# MapReduce and the New Software Stack

## Define,Applications,Algorithm

Dr Khajevand

Mohammad.F T.Saeedi

Apr 2021

University of Qazvin

# Content

Parallel Computing

# What is Parallel Computing?

Parallel computing is a type of computation in which many calculations or the execution of processes are carried out simultaneously ...

Large problems can often be divided into smaller ones, which can then be solved at the same time

# Distributed File System (DFS)

Distributed File Systems Divide a huge data into smaller data

Supports data duplication on different nodes

If the node or rack fails, it can replace another parallel data

# Cluster Computing

Thousands or millions of simple and inexpensive computers that are connected to each other in cluster format.

A common architecture for very large-scale applications is a cluster of compute nodes (processor chip, main memory, and disk). Compute nodes are mounted in racks, and the nodes on a rack are connected, typically by gigabit Ethernet Racks are also connected by a high-speed network or switch.

# Computing Node

A node inside a computing cluster is called a computing node

The core of mapreduce operations is performed on these systems

# Aplication of Mapreduce programing

# MapReduce

**DFS**

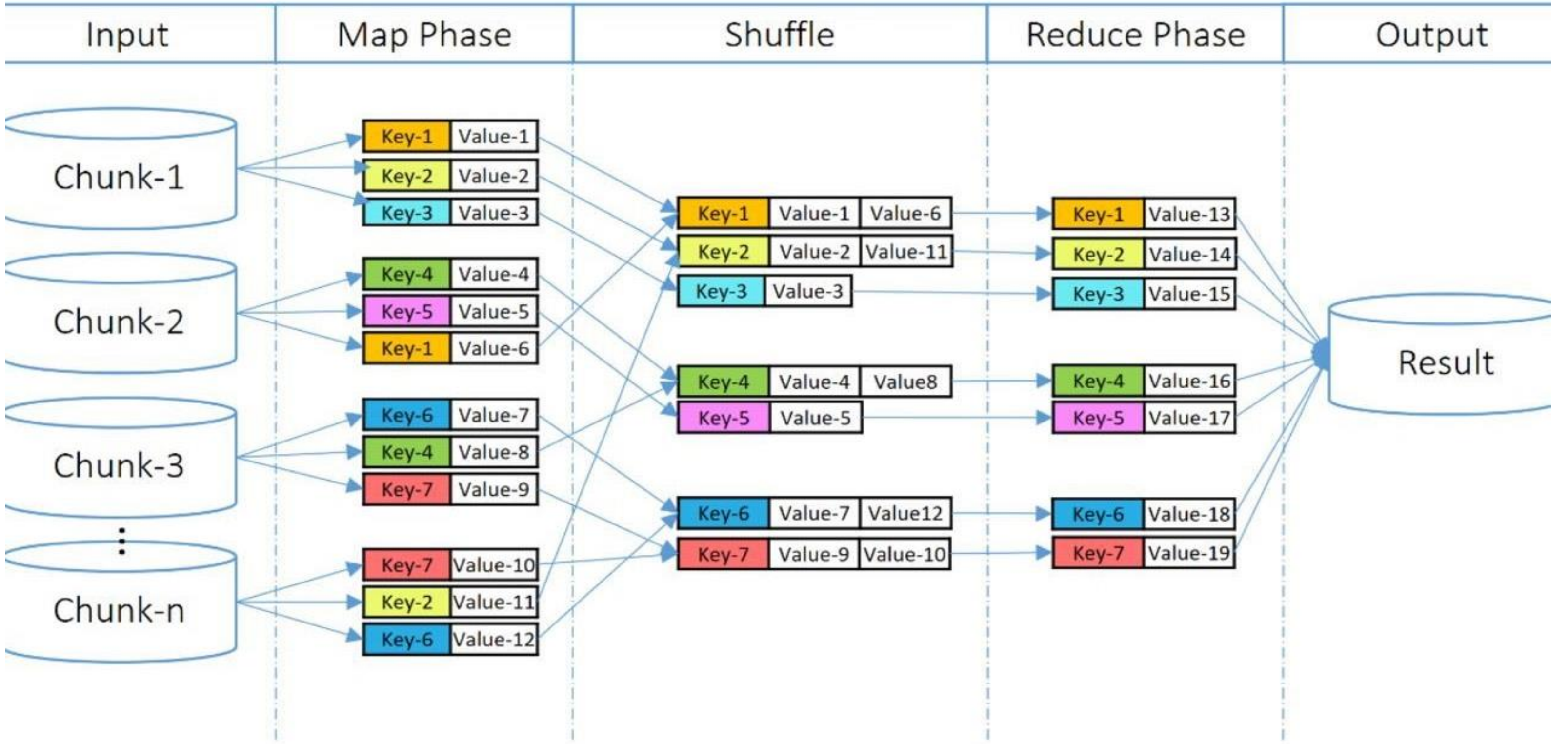
**Chunks**

**Map Phase**

**Shuffle**

**Reduce Phase**

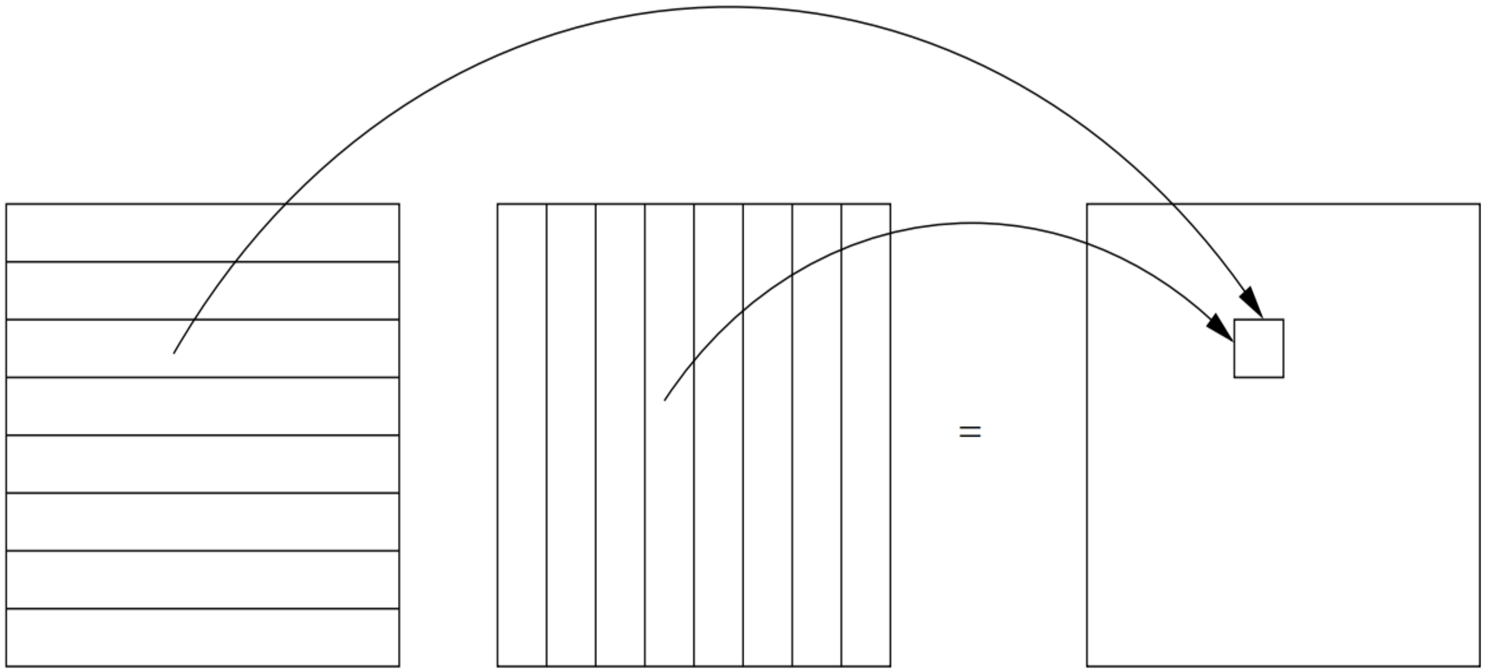
This programming system allows one to exploit parallelism inherent in cluster computing, and manages the hardware failures that can occur during a long computation on many nodes.



