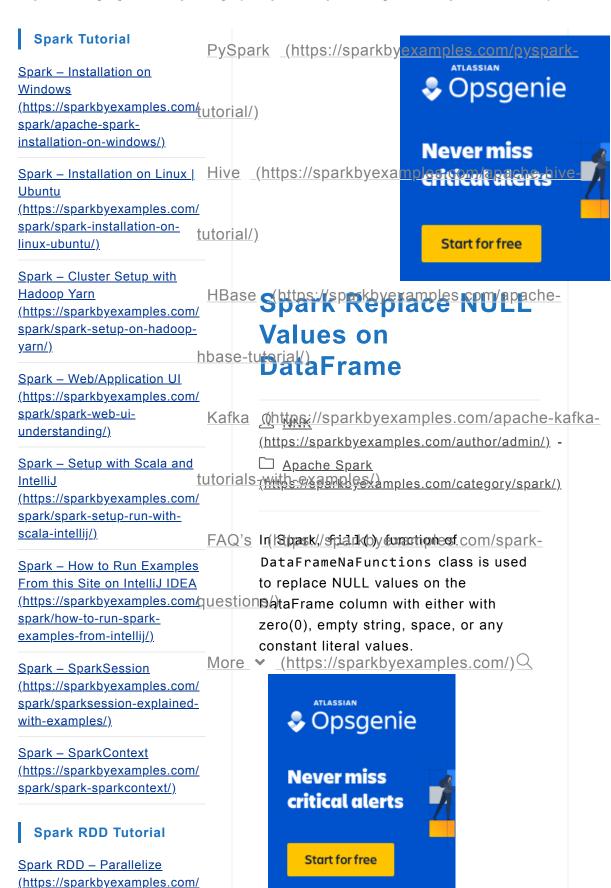
apache-spark-rdd/how-tocreate-an-rdd-using-

parallelize/)

Spark by {ExampleSpathtypshttpsp/tsplatkbyexamptels.som()m/)



(i)

<u>Spark RDD – Read text file</u> (https://sparkbyexamples.com/ apache-spark-rdd/spark-readmultiple-text-files-into-a-singlerdd/)

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

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

<u>Spark RDD – Create Empty</u> <u>RDD</u>

(https://sparkbyexamples.com/ apache-spark-rdd/spark-howto-create-an-empty-rdd/)

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

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

<u>Spark RDD – Pair Functions</u> (https://sparkbyexamples.com/ apache-spark-rdd/spark-pairrdd-functions/)

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

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

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

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

While working on Spark DataFrame we often need to replace null values as certain operations on null values return NullpointerException hence, we need to graciously handle nulls as the first step before processing. Also, while writing to a file, it's always best practice to replace null values, not doing this result nulls on the output file.

As part of the cleanup, some times you may need to Drop Rows with NULL
Values in Spark DataFrame
https://sparkbyexamples.com/spark/spark/sparkbyexamples.com/spark/spark/sparkbyexamples.com/sparkbyexamples.com/sparkbye

DataFrame API provides
DataFrameNaFunctions class with
fill() function to replace null values
on DataFrame. This function has several
overloaded signatures that take different
data types as parameters.

k-filter-rows-with-null-values/)

In this article, we use a subset of these and learn different ways to replace null values with an empty string, constant value, and zero(0) on Dataframe columns integer, string, array, and map with Scala examples.



- Replace NULL Values with Zero (0)
- Replace NULL Values with Empty String

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

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

Spark RDD – Convert RDD to DataFrame (https://sparkbyexamples.com/apache-spark-rdd/convert-spark-rdd-to-dataframe-dataset/)

Spark SQL Tutorial

Spark SQL - Create

DataFrame
(https://sparkbyexamples.com/
spark/different-ways-to-createa-spark-dataframe/)

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

<u>Spark SQL – Add and Update</u>
<u>Column (withColumn)</u>
(https://sparkbyexamples.com/
spark/spark-dataframewithcolumn/)

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

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

<u>Spark SQL – Where | Filter</u> (https://sparkbyexamples.com/ spark/spark-dataframe-wherefilter/)

<u>Spark SQL – When Otherwise</u> (https://sparkbyexamples.com/ <u>spark/spark-case-when-otherwise-example/)</u> Replace NULL on List and Map

Before we start, Let's read a CSV into
Spark DataFrame
(https://sparkbyexamples.com/spark/spark-read-csv-file-into-dataframe/) file,
where we have no values on certain
rows of String and Integer columns,
spark assigns null values to these no
value columns.

