MapReduce

Matrix Multiplication

A.Donoorkans
229317V

Introduction

- MapReduce operates on key-value pairs
- Accepts key-value pairs as input
- Produces key-value pairs as output

Matrix Multiplication

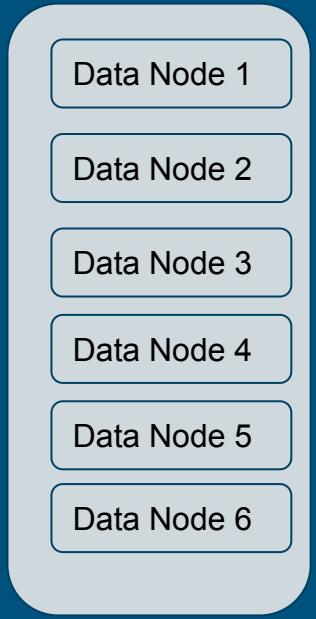
1. Input

$$\begin{pmatrix} 10 & 20 & 30 \\ 40 & 50 & 60 \end{pmatrix} \quad \begin{pmatrix} 70 & 80 \\ 90 & 100 \\ 110 & 120 \end{pmatrix}$$

File 1 - ab_matrix.txt

```
A 1 1 10
A 1 2 20
A 1 3 30
A 2 1 40
A 2 2 50
A 2 3 60
B 1 1 70
B 1 2 80
B 2 1 90
B 2 2 100
B 3 1 110
B 3 2 120
```

Hadoop Cluster



2. Split Phase

Matrix A

10	20	30
40	50	60

Key	Value
(A, 1, 1)	10

Key	Value
(A, 1, 2)	20

Key	Value
(A, 1, 3)	30

Key	Value
(A, 2, 1)	40

Key	Value
(A, 2, 2)	50

Key	Value
(A, 2, 3)	60

Matrix B

70	80
90	100
110	120

Key	Value
(B, 1, 1)	70

Key	Value
(B, 1, 2)	80

Key	Value
(B, 2, 1)	90

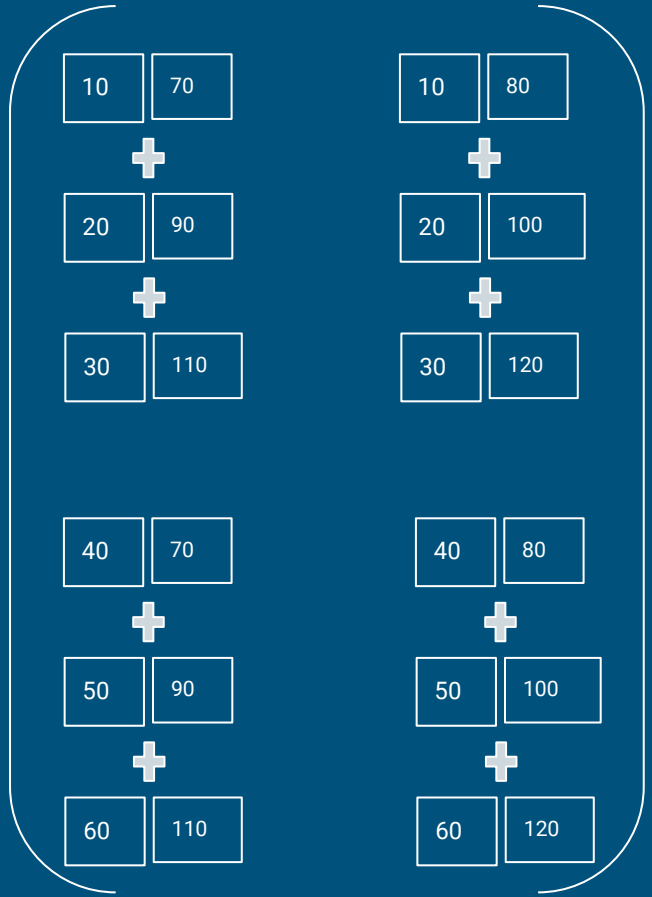
Key	Value
(B, 2, 2)	100

Key	Value
(B, 3, 1)	110

Key	Value
(B, 3, 2)	120

Matrix Multiplication

$$\begin{pmatrix} 10 & 20 & 30 \\ 40 & 50 & 60 \end{pmatrix} \begin{pmatrix} 70 & 80 \\ 90 & 100 \\ 110 & 120 \end{pmatrix} =$$



3. Map Phase

Matrix A is (2 x 3) rows(i) = 2, Columns (j) = 3

(key, value) -> Mapper_A -> ((i, k), (A, j, Aij)) for all k

k.values = [1, 2]

k = 1	i=1	j=1	((1, 1), (A, 1, 10))
		j=2	((1, 1), (A, 2, 20))
		j=3	((1, 1), (A, 3, 30))
	i=2	j=1	((2, 1), (A, 1, 40))
		j=2	((2, 1), (A, 2, 50))
		j=3	((2, 1), (A, 3, 60))
k = 2	i=1	j=1	((1, 2), (A, 1, 10))
		j=2	((1, 2), (A, 2, 20))
		j=3	((1, 2), (A, 3, 300))
	i=2	j=1	((2, 2), (A, 1, 30))
		j=2	((2, 2), (A, 2, 40))
		j=3	((2, 2), (A, 3, 20))

Matrix B is (3 x 2) rows(j) = 3, Columns (k) = 2

(key, value) -> Mapper_B -> ((i, k), (A, j, Aij)) for all i

i.values = [1, 2]

i = 1	j=1	k=1	((1, 1), (B, 1, 70))
		k=2	((1, 2), (B, 1, 80))
	j=2	k=1	((1, 1), (B, 2, 90))
		k=2	((1, 2), (B, 2, 100))
	j=3	k=1	((1, 1), (B, 3, 110))
		k=2	((1, 2), (B, 3, 120))
i = 2	j=1	k=1	((2, 1), (B, 1, 70))
		k=2	((2, 2), (B, 1, 80))
	j=2	k=1	((2, 1), (B, 2, 90))
		k=2	((2, 2), (B, 2, 100))
	j=3	k=1	((2, 1), (B, 3, 110))
		k=2	((2, 2), (B, 3, 120))

4. Shuffling Phase

k = 1	i=1	j=1	((1, 1), (A, 1, 10))
		j=2	((1, 1), (A, 2, 20))
		j=3	((1, 1), (A, 3, 30))
	i=2	j=1	((2, 1), (A, 1, 40))
		j=2	((2, 1), (A, 2, 50))
		j=3	((2, 1), (A, 3, 60))
k = 2	i=1	j=1	((1, 2), (A, 1, 10))
		j=2	((1, 2), (A, 2, 20))
		j=3	((1, 2), (A, 3, 30))
	i=2	j=1	((2, 2), (A, 1, 40))
		j=2	((2, 2), (A, 2, 50))
		j=3	((2, 2), (A, 3, 60))
i = 1	j=1	k=1	((1, 1), (B, 1, 70))
		k=2	((1, 2), (B, 1, 80))
	j=2	k=1	((1, 1), (B, 2, 90))
		k=2	((1, 2), (B, 2, 100))
	j=3	k=1	((1, 1), (B, 3, 110))
		k=2	((1, 2), (B, 3, 120))
i = 2	j=1	k=1	((2, 1), (B, 1, 70))
		k=2	((2, 2), (B, 1, 80))
	j=2	k=1	((2, 1), (B, 2, 90))
		k=2	((2, 2), (B, 2, 100))
	j=3	k=1	((2, 1), (B, 3, 110))
		k=2	((2, 2), (B, 3, 120))

Key	Value
(1, 1)	(A, 1, 10)
(1, 1)	(A, 2, 20)
(1, 1)	(A, 3, 30)
(1, 1)	(B, 1, 70)
(1, 1)	(B, 2, 90)
(1, 1)	(B, 3, 110)

Key	Value
(1, 2)	(A, 1, 10)
(1, 2)	(A, 2, 20)
(1, 2)	(A, 3, 30)
(1, 2)	(B, 1, 80)
(1, 2)	(B, 2, 100)
(1, 2)	(B, 3, 120)

Key	Value
(2, 1)	(A, 1, 40)
(2, 1)	(A, 2, 50)
(2, 1)	(A, 3, 60)
(2, 1)	(B, 1, 70)
(2, 1)	(B, 2, 90)
(2, 1)	(B, 3, 110)

Key	Value
(2, 2)	(A, 1, 40)
(2, 2)	(A, 2, 50)
(2, 2)	(A, 3, 60)
(2, 2)	(B, 1, 80)
(2, 2)	(B, 2, 100)
(2, 2)	(B, 3, 120)

4. Reducer Phase

Reducer (key, value) = (i, k)

(i, k) => Summation (Aij * Bjk) for j

output => ((i, k), sum)

Key	Value
(1, 1)	(A, 1, 10)
(1, 1)	(A, 2, 20)
(1, 1)	(A, 3, 30)
(1, 1)	(B, 1, 70)
(1, 1)	(B, 2, 90)
(1, 1)	(B, 3, 110)

Key	Value
(1, 2)	(A, 1, 10)
(1, 2)	(A, 2, 20)
(1, 2)	(A, 3, 30)
(1, 2)	(B, 1, 80)
(1, 2)	(B, 2, 100)
(1, 2)	(B, 3, 120)

Key	Value
(2, 1)	(A, 1, 40)
(2, 1)	(A, 2, 50)
(2, 1)	(A, 3, 60)
(2, 1)	(B, 1, 70)
(2, 1)	(B, 2, 90)
(2, 1)	(B, 3, 110)

Key	Value
(2, 2)	(A, 1, 40)
(2, 2)	(A, 2, 50)
(2, 2)	(A, 3, 60)
(2, 2)	(B, 1, 80)
(2, 2)	(B, 2, 100)
(2, 2)	(B, 3, 120)

Key	Value
(1, 1)	5800

Key	Value
(1, 2)	6400

Key	Value
(2, 1)	13900

Key	Value
(2, 2)	15400