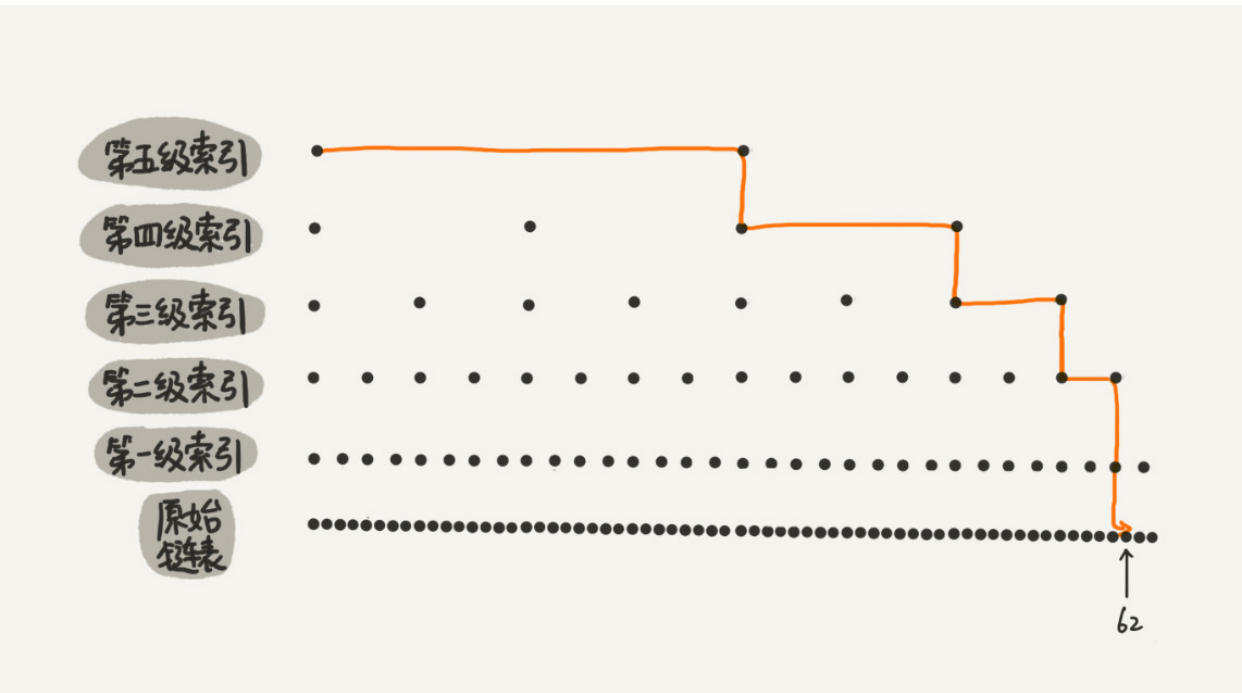
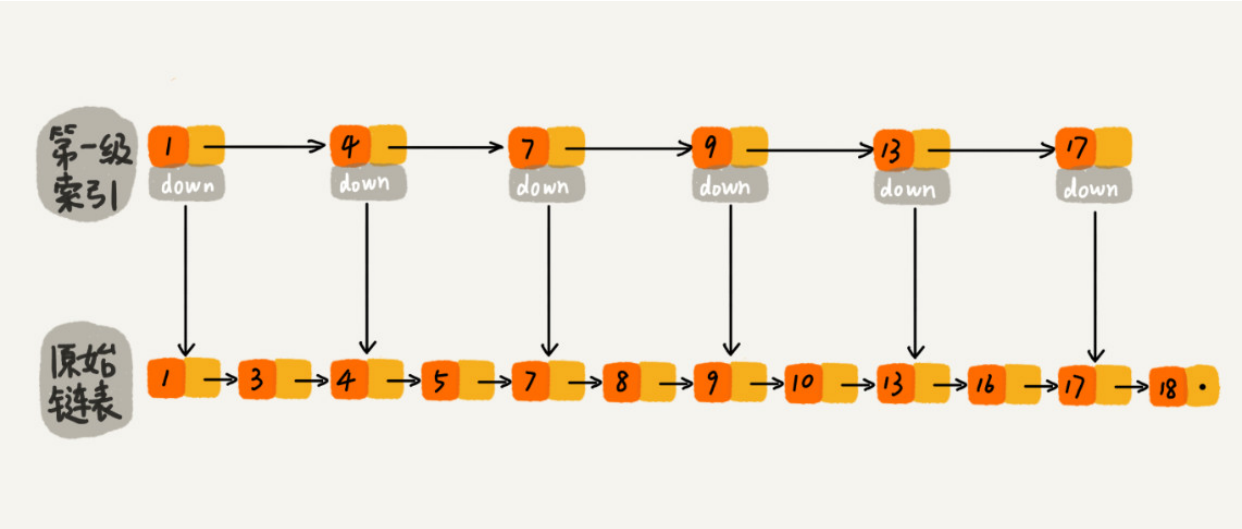


