Amazon's Dynamo

Note Title 08-Dec-22

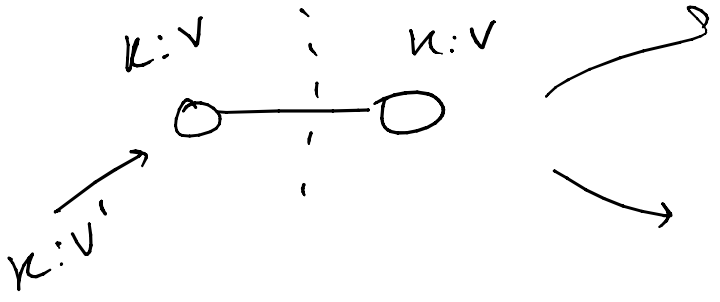
High availability

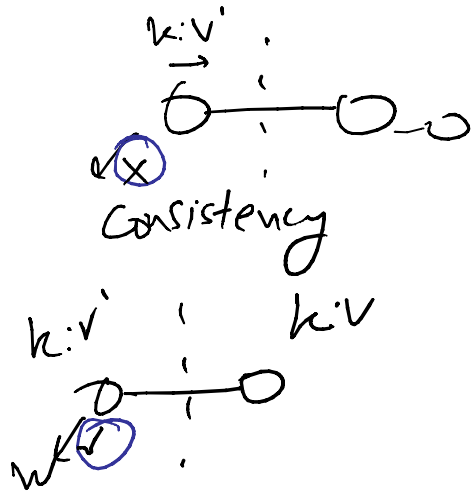
99.9% < 300ms

CAP theorem

Network Partition

PACELC





availability

put (k, val)   ← userId
↑ shopping cart

val = get (k)

