

# Indexing Choices and Cost estimation

---

**CS 4750**  
**Database Systems**

# Thinking About Indexing Choices

Assume each disk block holds  $t$  tuples of a relation `customer`. A relation `customer` has 100 tuples; and 10 disk blocks are used to hold tuples of the relation. An attribute `city` indicates a city where a customer lives. The `customer` relation is sorted by city (ascending). There are 80 distinct values of the city in the `customer` relation.

For each of the following data access methods, estimate the I/O cost to answer the query: *Find information of customers who live in Charlottesville*

1. Use **no index** (and thus need to go directly to the data file and do sequential scan)
2. Use **clustered index** (assume search key of a clustered index is based on `city` of the `customer` relation)
3. Use **unclustered index** (assume search key of an unclustered index is based on an attribute that is not `city`)

# Thinking About Indexing Choices (2)

$T(\text{customer}) = 100$       # tuples  
 $B(\text{customer}) = 10$       # blocks  
 $V(\text{customer}, \text{city}) = 80$       # distinct values of attribute city

Estimate the I/O cost to answer the query: *Find information of customers who live in Charlottesville*

1. Use **no index** (and thus need to go directly to the data file and do **sequential** scan)

$$\text{Cost} = \# \text{blocks} = 10 \quad \text{I/Os}$$

# Thinking About Indexing Choices (3)

$T(\text{customer}) = 100$                    # tuples  
 $B(\text{customer}) = 10$                 # blocks  
 $V(\text{customer}, \text{city}) = 80$        # distinct values of attribute city

Estimate the I/O cost to answer the query: *Find information of customers who live in Charlottesville*

2. Use **clustered index** (assume search key of a clustered index is based on city of the customer relation) – **same order**

Selectivity estimate = proportion of how likely that we will find tuples that satisfy the condition  
 $= 1/V(\text{customer}, \text{city}) = 1/80 = 1.25\%$

Cost = selectivity estimate  $\times$  #blocks  
 $= 1.25\% \times 10$  I/Os

# Thinking About Indexing Choices (4)

$T(\text{customer}) = 100$                    # tuples  
 $B(\text{customer}) = 10$                 # blocks  
 $V(\text{customer}, \text{city}) = 80$        # distinct values of attribute city

Estimate the I/O cost to answer the query: *Find information of customers who live in Charlottesville*

3. Use **unclustered index** (assume search key of an unclustered index is based on an attribute that is not city) – **different (or no) order**

Selectivity estimate = proportion of how likely that we will find tuples that satisfy the condition  
=  $1/V(\text{customer}, \text{city}) = 1/80 = 1.25\%$

Cost = selectivity estimate  $\times$  #tuples  
=  $1.25\% \times 100$  I/Os