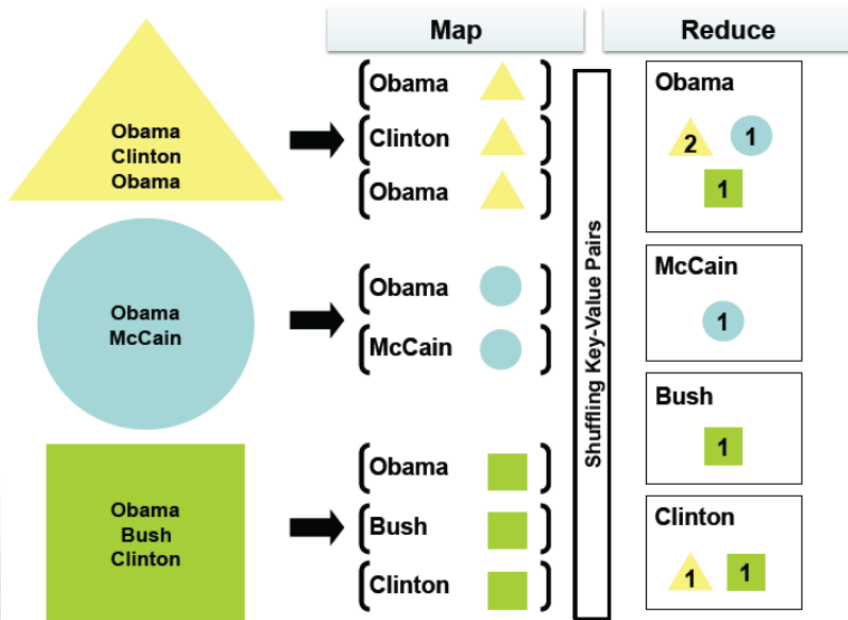


• 实现思想



这个图的坑点主要有二：

一是MapReduce的键值对不明确，这里的应该明确的指出Key和value是谁，这张图Map阶段中并没有value。而value应该是所出现文档和在文档中出现的数量，就是Reduce中的图形和其中数字。

二是Map阶段的key是错误的，Map阶段的key应该是词，而不是词和文章。不同的key会导致分发逻辑的不同，导致结果也不同。

综上，莫老师是希望大家考试时在写MapReduce的时候把键值对的键和值都写清楚。



大矩阵乘法

- 大矩阵乘法
 - 两个“超级大”的矩阵相乘
 - 如何计算
 - $A : m \times k, a_{ij}$
 - $B : k \times n, b_{ij}$

1	1	1
1	2	2
1	3	3
2	1	4
2	2	5
3	1	7
3	2	8
3	3	9
4	1	10
4	2	11
4	3	12

1	1	10
1	2	15
2	2	2
3	1	11
3	2	9

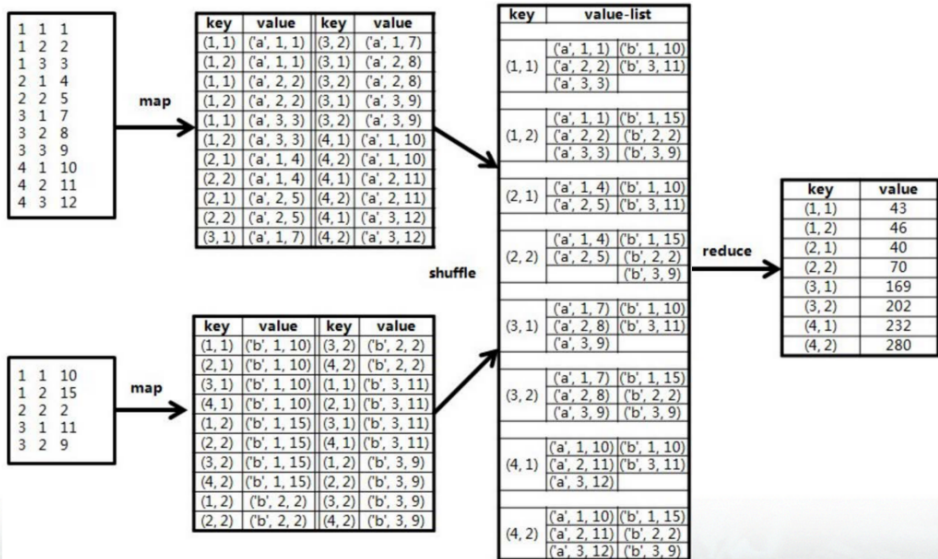
这个图的重点在于：大矩阵的格式。这两个并不是两个3列的矩阵。每一行的三个数分别代表行标列标和这个位置上的值，需要根据这个三元组还原出来整个矩阵。没有值的地方就是0.



北京大学



大矩阵乘法



大矩阵乘法键值对的构成如下:

键是一个二元组, 表示的是这个键值对是对结果矩阵哪个位置上的结果负责。

第一个数是结果矩阵, 第二个数是列标。在shuffle的时候会将所有key相同的键值对汇集到一起。

值是一个三元组, 第一个值代表这个键值对是来自哪一个矩阵的: a/b。第二个值是表示这个是在这一行/列中第几个元素, 方便两个数对应相乘。第三个数是这个值。在reduce的时候会把相应的键值对相乘



北京大学