

Ejercicio 3

Dadas las siguientes listas, crea un RDD para cada una y devuelve de los 2 RDDs su:

- Union
- Intersección
- Cartesiano

Listas:

- nameRDD1 = sc.parallelize(['Ana','Bob'])
- nameRDD2 = sc.parallelize(['Bob','Caren','Deric'])

Importaciones y creacion se SparkSession y SparkContext

```
In [9]: import findspark
findspark.init()

import pandas as pd
import pyspark
from pyspark.sql import SparkSession

spark = SparkSession.builder.getOrCreate()
sc = spark.sparkContext
```

Punto #1

Creacion de las siguietne listas:

- nameRDD1 = sc.parallelize(['Ana','Bob'])
- nameRDD2 = sc.parallelize(['Bob','Caren','Deric'])

```
In [10]: nameRDD1 = sc.parallelize(['Ana', 'Bob'])
nameRDD2 = sc.parallelize(['Bob', 'Caren', 'Deric'])
```

Punto #2

devuelve de los 2 RDDs su:

- Union
- Intersección
- Cartesiano

Union

```
In [11]: nameRDD1.union(nameRDD2).collect()
```

```
Out[11]: ['Ana', 'Bob', 'Bob', 'Caren', 'Deric']
```

Intersección

```
In [12]: nameRDD1.intersection(nameRDD2).collect()
```

```
Out[12]: ['Bob']
```

Cartesiano

```
In [13]: nameRDD1.cartesian(nameRDD2).collect()
```

```
Out[13]: [('Ana', 'Bob'),  
          ('Ana', 'Caren'),  
          ('Ana', 'Deric'),  
          ('Bob', 'Bob'),  
          ('Bob', 'Caren'),  
          ('Bob', 'Deric')]
```