National Technical University of Athens

Department of Electrical and Computer Engineering

Advanced Database Systems
Map/Reduce pseudocode
2nd semester of 2020

Masouris Athanasios

03115189 9th semester

Data processing

```
MAP (key, value):
    #key: line id
    #value: string with data separated by ","
    longitude = value.split(",")[3]  #4th index in the string
    latitude = value.split(",")[4]   #5th index in the string
    emit (key, (longitude,latitude))
```

Data labeling

```
MAP (key, value):
    #key: line id
    #value: point , a tuple of  (longitude, latitude)
    emit(key,(value, cluster(value, centroids)))
#where cluster(x,centroids) is a function that clusters the point x to the nearest centroid
```

Centroids updating

```
MAP(key, value):
    #key: line id
    #value: (point, clusterId), με point: (longitude, latitude)
    emit(clusterId, (point, 1))

REDUCE(key,values):
    #key: clusterId
    #values: list of (point,1) tuples
    sumLong = 0
    sumLat = 0
    size = 0
    for each (point, 1) in values:
      sumLong += point[0]
      sumLat += point[1]
      size += 1
    emit ((sumLong, sumLat), size)

MAP(key,value):
    #key: (sumLong, sumLat)
    #value: size
    emit(sumLong/size, sumLat/s)
```