数据结构纵览

2024 年 8 月

feecle8146

目录

今天我们将复习、学习、练习以下内容：

1. 线段树，树状数组，倍增，ST 表。
2. 基本的树上数据结构问题。
3. 单调栈和笛卡尔树。
4. 经典模型简介：分治和偏序问题。

同时，也将见到若干个综合例题。

线段树

区间加区间求和线段树的正确性

在联赛阶段，很容易出现的问题是：写出了很长的数据结构代码，却不知道如何调试。

为了避免该情况，我们需要对各类数据结构都有清晰的理解，并知道：如果我的最终结果错了，如何找到数据结构本身错在哪。

对于“区间加区间求和线段树”，其正确性依赖于如下的等式：

$$\sum_{i=l_p}^{r_p} a_{i\,(\text{真实值})} = sum_p + (r_p - l_p + 1) \times \sum_{q\text{是}p\text{的祖先},q\neq p} tag_q$$

一般线段树的正确性

我们把类似的“某个算法保持进行完操作后满足的等式”统称为循环不变式。

你能不能写出一般线段树（有 pushdown 也有 pushup，不妨假设 pushup 的是区间和）的循环不变式？

$$\sum_{i=l_p}^{r_p} a_{i\,(\text{真实值})}$$

$= sum_p$ 在被祖先标记按照由深到浅(也就是由旧到新)的顺序作用后的值

设计 pushup pushdown 的方法

$$\sum_{i=l_p}^{r_p} a_{i\,(\text{真实值})}$$

$= sum_p$ 在被祖先标记按照由深到浅(也就是由旧到新)的顺序作用后的值

从这个循环不变式，我们可以看出：

1. 标记需要复合封闭。如果还有区间询问，还要保证打标记之后，询问的信息变化是可控的。
标记 tag 可以看作值 x 到值 y 的函数：x 在打了 tag 后变成 y，就说 $tag(x) = y$。而 x 打了 tag_1 之后再打 tag_2，其实就是 $x \rightarrow tag_2(tag_1(x))$。
复合封闭性，就是无论打多少个标记，总可以用同一种形式来表示。

线段树

设计 pushup pushdown 的方法

$$\sum_{i=l_p}^{r_p} a_{i\,(\text{真实值})}$$

$= sum_p$ 在被祖先标记按照由深到浅(也就是由旧到新)的顺序作用后的值

从这个循环不变式，我们可以看出：

2. 信息需要可合并。也就是，能够 pushup。
如果只维护题目要求的信息 a 不可合并，不妨想想：如果已知 a_l 和 a_r，还需要知道什么（叫做 b）才能算出 a_p。进一步，如果 (a,b) 还是不能算出 b，想想：如果已知 b_l 和 b_r，还需要知道什么才能算出 b_p……
如果在常数步内能结束上面的迭代，信息就是可合并的。

