

多叉树

一、二叉树存在的问题

二叉树需要加载到内存的，如果二叉树的节点少，没有什么问题，但是如果二叉树的节点很多(比如 1 亿)，就存在如下问题：

问题 1：在构建二叉树时，需要多次进行 i/o 操作(海量数据存在数据库或文件中)，节点海量，构建二叉树时，速度有影响

问题 2：节点海量，也会造成二叉树的高度很大，会降低操作速度。

二、计算机中数据的存储原理

页是计算机管理存储的逻辑块，硬件及操作系统往往将主存和磁盘存储区分割为连续的大小相等的块。每个块都称为一页(在许多操作系统中，页大小通常为 4K)，主存和磁盘以页为单位交换数据。

文件系统及数据库系统的设计利用了磁盘预读原理，将一个节点的大小设计为一个页，这样每个节点只需要一次 IO 就可以完全载入。

众所周知，IO 操作的效率很低，那么，当在大量数据存储中，查询时我们不能一下子将所有数据加载到内存中，只能逐一加载磁盘页，每个磁盘页对应树的节点。造成大量磁盘 IO 操作（最坏情况下为树的高度）。平衡二叉树由于树深度过大而造成磁盘 IO 读写过于频繁，进而导致效率低下。

所以，我们为了减少磁盘 IO 的次数，就必须降低树的深度，将“瘦高”的树变得“矮胖”。一个基本的想法就是：

- **每个节点存储多个元素**
- **摒弃二叉树结构，采用多叉树**