

## Control 2 - Indices y algoritmos

El plazo para la entrega es el viernes 14 de octubre, hasta las 23:59. La entrega debe hacerse en un **pdf** y subirse a Canvas (se creará una tarea con el form para subir la entrega).

### Pregunta 1

Considere la relación

`Cervezas(id int PRIMARY KEY, nombre varchar(20), marca varchar(20), estilo varchar(15), precio int)`

indexada con un hash-join clustered sobre el atributo `id`, un hash-join unclustered sobre `marca` y un B+tree unclustered sobre `precio`.

- Indique el costo I/O de las siguientes consultas asumiendo que la relación tiene 10000 tuplas, que el `id` va desde el 0 hasta el 9999, que en una pagina caben 80 tuplas o 250 punteros, que la altura del B+Tree es  $h$ , que sus hojas están ocupadas a un 60 % y que el hash index tiene solo una pagina por bucket.
  - `SELECT * FROM Cervezas WHERE id = 20`
  - `SELECT * FROM Cervezas WHERE id <=3`
  - `SELECT * FROM Cervezas WHERE precio >2000 AND precio <3000` (asuma que existen menos de 80 tuplas que cumplen esta condición)
  - `SELECT * FROM Cervezas WHERE marca = 'Tamango Brebajes'`
  - `SELECT * FROM Cervezas WHERE estilo = 'Hazy IPA'`
- Bajo las mismas condiciones de la pregunta anterior, explique como podría resolver la consulta `SELECT * FROM Cervezas WHERE precio <3000 AND marca = 'Tamango Brebajes'` (haciendo un uso inteligente de los indices) e indique el costo I/O de su propuesta. Asuma que la cantidad de cervezas que valen menos de 3000 caben en una página.

### Pregunta 2

Considere que tiene las relaciones `R(a int PRIMARY KEY, b int)` y `S(b int PRIMARY KEY, c int)`. Escriba en pseudo-código un programa para computar  $R \bowtie S$  tal que recorra  $R$  y  $S$  solo una vez. *Hint: es probable que tenga que pre-procesar la(s) relacion(es) de alguna manera...*

### Pregunta 3

Considere las relaciones `R(a int, b int)` y `S(a int, c int)` e imagine que casi todas las consultas que se hacen implican hacer un join entre  $R$  y  $S$ . Imagine también que  $R$  ya se encuentra indexada según el atributo  $b$ . Responda lo siguiente:

- Explique cómo resolver  $R \bowtie S$  y cuál es su costo en operaciones I/O
- Recomendaría usted incluir algún otro índice en las relaciones? Si la respuesta es sí, indique el tipo de índice y cómo quedaría el costo I/O de la consulta. Si la respuesta es no, indique por qué no.
- ¿Existe alguna manera en la que se puede disminuir el costo I/O de la consulta sin crear nuevos índices? Justifique su respuesta.