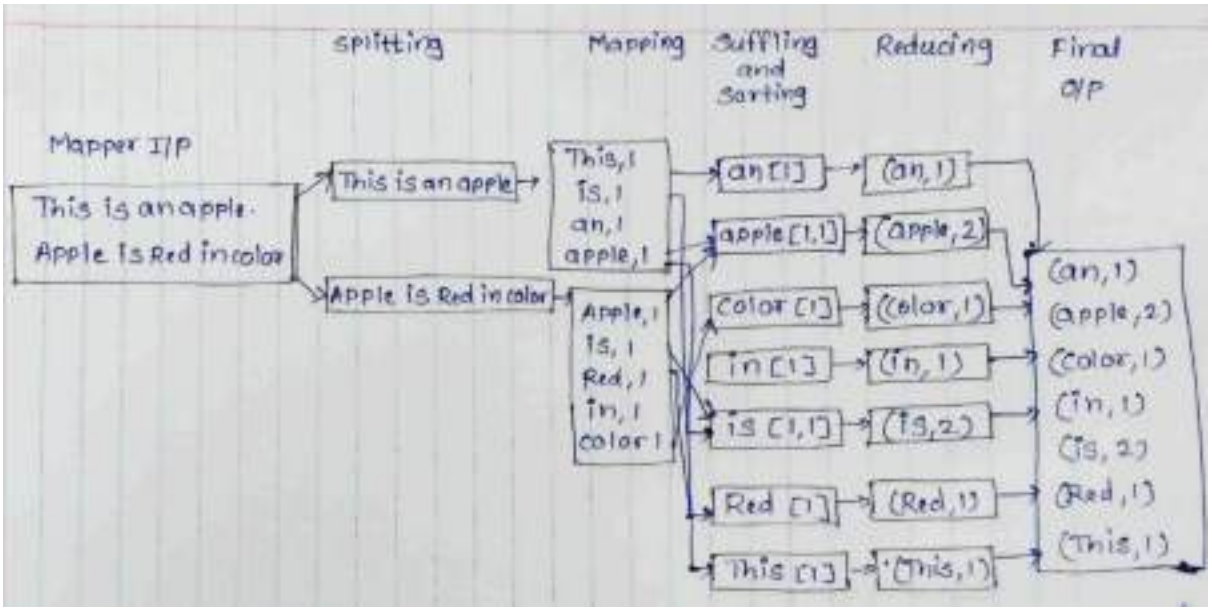
Q.1) What is the function of Map tasks in the Map Reduce framework? Explain with the help of example [5M] Dec 22

Q.2) Write a map reduce pseudocode for word count problem. Apply map reduce working on the following document. “This is an apple. Apple is Red in color” [10M]

(DEC-2022)

Ans:



Q.3)

