

3.ImportJson

March 4, 2025

```
[2]: from pyspark.sql import SparkSession
import json
```

```
[3]: # Criar uma sessão Spark
spark = SparkSession.builder.appName("APP_JSONExample").getOrCreate()

# Criar um ficheiro JSON
data = [
    {"id": 1, "nome": "Alice", "idade": 25},
    {"id": 2, "nome": "Bob", "idade": 30},
    {"id": 3, "nome": "Carlos", "idade": 22},
    {"id": 4, "nome": "Diana", "idade": 28},
    {"id": 5, "nome": "Eva", "idade": 35},
    {"id": 6, "nome": "Fernando", "idade": 40}
]
```

```
[4]: json_filename = "Json/dados2.json"
with open(json_filename, "w") as f:
    json.dump(data, f)
```

```
[5]: # Carregar o JSON com PySpark
df = spark.read.json(json_filename)
```

```
[6]: # Mostrar os dados
df.show()
```

```
+---+-----+-----+
| id|idade|   nome|
+---+-----+-----+
|  1|   25|  Alice|
|  2|   30|   Bob|
|  3|   22| Carlos|
|  4|   28|  Diana|
|  5|   35|   Eva|
|  6|   40|Fernando|
+---+-----+-----+
```

```
[22]: # Mostrar o top 5
df.limit(5).show()
```

```
+---+-----+-----+
| id|idade|  nome|
+---+-----+-----+
|  1|  25| Alice|
|  2|  30|  Bob|
|  3|  22|Carlos|
|  4|  28| Diana|
|  5|  35|  Eva|
+---+-----+-----+
```

```
[23]: # Mostrar o top 5
df.sort("id").show(5)
```

```
+---+-----+-----+
| id|idade|  nome|
+---+-----+-----+
|  1|  25| Alice|
|  2|  30|  Bob|
|  3|  22|Carlos|
|  4|  28| Diana|
|  5|  35|  Eva|
+---+-----+-----+
only showing top 5 rows
```

```
[ ]:
```

```
[ ]:
```