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

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/)

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

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

IIT Madras Data Science & AI

Learn from IIT Madras Professors & Industry Practitioners. Advance Your Career Now!

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

Spark withColumnRenamed to Rename Column

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

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

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

In Spark withColumnRenamed() is used to rename one column or multiple DataFrame column names. Depends on the DataFrame schema, renaming columns might get simple to complex, especially when a column is nested with struct type it gets complicated.

More ▼ (https://sparkbyexamples.com/)

In this article, I will explain how to rename a DataFrame column with multiple use cases like rename selected multiple columns, nested struct columns, all columns with Scala examples.

1. Using withColumnRenamed – To rename Spark DataFrame column name
2. Using withColumnRenamed – To rename multiple columns
3. Using StructType – To rename nested column on Spark DataFrame
4. Using Select – To rename nested columns
5. Using withColumn – To rename nested columns
6. Using col() function – To Dynamically rename all or multiple columns
7. Using toDF() – To rename all or multiple columns

First, let's create our data for our examples, we are using Row class as we convert this data to Spark DataFrame (https://sparkbyexamples.com/spark/different-ways-to-create-a-spark-dataframe/).

```
val data = Seq(Row(Row("James ",
  Row(Row("Michael ","Rose",""),
  Row(Row("Robert ","","Williams
  Row(Row("Maria ","Anne","Jones
  Row(Row("Jen","Mary","Brown"),
)
```

Our base schema (https://sparkbyexamples.com/spark/spark-schema-explained-with-examples/)with nested structure.

Spark SQL Tutorial

```
val schema = new StructType()
  .add("name",new StructType()
    .add("firstname",StringType)
    .add("middlename",StringType
    .add("lastname",StringType))
  .add("dob",StringType)
  .add("gender",StringType)
  .add("salary",IntegerType)
```

Let's create the DataFrame by using
parallelize
(https://sparkbyexamples.com/apache-
spark-rdd/how-to-create-an-rdd-using-
parallelize/)and provide the above
schema.

```
val df = spark.createDataFrame(s
df.printSchema()
```

Below is our schema structure. I am not
printing data here as it is not necessary
for our examples. This schema has a
nested structure.

```
root
 |-- name: struct (nullable = true)
 |    |-- firstname: string (nullable = true)
 |    |-- middlename: string (nullable = true)
 |    |-- lastname: string (nullable = true)
 |-- dob: string (nullable = true)
 |-- gender: string (nullable = true)
 |-- salary: integer (nullable = true)
```

1. Using Spark withColumnRenamed – To rename DataFrame column name

Spark has a `withColumnRenamed()` function on DataFrame to change a column name. This is the most straight forward approach; this function takes two parameters; the first is your existing column name and the second is the new column name you wish for.

Syntax:

```
def withColumnRenamed(existingNa
```

`existingName` – The existing column name you want to change

`newName` – New name of the column

Returns a new DataFrame (Dataset[Row]) with a column renamed. This is a no-op if the schema doesn't contain `existingName`.

Example:

```
df.withColumnRenamed("dob","Date
    .printSchema()
```

The above example changes the column name from "dob" to "DateOfBirth" on spark DataFrame. Note that `withColumnRenamed` function returns a new DataFrame and doesn't modify the current DataFrame.

```
root
 |-- name: struct (nullable = true)
 |    |-- firstname: string (nullable = true)
 |    |-- middlename: string (nullable = true)
 |    |-- lastname: string (nullable = true)
 |-- DateOfBirth: string (nullable = true)
 |-- gender: string (nullable = true)
 |-- salary: integer (nullable = true)
```

2. Using withColumnRenamed – To rename multiple columns

To change multiple column names, we should chain `withColumnRenamed` functions as shown below.

```
val df2 = df.withColumnRenamed("
            .withColumnRenamed("s
df2.printSchema()
```

This creates a new DataFrame "df2" after renaming dob and salary columns.

3. Using Spark StructType – To rename a nested column in Dataframe

Changing a column name on nested data is not straight forward and we can do this by creating a new schema with new DataFrame columns using StructType

(https://sparkbyexamples.com/spark/spar
k-sql-structtype-on-dataframe/) and use
it using cast function as shown below.

```
val schema2 = new StructType()
    .add("fname",StringType)
    .add("middlename",StringType
    .add("lname",StringType)
```

```
df.select(col("name").cast(schem
  col("dob"),
  col("gender"),
  col("salary"))
    .printSchema()
```

This statement renames firstname to
fname and lastname to lname within
name structure.

```
root
 |-- name: struct (nullable = true)
 |    |-- fname: string (nullable = true)
 |    |-- middlename: string (nullable = true)
 |    |-- lname: string (nullable = true)
 |-- dob: string (nullable = true)
 |-- gender: string (nullable = true)
 |-- salary: integer (nullable = true)
```