# Matrix Multiplication

## One Pass

$$N = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix} \qquad M = \begin{bmatrix} 1 & 3 \\ 5 & 7 \end{bmatrix} \qquad x = \sum_{j=1}^{n} M_{ij} \times N_{jk}$$



# Matrix Multiplication

## Map Task

### One Pass

$$N = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix} \quad M = \begin{bmatrix} 1 & 3 \\ 5 & 7 \end{bmatrix} \quad x = \sum_{j=1}^{n} M_{ij} \times N_{jk}$$

**Enter Map**

| | | |
|---|---|---|
| $i$ | $j$ | $M_{ij}$ |
| 1 | 1 | $M_{ij}$ |
| 1 | 2 | $M_{ij}$ |
| 2 | 1 | $M_{ij}$ |
| 2 | 2 | $M_{ij}$ |
| $j$ | $k$ | $N_{jk}$ |
| 1 | 1 | $N_{jk}$ |
| 1 | 2 | $N_{jk}$ |
| 2 | 1 | $N_{jk}$ |
| 2 | 2 | $N_{jk}$ |

**Exit Map**

| | |
|---|---|
| $(i, K)$ | $(M, j, M_{ij})$ |
| (1,1) | ( $M$, 1, 1 ) |
| (1,2) | ( $M$, 1, 1 ) |
| (1,1) | ( $M$, 2, 3 ) |
| (1,2) | ( $M$, 2, 3 ) |
| (2,1) | ( $M$, 1, 5 ) |
| (2,2) | ( $M$, 1, 5 ) |
| (2,1) | ( $M$, 2, 7 ) |
| (2,2) | ( $M$, 2, 7 ) |
| (1,1) | ( $N$, 1, 2 ) |
| (1,2) | ( $N$, 1, 4 ) |
| (1,1) | ( $N$, 2, 6 ) |
| (1,2) | ( $N$, 2, 8 ) |
| (2,1) | ( $N$, 1, 2 ) |
| (2,2) | ( $N$, 1, 4 ) |
| (2,1) | ( $N$, 2, 6 ) |
| (2,2) | ( $N$, 2, 8 ) |

# Matrix Multiplication

# Reduce Task

## One Pass

$$N = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix} \qquad M = \begin{bmatrix} 1 & 3 \\ 5 & 7 \end{bmatrix} \qquad x = \sum_{j=1}^{n} M_{ij} \times N_{jk}$$

### Enter Reduce

$(1,1) \quad , [(M, 1,1), (M, 2,3)], [(N, 1,2), (N, 2,6)]$

$(1,2) \quad , [(M, 1,1), (M, 2,3)], [(N, 1,4), (N, 2,8)]$

$(2,1) \quad , [(M, 1,5), (M, 2,7)], [(N, 1,2), (N, 2,6)]$

$(2,2) \quad , [(M, 1,5), (M, 2,7)], [(N, 1,4), (N, 2,8)]$

### Exit Reduce

$(1,1) \quad , [(1 \times 2) + (3 \times 6)]$

$(1,2) \quad , [(1 \times 4) + (3 \times 8)]$

$(2,1) \quad , [(5 \times 2) + (7 \times 6)]$

$(2,2) \quad , [(5 \times 4) + (7 \times 8)]$

# Thank You For Attention

Dr Khajevand

Mohammad.F T.Saeedi

Jun 2020

University of Qazvin