

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

PySpark Tutorial

PySpark Tutorial For Beginners

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

PySpark – Features

(https://sparkbyexamples.com/pyspark-tutorial/#features)

PySpark – Advantages

(https://sparkbyexamples.com/pyspark-tutorial/#advantages)

PySpark – Modules & Packages

(https://sparkbyexamples.com/pyspark-tutorial/#modules-packages)

PySpark – Cluster Managers

(https://sparkbyexamples.com/pyspark-tutorial/#cluster-manager)

PySpark – Install on Windows

(https://sparkbyexamples.com/pyspark-tutorial/#pyspark-installation)

PySpark – Web/Application UI

(https://sparkbyexamples.com/spark/spark-web-ui-understanding/)

PySpark – SparkSession

(https://sparkbyexamples.com/pyspark/pyspark-what-is-sparksession/)

PySpark – RDD

(https://sparkbyexamples.com/pyspark-rdd)

PySpark – Parallelize

(https://sparkbyexamples.com/pyspark/pyspark-parallelize-create-rdd/)

PySpark – repartition() vs coalesce()

(https://sparkbyexamples.com/pyspark/pyspark-repartition-vs-coalesce/)

PySpark – Broadcast Variables

(https://sparkbyexamples.com/pyspark/pyspark-broadcast-variables/)

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

PySpark – Loop/Iterate Through Rows in DataFrame

PySpark provides map(), mapPartitions() to loop/iterate through rows in RDD/DataFrame to perform the complex transformations, and these two returns the same number of records as in the original DataFrame but the number of columns could be different (after add/update).

Learn more

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

PySpark – Loop/Iterate Through Rows in DataFrame

PySpark provides map(), mapPartitions() to loop/iterate through rows in RDD/DataFrame to perform the complex transformations, and these two returns the same number of records as in the original DataFrame but the number of columns could be different (after add/update).

Learn more

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

PySpark

(https://sparkbyexamples.com/category/pyspark/)

FAQ's (https://sparkbyexamples.com/spark-questions/)

More (https://sparkbyexamples.com/)

[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/>).

nothing, In this article, I will explain how to use these methods to get DataFrame column values and process.

- [Using map\(\) to loop through DataFrame](#)
- [Using foreach\(\) to loop through DataFrame](#)
- [Using pandas\(\) to Iterate](#)
- [Collect Data As List and Loop Through in Python](#)

## PySpark Loop Through

### Rows in DataFrame

### Examples

In order to explain with examples, let's create a DataFrame

```
from pyspark.sql import SparkSession
spark = SparkSession.builder.appName('PySpark').getOrCreate()

data = [(('James', 'Smith', 'M', 30), ('Robert', 'Williams', 'M', 62)),
]
columns = ["firstname", "lastname", "gender", "salary"]
df = spark.createDataFrame(data, columns)
df.show()
```

firstname	lastname	gender	salary
James	Smith	M	30
Anna	Rose	F	25
Robert	Williams	M	62

Mostly for simple computations, instead of iterating through using map() and foreach(), you should use either DataFrame select() (<https://sparkbyexamples.com/pyspark/select-columns-from-pyspark-dataframe/>) or DataFrame withColumn() (<https://sparkbyexamples.com/pyspark/pyspark-withcolumn/>) in conjunction with PySpark SQL functions.



[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\(\).](#)

```
from pyspark.sql.functions import concat_ws, col, lit

df.select(concat_ws(",", df.first_name, df.last_name, df.gender, lit(df.salary)))

+-----+-----+-----+
|              name | gender | new_salary |
+-----+-----+-----+
|   James, Smith   |      M |          75000 |
|   Anna, Rose     |      F |          60000 |
| Robert, Williams |      M |          50000 |
+-----+-----+-----+
```

Below I have map() example to achieve same output as above.

Average User Earns  
**\$50/Month**

Salad.com

## Using map() to Loop Through Rows in DataFrame

[PySpark map\(\) Transformation](#) (<https://sparkbyexamples.com/pyspark/pyspark-map-transformation/>) is used to loop/iterate through the PySpark DataFrame/RDD by applying the transformation function (lambda) on every element (Rows and Columns) of RDD/DataFrame. PySpark doesn't have a map() in DataFrame instead it's in RDD hence we need to convert DataFrame to RDD first and then use the map(). It returns an RDD and you should [Convert RDD to PySpark DataFrame](#) (<https://sparkbyexamples.com/pyspark/convert-pyspark-rdd-to-dataframe/>) if needed.

If you have a heavy initialization use PySpark mapPartitions() transformation instead of map(), as with

<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/)

mapPartitions() heavy initialization executes only once for each partition instead of every record.

```
# Referring columns by index.
rdd=df.rdd.map(lambda x:
    (x[0]+", "+x[1],x[2],x[3]*2)
)
df2=rdd.toDF(["name","gender","age","score"])
df2.show()
```

The above example iterates through every row in a DataFrame by applying transformations to the data, since I need a DataFrame back, I have converted the result of RDD to DataFrame with new column names. Note that here I have used index to get the column values, alternatively, you can also refer to the DataFrame column names while iterating.

```
# Referring Column Names
rdd2=df.rdd.map(lambda x:
    (x["firstname"]+", "+x["lastname"])
)
```

Another alternative

```
# Referring Column Names
rdd2=df.rdd.map(lambda x:
    (x.firstname+", "+x.lastname)
)
```

You can also create a custom function to perform an operation. Below func1() function executes for every DataFrame row from the lambda function.

