

```

In [29]: %%file q1.py
from mrjob.job import MRJob
from mrjob.step import MRStep
class CalculateVariance(MRJob):
    def map_init(self):
        self.marks={}

    def mapper(self,_,line):
        course,marks=line.split(" ")
        marks=int(marks)
        self.marks.setdefault(course,[0,0,0]) #[0] has at 0th index sum , at 1th index square sum and 2nd index count
        self.marks[course][0]=self.marks[course][0]+marks
        self.marks[course][1]=self.marks[course][1]+marks*marks
        self.marks[course][2]=self.marks[course][2]+1

    def mapper_finaly(self):
        for c , val in self.marks.items():
            yield c,val

    def combiner(self,c,val):
        sumT=0
        sqSumT=0
        countT=0
        for summ,sqSum,count in val:
            sumT=sumT+summ
            sqSumT=sqSumT+sqSum
            countT=countT+count

        yield c,[sumT,sqSumT,countT]

    def reducer(self,c,val):
        sumT=0
        sqSumT=0
        countT=0
        for summ,sqSum,count in val:
            sumT=sumT+summ
            sqSumT=sqSumT+sqSum
            countT=countT+count

        if (countT-1) > 0 :
            var=((sqSumT-(sumT*sumT)/countT)/(countT-1))*0.5
        else:
            var=0
        yield c,var

    def steps(self):
        return [
            MRStep(mapper_init=self.map_init,
                    mapper=self.mapper,
                    mapper_final=self.mapper_finaly,
                    combiner=self.combiner,
                    reducer=self.reducer)
        ]

if __name__ == '__main__':
    CalculateVariance.run()

```

Overwriting q1.py

```
In [30]: !python q1.py rec.txt
```

```

"Algorithm"      49.49747468305833
"Database"       11.086778913041726
"English"        1.8257418583505538

```

No configs found; falling back on auto-configuration

No configs specified for inline runner

Creating temp directory C:\Users\DELL\AppData\Local\Temp\q1.DELL.20240312.090208.391942

Running step 1 of 1...

job output is in C:\Users\DELL\AppData\Local\Temp\q1.DELL.20240312.090208.391942\output

Streaming final output from C:\Users\DELL\AppData\Local\Temp\q1.DELL.20240312.090208.391942\output...

Removing temp directory C:\Users\DELL\AppData\Local\Temp\q1.DELL.20240312.090208.391942...

```
In [ ]:
```