

## Assignment 2: MapReduce on Spark

Name: Waghe Shubham Yatindra

Roll No.: **13MF3IM17**

### Question 2: A2Q2 :

- **Map function:**
  - Input format <number>
  - Output format : Key : <number>, Value <1>
- **Reduce Function:**
  - Input format <k, val1>,<k,val2>
  - Output format : Key : <k, val1+val2>
    - SortByKey on this output
- **Percentile :**
  - Calculate the index of the 25th percentile, 50th percentile and 75th percentile
  - Calculate Nth percentile\* 0.01 \* totalCount index by adding the count in loop and breaking out when the required breakPoint is achieved
  - If the division is integer and the count ends exactly at breakPoint then,  
Percentile = (current number + next number)/2