

ex57

August 18, 2022

```
[ ]: from pyspark import SparkConf, SparkContext
     from pyspark.sql import SparkSession
     from graphframes import GraphFrame

     conf = SparkConf().setAppName("ex57")
     sc = SparkContext(conf=conf)
     ssql = SparkSession.builder.getOrCreate()
```

```
[2]: edgesPath = "data/Ex57/data/edges.csv"
     vertexesPath = "data/Ex57/data/vertexes.csv"
     outputPath = "out57/"
```

```
[3]: eDF = ssql.read.load(
      edgesPath,
      format="csv",
      header=True,
      inferSchema=True
    )

    vDF = ssql.read.load(
      vertexesPath,
      format="csv",
      header=True,
      inferSchema=True
    )
```

```
[4]: eDF.show(), eDF.printSchema()
     vDF.show(), vDF.printSchema()
```

```
+---+---+-----+
|src|dst|linktype|
+---+---+-----+
| u1| u2|  friend|
| u1| u4|  friend|
| u1| u5|  friend|
| u2| u1|  friend|
| u2| u3| follow|
| u3| u2| follow|
```

```
| u4| u1| friend|
| u4| u5| friend|
| u5| u1| friend|
| u5| u4| friend|
| u5| u6| follow|
| u6| u3| follow|
+---+---+-----+
```

```
root
|-- src: string (nullable = true)
|-- dst: string (nullable = true)
|-- linktype: string (nullable = true)
```

```
+---+-----+---+
| id| name|age|
+---+-----+---+
| u1|Alice| 34|
| u2| Bob| 36|
| u3| John| 30|
| u4|David| 29|
| u5| Paul| 32|
| u6| Adel| 36|
| u7| Eddy| 60|
+---+-----+---+
```

```
root
|-- id: string (nullable = true)
|-- name: string (nullable = true)
|-- age: integer (nullable = true)
```

[4]: (None, None)

[5]: `g = GraphFrame(vDF, eDF)`

```
/home/webbelle/univenv/lib/python3.10/site-
packages/pyspark/sql/dataframe.py:148: UserWarning: DataFrame.sql_ctx is an
internal property, and will be removed in future releases. Use
DataFrame.sparkSession instead.
  warnings.warn(
```

[6]: `shortestPathsDF = g.shortestPaths(["u1"])`

```
/home/webbelle/univenv/lib/python3.10/site-
packages/pyspark/sql/dataframe.py:127: UserWarning: DataFrame constructor is
internal. Do not directly use it.
  warnings.warn("DataFrame constructor is internal. Do not directly use it.")
```

```
[7]: shortestPathsDF.show(), shortestPathsDF.printSchema()
```

```
+---+-----+---+-----+
| id| name|age|distances|
+---+-----+---+-----+
| u6| Adel| 36|{u1 -> 3}|
| u3| John| 30|{u1 -> 2}|
| u2| Bob| 36|{u1 -> 1}|
| u4|David| 29|{u1 -> 1}|
| u5| Paul| 32|{u1 -> 1}|
| u1|Alice| 34|{u1 -> 0}|
| u7| Eddy| 60|      {}|
+---+-----+---+-----+
```

```
root
|-- id: string (nullable = true)
|-- name: string (nullable = true)
|-- age: integer (nullable = true)
|-- distances: map (nullable = true)
|    |-- key: string
|    |-- value: integer (valueContainsNull = false)
```

```
[7]: (None, None)
```

```
[8]: finalDF = shortestPathsDF.filter("distances['u1']<3 AND id<>'u1'")
finalDF.show()
```

```
+---+-----+---+-----+
| id| name|age|distances|
+---+-----+---+-----+
| u3| John| 30|{u1 -> 2}|
| u2| Bob| 36|{u1 -> 1}|
| u4|David| 29|{u1 -> 1}|
| u5| Paul| 32|{u1 -> 1}|
+---+-----+---+-----+
```

```
[10]: resultDF = finalDF.selectExpr("name", "distances['u1'] AS NumHops")
resultDF.show()
```

```
+-----+-----+
| name|NumHops|
+-----+-----+
| John|      2|
| Bob|      1|
|David|      1|
| Paul|      1|
```

+-----+-----+

```
[11]: resultDF.write.csv(outputPath, header=True)
```