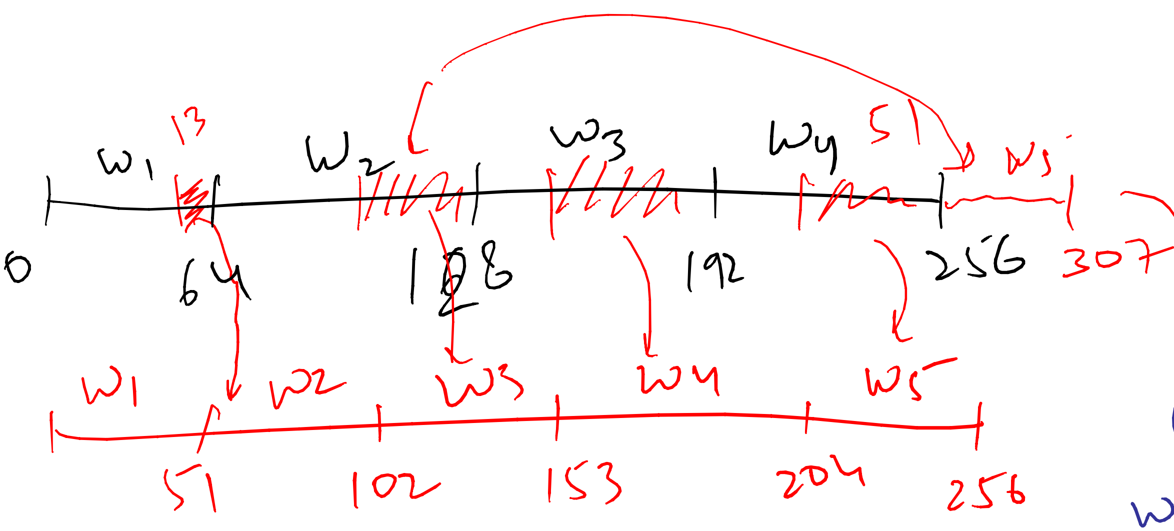
Decentralized storage system

→ Master : Load balancing
failure detection
Addressing
k → server

