

设计索引的时候，我们一般要考虑哪些因素呢？（中）

今天我们继续来说一下，在设计索引的时候要考虑哪些因素。之前已经说了，你设计的索引最好是让你的各个where、order by和group by后面跟的字段都是联合索引的最左侧开始的部分字段，这样他们都能用上索引。

但是在设计索引的时候还得考虑其他的一些问题，首先一个就是字段基数问题，举个例子，有一个字段他一共在10万行数据里有10万个值对吧？结果呢？这个10万值，要不然就是0，要不然就是1，那么他的基数就是2，为什么？因为这个字段的值就俩选择，0和1。

假设你要是针对上面说的这种字段建立索引的话，那就还不如全表扫描了，因为你的索引树里就仅仅包含0和1两种值，根本没法进行快速的二分查找，也根本就没有太大的意义了，所以这种时候，选用这种基数很低的字段放索引里意义就不大了。

一般建立索引，尽量使用那些基数比较大的字段，就是值比较多的字段，那么才能发挥出B+树快速二分查找的优势来。

其次的话，你尽量是对那些**字段的类型比较小的列来设计索引**，比如说什么tinyint之类的，因为他的字段类型比较小，说明这个字段自己本身的价值占用磁盘空间小，此时你在搜索的时候性能也会比较好一点。

不过当然了，这个所谓的字段类型小一点的列，也不是绝对的，很多时候你就是要针对varchar(255)这种字段建立索引，哪怕多占用一些磁盘空间，那你也得去设计这样的索引，比较关键的其实还是尽量别把基数太低的字段包含在索引里，因为意义不是太大。

那当然了，万一要是你真的有那种varchar(255)的字段，可能里面的值太大了，你觉得都放索引树里太占据磁盘空间了，此时你仔细考虑了一下，发现完全可以换一种策略，也就是仅仅针对这个varchar(255)字段的前20个字符建立索引，就是说，对这个字段里的每个值的前20个字符放在索引树里而已。

此时你建立出来的索引其实类似于KEY my_index(name(20),age,course)，就这样的一个形式，假设name是varchar(255)类型的，但是在索引树里你对name的值仅仅提取前20个字符而已。

此时你在where条件里搜索的时候，如果是根据name字段来搜索，那么此时就会先到索引树里根据name字段的前20个字符去搜索，定位到之后前20个字符的前缀匹配的部分数据之后，再回到聚簇索引提取出来完整的name字段值进行比对就可以了。

但是假如你要是order by name, 那么此时你的name因为在索引树里仅仅包含了前20个字符, 所以这个排序是没法用上索引了! group by也是同理的。所以这里大家要对前缀索引有一个了解。

好了, 同学们, 今天给大家重点讲了**索引字段的基数和前缀索引**的知识, 大家就记住两点, 对于那种字段基数很低的列尽量别包含到索引里去, 没多大用;

另外就是对于那种比较长的字符串类型的列, 可以设计前缀索引, 仅仅包含部分字符到索引树里去, where查询还是可以用的, 但是order by和group by就用不上了。

End

内部资源仅限自己学习
www.pp1sunny.top