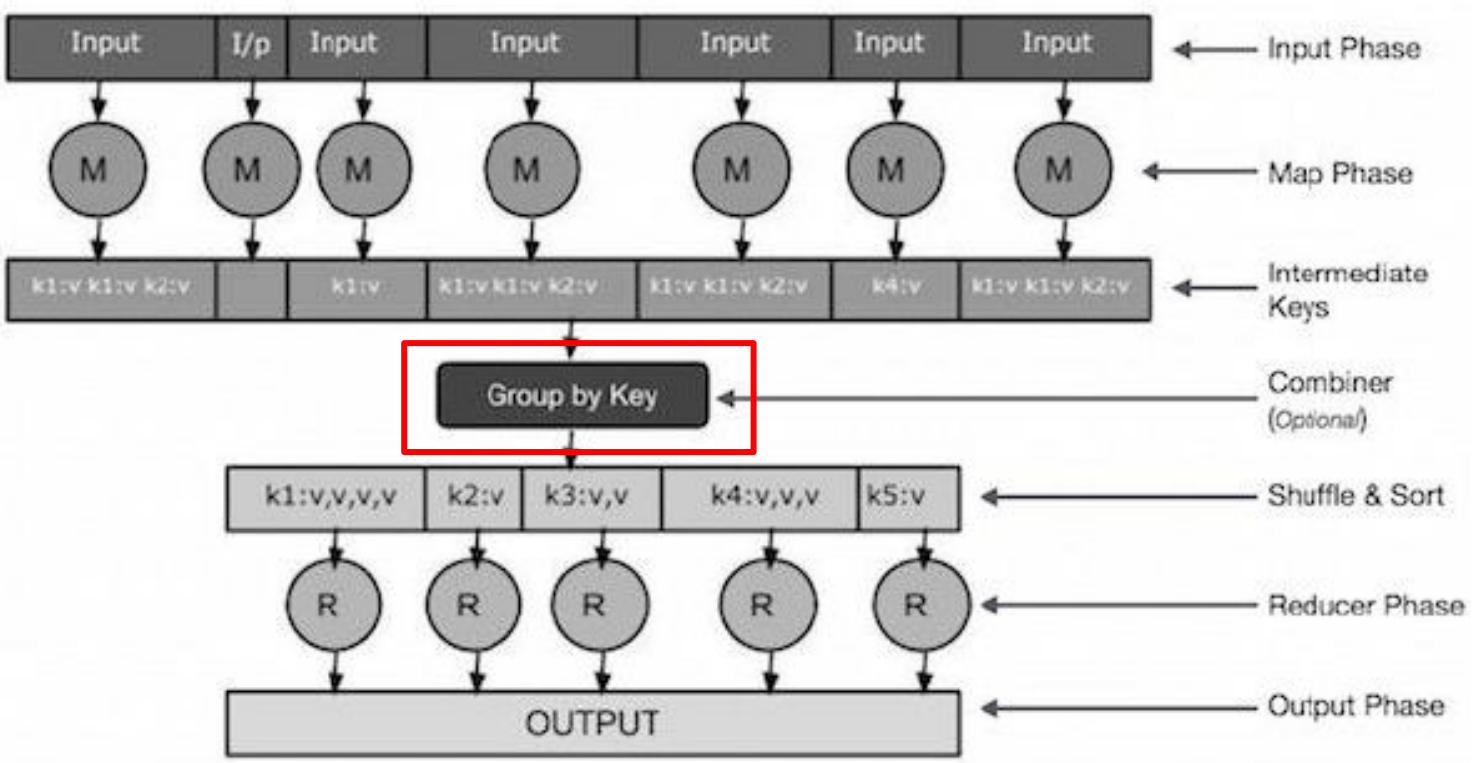
Assignment 7 NoSQL2

Computing Lab - II (CS69012)
IIT Kharagpur

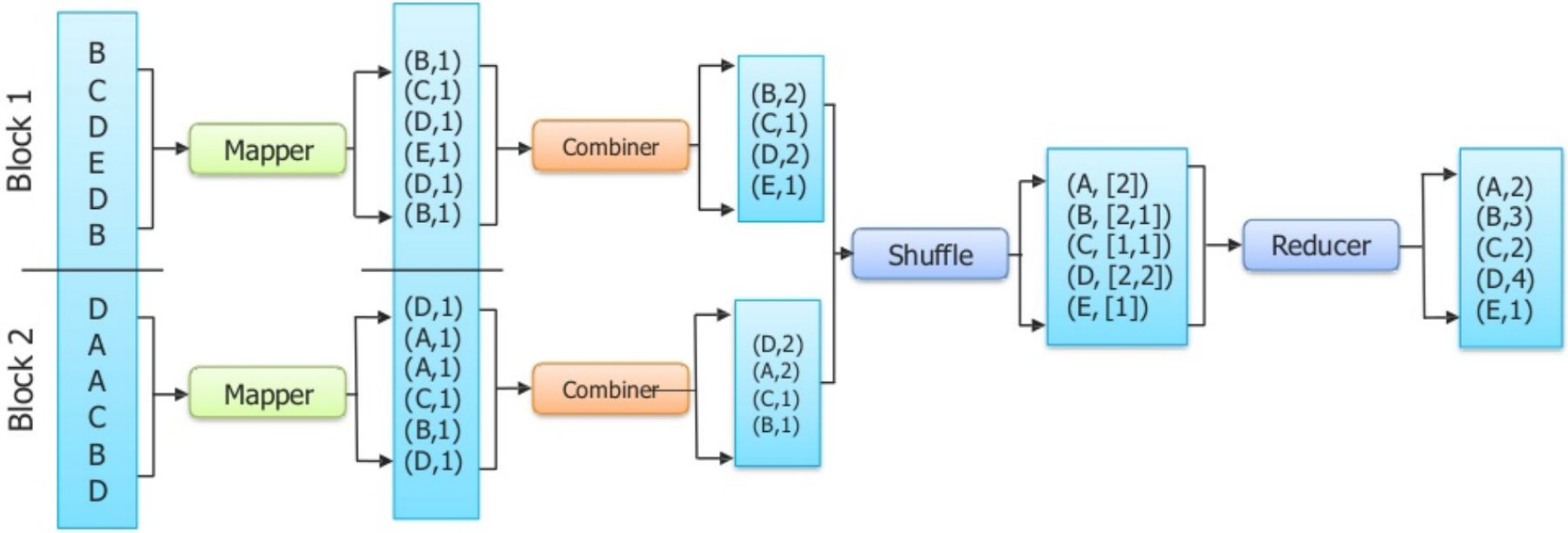
Understanding the Combiner


Input
I/p
Input
Input
Input
Input
Input
Input Phase
M
Map Phase
k1:v k1:v k2:v
k1:v
k1:v k1:v k2:v
k4:v
Intermediate Keys
Group by Key
Combiner
(Optional)
k1:v,v,v,v
k2:v
k3:v,v
k4:v,v,v
k5:v
Shuffle & Sort
R
Reducer Phase
OUTPUT
Output Phase

Hadoop Combiner

- Hadoop Combiner is also known as “Mini-Reducer” that summarizes the Mapper output record with the same Key before passing to the Reducer.
- The combiner in MapReduce is also known as ‘Mini-reducer’

Combiners


Block 1
B
C
D
E
D
B
Mapper
(B,1)
(C,1)
(D,1)
(E,1)
(D,1)
(B,1)
Combiner
(B,2)
(C,1)
(D,2)
(E,1)
Block 2
D
A
A
C
B
D
Mapper
(D,1)
(A,1)
(A,1)
(C,1)
(B,1)
(D,1)
Combiner
(D,2)
(A,2)
(C,1)
(B,1)
Shuffle
(A, [2])
(B, [2,1])
(C, [1,1])
(D, [2,2])
(E, [1])
Reducer
(A,2)
(B,3)
(C,2)
(D,4)
(E,1)