线段树

序列操作

P2572

给定一个 01 数组，支持区间赋值、区间取反、求区间 1 的个数、区间最长 1 连续段。

$$n, q \le 200000$$

K 次方和

无来源

给定一个数组，支持区间加、区间求 K 次方和。

$$n, q \le 200000, K \le 5$$

最大值个数

无来源

给定一个数组，支持区间加、查询区间最大值的出现次数。

$$n, q \le 500000$$

线段树

基础例子

序列操作

P4247

给定一个数组，支持区间加、区间取反、区间求选出 K 个下标不同的数，所有选法得到的积的和。

$$n, q \le 50000, K \le 20$$

求数值

AT_abl_e

给定一个数字串，只包含 0~9。要求支持区间赋值，求出数字串从左到右读出的值。

$$n, q \le 500000$$

经典问题

无来源

给定一个数组 a，支持区间加、求 $\sum_{i=l}^{r}\sum_{j=i}^{r}\sum_{k=i}^{j} a_k$。

$$n, q \le 500000$$

试看看！

幻梦 P8969

给定一个数组，支持区间加、区间 $x \to popcount(x)$ 、单点查询。

$$n, q \le 300000, \text{保证总有} a_i \le 10^{18}$$

请你设计标记，并说明 maketag 和 pushdown 函数的写法，并分析时间复杂度。

提示：

1. $popcount(x) \le 60$。
2. 标记本质上是一种“函数”。

试看看！

幻梦 P8969

给定一个数组，支持区间加、区间 $x \to popcount(x)$ 、单点查询。

$$n, q \le 300000, \text{保证总有} a_i \le 10^{18}$$

如果一个区间没有被 $popcount$ 过，维护加法标记足矣；如果被 $popcount$ 了，考虑第一次 $popcount$ 时的情况。

在第一次 $popcount$ 之前，只有加法操作 $x \to x + A$。

$$x \to x + A \to popcount(x + A) \to \cdots$$

蓝色部分有何性质？

试看看！

幻梦 P8969

给定一个数组，支持区间加、区间 $x \to popcount(x)$ 、单点查询。

$$n, q \le 300000, \text{保证总有} a_i \le 10^{18}$$

$$popcount(x + A) \to \cdots$$

这部分如果看作一个函数，它的定义域大小只有 60，完全可以直接存储下来这个函数作用在 0 ~ 60 的所有数分别的结果。

因此，只要作用过 $popcount$，就可以在标记中维护 A 和数组 $f[0 \dots 60]$，表示 $x \to f[popcount(x + A)]$。函数的复合也是容易的。

线段树

线段树上二分的一般方法

经典问题 无来源

给定一个数组，支持区间加、区间对 u 取 max，求出从左往右数，下标 $\geq x$ 的第一个值 $\geq y$ 的位置。

$$n,q \leq 500000$$

设函数 $Find(p,x,y)$ 表示只考虑 p 代表的区间内，询问 (x,y) 的答案。

维护区间最大值于变量 mx_p。

思考：这段代码的复杂度是什么？

```
int Find(int p, int l, int r, int x, int y) {
    if (mx[p] < y || r < x) return -1;
    if (l == r) return l;
    pushdown(p);
    int mid = (l + r) >> 1, res = Find(p * 2, l,
mid, x, y);
    if (res != -1) return res; // 注意
    return Find(p * 2 + 1, mid + 1, r, x, y);
}
```

见风使舵 CF773E

对于长为 n 的数组 a，维护变量 x，初始为 0。然后，依次扫描 $a_1, \ldots, a_n$：若 $x > a_i$ 则 $x \to x - 1$；若 $x = a_i$ 则 x 不变；若 $x < a_i$ 则 $x \to x + 1$。定义该过程结束后 x 的值为 $F(a)$。

定义 $G(a)$ 为：任意排列 $a_1, \ldots, a_n$ 的前提下，$F(a)$ 的最大值。

现在给出序列 $b_1, \ldots, b_n$，对每个 i 求 $G(b[1 \ldots i])$。

$$n, |b_i| \leq 500000$$

第一步，我们需要确定如何排列一个数组 a，才能使 $F(a)$ 最大。

线段树

见风使舵 CF773E

现在给出序列 $b_1, \ldots, b_n$，对每个 i 求 $G(b[1 \ldots i])$。

$n, |b_i| \le 500000$

提示：Exchange argument。考虑 a_i, a_{i+1} 交换后什么时候一定不劣。

若 $a_i > a_{i+1}$，交换一定不劣，因此最优的 a 一定是递增排列的。

在递增排列的基础上，x 的变化又有何性质？

一定是先单调严格减少，然后单调不减。

见风使舵 CF773E

现在给出序列 $b_1, \dots, b_n$，对每个 i 求 $G(b[1 \dots i])$。

$n, |b_i| \le 500000$

最优的 a 一定是递增排列的，x 一定是先单调严格减少，然后单调不减。

1. 如何求减少、增加的分界点？
2. 如何求最终 x 的值？

提示：当你不知道干什么的时候，具体写出形式化的表达式。

线段树

见风使舵 CF773E

现在给出序列 $b_1, \ldots, b_n$，对每个 i 求 $G(b[1 \ldots i])$。

$n, |b_i| \le 500000$

最优的 a 一定是递增排列的，x 一定是先单调严格减少，然后单调不减。

1. 如何求减少、增加的分界点？
   分界点之前，$x_i = -i$，故就是找第一个 $a_i \ge -i$ 的 i。
2. 如何求最终 x 的值？
   先找到分界点 p，$x_p = -p$。之后，$x_j = \min(x_{j-1} + 1, a_j)$。

