BloomFilter: Core Idea

Lossless: RLE, WAH, decomposed bitmap

Lossy: approximate bit map, bloom filter

Bloom filter & hash + bit lists
setBit(424342)

- $h_1(424342) = 12$
- $h_2(424342) = 7$
- $h_3(424342) = 32$

# of inserts so far, N=1

k hash functions, here k=3

bit list of size M, here M=42
setBit(983764)

- \( h_1(983764) = 30 \)
- \( h_2(983764) = 17 \)
- \( h_3(983764) = 19 \)

- \( \text{# of inserts so far, } N=2 \)  
- \( k \) hash functions, here \( k=3 \)  
- \( \text{bit list of size } M, \text{ here } M=42 \)
setBit(12781), which color is bit 30?

h1(12781) = 0
h2(12781) = 10
h3(12781) = 30

# of inserts so far, N=3
k hash functions, here k=3
bit list of size M, here M=42
setBit(12781), it is yellow!

# of inserts so far, N=3

k hash functions, here k=3

bit list of size M, here M=42
BloomFilter Lookup Semantics

for one hash-function \( h_i(\text{Key}) = 0 \):

key not in bitmap
BloomFilter Lookup Semantics

1) for one hash-function $h_i(\text{Key}) = 0$:
   
   key not in bitmap

2) for all hash-function $h_i(\text{Key}) = 1$:
   
   key maybe in bitmap

   ②
   key is in bitlist
   → key is not in bitlist
   = False positive
lookup(983764)

h1(983764) = 30
h2(983764) = 17
h3(983764) = 19

# of inserts so far, N=3
k hash functions, here k=3
bit list of size M, here M=42
lookup(2956)

h1(2956) = 7
h2(2956) = 16
h3(2956) = 41

# of inserts so far, N=3
k hash functions, here k=3
bit list of size M, here M=42

2956 not in bloom filter
lookup(7431)

h1(7431) = 10    h2(7431) = 12    h3(7431) = 19

1 0 0 0 0 0 0 0 1 0 0 1 0 1 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0

bucket 0    bucket M-1

# of inserts so far, N=3
k hash functions, here k=3
bit list of size M, here M=42

=> 7431 may be in bloom filter
=> False positive