

Question 1

- 20000 tuples
- 600 data pages
- Prefix: 20 bytes
- Full: 30 bytes
- 200 different types
- 50 producers
- Rid has 10 bytes
- Pointer has 6 bytes
- Leaf pages are filled about 70%
- Index page has 4000 Bytes

- 1) We have $200 * 50 = 10000$ possible different values. Then as its uniformly distributed we have 10000 data entries

The number of rids per data entry is $\left(\frac{\text{number of tuples}}{\text{diff values}}\right) = \frac{20000}{10000} = 2$

The average length of a data entry is $\text{size of key} + nb(\text{rids} * \text{size}(\text{rids})) = 30 * 2 + 2 * 10 = 80$

- 2) The average number of data entry per intermediate page is $\frac{4000 * 0.7}{46} = 107$ Average nb of data entry per page is $\frac{0.7 * 4000}{98} = 28$

Number of leaf pages is $\frac{10000}{28} = 357$

- $\text{Entry size in leaf pages} = \text{size of key} + \text{size of rid} = 30 + 10 = 40$
- $\text{Size of tuples} = 4 + 30 * 3 + 4 = 98$
- $\text{Index entry size in root and intermediate pages} = \text{size of key} + \text{size of pointer} = 20 * 2 + 6 = 46$
-