线段树

见风使舵 CF773E

现在给出序列 $b_1, \dots, b_n$，对每个 i 求 $G(b[1 \dots i])$。

$n, |b_i| \le 500000$

先找到分界点 p，$x_p = -p$。之后，$x_j = \min(x_{j-1} + 1, a_j)$。这一步怎么维护？

将式子展开：$x_j = \min(x_{j-1} + 1, a_j) = \min(x_{j-2} + 2, a_{j-1} + 1, a_j) = \cdots = \min(x_0 + (n - 0), a_1 + (n - 1), a_2 + (n - 2), \dots, a_n)$。

提出 n，就是求 $a_i - i$ 的最小值，其中 a_i 是从小到大排列的。

“从小到大排列”，说明我们应该以权值为下标建线段树。

线段树

见风使舵 CF773E

现在给出序列 $b_1, \dots, b_n$，对每个 i 求 $G(b[1 \dots i])$。

$n, |b_i| \le 500000$

为了避面相等元素造成干扰，不妨先“不等离散化”，人为把值相同的元素赋予不同排名。

$a_i - i, a_i + i$ 在插入新元素后的变化，无非是几个区间加。

第一步就是线段树上二分，第二步就是区间最小值。

时间复杂度：$O(n \log n)$。

小结

关于线段树的基本用法，应掌握以下几点。

1. 显式写出循环不变式，来帮助自己理解、证明算法；
2. 标记封闭性，设计标记的方法，知道标记的本质是一个函数；
3. 信息可合并性，设计信息的方法。
4. 维护复杂式子不如分离变量再维护简单式子。

现在，面对简单的线段树问题，你应当无所畏惧了。

树状数组

类比之前线段树的讨论，请你写出单点修改区间求和树状数组的循环不变式。

$$\forall i, c_i = \sum_{k=i-lowbit(i)+1}^{i} a_k$$

树状数组只能维护单点修改，区间加单点求值怎么办？区间加区间求和呢？

不论维护什么信息，目标都是把区间修改变成单点修改，求和的式子分离变量。

倍增和 ST 表

基本思想

倍增维护的是树上 x 朝上 2^k 个位置的信息。

ST 表是序列上的倍增，且特指用来维护幂等信息的一类倍增。此时，区间询问，只需要合并 1 次信息。

对于树上有可减性的信息，可以将 $x \to y$ 的路径拆分成 $O(1)$ 条到根链。需要特别注意 LCA 周围。

对于没有可减性但有可合并性的信息，可以将 $x \to y$ 的路径拆分成 $O(\log n)$ 条直上直下的、长度为 2^k 的链，有时需要注意顺序问题。

可合并性与线段树设计信息时遵循的原则是完全一样的。

数据传输 P8820

给出一棵树，点 x 有点权 v_x。

q 次询问，每次问 x, y，你需要求出一个序列 $a_1 = x, a_2, \dots, a_k = y$，使得 $dis(a_i, a_{i+1}) \le K$ 且 $\sum v_{a_i}$ 最小。

$$n, q \le 200000, K \le 3$$

设计信息时，树的结构已经不重要，可以把链想象成一个序列。

$k = 3$ 时，最优解可能跳到离路径距离为 1 的点上，但这个点一定是路径上的点周围点权最小者。
对于序列的一个区间 $[l, r]$，记录四个数 $ans[x][y]$，分别表示从 $l + x$ 跳到 $r - y$（或者周围一圈的点）的最小代价。

保卫王国 P5024

给出一棵树，点 x 有点权 v_x。每次询问给定两个点必须选/不选，求此时的最小点覆盖（选权值和最小的点，满足每条边都至少一端被选）。

$$n, q \leq 300000$$

在分类讨论上而不是思维上上强度的题目，我们称为“屑题”。

dfs 序和树上差分

基本思想

希望将树上问题转化为序列问题。

x 子树的 dfn 恰好构成了区间 $[dfn_x, dfn_x + sz_x)$。这就把子树限制转化为了区间限制。

树上差分是树上前缀和的逆操作。类比序列问题，我们希望把链加单点查化为单点加查询“前缀和”：在树上，一般认为这里的前缀和是子树和。

1. 对于链加点权，一般写成 x, y 加，lca, fa_{lca} 减。
2. 对于链加边权，一般令点权等于 x, fa_x 间的边权，x, y 加，lca 减两倍。

