



# Learning PySpark

**1 – Create Dataframe manually with user  
defined values in PySpark**

## • Initialize a spark session

This step is only needed if you are using a local spark environment and not Databricks or any other managed spark implementation

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

# Initialize a Spark session
spark = SparkSession.builder \
    .appName("CreateDataFrame") \
    .getOrCreate()
```

✓ 18.9s

## • Define data and schema

```
1 data = [("Alice", 25, "New York"),
2         ("Bob", 30, "San Francisco"),
3         ("Charlie", 22, "Los Angeles")]
4
5 schema = ['name', 'age', 'city']
6
7
```

- Create the dataframe

```
1 df = spark.createDataFrame(data=data, schema = schema)
2 df.show()
```

▶ (3) Spark Jobs

▶  df: pyspark.sql.dataframe.DataFrame = [name: string, age: long ... 1 more field]

```
+-----+---+-----+
|  name|age|          city|
+-----+---+-----+
| Alice| 25|    New York|
|   Bob| 30|San Francisco|
|Charlie| 22|  Los Angeles|
+-----+---+-----+
```

## • Define schema with datatype

Unlike the schema definition in previous page we can also define schema along with datatype

```
1  from pyspark.sql.types import StructType, StructField, IntegerType, StringType
2
3  data = [("Alice", 25, "New York"),
4         |      ("Bob", 30, "San Francisco"),
5         |      ("Charlie", 22, "Los Angeles")]
6
7  schema = StructType(
8      [
9          StructField(name = 'name', dataType = StringType()),
10         StructField(name = 'age', dataType = IntegerType()),
11         StructField(name = 'city', dataType = StringType())
12     ]
13 )
```

```
1  df = spark.createDataFrame(data, schema = schema)
2  df.printSchema()
3  df.show()
```

► (3) Spark Jobs

►  df: pyspark.sql.dataframe.DataFrame = [name: string, age: long ... 1 more field]

root

```
|-- name: string (nullable = true)
|-- age: long (nullable = true)
|-- city: string (nullable = true)
```

```
+-----+---+-----+
|  name|age|      city|
+-----+---+-----+
| Alice| 25|   New York|
|   Bob| 30|San Francisco|
|Charlie| 22|  Los Angeles|
+-----+---+-----+
```