一、什么是跳表？

为一个值有序的链表建立多级索引，比如每2个节点提取一个节点到上一级，我们把抽出来的那一级叫做索引或索引层。如下图所示，其中down表示down指针，指向下一级节点。以此类推，对于节点数为n的链表，大约可以建立 $\log_2 n - 1$ 级索引。像这种为链表建立多级索引的数据结构就称为跳表。



二、跳表的时间复杂度？

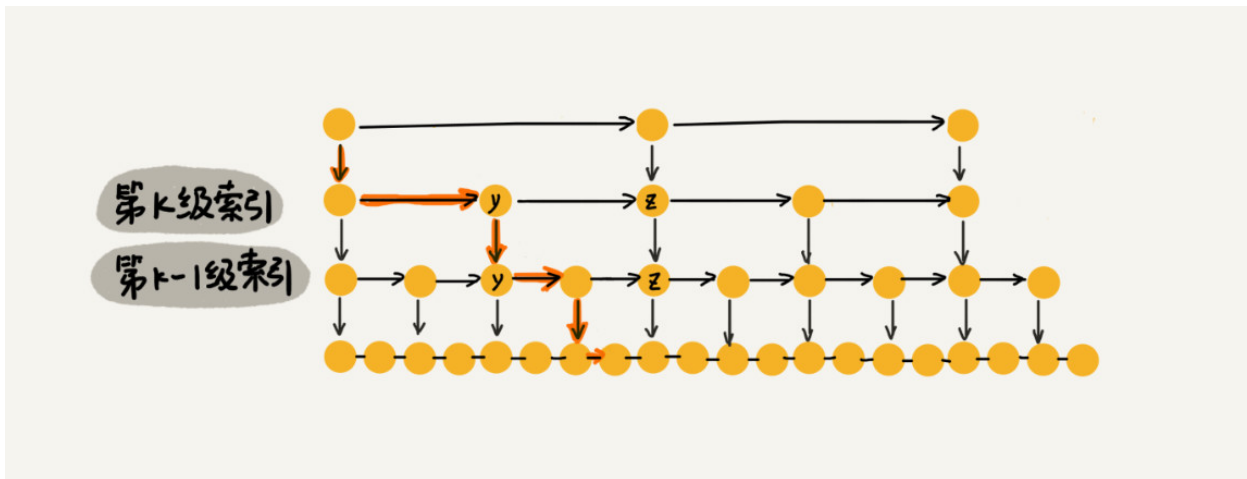
1. 计算跳表的高度

如果链表有n个节点，每2个节点抽取抽出一个节点作为上一级索引的节点，那第1级索引的节点个数大约是 $n/2$ ，第2级索引的节点个数大约是 $n/4$ ，依次类推，第

k级索引的节点个数就是 $n/(2^k)$ 。假设索引有h级别，最高级的索引有2个节点，则有 $n/(2^h)=2$ ，得出 $h=\log_2 n-1$ ，包含原始链表这一层，整个跳表的高度就是 $\log_2 n$ 。

2. 计算跳表的时间复杂度

假设我们在跳表中查询某个数据的时候，如果每一层都遍历m个节点，那在跳表中查询一个数据的时间复杂度就是 $O(m \cdot \log n)$ 。那这个m是多少呢？如下图所示，假设我们要查找的数据是x，在第k级索引中，我们遍历到y节点之后，发现x大于y，小于后面的节点z，所以我们通过y的down指针，从第k级下降到第k-1级索引。在第k-1级索引中，y和z之间只有3个节点（包含y和z），所以，我们在k-1级索引中最多只需要遍历3个节点，以此类推，每一级索引都最多只需要遍历3个节点。所以 $m=3$ 。因此在跳表中查询某个数据的时间复杂度就是 $O(\log n)$ 。



