

Maximum

January 4, 2024

```
[1]: import pyspark
import os
import sys
from pyspark import SparkContext
from pyspark import SparkConf
os.environ['PYSPARK_PYTHON'] = sys.executable
os.environ['PYSPARK_DRIVER_PYTHON'] = sys.executable

from pyspark.sql import SparkSession
```

```
[2]: spark = SparkSession.builder.config("spark.driver.memory", "16g").
    →appName('square').getOrCreate()
```

```
[3]: import pandas as pd
from pyspark.sql import functions as F
df_pd = pd.DataFrame(
    data={'integers': [1, 2, 3],
          'floats': [-1.0, 0.5, 2.7],
          'integer_arrays': [[1, 2], [3, 4, 5], [6, 7, 8, 9]]}
)
df = spark.createDataFrame(df_pd)
df.printSchema() # It will print the Schema
df.show()
print(df.select("integers").rdd.max()[0])

print(df.agg(F.avg(df['integers']))).show()
```

```
root
|-- integers: long (nullable = true)
|-- floats: double (nullable = true)
|-- integer_arrays: array (nullable = true)
|    |-- element: long (containsNull = true)
```

```
+-----+-----+-----+
|integers|floats|integer_arrays|
+-----+-----+-----+
|         1| -1.0|         [1, 2]|
```

	2	0.5	[3, 4, 5]
	3	2.7	[6, 7, 8, 9]
+-----+-----+-----+			

```

3
+-----+
|avg(integers)|
+-----+
|           2.0|
+-----+

```

None

[]: