数据结构及其应用

Vxlimo

福建师范大学附属中学

2023 年 7 月 10 日

写在前面

写在前面

- 请认真听课并尝试理解，如果该部分你已经会了或者掉线了不想听，请不要打扰到其他同学。

写在前面

- 请认真听课并尝试理解，如果该部分你已经会了或者掉线了不想听，请不要打扰到其他同学。
- 在讲课过程中如有任何问题请立即提出，但是建议提问之前和周围同学探讨一下，尽量不要出现无意义问题。

目录

1. 一些基本的数据结构
   - 树状数组和线段树
   - 并查集
   - 堆
2. 一些应用
   - 重链剖分
   - 数据结构的可持久化
3. 题目选讲
   - [NOIP2017 提高组] 列队
   - [十二省联考 2019] 春节十二响
   - [NOI2021] 轻重边
   - [JLOI2015] 城池攻占
4. 选讲内容
   - 基于重链剖分的毛毛虫剖分

简单介绍

简单介绍

- 这两个简单数据结构我想应该绝大部分同学都已经完全掌握了，所以不再赘述。

简单介绍

- 这两个简单数据结构我想应该绝大部分同学都已经完全掌握了，所以不再赘述。
- 本质来说，这两种数据结构就是在做一件事情：维护一个序列的某些信息，同时支持一定程度的修改。

简单介绍

- 这两个简单数据结构我想应该绝大部分同学都已经完全掌握了，所以不再赘述。
- 本质来说，这两种数据结构就是在做一件事情：维护一个序列的某些信息，同时支持一定程度的修改。
- 利用 $\log$ 来达成时间和空间上的平衡，复杂度均为 $\mathcal{O}(n \log n)$。

简单介绍

- 这两个简单数据结构我想应该绝大部分同学都已经完全掌握了，所以不再赘述。
- 本质来说，这两种数据结构就是在做一件事情：维护一个序列的某些信息，同时支持一定程度的修改。
- 利用 $\log$ 来达成时间和空间上的平衡，复杂度均为 $\mathcal{O}(n\log n)$。
- 显然，树状数组能做的，线段树也一定能做。但是树状数组更好写，并且常数也更小。

简单介绍

- 这两个简单数据结构我想应该绝大部分同学都已经完全掌握了，所以不再赘述。
- 本质来说，这两种数据结构就是在做一件事情：维护一个序列的某些信息，同时支持一定程度的修改。
- 利用 $\log$ 来达成时间和空间上的平衡，复杂度均为 $\mathcal{O}(n\log n)$。
- 显然，树状数组能做的，线段树也一定能做。但是树状数组更好写，并且常数也更小。
- 简单讲一下线段树的标记永久化操作，部分条件下有用。

标记永久化

标记永久化

- 线段树区间修改的时候，会用到标记和标记下传。

标记永久化

- 线段树区间修改的时候，会用到标记和标记下传。
- 在一些情况下，标记也可以不需要下传。

标记永久化

- 线段树区间修改的时候，会用到标记和标记下传。
- 在一些情况下，标记也可以不需要下传。
- 例如区间加，区间修改等，大多只需要多记录一下修改时间，查询的时候根据修改时间讨论一下就可以了。

标记永久化

- 线段树区间修改的时候，会用到标记和标记下传。
- 在一些情况下，标记也可以不需要下传。
- 例如区间加，区间修改等，大多只需要多记录一下修改时间，查询的时候根据修改时间讨论一下就可以了。
- 在一些标记下传会破坏复杂度，比如可持久化的时候就很重要。

简单介绍

简单介绍

- 并查集是一种，维护“祖先关系”的简单数据结构，同样不做过多说明。

简单介绍

- 并查集是一种，维护“祖先关系”的简单数据结构，同样不做过多说明。
- 这里我们主要对它的复杂度进行一个分析，同时介绍一个很强大的思想。

简单介绍

- 并查集是一种，维护“祖先关系”的简单数据结构，同样不做过多说明。
- 这里我们主要对它的复杂度进行一个分析，同时介绍一个很强大的思想。
- 启发式合并。

复杂度是 $\mathcal{O}(\log n)$?

复杂度是 $\mathcal{O}(\log n)$?

- 第一次接触并查集时，大多使用的是一个叫“路径压缩”的方法来保证它的复杂度。

复杂度是 $\mathcal{O}(\log n)$?

- 第一次接触并查集时，大多使用的是一个叫“路径压缩”的方法来保证它的复杂度。
- 那这为什么是对的?

复杂度是 $\mathcal{O}(\log n)$?

- 第一次接触并查集时，大多使用的是一个叫“路径压缩”的方法来保证它的复杂度。
- 那这为什么是对的?
- 实际上，这个做法在少数情况下是会被卡的。有经验的出题人甚至会特意针对这种做法。

复杂度是 $\mathcal{O}(\log n)$?

- 第一次接触并查集时，大多使用的是一个叫“路径压缩”的方法来保证它的复杂度。
- 那这为什么是对的?
- 实际上，这个做法在少数情况下是会被卡的。有经验的出题人甚至会特意针对这种做法。
- 这里我们不打算讲，有兴趣的同学可以自行去 OI Wiki 或者其他网站上查找。

复杂度是 $\mathcal{O}(\log n)$?

- 第一次接触并查集时，大多使用的是一个叫“路径压缩”的方法来保证它的复杂度。
- 那这为什么是对的?
- 实际上，这个做法在少数情况下是会被卡的。有经验的出题人甚至会特意针对这种做法。
- 这里我们不打算讲，有兴趣的同学可以自行去 OI Wiki 或者其他网站上查找。
- 所以正确做法是什么?

启发式合并

启发式合并

- 在这之前，先来介绍一下启发式合并到底怎么启发了。就以并查集为例子。

启发式合并

- 在这之前，先来介绍一下启发式合并到底怎么启发了。就以并查集为例子。
- 每次，比较两个需要合并的集合的大小，然后把小的那个合并到大的那个上面。

启发式合并

- 在这之前，先来介绍一下启发式合并到底怎么启发了。就以并查集为例子。
- 每次，比较两个需要合并的集合的大小，然后把小的那个合并到大的那个上面。
- 就这么简单。简单到我们需要严谨证明一下。

启发式合并

- 在这之前，先来介绍一下启发式合并到底怎么启发了。就以并查集为例子。
- 每次，比较两个需要合并的集合的大小，然后把小的那个合并到大的那个上面。
- 就这么简单。简单到我们需要严谨证明一下。
- 对于合并部分，考虑每个元素，由于每次它所在的集合合并都会导致集合大小至少 $\times 2$，因此最多合并次数是 $\log n$ 次。

启发式合并

- 在这之前，先来介绍一下启发式合并到底怎么启发了。就以并查集为例子。
- 每次，比较两个需要合并的集合的大小，然后把小的那个合并到大的那个上面。
- 就这么简单。简单到我们需要严谨证明一下。
- 对于合并部分，考虑每个元素，由于每次它所在的集合合并都会导致集合大小至少 $\times 2$，因此最多合并次数是 $\log n$ 次。
- 对 n 个元素，那就是 $\mathcal{O}(n \log n)$ 的了。

启发式合并

启发式合并

- 对于查询？可能不是那么直观。

启发式合并

- 对于查询？可能不是那么直观。
- 这里有一个结论，启发式合并最多导致整棵树的深度 $+1$。

启发式合并

- 对于查询？可能不是那么直观。
- 这里有一个结论，启发式合并最多导致整棵树的深度 $+1$。
- 也就是说，启发式合并中，小的子树的深度一定不大于大的子树。

启发式合并

- 对于查询？可能不是那么直观。
- 这里有一个结论，启发式合并最多导致整棵树的深度 $+1$。
- 也就是说，启发式合并中，小的子树的深度一定不大于大的子树。
- 证明用数学归纳法分奇偶讨论，

  $dep_{2k} = dep_{2k+1}, dep_{2k+2} = dep_{2k} + 1 / dep_{2k+1}$。

启发式合并

- 对于查询？可能不是那么直观。
- 这里有一个结论，启发式合并最多导致整棵树的深度 $+1$。
- 也就是说，启发式合并中，小的子树的深度一定不大于大的子树。
- 证明用数学归纳法分奇偶讨论，
  $dep_{2k} = dep_{2k+1}, dep_{2k+2} = dep_{2k} + 1/dep_{2k+1}$。
- 于是，总深度一定是 $\log n$ 级别的，也就证明完毕了。

左偏树

左偏树

- 普通的二叉堆我们就不介绍了。这里介绍一种简单的可并堆：左偏树。

左偏树

- 普通的二叉堆我们就不介绍了。这里介绍一种简单的可并堆：左偏树。
- 可并堆，顾名思义，可以在较低的时间复杂度下完成对两个堆的合并。左偏树是其中较为简单的一种。

左偏树

- 普通的二叉堆我们就不介绍了。这里介绍一种简单的可并堆：左偏树。
- 可并堆，顾名思义，可以在较低的时间复杂度下完成对两个堆的合并。左偏树是其中较为简单的一种。
- 何为左偏？我们定义外节点为左儿子或右儿子不存在的点，$dist$ 为一个节点到离他最近的外节点的距离 $+1$，空节点的 $dist = 0$。

左偏树

- 普通的二叉堆我们就不介绍了。这里介绍一种简单的可并堆：左偏树。
- 可并堆，顾名思义，可以在较低的时间复杂度下完成对两个堆的合并。左偏树是其中较为简单的一种。
- 何为左偏？我们定义外节点为左儿子或右儿子不存在的点，$dist$ 为一个节点到离他最近的外节点的距离 $+1$，空节点的 $dist = 0$。
- 其中 $+1$ 只是为了写代码方便，没有其他含义。

左偏树

- 普通的二叉堆我们就不介绍了。这里介绍一种简单的可并堆：左偏树。
- 可并堆，顾名思义，可以在较低的时间复杂度下完成对两个堆的合并。左偏树是其中较为简单的一种。
- 何为左偏？我们定义外节点为左儿子或右儿子不存在的点，$dist$ 为一个节点到离他最近的外节点的距离 $+1$，空节点的 $dist = 0$。
- 其中 $+1$ 只是为了写代码方便，没有其他含义。
- 左偏树满足 $dist_{lson} \geq dist_{rson}$，稍微画一下图就可以发现，它的确是“左偏”的。

几个性质

几个性质

- 首先，$dist_u = dist_{rson} + 1$，这一点应该很显然。

几个性质

- 首先，$dist_u = dist_{rson} + 1$，这一点应该很显然。
- 其次，一棵 n 个点左偏树，$dist$ 的最大值应该在 $\log n$ 级别。

几个性质

- 首先，$dist_u = dist_{rson} + 1$，这一点应该很显然。
- 其次，一棵 n 个点左偏树，$dist$ 的最大值应该在 $\log n$ 级别。
- 假设根的 $dist = x$，那么这棵树至少有 $x - 1$ 层是满二叉树，这样就至少有 $2^x - 1$ 个节点了。

几个性质

- 首先，$dist_u = dist_{rson} + 1$，这一点应该很显然。
- 其次，一棵 n 个点左偏树，$dist$ 的最大值应该在 $\log n$ 级别。
- 假设根的 $dist = x$，那么这棵树至少有 $x - 1$ 层是满二叉树，这样就至少有 $2^x - 1$ 个节点了。
- 需要注意的是，左偏树的深度是没有保证的。一条向左的链同样是正确的左偏树，这一点请牢记于心。

merge 操作

$merge$ 操作

- $merge$ 操作是左偏树的核心，剩下所有操作都是基于 $merge$。

$merge$ 操作

- $merge$ 操作是左偏树的核心，剩下所有操作都是基于 $merge$。
- 看一下代码。

其他操作

其他操作

- 插入元素：一个点也算作一个堆，直接合并即可。

其他操作

- 插入元素：一个点也算作一个堆，直接合并即可。
- 删除堆顶：合并根的左右儿子。

其他操作

- 插入元素：一个点也算作一个堆，直接合并即可。
- 删除堆顶：合并根的左右儿子。
- 不改变相对大小的全局操作（加乘）：在根上打标记，$merge$ 时 pushdown。

其他操作

其他操作

- 删除任意节点：儿子合并后向上更新 $dist$，并交换左右儿子维护左偏性质，直到 $dist$ 不改变为止。

其他操作

- 删除任意节点：儿子合并后向上更新 $dist$，并交换左右儿子维护左偏性质，直到 $dist$ 不改变为止。
- 令删除的节点为 u，若 u 为右儿子，$dist_{fa_u} = dist_u + 1$。

其他操作

- 删除任意节点：儿子合并后向上更新 $dist$，并交换左右儿子维护左偏性质，直到 $dist$ 不改变为止。
- 令删除的节点为 u，若 u 为右儿子，$dist_{fa_u} = dist_u + 1$。
- 若 u 为左儿子，当且仅当它和右儿子的 $dist$ 一样的时候，fa_u 的 $dist$ 才会更新。

其他操作

- 删除任意节点：儿子合并后向上更新 $dist$，并交换左右儿子维护左偏性质，直到 $dist$ 不改变为止。
- 令删除的节点为 u，若 u 为右儿子，$dist_{fa_u} = dist_u + 1$。
- 若 u 为左儿子，当且仅当它和右儿子的 $dist$ 一样的时候，fa_u 的 $dist$ 才会更新。
- 因此每层 $dist$ 均 $+1$。结合 $dist$ 的性质可知，操作次数在 $\log n$ 级别。

[SCOI2011] 棘手的操作

[SCOI2011] 棘手的操作

有 n 个节点，一开始相互不连通。

第 i 个节点的初始权值为 a_i，接下来有如下一些操作：

- U x y：加一条边，连接第 x 个节点和第 y 个节点。
- A1 x v：将第 x 个节点的权值增加 v。
- A2 x v：将第 x 个节点所在的连通块的所有节点的权值都增加 v。
- A3 v：将所有节点的权值都增加 v。
- F1 x：输出第 x 个节点当前的权值。
- F2 x：输出第 x 个节点所在的连通块中，权值最大的节点的权值。
- F3：输出所有节点中，权值最大的节点的权值。

$n, q \leq 3 \times 10^5$。

重链剖分

重链剖分

- 对于一棵树，先给出若干定义如下：

重链剖分

- 对于一棵树，先给出若干定义如下：
- 重儿子：一个点的子节点中，子树大小最大的子节点。

重链剖分

- 对于一棵树，先给出若干定义如下：
- 重儿子：一个点的子节点中，子树大小最大的子节点。
- 轻儿子：不是重儿子的其他子节点。

重链剖分

- 对于一棵树，先给出若干定义如下：
- 重儿子：一个点的子节点中，子树大小最大的子节点。
- 轻儿子：不是重儿子的其他子节点。
- 重边：一个点与其重儿子的连边。重链：首尾衔接的重边。

