# Spark by {Examplesparkhttpthttps://aptalkbyexamplescom/pm/)

#### **PySpark Tutorial**

Py\$park (https://sparkbyexamples.com/pyspark-

PySpark Tutorial For

Beginners

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

PySpark - Features

pyspark-tutorial/#features)

PySpark - Advantages

(https://sparkbyexamples.com/ pyspark-tutorial/#advantages)

# Incredible Savings

Data Science-Amazing new deal on e know re: data science

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

humblebundle.com

### Incredible Savings Data Science Mercury Le

Data Science-Amazing new de everything U need to know rescience

humblebundle.com

OPFN

PySpark - Modules &

Packages

(https://sparkbyexamples.com/

pyspark-tutorial/#modules-

<u>packages)</u>

HBase p(https://sparkbyexamples.com/apache-

**Empty DataFrame &** 

hbase-tutarial()

PySpark - Cluster Managers

pyspark-tutorial/#cluster-

manager)

(https://sparkbyexamples.com/ Kafka (https://sparkbyexamples.com/apache-(https://sparkbyexamples.com/author/admin/) -

PySpark

PySpark - Install on Windowskafka-tutorials/withkeyann best com/category/pyspark/)

(https://sparkbyexamples.com/

pyspark-tutorial/#pyspark-

installation)

PySpark - Web/Application UI

(https://sparkbyexamples.com/

spark/spark-web-ui-

understanding/)

FAQ's Inhibipsattionerklowile exphaine bovo to/sparkcreate an empty PySpark

DataFrame/RDD manually with or

questionwithout schema (column names) in different ways. Below I have explained

one of the many scenarios where we

PySpark - SparkSession (https://sparkbyexamples.com/ More Mee(Inthoridate) pyspark/pyspark-what-issparksession/)

PySpark - RDD

(https://sparkbyexamples.com/

pyspark-rdd)

PySpark - Parallelize (https://sparkbyexamples.com/ pyspark/pyspark-parallelizecreate-rdd/)

PySpark - repartition() vs coalesce() (https://sparkbyexamples.com/ pyspark/pyspark-repartition-

PySpark - Broadcast

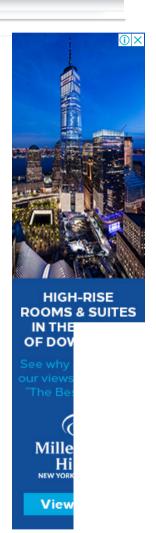
vs-coalesce/)

(https://sparkbyexamples.com/ pyspark/pyspark-broadcastvariables/)



While working with files, sometimes we may not receive a file for processing. however, we still need to create a

Valiables pookies to ensure that we give you the best experience on our website. If you continue to use this site we will assume that you are happy schema we expect with five don't create with the same schema, our



PySpark - Accumulator (https://sparkbyexamples.com/ pyspark/pyspark-accumulatorwith-example/)

#### **PySpark DataFrame**

PvSpark - Create a **DataFrame** 

(https://sparkbyexamples.com/ pyspark/different-ways-tocreate-dataframe-in-pyspark/)

PySpark - Create an empty DataFrame

(https://sparkbyexamples.com/ pyspark/pyspark-create-anempty-dataframe/)

PySpark - Convert RDD to **DataFrame** 

(https://sparkbyexamples.com/ pyspark/convert-pyspark-rddto-dataframe/)

PySpark - Convert DataFrame to Pandas

(https://sparkbyexamples.com/ pyspark/convert-pysparkdataframe-to-pandas/)

PySpark - show() (https://sparkbyexamples.com/ pyspark/pyspark-show-displaydataframe-contents-in-table/)

PySpark - StructType & StructField (https://sparkbyexamples.com/ pyspark/pyspark-structtypeand-structfield/)

PySpark - Row Class (https://sparkbyexamples.com/ pyspark/pyspark-row-usingrdd-dataframe/)

PySpark - Column Class (https://sparkbyexamples.com/ pyspark/pyspark-columnfunctions/)

PySpark - select() (https://sparkbyexamples.com/ pyspark/select-columns-frompyspark-dataframe/)

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

pyspark/pyspark-withcolumn/)

operations/transformations (like union's) on DataFrame fail as we refer to the columns that may not present.

To handle situations similar to these. we always need to create a DataFrame with the same schema, which means the same column names and datatypes regardless of the file exists or empty file processing.

# 1. Create Empty RDD in **PvSpark**

Create an empty RDD by using emptyRDD() of SparkContext for example

spark.sparkContext.emptyRDD().

from pyspark.sql import SparkSe spark = SparkSession.builder.ap #Creates Empty RDD emptyRDD = spark.sparkContext.e print(emptyRDD) #Diplays #EmptyRDD[188] at emptyRDD

Alternatively you can also get empty RDD by using spark.sparkContext.parallelize( []).