dfs 序和树上差分

综合运用

简单环 P5203

给定一棵树，和 $m-n+1$ 条非树边，构成一个无向图。询问有多少个简单环恰好包含 2 条非树边（点相同但是经过的重边不同，算不同的环）。

$$n \le m \le 10^5$$

因为每个符合要求的简单环都包含两条非树边，自然想到先探究什么样的两条非树边能取出环，以及能取出几个环。

提示：在树上加边问题中，考虑非树边覆盖的树边是常见的思考方式。

dfs 序和树上差分

简单环 P5203

给定一棵树，和 $m-n+1$ 条非树边，构成一个无向图。询问有多少个简单环恰好包含 2 条非树边（点相同但是经过的重边不同，算不同的环）。

$$n \le m \le 10^5$$

若两条非树边覆盖的树边有交，则存在恰好一条简单环，否则不存在简单环。因此，只需要算有几对非树边覆盖的边有交。

若有交，交一定是一条链。如果枚举交包含的一条边，树上差分算出有 k 条非树边边经过这条边，则这条边会带来 $C(k,2)$ 的答案，可惜这样会算重。

dfs 序和树上差分

综合运用

简单环 P5203

给定一棵树，和 $m-n+1$ 条非树边，构成一个无向图。询问有多少个简单环恰好包含 2 条非树边（点相同但是经过的重边不同，算不同的环）。

$$n \le m \le 10^5$$

点减边容斥：对于树上一个连通块只应该被算一次的问题，因为连通块的点数 - 边数总是 1，所以算点的答案和 - 边的答案和即可。

在本题中，我们希望每条长度 ≥ 1（边数）的路径都被算恰好一次，而每条边的贡献已经算出来了。还需要减去什么？

每两条边的贡献

dfs 序和树上差分

简单环 P5203

给定一棵树，和 $m-n+1$ 条非树边，构成一个无向图。询问有多少个简单环恰好包含 2 条非树边（点相同但是经过的重边不同，算不同的环）。

$$n \le m \le 10^5$$

对于直上直下的两条边，可以把贡献记在最深的点上，树上差分计算。

对于跨 LCA 的，每条路径至多一处，可以用 map 直接计算。

代码：[P5203 - 洛谷专栏 (luogu.com.cn)]

单调栈

单调栈的使用思路

单调栈回答了如下的问题：给定一个序列 a。对于每个 $1 \le i \le n$，维护出 $a[1 \dots i]$ 这个前缀的所有后缀最大值的信息。

由此，我们可以得到 l_i, r_i（i 左右两侧最近的值大于 a_i 的下标）。这两个数组对许多与区间最值有关的问题都紧密相关。

同时，单调栈是一个只有尾部变化的数组，因此在栈上可以进行前缀和、递推等丰富的操作。

下面，我们看一些应用。

单调栈

基本运用

区间后缀最大值和

无来源

给你一个数组，多次询问区间的所有后缀的最大值之和。

$$n, q \leq 300000$$

区间最大值之和

无来源

给你一个数组，求所有区间 $[l, r]$ 的最大值之和。

要求线性。

单调栈

带删数的区间最大值之和 CF1988E

定义数组 a 所有区间的最大值之和为 $f(a)$。

设 a 删掉第 i 个元素得到的数组为 b_i，求所有 $f(b_i)$。

如何将上述问题的做法泛化到本题？

提示：枚举最小值位置 j，考虑 j 对哪些 b_i 有贡献。

分 i 所在区间（与 l_j, r_j 位置关系）讨论，用差分计算答案。

单调栈

体育馆问题

CF1601E

有一个体育馆，第 $i\ (1 \le i \le n)$ 天，票价为 a_i 元。每张你手上的票可以管任意连续的 k（定值）天，也就是说，如果你在第 i 天买了这张票，你可以任意选择 A，满足 $i \le A$，这样，第 $A, A+1, \ldots, A+k-1$ 天都可以用这张票进入体育馆。

q 次询问，每次给出 l, r，问如果第 l 天某人来到这个城市（也就是无法在第 l 天前买票），并且要在第 $l, l+1, \ldots, r$ 天进入体育馆，至少要花多少钱。

$$n, q \le 300000$$

提示：先形式化地写出问题。

体育馆问题 CF1601E

$$\sum_{i\in \mathrm{N}}[l+ik\le r]\min_{l\le j\le l+ik} a_j$$

注意到最小值的计算除了第一天 a_l，都是 k 个 k 个一组的。同时，不妨设 l,r 模 k 同余。

$\min_{l\le j\le l+ik} a_j$ 可以写成 $\min(a_l, b_{l+k}, \dots, b_{l+ik})$，其中 $b_i=\min_{i-k<j\le i} a_j$。可以用单调队列 $O(n)$ 求出 b。

