Spark by {Examplessom/n/)

PySpark Tutorial

PySpark _(https://sparkbyexamples.com/pyspark-

PySpark Tutorial For Beginners (https://sparkbyexamples.com/py spark-tutorial/)

tutorial/)

PySpark - Features

(https://sparkbyexamples.com/py

spark-tutorial/#features)

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

Best DSA Course

Course covering DS Algo, Comp Science Fundamentals, Low leve and System Design

PySpark - Advantages

(https://sparkbyexamples.com/py tutorial/)

spark-tutorial/#advantages)

PySpark - Modules & Packages

packages)

PySpark - Cluster Managers

(https://sparkbyexamples.com/py

spark-tutorial/#cluster-manager)

PySpark - Install on Windows

(https://sparkbyexamples.com/py

spark-tutorial/#pyspark-

installation)

(https://sparkbyexamples.com/py HBase phttps://sparkbyexamples.com/apache-hbase-spark-tutorial/#modules-

array column to a

Programming Pathshala

tutorial/) String

Kafka (https://sparkbyexamples.com/apache-kafka-

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

PySpark

tutorials-with-examples/amples.com/category/pyspark/)

PySpark - Web/Application UI (https://sparkbyexamples.com/sp

PySpark - SparkSession spark/pyspark-what-is-

sparksession/)

PySpark - RDD

(https://sparkbyexamples.com/py spark-rdd)

PySpark - Parallelize

(https://sparkbyexamples.com/py spark/pyspark-parallelize-create-

rdd/)

PySpark - repartition() vs coalesce()

(https://sparkbyexamples.com/py spark/pyspark-repartition-vscoalesce/)

PySpark - Broadcast Variables (https://sparkbyexamples.com/py spark/pyspark-broadcastvariables/)

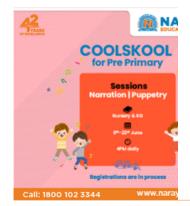
PySpark - Accumulator

(https://sparkbyexamples.com/py spark/pyspark-accumulator-withexample/)

PySpark DataFrame

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

ark/spark-web-ui-understanding/) FAQ's InhthissP//Sparklaytiquempweb explainprow-touestions/) convert an array of String column on DataFrame to a String column (separated or (https://sparkbyexamples.com/py More con(catenated with accommodespace)/prany delimiter character) using PySpark function concat ws() (translates to concat with separator), and with SQL expression using Scala example.



When curating data on DataFrame we may want to convert the Dataframe with complex struct datatypes

(https://sparkbyexamples.com/spark/sparksql-structtype-on-dataframe/), arrays (https://sparkbyexamples.com/spark/sparkarray-arraytype-dataframe-column/)and maps to a flat structure. here we will see how to convert array type to string type.

<u>spark/different-ways-to-create-dataframe-in-pyspark/)</u>

<u>PySpark – Create an empty</u> <u>DataFrame</u>

(https://sparkbyexamples.com/py spark/pyspark-create-an-emptydataframe/)

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

<u>PySpark – Convert DataFrame to</u> <u>Pandas</u>

(https://sparkbyexamples.com/py spark/convert-pyspark-dataframeto-pandas/)

<u>PySpark – show()</u> (https://sparkbyexamples.com/py spark/pyspark-show-displaydataframe-contents-in-table/)</u>

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

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

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

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

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

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

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

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

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

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

Before we start, first let's <u>create a</u> DataFrame

(https://sparkbyexamples.com/pyspark/different-ways-to-create-dataframe-in-pyspark/) with array of string column.

In this example "languagesAtSchool" is a column of type array. In the next section, we will convert this to a String. This example yields below schema and DataFrame.

Convert an array of String to String column using concat_ws()

In order to convert array to a string, PySpark SQL provides a built-in function concat_ws() which takes delimiter of your choice as a first argument and array column (type Column) as the second argument.

<u>spark/pyspark-orderby-and-sort-explained/)</u>

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

PySpark - join()

(https://sparkbyexamples.com/py spark/pyspark-join-explainedwith-examples/)

PySpark – union() & unionAll() (https://sparkbyexamples.com/py spark/pyspark-union-andunionall/)

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

PySpark – UDF (User Defined Function)
(https://sparkbyexamples.com/py

(https://sparkbyexamples.com/py spark/pyspark-udf-user-definedfunction/)

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

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

pyspark - foreach()
(https://sparkbyexamples.com/py
spark/pyspark-loop-iteratethrough-rows-in-dataframe/#useforeach-loop-through-dataframe)

PySpark - sample() vs sampleBy() (https://sparkbyexamples.com/py spark/pyspark-samplingexample/)

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

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

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

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

Install the Business Brows

Syntax

```
concat_ws(sep, *cols)
```

Usage

In order to use concat_ws() function, you need to import it using pyspark.sql.functions.concat_ws. Since this function takes the Column type as a second argument, you need to use col().

```
from pyspark.sql.functions import o
df2 = df.withColumn("languagesAtSch
    concat_ws(",",col("languagesAtSch
df2.printSchema()
df2.show(truncate=False)
```

This yields below output

Using PySpark SQL

expression

You can also use concat_ws() function with SQL expression.



Compatible Across Platforms

Manage Chrome Brov using existing tools an policies across any O device.