三、跳表的空间复杂度及如何优化？

1. 计算索引的节点总数

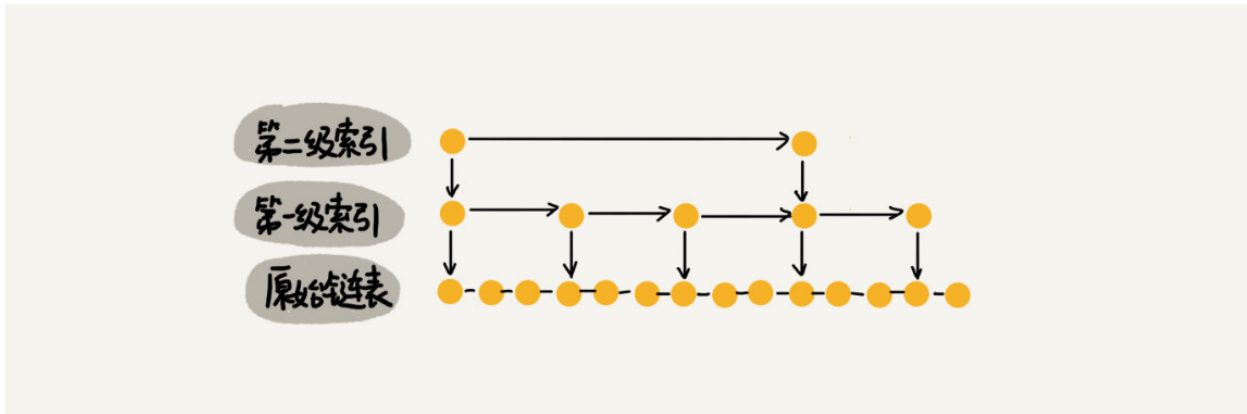
如果链表有n个节点，每2个节点抽取出一个节点作为上一级索引的节点，那每一级索引的节点数分别为： $n/2$ ， $n/4$ ， $n/8$ ， \dots ，8，4，2，等比数列求和 $n-1$ ，所以跳表的空间复杂度为 $O(n)$ 。

