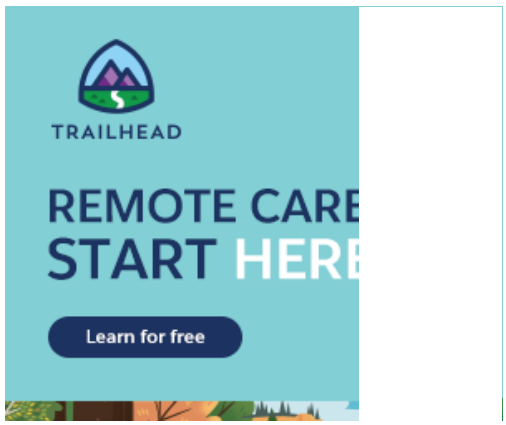
Delicious Cakes
Ferns N Petals

Spark – How to Sort DataFrame column explained

NNK (https://sparkbyexamples.com/author/admin/) -
Apache Spark (https://sparkbyexamples.com/category/spark/)

In Spark, you can use either `sort()` or `orderBy()` function of DataFrame/Dataset to sort by ascending or descending order based on single or multiple columns, you can also do sorting using Spark SQL sorting functions (https://sparkbyexamples.com/spark/spark-sql-sort-functions/), In this article, I will explain all these different ways using Scala examples.



- Using sort() function
- Using orderBy() function
- Ascending order

Spark Tutorial

Spark – Installation on Windows (https://sparkbyexamples.com/spark/apache-spark-installation-on-windows/)

Spark – Installation on Linux | Ubuntu (https://sparkbyexamples.com/spark/spark-installation-on-linux-ubuntu/)

Spark – Cluster Setup with Hadoop Yarn (https://sparkbyexamples.com/spark/spark-setup-on-hadoop-yarn/)

Spark – Web/Application UI (https://sparkbyexamples.com/spark/spark-web-ui-understanding/)

Spark – Setup with Scala and IntelliJ (https://sparkbyexamples.com/spark/spark-setup-run-with-scala-intellij/)

Spark – How to Run Examples From this Site on IntelliJ IDEA (https://sparkbyexamples.com/spark/how-to-run-spark-examples-from-intellij/)

Spark – SparkSession (https://sparkbyexamples.com/spark/sparksession-explained-with-examples/)

Spark – SparkContext (https://sparkbyexamples.com/spark/spark-sparkcontext/)

Spark RDD Tutorial

Spark RDD – Parallelize (https://sparkbyexamples.com/apache-spark-rdd/how-to-create-an-rdd-using-parallelize/)

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

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

Spark RDD – Create RDD (https://sparkbyexamples.com/apache-spark-rdd/different-ways-to-create-spark-rdd/)

Spark RDD – Create Empty RDD (https://sparkbyexamples.com/apache-spark-rdd/spark-how-to-create-an-empty-rdd/)

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

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

Spark RDD – Pair Functions (https://sparkbyexamples.com/apache-spark-rdd/spark-pair-rdd-functions/)

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

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

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

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

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

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

Spark RDD – Convert RDD to DataFrame (https://sparkbyexamples.com/

- Descending order
- SQL Sort functions

Before we start, first let's create a DataFrame.

```
import spark.implicits._

val simpleData = Seq(("James"
  ("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
)
val df = simpleData.toDF("emp
df.show()
```

This Yields below output.

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

DataFrame sorting using the sort() function

apache-spark-rdd/convert-spark-rdd-to-dataframe-dataset/)

Spark SQL Tutorial

Spark SQL – Create DataFrame (https://sparkbyexamples.com/spark/different-ways-to-create-a-spark-dataframe/)

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

Spark SQL – Add and Update Column (withColumn) (https://sparkbyexamples.com/spark/spark-dataframe-withcolumn/)

Spark SQL – Rename Nested Column (https://sparkbyexamples.com/spark/rename-a-column-on-spark-dataframes/)

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

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

Spark SQL – When Otherwise (https://sparkbyexamples.com/spark/spark-case-when-otherwise-example/)

