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%

-50%

-30%

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

shop.adidas.co.in

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

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

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

More ∨   (https://sparkbyexamples.com/) 🔍

PySpark withColumnRenamed to Rename Column on DataFrame

👤 NNK (https://sparkbyexamples.com/author/admin/) -
📁 PySpark (https://sparkbyexamples.com/category/pyspark/)

Use PySpark withColumnRenamed() to rename a DataFrame column, we often need to rename one column or multiple (or all) columns on PySpark DataFrame, you can do this in several ways. When columns are nested it becomes complicated.

Since DataFrame's are an immutable collection, you can't rename or update a column instead when using withColumnRenamed() it creates a new DataFrame with updated column names, In this PySpark article, I will cover different ways to rename columns with several use cases like rename nested column, all columns, selected multiple columns with Python/PySpark examples.

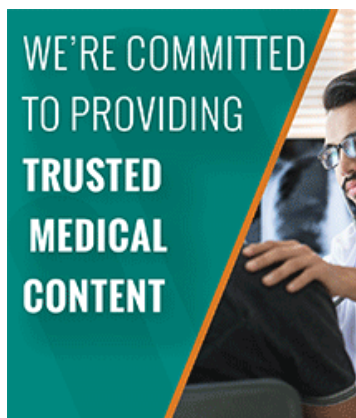Refer to this page, If you are looking for a Spark with Scala example (https://sparkbyexamples.com/spark/rename-a-column-on-spark-dataframes/).

1. PySpark withColumnRenamed – To rename DataFrame column name
2. PySpark withColumnRenamed – To rename multiple columns
3. Using StructType – To rename nested column on PySpark 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 set to work with.

```
dataDF = [(('James','','Smith')
  (('Michael','Rose',''),'2000-(
  (('Robert','','Williams'),'19
  (('Maria','Anne','Jones'),'19
  (('Jen','Mary','Brown'),'1980
]
```

Our base schema with nested structure.

```
from pyspark.sql.types import St
schema = StructType([
        StructField('name', Str
            StructField('first
            StructField('middl
            StructField('lastn
            ])),
        StructField('dob', Str
        StructField('gender',
        StructField('gender',
        ])
```

Let's create the DataFrame by using parallelize and provide the above schema.

```
from pyspark.sql import SparkSe
spark = SparkSession.builder.ap
df = spark.createDataFrame(data
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. PySpark withColumnRenamed – To rename DataFrame column name

PySpark 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.

PySpark withColumnRenamed() Syntax:

```
withColumnRenamed(existingName,
```

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

`newName` – New name of the column

Returns a new DataFrame with a column renamed.

Example

```
df.withColumnRenamed("dob","Dat
```

The above statement changes column "dob" to "DateOfBirth" on PySpark 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. PySpark withColumnRenamed – To rename multiple columns

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

store all columns to rename in a list and
loop through to rename all columns, I
will leave this to you to explore.

```
df2 = df.withColumnRenamed("dob
    .withColumnRenamed("salary"
'f2.printSchema()
```

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

Using PySpark

ructType – To rename a

...sted column in

ataframe

nging a column name on nested
a is not straight forward and we can
his by creating a new schema with
DataFrame columns using
ctType
s://sparkbyexamples.com/spark/sp
sql-structtype-on-dataframe/) and
it using cast function as shown
ow.

```
chema2 = StructType([
    StructField("fname",StringT}
    StructField("middlename",St
    StructField("lname",StringT}
```

```
f.select(col("name").cast(sche
    col("dob"), col("gender"),
    .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.

```
from pyspark.sql.functions impo
df.select(col("name.firstname")
   col("name.middlename").alias(
   col("name.lastname").alias("l
   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)
```

5. Using PySpark DataFrame withColumn – To rename nested columns

When you have nested columns on PySpark 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

```
from pyspark.sql.functions impo
df4 = df.withColumn("fname",col
        .withColumn("mname",col("
        .withColumn("lname",col("
        .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.

