2”. This snippet yields below schema.

```
root
 |-- _1: string (nullable = true)
 |-- _2: long (nullable = true)
```

```
+-----+-----+
|_1      |_2 |
+-----+-----+
|Finance |10 |
|Marketing|20 |
|Sales   |30 |
|IT      |40 |
+-----+-----+
```

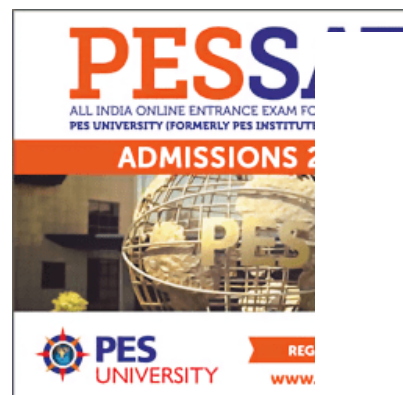
toDF() has another signature that takes arguments to define column names as shown below.

```
deptColumns = ["dept_name", "dept_id"]
df2 = rdd.toDF(deptColumns)
df2.printSchema()
df2.show(truncate=False)
```

Outputs below schema.

```
root
 |-- dept_name: string (nullable = true)
 |-- dept_id: long (nullable = true)
```

```
+-----+-----+
|dept_name|dept_id|
+-----+-----+
|Finance  |10     |
|Marketing|20     |
|Sales    |30     |
|IT       |40     |
+-----+-----+
```



[\(https://sparkbyexamples.com/pyspark/pyspark-sampling-example/\)](https://sparkbyexamples.com/pyspark/pyspark-sampling-example/).

[PySpark – fillna\(\) & fill\(\). \(https://sparkbyexamples.com/pyspark/pyspark-fillna-fill-replace-null-values/\)](https://sparkbyexamples.com/pyspark/pyspark-fillna-fill-replace-null-values/).

[PySpark – pivot\(\)_\(Row to Column\). \(https://sparkbyexamples.com/pyspark/pyspark-pivot-and-unpivot-dataframe/\)](https://sparkbyexamples.com/pyspark/pyspark-pivot-and-unpivot-dataframe/).

[PySpark – partitionBy\(\). \(https://sparkbyexamples.com/pyspark/pyspark-partitionby-example/\)](https://sparkbyexamples.com/pyspark/pyspark-partitionby-example/).

[PySpark – ArrayType Column \(Array\). \(https://sparkbyexamples.com/pyspark/pyspark-arraytype-column-with-examples/\)](https://sparkbyexamples.com/pyspark/pyspark-arraytype-column-with-examples/).

[PySpark – MapType \(Map/Dict\). \(https://sparkbyexamples.com/pyspark/pyspark-maptype-dict-examples/\)](https://sparkbyexamples.com/pyspark/pyspark-maptype-dict-examples/).

PySpark SQL Functions

[PySpark – Aggregate Functions \(https://sparkbyexamples.com/pyspark/pyspark-aggregate-functions/\)](https://sparkbyexamples.com/pyspark/pyspark-aggregate-functions/).

[PySpark – Window Functions \(https://sparkbyexamples.com/pyspark/pyspark-window-functions/\)](https://sparkbyexamples.com/pyspark/pyspark-window-functions/).

[PySpark – Date and Timestamp Functions \(https://sparkbyexamples.com/pyspark/pyspark-sql-date-and-timestamp-functions/\)](https://sparkbyexamples.com/pyspark/pyspark-sql-date-and-timestamp-functions/).

[PySpark – JSON Functions \(https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/\)](https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/).

PySpark Datasources

[PySpark – Read & Write CSV File \(https://sparkbyexamples.com/pyspark/pyspark-read-csv-file-into-dataframe/\)](https://sparkbyexamples.com/pyspark/pyspark-read-csv-file-into-dataframe/).

[PySpark – Read & Write Parquet File \(https://sparkbyexamples.com/pyspark/pyspark-read-write-parquet-file/\)](https://sparkbyexamples.com/pyspark/pyspark-read-write-parquet-file/).

