

---

## 网易 2020 校招笔试- 大数据开发工程师（正式批）

---

### 1.[单选题]

hive 的元数据是用来存储表的名字，表的列和分区及其属性，一般不会把它存在？

- A. Mysql
- B. Derby
- C. HDFS
- D. 其余选项都可以

### 2.[单选题]

hive 上查看 sql 执行计划的命令是什么？

- A. show
- B. explain
- C. describe
- D. list

### 3.[单选题]

数据结构中，沿着某条搜索路线，依次对树中每个结点均做一次且仅做一次访问。对二叉树的结点从 1 开始进行连续编号，要求每个结点的编号大于其左、右孩子的编号，同一结点的左右孩子中，其左孩子的编号小于其右孩子的编号，可采用（ ）次序的遍历实现编号。

- A. 先序
- B. 中序
- C. 后序
- D. 从根开始按层次遍历

### 4.[单选题]

现在假设对  $N$  个元素的链表做顺序查找时，若查找每个元素的概率相同，则平均查找长度为（ ）？

- A.  $(N+1)/2$
- B.  $N/2$
- C.  $N$
- D.  $[(1+N)*N] / 2$

### 5.[单选题]

以下哪个 Spark 函数不属于转换

- A. map

- B. flatMap
- C. reduce
- D. reduceByKey

## 6.[单选题]

假设有选课表 `course_relation(student_id, course_id)`，其中 `student_id` 表示学号，`course_id` 表示课程编号，如果小易现在想获取每个学生所选课程的个数信息，请问如下的 sql 语句正确的是（ ）

- A. `select student_id, sum(course_id) from course_relation;`
- B. `select student_id, sum(course_id) from lcourse_relation group by student_id;`
- C. `select student_id, count(course_id) from course_relation;`
- D. `select student_id, count(course_id) from course_relation group by student_id;`

## 7.[单选题]

关于 Linux 下面说法正确的是？

- A. 可以使用一般身份用户执行 `ifconfig eth0` 指令
- B. 可以将“.”加入到 `PATH` 的查询目录中
- C. `init` 是 linux 启动的第一个进程
- D. linux 操作系统可以对目录进行硬链接

## 8.[单选题]

`(umask 022;touch a)` 等价于 `(touch a; chmod XXX)` ,请问 XXX 用哪个数字表示

- A. 644
- B. 755
- C. 022
- D. 700

## 9. [单选题]

位图这样的数据结构在操作系统当中的哪个场景下使用？

- A. 文件目录的查找
- B. 磁盘空间的管理
- C. 主存空间的共享
- D. 文件的保护和保密

## 10.[单选题]

小易有个 32G 字节的文件，需要从电脑 1 传送到电脑 2，假设两者之间是 1024Mbit/s 的网络，那么需要多久能传输完

- A. 32 秒
- B. 128 秒

- C. 256 秒
- D. 512 秒

## 11.[问答题]

题目描述

简要说明列式存储数据库的架构特点，主要优势和适用场景。

## 12.[问答题]

题目描述

流式计算中存在数据丢失和重复消费的风险。从实现原理层面说明一个流式计算系统如何实现数据消息的最多消费一次，最少消费一次，甚至是精确消费一次？

## 13.[编程题]翻倍

时间限制：C/C++ 2 秒，其他语言 4 秒

空间限制：C/C++ 256M，其他语言 512M

小易给你数字  $A, B (A < B)$  和系数  $p, q$ 。每次操作你可以将  $A$  变成  $A + p$  或者将  $p$  变成  $p \times q$ 。问至少几次操作使得  $B \leq A$ 。

输入描述:

第一行数据组数  $T$ ，对于每组数据，一行四个整数  $A, B, p, q$ 。  
 $1 \leq A, p, B \leq 10^9, 2 \leq q \leq 10, 1 \leq T \leq 5$ 。

输出描述:

对于每组数据，输出一个数字表示答案

输入例子 1:

2  
1 5 7 2  
3 5 1 2

输出例子 1:

1  
2

输入例子 2:

2  
1 15 4 2

---

12 19 3 2

输出例子 2:

