

Q1: What is the role of the mapper?

The mapper's job is to count and group data locally before sending it out. This way, it sends less data over the network, making the process faster and more efficient.

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()
    H = new AssociativeArray()
    str ← "using MapReduce in Hadoop"
    docId ← 0
    pageNo ← 0
    pageContent ← ""

  method map(word w; docid a, pageNumber b)
    if a != docId then
      pageContent = ""
    if b != pageNo then
      pageContent = ""
    pageContent += " " + w
    if pageContent includes str, then
      list pages P ← H{docId}
      if P is not contains pageNo then
        P add pageNo
      H{docId} ← P

  method close()
    for all term t in H do
      Emit(t, H{t})

class Reducer
  method reduce(docid a, pageNumber [p1, p2, ...])
    Emit(a, pageNumber
```