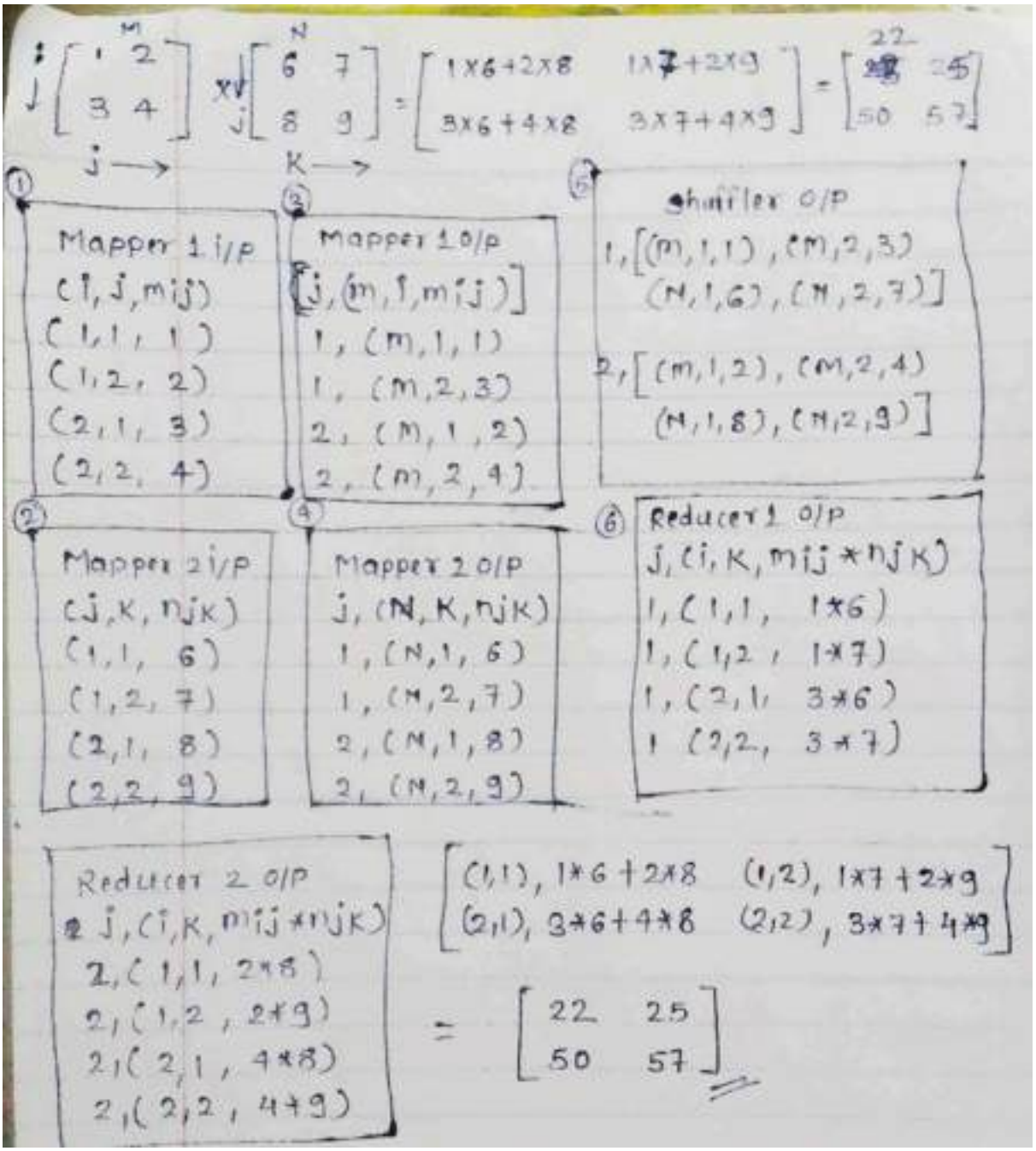
Write a map reduce pseudo code to multiply two matrices. Apply map reduce working to perform following matrix multiplication. **[10]**

1 2 6 7

X

3 4 8 9

May 2023



Note: Also include MapReduce PseudoCode.

Q.4)

Explain natural join and grouping and aggregation relational algebraic operation using MapReduce. **[10]**

Dec 2023 : Ans:

Natural join using MapReduce


```
map (Key, value):
    if Key == R :  <- means if Table 1 then will
        for (a,b) in value :      i.e. 'R'
            emit ( b, (R,a)).     iterate all values
                                  from that table
    else:
        for (b,c) in value:
            emit (b, (S,c))
```

```
reduce (Key, values):
    list_R = [a for (x,a) in values if x==R]
    list_S =[c for (x,c) in values if x==S]
    for a in list_R:
        for c in list_S:
            emit (Key, (a, Key, c))
```

Map Worker 1

Table 1

A	B
1	2
2	3
5	6

Table 2

B	C
2	3
4	4
6	1

Map Worker 2

Table 1

A	B
6	1
6	3
7	6

Table 2

B	C
9	8
3	4
7	1

MW1

Key	Value
2	$(T_1,1), (T_2,3)$
3	$(T_1, 2)$
4	$(T_2, 4)$
6	$(T_1,5), (T_1, 1)$

MW2

Key	Value
1	$(T_1, 6)$
9	$(T_2, 8)$
3	$(T_1, 6), (T_2, 4)$
6	$(T_1, 7)$
7	$(T_2, 1)$

A hash function will be applied.

MW1

Key	Value
2	$(T_1,1), (T_2,3)$
3	$(T_1, 2)$

M

Key	value
4	$(T_2,4)$
6	$(T_1,5), (T_2,1)$

MW2

Key	value
1	$(T_1, 6)$
9	$(T_2, 8)$
3	$(T_1, 6), (T_2, 4)$

Key	value
6	$(T_1, 7)$
7	$(T_2, 1)$

swapping

We will swap to avoid redundancy caused due to duplicate keys.

RW1

Key	value
2	$(T_1,1),(T_2,3)$
3	$(T_1,2)$

Key	value
1	$(T_1,6)$
9	$(T_2,8)$
3	$(T_1,6)$ $(T_2,4)$

RW2

Key	value
4	$(T_2,4)$
6	$(T_1,5)(T_2,1)$

Key	value
6	$(T_1,7)$
7	$(T_2,1)$

We will clubbed the two tables.

