

## blooming filtering چیست؟

قبل از اینکه به بررسی چستی بلوم فیلترینگ پردازم بهتره بگم چه موقع کاربرد داره، به طور ساده وقتی بخوایم در مجموعه ای جستجو کنیم یک زمان زیادی رو میطلبه اما اگر نتیجه نهایی جستجو نبودن اون عضو در مجموعه باشه ما باید  $n$  واحد زمانی صبر کنیم تا بفهمیم که این اصلا به صرفه نیست. با کمک بلوم فیلترینگ ما میتونیم یک چیز رو مطمئن شویم آیا این عضو در مجموعه هست یا نه. اگر خروجی مبتنی بر نبودن بود یعنی قطعا نیستش.

البته در اینجا مفهومی به اسم خطای منفی و مثبت هم وجود دارد. به عنوان مثال ما میدونیم که  $X$  وارد مجموعه ما شده است و اگر در آن لحظه در مجموعه نباشد خطای مثبت رخ داده است و وقتی  $Y$  وارد مجموعه ما نشده است و ما میگوییم شده است به معنای خطای منفی است.

پایه دیگر بلوم فیلترینگ هش کردن میباشد با کمک این مکانیزم داده ها رو هش میکنیم و این روند  $k$  بار تکرار میشود. حالا به طور ساده بلوم فیلترینگ یک عدد لیست به طول  $m$  است که حاصل از نتیجه هش شدن عضو هاست.

به عنوان مثال عضو های ما به این شکل هستند  $\{a, b, c\}$  و جدول حاصل به شکل زیر میباشد

0	1	0	0	1	1
---	---	---	---	---	---

حالا اگر بخوایم بررسی کنیم عضو  $d$  در جدول است باید چیکار کنیم؟ کافیه ابتدا به اندازه  $k$  تابع هش بدیم و نتیجه اون رو بگیریم. که نتیجه زیر رو میده

1	0	1	0	0	0
---	---	---	---	---	---

که با توجه یک مقایسه دیداری میشه متوجه شد که عضو  $d$  در مجموعه ما نبوده. حالا همان طور که درباره خطای منفی و مثبت گفتیم در اینجا به این معناست که نبود یک عضو در مجموع قطعیت ولی بودن آن نه. یکم بیشتر پیش بریم و دربارش صحبت کنیم:

حالا فرض کنید مجموعه ما بروز رسانی میشه و عضو  $b$  از اون حذف میشه اما باز هم لیست  $m$  تایی ما ثابت هستش. دوباره میریم و  $b$  رو بررسی میکنیم که آیا در مجموعه هستش در قدم اول به  $k$  تابع هش میدیم و نتیجه رو میگیریم:

0	1	0	0	0	1
---	---	---	---	---	---

با مقایسه دیداری 1 ها میشه متوجه شد که  $b$  احتمالا در مجموعه ما هست و وقتی جستجو انجام میشه و در طول مجموعه میگرده باعث میشه که نتیجه بده که نیست. حالا در اینجا خطای مثبت رخ داده یعنی در بخش لیست بلوم فیلترینگ گفتیم احتمالا هست ولی وقتی بررسی کردیم متوجه شدیم که  $b$  از عناصر ما حذف شده. در اینجا یک نکته هستش لیست بلو فیلترینگ فقط موقع اضافه کردن عنصر جدید به لیست بروز رسانی میشه نه موقع حذف اون به خاطر اینکه یک عضو با عضو دیگر ممکنه جاهای مشترکی از لیست رو اشغال کنند.

## رابطه کاساندرای و بلوم فیلتر:

کاساندرای با کمک بلوم فیلترینگ میتونه توی حافظه رم (سریع اما کم حجم) این لیست رو ذخیره کنه و مسئله وجود یا نبودن SSTables ها رو تشخیص بده. دقیقا موارد بلوم فیلترینگ در کاساندرای شبیه موارد بالایی که در بخش بلوم فیلترینگ توضیح دادم هست. در این با هم مسئله اینکه فقط موقع اضافه شدن یک عضو به مجموعه اون لیست بروز رسانی میشه هم وجود داره که مسئله خطای مثبت و منفی رو دوباره میاره.

منابع:

[http://cassandra.apache.org/doc/latest/operating/bloom\\_filters.html](http://cassandra.apache.org/doc/latest/operating/bloom_filters.html)

<https://stackoverflow.com/questions/39327427/what-is-role-of-bloom-filter-in-cassandra>