

## 7. WriteJson

March 4, 2025

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

```
[4]: # 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}
]
```

```
[5]: json_filename = "dados3.json"
with open(json_filename, "w") as f:
    json.dump(data, f)
```

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

```
[7]: # 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|
+---+-----+-----+
```

```
[8]: from pyspark.sql import SparkSession

# Criar a sessão do Spark
spark = SparkSession.builder.appName("ExemploPySpark").getOrCreate()

# Verificar a versão do Spark
print(spark.version)
```

3.5.0

```
[9]: from pyspark.sql import Row

# Criar uma lista de dados
dados = [Row(nome="Ana", idade=25), Row(nome="Bruno", idade=30),
        ↪Row(nome="Carlos", idade=35)]

# Criar um DataFrame
df = spark.createDataFrame(dados)

# Mostrar os dados
df.show()
```

```
+-----+-----+
| nome|idade|
+-----+-----+
|  Ana|   25|
| Bruno|   30|
| Carlos|  35|
+-----+-----+
```

```
[10]: df_csv = spark.read.csv("Aula 2/CSV/dados.csv", header=True, inferSchema=True)
df_csv.show()
```

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

```
[11]: df.select ("nome" ).show()
```

```
+-----+  
|  nome |  
+-----+  
|   Ana |  
| Bruno |  
| Carlos|  
+-----+
```

[ ]: