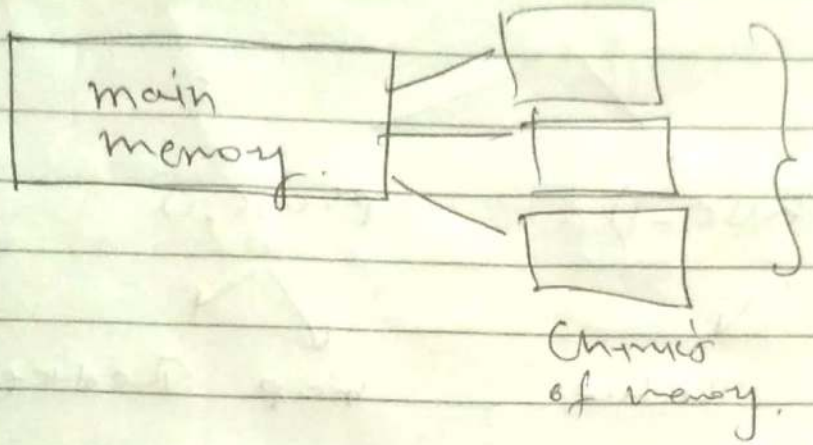
SON Algorithm

* SON Algorithm *



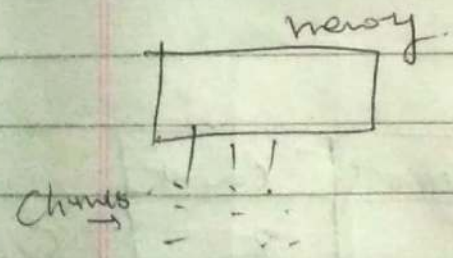
Features :-

P (1) Read.

F (2) Frequent data items.

C (3) Candidate key.

Process :-



Read main memory.

↓

divide into chunks.

↓

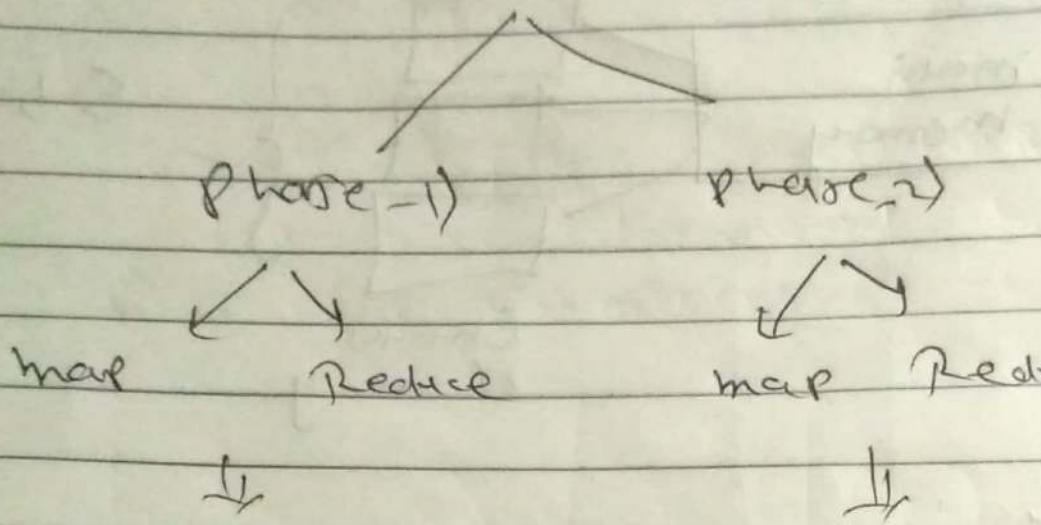
Finding frequent itemsets.

↓

Find candidate key.

Candidate key :→ data items which is common in all subset's called as candidate key.

SON algorithm has 2 map-Reduce phases:



Example:-

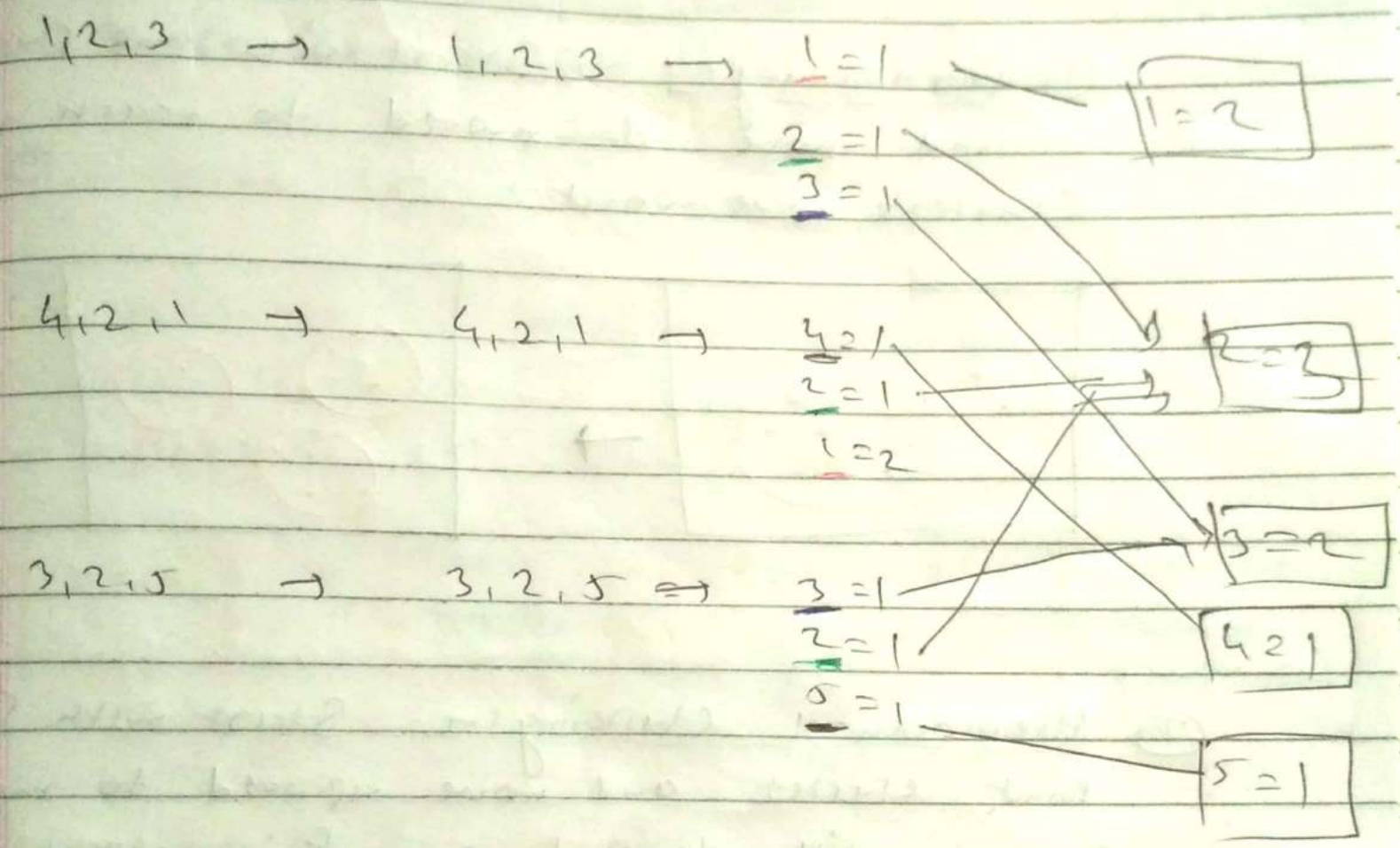
T_id	itemset.
1	1, 2, 3
2	4, 2, 1
3	3, 2, 5

⇒ Frequent Count = candidate reg.

itemset	count	
1	2	Frequ
2	3	freq
3	2	item
4	1	X
5	1	x.

Map Reduce by sort

phase-1)



Here, phase-1) is for finding frequent data items or candidate key &

phase-2) is only for verification of phase-1)

∴ Candidate key - 2=3 as it is frequently occurs and common in more subsets.