

intersection Transformation

In [6]: `from pyspark.sql import SparkSession`

```
spark = SparkSession \
    .builder \
    .master("local[*]") \
    .appName("Sample Transformation") \
    .getOrCreate()
```

#Two lists

```
list1 = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
list2 = [11, 12, 13, 4, 15, 16, 7, 18, 9, 20]
```

```
rdd = spark.sparkContext.parallelize(list1, 5)
rdd1 = spark.sparkContext.parallelize(list2, 2)
```

#with intersection(), we find the intersection of two datasets.

#Therefore, the output will not contain any duplicates.

```
intersection_rdd = rdd.intersection(rdd1)
```

```
print(intersection_rdd.collect())
```

```
[7, 9, 4]
```

In [4]: *#Finding the intersection between two strings*

```
str_rdd_1 = spark.sparkContext.parallelize(['hi', 'John', 'how', 'are', 'you', 'doing'])
str_rdd_2 = spark.sparkContext.parallelize(['hi', 'David', 'how', 'are', 'you', 'coping'])
```

```
str_rdd = str_rdd_1.intersection(str_rdd_2)
print(str_rdd.collect())
```

```
['how', 'you', 'are', 'hi']
```