The file we are using here is available at GitHub small_zipcode.csv (https://github.com/spark-examples/spark-scala-examples/blob/master/src/main/resources/small_zipcode.csv)

```
val filePath="src/main/resources
val df=spark.read.options(
   Map("inferSchema"->"true","del
   .csv(filePath)
   df.printSchema()
   df.show(false)
```

This yields the below output. As you see columns type, city and population columns have null values.

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

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

<u>Spark SQL- Pivot Table</u>
<u>DataFrame</u>
(https://sparkbyexamples.com/
spark/how-to-pivot-table-andunpivot-a-spark-dataframe/)</u>

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

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

<u>Spark SQL – Schema</u> (https://sparkbyexamples.com/ spark/spark-schema-explainedwith-examples/)

<u>Spark SQL – Groupby</u> (https://sparkbyexamples.com/ spark/using-groupby-ondataframe/)

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

<u>Spark SQL – Join Types</u> (https://sparkbyexamples.com/ spark/spark-sql-dataframejoin/)

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

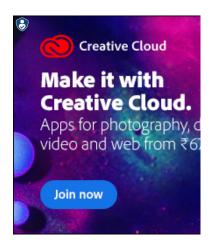
<u>Spark SQL - map() vs</u> <u>mapPartitions()</u> (https://sparkbyexamples.com/ <u>spark/spark-map-vs-</u> <u>mappartitions-transformation/)</u>

```
root
|-- id: integer (nullable = tru
|-- zipcode: integer (nullable
|-- type: string (nullable = tr
|-- city: string (nullable = tr
|-- state: string (nullable = t
 |-- population: integer (nullab
+---+-----
|id |zipcode|type |city
1 | 704
         |STANDARD|null
2 | 704
         | null | PASEO COST
3 | 709
         null
                  BDA SAN LU
4 | 76166 | UNIQUE | CINGULAR W
|5 |76177 |STANDARD|null
+---+-----
```

Now, let's see how to replace these null values.

Spark Replace NULL Values with Zero (0)

Spark fill(value:Long) signatures that are available in DataFrameNaFunctions is used to replace NULL values with numeric values either zero(0) or any constant value for all integer and long datatype columns of Spark DataFrame or Dataset.



Syntax:

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

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

<u>Spark SQL – Persist and</u>
<u>Cache</u>
(https://sparkbyexamples.com/spark/spark-dataframe-cache-

and-persist-explained/)

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

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

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

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

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

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

<u>Spark SQL – Sampling</u> (https://sparkbyexamples.com/ spark/spark-sampling-withexamples/)

for example

```
//Replace all integer and long c
df.na.fill(0)
    .show(false)

//Replace with specific columns
df.na.fill(0,Array("population")
    .show(false)
```

Above both statements yields the same below output. Note that it replaces only Integer columns.

Spark Replace Null

Values with Empty String

Spark fill(value:String) signatures are used to replace null values with an empty string or any constant values
String on DataFrame or Dataset columns.

Syntax:

<u>Spark SQL – Partitioning</u> (https://sparkbyexamples.com/ spark/spark-partitioningunderstanding/)</u>

Spark SQL Functions

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

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

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

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

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

<u>Spark SQL Aggregate</u>
<u>Functions</u>
(https://sparkbyexamples.com/spark/spark-sql-aggregate-functions/)

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

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

Spark Data Source API

<u>Spark – Read & Write CSV file</u> (https://sparkbyexamples.com/ spark/spark-read-csv-file-intodataframe/)

<u>Spark – Read and Write JSON</u> <u>file</u> (https://sparkbyexamples.com/

The first syntax replaces all nulls on all String columns with a given value, from our example it replaces nulls on columns type and city with an empty string.

```
df.na.fill("").show(false)
```

Yields below output. This replaces all NULL values with empty/blank string

Now, let's use the second syntax to replace the specific value on specific columns, below example replace column type with empty string and column city with value "unknown".

Yields below output. This replaces null values with an empty string for type column and replaces with a constant value "unknown" for city column.