4. Using Select – To rename nested elements.

Let's see another way to change nested
columns by transposing the structure to
flat.

```
df.select(col("name.firstname").
  col("name.middlename").as("mna
  col("name.lastname").as("lname
  col("dob"),col("gender"),col("
  .printSchema()
```

```
root
 |-- fname: string (nullable = true)
 |-- mname: string (nullable = true)
 |-- lname: string (nullable = true)
 |-- dob: string (nullable = true)
 |-- gender: string (nullable = true)
 |-- salary: integer (nullable = true)
```

**Spark Streaming &
Kafka**

5. Using Spark DataFrame withColumn – To rename nested columns

When you have nested columns on Spark DatFrame and if you want to rename it, use `withColumn` on a data frame object to create a new column from an existing and we will need to drop the existing column. Below example creates a "fname" column from "name.firstname" and drops the "name" column

```
val df4 = df.withColumn("fname",
      .withColumn("mname",col("n
      .withColumn("lname",col("n
      .drop("name")
df4.printSchema()
```

6. Using col() function – To Dynamically rename all or multiple columns

Another way to change all column names on Dataframe is to use `col()` function.

```
val old_columns = Seq("dob","gende
    val new_columns = Seq("DateOfB
    val columnsList = old_columns.
    val df5 = df4.select(columnsLi
    df5.printSchema()
```

7. Using toDF() – To change all columns in a Spark DataFrame

When we have data in a flat structure (without nested) , use `toDF()` with a new schema to change all column names.

```scala
val newColumns = Seq("newCol1","
val df = df.toDF(newColumns:_*)
```

The complete code can be downloaded from GitHub (https://github.com/sparkbyexamples/spark-examples/blob/master/spark-sql-examples/src/main/scala/com/sparkbyexamples/spark/dataframe/RenameColDataFrame.scala)

Conclusion:

This article explains `withColumnRenamed()` function and different ways to rename a single column, multiple, all, and nested columns on Spark DataFrame. Besides what explained here, we can also change column names using Spark SQL and the same concept can be used in PySpark.

Hope you like this article!!

Happy Learning.

Share this:

2

(https://sparkbyexamples.com/spark/rename-a-column-on-spark-dataframes/?share=pocket&nb=1)

(https://sparkbyexamples.com/spark/rename-a-column-on-spark-dataframes/?share=linkedin&nb=1)

(https://sparkbyexamples.com/spark/rename-a-column-on-spark-dataframes/?share=twitter&nb=1)

TAGS: **SPARK.CREATEDATAFRAME (HTTPS://SPARKBYEXAMPLES.COM/TAG/CREATEDATAFRAME/)**, **STRUCTTYPE (HTTPS://SPARKBYEXAMPLES.COM/TAG/STRUCTTYPE/)**, **TODF() (HTTPS://SPARKBYEXAMPLES.COM/TAG/TODF/)**, **WITHCOLUMN (HTTPS://SPARKBYEXAMPLES.COM/TAG/WITHCOLUMN/)**, **WITHCOLUMNRENAMED (HTTPS://SPARKBYEXAMPLES.COM/TAG/WITHCOLUMNRENAMED/)**

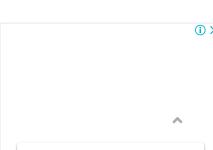## NNK (Https://Sparkbyexamples.Com/Author/Admin/)

(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/)

Leave a Reply

Invalid request signature.

Advance Certification Program

Learn from IIT Madras Professors & Industry Practitioners. Advance Your Career Now!

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/categor

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

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

y/hbase/)

Apache Cassandra (https://sparkbyexamples.com/categor y/cassandra/)

Snowflake Database (https://sparkbyexamples.com/categor y/snowflake/)

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

PySpark (https://sparkbyexamples.com/categor y/pyspark/)

(https://sparkbyexamples.com/spark/spar k-isin-is-not-in-operator-example/)

Spark – Get Size/Length of Array & Map Column (https://sparkbyexamples.com/spark/spar k-get-size-length-of-array-map-column/)

Spark Using Length/Size Of a DataFrame Column (https://sparkbyexamples.com/spark/spar k-using-length-size-of-a-dataframe-column/)

Spark rlike() Working with Regex Matching Examples (https://sparkbyexamples.com/spark/spar k-rlike-regex-matching-examples/)

Spark Check String Column Has Numeric Values (https://sparkbyexamples.com/spark/spar k-check-string-column-has-numeric-values/)

Spark Check Column Data Type is Integer or String (https://sparkbyexamples.com/spark/spar k-check-column-data-type-is-integer-or-string/)

(https: (https:

//www. //www.

(https: facebo linkedi (https:

//twitte ok.co n.com/ //githu

r.com/ m/spar in/n- b.com/

sparkb kbyex nk- spark-

yexam ample b860a examp

ples) s/) 8193/) les/)