# Roll Number | School Name | Name | Age | Gender | Class | Subject | Marks

```
schoolRaw = sc.textFile("file:///home/srikarthik/tejdata/students-db.txt")
```

```
schooldata = schoolRaw.map(lambda x : x.split("|"))
```

# a) Who got the highest for each class in each school?

```
schoolWise = schooldata.map(lambda x : ((int(x[0]),x[1],x[2],x[-3]),int(x[-1]))).reduceByKey(lambda x,y : x+y).sortBy(lambda x: x[1],ascending = False)

schoolWise2 = schoolWise.map(lambda x : ((x[0][1],x[0][-1]),(x[1],x[0][0],x[0][2]))).groupByKey().map(lambda x : (x[0],sorted(x[1],key = lambda k:k[0],reverse = True )))

schoolWiseFirstInEachClass = schoolWise2.sortByKey().map(lambda x: (x[0][0]+','+x[0][1]+","+ ",".join(str(i) for i in x[1][0])))

 for i in schoolWiseFirstInEachClass.collect(): print i
```

# Result:

```
DAV,eight,430,3,Brick
DAV,ninth,295,4,Prathibha
DAV,seventh,336,1,Praveen
DAV,tenth,315,1,Kyle
DPS,eight,414,3,Brick
DPS,ninth,383,2,Macency
DPS,seventh,391,4,Hasini
DPS,tenth,313,4,Hasini
HPS,eight,296,3,Brick
HPS,ninth,430,3,Rudima
HPS,seventh,278,2,Prajval
HPS,tenth,360,3,Kalpana
```

b)Who got the highest across all the schools for each class?

```
OverAllFirstInEachClass = schoolWise.map(lambda x : (x[0][-
1],(x[0][1],x[1],x[0][0],x[0][2]))).groupByKey().map(lambda x : (x[0],sorted(x[1],key = lambda
k:k[1],reverse = True ))).map(lambda x: (x[0]+','+ ",".join(str(i) for i in x[1][0])))
```

```
for i in OverAllFirstInEachClass.take(10):print i
```

```
        #Result:

                                tenth,HPS,360,3,Kalpana

                                eight,DAV,430,3,Brick

                                seventh,DPS,391,4,Hasini

                                ninth,HPS,430,3,Rudima
```

c) Sort the students according to the total marks for each school?

```
MarksWiseForEachSchool = schoolWise.map(lambda x: ((x[0][1],-x[-1]),x[0][2])).sortByKey(False)
```

```
for i in MarksWiseForEachSchool.take(10): print i
```

# Sample Result:

```
                ((u'HPS', 430), u'Rudima')

                ((u'HPS', 360), u'Kalpana')

                ((u'HPS', 359), u'Tobi')

                ((u'HPS', 322), u'Charley')

                ((u'HPS', 308), u'Kyle')

                ((u'HPS', 296), u'Brick')

                ((u'HPS', 278), u'Prajval')

                ((u'HPS', 273), u'Hasini')

                ((u'HPS', 260), u'William')
```

```
                    ((u'HPS', 260), u'Prathibha')
```

d) Did boys fare better or girls for each class?

```
TotalBoysAndTotalMarksClassWise = schooldata.filter(lambda x: x[4]=='M').map(lambda x : ((x[4],x[-3]),int(x[-1]))).aggregateByKey((0,0),lambda x,y : (x[0]+1,+x[1]+y), lambda x,y: (x[0]+y[0],x[1]+y[1])).sortBy(lambda x: x[1],ascending = False)
```

```
for i in TotalBoysAndTotalMarksClassWise.collect(): print i
```

```
          # Sample Result:

                    ((u'M', u'ninth'), (36, 1709))

                    ((u'M', u'tenth'), (36, 1645))

                    ((u'M', u'seventh'), (36, 1619))

                    ((u'M', u'eight'), (36, 1612))
```

```
BoysMeanMarksClassWise = TotalBoysAndTotalMarksClassWise.map(lambda x: (x[0][1],float(x[1][1]/x[1][0])))
```

```
for i in BoysMeanMarksClassWise.collect(): print i
```

```
          # Sample Result:

                    (u'ninth', 47.0)

                    (u'tenth', 45.0)

                    (u'seventh', 44.0)

                    (u'eight', 44.0)
```

```
TotalGirlsAndTotalMarksClassWise = schooldata.filter(lambda x: x[4]=='F').map(lambda x : ((x[4],x[-
3]),int(x[-1]))).aggregateByKey((0,0),lambda x,y : (x[0]+1,+x[1]+y), lambda x,y:
(x[0]+y[0],x[1]+y[1])).sortBy(lambda x: x[1],ascending = False)
```

```
for i in TotalGirlsAndTotalMarksClassWise.take(10): print i
```

```
        #Result:
                        ((u'F', u'eight'), (36, 1956))
                        ((u'F', u'ninth'), (36, 1856))
                        ((u'F', u'seventh'), (36, 1737))
                        ((u'F', u'tenth'), (36, 1678))
```

```
GirlsMeanMarksClassWise = TotalGirlsAndTotalMarksClassWise.map(lambda x:
(x[0][1],float(x[1][1]/x[1][0])))
```

```
for i in GirlsMeanMarksClassWise.collect(): print i
        #Result:
                (u'eight', 54.0)
                (u'ninth', 51.0)
                (u'seventh', 48.0)
                (u'tenth', 46.0)
```

```
compareResults = BoysMeanMarksClassWise.join(GirlsMeanMarksClassWise)
```

```
compareResults.collect()
[(u'eight', (44.0, 54.0)), (u'tenth', (45.0, 46.0)), (u'seventh', (44.0, 48.0)), (u'ninth', (47.0, 51.0))]
```

From above results we can interpret that girls fare better than boys.