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/)

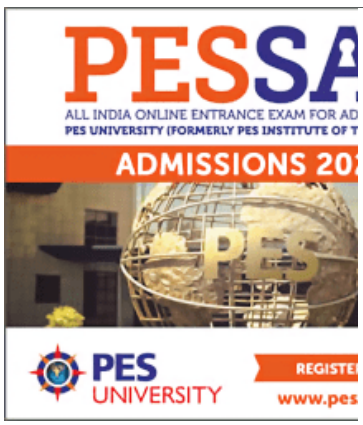
PySpark – Broadcast Variables (https://sparkbyexamples.com/pyspark/pyspark-broadcast-variables/)

Data Science with Internship - 25,000+ Certified
Datamites - Data Science Course...

PySpark orderBy() and sort() explained

NNK (https://sparkbyexamples.com/author/sparkbyexamp
PySpark (https://sparkbyexamples.com/category/pyspark/)

You can use either `sort()` or `orderBy()` function of PySpark DataFrame to sort DataFrame by ascending or descending order based on single or multiple columns, you can also do sorting using PySpark SQL sorting functions, In this article, I will explain all these different ways using PySpark examples.



- Using sort() function
- Using orderBy() function
- Ascending order
- Descending order
- SQL Sort functions

Explore our collection of insights, trends and ad innovations from Google Marketing Livestream

Learn more

Think with Googl

PySpark – Accumulator (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/)

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/)

PySpark – Convert 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/)

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

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

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

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

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

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

Related: How to sort DataFrame by using Scala (https://sparkbyexamples.com/spark/spark-how-to-sort-dataframe-column-explained/)

Before we start, first let's create a DataFrame (https://sparkbyexamples.com/pyspark/different-ways-to-create-dataframe-in-pyspark/).

```
simpleData = [("James","Sales","
    ("Michael","Sales","NY",860
    ("Robert","Sales","CA",8100
    ("Maria","Finance","CA",900
    ("Raman","Finance","CA",990
    ("Scott","Finance","NY",830
    ("Jen","Finance","NY",79000
    ("Jeff","Marketing","CA",80
    ("Kumar","Marketing","NY",9
  ]
columns= ["employee_name","depar
df = spark.createDataFrame(data
df.printSchema()
df.show(truncate=False)
```

This Yields below output.

```
root
 |-- employee_name: string (nul
 |-- department: string (nullab
 |-- state: string (nullable = 
 |-- salary: integer (nullable 
 |-- age: integer (nullable = fa
 |-- bonus: integer (nullable =

+-------------+----------+-----
|employee_name|department|state
+-------------+----------+-----
|        James|     Sales|   NY
|      Michael|     Sales|   NY
|       Robert|     Sales|   CA
|        Maria|   Finance|   CA
|        Raman|   Finance|   CA
|        Scott|   Finance|   NY
|          Jen|   Finance|   NY
|         Jeff| Marketing|   CA
|        Kumar| Marketing|   NY
+-------------+----------+-----
```

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/)

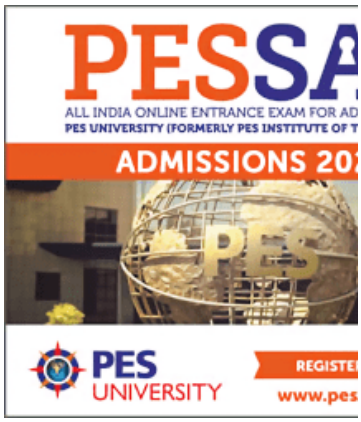
PySpark – flatMap() (https://sparkbyexamples.com/pyspark/pyspark-flatmap-transformation/)

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

PySpark – sample() vs sampleBy()

DataFrame sorting using the sort() function

PySpark DataFrame class provides `sort()` function to sort on one or more columns. By default, it sorts by ascending order.



Syntax

```
sort(self, *cols, **kwargs):
```

Example

```
df.sort("department","state").sh
df.sort(col("department"),col("s
```

The above two examples return the same below output, the first one takes the DataFrame column name as a string and the next takes columns in Column type. This table sorted by the first `department` column and then the `state` column.