RW1

Key	value
2	$(T_1,1),(T_2,3)$
3	$(T_1,2),(T_1,6),(T_2,4)$
1	$(T_1,6)$ X
9	$(T_2,8)$ X

RW2

Key	value
4	$(T_2,4)$ X
6	$(T_1,5),(T_2,1),(T_1,7)$
7	$(T_2,1)$ X

RW1

B	A	C
2	1	3
3	2	4
3	6	4

RW2

B	A	C
6	5	1
6	7	1

Grouping and Aggregation by mapReduce

```
map (key, value):
    for (a,b,c) in value:
        emit (a,b)

reduce (key, values):
    emit (key, theta (values))
```

Map Worker 1

Table 1

A	B	C	D
1	2	10	2
2	3	2	8
1	2	9	6

Table 2

A	B	C	D
4	2	7	9
6	8	3	3
3	2	5	7

Map Worker 2

Table 1

A	B	C	D
1	2	1	4
2	2	8	7
1	3	9	1

Table 2

A	B	C	D
8	2	5	6
2	2	4	2
3	4	2	3

Group by (A,B) and apply max as the aggregation on C. (simply discard D)

We need to iterate over the values and emit Key value part which is coming under as the set of attributes which are coming under grouping operation. And b part (i.e value) which is coming under aggregation operation.

In the given problem statement we need to group a & b that means under key part, a & b will come. We need to aggregate 'c' that means under value section we need to mention elements that are present inside 'c' attribute.

MW1

Key	Value
(1,2)	[10], 9]
(2,3)	[2]
(4,2)	[7]
(6,8)	[3]
(3,2)	[5]

MW2

Key	Value
(1,2)	[1]
(2,2)	[8,4]
(1,3)	[9]
(3,2)	[5]
(3,4)	[2]

A hash function will be applied

MW1

Key	value
(1,2)	[10], [9]
(8,3)	[2]
(4,2)	[7]

Key	Value
(6,8)	[3]
(3,2)	[5]

MW2

Key	Value
(1,2)	[1]
(2,2)	[8,4]
(1,3)	[9]

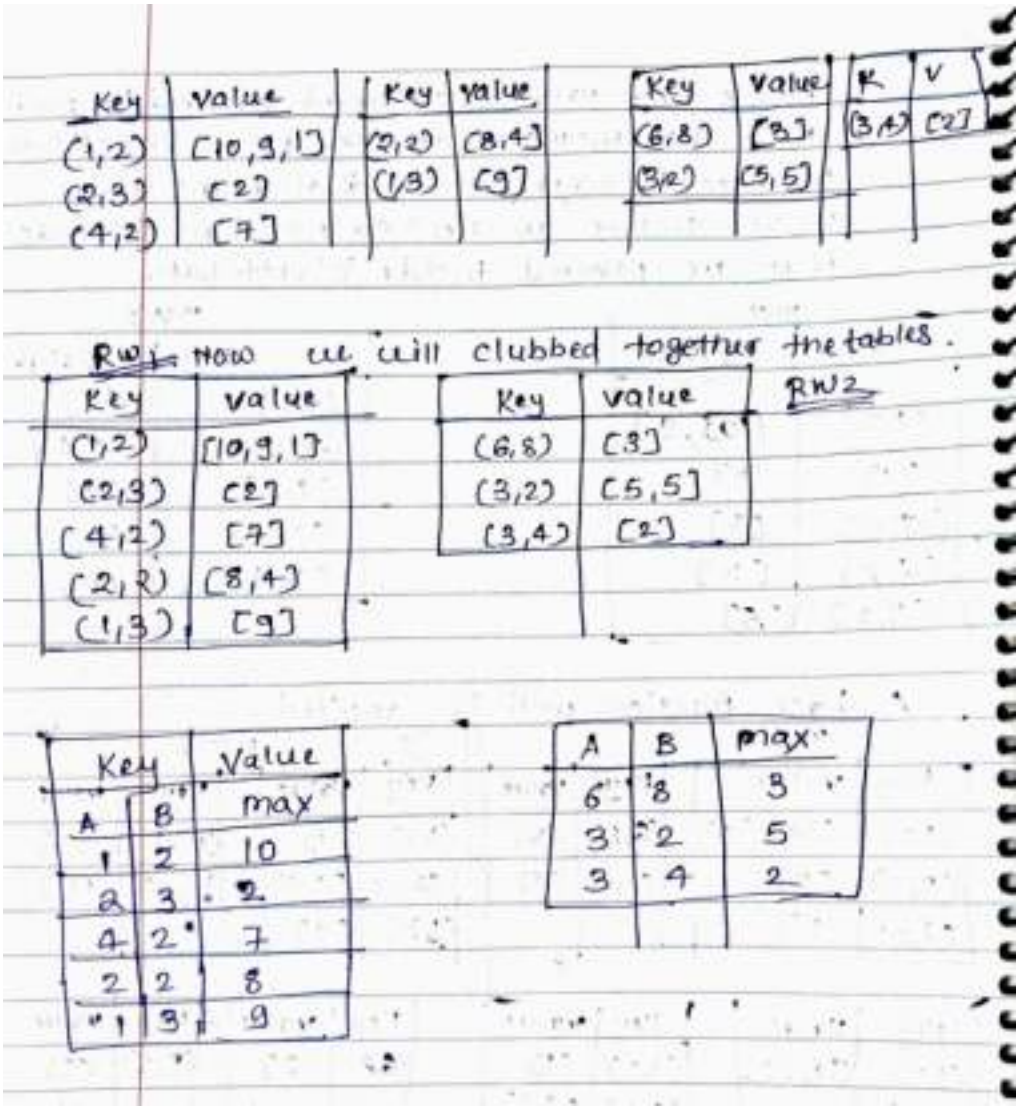
Key	value
(3,2)	[5]
(3,4)	[2]

Key	Value
(1,2)	[10] [9]
(2,3)	[2]
(4,2)	[7]

Key	value
(1,2)	[1]
(2,2)	[8,4]
(1,3)	[9]

Key	Value
(6,8)	[3]
(3,2)	[5]

Key	value
(3,2)	[5]
(3,4)	[2]



Q.5) Explain how node failure is handled in Hadoop. [5M] May 2023

Coping with Node Failures:

There are 3 types of node failures:

- Master node failure
- Map worker node failure
- Reducer worker node failure.

- If the master node fails then entire process has to be restarted. This is worst kind of failure.
- If map worker node fails then master will assign the task to some othe available worker node even if the task had completed.
- If reducer work node fails then the tasks are simply rescheduled on some other Reduce worker later.

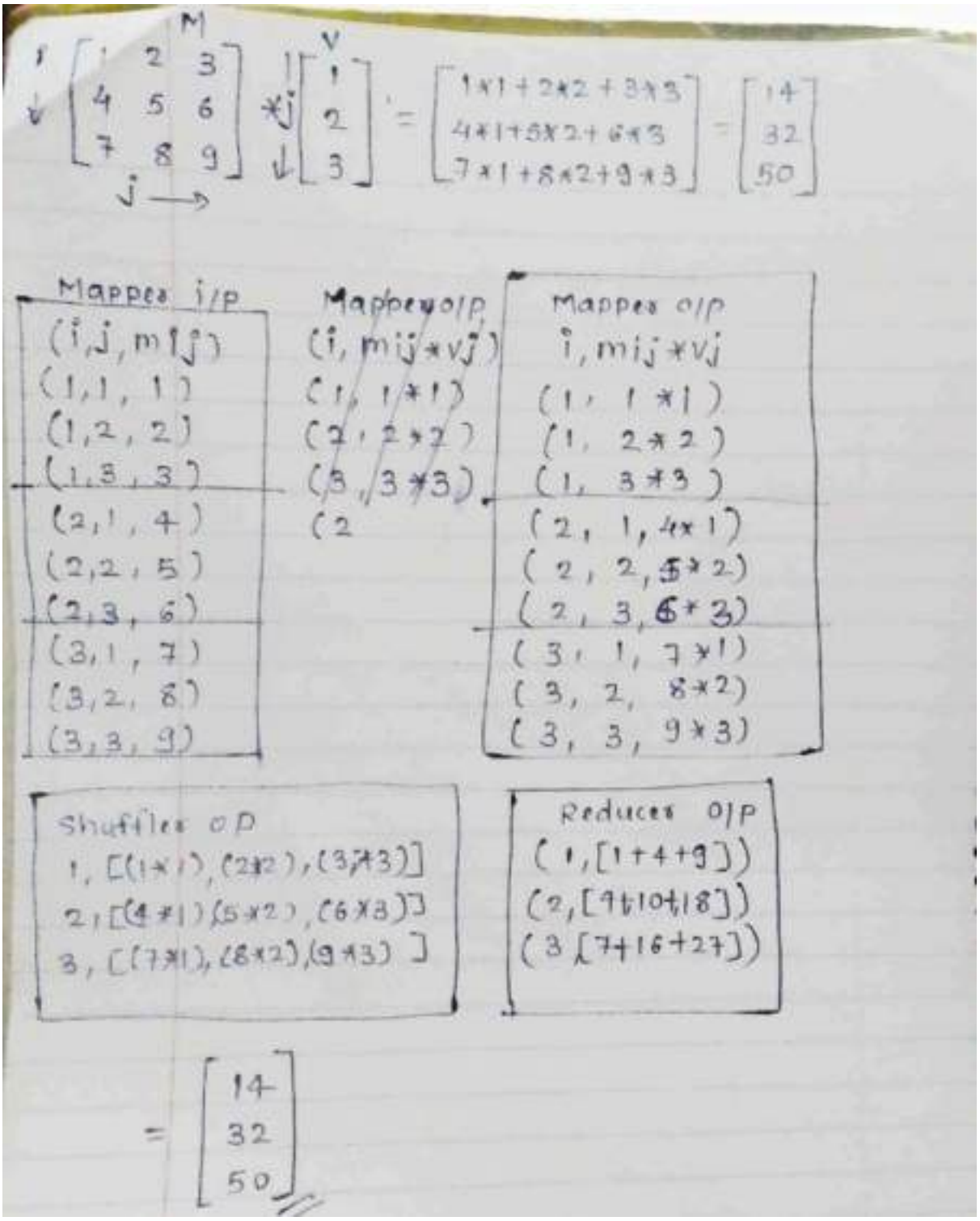
What happens if one of the Datanodes gets failed in HDFS?

Namenode periodically receives a heartbeat and a Block report from each Datanode in the cluster. Every Datanode sends heartbeat message after every **3 seconds** to Namenode. The health report is just information about a particular Datanode that is working properly or not. In the other words we can say that particular Datanode is alive or not. A block report of a particular Datanode contains information about all the blocks on that resides on the corresponding Datanode. When Namenode doesn't receive any heartbeat message for **10 minutes**(ByDefault) from a particular Datanode then corresponding Datanode is considered Dead or failed by Namenode. Since blocks will be under replicated, the system starts the replication process from one Datanode to another by taking all block information from the Block report of corresponding Datanode. The Data for replication transfers directly from one Datanode to another without data passing through Namenode.

Write a map reduce pseudo code to multiply two matrices. Apply map reduce working to perform following matrix multiplication. [10]

$$M = \begin{matrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{matrix} \quad X \quad V = \begin{matrix} 1 \\ 2 \\ 3 \end{matrix}$$

Dec 2023



Note: Also include MapReduce PseudoCode.

Q.6) Explain MapReduce execution pipeline with suitable example [5M] and

Q.7) **Ans:**

Write a map reduce pseudo code for word count problem. Illustrate with an example [10] showing all the steps.

May 2024

Ans:

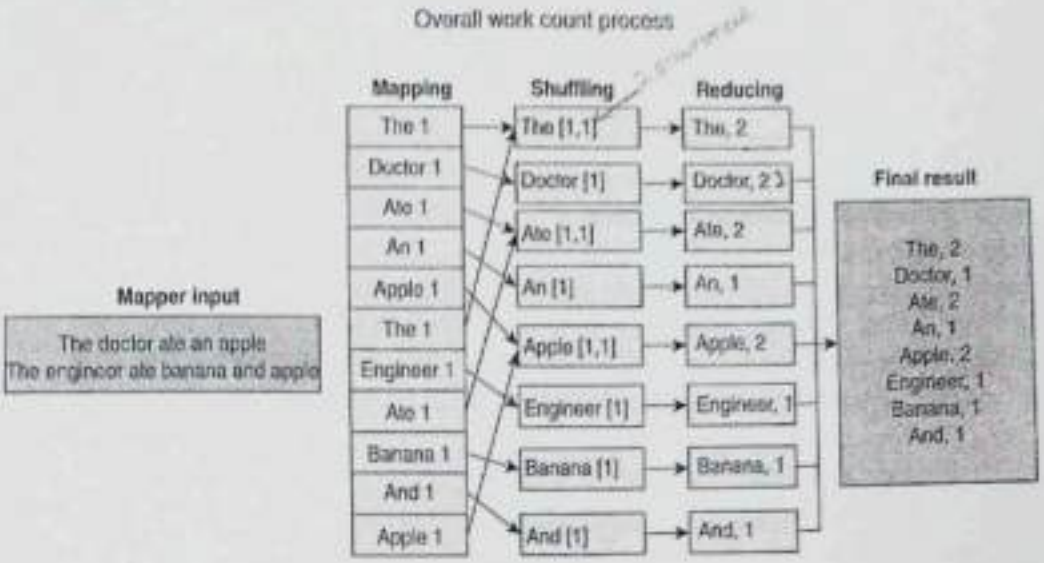


Figure 4.2 Word count using MapReduce algorithm.