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

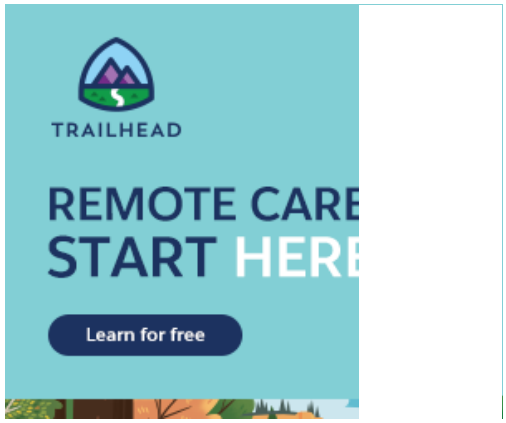
Spark SQL – Distinct (https://sparkbyexamples.com/spark/spark-remove-duplicate-rows/)

Spark SQL- Pivot Table DataFrame (https://sparkbyexamples.com/spark/how-to-pivot-table-and-unpivot-a-spark-dataframe/)

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

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

Syntax




```
sort(sortCol : scala.Predef.Str
sort(sortExprs : org.apache.spa
```

Example

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

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.

```
+-------------+----------+-----
|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
+-------------+----------+-----
```

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

Spark SQL – Schema (https://sparkbyexamples.com/spark/spark-schema-explained-with-examples/)

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

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

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

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

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

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

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

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

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

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

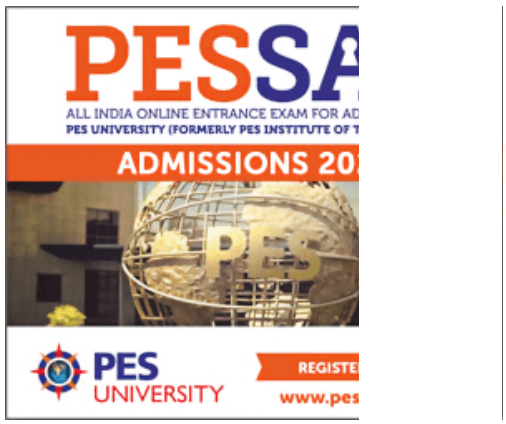
DataFrame sorting using orderBy() function

Alternatively, Spark DataFrame/Dataset class also provides `orderBy()` function to sort on one or more columns. By default, it also orders by ascending.

Syntax

```
orderBy(sortCol : scala.Predef.
orderBy(sortExprs : org.apache.
```

Example




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

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` (https://sparkbyexamples.com/spark/spark-sql-sort-functions/#asc) method of the `Column` function. for example

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

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

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

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

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

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

Spark SQL – Partitioning (https://sparkbyexamples.com/spark/spark-partitioning-understanding/)

Spark SQL Functions

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

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

Spark SQL Array Functions (https://sparkbyexamples.com/spark/spark-sql-array-functions/)

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

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

Spark SQL Aggregate Functions (https://sparkbyexamples.com/

The above two 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
+-------------+----------+-----
```

Sort by Descending (DESC)

If you wanted to specify the sorting by descending order on DataFrame, you can use the `desc` (https://sparkbyexamples.com/spark/spark-sql-sort-functions/#desc) method of the `Column` function. for example

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

This yields the below output.

```
+-------------+----------+-----
|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
+-------------+----------+-----
```

spark/spark-sql-aggregate-functions/)

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

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

Spark Data Source API

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

Spark – Read and Write JSON file (https://sparkbyexamples.com/spark/spark-read-and-write-json-file/)

Spark – Read & Write Parquet file (https://sparkbyexamples.com/spark/spark-read-write-dataframe-parquet-example/)

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

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

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

Spark – Read & Write HBase using “hbase-spark” Connector (https://sparkbyexamples.com/spark/spark-read-write-using-hbase-spark-connector/)

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

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

Using Sorting functions

Spark SQL function provides several sorting functions, below are some examples of how to use asc (https://sparkbyexamples.com/spark/spark-sql-sort-functions/#asc) and desc (https://sparkbyexamples.com/spark/spark-sql-sort-functions/#desc) functions. Besides these Spark also provides asc_nulls_first (https://sparkbyexamples.com/spark/spark-sql-sort-functions/#asc-nulls-first) and asc_nulls_last (https://sparkbyexamples.com/spark/spark-sql-sort-functions/#asc-nulls-last) functions and equivalent for descending.