(k) → replicas

millions   100–1000

read (x) → 42

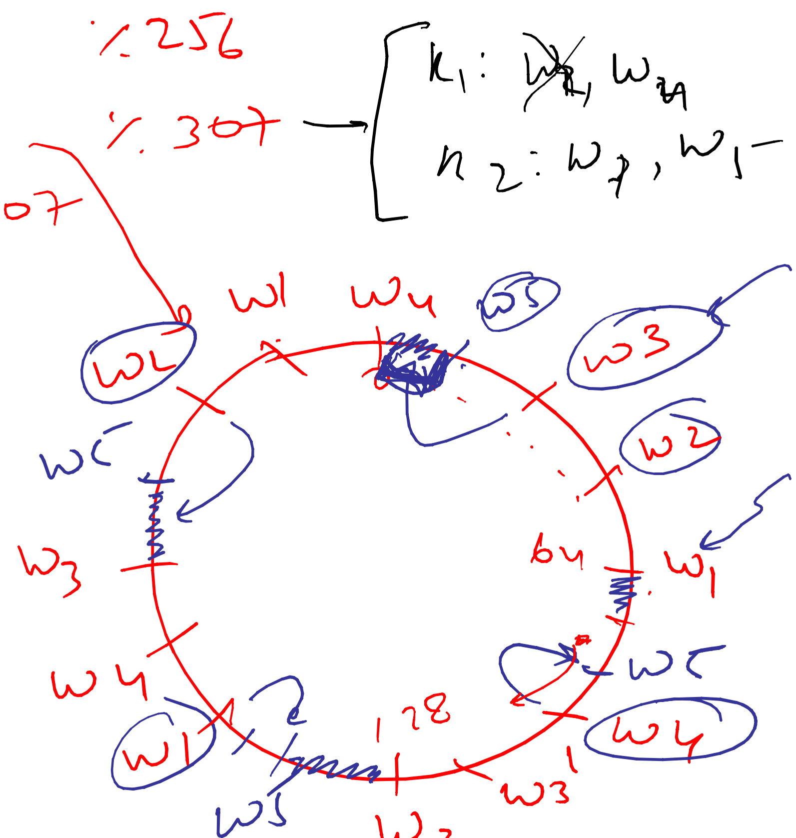


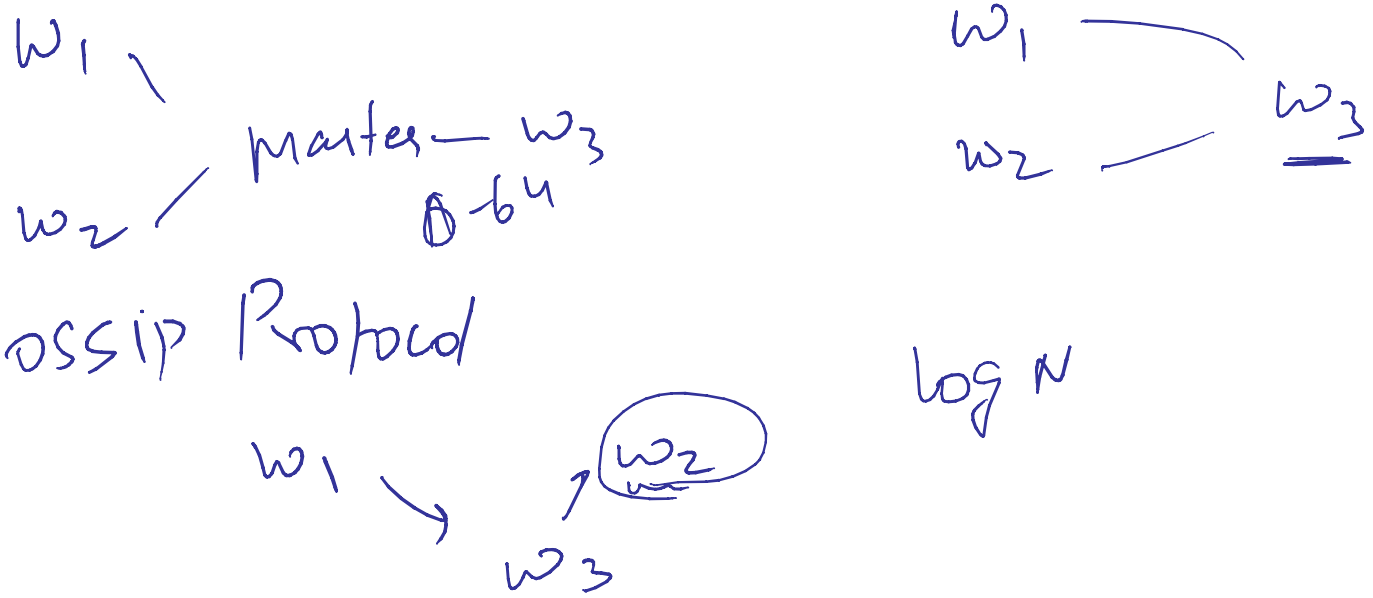
O(Q ranges)

