**Aim : Matrix Multiplication using MapReduce.**

MapReduce is a technique in which a huge program is subdivided into small tasks and run parallelly to make computation faster, save time, and mostly used in distributed systems. It has 2 important parts:

Mapper : It takes raw data input and organizes into key, value pairs. For example, In a dictionary, you search for the word “Data” and its associated meaning is “facts and statistics collected together for reference or analysis”. Here the Key is Data and the Value associated with is facts and statistics collected together for reference or analysis.

Reducer : It is responsible for processing data in parallel and produce final output.

MStep 1 : Open terminal > Create two text files of matrix m1.txt and m2.txt.

[cloudera@quickstart ~]$ cd ~/Desktop

[cloudera@quickstart Desktop]$ mkdir matrix\_mul

[cloudera@quickstart Desktop]$ cd matrix\_mul

[cloudera@quickstart matrix\_mul]$ gedit m1.txt

m1

**1 2 3**

**4 5 6**

[cloudera@quickstart matrix\_mul]$ gedit m2.txt

m2

**4 5 6**

**7 8 9**

**1 2 6**

Step 2 : Create Mapper file mapper.py

[cloudera@quickstart matrix\_mul]$ gedit mapper.py



[cloudera@quickstart matrix\_mul]$ cat m1.txt m2.txt | python **mapper.py Mapper Output :**

0 0 0 1

0 1 0 1

0 2 0 1

0 0 1 2

0 1 1 2

0 2 1 2

0 0 2 3

0 1 2 3

0 2 2 3

0 0 0 4

0 1 0 4

0 2 0 4

0 0 1 5

0 1 1 5

0 2 1 5

0 0 2 6

0 1 2 6

0 2 2 6

0 0 0 4

0 1 0 4

0 2 0 4

0 0 1 5

0 1 1 5

0 2 1 5

0 0 2 6

0 1 2 6

0 2 2 6

0 0 0 7

0 1 0 7

0 2 0 7

0 0 1 8

0 1 1 8

0 2 1 8

0 0 2 9

0 1 2 9

0 2 2 9

0 0 0 1

0 1 0 1

0 2 0 1

0 0 1 2

0 1 1 2

0 2 1 2

0 0 2 6

0 1 2 6

0 2 2 6

Step 3 : Create Reducer file reducer.py

[cloudera@quickstart matrix\_mul]$ gedit reducer.py



Step 4 : to get matrix multiplication :

[cloudera@quickstart matrix\_mul]$ cat m1.txt m2.txt | python mapper.py | python reducer.py

**Output :**

**[7, 37, 91]**

**[61, 127, 217]**