```
df.select($"employee_name",asc(
```

Using Sorting on SQL

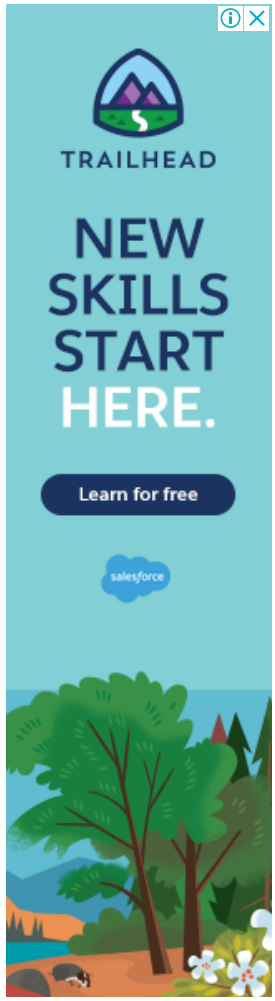
```
df.createOrReplaceTempView("EMP
spark.sql(" select employee_nam
```

The above two examples return the same output as above.

Dataframe Sorting

Complete Example

-30% -30%
shop.adidas.co.



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

Spark Streaming & Kafka

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

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

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

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

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

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

```
package com.sparkbyexamples.spa

import org.apache.spark.sql.Spa
import org.apache.spark.sql.fun
object SortExample extends App

  val spark: SparkSession = Spa
    .master("local[1]")
    .appName("SparkByExamples.c
    .getOrCreate()

  spark.sparkContext.setLogLeve

  import spark.implicits._

  val simpleData = Seq(("James"
    ("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
  )
  val df = simpleData.toDF("emp
  df.printSchema()
  df.show()

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

  df.orderBy("department","stat
  df.orderBy(col("department"),

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

  df.sort(col("department").asc
  df.orderBy(col("department").
 df.select($"employee_name",asc
  df.createOrReplaceTempView("E
  spark.sql(" select employee_n

}
```

This complete example is also available at GitHub project (https://github.com/spark-examples/spark-scala-examples/blob/master/src/main/scala/co

parkbyexamples/spark/dataframe/fu
ons/SortExample.scala) for
rence.

nclusion

e you have learned how to Sort
rk DataFrame columns using sort(),
erBy() and using SQL sort functions
used this function with Spark SQL
g with Ascending and Descending
ing orders.

py Learning !!

e this:

(https://sparkbyexamples.com/spark/spark-
-to-sort-dataframe-column-explained/?
re=facebook&nb=1)

(https://sparkbyexamples.com/spark/spark-
-to-sort-dataframe-column-explained/?
re=reddit&nb=1)

(https://sparkbyexamples.com/spark/spark-
-to-sort-dataframe-column-explained/?
re=pinterest&nb=1)

(https://sparkbyexamples.com/spark/spark-how-to-sort-dataframe-column-explained/?share=tumblr&nb=1)

(https://sparkbyexamples.com/spark/spark-how-to-sort-dataframe-column-explained/?share=pocket&nb=1)

(https://sparkbyexamples.com/spark/spark-how-to-sort-dataframe-column-explained/?share=linkedin&nb=1)

(https://sparkbyexamples.com/spark/spark-how-to-sort-dataframe-column-explained/?share=twitter&nb=1)

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

NNK

(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...

Personalized Tissue

Custom Tissue Holder With Name & Charm in Wine, Tan, Black & Many More Variants. Shop Now!

The Messy Corner

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.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