

✓ ¡Felicitaciones! ¡Aprobaste!

Calificación recibida 100 % Para Aprobar 80 % o más

Ir al siguiente elemento

Problem Set #4

Calificación de la entrega más reciente: 100 %

1. Which of the following is *not* a property that you expect a well-designed hash function to have?

1 / 1 punto

- ☒ The hash function should "spread out" every data set (across the buckets/slots of the hash table).
- ☐ The hash function should be easy to store (constant space or close to it).
- ☐ The hash function should be easy to compute (constant time or close to it).
- ☐ The hash function should "spread out" most (i.e., "non-pathological") data sets (across the buckets/slots of the hash table).

✓ Correcto

As discussed in lecture, unfortunately, there is no such hash function.

2. Suppose we use a hash function h to hash n distinct keys into an array T of length m . Assuming simple uniform hashing --- that is, with each key mapped independently and uniformly to a random bucket --- what is the expected number of keys that get mapped to the first bucket? More precisely, what is the expected cardinality of the set $\{k : h(k) = 1\}$.

1 / 1 punto

- ☐ $1/n$
- ☐ $m/(2n)$
- ☒ n/m
- ☐ $n/(2m)$
- ☐ $1/m$
- ☐ m/n

✓ Correcto

Use linearity of expectation, with one indicator variable for each key. The probability that one key hashes to the first bucket is $1/m$, and by linearity of expectation the total expected number of keys that hash to the first bucket is just n/m .

3. Suppose we use a hash function h to hash n distinct keys into an array T of length m . Say that two distinct keys x, y *collide* under h if $h(x) = h(y)$. Assuming simple uniform hashing --- that is, with each key mapped independently and uniformly to a random bucket --- what is the probability that a given pair x, y of distinct keys collide?

1 / 1 punto

- ☐ $1/n$
- ☒ $1/m$
- ☐ $1/m^2$
- ☐ $1/n^2$
- ☐ $1/(m-1)$

✓ Correcto

For each of x and y , there are m choices for which bucket they map to. This means m^2 choices for the pair of buckets $(h(x), h(y))$. Under uniform hashing, each of these choices is equally likely (i.e., occurs with probability $1/m^2$). Since exactly m of these pairs yield a collision, the probability of a collision is $m/m^2 = 1/m$.

4. Suppose we use a hash function h to hash n distinct keys into an array T of length m . Assuming simple uniform hashing --- that is, with each key mapped independently and uniformly to a random bucket --- what is the expected number of pairs of distinct keys that collide? (As above, distinct keys x, y are said to collide if $h(x) = h(y)$.)

1 / 1 punto

- ☒ $n(n-1)/2m$
- ☐ n/m^2
- ☐ n/m
- ☐ $n(n-1)/m$
- ☐ n^2/m

✓ Correcto

There are n choose 2 (i.e., $n(n-1)/2$) pairs of distinct keys. By the previous problem, each pair has a $1/m$ chance of colliding. The answer now follows from the linearity of expectation.

5. To interpret our heuristic analysis of bloom filters in lecture, we considered the case where we were willing to use 8 bits of space per object in the bloom filter. Suppose we were willing to use twice as much space (16 bits per object). What can you say about the corresponding false positive rate, according to our heuristic analysis (assuming that the number k of hash tables is set optimally)? [Choose the strongest true statement.]

1 / 1 punto

- ☐ Less than 1%
- ☒ Less than .1%
- ☐ Less than .01%
- ☐ Less than .001%

✓ Correcto