Algorithm

Pseudo Code for Word Count Using MapReduce Algorithm

Word Count using MapReduce

→ count occurrences of each word input.

M = large corpus of text.

Mapper:

each word in M → ("word", 1)

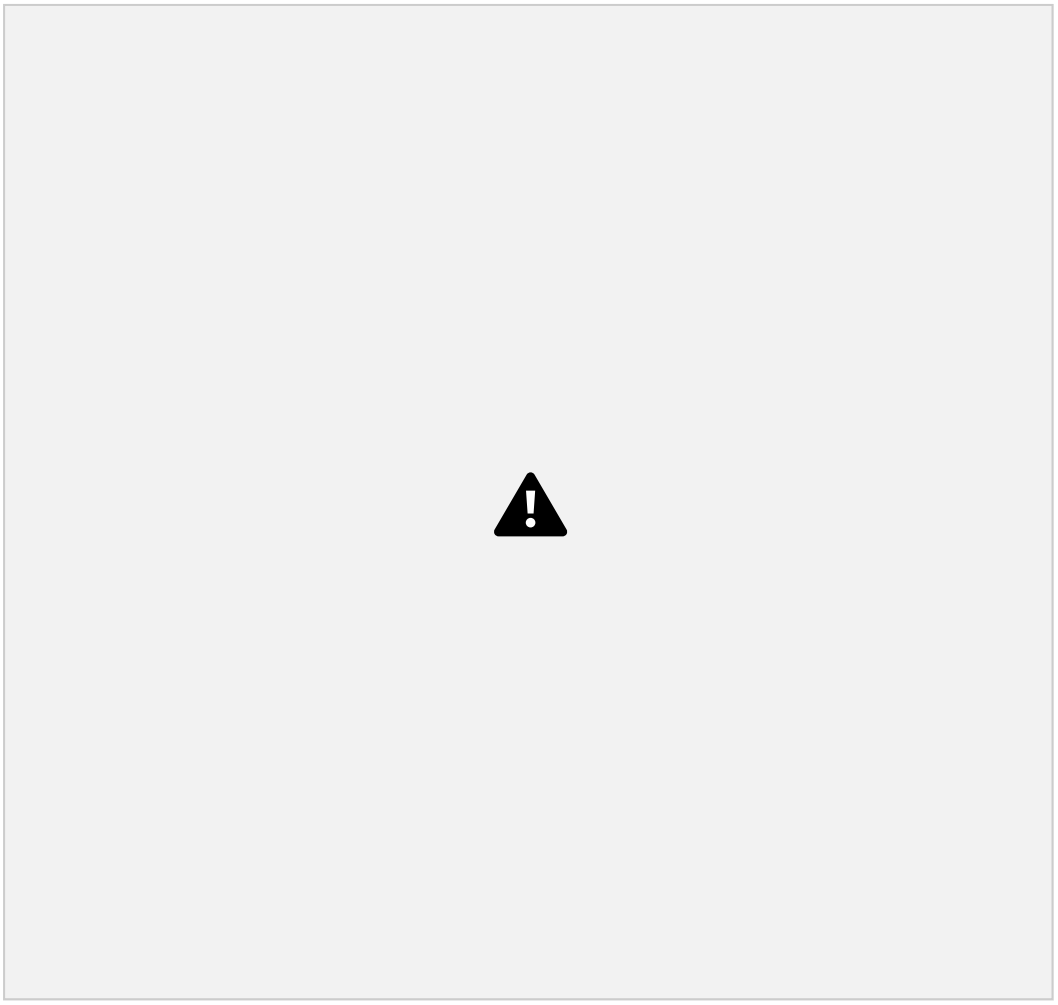
Reducer:

for all ("word", *v_i*)

→ ("word", sum_*iv_i*)

"aggregate" in Hadoop.

OR



Q.8)

May 2024 **Ans:**