

---

## 网易 2020 校招笔试- 数据分析师（提前批）

---

### 1.[单选题]

有一类二叉树用三叉链表来存储的时候除了带有指向左右孩子节点的两个指针, 还有指向父节点的指针, 那么这样一棵二叉树有 2 个节点, 那么有多少指针指向 NULL (注: 根节点的父指针指向 NULL, 对于不存在的节点表示为 NULL) ?

- A. 1
- B. 2
- C. 3
- D. 4
- E. 5

### 2.[单选题]

下列最短路径算法的叙述中正确的是 ( )

- A. Dijkstra 算法通常用于求每一对顶点间的最短路径;
- B. Dijkstra 算法不允许图中带有负权值的边, 而 Floyd 算法则可以适用;
- C. Floyd 算法通常用于求某一顶点到其他各顶点的最短路径;
- D. Floyd 算法允许有包含负权值的边组成的回路, 而 Dijkstra 算法不允许;

### 3.[单选题]

一个盒子中有三个大小相同的球, 这三个球可能是红和蓝两种颜色, 并且一个球是红的还是蓝的是等可能的。已知其中有一个是红色的, 那么至少有一个球是蓝色的概率是多少 ( )

- A.  $7/8$
- B.  $3/4$
- C.  $6/7$
- D.  $1/3$

### 4.[单选题]

在 10 件产品中有五件是残次品, 从中任取五件, 求其中至少有两件是残次品的概率( )

- A.  $4/5$
- B.  $113/126$
- C.  $105/126$
- D.  $1/12$

### 5.[单选题]

小明在玩掷骰子的游戏, 将一颗骰子掷了两次, 两次骰子的点数之和为 7, 求其中一次掷到

6 点的概率 ( )

- A. 1/6
- B. 1/18
- C. 1/12
- D. 1/3

### 6.[单选题]

在 5 张卡片上按顺序写上 laval 这五个字母，并依次放入 5 个盒中，有人从中任意取出两张卡片使用，但是在放回时，忘记了两张卡片各自的位置，求此人将卡片随意放回两个空盒子后卡片顺序仍为 laval 的概率 ( )

- A. 3/5
- B. 1/2
- C. 1/5
- D. 2/5

### 7.[单选题]

以下是泊松分布概率密度函数的是

A.

$$\binom{n}{k} p^k (1-p)^{n-k}, \quad k=0, 1, 2$$

B.

$$\frac{\lambda^k e^{-\lambda}}{k!}, \quad k=0, 1, 2, \dots$$

C.

$$\frac{1}{\theta} e^{-x/\theta} \quad x > 0$$

D.

$$1/(b-a), \quad x \in [a, b]$$

### 8.[单选题]

---

判断一个数组或序列是正序,倒序还是乱序,需要我们将这个数组完整的遍历一遍通过构建有序序列,对于未排序数据,在已排序序列中从后向前扫描,找到相应的位置并插入的排序算法是 ( )

- A. 选择排序
- B. 希尔排序
- C. 插入排序
- D. 归并排序

## 9.[单选题]

队列是一种特殊的线性表,特殊之处在于它只允许在表的前端(front)进行删除操作,而在表的后端(rear)进行插入操作,和栈一样,队列是一种操作受限制的线性表,进行插入操作的端称为队尾,进行删除操作的端称为队头,若用一个大小为6的数组来实现循环队列,数组下标为[0,5],且当前rear和front的值分别为0和3,当从队列中删除一个元素,再加入两个元素后,rear和front的值分别为多少? ( )

- A. 1 和 5
- B. 2 和 4
- C. 4 和 2
- D. 5 和 1

## 10.[单选题]

访问记录 visit 表中包含日期 date,用户 user\_id,访问的页面 url 3 个字段,以下哪个选项不能计算'2019-03-11'这一天访问过页面的所有用户数

- A. `SELECT count(user_id) FROM (SELECT user_id FROM visit WHERE date = '2019-03-11' GROUP BY user_id) f`
- B. `SELECT count(user_id) FROM (SELECT DISTINCT user_id FROM visit WHERE date = '2019-03-11') f`
- C. `SELECT count(DISTINCT user_id) FROM visit WHERE date = '2019-03-11'`
- D. `SELECT count(user_id) FROM visit WHERE date = '2019-03-11' GROUP BY date`

## 11.[问答题]

题目描述

用户分析是电商数据分析中重要的模块,在对用户特征深度理解和用户需求充分挖掘基础上,进行全生命周期的运营管理(拉新—>活跃—>留存—>价值提升—>忠诚),请尝试回答以下3个问题:

- ① 用户第一单购买的行为往往反映了用户对平台的信任度和消费能力。现在数据库中有一张用户交易表 order,其中有userid(用户ID)、amount(消费金额)、paytime(支付时间),请写出对应的SQL语句,查出每个用户第一单的消费金额。