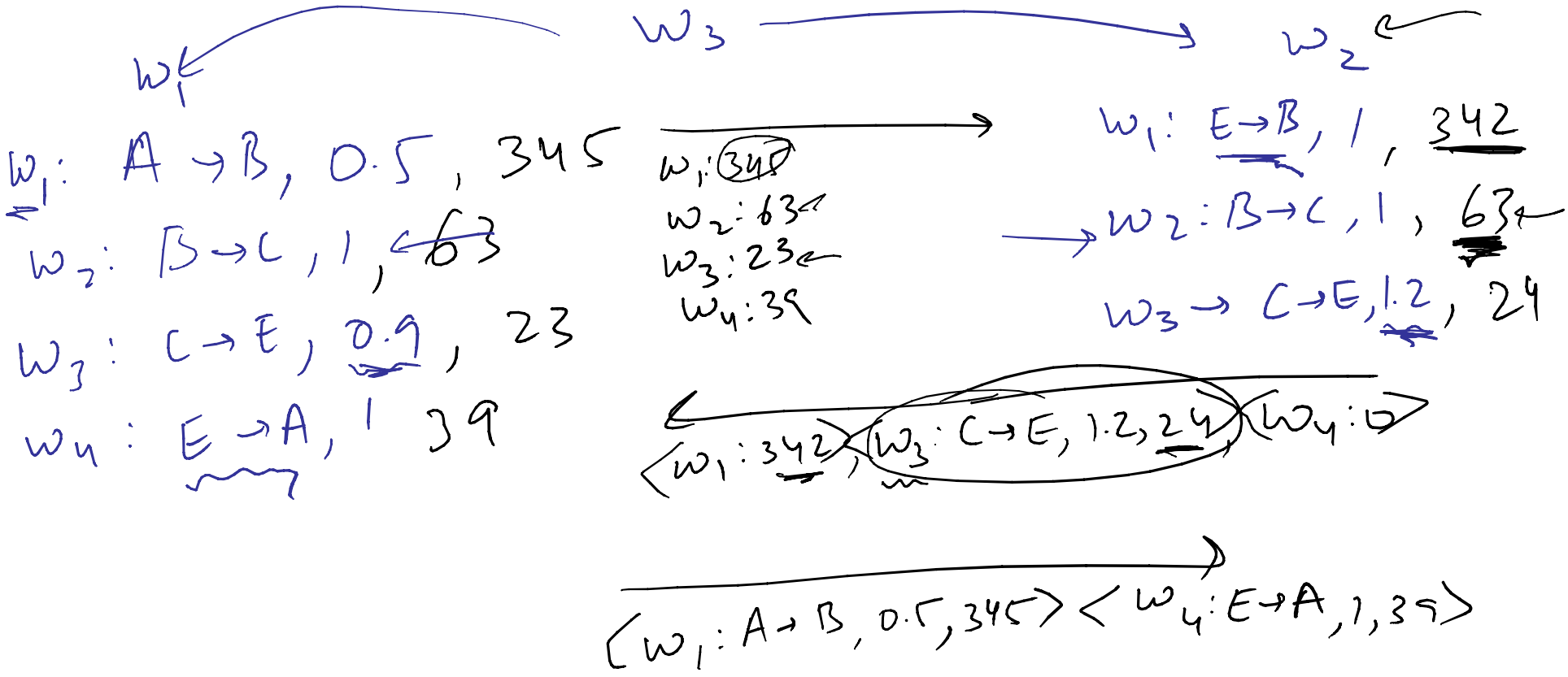
Q/N

Q>>N

Token ring

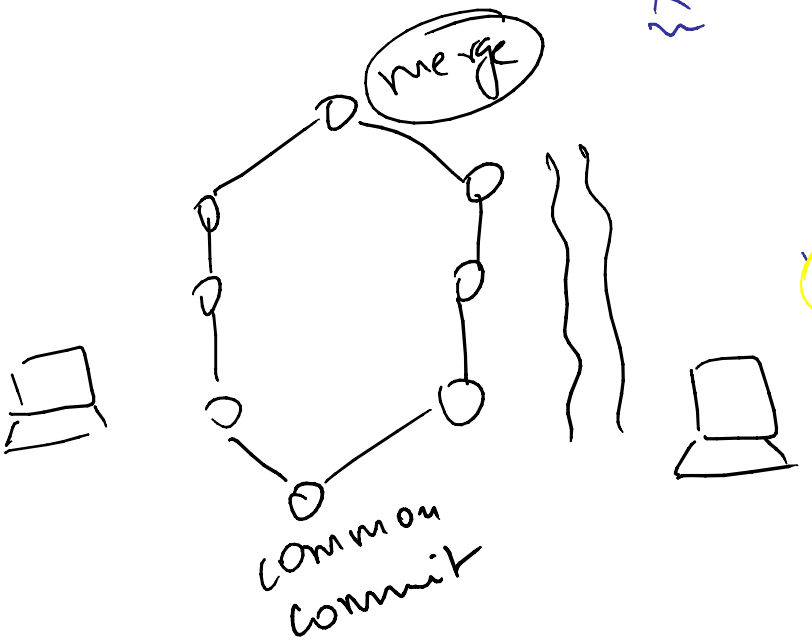


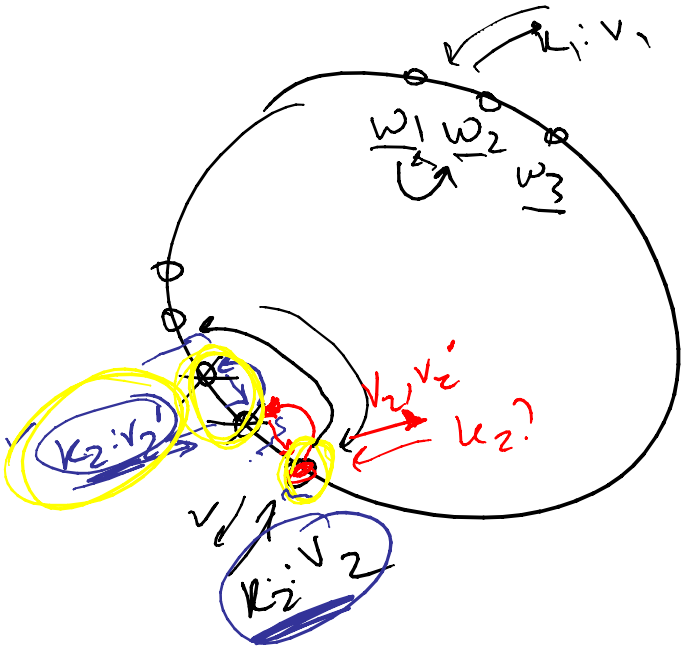

w_1
w_2
master — w_3
0-64
Gossip Protocol
w_1
w_2
w_3
log N
w_1
w_2
w_3


w_3
w_1
w_2
w_1: A → B, 0.5, 345
w_2: B → C, 1, 63
w_3: C → E, 0.9, 23
w_4: E → A, 1, 39
w_1: 345
w_2: 63
w_3: 23
w_4: 39
w_1: E → B, 1, 342
w_2: B → C, 1, 63
w_3 → C → E, 1.2, 24
w_1: 342, w_3: C → E, 1.2, 24, w_4: D
[w_1: A → B, 0.5, 345> < w_4: E → A, 1, 39>

