

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/>)

-30%

-40%

tutorial/)

Hive (<https://sparkbyexamples.com/apache-hive-tutorial/>)

tutorial/)

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

hbase-tutorial/)

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

(<https://sparkbyexamples.com/author/admin/>)

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

FAQ's PySpark RDD/DataFrame collect()

is an action operation that is used to retrieve all the elements of the dataset (from all nodes) to the driver node. We should use the collect() on smaller dataset usually after filter()

question/)

More <https://sparkbyexamples.com/pyspark-pyspark-where-filter/>), group()

(<https://sparkbyexamples.com/pyspark/pyspark-groupby-explained-with-example/>) e.t.c. Retrieving larger datasets results in OutOfMemory error.

In this PySpark article, I will explain the usage of collect() with DataFrame example, when to avoid it, and the difference between collect() and select().

Related Articles:

- [How to Iterate PySpark DataFrame through Loop](#)
(<https://sparkbyexamples.com/pyspark>)



Your Website
Is Not 100% Visible



shop.adidas.co.in

PySpark Collect() – Retrieve data from DataFrame

[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/)

[k/pyspark-loop-iterate-through-rows-in-dataframe/\)](https://sparkbyexamples.com/pyspark-loop-iterate-through-rows-in-dataframe/)

- [How to Convert PySpark DataFrame Column to Python List
\(https://sparkbyexamples.com/pyspark/convert-pyspark-dataframe-column-to-python-list/\)](https://sparkbyexamples.com/pyspark/convert-pyspark-dataframe-column-to-python-list/)

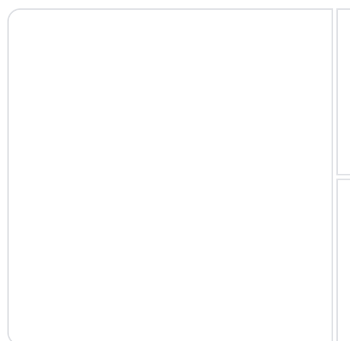
In order to explain with example, first, let's create a DataFrame [\(https://sparkbyexamples.com/pyspark/different-ways-to-create-dataframe-in-pyspark/\)](https://sparkbyexamples.com/pyspark/different-ways-to-create-dataframe-in-pyspark/).

```
import pyspark
from pyspark.sql import SparkSession

spark = SparkSession.builder.appName("PySpark Demo").getOrCreate()

dept = [("Finance",10), \
        ("Marketing",20), \
        ("Sales",30), \
        ("IT",40) \
       ]
deptColumns = ["dept_name","dept_no"]
deptDF = spark.createDataFrame(dept,deptColumns)
deptDF.show(truncate=False)
```

[show\(\)](https://sparkbyexamples.com/pyspark/pyspark-show-display-dataframe-contents-in-table/) function on DataFrame prints the result of DataFrame in a table format [\(https://sparkbyexamples.com/pyspark/pyspark-show-display-dataframe-contents-in-table/\)](https://sparkbyexamples.com/pyspark/pyspark-show-display-dataframe-contents-in-table/). By default, it shows only 20 rows. The above snippet returns the data in a table.



Royal Enfield

[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\(\).](#)

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

Now, let's use the `collect()` to retrieve the data.

```
dataCollect = deptDF.collect()
print(dataCollect)
```

`deptDF.collect()` retrieves all elements in a DataFrame as an Array of Row type
(<https://sparkbyexamples.com/pyspark/pyspark-row-using-rdd-dataframe/>) to the driver node. printing a resultant array yields the below output.

```
[Row(dept_name='Finance', dept_id=10),
 Row(dept_name='Marketing', dept_id=20),
 Row(dept_name='Sales', dept_id=30),
 Row(dept_name='IT', dept_id=40)]
```

Note that `collect()` is an action hence it does not return a DataFrame instead, it returns data in an Array to the driver. Once the data is in an array, you can use python for loop to process it further.

```
for row in dataCollect:
    print(row['dept_name'] + ",")
```

If you wanted to get first row and first column from a DataFrame.

```
#Returns value of First Row, First Column
deptDF.collect()[0][0]
```

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

Let's understand what's happening on above statement.

