**Q1: State two differences between an in-mapper combiner and a combiner?**

The following are the differences between an in-mapper combiner and a combiner:

1. We have full control over when the local aggregation takes places and how it exactly takes place in in-mapper combiner and for combiner everything is handled by the frame and we don’t have full control.
2. In-mapper combiner are efficient as there is no intermediate data and intermediate processing and overheads whereas combiner is slower and in-efficient compared to in-mapper combiner.

**Q2: Write a MapReduce program to solve the following problem:**

**Given many large documents, identify the pages in each document that have the following sentence "using MapReduce in Hadoop". You may assume each item in the document has three parts (word, docid, pageNumber).**

class Mapper{

method Initialize(){

map = new HashMap //hash map of sets

}

method Map(String word, int docid, int pageNumber){

if(word.contains("using MapReduce in Hadoop")){

if(map.contains(docid)){

map.get(docid).add(pageNumber)

}else{

map.put(docid, {pageNumber})

}

}

}

method Close(){

for all term t in map {

Emit(t, map.get(t))

}

}

}

class Reducer{

method Reduce(int docid, array list){

set = new Set; // handle duplicate items

for all set in list{

set.addAll(set)

}

Emit(docid, set);

}

}