3

3

## 14. [编程题]跳柱子

时间限制：C/C++ 2 秒，其他语言 4 秒

空间限制：C/C++ 256M，其他语言 512M

小易有  $n$  根柱子，第  $i$  根柱子的高度为  $h_i$ 。一开始小易站在第一根柱子上。小易能从第  $i$  根柱子跳到第  $j$  根柱子，当且仅当  $h_j \leq h_i$  且  $1 \leq j - i \leq k$ 。其中  $k$  为指定的一个数字。

另外小易拥有一次释放超能力的机会。这个超能力能让小易从柱子  $i$  跳到任意满足  $1 \leq j - i \leq k$  的柱子  $j$  而无视柱子高度的限制。现在小易想知道，小易是否能到达第  $n$  根柱子。

输入描述:

第一行数据组数  $T$

对于每组数据，第一行数字  $n, k$ ，接下来一行  $n$  个数字表示  $h_i$ 。

$1 \leq n \leq 1000, 1 \leq h_i \leq 10^9, 1 \leq T \leq 10, 1 \leq k \leq n$

输出描述:

对于每组数据，输出 YES 或 NO

输入例子 1:

1

5 3

6 2 4 3 8

输出例子 1:

YES

输入例子 2:

1

5 2

1 8 2 3 4

输出例子 2:

NO

## 15.[编程题]人数统计

时间限制：C/C++ 2 秒，其他语言 4 秒

空间限制：C/C++ 256M，其他语言 512M

小易的公司一共有 $n$ 名员工, 第 $i$ 个人每个月的薪酬是 $x_i$ 万元。

现在小易的老板向小易提了 $m$ 次询问, 每次询问老板都会给出一个整数 $k$ , 小易要快速回答老板工资等于 $k$ 的员工数量。

输入描述:

第一行, 两个空格间隔的整数 $n$ 和 $m$ , 表示人数和提问的次数

第二行,  $n$ 个用空格间隔的整数 $x_i$ , 表示每名员工的薪酬

接下来有 $m$ 行, 每行一个整数, 表示老板的一次提问。

$1 \leq m \leq 80000, 1 \leq n \leq 100000, 1 \leq x_i \leq 500,000,000$

输出描述:

$m$  行, 每行一个整数, 表示对应提问的答案

输入例子 1:

7 4  
6 2 1 2 6 2 5  
6  
5  
8  
2

输出例子 1:

2  
1  
0  
3

## 16.[编程题]积木

时间限制：C/C++ 2 秒，其他语言 4 秒

空间限制：C/C++ 256M，其他语言 512M

小易有 $n$ 堆积木, 第 $i$ 堆积木有 $h_i$ 块。小易还拥有容量无限的背包。

一开始小易站在第一堆积木旁边。每次小易可以选择进行下列三种操作中的一种:

- 1、从背包里掏出一块积木 (如果有的话) 放到当前这一堆里
- 2、从当前这一堆积木里掏出一块塞到背包里 (如果当前积木堆不为空的话)
- 3、从当前这一堆走到下一堆。

一开始小易的背包里有 $m$ 块积木。小易希望把这些个积木变成严格递增的 (即 $h_1 < h_2 < h_3 \cdots < h_n$ )。小易希望知道这是否有可能完成。(所有操作结束后不需要保证背包里没有积木了, 可以有积木堆为空)。

输入描述:

---

第一行数据组数 $T$

对于每组数据，第一行数字 $n, m$ ，接下来一行 $n$ 个数字表示 $h_i$ 。

$1 \leq n \leq 100000, 0 \leq h_i \leq 10^9, 1 \leq T \leq 10, 0 \leq m \leq 10^9$

输出描述:

对于每组数据输出一行，输出结果 YES 或 NO

输入例子 1:

1  
5 3  
2 2 3 3 1

输出例子 1:

YES

输入例子 2:

1  
5 2  
0 0 1 2 1

输出例子 2:

NO

附选择题正确答案

1. C
2. B
3. C
4. A
5. C
6. D
7. C
8. A
9. B
10. C