

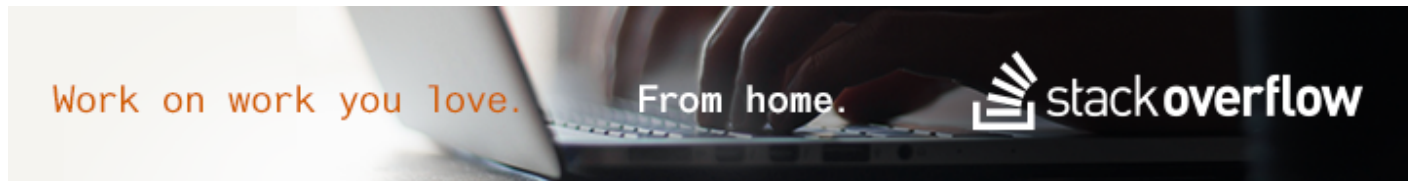
Announcing Stack Overflow Documentation

We started with Q&A. Technical documentation is next, and we need your help.

Whether you're a beginner or an experienced developer, you *can* contribute.

[Sign up and start helping →](#)[Learn more about Documentation →](#)

Trim string column in PySpark dataframe



I'm beginner on Python and Spark. After creating a `DataFrame` from `csv` file, I would like to know how I can trim a column. I've try:

```
df = df.withColumn("Product", df.Product.strip())
```

`df` is my data frame, `Product` is a column in my table

But I see always the error:

```
Column object is not callable
```

Do you have any suggestions?

[python](#) [apache-spark](#) [trim](#) [pyspark-sql](#)

edited Feb 2 at 19:27



desertnaut

1,214 2 15

asked Feb 2 at 14:15



minh-hieu.pham

134 2 12

Did the solution work??? – desertnaut Feb 5 at 14:11

1 Answer

Starting from version **1.5**, Spark SQL provides two specific functions for trimming white space, `ltrim` and `rtrim` (search for "trim" in the [DataFrame documentation](#)); you'll need to import `pyspark.sql.functions` first. Here is an example:

```
from pyspark.sql import SQLContext
from pyspark.sql.functions import *
sqlContext = SQLContext(sc)

df = sqlContext.createDataFrame([(' 2015-04-08 ', ' 2015-05-10 ')], ['d1', 'd2']) # create
a dataframe - notice the extra whitespaces in the date strings
df.collect()
# [Row(d1=u' 2015-04-08 ', d2=u' 2015-05-10 ')]
df = df.withColumn('d1', ltrim(df.d1)) # trim left whitespace from column d1
df.collect()
# [Row(d1=u'2015-04-08 ', d2=u' 2015-05-10 ')]
df = df.withColumn('d1', rtrim(df.d1)) # trim right whitespace from d1
df.collect()
# [Row(d1=u'2015-04-08', d2=u' 2015-05-10 ')]
```

answered Feb 2 at 18:45



desertnaut

1,214 2 15