```
IN progress
```

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

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

```
newColumns = ["newCol1","newCol
df.toDF(*newColumns).printSchema
```

Source code

```python
import pyspark
from pyspark.sql import SparkSe
from pyspark.sql.types import S
from pyspark.sql.functions impo

spark = SparkSession.builder.ap

dataDF = [(('James','','Smith')
  (('Michael','Rose',''),'2000-
  (('Robert','','Williams'),'19
  (('Maria','Anne','Jones'),'19
  (('Jen','Mary','Brown'),'1980
]

schema = StructType([
      StructField('name', Str
          StructField('first
          StructField('middl
          StructField('lastn
          ])),
      StructField('dob', Str
      StructField('gender',
      StructField('salary',
      ])

df = spark.createDataFrame(data
df.printSchema()

# Example 1
df.withColumnRenamed("dob","Dat
# Example 2
df2 = df.withColumnRenamed("dob
    .withColumnRenamed("salary"
df2.printSchema()

# Example 3
schema2 = StructType([
    StructField("fname",StringT
    StructField("middlename",St
    StructField("lname",StringT

df.select(col("name").cast(sche
  col("dob"),
  col("gender"),
  col("salary")) \
    .printSchema()

# Example 4
df.select(col("name.firstname")
  col("name.middlename").alias(
  col("name.lastname").alias("l
  col("dob"),col("gender"),col(
  .printSchema()

# Example 5
```

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

#Example 7
newColumns = ["newCol1","newCol
df.toDF(*newColumns).printSchem

# Example 6
'''
not working
old_columns = Seq("dob","gender
new_columns = Seq("DateOfBirth"
columnsList = old_columns.zip(n
df5 = df4.select(columnsList:_*
df5.printSchema()
'''
```

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 different ways to rename all, a single, multiple, and nested columns on PySpark DataFrame.

I hope you like this article!!

Happy Learning.

TAGS: **SPARK.CREATEDATAFRAME (HTTPS://SPARKBYEXAMPLES.COM/TAG/CREATED ATAFRAME/)**, **STRUCTTYPE (HTTPS://SPARKBYEXAMPLES.COM/TAG/STRUCTT YPE/)**, **TODF() (HTTPS://SPARKBYEXAMPLES.COM/TAG/TODF/)**, **WITHCOLUMN (HTTPS://SPARKBYEXAMPLES.COM/TAG/WITHCOL UMN/)**, **WITHCOLUMNRENAMED (HTTPS://SPARKBYEXAMPLES.COM/TAG/WITHCOL UMNRENAMED/)**

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/)
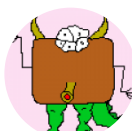
❯ THIS POST HAS 14 COMMENTS

Anonymous

10 APR 2021   REPLY

Really helped me. Thanks a lot

Anonymous

24 FEB 2021   REPLY

Veri good tutorials , keep it up

NNK 24 FEB 2021

Thanks for liking PySpark withColumnRenamed Example

GonzaloGF

30 JAN 2021

hi NNK I would like to tell you that you have done a great job with this website, I thank you from the bottom of my heart because I am learning a lot and in a simple way, since you explain everything very well.

NNK 31 JAN 2021

Thank you for your comments. It means a lot to me and motivates me to write more.

Avula Sreenu

27 NOV 2020

Thank you so much bro, the way you have to show simple easy to understandable tutorial

---

NNK 27 NOV 2020

Thanks Sreenu. Glad you like the articles.

Janusz   12 NOV 2020

NNK, Example 7 is working when you use asterisk:
newColumns = ["newCol1″,"newCol2″,"newCol3″,"newCol4"]
new_df = df.toDF(*newColumns).printSchema()
Please update the article 😊

NNK **12 NOV 2020**   REPLY

Thanks Janusz. I've added this statement. appreciate your comment. Hope articles are helpful.

Anonymous

19 OCT 2020   REPLY

Thank you so much

Anonymous

9 SEP 2020   REPLY

Thank you so much for the entire article. You have explained it in such simple words that it is so easy to understand. Keep up the good work! 😊

NNK **9 SEP 2020**   REPLY

Thank you for your wonderful words. Happy helping the Spark community.

Manjunath

18 AUG 2020   REPLY

Hi NNK(I dont have your full name, I m sorry) Thanks for your postings,

I appreciate them heart fully.
Learning Pyspark on my own.

Anonymous

18 AUG 2020

You are simply Awesome.

Leave a Reply

Follow Us

(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/)