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

PySpark SQL Functions

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

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

PySpark – Date and Timestamp Functions

(https://sparkbyexamples.com/pysp ark/pyspark-sql-date-andtimestamp-functions/)

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

PySpark Datasources

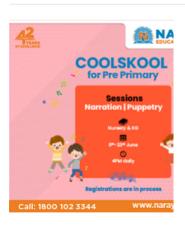
<u>PySpark</u> – Read & Write CSV File (https://sparkbyexamples.com/pyspark/pyspark-read-csv-file-into-dataframe/)

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

(https://sparkbyexamples.com/py spark/pyspark-read-and-writeparquet-file/)

PySpark - Read & Write JSON

(https://sparkbyexamples.com/py spark/pyspark-read-json-file-intodataframe/)





Complete Example

Below is a complete PySpark DataFrame example of converting an array of String column to a String using a Scala example.

```
import pyspark
from pyspark.sql import SparkSessid
spark = SparkSession.builder.appNan
columns = ["name", "languagesAtSchool
data = [("James,,Smith",["Java","Sd
    ("Michael, Rose, ", ["Spark", "Java
    ("Robert,, Williams", ["CSharp", '
df = spark.createDataFrame(data=dat
df.printSchema()
df.show(truncate=False)
from pyspark.sql.functions import
df2 = df.withColumn("languagesAtSch
   concat_ws(",",col("languagesAtSc
df2.printSchema()
df2.show(truncate=False)
df.createOrReplaceTempView("ARRAY_S
spark.sql("select name, concat_ws(
    " currentState from ARRAY STRIN
    .show(truncate=False)
```

This example is also available at the PySpark Github example project (https://github.com/spark-examples/pyspark-examples/blob/master/pyspark-array-string.py) for reference.

Hope it helps you !! Thanks for reading.



(https://sparkbyexamples.com/pyspark/pysparkconvert-array-column-to-string-column/? share=facebook&nb=1)

(https://sparkbyexamples.com/pyspark/pysparkconvert-array-column-to-string-column/? share=reddit&nb=1)

ρ (https://sparkbyexamples.com/pyspark/pysparkconvert-array-column-to-string-column/? share=pinterest&nb=1)

t (https://sparkbyexamples.com/pyspark/pysparkconvert-array-column-to-string-column/? share=tumblr&nb=1)

(https://sparkbyexamples.com/pyspark/pysparkconvert-array-column-to-string-column/? share=pocket&nb=1)

in (https://sparkbyexamples.com/pyspark/pysparkconvert-array-column-to-string-column/? share=linkedin&nb=1)

(https://sparkbyexamples.com/pyspark/pysparkconvert-array-column-to-string-column/? share=twitter&nb=1)

AGS: CONCAT_WS()

ITTPS://SPARKBYEXAMPLES.COM/TAG/CONCAT_WS/)

PySpark Built-In Functions

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

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

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

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

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

<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)

<u>PySpark - regexp_replace()</u>
(https://sparkbyexamples.com/pyspark/pyspark-replace-column-



<u>NNK</u>

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

(https://sp arkbyexa 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...

<u>values/#regexp_replace-replace-string-columns)</u>

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

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

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

<u>PySpark - date_format()</u>
(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())

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

<u>PySpark – array_contains()</u> (<u>https://sparkbyexamples.com/pyspark/pyspark-arraytype-column-withexamples/#array_contains)</u>

<u>PySpark – array()</u> (https://sparkbyexamples.com/pyspark/pyspark-arraytype-column-withexamples/#array)</u>

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

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

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)

<u>PySpark - map_values()</u> (https://sparkbyexamples.com/pysp

ark/pyspark-maptype-dictexamples/#map_values)

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

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

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

<u>PySpark - row_number()</u>
(https://sparkbyexamples.com/pyspark/pyspark-window-functions/#row_number)

<u>PySpark - rank()</u>
(https://sparkbyexamples.com/pyspark/pyspark-window-functions/#rank)

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

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

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

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

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

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

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

<u>PySpark – schema of json()</u> (https://sparkbyexamples.com/pyspark/pyspark-json-functions-with-examples/#schema of json) Apache Hadoop

(https://sparkbyexamples.com/category/h adoop/)

Apache Spark

(https://sparkbyexamples.com/category/s park/)

Apache Spark Streaming

(https://sparkbyexamples.com/category/s park/apache-spark-streaming/)

Apache Kafka

(https://sparkbyexamples.com/category/k afka/)

Apache HBase

(https://sparkbyexamples.com/category/h base/)

Apache Cassandra

(https://sparkbyexamples.com/category/c assandra/)

Snowflake Database

(https://sparkbyexamples.com/category/s nowflake/)

H2O Sparkling Water

(https://sparkbyexamples.com/category/h 2o-sparkling-water/)

PySpark

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

Spark regexp_replace() – Replace String

(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

(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/)

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

ples)



//www. //www.

(htt	ps:	facebo	linkedi	(https:
			n.com/	
r.co	om/	m/spar	in/n-	b.com/
spa	rkb	kbyex	nk-	spark-
yex	am	ample	b860a	examp

s/)

8193/)

les/)

Copyright sparkbyexamples.com