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

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

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

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 Replace Column Values in DataFrame

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

You can replace column values of PySpark DataFrame by using SQL string functions *regexp_replace*(), translate(), and overlay() with Python examples.

In this article, I will cover examples of how to replace part of a string with another string, replace all columns, change values conditionally, replace values from a python dictionary, replace column value from another DataFrame column e.t.c

First, let's create a PySpark DataFrame with some addresses and will use this DataFrame to explain how to replace column values.

```
from pyspark.sql import SparkSessi
spark = SparkSession.builder.master
address = [(1,"14851 Jeffrey Rd","D
    (2,"43421 Margarita St","NY"),
    (3,"13111 Siemon Ave","CA")]
df =spark.createDataFrame(address,[
df.show()
```

1. PySpark Replace String Column Values

By using PySpark SQL function `regexp_replace()` you can replace a column value with a string for another string/substring. `regexp_replace()` uses **Java regex** for matching, if the regex does not match it returns an empty string, the below example replace the street name `Rd` value with `Road` string on `address` column.

```
#Replace part of string with anothe
from pyspark.sql.functions import r
df.withColumn('address', regexp_rep
    .show(truncate=False)

#+---+------------------+-----+
#|id |address           |state|
#+---+------------------+-----+
#|1  |14851 Jeffrey Road|DE   |
#|2  |43421 Margarita St|NY   |
#|3  |13111 Siemon Ave  |CA   |
#+---+------------------+-----+
```

2. Replace Column Values Conditionally

In the above example, we just replaced `Rd` with `Road`, but not replaced `St` and `Ave` values, let's see how to replace column values conditionally in PySpark Dataframe by using when().otherwise() SQL condition function (https://sparkbyexamples.com/pyspark/pyspark-when-otherwise/).

```
#Replace string column value condit
from pyspark.sql.functions import w
df.withColumn('address',
    when(df.address.endswith('Rd'),
    .when(df.address.endswith('St'),
    .when(df.address.endswith('Ave')
    .otherwise(df.address)) \
    .show(truncate=False)


#+---+--------------------+-----+
#|id |address             |state|
#+---+--------------------+-----+
#|1  |14851 Jeffrey Road  |DE   |
#|2  |43421 Margarita Street|NY |
#|3  |13111 Siemon Avenue |CA   |
#+---+--------------------+-----+
```

3. Replace Column Value with Dictionary (map)

You can also replace column values from the **python dictionary (map)**. In the below example, we replace the string value of the `state` column with the full abbreviated name from a dictionary **key-value pair**, in order to do so I use PySpark map() transformation to loop through each row of DataFrame (https://sparkbyexamples.com/pyspark/pyspark-map-transformation/).

```
#Replace values from Dictionary
stateDic={'CA':'California','NY':'N
df2=df.rdd.map(lambda x:
    (x.id,x.address,stateDic[x.stat
    ).toDF(["id","address","state"]
df2.show()

#+---+-----------------+---------
#| id|          address|     state
#+---+-----------------+---------
#|  1|   14851 Jeffrey Rd|  Delaware
#|  2|43421 Margarita St|  New York
#|  3|   13111 Siemon Ave|California
#+---+-----------------+---------
```

4. Replace Column Value Character by Character

By using `translate()` string function you can **replace character by character of DataFrame column** value. In the below example, every character of **1** is replaced with **A**, **2** replaced with **B**, and **3** replaced with **C** on the `address` column.

```
#Using translate to replace charact
from pyspark.sql.functions import t
df.withColumn('address', translate(
  .show(truncate=False)

#+---+-----------------+-----+
#|id |address            |state|
#+---+-----------------+-----+
#|1  |A485A Jeffrey Rd   |DE   |
#|2  |4C4BA Margarita St|NY   |
#|3  |ACAAA Siemon Ave   |CA   |
#+---+-----------------+-----+
```