# **Earn Money from** Crypto

PySpark – withColumn()
We use cookies to ensure that we give you the best experience on our website. If you continue to use this site we will assume that you are happy (https://sparkbyexamples.com/ with it

Accept

<u>PySpark –</u>
withColumnRenamed()
(https://sparkbyexamples.com/
pyspark/pyspark-renamedataframe-column/)

<u>PySpark – where() & filter()</u> (https://sparkbyexamples.com/ pyspark/pyspark-where-filter/)

PySpark – drop() & dropDuplicates() (https://sparkbyexamples.com/pyspark/pyspark-distinct-to-drop-duplicates/)

<u>PySpark – orderBy() and sort()</u> (https://sparkbyexamples.com/ <u>pyspark/pyspark-orderby-and-</u> sort-explained/)

<u>PySpark – groupBy()</u>
(https://sparkbyexamples.com/
pyspark/pyspark-groupbyexplained-with-example/)</u>

<u>PySpark – join()</u>
(https://sparkbyexamples.com/
<u>pyspark/pyspark-join-</u>
explained-with-examples/)

<u>PySpark – union() & unionAll()</u> (https://sparkbyexamples.com/ <u>pyspark/pyspark-union-and-unionall/)</u>

<u>PySpark – unionByName()</u>
(https://sparkbyexamples.com/
spark/spark-merge-twodataframes-with-differentcolumns/)

<u>PySpark – UDF (User Defined Function)</u>
(https://sparkbyexamples.com/pyspark/pyspark-udf-user-defined-function/)

<u>PySpark – map()</u>
(https://sparkbyexamples.com/
pyspark/pyspark-maptransformation/)</u>

<u>PySpark – flatMap()</u>
(<u>https://sparkbyexamples.com/pyspark/pyspark-flatmap-</u>
transformation/)

<u>pyspark – foreach()</u>
(https://sparkbyexamples.com/
<u>pyspark/pyspark-loop-iterate-</u>
through-rows-in-

dataframe/#use-foreach-loop give you the through-dataframe)

<u>PySpark – sample() vs</u> <u>sampleBy()</u>

```
#Creates Empty RDD using paralle
rdd2= spark.sparkContext.paralle
print(rdd2)

#EmptyRDD[205] at emptyRDD at Note
#ParallelCollectionRDD[206] at
```

**Note:** If you try to perform operations on empty RDD you going to get ValueError("RDD is empty").

# 2. Create Empty

# DataFrame with Schema

# (StructType)

In order to create an empty PySpark

DataFrame manually with schema (
column names & data types) first,

Create a schema using StructType and

StructField

(https://oparkhysysemples.com/pycnark/

(https://sparkbyexamples.com/pyspark/ pyspark-structtype-and-structfield/) .

```
#Create Schema
from pyspark.sql.types import S
schema = StructType([
    StructField('firstname', Stri
    StructField('middlename', Str
    StructField('lastname', Stri
    ])
```

and pass it to createDataFrame() of <a href="SparkSession">SparkSession</a>
(https://sparkbyexamples.com/pyspark/
pyspark-what-is-sparksession/) along
with the schema for column names & data types.

Now use the empty RDD created above

```
#Create empty DataFrame from em
df = spark.createDataFrame(empt)
df.printSchema()
```

site. If you continue to ese this site we will assume that you are happy

This yields below schema of the empty DataFrame. Accept

(https://sparkbyexamples.com/ pyspark/pyspark-samplingexample/)

PySpark – fillna() & fill() (https://sparkbyexamples.com/ pyspark/pyspark-fillna-fillreplace-null-values/)

<u>PySpark – pivot() (Row to</u> Column)

(https://sparkbyexamples.com/ pyspark/pyspark-pivot-andunpivot-dataframe/)

<u>PySpark – partitionBy()</u>
(https://sparkbyexamples.com/
<u>pyspark/pyspark-partitionby-example/)</u>

<u>PySpark – ArrayType Column</u> (Array)

(https://sparkbyexamples.com/ pyspark/pyspark-arraytypecolumn-with-examples/)

PySpark — MapType (Map/Dict) (https://sparkbyexamples.com/ pyspark/pyspark-maptype-dictexamples/)

#### **PySpark SQL Functions**

<u>PySpark – Aggregate Functions</u> (<u>https://sparkbyexamples.com/pyspark/pyspark-aggregate-functions/</u>)

<u>PySpark – Window Functions</u> (<u>https://sparkbyexamples.com/pyspark/pyspark-window-functions/</u>)

