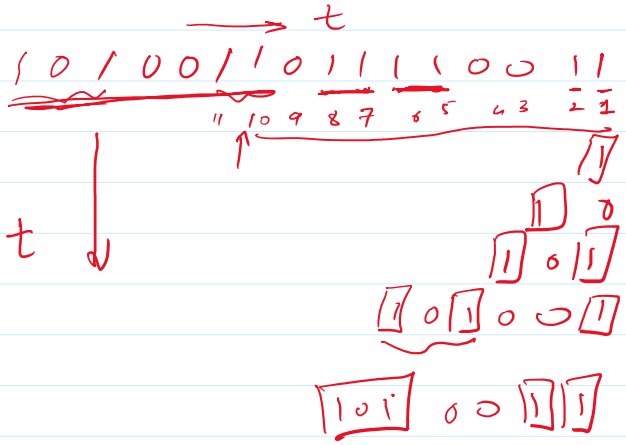
BDA 12th September

11 September 2024 16:39

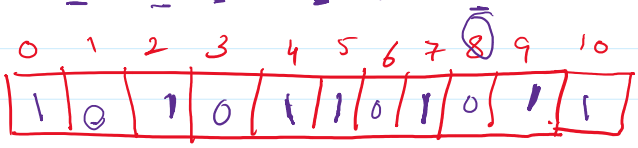
Online session from 11:15am to 12:15pm



How many 1's in last 10 transactions

$8 = (1+1+2+2+4/2)$

Filtering : 2, 3, 1, 6, 4, 5, 9



11 ??

8 ??

$(8*2+3)\%11$

$19\%11$

$= 8$

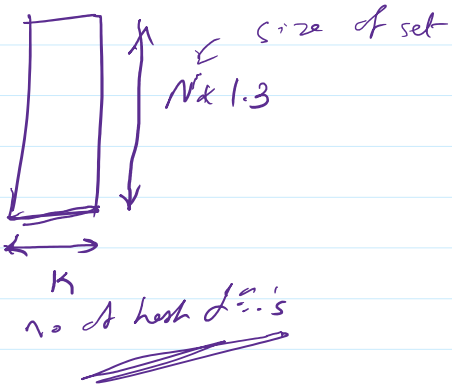
$(Key * 2 + 3) \% 11$

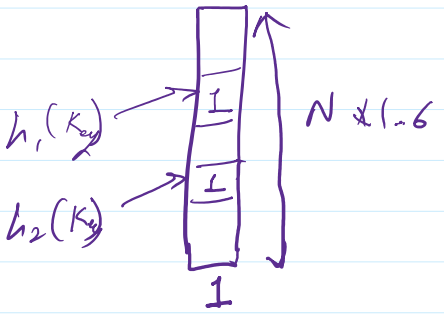
8 ⇒ Not seen earlier

11 ⇒ Possibly seen earlier ⇒

To reduce false +ves, use multiple hash f^n

Memory space needed with each f^n having its own bucket/table



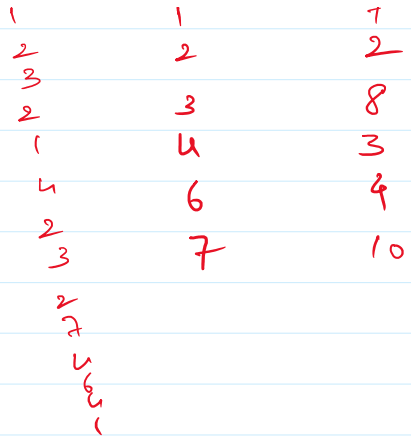


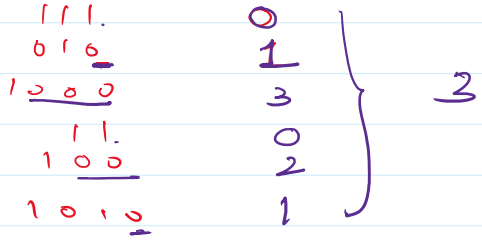
FM Algorithm ⇒ To count no of distinct elements.

1, 2, 3, 2, 1, 4, 2, 3, 2, 7, 4, 6, 4, 1

Hash f^n $(Key * 6 + 1) \bmod 11$

	no	hash f^n val	Binary	r(a)	R max{r(a)}
1	1	7	111	0	
2	2	2	010	1	
3	3	8	1000	3	2





No of distinct elements = $2^3 = 8$

Assignment

1. Parameters & Methodology use to determine duplicate assignments.
2. Frequent Itemset using (Dhaval : Apriori). As volume ↑ what are the drawbacks of Apriori & how are they overcome using
   - → PCY
   - → Multihash