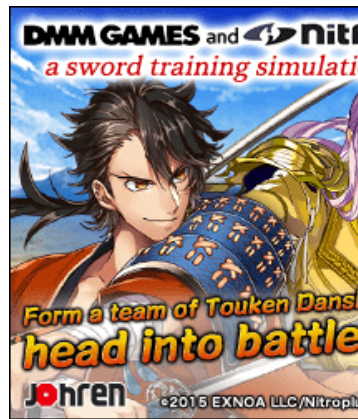
2.2 Using PySpark

createDataFrame() function

SparkSession class provides createDataFrame() method to create DataFrame and it takes rdd object as an argument.

```
deptDF = spark.createDataFrame(  
    deptDF.printSchema()  
    deptDF.show(truncate=False)
```

This yields the same output as above.



2.3 Using

createDataFrame() with StructType schema

When you infer the schema, by default the datatype of the columns is derived from the data and set's nullable to true for all columns. We can change this behavior by supplying schema (https://sparkbyexamples.com/spark/spark-schema-explained-with-examples/) using StructType (https://sparkbyexamples.com/pyspark/pyspark-structtype-and-structfield/) – where we can specify a column name, data type and nullable for each field/column.

If you wanted to know more about StructType, please go through [how to use StructType and StructField to](#)

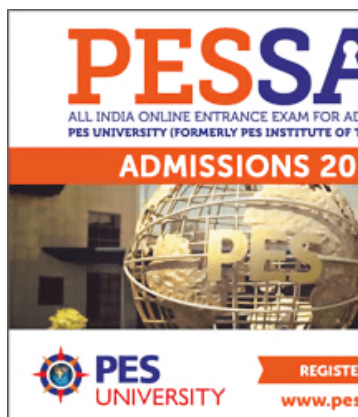
We use cookies to ensure that we give you the best experience on our website. If you continue to use this site we will assume that you are happy with it.

Accept



[pyspark/pyspark-read-and-write-parquet-file/](#)

[PySpark – Read & Write JSON file](#)
(<https://sparkbyexamples.com/pyspark/pyspark-read-json-file-into-dataframe/>)



PySpark Built-In Functions

[PySpark – when\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-when-otherwise/>)

[PySpark – expr\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-sql-expr-expression-function/>)

[PySpark – lit\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-lit-add-literal-constant/>)

[PySpark – split\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-convert-string-to-array-column/>)

[PySpark – concat_ws\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-convert-array-column-to-string-column/>)

[Pyspark – substring\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-substring-from-a-column/>)

[PySpark – translate\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-replace-column-values/#translate-replace-character-by-character>)

[PySpark – regexp_replace\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-replace-column->

[define the custom schema](#)
(<https://sparkbyexamples.com/pyspark/pyspark-structtype-and-structfield/>)

```
from pyspark.sql.types import StructType, StructField
deptSchema = StructType([
    StructField('dept_name', StringType),
    StructField('dept_id', StringType)
])
```

```
deptDF1 = spark.createDataFrame(deptSchema, deptSchema.fields)
deptDF1.printSchema()
deptDF1.show(truncate=False)
```

; also yields the same output.

3. Complete Example

```
import pyspark
from pyspark.sql import SparkSession

spark = SparkSession.builder.appName('Pyspark StructType Example').getOrCreate()

dept = [("Finance", 10), ("Marketing", 20)]
rdd = spark.sparkContext.parallelize(dept)

df = rdd.toDF()
df.printSchema()
df.show(truncate=False)

deptColumns = ["dept_name", "dept_id"]
df2 = rdd.toDF(deptColumns)
df2.printSchema()
df2.show(truncate=False)

deptDF = spark.createDataFrame(deptSchema, deptSchema.fields)
deptDF.printSchema()
deptDF.show(truncate=False)

from pyspark.sql.types import StructType, StructField
deptSchema = StructType([
    StructField('dept_name', StringType),
    StructField('dept_id', StringType)
])

deptDF1 = spark.createDataFrame(deptSchema, deptSchema.fields)
deptDF1.printSchema()
deptDF1.show(truncate=False)
```

We use cookies to ensure that we give you the best experience on our website. If you continue to use this site we will assume that you are happy

with it.

Accept



[values/#regexp_replace-replace-string-columns](#)).