重链剖分

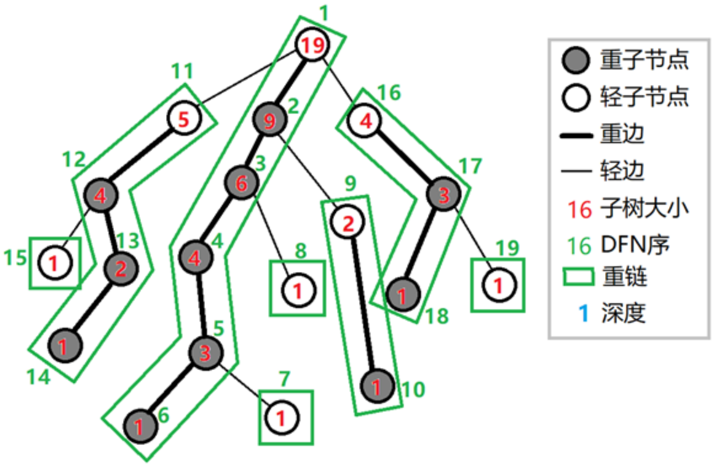
- 对于一棵树，先给出若干定义如下：
- 重儿子：一个点的子节点中，子树大小最大的子节点。
- 轻儿子：不是重儿子的其他子节点。
- 重边：一个点与其重儿子的连边。重链：首尾衔接的重边。
- 轻边：剩下的边。

重链剖分

- 对于一棵树，先给出若干定义如下：
- 重儿子：一个点的子节点中，子树大小最大的子节点。
- 轻儿子：不是重儿子的其他子节点。
- 重边：一个点与其重儿子的连边。重链：首尾衔接的重边。
- 轻边：剩下的边。
- 把单个点也看成重链，那么整棵树就被剖分成若干条重链。

重链剖分

重链剖分



重要性质

重要性质

- 任意一条父子链的轻边条数在 $\log n$ 级别。

重要性质

- 任意一条父子链的轻边条数在 $\log n$ 级别。
- 证明考虑每走过一条轻边，子树大小必定至少翻倍。

重要性质

- 任意一条父子链的轻边条数在 $\log n$ 级别。
- 证明考虑每走过一条轻边，子树大小必定至少翻倍。
- 有什么用？回到上一张图看看。

重要性质

- 任意一条父子链的轻边条数在 $\log n$ 级别。
- 证明考虑每走过一条轻边，子树大小必定至少翻倍。
- 有什么用？回到上一张图看看。
- 我们注意到，如果让每条重链的 dfn 序连续，那么每条父子链都可以按照 dfn 序被拆分成连续的 $\log n$ 段区间。

重要性质

- 任意一条父子链的轻边条数在 $\log n$ 级别。
- 证明考虑每走过一条轻边，子树大小必定至少翻倍。
- 有什么用？回到上一张图看看。
- 我们注意到，如果让每条重链的 dfn 序连续，那么每条父子链都可以按照 dfn 序被拆分成连续的 $\log n$ 段区间。
- 推广一下，树上任意一条链都可以按照 dfn 序被拆分成连续的 $\log n$ 段区间。

重要性质

- 任意一条父子链的轻边条数在 $\log n$ 级别。
- 证明考虑每走过一条轻边，子树大小必定至少翻倍。
- 有什么用？回到上一张图看看。
- 我们注意到，如果让每条重链的 dfn 序连续，那么每条父子链都可以按照 dfn 序被拆分成连续的 $\log n$ 段区间。
- 推广一下，树上任意一条链都可以按照 dfn 序被拆分成连续的 $\log n$ 段区间。
- 如果我们结合上区间修改相关的数据结构？线段树？

应用

应用

- 于是，借助树剖和线段树，现在我们可以实现树上任意一条链的一些操作。

应用

- 于是，借助树剖和线段树，现在我们可以实现树上任意一条链的一些操作。
- 顺带一提，由于 dfn 序本身的特殊性，子树的 dfn 序必然是一个连续区间，因此子树操作也可以实现。

应用

- 于是，借助树剖和线段树，现在我们可以实现树上任意一条链的一些操作。
- 顺带一提，由于 dfn 序本身的特殊性，子树的 dfn 序必然是一个连续区间，因此子树操作也可以实现。
- $\log n$ 个重链区间，加上线段树自带的 $\log n$，单次操作的复杂度是 $\mathcal{O}(\log^2 n)$。

