

You submitted this quiz on **Mon 16 Feb 2015 5:29 PM CST**. You got a score of **3.00** out of **3.00**.

Question 1

Suppose we have transactions that satisfy the following assumptions:

- s , the support threshold, is 10,000.
- There are one million items, which are represented by the integers $0, 1, \dots, 999999$.
- There are N frequent items, that is, items that occur 10,000 times or more.
- There are one million pairs that occur 10,000 times or more.
- There are $2M$ pairs that occur exactly once. M of these pairs consist of two frequent items, the other M each have at least one nonfrequent item.
- No other pairs occur at all.
- Integers are always represented by 4 bytes.

Suppose we run the a-priori algorithm to find frequent pairs and can choose on the second pass between the triangular-matrix method for counting candidate pairs (a triangular array $\text{count}[i][j]$ that holds an integer count for each pair of items (i, j) where $i < j$) and a hash table of item-item-count triples. Neglect in the first case the space needed to translate between original item numbers and numbers for the frequent items, and in the second case neglect the space needed for the hash table. Assume that item numbers and counts are always 4-byte integers.

As a function of N and M , what is the minimum number of bytes of main memory needed to execute the a-priori algorithm on this data? Demonstrate that you have the correct formula by selecting, from the choices below, the triple consisting of values for N , M , and the (approximate, i.e., to within 10%) minimum number of bytes of main memory, S , needed for the a-priori algorithm to execute with this data.

Your Answer	Score	Explanation
<input type="radio"/> $N = 40,000$; $M = 60,000,000$; $S = 3,200,000,000$		
<input checked="" type="radio"/> $N = 100,000$; $M = 100,000,000$; $S = 1,200,000,000$	✓ 1.00	
<input type="radio"/> $N = 20,000$; $M = 60,000,000$; $S = 1,000,000,000$		
<input type="radio"/> $N = 50,000$; $M = 80,000,000$; $S = 1,500,000,000$		
Total	1.00 / 1.00	

Question 2

Imagine there are 100 baskets, numbered 1,2,...,100, and 100 items, similarly numbered. Item i is in basket j if and only if i divides j evenly. For example, basket 24 is the set of items $\{1,2,3,4,6,8,12,24\}$. Describe all the association rules that have 100% confidence. Which of the following rules has 100% confidence?

Your Answer	Score	Explanation
<input type="radio"/> $\{2,3,5\} \rightarrow 45$		
<input type="radio"/> $\{4,6\} \rightarrow 24$		
<input checked="" type="radio"/> $\{3,5\} \rightarrow 1$	✓ 1.00	
<input type="radio"/> $\{8,12\} \rightarrow 96$		
Total	1.00 / 1.00	

Question 3

Suppose ABC is a frequent itemset and BCDE is NOT a frequent itemset. Given this information, we can be sure that certain other itemsets are frequent and sure that certain itemsets are NOT frequent. Other itemsets may be either frequent or not. Which of the following is a correct classification of an itemset?

Your Answer	Score	Explanation
<input type="radio"/> ABCDE can be either frequent or not frequent.		
<input type="radio"/> ABCDEF can be either frequent or not frequent.		
<input type="radio"/> BCE is frequent.		
<input checked="" type="radio"/> BC is frequent.	✓ 1.00	
Total	1.00 / 1.00	