[PySpark – overlay\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-replace-column-values/#overlay-function>)

[PySpark – to_timestamp\(\)](#)
(https://sparkbyexamples.com/spark/pyspark-to_timestamp-convert-string-to-timestamp-type/)

[PySpark – to_date\(\)](#)
(https://sparkbyexamples.com/pyspark/pyspark-to_date-convert-timestamp-to-date/)

[PySpark – date_format\(\)](#)
(https://sparkbyexamples.com/pyspark/pyspark-date_format-convert-date-to-string-format/)

[PySpark – datediff\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-difference-between-two-dates-days-months-years/#datediff>)

[PySpark – months_between\(\)](#)
([https://sparkbyexamples.com/pyspark/pyspark-difference-between-two-dates-days-months-years/#months_between\(\)](https://sparkbyexamples.com/pyspark/pyspark-difference-between-two-dates-days-months-years/#months_between()))

[PySpark – explode\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-explode-nested-array-into-rows/>)

[PySpark – array_contains\(\)](#)
(https://sparkbyexamples.com/pyspark/pyspark-arraytype-column-with-examples/#array_contains)

[PySpark – array\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-arraytype-column-with-examples/#array>)

[PySpark – collect_list\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-aggregate-functions/#collect-list>)

[PySpark – collect_set\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-aggregate-functions/#collect-set>)

[PySpark – create_map\(\)](#)
(<https://sparkbyexamples.com/pyspark/pyspark-convert-dataframe-columns-to-maptype-dict/>)

The complete code can be downloaded from [GitHub \(https://github.com/spark-examples/pyspark-examples/blob/master/pyspark-rdd-to-dataframe.py\)](https://github.com/spark-examples/pyspark-examples/blob/master/pyspark-rdd-to-dataframe.py)

4. Conclusion:

In this article, you have learned how to convert PySpark RDD to DataFrame, we would need these frequently while working in PySpark as these provides optimization and performance over RDD.

Happy Learning !!

Share this:



(<https://sparkbyexamples.com/pyspark/convert-pyspark-rdd-to-dataframe/?share=facebook&nb=1>)



(<https://sparkbyexamples.com/pyspark/convert-pyspark-rdd-to-dataframe/?share=reddit&nb=1>)



(<https://sparkbyexamples.com/pyspark/convert-pyspark-rdd-to-dataframe/?share=pinterest&nb=1>)



(<https://sparkbyexamples.com/pyspark/convert-pyspark-rdd-to-dataframe/?share=tumblr&nb=1>)



(<https://sparkbyexamples.com/pyspark/convert-pyspark-rdd-to-dataframe/?share=pocket&nb=1>)



(<https://sparkbyexamples.com/pyspark/convert-pyspark-rdd-to-dataframe/?share=linkedin&nb=1>)



(<https://sparkbyexamples.com/pyspark/convert-pyspark-rdd-to-dataframe/?share=twitter&nb=1>)

TAGS: DATAFRAME

([HTTPS://SPARKBYEXAMPLES.COM/TAG/DATAFRAME/](https://sparkbyexamples.com/tag/dataframe/)), **DATASET**
([HTTPS://SPARKBYEXAMPLES.COM/TAG/DATASET/](https://sparkbyexamples.com/tag/dataset/)), **RDD**
([HTTPS://SPARKBYEXAMPLES.COM/TAG/RDD/](https://sparkbyexamples.com/tag/rdd/)), **SPARK.CREATEDATAFRAME**
([HTTPS://SPARKBYEXAMPLES.COM/TAG/CREATEDATAFRAME/](https://sparkbyexamples.com/tag/createdataframe/)), **TODF()**
([HTTPS://SPARKBYEXAMPLES.COM/TAG/TODF/](https://sparkbyexamples.com/tag/todf/))

We use cookies to ensure that we give you the best experience on our website. If you continue to use this site we will assume that you are happy with it.

Accept



[PySpark – map_keys\(\).
\(https://sparkbyexamples.com/py-spark/pyspark-maptype-dict-examples/#map_keys\)](https://sparkbyexamples.com/py-spark/pyspark-maptype-dict-examples/#map_keys)

[PySpark – map_values\(\).
\(https://sparkbyexamples.com/py-spark/pyspark-maptype-dict-examples/#map_values\)](https://sparkbyexamples.com/py-spark/pyspark-maptype-dict-examples/#map_values)

[PySpark – struct\(\).
\(https://sparkbyexamples.com/py-spark/pyspark-structtype-and-structfield/#update-struct-function\)](https://sparkbyexamples.com/py-spark/pyspark-structtype-and-structfield/#update-struct-function)

[PySpark – countDistinct\(\).
\(https://sparkbyexamples.com/py-spark/pyspark-count-distinct-from-dataframe/\)](https://sparkbyexamples.com/py-spark/pyspark-count-distinct-from-dataframe/)