思考：此时能否再简化问题？
提示：发现独立性。

单调栈

综合运用

体育馆问题 CF1601E

$$\sum_{i\in\mathrm{N}}[l+ik\le r]\min_{l\le j\le l+ik}a_j$$

发现独立性。

注意到模 k 不同余的 a,b 对询问互不影响，所以可以分开处理每种模 k 的余数。

分开处理后，即为：给定一个序列，区间询问区间每个位置的 min(区间前缀最小值, v) 之和。
在单调栈上二分，找到 min 的分界点，前缀和回答。

笛卡尔树

笛卡尔树的使用思路

笛卡尔树构建稍微比单调栈麻烦一点，但其在单调栈的基础上提供了更丰富的信息：**将区间最值的问题转化为了树上 LCA 相关问题，将 l_i, r_i 相关问题转化为了树上左链、右链相关问题**，使得我们能用树形结构的特殊性质处理它。

所谓“Kruskal 重构树”其实也与笛卡尔树没有本质区别，只是一个维护的是图上（准确地说，最小生成树的链上）信息，一个维护的是序列上信息。

笛卡尔树

带删数的区间最大值之和 CF1988E

定义数组 a 所有区间的最大值之和为 $f(a)$。

设 a 删掉第 i 个元素得到的数组为 b_i，求所有 $f(b_i)$。

删去某个元素后，笛卡尔树的改变是什么样的？

类似于无旋 Treap 的 merge。而且可以发现，直接暴力 $O(\text{链长})$ merge，时间复杂度也是对的。

分治

序列分治梗概

统计序列所有区间的信息的分治：定义 Solve(L, R) 表示求出 $L \le l \le r \le R$ 的整体结果。令 $mid = \frac{L+R}{2}$（取中点目的是减少递归层数），先求出 $L \le l \le mid, mid < r \le R$ 的整体结果，再调用 Solve(L, mid) 和 Solve(mid+1, r)。

回答若干个针对区间的询问的分治：定义 Solve(L, R) 表示求出 $L \le l \le mid, mid < r \le R$ 的所有询问的结果，其中 $mid = \frac{L+R}{2}$。递归过程同上，可以在 solve 过程中同时传一个 vector 存储待回答的询问。

分治

分治的目的

为什么分治能简化问题？

1. 原来的限制 $l < r$ 简化成了 $l \in A, r \in B$。
2. 对于与序列的区间 $[l, r]$ 相关的问题，左右两边可能具有独立性，且只需要在最后进行一次信息合并。相对地，线段树就需要 $\log n$ 次。
3. 对于与序列的区间 $[l, r]$ 的所有子区间 $[l', r']$ 相关的问题，将问题拆成了一个前缀、一个后缀和跨过的部分。前缀和后缀比区间简单，而跨过的就去掉了 $l' \le r'$ 的限制。

分治

P6240

给定一个序列，每个元素是物品，具有体积 w_i 和价值 c_i。q 次询问，每次询问一个区间 $[l,r]$ 和背包大小 V，请你求出用这个背包去装区间内的物品（每个物品只能装一次），价值和最大是多少。

$$n \le 10^4, q \le 10^5, V \le 500$$

分治，在 $l \in [L,mid], r \in [mid+1,R]$ 的时候处理询问 $[l,r]$。具体如何处理？

对于询问 (l,r)，答案就是 $[l,r]$ 区间背包的一项，可以由左侧后缀背包和右侧前缀背包 $O(V)$ 合并而来。

分治

P6240

给定一个序列，每个元素是物品，具有体积 w_i 和价值 c_i。q 次询问，每次询问一个区间 $[l,r]$ 和背包大小 V，请你求出用这个背包去装区间内的物品（每个物品只能装一次），价值和最大是多少。

$$n \le 10^4, q \le 10^5, V \le 500$$

思考：

1. 上述算法的时间复杂度？

$$O(nV\log n + qV + q\log n)$$

2. 本题中分治和线段树相比体现了什么特点？

   只需一次合并的一项，而线段树需 log 次合并的 $O(V)$ 项

最大独立集　　P7482

给定一个序列，求所有区间的最大独立集之和。

$$n \le 10^5$$

看题，便知可立刻写出最大独立集的二信息合并形式：

预处理出 $l0, l1, r0, r1$ 分别表示左侧不选/选（这里，为了避免额外的取 max 操作，可以认为 1 表示选/不选均可）mid 的最大独立集，右侧不选/选 $mid+1$ 的最大独立集，则