② 当你发现本月的支付用户数环比上月大幅下跌（超 30%），你会如何去探查背后的原因？请描述你的思路和其中涉及的关键指标

③ 为了更好的理解用户，我们通常会基于用户的特征对用户进行分类，便于更加精细化的理解用户，设计产品和运营玩法，请你设计对应的聚类方法，包括重点的用户特征的选择及聚类算法并说明其基本原理和步骤

## 12.[问答题]

题目描述

网易严选是网易旗下原创生活类自营电商品牌，深度贯彻“好的生活，没那么贵”的品牌理念。商品覆盖居家、餐厨、配件、服装、洗护、母婴、原生态饮食等几大类目，兼具品质和性价比，得到用户的广泛好评。若你是网易严选负责商品的数据分析师，当面对以下业务问题时，你会如何解决？

① 用户增长团队期望选择一批合适的商品用于吸引新客，期望你帮助从数据的角度筛选出一批合适的商品，你会如何帮助他们进行筛选？请描述你的思路。

② 商品研发负责人期望能有一套指标帮助衡量开发的商品表现，请你帮助设计对应的评估方案，包括设计思路、涉及的数据指标等。

## 13.[编程题]序列维护

时间限制：C/C++ 3 秒，其他语言 6 秒

空间限制：C/C++ 256M，其他语言 512M

小易在维护数据的时候遇到一个需求，具体来说小易有一系列数据，这些数据构成了一个长度为  $n$  的数字序列，接下来小易会在这个序列上进行  $q$  次操作。

每次操作有一个查询的数字  $x$ ，小易需要将序列数据中所有大于等于  $x$  的数字都减一，并输出在本次操作中有多少个数字被减一了。

小易犯了难，希望你能帮帮他。

输入描述：

第一行  $n, q$ ，表示数字个数和操作个数。

接下来一行  $n$  个数表示初始的数字。

接下来  $q$  行，每行一个数，表示指定的数字  $x$ 。

$$1 \leq n, q \leq 200000, 1 \leq a_i, x \leq n$$

输出描述：

对于每个询问，输出一个数字表示答案

---

输入例子 1:

4 3  
1 2 3 4  
4  
3  
1

输出例子 1:

1  
2  
4

输入例子 2:

3 2  
1 2 3  
3  
3

输出例子 2:

1  
0

## 14.[编程题]最大公约数

时间限制：C/C++ 1 秒，其他语言 2 秒

空间限制：C/C++ 256M，其他语言 512M

小易学习了辗转相除法之后，就开始实践这个算法在求解最大公约数上。

牛牛给小易出了一道不同寻常的求解最大公约数：求解  $a$  和  $b$  的最大公约数，但是  $a$  和  $b$  的范围特别大。

小易遇到了困难，向聪明的你寻求帮助，希望你能帮帮他。

输入描述：

第一行数字  $a$ ，第二行数字  $b$ 。

$$1 \leq a \leq 10^{10^5}, 1 \leq b \leq 10^{18}$$

输出描述：

一行一个数字表示答案

输入例子 1:

---

6

4

输出例子 1:

2

输入例子 2:

7951346523609888

6998915114363550

输出例子 2:

1013754

## 15.[编程题]按位或

时间限制：C/C++ 1 秒，其他语言 2 秒

空间限制：C/C++ 256M，其他语言 512M

小易有一个初始为空的数字集合，支持两种操作：

1、加入数字  $x$  到集合中。

2、询问集合中是否存在一个子集，满足子集中所有数字的 Or 值恰好为  $k$ 。

Or 为二进制按位或操作，C++中表示为" $|$ "。

小易希望你能解决这个问题。

输入描述：

第一行数字  $q$ ,表示操作个数

接下来  $q$  行，每行两个数字：

1  $x$  表示插入数字  $x$

2  $x$  表示询问数字  $x$ (即题设中询问的数值  $k$ )

$$1 \leq q \leq 100000, 1 \leq x \leq 100000$$

输出描述：

对于每个询问，输出"YES"或者"NO"表示是否存在。

输入例子 1:

9

1 4

2 5

1 9

1 15

2 4

1 11

2 10

---

2 7

2 9

输出例子 1:

NO

YES

NO

NO

YES

## 16. [编程题]最大最小值

时间限制：C/C++ 1 秒，其他语言 2 秒

空间限制：C/C++ 256M，其他语言 512M

小易给定了一个长度为  $n$  的数字序列，对于每一个  $1 \leq k \leq n$ ，小易希望能求解出所有长度为  $k$  的连续子序列的最大值中的最小值。

输入描述:

第一行数字  $n$

接下来一行是一个长度为  $n$  的数字序列

$$1 \leq n \leq 100000, 0 \leq a_i \leq 10^9$$

输出描述:

一行  $n$  个数字，第  $i$  个数字表示  $k = i$  时的答案。

输入例子 1:

6

1 3 2 4 6 5

输出例子 1:

1 3 3 4 6 6

例子说明 1:

当  $k = 2$  的时候

子序列分别是:

1 3 最大值为 3

3 2 最大值为 3

2 4 最大值为 4

4 6 最大值为 6

6 5 最大值为 6

---

所有最大值中的最小值为 3

附选择题正确答案

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