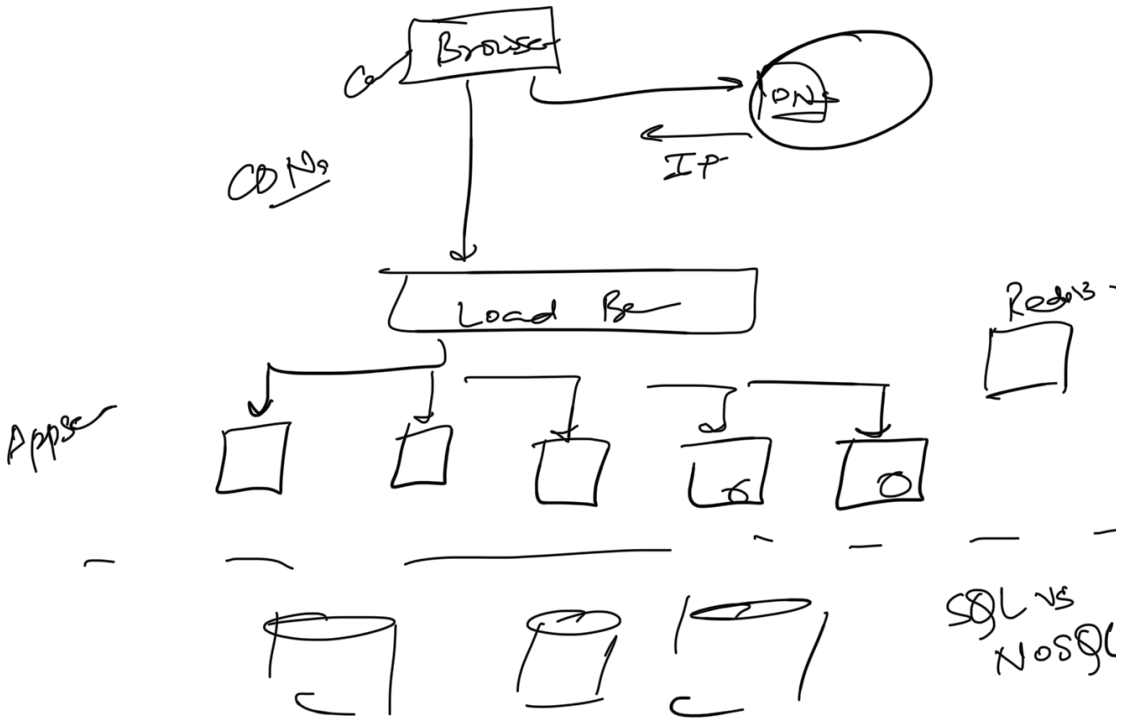
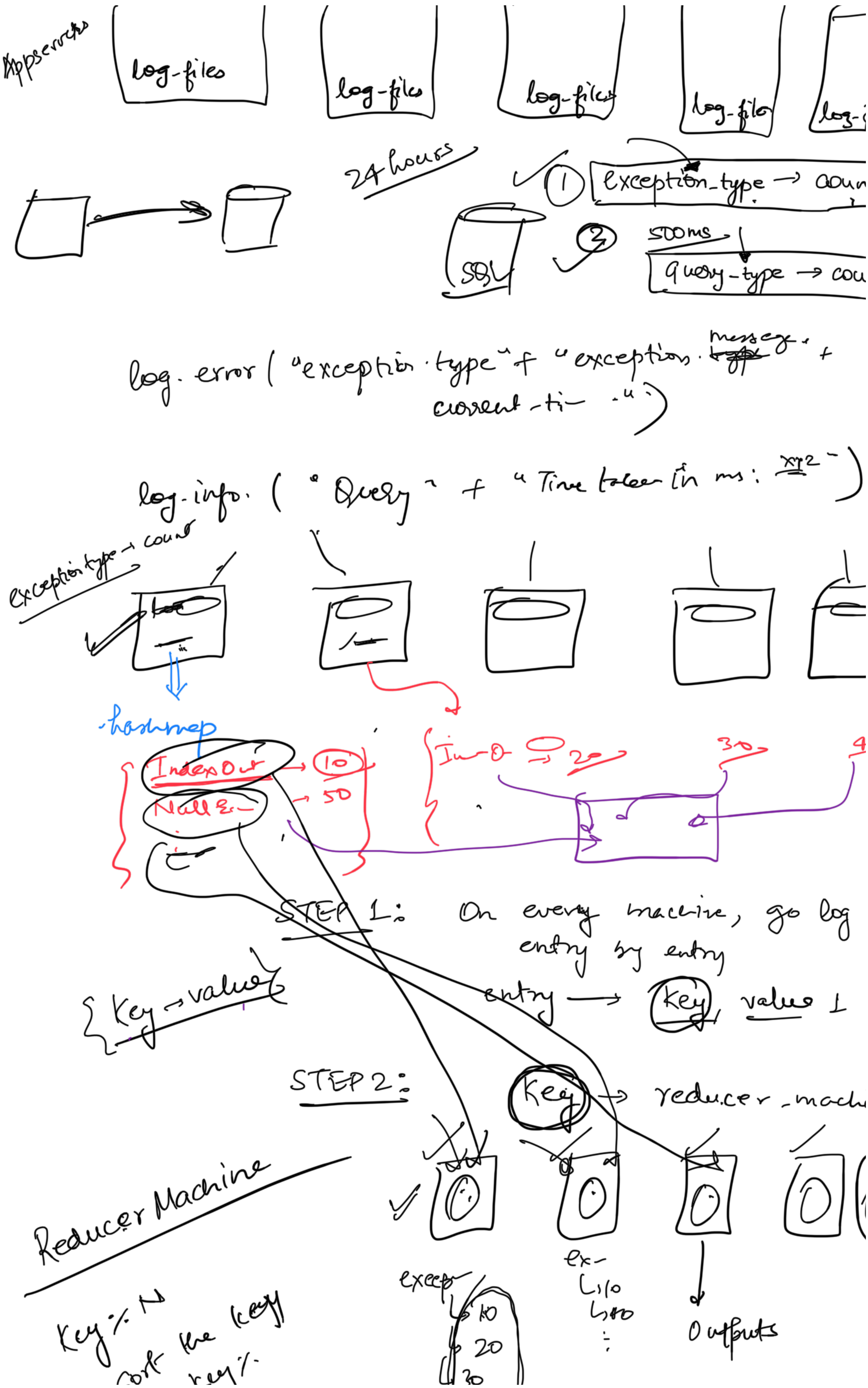
SYSTEM DESIGN



① Zookeeper , ② Elastic Search or lucene , ③ Messaging queues
Kafka
Rab

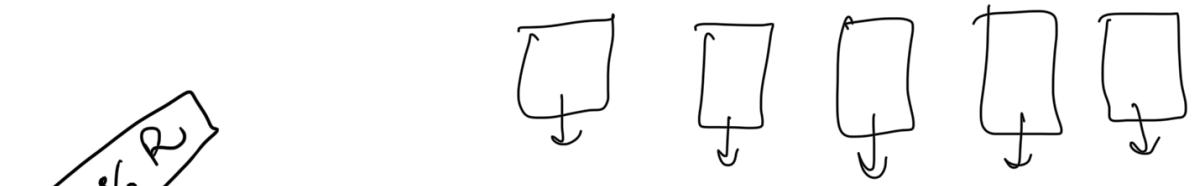
④ Large file storage
S3, HDFS

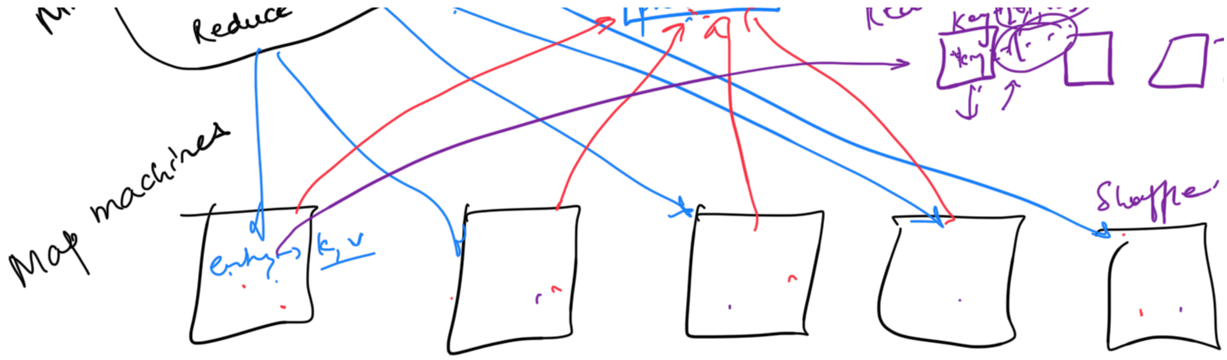


Appservers
log-files
log-files
log-files
log-files
24 hours
1 exception_type → count
SQL
2 500ms
query_type → cou
log.error("exception.type" + "exception.message" + current_time .")
log.info.("Query" + "Time taken in ms: xyz")
exception type → count
hashmap
IndexOut → 10
NullE → 50
STEP 1: On every machine, go log entry by entry
{Key → value}
entry → Key, value 1
STEP 2: Key → reducer_machine
Reducer Machine
Key % N
sort the key
except
10
20
30
ex-
1, 10
1, 20
Outputs

STEP 2: Merge the values for common k

* exception-key → total-count

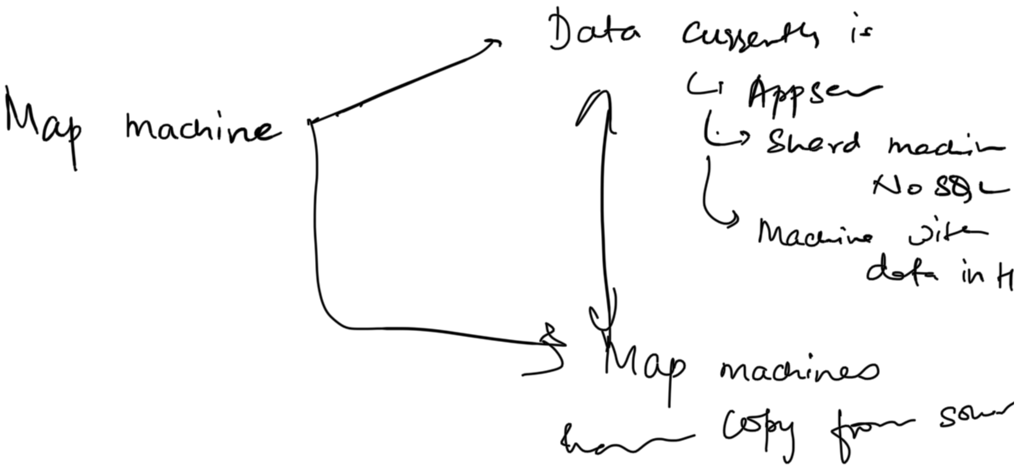


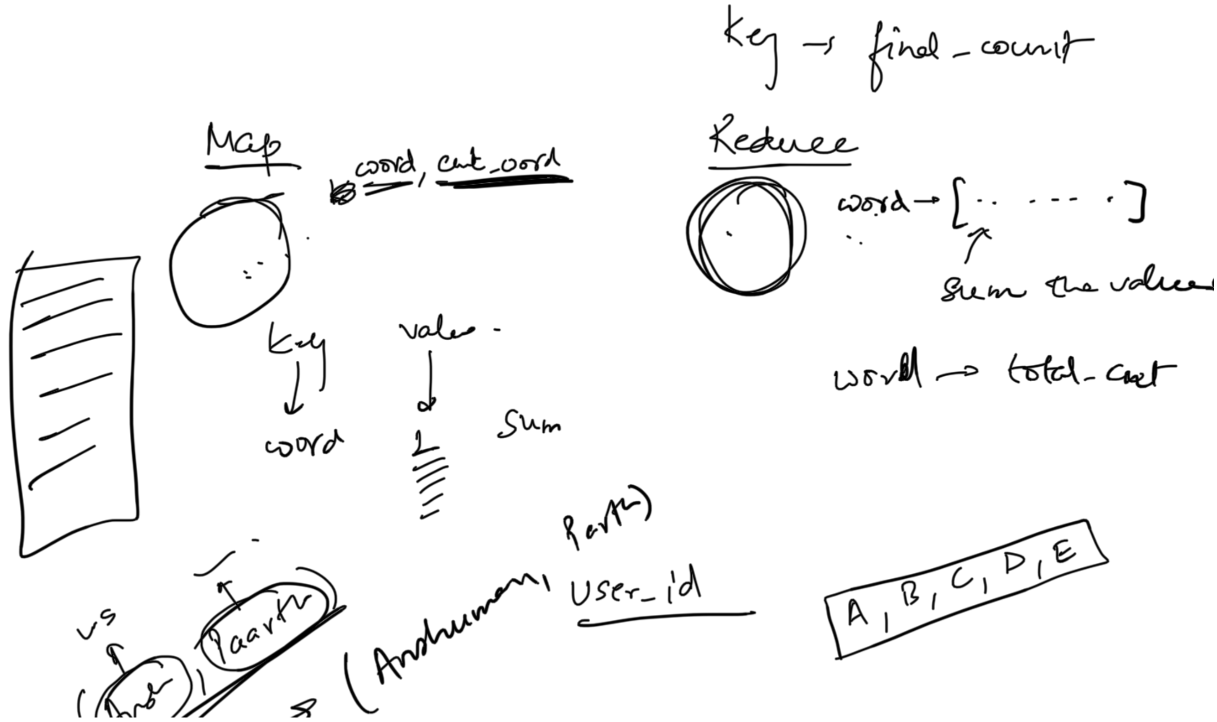


Phase 1 : Map completed

Phase 2 : Shuffle/sort

Phase 3 : Reduce (output is written)





(A,B) → A → B

✓ Map

✓ Reduce

common fr

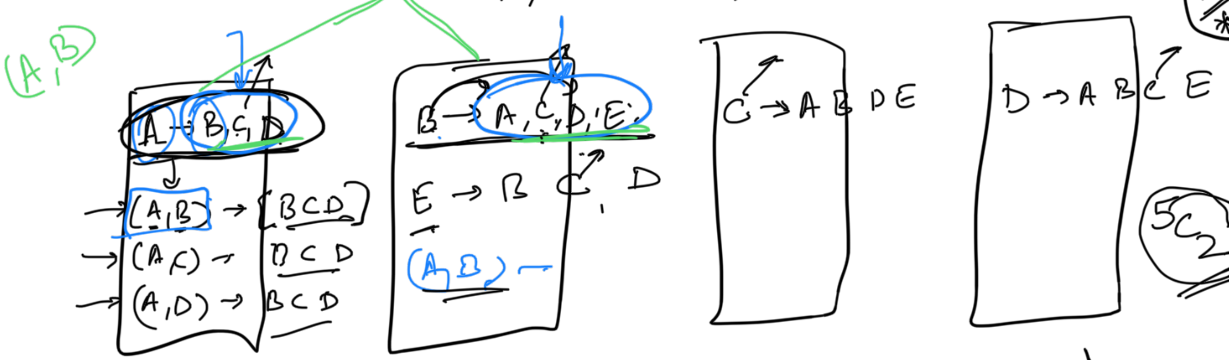
$(X, Y) \to [$

entry

$X \to [X_1, X_2, X_3, X_4]$

$(\min(X, X_1), \max(X, X_1)) \to [X_1\ X_2\ X_3\ X_4]$

$(\min(X, X_2), \max(X, X_2)) \to [X_1\ X_2\ X_3\ X_4$

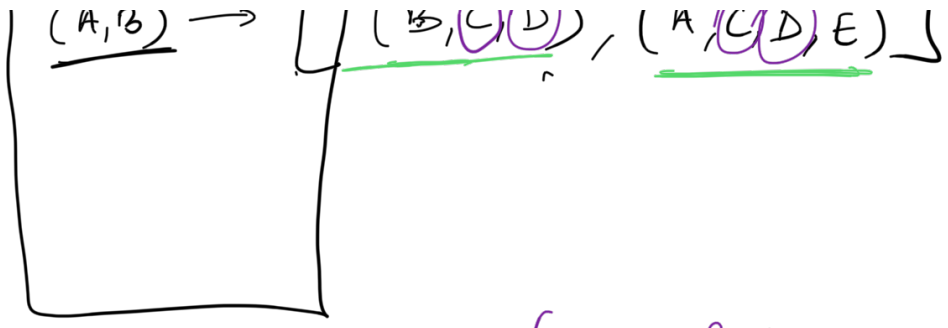


(A,B) → BCD
(A,C) → BCD
A,D → BCD

(A,B) → A C D E
(B,C) → A C D E
(B,C) → A C D E
(B,E) → A C D E
(B,E) → B C D
(C,E) → B C D
(D,E) → B C D

(A,C) → A B D E
(B,C) → A B D E
(C,D) → "
(C,E) → "

(A,B) ↓ (C,D)



Map

A → [B, C, D]

(A,B) → B C D

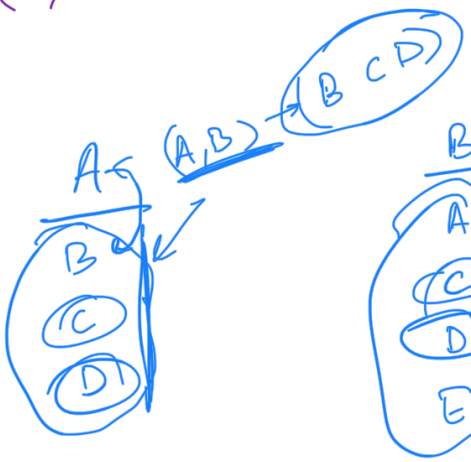
(A,C) → D C ↑

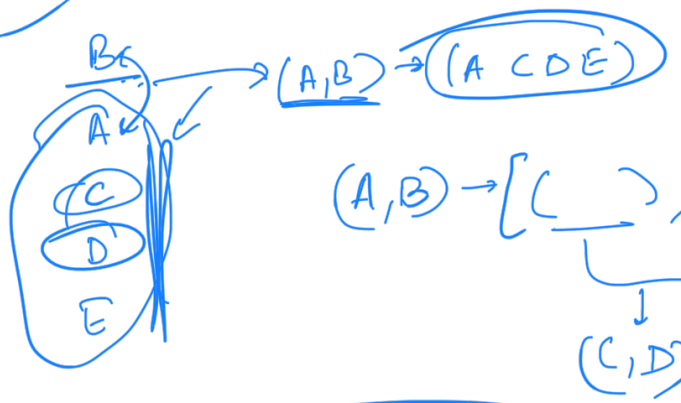
Reduce

(A,B) → [X1, X2]

intersection(X,

(A,B) → (C,D)





(A,B) → (C,D)

Distributed computing → Parallel

# of friend pairs → 2 machines

1 Trillion

[ ] [ ]

1B users

↓ 1000 friends

1 Trillion x

Time taken to talk to 2 ma

Time taken to

Map: 1000 * 1000 ← < 1s.

Show ? similar time ← < 1s.

Reduce: [1 million key → [L1, L2] ← O(1000) ~~O(log(~~

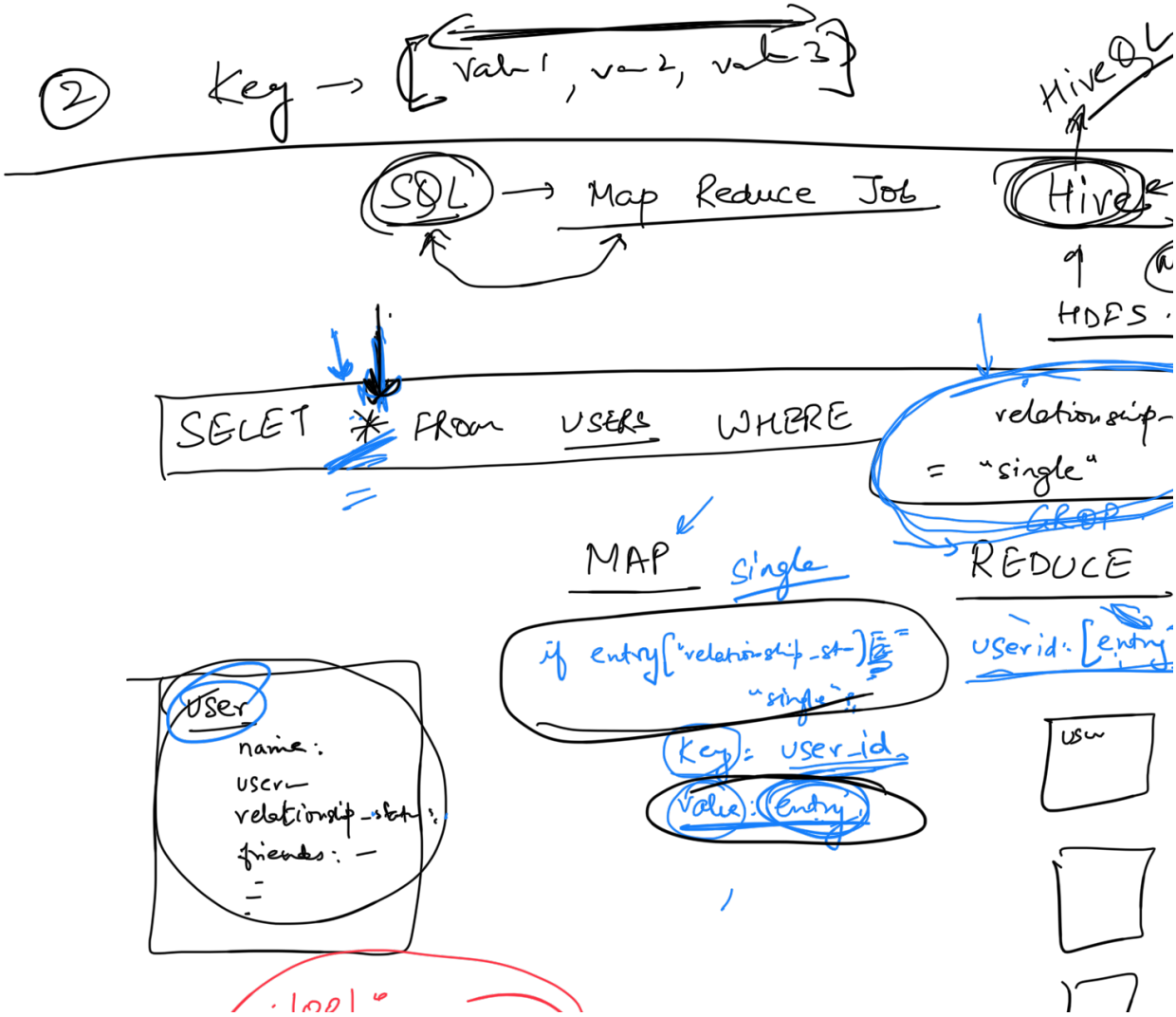
L1: 1000, L2: 1000

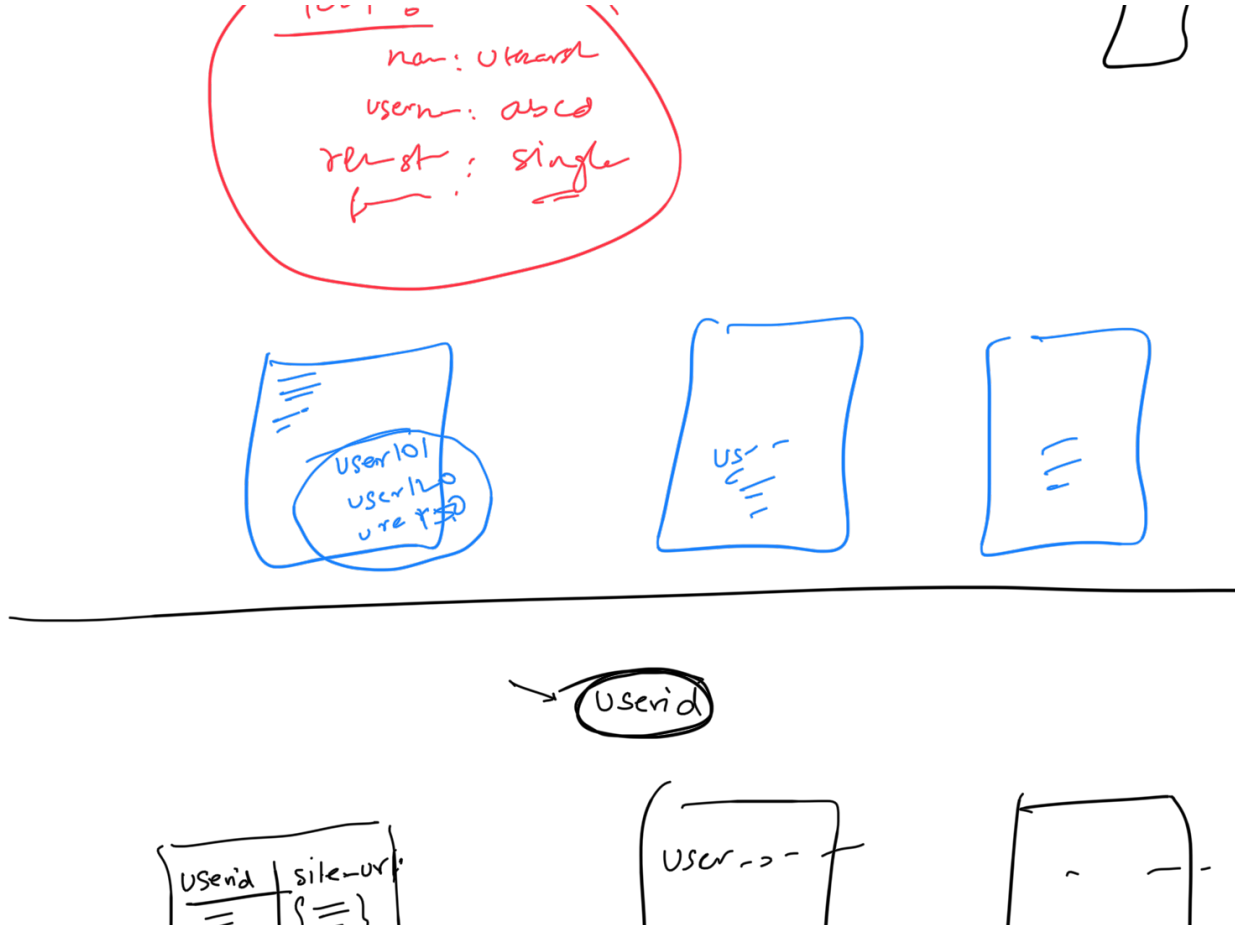
10^9 iterations.

3s - 10s.

① Entry → key, value

② key → [val 1, val 2, val 3]





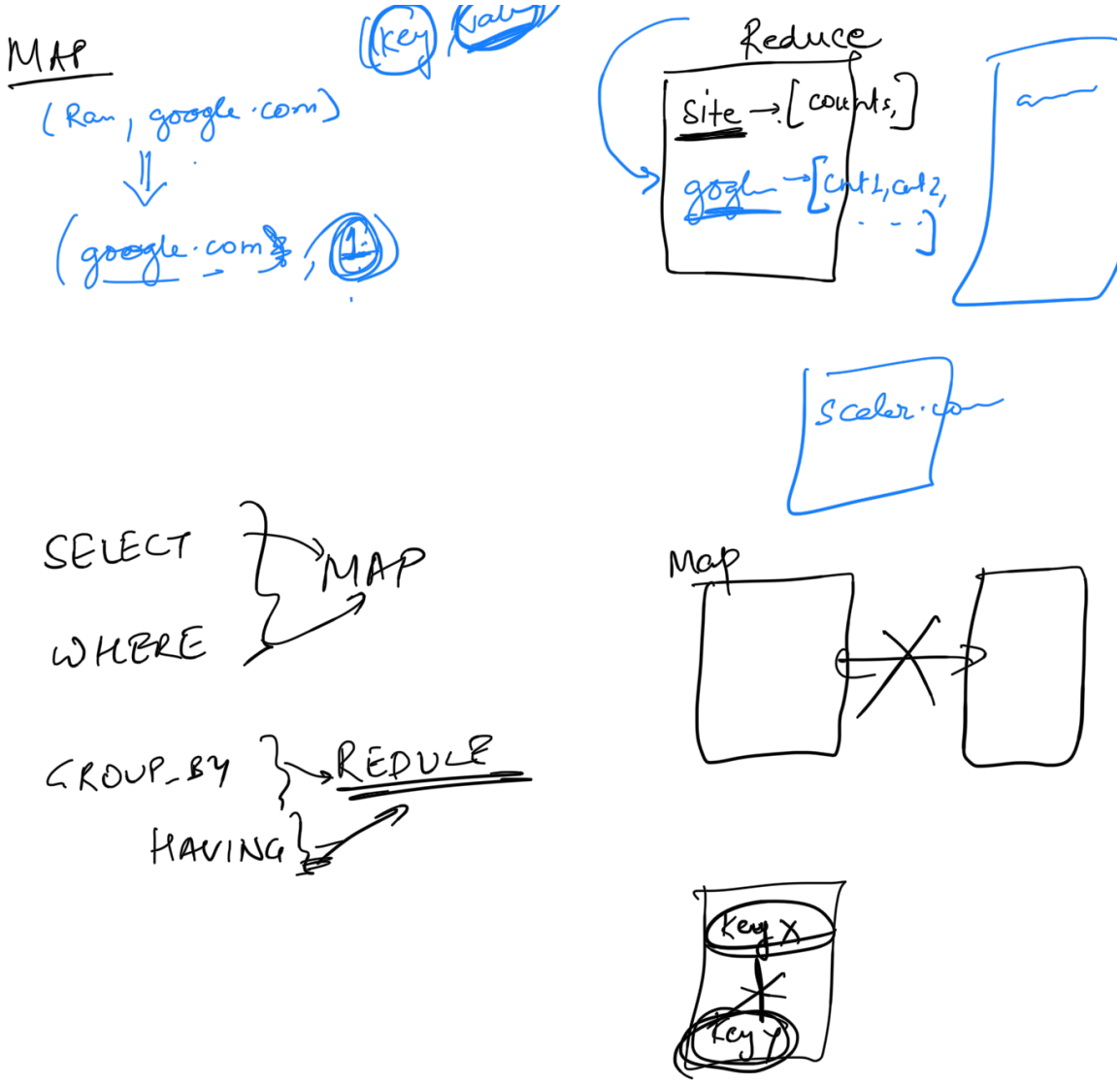
SELECT site_url, count(1) FROM users
GROUP BY site_url

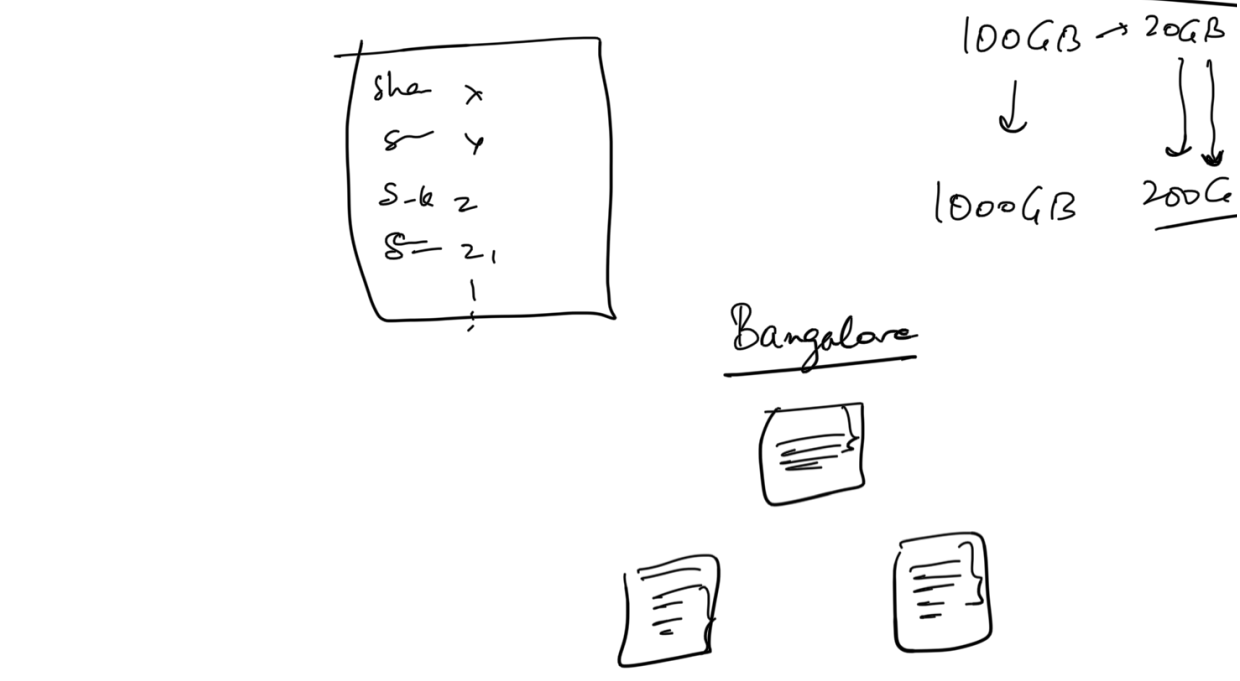
MAP
f(user_id, site_url) → (site_url, 1)

swaggy

REDUCE
site_url
↳ [cnt1, cnt2, cnt3, c

user_id


MAP
(Ram, google.com)
(google.com, 1)
Reduce
site → [counts,]
google → [cnt1, cnt2, ...]
scaler.com
SELECT
WHERE
MAP
GROUP_BY
HAVING
REDUCE
Map
Key X
Key Y


Sha x
S y
S-k z
S z1
100GB → 20GB
1000GB
200G
Bangalore