$$ans(i,j) = \max(l1_i + r0_j, l0_i + r1_j)$$

分类讨论，分离变量。

分治

最大独立集 P7482

给定一个序列，求所有区间的最大独立集之和。

$$n \le 10^5$$

$$ans(i,j) = \max(l1_i + r0_j, l0_i + r1_j)$$

枚举 i。

若 $l1_i + r0_j \ge l0_i + r1_j$，就是 $l1_i - l0_i \ge r1_j - r0_j$。要求所有这样的 j 的 $l1_i + r0_j$ 之和，也就是分别是个数、$r0$ 之和。提前把 j 按照 $r1 - r0$ 排序，二分出分界点，维护前缀和即可；另一侧同理。

时间复杂度 $O(n \log^2 n)$。

分治

小结

前面的例题已经向我们展示了分治解决区间计数问题的范式：

1. 转为两个信息合并。
2. 对合并方式分类讨论，再分离变量，分别计算。

这里，不要惧怕分类讨论繁复，毕竟“讨论”之难，怎样也不如“思维”之难。是谓：

胆大心细不畏繁，勇刃难题换笑颜。

更多练习题：CF1156E

偏序问题

偏序问题的定义

偏序问题，是形态如下的问题：

偏序问题的一般形式

k 维空间里有 n 个点，第 i 个点的坐标为 $\left(x_{i,j}\right)$ 其中 $1 \le j \le k$。

有 q 次询问，每次询问包含一个点 (y_j)，求出有多少个给出的点满足 $\forall 1 \le j \le k, x_{i,j} \le y_j$。

有时，上述问题也简称“k 维数点”。

偏序问题

偏序问题的定义

有时题目要求的不是点数，而是点的权值和/异或和或更为复杂的信息合并，维护方法没有本质区别。

需要注意，偏序问题

1. 是“数点”，不是数“点对”或其他。
2. 只能处理“每维分别小于（当然小于等于也可以）”的限制。

解决偏序问题的通法叫做 CDQ 分治，可以在 $k-1$ 个 log 的时间复杂度内解决 k 维偏序。

偏序问题

偏序问题的例子

在省选及以前的阶段，OI 中的数据结构问题有很大一部分都用到了偏序问题的思想，所以学习它是非常有用的。

1. 逆序对计数
   把 (i, a_i) 看成点，则每个 i 对应的 j 的数量是二维数点。
   注意：不是“逆序对”是二维数点，而是对于一个 i 数 j 是二维数点。
2. 矩形求和
   给出二维平面上的一些点，每次询问一个横平竖直的矩形，求出矩形内有几个点。
   拆分成前缀的和差。（由此也看出，如果不是前缀问题，则需要信息有可减性）

偏序问题

区间数颜色 P4113

给出一个序列，多次询问区间中有几种不同的 a_i。

$$n, q \leq 10^6$$

带修数点 经典问题

平面上初始没有点，你需要支持动态加点、删点，同时求矩形内点数。

$$n, q \leq 10^5$$

三角形求和 经典问题

给定平面上 n 个点，每次询问给出 u, v, w，求满足 $x + y \leq u, x \geq v, y \geq w$ 的点的个数。

$$n, q \leq 10^6$$

偏序问题

小结

前几个偏序问题的应用中，有几个需要注意的地方：

1. 对于“出现几次”或类似的问题，考虑 lst。
2. 带修问题就是加时间维。
3. 并不是把问题转化为偏序问题就万事大吉了，有时候如果发现的性质太少，偏序问题维数过高，反而可能阻碍问题的继续思考。