5. Replace Column with Another Column Value

By using `expr()` and `regexp_replace()` you can **replace column value with a value from another DataFrame column**. In the below example, we match the value from `col2` in `col1` and replace with `col3` to create `new_column`. Use expr() to provide SQL like expressions (https://sparkbyexamples.com/pyspark/pysp

ark-sql-expr-expression-function/) and is used to refer to another column to perform operations.

```
#Replace column with another column
from pyspark.sql.functions import e
df = spark.createDataFrame(
   [("ABCDE_XYZ", "XYZ","FGH")],
     ("col1", "col2","col3")
  )
df.withColumn("new_column",
             expr("regexp_replace(
             .alias("replaced_valu
             ).show()

#+---------+----+----+----------+
#|     col1|col2|col3|new_column|
#+---------+----+----+----------+
#|ABCDE_XYZ| XYZ| FGH| ABCDE_FGH|
#+---------+----+----+----------+
```

6. Replace All or Multiple Column Values

If you want to replace values on all or selected DataFrame columns, refer to How to Replace NULL/None values on all column in PySpark (https://sparkbyexamples.com/pyspark/pyspark-fillna-fill-replace-null-values/) or How to replace empty string with NULL/None value (https://sparkbyexamples.com/pyspark/pyspark-replace-empty-value-with-none-on-dataframe/)

7. Using overlay() Function

Replace column value with a string value from another column.

```
#Overlay
from pyspark.sql.functions import 
df = spark.createDataFrame([("ABCDE
df.select(overlay("col1", "col2", 7

#+---------+
#|overlayed|
#+---------+
#|ABCDE_FGH|
#+---------+
```

8. Complete Example

```python
from pyspark.sql import SparkSessid
spark = SparkSession.builder.master

address = [(1,"14851 Jeffrey Rd","D
    (2,"43421 Margarita St","NY"),
    (3,"13111 Siemon Ave","CA")]

df =spark.createDataFrame(address,[
df.show()

#Replace string
from pyspark.sql.functions import r
df.withColumn('address', regexp_rep
  .show(truncate=False)

#Replace string
from pyspark.sql.functions import w
df.withColumn('address',
      when(df.address.endswith('Rd'
      .when(df.address.endswith('St'
      .when(df.address.endswith('Ave
      .otherwise(df.address)) \
      .show(truncate=False)


#Replace values from Dictionary
stateDic={'CA':'California','NY':'N
df2=df.rdd.map(lambda x:
    (x.id,x.address,stateDic[x.stat
    ).toDF(["id","address","state"]
df2.show()

#Using translate
from pyspark.sql.functions import t
df.withColumn('address', translate(
  .show(truncate=False)

#Replace column with another column
from pyspark.sql.functions import e
df = spark.createDataFrame([("ABCDE
df.withColumn("new_column",
            expr("regexp_replace(
            .alias("replaced_valu
            ).show()

#Overlay
from pyspark.sql.functions import o
df = spark.createDataFrame([("ABCDE
df.select(overlay("col1", "col2", 7
```

Conclusion

In conclusion `regexp_replace()` function
is used to replace a string in a DataFrame
column with another value, `translate()`
function to replace character by character
of column values, `overlay()` function to

overlay string with another column string from start position and number of characters. Finally, you have also learned how to replace column values from a dictionary using Python examples.

Happy Learning !!

References

- https://kb.databricks.com/data/null-empty-strings.html (https://kb.databricks.com/data/null-empty-strings.html)

TAGS: OVERLAY() (HTTPS://SPARKBYEXAMPLES.COM/TAG/OVERLAY/), REGEXP_REPLACE() (HTTPS://SPARKBYEXAMPLES.COM/TAG/REGEXP_REPLACE/)

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

Enter your comment here...

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-sql-like-using-wildcard-example/)

Spark isin() & IS NOT IN Operator Example
(https://sparkbyexamples.com/spark-isin-is-not-in-operator-example/)

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

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

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

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

Spark Check Column Data Type is Integer
or String
(https://sparkbyexamples.com/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: (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/)