**BBA-LLB (Hons) / BA-LLB (Hons)**  
**#1 University in Karnataka**  
KSRU



**ADMISSIONS 2021 NOW OPEN**

**PESSA**  
ALL INDIA ONLINE ENTRANCE EXAM  
FOR ADMISSION TO PES UNIVERSITY

**REGISTER TODAY**



[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-expression-function/>)

## Average User Earns \$50/Month

Salad.com

```
# By Calling function
def func1(x):
    firstName=x.firstname
    lastName=x.lastName
    name=firstName+","+lastName
    gender=x.gender.lower()
    salary=x.salary*2
    return (name,gender,salary)

rdd2=df.rdd.map(lambda x: func1
```

## Using foreach() to Loop Through Rows in DataFrame

Similar to map(), foreach() also applied to every row of DataFrame, the difference being foreach() is an action and it returns nothing. Below are some examples to iterate through DataFrame using for each.

```
# Foreach example
def f(x): print(x)
df.foreach(f)

# Another example
df.foreach(lambda x:
    print("Data ==>"+x["firstna
    )
```

## Using pandas() to Iterate

If you have a small dataset, you can also [Convert PySpark DataFrame to Pandas](#)

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

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

[PySpark – concat\\_ws\(\).  
\(https://sparkbyexamples.com/pyspark/pyspark-convert-array-column-to-string-column/\)](https://sparkbyexamples.com/pyspark/pyspark-convert-array-column-to-string-column/)

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

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

[PySpark – regexp\\_replace\(\).  
\(https://sparkbyexamples.com/pyspark/pyspark-replace-column-values/#regexp\\_replace-replace-string-columns\)](https://sparkbyexamples.com/pyspark/pyspark-replace-column-values/#regexp_replace-replace-string-columns)

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

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

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

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

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

[\(https://sparkbyexamples.com/pyspark/convert-pyspark-dataframe-to-pandas/\)](https://sparkbyexamples.com/pyspark/convert-pyspark-dataframe-to-pandas/) and use pandas to iterate through. Use `spark.sql.execution.arrow.enabled` config to enable Apache Arrow with Spark. Apache Spark uses Apache Arrow which is an in-memory columnar format to transfer the data between Python and JVM.

```
# Using pandas
import pandas as pd
spark.conf.set("spark.sql.execution.arrow.enabled", "true")
pandasDF = df.toPandas()
for index, row in pandasDF.iterrows():
    print(row['firstname'], row['lastname'])
```

## Collect Data As List and Loop Through

You can also [Collect the PySpark DataFrame to Driver](https://sparkbyexamples.com/pyspark/pyspark-collect/) [\(https://sparkbyexamples.com/pyspark/pyspark-collect/\)](https://sparkbyexamples.com/pyspark/pyspark-collect/) and iterate through Python, you can also use `toLocalIterator()`.

```
# Collect the data to Python List
dataCollect = df.collect()
for row in dataCollect:
    print(row['firstname'] + ", " + row['lastname'])

#Using toLocalIterator()
dataCollect=df.rdd.toLocalIterator()
for row in dataCollect:
    print(row['firstname'] + ", " + row['lastname'])
```

## Conclusion

In this article, you have learned iterating/loop through Rows of PySpark DataFrame could be done using `map()`, `foreach()`, converting to Pandas, and finally converting DataFrame to Python List. If you want to do similar computations, use either `select` or `withColumn()`.



[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](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/>)

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

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

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

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

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

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

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

Happy Learning !!

## References

- <https://spark.apache.org/docs/2.2.0/api/python/pyspark.sql.html#pyspark.sql.DataFrame.foreach>  
(<https://spark.apache.org/docs/2.2.0/api/python/pyspark.sql.html#pyspark.sql.DataFrame.foreach>)

Share this:



(<https://sparkbyexamples.com/pyspark/pyspark-loop-iterate-through-rows-in-dataframe/?share=facebook&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-loop-iterate-through-rows-in-dataframe/?share=reddit&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-loop-iterate-through-rows-in-dataframe/?share=pinterest&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-loop-iterate-through-rows-in-dataframe/?share=tumblr&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-loop-iterate-through-rows-in-dataframe/?share=pocket&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-loop-iterate-through-rows-in-dataframe/?share=linkedin&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-loop-iterate-through-rows-in-dataframe/?share=twitter&nb=1>)

**TAGS:** [COLLECT\(\)](#),  
(<https://sparkbyexamples.com/tag/collect/>), [FOREACH\(\)](#),  
(<https://sparkbyexamples.com/tag/foreach/>), [MAP\(\)](#),  
(<https://sparkbyexamples.com/tag/map-2/>)



**NNK**

(<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/>)



[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)

[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)

## Leave a Reply

Enter your comment here...

### 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/>)

### Recent Posts

Spark regexp\_replace() – Replace String Value  
([https://sparkbyexamples.com/spark/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/>)

### 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





ry/cassandra/)

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

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

PySpark  
(https://sparkbyexamples.com/category/pyspark/)

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/)

f

(https://

in

/www.f

(https://

(https://

acebo

/www.li

(https://

/twitter

ok.co

nkedin

/github

.com/s

m/spar

.com/i

.com/s

parkby

kbyexa

n/n-nk-

park-

examp

mples/

b860a

examp

les)

)

8193/)

les/)

Copyright sparkbyexamples.com

^