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 – Accumulator (https://sparkbyexamples.com/pyspark/pyspark-accumulator-with-example/)

PySpark DataFrame

PySpark – Create a DataFrame (https://sparkbyexamples.com/py

Succeed online.

Our helpful guides are here 24/7 to get your business idea online.

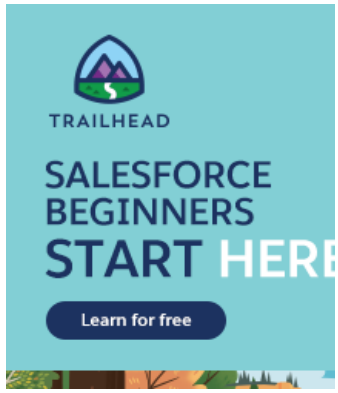
in.godaddy.com

PySpark – Difference between two dates (days, months, years)

NNK (https://sparkbyexamples.com/author/admin/) -

PySpark (https://sparkbyexamples.com/category/pyspark/)

Using PySpark SQL functions `datediff()`, `months_between()` you can calculate the difference between two dates in days, months, and year, let's see this by using a DataFrame example. You can also use these to calculate age.



datediff() Function

spark/different-ways-to-create-dataframe-in-pyspark/)

PySpark – Create an empty DataFrame (https://sparkbyexamples.com/pyspark/pyspark-create-an-empty-dataframe/)

PySpark – Convert RDD to DataFrame (https://sparkbyexamples.com/pyspark/convert-pyspark-rdd-to-dataframe/)

PySpark – Convert DataFrame to Pandas (https://sparkbyexamples.com/pyspark/convert-pyspark-dataframe-to-pandas/)

PySpark – show() (https://sparkbyexamples.com/pyspark/pyspark-show-display-dataframe-contents-in-table/)

PySpark – StructType & StructField (https://sparkbyexamples.com/pyspark/pyspark-structtype-and-structfield/)

PySpark – Row Class (https://sparkbyexamples.com/pyspark/pyspark-row-using-rdd-dataframe/)

PySpark – Column Class (https://sparkbyexamples.com/pyspark/pyspark-column-functions/)

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

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

PySpark – withColumn() (https://sparkbyexamples.com/pyspark/pyspark-withcolumn/)

PySpark – withColumnRenamed() (https://sparkbyexamples.com/pyspark/pyspark-rename-dataframe-column/)

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

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

PySpark – orderBy() and sort() (https://sparkbyexamples.com/py

First Let's see getting the difference between two dates using `datediff()` PySpark function.

```
from pyspark.sql.functions import *
data = [("1","2019-07-01"),("2","20
df=spark.createDataFrame(data=data,

df.select(
      col("date"),
      current_date().alias("current
      datediff(current_date(),col("
    ).show()
```

Output:

```
+----------+------------+--------+
|      date|current_date|datediff|
+----------+------------+--------+
|2019-07-01|  2021-02-26|     606|
|2019-06-24|  2021-02-26|     613|
|2019-08-24|  2021-02-26|     552|
+----------+------------+--------+
```

months_between() Function

Now, Let's see how to get month and year differences between two dates using `months_between()` function.

```
from pyspark.sql.functions import *
df.withColumn("datesDiff", datediff
  .withColumn("montsDiff", months_b
  .withColumn("montsDiff_round",rou
  .withColumn("yearsDiff",months_be
  .withColumn("yearsDiff_round",rou
  .show()
```

Yields below output. Note that here we use round() function and lit() functions on top of months_between() to get the year between two dates.

spark/pyspark-orderby-and-sort-explained/)

PySpark – groupBy() (https://sparkbyexamples.com/pyspark/pyspark-groupby-explained-with-example/)

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

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

PySpark – unionByName() (https://sparkbyexamples.com/spark/spark-merge-two-dataframes-with-different-columns/)

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

PySpark – map() (https://sparkbyexamples.com/pyspark/pyspark-map-transformation/)

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

pyspark – foreach() (https://sparkbyexamples.com/pyspark/pyspark-loop-iterate-through-rows-in-dataframe/#use-foreach-loop-through-dataframe)