偏序问题

扫描线

二维偏序是最简单的偏序问题，通常采用扫描线来解决。

想象一根竖着的数轴扫过平面，在扫到 $x = i$ 时，数轴的 j 位置就存储了 (i, j) 位置的信息。

放在区间上，扫描线可以理解成：枚举 r，同时维护一个数据结构，在扫到 $r = r_0$ 时，数据结构的 l 位置存储 $[l, r_0]$ 的区间信息。

扫描线解决二维偏序的方法是：按 x 从小到大考虑点和询问，用数据结构维护 y 一维。可以结合平面上矩形数点来理解。

对于一般问题，x, y 不一定是对称的！有时，需要换维度扫描线。

偏序问题

矩形面积并 经典问题

平面上有 n 个矩形，求其面积并。

$$n \le 10^5, x_i, y_i \le 10^9$$

首先离散化。注意：此处离散化需要保留权值。
对 x 一维扫描线，对 y 一维维护线段树。在扫到 $x = k$（实际上对应了离散化数组里的 $v_k \sim v_{k+1}$ 这一段）时，希望在线段树上 u 位置处维护出 (k, u) 的覆盖情况。
维护方法是：对于矩形 (x_0, y_0, x_1, y_1)，在 x_0 处给 $[y_0, y_1]$ 加一（注意离散化后是个左闭右开区间），在 $x_1 + 1$ 处减一。询问即为 ≥ 1 的位置个数。
≥ 1 的位置个数可以转化为 $= 0$ 的位置个数，而这又可以转为最小值个数。

偏序问题

连续段问题 CF526F

给定一个排列，问有几个区间满足 $r-l=\max\limits_{i\in[l,r]} a_i - \min\limits_{i\in[l,r]} a_i$。

$$n \le 3\times 10^5$$

希望在 r 处时，线段树上 l 位置上恰好等于 $\max\limits_{i\in[l,r]} a_i - \min\limits_{i\in[l,r]} a_i - (r-l)$。

从 $\max\limits_{i\in[l,r]} a_i$ 到 $\max\limits_{i\in[l,r+1]} a_i$ 的变化不就是单调栈所求吗！在弹栈的时候顺便执行一下线段树区间加（区间为 $(st_{top-1}, st_{top}]$）操作就行了。

同理，$\max\limits_{i\in[l,r]} a_i - \min\limits_{i\in[l,r]} a_i - (r-l)$ 这个整体也是好维护的，但如何计数？

偏序问题

连续段问题 CF526F

给定一个排列，问有几个区间满足 $r-l=\max\limits_{i\in[l,r]} a_i - \min\limits_{i\in[l,r]} a_i$。

$$n \le 3\times 10^5$$

提示：“区间 0 的个数”难以维护，但回想之前所学，什么值的个数容易维护？类比“矩形面积并”问题。

注意到 $\max\limits_{i\in[l,r]} a_i - \min\limits_{i\in[l,r]} a_i - (r-l) \ge 0$，因此可以维护区间最小值个数。

偏序问题

CDQ 分治

CDQ 分治的目的是，将 k 维偏序问题转化为 $k-1$ 维偏序问题。当 $k=2$ 时，就可以用扫描线解决了。

对第一维进行类似序列分治的分治。不妨假设询问的是 $x_i < x, \ldots$ 的点数，则点 x_i 对于询问 x 的贡献在 $[x_i, x]$ 这个区间跨过 mid 时计算。

可以总结为：CDQ 分治特指考虑左侧对右侧的影响的分治。

在偏序问题上，分治第一维后，第一维的限制不复存在，也就把问题转为了总大小乘了 log 的 $k-1$ 维偏序问题。

偏序问题

讨论

CDQ 分治思想简单，但当维数 ≥ 4 时就又难写又慢了。各位在准备写高维偏序代码前，一定要多想能不能多发现一些能够降维的性质。（例如：若 a 有单调性，则 $i \leq p$ 和 $a_i \leq q$ 是同一维限制！）

如果信息不可减，偏序问题什么时候仍然可做？

至多只有一维不是前缀/后缀限制

偏序问题

例子

如何正确地排序 P8253

有一个 $4 \times n$ 的数组 $a_{i,j}$，定义

$$f(i,j) = \min_k\left(a_{k,i} + a_{k,j}\right) + \max_k\left(a_{k,i} + a_{k,j}\right)$$