原始链表大小为n，每2个结点抽1个，每层索引的结点数：

$$\frac{n}{2}, \frac{n}{4}, \frac{n}{8}, \dots, 8, 4, 2$$

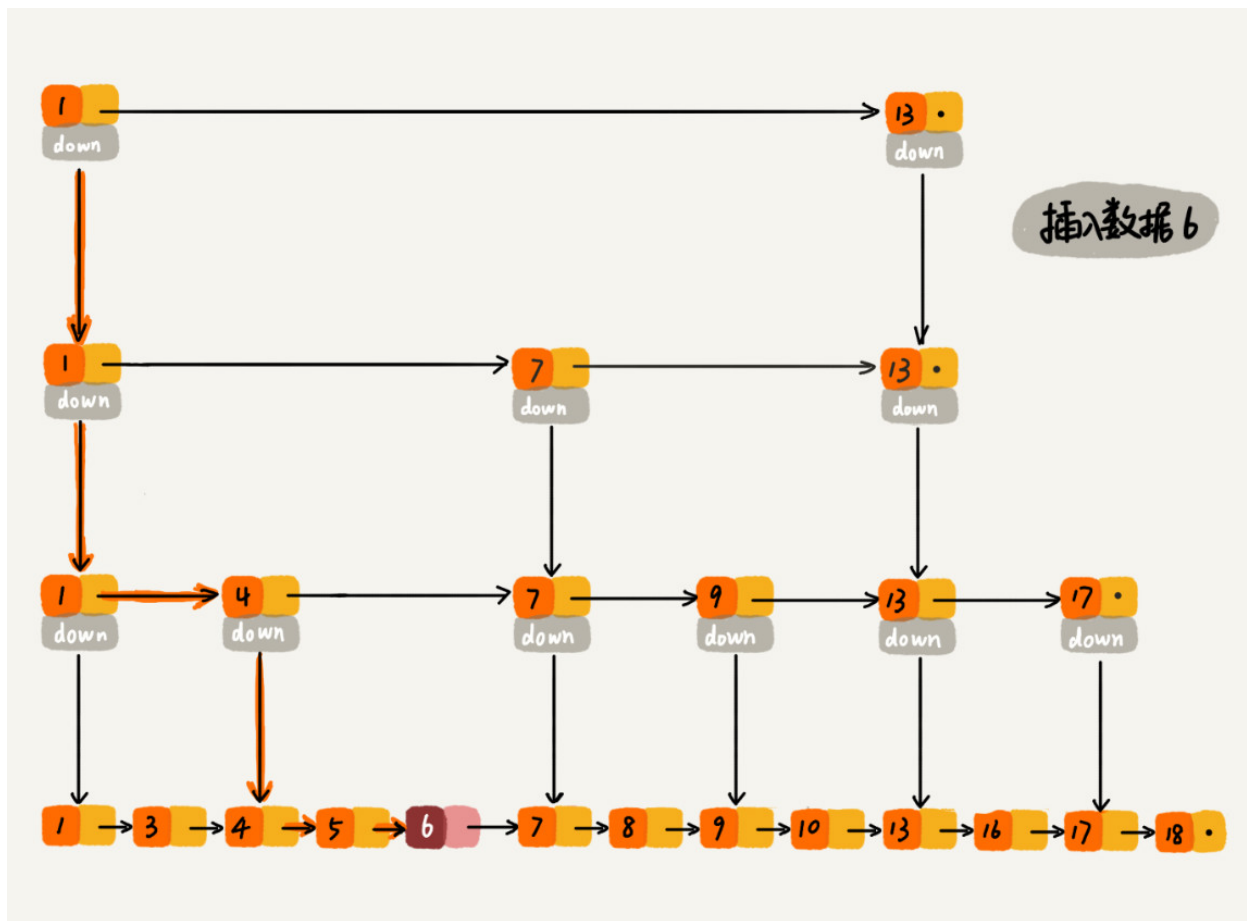
2.如何优化时间复杂度

如果链表有 n 个节点，每3或5个节点抽取出一个节点作为上一级索引的节点，那每一级索引的节点数分别为（以3为例）： $n/3$ ， $n/9$ ， $n/27$ ， \dots ，27，9，3，1，等比数列求和 $n/2$ ，所以跳表的空间复杂度为 $O(n)$ ，和每2个节点抽取一次相比，时间复杂度要低不少呢。



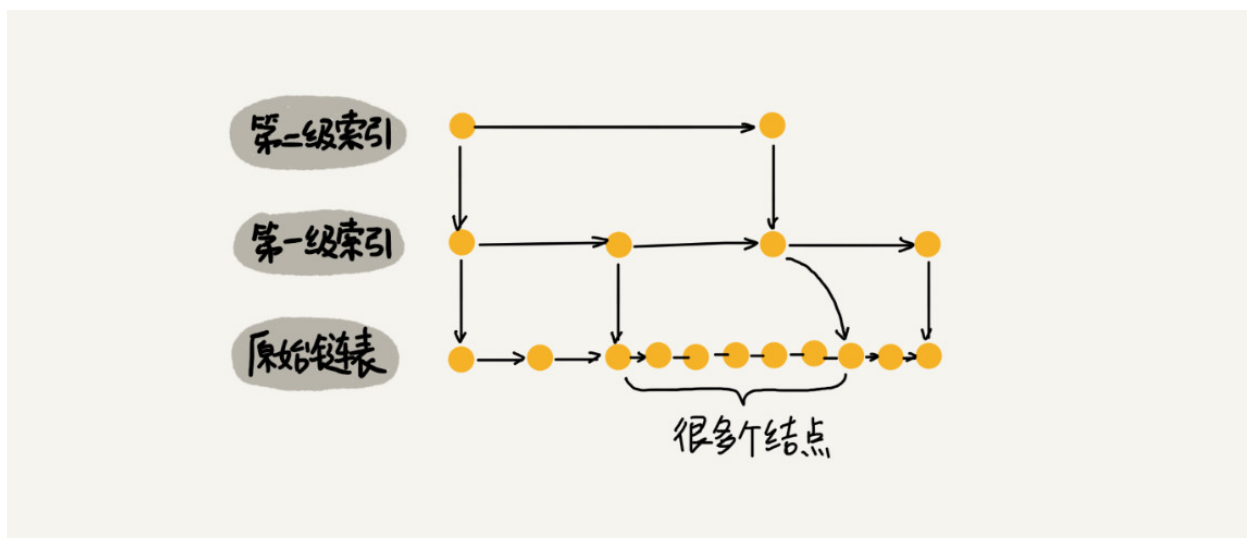
四、高效的动态插入和删除？

跳表本质上就是链表，所以仅插作，插入和删除操作时间复杂度就为 $O(1)$ ，但在实际情况中，要插入或删除某个节点，需要先查找到指定位置，而这个查找操作比较费时，但在跳表中这个查找操作的时间复杂度是 $O(\log n)$ ，所以，跳表的插入和删除操作的时间复杂度也是 $O(\log n)$ 。



五、跳表索引动态更新？

当往跳表中插入数据的时候，可以选择同时将这个数据插入到部分索引层中，那么如何选择这个索引层呢？可以通过随机函数来决定将这个节点插入到哪几级索引中，比如随机函数生成了值K，那就可以把这个节点添加到第1级到第K级索引中。



极客时间文档：<https://time.geekbang.org/column/article/42896>