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

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

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

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

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

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

PySpark SQL Functions

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

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/)

PySpark – JSON Functions (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/)

PySpark – Read & Write Parquet File (https://sparkbyexamples.com/

```
+-------------+----------+-----
|employee_name|department|state
+-------------+----------+-----
|Maria        |Finance   |CA
|Raman        |Finance   |CA
|Jen          |Finance   |NY
|Scott        |Finance   |NY
|Jeff         |Marketing |CA
|Kumar        |Marketing |NY
|Robert       |Sales     |CA
|James        |Sales     |NY
|Michael      |Sales     |NY
+-------------+----------+-----
```

DataFrame sorting using orderBy() function

PySpark DataFrame also provides `orderBy()` function to sort on one or more columns. By default, it orders by ascending.

Example

```
df.orderBy("department","state")
df.orderBy(col("department"),col
```

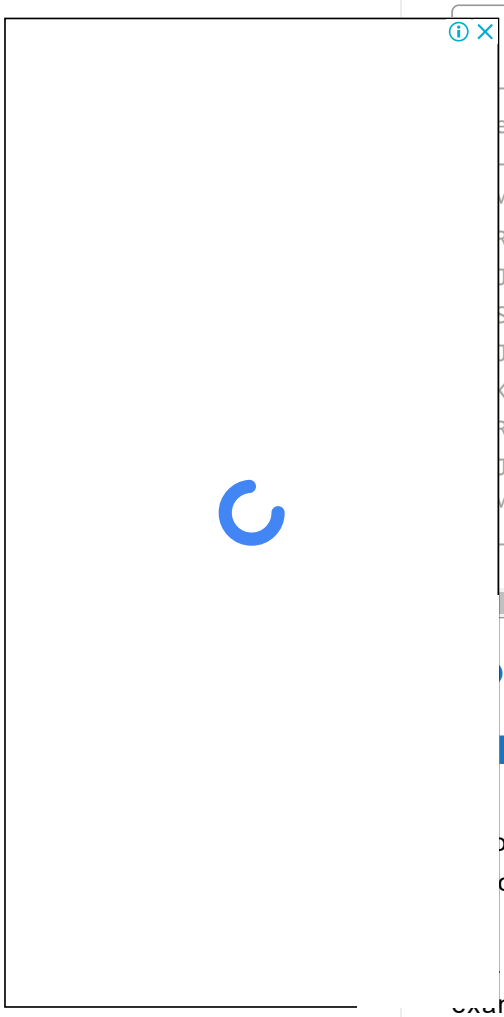
This returns the same output as the previous section.

Sort by Ascending (ASC)

If you wanted to specify the ascending order/sort explicitly on DataFrame, you can use the `asc` method of the `Column` function. for example

pyspark/pyspark-read-and-write-parquet-file/)

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



```
df.sort(df.department.asc(),df.s
df.sort(col("department").asc(),
df.orderBy(col("department").asc
```

The above three examples return the same output.

```
+-------------+----------+-----
|employee_name|department|state
+-------------+----------+-----
|Maria        |Finance   |CA
|Raman        |Finance   |CA
|Jen          |Finance   |NY
|Scott        |Finance   |NY
|Jeff         |Marketing |CA
|Kumar        |Marketing |NY
|Robert       |Sales     |CA
|James        |Sales     |NY
|Michael      |Sales     |NY
+-------------+----------+-----
```

rt by Descending ESC)

ou wanted to specify the sorting by cending order on DataFrame, you use the `desc` method of the umn function. for example. From our xample, let's use desc on the state column.

```
df.sort(df.department.asc(),df.s
df.sort(col("department").asc(),
df.orderBy(col("department").asc
```

This yields the below output for all three examples.

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/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())

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)

```
+-------------+----------+-----
|employee_name|department|state
+-------------+----------+-----
|Scott        |Finance   |NY
|Jen          |Finance   |NY
|Raman        |Finance   |CA
|Maria        |Finance   |CA
|Kumar        |Marketing |NY
|Jeff         |Marketing |CA
|James        |Sales     |NY
|Michael      |Sales     |NY
|Robert       |Sales     |CA
+-------------+----------+-----
```

Besides `asc()` and `desc()` functions, PySpark also provides `asc_nulls_first()` and `asc_nulls_last()` and equivalent descending functions.

