

In [ ]:

```
import org.apache.spark.SparkContext
import org.apache.spark.SparkConf
```

In [ ]:

```
val data = Array(1, 2, 3, 4, 5)
val distData = sc.parallelize(data)
distData.collect().foreach(println)
```

1. SELECT name FROM Customer WHERE month(startDate)=7

In [ ]:

```
val rddFromFile = sc.textFile("C:/Users/EASYFRONT/Documents/BD/requetessql/data/Customer.txt")
val rddsplitted = rddFromFile.map( x => x.split(","))
val rddmonth = rddsplitted.map(x => (x(2),x(1).split("/")(1)) )
val rddfiltered = rddmonth.filter(x => x._2=="07")
val rddnames = rddfiltered.map( x=> x._1)
rddnames.collect()
```

1. SELECT DISTINCT name FROM Customer WHERE month(startDate)=7

In [ ]:

```
rddnames.count()
```

In [ ]:

```
val rddnamesdistinct = rddnames.distinct()
rddnamesdistinct.collect()
```

In [ ]:

```
rddnamesdistinct.count()
```

1. SELECT O.cid, SUM(total), COUNT(DISTINCT total) FROM Order O GROUP BY O.cid

In [ ]:

```
val rddFromFile2 = sc.textFile("C:/Users/EASYFRONT/Documents/BD/requetessql/data/Order.txt")
val rddsplitted2 = rddFromFile2.map( x => x.split(","))
val rddgroupby = rddsplitted2.groupByKey()
val rddtemp = rddgroupby.map(x => (x._1, x._2.map(_._2.map(_.toInt))))
val rddsum = rddtemp.map(x=> (x._1, x._2.map(_._2.map(_.sum).sum, x._2.map(_._2.map(_.distinct.size))) )
rddsum.take(1)
```

In [ ]:

```
//test on small rdd
val rddFromFile2 = sc.parallelize(List(("5",List(Vector(1,1)))) )
val rddsum2 = rddFromFile2.map(x=> (x._1, x._2.map(_.sum).sum, x._2.map(_.distinct.size)))
rddsum2.map (x => (x._1, x._2, x._3.toList(0)))
rddsum2.take(1)
```

1. SELECT C.cid, O.total FROM Customer C, Order O WHERE month(startDate)=7 and C.cid=O.cid

In [ ]:

```
val rddFromFile = sc.textFile("C:/Users/EASYFRONT/Documents/BD/requetessql/data/Customer.txt")
val rddsplitted = rddFromFile.map( x => x.split(","))
val rddmonth = rddsplitted.map(x => (x(0),x(1).split("/")(1)) )
val rddFromFile2 = sc.textFile("C:/Users/EASYFRONT/Documents/BD/requetessql/data/Order.txt")
val rddsplitted2 = rddFromFile2.map( x => x.split(",")).map(x=> (x(0),x(1)))
val rddjoin = rddmonth.join(rddsplitted2)
val rddfilter = rddjoin.filter(x=> x._2._1=="07")
val nb = rddfilter.count()
print(nb)
```