- `deptDF.collect()` returns Array of Row type.
- `deptDF.collect()[0]` returns the first element in an array (1st row).
- `deptDF.collect()[0][0]` returns the value of the first row & first column.

In case you want to just return certain elements of a DataFrame, you should call [PySpark select\(\) transformation
\(https://sparkbyexamples.com/pyspark/select-columns-from-pyspark-dataframe/\)](https://sparkbyexamples.com/pyspark/select-columns-from-pyspark-dataframe/) first.

```
dataCollect = deptDF.select("de
```

When to avoid Collect()

Usually, `collect()` is used to retrieve the action output when you have very small result set and calling `collect()` on an RDD/DataFrame with a bigger result set causes out of memory as it returns the entire dataset (from all workers) to the driver hence we should avoid calling `collect()` on a larger dataset.

collect () vs select ()

`select()` is a transformation that returns a new DataFrame and holds the columns that are selected whereas `collect()` is an action that returns the entire data set in an Array to the driver.

[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-values/#regexp_replace-replace-string-columns)

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

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

Complete Example of PySpark collect()

Below is complete PySpark example of using collect() on DataFrame, similarly you can also create a program using collect() with RDD.

```
import pyspark
from pyspark.sql import SparkSession

spark = SparkSession.builder.appName('PySparkCollect').getOrCreate()

dept = [("Finance",10), \
        ("Marketing",20), \
        ("Sales",30), \
        ("IT",40) \
        ]

deptColumns = ["dept_name","dept_no"]
deptDF = spark.createDataFrame(dept,deptColumns)
deptDF.printSchema()
deptDF.show(truncate=False)

dataCollect = deptDF.collect()

print(dataCollect)

dataCollect2 = deptDF.select("dept_name","dept_no")
print(dataCollect2)

for row in dataCollect:
    print(row['dept_name'] + ",")
```

This example is also available at [PySpark Github](#)
(<https://github.com/spark-examples/pyspark-examples/blob/master/pyspark-collect.py>) project.

Conclusion

In this PySpark article, you have learned the collect() function of the RDD/DataFrame is an action operation that returns all elements of the DataFrame to spark driver program and also learned it's not a good practice to use it on the bigger dataset.

Happy Learning !!

[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/>)

[PySpark – 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)

Share this:



(<https://sparkbyexamples.com/pyspark/pyspark-collect/?share=facebook&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-collect/?share=reddit&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-collect/?share=pinterest&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-collect/?share=tumblr&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-collect/?share=pocket&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-collect/?share=linkedin&nb=1>)



(<https://sparkbyexamples.com/pyspark/pyspark-collect/?share=twitter&nb=1>)

TAGS: [COLLECT\(\)](#)

(<https://sparkbyexamples.com/tag/collect/>)



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/>)

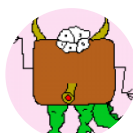
THIS POST HAS 3 COMMENTS



Anonymous

8 APR 2021 [REPLY](#)

Thank you 😊 Much appreciated for your quick response



Anonymous

6 APR 2021 [REPLY](#)

Pls explain how collect()
[0][0] works

[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)

[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)



NNK 6 APR 2021 [REPLY](#)

collect() returns Array of Row type. collect()[0] means first element in a array (1st row) and collect[0][0] means first column of first row.

On above DataFrame, if you run print(deptDF.collect()[0][0]), it prints “**Finance**” and print(deptDF.collect()[0][1]) print **10**.

[Leave a Reply](#)

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/>)

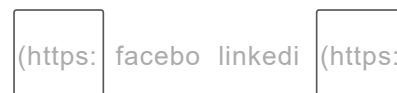
About SparkByExamples.Com

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.youtube.com/channel/UC860aexamples/>



<https://twitter.com/sparkbyexamples>

<https://www.facebook.com/sparkbyexamples/>

<https://www.linkedin.com/company/sparkbyexamples/>

<https://github.com/sparkbyexamples>

<https://www.youtube.com/channel/UC860aexamples/>