

Task 6: Use Broadcast Variables to Display Movie Names Instead of ID Numbers

```
from pyspark.sql import SparkSession
from pyspark import SparkContext
```

```
sc = SparkContext("local", "Broadcast Example")
spark = SparkSession.builder.getOrCreate()
```

```
# Broadcast variable
movies = {1: "Inception", 2: "Interstellar", 3: "Tenet"}
broadcast_movies = sc.broadcast(movies)
```

```
# Sample RDD with movie IDs
data = [(1, 5), (2, 4), (3, 3)]
rdd = sc.parallelize(data)
```

```
# Map movie IDs to names
result = rdd.map(lambda x: (broadcast_movies.value.get(x[0]), x[1]))
print(result.collect())
```

```
sc.stop()
```

```
→ [(('Inception', 5), ('Interstellar', 4), ('Tenet', 3))]
```

Start coding or [generate](#) with AI.

