Ques10 (/)

Q (/local/search/page/)   Ask (/p/new/post/)   Login (/site/login/?next=/p/20554/explain-bloom-filter-with-the-help-of-an-example/)

Explain Bloom Filter with the help of an example.

written 5.3 years ago by   •   modified 2.9 years ago

teamques10 (/u/1/teamques... ✔ ★ 48k

big data analytics (/t/big data analytics/)

0

14k views

ADD COMMENT   FOLLOW   SHARE   EDIT (/p/edit/20554/)

1 Answer

written 5.3 years ago by

teamques10 (/u/1/teamques... ✔ ★ 48k

A Bloom filter consists of:

1. An array of n bits, initially all 0's.
2. A collection of hash functions h1, h2, . . . , hk. Each hash function maps "key" values to n buckets, corresponding to the n bits of the bit-array.
3. A set S of m key values.

The purpose of the Bloom filter is to allow through all stream elements whose keys are in S, while rejecting most of the stream elements whose keys are not in S.

The model to use is throwing darts at targets. Suppose we have x targets and y darts. Any dart is equally likely to hit any target. After throwing the darts, how many targets can we expect to be hit at least once?
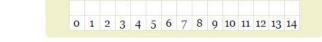
- The probability that a given dart will not hit a given target is $(x − 1)/x$
- The probability that none of the y darts will hit a given target is $((x−1)/x)^y$
- We can write this expression as $(1 − 1 x)^x(y x)$.
- Using the approximation $(1−\varrho)1/\varrho = 1/e$ for small E we conclude that the probability that none of the y darts hit a given target is $e−y/x$.

The base data structure of a Bloom filter is a Bit Vector. Here's a small one we'll use to demonstrate:

Each empty cell in that table represents a bit, and the number below it its index. To add an element to the Bloom filter, we simply hash it a few times and set the bits in the bit vector at the index of those hashes to 1.

It's easier to see what that means than explain it, so enter some strings and see how the bit vector changes. Fnv and Murmur are two simple hash functions:

3

790 views

When you add a string, you can see that the bits at the index given by the hashes are set to 1. I've used the color green to show the newly added ones, but any colored cell is simply.

To test for membership, you simply hash the string with the same hash functions, then see if those values are set in the bit vector. If they aren't, you know that the element isn't in the set. If they are, you only know that it might be, because another element or some combination of other elements could have set the same bits. Again, let's demonstrate:

Test an element for membership:

fnv:
murmur:

Is the element in the set? no

Probability of a false positive: 0%

And that's the basics of a bloom filter.

ADD COMMENT　　SHARE　EDIT (/p/edit/20555/)

Please log in (/site/login/?next=/p/20554/explain-bloom-filter-with-the-help-of-an-example/) to add an answer.

Start with Confluenc

Get started with Confluence for fr

COMMUNITY
Users (/user/list/)
Levels (/info/levels/)
Badges (/b/list/)

CONTENT
All posts (/t/latest)
Tags (/t/)
Dashboard (/dashboard/)

COMPANY
About (/info/about/)
Team (/info/team/)
Privacy (/info/privacy/)

Submit question paper solutions and earn money › (/info/solutions/)