<u>spark/spark-read-and-write-json-file/)</u>

<u>Spark – Read & Write Parquet</u> <u>file</u> <u>(https://sparkbyexamples.com/</u>

dataframe-parquet-example/)

spark/spark-read-write-

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

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

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

<u>Spark – Read & Write HBase</u> <u>using "hbase-spark" Connector</u> (https://sparkbyexamples.com/ <u>spark/spark-read-write-using-hbase-spark-connector/)</u>

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

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

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

Spark Streaming & Kafka

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

4	.	+	
id	zipcode	type	city
		+ STANDARD	
2	704		PASEO COST
3	709		BDA SAN LU
4	76166	UNIQUE	CINGULAR W
5	76177	STANDARD	unknown
+	+	+	+
4			•

Replace NULL Values on List and Map Columns

Compelte Code

Below is complete code with Scala example. You can use it by copying it from here or use the GitHub to download the source code.



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

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

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

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

<u>Spark SQL Batch – Consume & Produce Kafka Message</u>
(https://sparkbyexamples.com/
spark/spark-batch-processingproduce-consume-kafka-topic/)

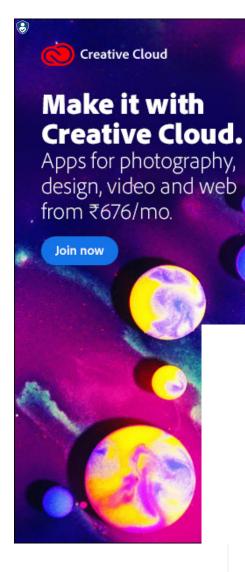


```
package com.sparkbyexamples.spar
import org.apache.spark.sql.Spar
object HandleNullExample extends
 val spark: SparkSession = Spar
    .master("local[1]")
    .appName("SparkByExamples.co
    .getOrCreate()
 val filePath="src/main/resourc
 val df = spark.read.options(Ma)
 df.printSchema()
 df.show(false)
 df.na.fill(0)
    .show(false)
 df.na.fill(0,Array("population")
    .show(false)
 df.na.fill("")
    .show(false)
 df.na.fill("unknown", Array("ci
    .na.fill("",Array("type"))
    .show(false)
 // Array and map columns
```

ps://github.com/sparkimples/spark-scalaimples/blob/master/src/main/scala/co
m/sparkbyexamples/spark/dataframe/Ha
ndleNullExample.scala) project for
reference.

Conclusion

In this Spark article, you have learned how to replace null values with zero or an empty string on integer and string



umns respectively.

anks for reading. If you recognize my ort or like articles here please do nment or provide any suggestions for provements in the comments stions!

ppy Learning !!

re this:

(https://sparkbyexamples.com/spark/spark-w-to-replace-null-values/?are=facebook&nb=1)

(https://sparkbyexamples.com/spark/sparkw-to-replace-null-values/?share=reddit&nb=1)

(https://sparkbyexamples.com/spark/sparkw-to-replace-null-values/?
are=pinterest&nb=1)

(https://sparkbyexamples.com/spark/sparkw-to-replace-null-values/?
are=tumblr&nb=1)

(https://sparkbyexamples.com/spark/sparkw-to-replace-null-values/?
are=pocket&nb=1)

(https://sparkbyexamples.com/spark/spark-now-to-replace-null-values/?share=linkedin&nb=1)

(https://sparkbyexamples.com/spark/sparkhow-to-replace-null-values/? share=twitter&nb=1)

TAGS: NULL (HTTPS://SPARKBYEXAMPLES.COM/TAG/NULL/)



(https://sp arkbyexa mples.co m/author/

admin/)

<u>NNK</u>

(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/)

Best Pediatrician in Bangalore

Cloudnine Hospital - Bellandur

Cloudnine hospital, is ranked as #1 maternity hospital.





STORE INFO



Categories

Apache Hadoop (https://sparkbyexamples.com/category/hadoop/)

Apache Spark (https://sparkbyexamples.com/categor y/spark/)

Apache Spark Streaming (https://sparkbyexamples.com/category/spark/apache-spark-streaming/)

Apache Kafka (https://sparkbyexamples.com/categor y/kafka/)

Apache HBase (https://sparkbyexamples.com/category/hbase/)

Apache Cassandra (https://sparkbyexamples.com/category/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/)

Recent Posts

Spark regexp_replace() - Replace String Value

(https://sparkbyexamples.com/spark/spar k-regexp_replace-replace-string-value/)

How to Run a PySpark Script from Python? (https://sparkbyexamples.com/pyspark/ru n-pyspark-script-from-pythonsubprocess/)

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/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-dataframecolumn/)

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/spar k-check-string-column-has-numericvalues/) 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:/ in

/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/)

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

Copyright sparkbyexamples.com