

```
In [107... from pyspark.sql import SparkSession

spark = (
    SparkSession
    .builder
    .master("local[*]")
    .appName("Spark Streaming")
    .getOrCreate()
)

spark
```

Out[107]: **SparkSession - in-memory**

**SparkContext**

[Spark UI](#)

**Version** v3.3.0  
**Master** local[\*]  
**AppName** Spark Streaming

```
In [108... streaming_df = (
    spark.read
    .json("datas/input.json", multiLine=True)
)
```

```
In [109... streaming_df.printSchema()

root
 |-- customerId: string (nullable = true)
 |-- data: struct (nullable = true)
 |   |-- devices: array (nullable = true)
 |   |   |-- element: struct (containsNull = true)
 |   |   |   |-- deviceId: string (nullable = true)
 |   |   |   |-- measure: string (nullable = true)
 |   |   |   |-- status: string (nullable = true)
 |   |   |   |-- temperature: long (nullable = true)
 |-- eventId: string (nullable = true)
 |-- eventOffset: long (nullable = true)
 |-- eventPublisher: string (nullable = true)
 |-- eventTime: string (nullable = true)
```

```
In [110... streaming_df.show()

+-----+-----+-----+-----+-----+
|customerId|data|eventId|eventOffset|eventPublisher|
+-----+-----+-----+-----+-----+
|c1|[{"d1": "c", "error": "...}|some343id|100|device|2024-0
```

In [111... **from** pyspark.sql.functions **import** explode

```
explode_df = streaming_df.withColumn("devices", explode("data.devices"))
explode_df.show(truncate=False)
```

```
+-----+-----+-----+-----+
|customerId|data                                     |eventId |eventOffs|
+-----+-----+-----+-----+
|c1        |[{d1, c, error, 15}, {d2, d, success, 20}]|some343id|100      |
|c1        |[{d1, c, error, 15}, {d2, d, success, 20}]|some343id|100      |
+-----+-----+-----+-----+
```

In [112... explode\_df.printSchema()

```
root
 |-- customerId: string (nullable = true)
 |-- data: struct (nullable = true)
 |   |-- devices: array (nullable = true)
 |   |   |-- element: struct (containsNull = true)
 |   |   |   |-- deviceId: string (nullable = true)
 |   |   |   |-- measure: string (nullable = true)
 |   |   |   |-- status: string (nullable = true)
 |   |   |   |-- temperature: long (nullable = true)
 |-- eventId: string (nullable = true)
 |-- eventOffset: long (nullable = true)
 |-- eventPublisher: string (nullable = true)
 |-- eventTime: string (nullable = true)
 |-- devices: struct (nullable = true)
 |   |-- deviceId: string (nullable = true)
 |   |-- measure: string (nullable = true)
 |   |-- status: string (nullable = true)
 |   |-- temperature: long (nullable = true)
```

In [113... **from** pyspark.sql.functions **import** col

```
flatened_df = (
    explode_df
    .drop('data')
    .withColumn('deviceId', col('devices.deviceId'))
    .withColumn('measure', col('devices.measure'))
    .withColumn('status', col('devices.status'))
    .withColumn('temperature', col('devices.temperature'))
)
```

```
flatened_df.show()
```

```
+-----+-----+-----+-----+-----+
|customerId| eventId|eventOffset|eventPublisher| eventTime|
+-----+-----+-----+-----+-----+
|          c1|some343id|      100|device|2024-01-01 11:11:11| {d1,
|          c1|some343id|      100|device|2024-01-01 11:11:11| {d2, d,
+-----+-----+-----+-----+-----+
```