求 $\sum_{i,j\in[1,n]} f(i,j)$。

$n \leq 2 \times 10^5$

分开处理 min 和 max。

讨论 min 和 max 分别在哪里取到，再分离变量，就是三维偏序问题。
需要注意相等时不要算重：可以认为 k 更小的值更小。

更多练习题：P4169, P3157, P2487

偏序问题

小结

我们刚刚学习了偏序问题的解法，特别地，学习了扫描线解决问题的思路。其中有以下要点值得注意：

1. 将“求 0 的个数”转为求“区间最小值个数”。
2. 对于有 max min 的式子，分类讨论在哪里取到再分离变量是常见处理方法。

综合运用

Beautiful Pair P4755

给定一个数组，问有多少个 (i,j) 满足 $a_i a_j \le \max\limits_{k\in[i,j]} a_k$。

$$n \le 10^5$$

提示：本题或许可以使用序列分治做出，但请思考一下笛卡尔树上的解法。

枚举 $\max a_k$ 取在哪（笛卡尔树上 LCA），则 a_i, a_j 分属两个子树。

枚举其中较小的一侧，由启发式合并的复杂度证明可以知道总枚举量为 $O(n\log n)$。

综合运用

天天爱打卡 P9871

小 T 跑步打卡。共有 n 天，每天可以跑步或不跑步。能量值初始为 0，若某天选择跑步，则能量值减少 d。不能连续超过 k 天跑步。

有 m 条奖励：若他在第 $l_i \sim r_i$ 天都选择跑步，会得到 v_i 能量值。

问 n 天后能量值最高是多少。

$$n \leq 10^9, m \leq 10^5$$

先对 l, r 离散化。思考：具体如何离散化？

综合运用

例子

天天爱打卡 P9871

选择跑步，则能量值减少 d。不能连续超过 k 天跑步。
有 m 条奖励：若他在第 $l_i \sim r_i$ 天都选择跑步，会得到 v_i 能量值。

以 l_i, r_i 划分数轴（均为闭区间），则跑步一定是一段一段跑。

设 dp_i 表示第 $1 \sim i$ 段的最优决策下能量值最大值，如何转移？

提示：枚举最后一段跑步的连续段。

若不跑，$dp_{i-1} \to dp_i$。
若跑，枚举 $j \le i$，$dp_{j-2} - d\left(s_i - s_{j-1}\right) + V(j,i) \to dp_i$。这里要求 $s_i - s_{j-1} \le k$，可以二分求出分界点。

综合运用

天天爱打卡 P9871

选择跑步，则能量值减少 d。不能连续超过 k 天跑步。
有 m 条奖励：若他在第 $l_i \sim r_i$ 天都选择跑步，会得到 v_i 能量值。

若不跑，$dp_{i-1} \to dp_i$。
若跑，枚举 $j \le i$，$dp_{j-2} - d(s_i - s_{j-1}) + V(j, i) \to dp_i$。

用扫描线的思想，在 r_i 处给 $j \le l_i$ 的 j 加上 v_i 的贡献，则上述转移就是区间最大值。

时间复杂度 $O(m \log m)$。
更多类似题目：P2605, CF1889C2

综合运用

Drying Plan CF1889C2

有 m 条线段，值域为 $[1,n]$。请删除不超过 K 条线段，最大化未被任何线段覆盖的整点数量。

$$n,m \le 10^5, K \le 10$$

提示：dp 题，最难的是状态设计，需要搞清楚

1. dp 的“对象”是什么？也即，本题中，$f(i)$ 是表示第 i 条线段还是第 i 个点相关的值？
2. 如何保证 dp 没有后效性？也即，只用前 i 个对象的信息，就能唯一确定 $f(i)$。
3. 如何处理 K 的限制？每条线段在哪里被算？左右端点有无影响？

综合运用

Drying Plan CF1889C2

有 m 条线段，值域为 $[1,n]$。请删除不超过 K 条线段，最大化未被任何线段覆盖的整点数量。

$$n,m \le 10^5, K \le 10$$

对 未被覆盖的整点序列 dp。

设 $f(i,k)$ 表示 i 本身未被任何线段覆盖，只考虑左端点 $\le i$ 的线段的话需要删去 k 条，此时未被覆盖的整点数量最大值。

转移需要找上一个未被覆盖的整点 j，k 减去 $cost(j,i)$，其中 $cost(j,i)$ 表示左端点 $\in (j,i]$ 且右端点 $\ge i$ 的线段数。由于 $cost$ 随 j 减少单增，故可以分段转移。转移过程需要求区间最小值，可以 ST 表。

综合运用

小结

综合运用数据结构来解决问题时，以下法则应当牢记于心：

想清楚数据结构中存储的数据的含义。
用什么数据结构最合适，我想求的答案满不满足我想用的数据结构的要求？
什么时间修改，什么时间询问？
下标与存储内容怎么对应？