Availability, Latency

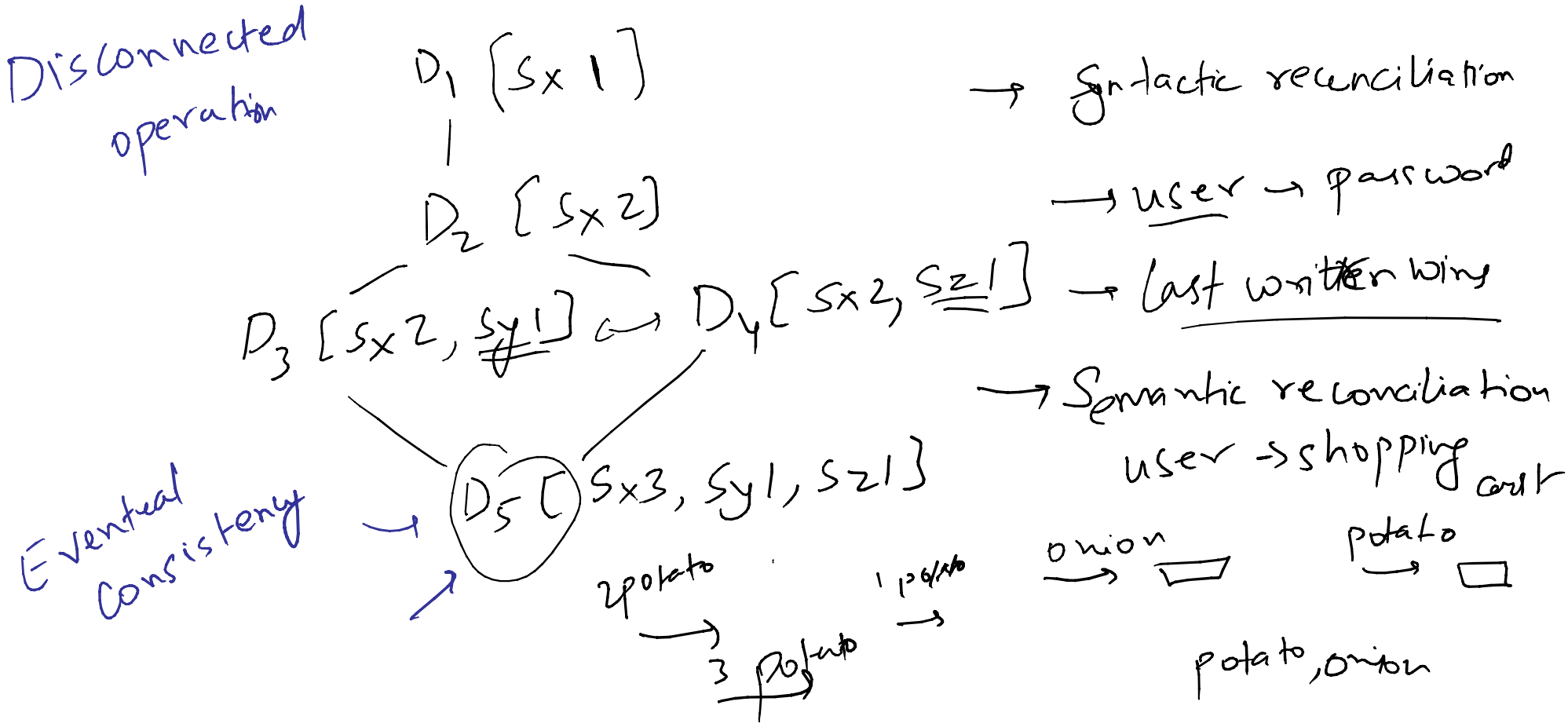
R=2

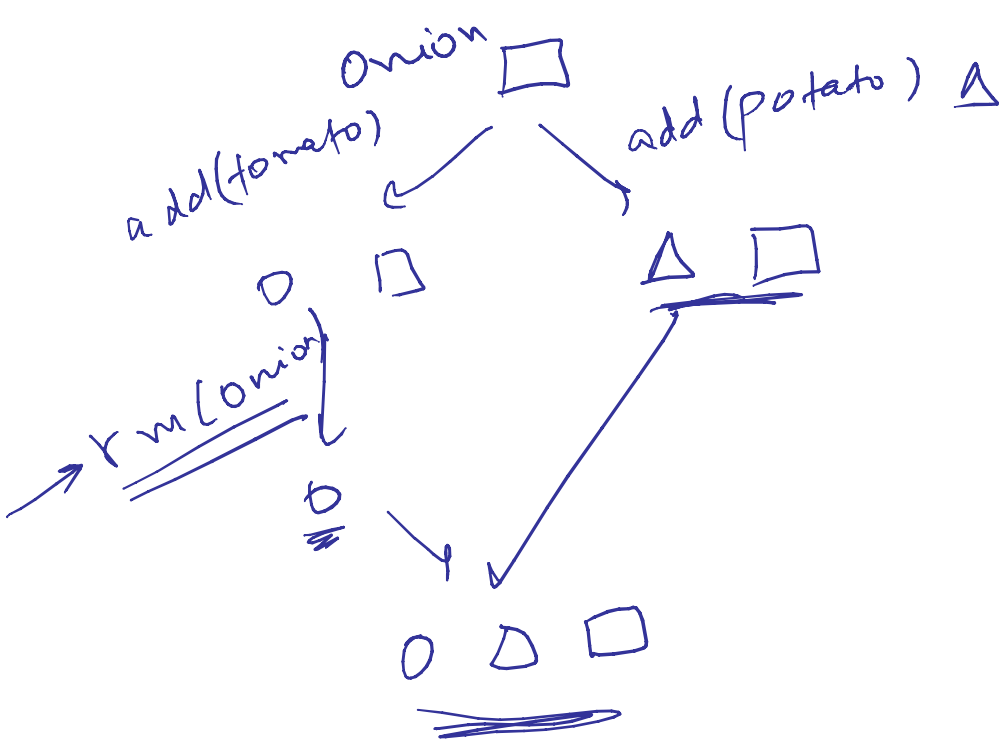




W=2

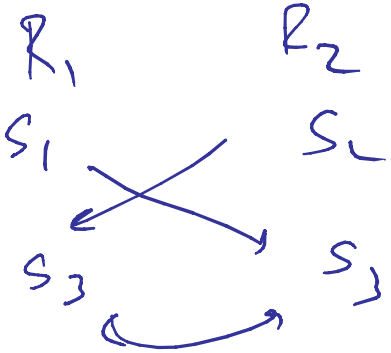
Fault tolerant

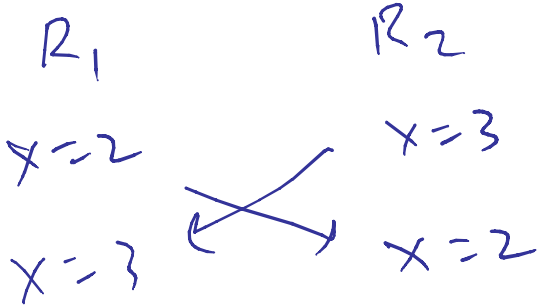

Disconnected operation
D_1 [Sx 1]
D_2 [Sx 2]
D_3 [Sx 2, Sy 1]
D_4 [Sx 2, Sz 1]
D_5 [Sx 3, Sy 1, Sz 1]
Eventual Consistency
→ Syntactic reconciliation