Using Raw SQL

Below is an example of how to sort DataFrame using raw SQL syntax.

```
df.createOrReplaceTempView("EMP"
spark.sql("select employee_name
```

The above two examples return the same output as above.

Dataframe Sorting

Complete Example

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)

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)

PySpark – rank() (https://sparkbyexamples.com/pyspark/pyspark-window-functions/#rank)

PySpark – dense_rank() (https://sparkbyexamples.com/pyspark/pyspark-window-functions/#dense_rank)

PySpark – percent_rank() (https://sparkbyexamples.com/pyspark/pyspark-window-functions/#percent_rank)

```
import pyspark
from pyspark.sql import SparkSes
from pyspark.sql.functions impor

spark = SparkSession.builder.app

simpleData = [("James","Sales","
    ("Michael","Sales","NY",8600
    ("Robert","Sales","CA",81000
    ("Maria","Finance","CA",900
    ("Raman","Finance","CA",990
    ("Scott","Finance","NY",830
    ("Jen","Finance","NY",79000
    ("Jeff","Marketing","CA",80
    ("Kumar","Marketing","NY",91
  ]
columns= ["employee_name","depar

df = spark.createDataFrame(data

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

df.sort("department","state").sh
df.sort(col("department"),col("s

df.orderBy("department","state"
df.orderBy(col("department"),co

df.sort(df.department.asc(),df.s
df.sort(col("department").asc()
df.orderBy(col("department").asc

df.sort(df.department.asc(),df.s
df.sort(col("department").asc()
df.orderBy(col("department").asc

df.createOrReplaceTempView("EMP"
spark.sql("select employee_name
```

This complete example is also available at PySpark sorting GitHub project (https://github.com/spark-examples/pyspark-examples/blob/master/pyspark-orderby.py) for reference.

Conclusion

Here you have learned how to Sort PySpark DataFrame columns using `sort()`, `orderBy()` and using SQL

PySpark – typedLit() (https://sparkbyexamples.com/pyspark/pyspark-lit-add-literal-constant/#typedlit)

PySpark – from_json() (https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/#from_json)

PySpark – to_json() (https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/#to_json)

PySpark – json_tuple() (https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/#json_tuple)

PySpark – get_json_object() (https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/#get_json_object)

PySpark – schema_of_json() (https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/#schema_of_json)

sort functions and used this function with PySpark SQL along with Ascending and Descending sorting orders.

Happy Learning !!

Share this:

(https://sparkbyexamples.com/pyspark/pyspark-orderby-and-sort-explained/?share=facebook&nb=1)

(https://sparkbyexamples.com/pyspark/pyspark-orderby-and-sort-explained/?share=reddit&nb=1)

(https://sparkbyexamples.com/pyspark/pyspark-orderby-and-sort-explained/?share=pinterest&nb=1)

2

(https://sparkbyexamples.com/pyspark/pyspark-orderby-and-sort-explained/?share=tumblr&nb=1)

(https://sparkbyexamples.com/pyspark/pyspark-orderby-and-sort-explained/?share=pocket&nb=1)

(https://sparkbyexamples.com/pyspark/pyspark-orderby-and-sort-explained/?share=linkedin&nb=1)

(https://sparkbyexamples.com/pyspark/pyspark-orderby-and-sort-explained/?share=twitter&nb=1)

TAGS: PYSPARK ORDERBY (HTTPS://SPARKBYEXAMPLES.COM/TAG/PYSPARK-ORDERBY/)

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

NNK (Https://Sparkbyexamples.Com/Author/Sparkbyexamp

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

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/catego

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/)

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/snowflake/)

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

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

(https://www.facebook.com/sparkbyexamples/) (https://www.linkedin.com/in/n-nk-b860a8193/) (https://twitter.com/sparkbyexamples) (https://github.com/spark-examples/)

Copyright sparkbyexamples.com