<u>PySpark – Date and Timestamp</u> <u>Functions</u>

(https://sparkbyexamples.com/py spark/pyspark-sql-date-andtimestamp-functions/)

PySpark – JSON Functions (https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/)

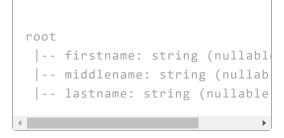
#### **PySpark Datasources**

<u>PySpark – Read & Write CSV</u> <u>File</u>

(https://sparkbyexamples.com/ pyspark/pyspark-read-csv-file-

into-dataframe/) to ensure that we give you the best experience on our website. If you continue to use this site we will assume that you are happy with it

PySpark - Read & Write
Parquet File
(https://sparkbyexamples.com/



# 3. Convert Empty RDD to

#### **DataFrame**

You can also create empty DataFrame by converting empty RDD to DataFrame using toDf().



#Convert empty RDD to Dataframe
df1 = emptyRDD.toDF(schema)
df1.printSchema()

# 4. Create Empty

#### DataFrame with Schema.

So far I have covered creating an empty DataFrame from RDD, but here will create it manually with schema and without RDD.

#Create empty DataFrame directly
df2 = spark.createDataFrame([],
df2.printSchema()

<u>pyspark/pyspark-read-and-write-parquet-file/)</u>

<u>PySpark – Read & Write JSON</u> file

(https://sparkbyexamples.com/ pyspark/pyspark-read-jsonfile-into-dataframe/)

# Earn Money from Crypto

Yamgo - Get Paid For Life

# PySpark Built-In Functions

<u>PySpark – when()</u> (https://sparkbyexamples.com/py spark/pyspark-when-otherwise/)</u>

<u>PySpark - expr()</u>
(https://sparkbyexamples.com/py
spark/pyspark-sql-exprexpression-function/)

<u>PySpark – lit()</u>

(https://sparkbyexamples.com/py spark/pyspark-lit-add-literalconstant/)

PySpark - split()

(https://sparkbyexamples.com/py spark/pyspark-convert-string-toarray-column/)

PySpark – concat\_ws() (https://sparkbyexamples.com/pyspark/pyspark-convert-array-column-to-string-column/)

<u>Pyspark – substring()</u>
(https://sparkbyexamples.com/pyspark/pyspark-substring-from-a-column/)</u>

PySpark – translate() (https://sparkbyexamples.com/pyspark/pyspark-replace-column-values/#translate-replace-

character-by-character)
We use cookies to ensure that we give you

<u>PySpark – regexp\_replace()</u> (https://sparkbyexamples.com/py spark/pyspark-replace-column-

# 5. Create Empty

# DataFrame without Schema (no columns)

To create empty DataFrame with out ema (no columns) just create a ety schema and use it while creating park DataFrame.

freate empty DatFrame with no
f3 = spark.createDataFrame([],
f3.printSchema()

print below empty schema
#root

#### Happy Learning !!

#### Share this:



(https://sparkbyexamples.com/pyspark/pysparkcreate-an-empty-dataframe/? share=facebook&nb=1)



(https://sparkbyexamples.com/pyspark/pyspark-create-an-empty-dataframe/?share=reddit&nb=1)

P

(https://sparkbyexamples.com/pyspark/pysparkcreate-an-empty-dataframe/? share=pinterest&nb=1)



(https://sparkbyexamples.com/pyspark/pyspark-create-an-empty-dataframe/?share=tumblr&nb=1)



(https://sparkbyexamples.com/pyspark/pyspark-create-an-empty-dataframe/?share=pocket&nb=1)

in

(https://sparkbyexamples.com/pyspark/pyspark-create-an-empty-dataframe/?share=linkedin&nb=1)



(https://sparkbyexamples.com/pyspark/pyspark-create-an-empty-dataframe/?share=twitter&nb=1)

TAGS: EMPTY DATAFRAME

(HTTPS://SPARKBYEXAMPLES.COM/TAG/EMPTY-DATAFRAME/), EMPTYRDD().

the best experience on our website. If you continue to use this site we will assume that you are happy D-2/).