[PySpark – sum\(\).avg\(\).
\(https://sparkbyexamples.com/py-spark/pyspark-dataframe-groupby-and-sort-by-descending-order/\)](https://sparkbyexamples.com/py-spark/pyspark-dataframe-groupby-and-sort-by-descending-order/)

[PySpark – row_number\(\).
\(https://sparkbyexamples.com/py-spark/pyspark-window-functions/#row_number\)](https://sparkbyexamples.com/py-spark/pyspark-window-functions/#row_number)

[PySpark – rank\(\).
\(https://sparkbyexamples.com/py-spark/pyspark-window-functions/#rank\)](https://sparkbyexamples.com/py-spark/pyspark-window-functions/#rank)

[PySpark – dense_rank\(\).
\(https://sparkbyexamples.com/py-spark/pyspark-window-functions/#dense_rank\)](https://sparkbyexamples.com/py-spark/pyspark-window-functions/#dense_rank)

[PySpark – percent_rank\(\).
\(https://sparkbyexamples.com/py-spark/pyspark-window-functions/#percent_rank\)](https://sparkbyexamples.com/py-spark/pyspark-window-functions/#percent_rank)

[PySpark – typedLit\(\).
\(https://sparkbyexamples.com/py-spark/pyspark-lit-add-literal-constant/#typedlit\)](https://sparkbyexamples.com/py-spark/pyspark-lit-add-literal-constant/#typedlit)

[PySpark – from_json\(\).
\(https://sparkbyexamples.com/py-spark/pyspark-json-functions-with-examples/#from_json\)](https://sparkbyexamples.com/py-spark/pyspark-json-functions-with-examples/#from_json)

[PySpark – to_json\(\).
\(https://sparkbyexamples.com/py-spark/pyspark-json-functions-with-examples/#to_json\)](https://sparkbyexamples.com/py-spark/pyspark-json-functions-with-examples/#to_json)

[PySpark – json_tuple\(\).
\(https://sparkbyexamples.com/py-spark/pyspark-json-functions-with-examples/#json_tuple\)](https://sparkbyexamples.com/py-spark/pyspark-json-functions-with-examples/#json_tuple)



NNK

[\(https://Sparkbyexamples.Com/Author/Admin/\)](https://Sparkbyexamples.Com/Author/Admin/)

[\(https://sparkbyexamples.com/author/admin/\)](https://sparkbyexamples.com/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

Enter your comment here...

We use cookies to ensure that we give you the best experience on our website. If you continue to use this site we will assume that you are happy with it.

Accept



[PySpark – get_json_object\(\)
\(https://sparkbyexamples.com/py-spark/pyspark-json-functions-with-examples/#get_json_object\)](https://sparkbyexamples.com/py-spark/pyspark-json-functions-with-examples/#get_json_object)

[PySpark – schema_of_json\(\)
\(https://sparkbyexamples.com/py-spark/pyspark-json-functions-with-examples/#schema_of_json\)](https://sparkbyexamples.com/py-spark/pyspark-json-functions-with-examples/#schema_of_json)

Categories

- Apache Hadoop
(https://sparkbyexamples.com/category/hadoop/)
- Apache Spark
(https://sparkbyexamples.com/category/spark/)
- Apache Spark Streaming
(https://sparkbyexamples.com/category/spark/apache-spark-streaming/)
- Apache Kafka
(https://sparkbyexamples.com/category/kafka/)
- Apache HBase
(https://sparkbyexamples.com/category/hbase/)
- Apache Cassandra
(https://sparkbyexamples.com/category/cassandra/)
- Snowflake Database
(https://sparkbyexamples.com/category/snowflake/)
- H2O Sparkling Water
(https://sparkbyexamples.com/category/h2o-sparkling-water/)
- PySpark
(https://sparkbyexamples.com/category/pyspark/)


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/run-pyspark-script-from-python-subprocess/)
- 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/spark-get-size-length-of-array-map-column/)
- 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/spark-rlike-regex-matching-examples/)
- Spark Check String Column Has Numeric Values
(https://sparkbyexamples.com/spark/spark-check-string-column-has-numeric-values/)
- Spark Check Column Data Type is Integer or String
(https://sparkbyexamples.com/spark/spark-check-column-data-type-is-integer-or-string/)


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


(https://www.facebook.com/sparkbyexamples/)


(https://www.linkedin.com/company/sparkbyexamples/)


(https://twitter.com/sparkbyexamples)


(https://www.youtube.com/channel/UC8b60a38193/)


(https://github.com/sparkbyexamples)