→ user → password
→ Last writer wins
→ Semantic reconciliation
user → shopping cart
onion
potato
2 potato
1 potato
3 potato
potato, onion



Data types

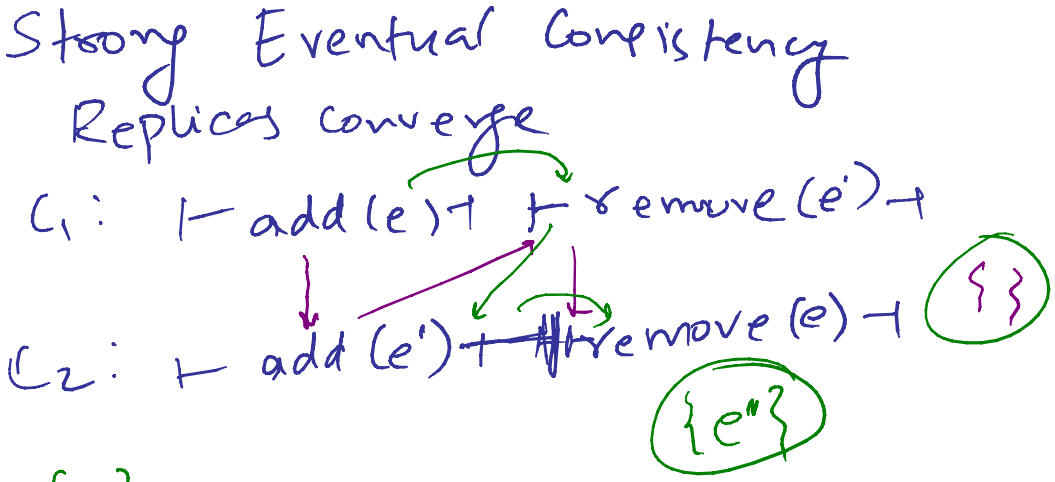
meaningful

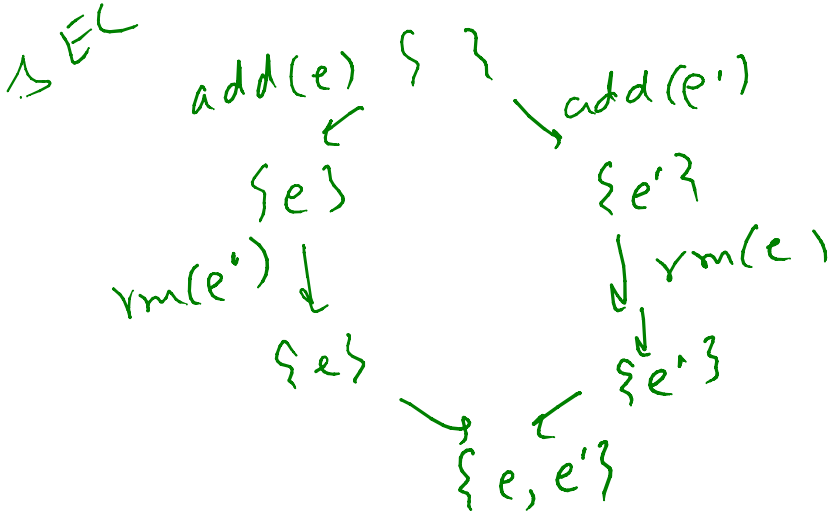




Strong Eventual Consistency

Replicas converge





CRDTs Commutative Replicated Data Types