Accept

values/#regexp\_replace-replacestring-columns)

PySpark - overlay()

(https://sparkbyexamples.com/py spark/pyspark-replace-columnvalues/#overlay-function)

PySpark - to timestamp() (https://sparkbyexamples.com/sp ark/pyspark-to\_timestampconvert-string-to-timestamptype/)

PySpark - to date() (https://sparkbyexamples.com/py spark/pyspark-to date-converttimestamp-to-date/)

PySpark - date format() (https://sparkbyexamples.com/py spark/pyspark-date formatconvert-date-to-string-format/)

PySpark - datediff() (https://sparkbyexamples.com/py spark/pyspark-differencebetween-two-dates-daysmonths-years/#datediff)

PySpark - months between() (https://sparkbyexamples.com/py spark/pyspark-differencebetween-two-dates-daysmonthsyears/#months between())

PySpark - explode() (https://sparkbyexamples.com/py spark/pyspark-explode-nestedarray-into-rows/)

PySpark - array contains() (https://sparkbyexamples.com/py spark/pyspark-arraytype-columnwith-examples/#array\_contains)

PySpark - array() (https://sparkbyexamples.com/py spark/pyspark-arraytype-columnwith-examples/#array)

PySpark - collect\_list() (https://sparkbyexamples.com/py spark/pyspark-aggregatefunctions/#collect-list)

PySpark - collect\_set() (https://sparkbyexamples.com/py spark/pyspark-aggregatefunctions/#collect-set)

(https://sparkbyexamples.com/py spark/pyspark-convertdataframe-columns-to-maptype-

PySpark - create map()

dict/)



(https://sp <u>arkbyexa</u> mples.co m/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...

with it

Accept

PySpark - map\_keys()
(https://sparkbyexamples.com/py
spark/pyspark-maptype-dictexamples/#map\_keys)

PySpark – map\_values() (https://sparkbyexamples.com/pyspark/pyspark-maptype-dict-examples/#map\_values)

PySpark – struct() (https://sparkbyexamples.com/py spark/pyspark-structtype-andstructfield/#update-structfunction)

<u>PySpark – countDistinct()</u>
(https://sparkbyexamples.com/pyspark/pyspark-count-distinct-from-dataframe/)</u>

PySpark – sum(), avg() (https://sparkbyexamples.com/py spark/pyspark-dataframegroupby-and-sort-bydescending-order/)

PySpark – row\_number()
(https://sparkbyexamples.com/py
spark/pyspark-windowfunctions/#row\_number)

<u>PySpark - rank()</u>
(https://sparkbyexamples.com/py
spark/pyspark-windowfunctions/#rank)

PySpark – dense\_rank()
(https://sparkbyexamples.com/py
spark/pyspark-windowfunctions/#dense\_rank)

PySpark - percent\_rank()
(https://sparkbyexamples.com/py
spark/pyspark-windowfunctions/#percent\_rank)

<u>PySpark - typedLit()</u>
(https://sparkbyexamples.com/py
spark/pyspark-lit-add-literalconstant/#typedlit)</u>

PySpark – from\_json() (https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/#from\_json)

PySpark – to\_json() (https://sparkbyexamples.com/py spark/pyspark-json-functionswith-examples/#to\_json)

PySpark - json\_tuple() ure that we give you the best experience on our website. If you continue to use this site we will assume that you are happy (https://sparkbyexamples.com/py with it.

spark/pyspark-json-functions-

with-examples/#json\_tuple)

PySpark – get json object()
(https://sparkbyexamples.com/py
spark/pyspark-json-functionswith-examples/#get json object)

PySpark – schema\_of\_json()
(https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-

examples/#schema of json)

#### Categories

Apache Hadoop (https://sparkbyexamples.com/catego ry/hadoop/)

Apache Spark (https://sparkbyexamples.com/catego ry/spark/)

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

Apache Kafka (https://sparkbyexamples.com/catego ry/kafka/)

Apache HBase (https://sparkbyexamples.com/catego ry/hbase/)

Apache Cassandra (https://sparkbyexamples.com/catego ry/cassandra/)

Snowflake Database (https://sparkbyexamples.com/category/snowflake/)

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

PySpark (https://sparkbyexamples.com/catego ry/pyspark/)

#### Recent Posts

Spark regexp\_replace() - Replace String Value (https://sparkbyexamples.com/spark/spark-regexp\_replace-replace-string-value/)

Python? (https://sparkbyexamples.com/pyspark/r un-pyspark-script-from-python-subprocess/)

How to Run a PySpark Script from

Spark SQL like() Using Wildcard
Example
(https://sparkbyexamples.com/spark

(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/sp ark-get-size-length-of-array-mapcolumn/)

Spark Using Length/Size Of a

DataFrame Column
(https://sparkbyexamples.com/spark/sp
ark-using-length-size-of-a-dataframe-

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/sp ark-check-string-column-has-numericvalues/)

Spark Check Column Data Type is Integer or String (https://sparkbyexamples.com/spark/spark-check-column-data-type-is-integerAbout 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

4

**(7**)

(https:/ in

(https://acebo /www.li (https:/

/www.f (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/)

We use cookies to ensure that we give you the best experience on our website. If you continue to use this site we will assume that you are happy with it.