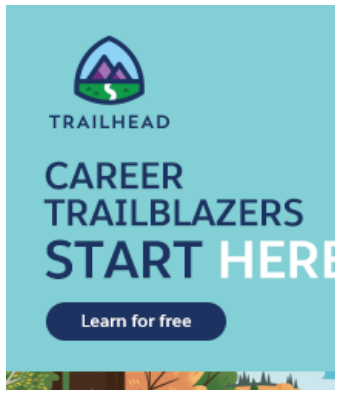
PySpark – sample() vs sampleBy() (https://sparkbyexamples.com/pyspark/pyspark-sampling-example/)

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

PySpark – pivot() (Row to Column) (https://sparkbyexamples.com/pyspark/pyspark-pivot-and-unpivot-dataframe/)

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

PySpark – ArrayType Column (Array) (https://sparkbyexamples.com/pyspark/pyspark-arraytype-column-with-examples/)




```
+---+----------+---------+---------
| id|      date|datesDiff|  montsDi
+---+----------+---------+---------
|  1|2019-07-01|      606|19.806451
|  2|2019-06-24|      613|20.064516
|  3|2019-08-24|      552|18.064516
+---+----------+---------+---------
```

Let's see another example of the difference between two dates when dates are not in PySpark DateType format `yyyy-MM-dd`. when dates are not in DateType format, all date functions return null. Hence, you need to first convert the input date to Spark DateType using `to_date()` function.

```
from pyspark.sql.functions import *
data2 = [("1","07-01-2019"),("2","0
df2=spark.createDataFrame(data=data
df2.select(
    to_date(col("date"),"MM-dd-yyyy
    current_date().alias("endDate")
    )
```

SQL Example

Let's see how to calculate the difference between two dates in years using PySpark SQL example. similarly you can calculate the days and months between two dates.

```
spark.sql("select round(months_betw
```

Complete Code:

PySpark – MapType (Map/Dict) (https://sparkbyexamples.com/pyspark/pyspark-maptype-dict-examples/)

PySpark SQL Functions

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

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

PySpark – Date and Timestamp Functions (https://sparkbyexamples.com/pyspark/pyspark-sql-date-and-timestamp-functions/)

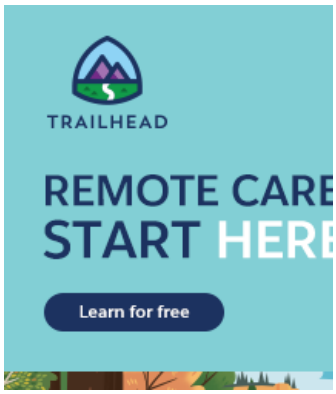
PySpark – JSON Functions (https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/)

PySpark Datasources

PySpark – Read & Write CSV File (https://sparkbyexamples.com/pyspark/pyspark-read-csv-file-into-dataframe/)

PySpark – Read & Write Parquet File (https://sparkbyexamples.com/pyspark/pyspark-read-and-write-parquet-file/)

PySpark – Read & Write JSON file (https://sparkbyexamples.com/pyspark/pyspark-read-json-file-into-dataframe/)




```
from pyspark.sql import SparkSessio

# Create SparkSession
spark = SparkSession.builder \
          .appName('SparkByExamples
          .getOrCreate()
data = [("1","2019-07-01"),("2","20

df=spark.createDataFrame(data=data,

from pyspark.sql.functions import *

df.select(
      col("date"),
      current_date().alias("current
      datediff(current_date(),col("
    ).show()

df.withColumn("datesDiff", datediff
      .withColumn("montsDiff", mont
      .withColumn("montsDiff_round"
      .withColumn("yearsDiff",month
      .withColumn("yearsDiff_round"
      .show()

data2 = [("1","07-01-2019"),("2","0
df2=spark.createDataFrame(data=data
df2.select(
    to_date(col("date"),"MM-dd-yyyy
    current_date().alias("endDate")
    )

spark.sql("select round(months_betw
```

Conclusion:

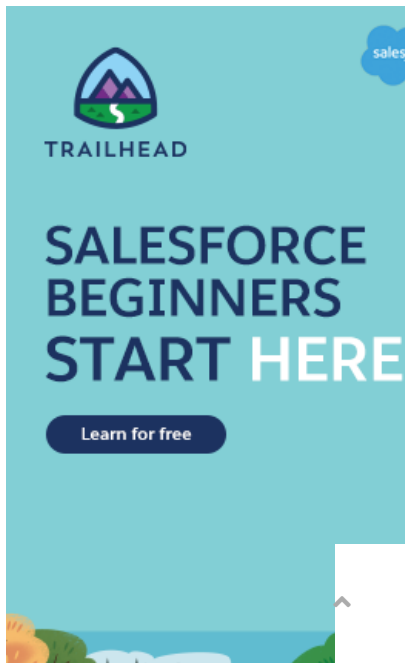
In this tutorial, you have learned how to alculate days, months, and years between vo dates using PySpark Date and Time unctions (https://sparkbyexamples.com/pyspark/pyspark-sql-date-and-timestamp-functions/) atediff(), months_between(). You can find ıore information about these functions at ıe following blog (https://databricks.com/blog/2015/09/16/apache-spark-1-5-dataframe-api-highlights.html)

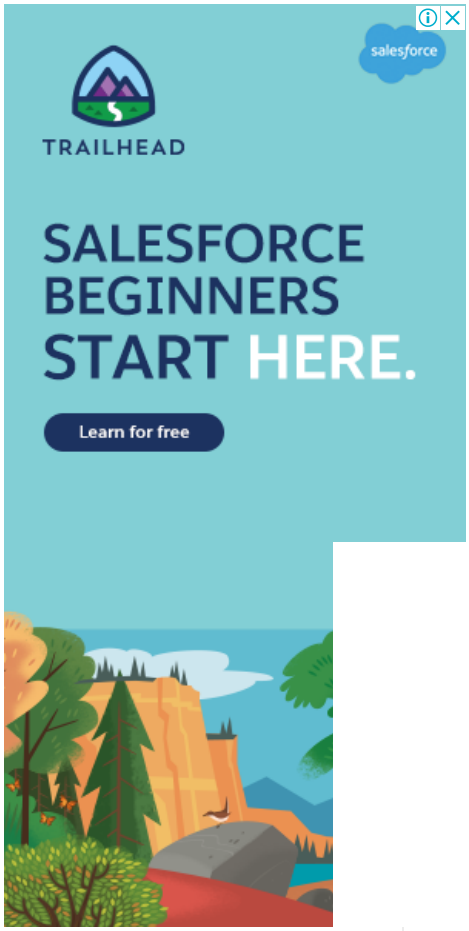
Happy Learning !!

Related Articles:

- PySpark SQL – How to Get Current Date & Timestamp

אם המדע גילה
ת אלוהים





(https://sparkbyexamples.com/pyspark/pyspark-current-date-timestamp/)
PySpark SQL – Working with Unix Time (https://sparkbyexamples.com/pyspark/pyspark-sql-working-with-unix-time-timestamp/)
PySpark Timestamp Difference (seconds, minutes, hours) (https://sparkbyexamples.com/pyspark/pyspark-timestamp-difference-seconds-minutes-hours/)
PySpark SQL – Convert Date to String Format (https://sparkbyexamples.com/pyspark/pyspark-date_format-convert-date-to-string-format/)
PySpark SQL – Convert String to Date Format (https://sparkbyexamples.com/pyspark/pyspark-to_date-convert-string-to-date-format/)
PySpark SQL – Convert Timestamp to Date (https://sparkbyexamples.com/pyspark/pyspark-to_date-convert-timestamp-to-date/)

PySpark Built-In Functions

PySpark – when() (https://sparkbyexamples.com/pyspark/pyspark-when-otherwise/)

PySpark – expr() (https://sparkbyexamples.com/pyspark/pyspark-sql-expr-expression-function/)

PySpark – lit() (https://sparkbyexamples.com/pyspark/pyspark-lit-add-literal-constant/)

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

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

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

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

PySpark – regexp_replace() (https://sparkbyexamples.com/pyspark/pyspark-replace-column-

Share this:

(https://sparkbyexamples.com/pyspark/pyspark-difference-between-two-dates-days-months-years/?share=facebook&nb=1)

