

# How can I convert the columns of a PySpark DataFrame into a MapType (Dict) format?

Authored by  
**stats writer**

June 24, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can I convert the columns of a PySpark DataFrame into a MapType (Dict) format?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=151150>

Converting the columns of a PySpark DataFrame into a MapType (Dict) format allows for easier manipulation and analysis of data. To do this, one can use the PySpark function "create\_map" which takes in a list of columns and creates a map from those columns. This map can then be used to convert the DataFrame columns into a dictionary format, providing a more structured and organized way to access and work with the data. This can be particularly useful for tasks such as feature engineering or data transformation. Overall, converting DataFrame columns into a MapType format expands the capabilities of PySpark and enables more efficient data handling.

To convert DataFrame columns to a MapType (dictionary) column in PySpark, you can use the create\_map function from the pyspark.sql.functions module. This function allows you to create a map from a set of key-value pairs, where the keys and values are columns from the DataFrame.

Let's create a DataFrame

```
from pyspark.sql import SparkSession
from pyspark.sql.types import StructType, StructField, StringType, IntegerType

spark = SparkSession.builder.appName('SparkByExamples.com').getOrCreate()
data =
schema = StructType()

df = spark.createDataFrame(data=data, schema=schema)
df.printSchema()
df.show(truncate=False)
```

This yields below output

```
root
|-- id: string (nullable = true)
|-- dept: string (nullable = true)
|-- salary: integer (nullable = true)
|-- location: string (nullable = true)
```

```
+-----+-----+-----+-----+
|id|dept|salary|location|
+-----+-----+-----+
|36636|Finance|3000|USA|
|40288|Finance|5000|IND|
|42114|Sales|3900|USA|
```

```
|39192|Marketing|2500 |CAN |
|34534|Sales |6500 |USA |
+-----+-----+-----+-----+
```

## Convert DataFrame Columns to MapType

Now, using `create_map()` SQL function let's convert PySpark DataFrame columns `salary` and `location` to `MapType`.

```
#Convert columns to Map
from pyspark.sql.functions import col,lit,create_map
df = df.withColumn("propertiesMap",create_map(
lit("salary"),col("salary"),
lit("location"),col("location")
)).drop("salary","location")
df.printSchema()
df.show(truncate=False)
```

This yields below output.

```
root
|-- id: string (nullable = true)
|-- dept: string (nullable = true)
|-- propertiesMap: map (nullable = false)
|   |-- key: string
|   |-- value: string (valueContainsNull = true)
```

```
+-----+-----+-----+-----+
|id|dept|propertiesMap|
+-----+-----+-----+
|36636|Finance||
|40288|Finance||
|42114|Sales||
|39192|Marketing||
|34534|Sales||
+-----+-----+-----+-----+
```

Happy Learning !!

## Related Articles

ARABPSYCHOLOGY.COM