Understanding the Assignment

Data Set:

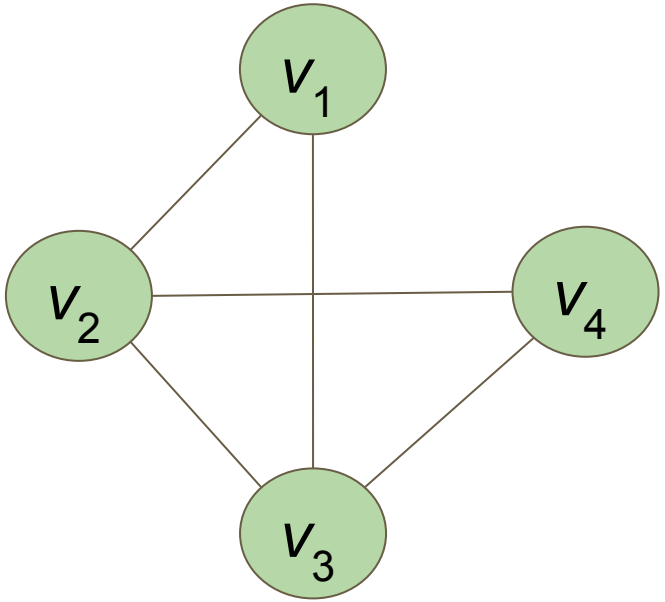
- The dataset describes **chat sessions data** from Facebook.com.
- There are 10 files.
  - “day1.txt”, “day2.txt”, …, “day10.txt”
  - “day***x***.txt” has the logs of chat sessions on $\boldsymbol{x}^{th}$ day
  - Total: 300 users

Task 1: Strongly connected chat network

- *300 users in total*
- Find the pair of users who have chatted at least 30 times in the 10 days
- You have to use combiner in this assignment
- (deliverables: mapper1.py, combiner1.py, reducer1.py, network.txt)

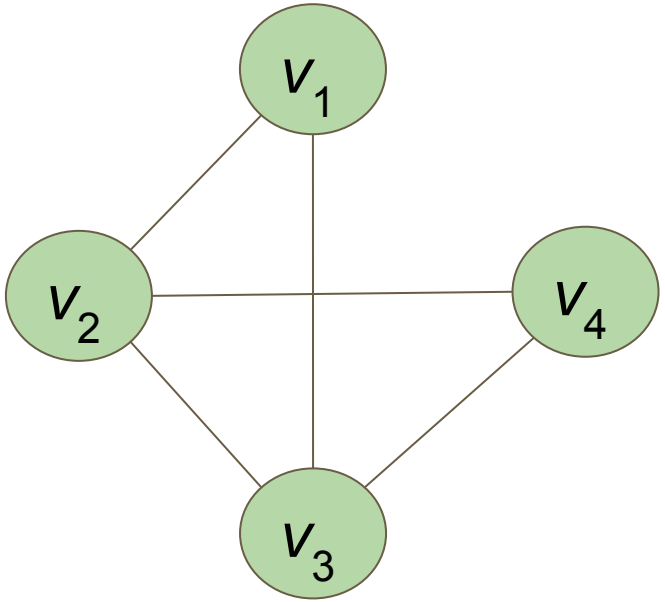
Task 2: Number of mutual friends

- Find the number of mutual friends for each pair of nodes in the chat network
- E.g. in the shown network
  - #mutualFriends(v_1,v_2)=1
  - #mutualFriends(v_2,v_3)=2
- (deliverables: mapper2.py, reducer2.py, friends.txt)



Task 3: Number of triangles

- Find out the number of triangles through each node in the chat network
- E.g. in the shown network
  - #triangles(v_1)=1
  - #triangles(v_2)=2
  - #triangles(v_3)=2
  - #triangles(v_4)=1
- (deliverables: mapper3.py, reducer3.py, triangles.txt)



Thank you.