应用

- 于是，借助树剖和线段树，现在我们可以实现树上任意一条链的一些操作。
- 顺带一提，由于 dfn 序本身的特殊性，子树的 dfn 序必然是一个连续区间，因此子树操作也可以实现。
- $\log n$ 个重链区间，加上线段树自带的 $\log n$，单次操作的复杂度是 $\mathcal{O}(\log^2 n)$。
- 并且，树剖的常数很小，很多时候吊打理论复杂度 $\mathcal{O}(\log n)$ 的 LCT。

应用

- 于是，借助树剖和线段树，现在我们可以实现树上任意一条链的一些操作。
- 顺带一提，由于 dfn 序本身的特殊性，子树的 dfn 序必然是一个连续区间，因此子树操作也可以实现。
- $\log n$ 个重链区间，加上线段树自带的 $\log n$，单次操作的复杂度是 $\mathcal{O}(\log^2 n)$。
- 并且，树剖的常数很小，很多时候吊打理论复杂度 $\mathcal{O}(\log n)$ 的 LCT。
- 看看板子题。

[ZJOI2008] 树的统计

[ZJOI2008] 树的统计

一棵树上有 n 个节点，每个节点都有一个权值 w。

有如下 q 次操作：

1. 把结点 u 的权值改为 t。
2. 询问从点 u 到点 v 的路径上的节点的最大权值。
3. 询问从点 u 到点 v 的路径上的节点的权值和。

从点 u 到点 v 的路径上的节点包括 u 和 v 本身。

$n \leq 3 \times 10^4, q \leq 2 \times 10^5$。

何为可持久化?

何为可持久化?

- 有的时候，题目会要求我们将操作回退，也即进行 undo/redo 操作。

何为可持久化?

- 有的时候，题目会要求我们将操作回退，也即进行 undo/redo 操作。
- 使数据结构能够执行这样的操作的改造，就叫可持久化。

何为可持久化?

- 有的时候，题目会要求我们将操作回退，也即进行 undo/redo 操作。
- 使数据结构能够执行这样的操作的改造，就叫可持久化。
- 总的来说，可持久化只有一个中心思想：存档只存改过的。以下以线段树为例。

可持久化线段树

可持久化线段树

- 我们知道，无论是区间修改还是单点修改，线段树上变化的点数都只有 $\log n$ 级别个。

可持久化线段树

- 我们知道，无论是区间修改还是单点修改，线段树上变化的点数都只有 $\log n$ 级别个。
- 因此一个新的线段树，其实只有 $\log n$ 个点和原来不同。我们只要对所有修改过的点新建一个版本，而不需要复制整棵树。

可持久化线段树

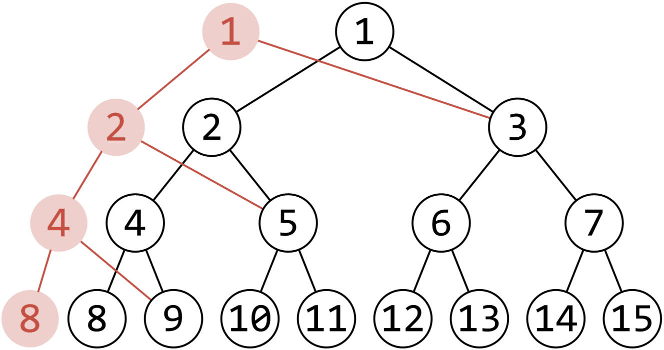
- 我们知道，无论是区间修改还是单点修改，线段树上变化的点数都只有 $\log n$ 级别个。
- 因此一个新的线段树，其实只有 $\log n$ 个点和原来不同。我们只要对所有修改过的点新建一个版本，而不需要复制整棵树。
- 需要注意的是，显然父子之间的编号关系被破坏了，因而可持久化时需要动态开点。

