# Union Transformation

In [1]:
```python
from pyspark.sql import SparkSession

if __name__ == "__main__":
    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, 14, 15, 16, 17, 18, 19, 20]

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

    #with union() they are both combined
    union_rdd = rdd.union(rdd1)

    print(union_rdd.collect())
```

```
22/10/12 01:25:21 WARN Utils: Your hostname, Vaishalis-MacBook-Pro.local resolves to a loopback address: 127.0.0.1; using 192.168.0.105 instead (on interface en0)
22/10/12 01:25:21 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
22/10/12 01:25:22 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
22/10/12 01:25:23 WARN Utils: Service 'SparkUI' could not bind on port 4040. Attempting port 4041.
[Stage 0:>                                          (0 + 7) / 7]
[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]
```

In [2]:
```python
#Reading two files
file1= spark.read.text("file.txt")

print("Printing files1_rdd: ", file1.show())

file2= spark.read.text("tech_overview.txt")
print("Printing files2_rdd: ", file2.show())
```

```
+--------------------+
|               value|
+--------------------+
|Machine learning ...|
|provides systems ...|
|To learn and impr...|
|Without being exp...|
+--------------------+

Printing files1_rdd:  None
+--------------------+
|               value|
+--------------------+
|mapPartitionsWith...|
|Similar to mapPar...|
|with an integer v...|
|the partition, so...|
|type (Int, Iterat...|
|when running on a...|
+--------------------+

Printing files2_rdd:  None
```

In [3]:
```python
#Joining them
str_rdd = file1.union(file2)
print(str_rdd.show())
```

```
+--------------------+
|               value|
+--------------------+
|Machine learning ...|
|provides systems ...|
|To learn and impr...|
|Without being exp...|
|mapPartitionsWith...|
|Similar to mapPar...|
|with an integer v...|
|the partition, so...|
|type (Int, Iterat...|
|when running on a...|
+--------------------+

None
```