(https://sparkbyexamples.com/pyspark/pyspark-difference-between-two-dates-days-months-years/?share=reddit&nb=1)

(https://sparkbyexamples.com/pyspark/pyspark-difference-between-two-dates-days-months-years/?share=pinterest&nb=1)

(https://sparkbyexamples.com/pyspark/pyspark-difference-between-two-dates-days-months-years/?share=tumblr&nb=1)

(https://sparkbyexamples.com/pyspark/pyspark-difference-between-two-dates-days-months-years/?share=pocket&nb=1)

(https://sparkbyexamples.com/pyspark/pyspark-difference-between-two-dates-days-months-years/?share=linkedin&nb=1)

(https://sparkbyexamples.com/pyspark/pyspark-difference-between-two-dates-days-months-years/?share=twitter&nb=1)

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

(https://sparkbyexa

NNK (Https://Sparkbyexamples.Com/Author/Admin/)

SparkByExamples.com is a Big Data and Spark examples community page, all examples are simple and

values/#regexp_replace-replace-string-columns)

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

PySpark – to_timestamp() (https://sparkbyexamples.com/spark/pyspark-to_timestamp-convert-string-to-timestamp-type/)

PySpark – to_date() (https://sparkbyexamples.com/pyspark/pyspark-to_date-convert-timestamp-to-date/)

PySpark – date_format() (https://sparkbyexamples.com/pyspark/pyspark-date_format-convert-date-to-string-format/)

PySpark – datediff() (https://sparkbyexamples.com/pyspark/pyspark-difference-between-two-dates-days-months-years/#datediff)

PySpark – months_between() (https://sparkbyexamples.com/pyspark/pyspark-difference-between-two-dates-days-months-years/#months_between())

PySpark – explode() (https://sparkbyexamples.com/pyspark/pyspark-explode-nested-array-into-rows/)

PySpark – array_contains() (https://sparkbyexamples.com/pyspark/pyspark-arraytype-column-with-examples/#array_contains)

PySpark – array() (https://sparkbyexamples.com/pyspark/pyspark-arraytype-column-with-examples/#array)

PySpark – collect_list() (https://sparkbyexamples.com/pyspark/pyspark-aggregate-functions/#collect-list)

PySpark – collect_set() (https://sparkbyexamples.com/pyspark/pyspark-aggregate-functions/#collect-set)

PySpark – create_map() (https://sparkbyexamples.com/pyspark/pyspark-convert-dataframe-columns-to-maptype-dict/)

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

PySpark – map_values() (https://sparkbyexamples.com/pysp

mples.com/author/admin/)

easy to understand and well tested in our development environment Read more .. (https://sparkbyexamples.com/about-sparkbyexamples/)

Leave a Reply

Enter your comment here...

האם המדע גילה את אלוהים

מהן אותן תגליות מופלאות שגרמו למדענים לדבר פתאום על אלוהים

פתח

JesusT

ark/pyspark-maptype-dict-examples/#map_values)

PySpark – struct() (https://sparkbyexamples.com/pyspark/pyspark-structtype-and-structfield/#update-struct-function)

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

PySpark – sum(), avg() (https://sparkbyexamples.com/pyspark/pyspark-dataframe-groupby-and-sort-by-descending-order/)

PySpark – row_number() (https://sparkbyexamples.com/pyspark/pyspark-window-functions/#row_number)

PySpark – rank() (https://sparkbyexamples.com/pyspark/pyspark-window-functions/#rank)

PySpark – dense_rank() (https://sparkbyexamples.com/pyspark/pyspark-window-functions/#dense_rank)

PySpark – percent_rank() (https://sparkbyexamples.com/pyspark/pyspark-window-functions/#percent_rank)

PySpark – typedLit() (https://sparkbyexamples.com/pyspark/pyspark-lit-add-literal-constant/#typedlit)

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

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

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

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

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

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/)

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/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/spark-get-size-length-of-array-map-column/)

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

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/spark-check-string-column-has-numeric-values/)

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

Copyright 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://www.facebook.com/sparkbyexamples) (https://www.linkedin.com/in/n-k-b860a8193/) (https://twitter.com/sparkbyexamples) (https://github.com/spark-examples)