可持久化线段树

- 我们知道，无论是区间修改还是单点修改，线段树上变化的点数都只有 $\log n$ 级别个。
- 因此一个新的线段树，其实只有 $\log n$ 个点和原来不同。我们只要对所有修改过的点新建一个版本，而不需要复制整棵树。
- 需要注意的是，显然父子之间的编号关系被破坏了，因而可持久化时需要动态开点。
- 下面是一张经典的图。

可持久化线段树

可持久化线段树



可持久化数据结构

可持久化数据结构

- 事实上，依据这个思想，任意一个数据结构，只要新建一个版本的复杂度不高于原本修改的复杂度，都可以可持久化。

可持久化数据结构

- 事实上，依据这个思想，任意一个数据结构，只要新建一个版本的复杂度不高于原本修改的复杂度，都可以可持久化。
- 这也是左偏树，无旋 Treap 等的存在意义之一。和二叉堆不同，修改不会改变树的结构，也因而可持久化。

在做题以前

在做题以前

- 今天讲的是数据结构，但是裸数据结构题几乎不会出现。

在做题以前

- 今天讲的是数据结构，但是裸数据结构题几乎不会出现。
- 要理解一个思想：数据结构只是一个工具，题目的核心环节应当是对条件的转换。

在做题以前

- 今天讲的是数据结构，但是裸数据结构题几乎不会出现。
- 要理解一个思想：数据结构只是一个工具，题目的核心环节应当是对条件的转换。
- 数据结构承担的仅仅是“维护某个需要维护的东西”的作用。

[NOIP2017 提高组] 列队

[NOIP2017 提高组] 列队

n 行 m 列个学生，从左往右，从上往下地依次编号为 $1 \sim n \times m$。

q 次事件，每次有 1 个坐标为 (i, j) 的学生离队。

离队后，所有学生依次执行以下两个操作：

- 第一列保持不动，所有学生向左填补空位。

- 第一行保持不动，所有学生向前填补空位。

显然这样空位来到 (n, m) 的位置，此时离队的学生回到此空位。

注意：编号不会随着学生位置改变。

求每次离队的学生的编号。

$n, m, q \leq 3 \times 10^5$。

[十二省联考 2019] 春节十二响

[十二省联考 2019] 春节十二响

给一棵 n 个点的树，每个点有一个权值 w_i。

你每次可以选取树上的一个点集，要求点集中的每个点不能是另一个点的祖先，而选出点集的代价为点集中权值最大点的权值。

问将所有点都选一遍的最小代价为多少，每次选的点集不能包含之前已经被选过的点。

$n \le 2 \times 10^5$。

提示：若树是一条链，怎么做。

[NOI2021] 轻重边

[NOI2021] 轻重边

给一棵 n 个点的树，树上的每一条边可能是轻边或者重边。

接下来对树进行 m 次操作，在操作开始前，树上所有边都是轻边。操作有以下两种：

1. 给定两个点 u 和 v，对于 u 到 v 路径上的所有点 x（包含 u 和 v），将与 x 相连的所有边变为轻边。然后再将 u 到 v 路径上的所有边变为重边。
2. 给定两个点 u 和 v，查询当前 u 到 v 的路径有多少条重边。

$n, m \leq 10^5$。

提示：当两端的点满足什么条件时，一条边是重边？

[JLOI2015] 城池攻占

[JLOI2015] 城池攻占

有 n 个城池构成一棵树，每个城池有一个防御力 h_i。

现在有 m 个骑士，每个骑士有一个初始攻击力 s_i，每个骑士都会从一个城池出发，一路向上攻打直到到根所在的城池。

显然地，如果一个骑士的攻击力低于城池的防御力，他就会在这个城池阵亡。

否则，他会攻克这个城池，并且攻击力会加上/乘上 v_i。

m 个骑士相互独立，也即一个骑士的攻打结果对其他骑士没有影响。

求对每个骑士，他攻克的城池数；以及对每个城池，有多少个骑士阵亡在这里。

$n, m \leq 3 \times 10^5$。

基于重链剖分的毛毛虫剖分