

You submitted this quiz on **Tue 17 Feb 2015 9:12 PM CST**. You got a score of **2.00** out of **2.00**.

Question 1

Suppose we perform the PCY algorithm to find frequent pairs, with market-basket data meeting the following specifications:

- s , the support threshold, is 10,000.
- There are one million items, which are represented by the integers 0,1,...,999999.
- There are 250,000 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 P pairs that occur exactly once and consist of 2 frequent items.
- No other pairs occur at all.
- Integers are always represented by 4 bytes.
- When we hash pairs, they distribute among buckets randomly, but as evenly as possible; i.e., you may assume that each bucket gets exactly its fair share of the P pairs that occur once.

Suppose there are S bytes of main memory. In order to run the PCY algorithm successfully, the number of buckets must be sufficiently large that most buckets are not large. In addition, on the second pass, there must be enough room to count all the candidate pairs. As a function of S , what is the largest value of P for which we can successfully run the PCY algorithm on this data? Demonstrate that you have the correct formula by indicating which of the following is a value for S and a value for P that is approximately (i.e., to within 10%) the largest possible value of P for that S .

Your Answer	Score	Explanation
<input type="radio"/> $S = 100,000,000$; $P = 540,000,000$		
<input type="radio"/> $S = 200,000,000$; $P = 1,600,000,000$		
<input checked="" type="radio"/> $S = 100,000,000$; $P = 200,000,000$	✓ 1.00	
<input type="radio"/> $S = 1,000,000,000$; $P = 10,000,000,000$		
Total	1.00 / 1.00	

Question 2

During a run of Toivonen's Algorithm with set of items {A,B,C,D,E,F,G,H} a sample is found to have the following maximal frequent itemsets: {A,B}, {A,C}, {A,D}, {B,C}, {E}, {F}. Compute the negative border. Then, identify in the list below the set that is NOT in the negative border.

Your Answer	Score	Explanation
<input checked="" type="radio"/> {A,B,D}	✓ 1.00	Correct! This set is not in the negative border because immediate proper subset {B,D} is not frequent.
<input type="radio"/> {B,D}		
<input type="radio"/> {E,F}		
<input type="radio"/> {B,E}		
Total	1.00 / 1.00	