

**Base de Datos (75.15 / 75.28 / 95.05)**

Evaluación Integradora - 12 de febrero de 2020

<b>TEMA 20192C4</b>						Padrón: _____
<b>CRT</b>		<b>Proc.</b>		<b>Proc.II</b>		Apellido: _____
<b>NoSQL</b>		<b>Rec.</b>		<b>DW</b>		Nombre: _____
Corrigió:						Cantidad de hojas: _____
<b>Nota:</b>						<input type="checkbox"/> Aprobado <input type="checkbox"/> Insuficiente

**Criterio de aprobación:** El examen está compuesto por 6 ítems, cada uno de los cuales se corrige como B/B-/Reg/Reg-/M. Se aprueba con nota mayor o igual a 4(cuatro), equivalente a desarrollar el 60% del examen correctamente.

1. (*Cálculo Relacional de Tuplas*) Los *Premios Nobel* son entregados todos los años en Suecia a partir del legado dejado por Alfred Nobel. Las siguientes tablas reflejan las distintas categorías de Premios Nobel existentes y sus ganadores históricos.

- **Categorías**(nombre\_categoria, año\_instauración, institución\_otorgante)
- **Ganadores**(categoria, año, apellido, nombre, país)

Considere la siguiente consulta en Álgebra Relacional sobre este esquema de base de datos:

$$(\pi_{\text{pais, nombre\_categoria}}(\text{Ganadores} \bowtie_{\text{categoria=nombre\_categoria}} \text{Categorías})) \div \pi_{\text{nombre\_categoria}}(\text{Categorías})$$

Se pide:

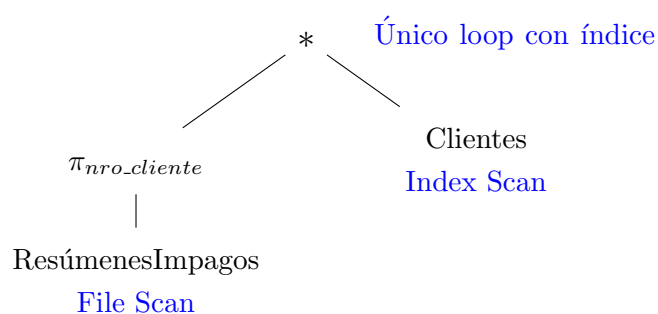
- a) Expresar en lenguaje coloquial el resultado de esta consulta.
- b) Traduzca la consulta a una expresión en Cálculo Relacional de Tuplas.

2. (*Procesamiento de consultas*) El *Banco de las Sierras* mantiene una tabla con todos los resúmenes vencidos y que aún no fueron pagados por los clientes. El gerente de cobranzas desearía contactar a todos aquellos clientes que figuran en este listado con el fin de enviarles un mail recordándoles su estado de morosidad. Considere para ello las siguientes dos tablas:

- **ResúmenesImpagos**(nro\_resumen, fecha\_emision, fecha\_venc, nro\_cliente, monto)
- **Cientes**(nro\_cliente, CUIT, apellido, nombre, domicilio, mail)

Adicionalmente, la tabla **Cientes** dispone de un índice primario por el atributo **nro\_cliente**.

El gerente construye el siguiente plan de ejecución para obtener el listado de clientes morosos:



k	2 <sup>k</sup>
10	1024
11	2048
12	4096
13	8192
14	16384
15	32768
16	65536
17	131072
18	262144
19	524288
20	1048576

Se pide:

- a) Estime el costo del plan de ejecución propuesto, considerando que se dispone de sólo **M=3** bloques de memoria. Para ello, suponga que la proyección se hace a través de sort-externo, almacenando en disco el listado ordenado de todos los **nro\_cliente** (uno por tupla), para luego en una segunda etapa eliminar duplicados.

Nota: Considere que al leer la tabla **ResúmenesImpagos**, la proyección elimina inmediatamente los atributos que no utilizará, antes de iniciar el ordenamiento en sí. Asuma que el atributo **nro\_cliente** ocupa el 10 % de una tupla de **ResúmenesImpagos**.

- b) Estime la cardinalidad del resultado en términos de cantidad de bloques.
- c) (*Bonus track*) Si dispusiera de **M=1000** bloques de memoria, ¿estima que habría una mejora significativa en el costo total? Justifique su respuesta.

Considere la siguiente información de catálogo:

RESUMENESIMPAGOS	CLIENTES
n(ResúmenesImpagos) = 1.000.000	n(Cientes) = 5.000.000
B(ResúmenesImpagos) = 200.000	B(Cientes) = 500.000
V(nro_cliente, ResúmenesImpagos) = 500.000	H(I(nro_cliente, Cientes)) = 4

3. (*Procesamiento de Consultas II*) Mencione las heurísticas de optimización de consultas estudiadas para bases de datos relacionales.

4. (NoSQL) Una aerolínea almacena información sobre sus vuelos diarios y los pasajeros que viajaron en ellos en una base de datos MongoDB. El siguiente documento JSON ejemplifica la estructura almacenada por cada vuelo, que incluye los siguientes campos:

- El código de vuelo.
- La fecha y hora de salida.
- El apellido y nombre del comandante.
- Un vector de strings indicando los códigos de los aeropuertos que conforman el recorrido (2 ó más, según haya paradas intermedias o no).
- La distancia recorrida en vuelo.
- Un vector de documentos embebidos conteniendo el número de documento, apellido, nombre y asiento de cada pasajero que viajó.

```
1  { "cod.vuelo": "AE2840",
2    "fecha.salida": Date("2020-01-18 11:30"),
3    "comandante": "GONZÁLEZ, AURELIANO",
4    "recorrido": [ "EZE", "PAN", "BOS" ],
5    "distancia.km": 830,
6    "pasajeros": [
7      { "nro.dni": 38291440,
8        "apell.nombre": "AUZQUI, BELÉN",
9        "asiento": "36A",
10       },
11      { "nro.dni": 41219085,
12        "apell.nombre": "RIZOU, ADOLFO ORLANDO",
13        "asiento": "28F",
14       }
15    ]
16  }
```

La empresa quisiera generar una colección que almacene, para cada persona que viajó al menos 5000 kilómetros en total, el listado de viajes que realizó y la cantidad de kilómetros totales recorridos, con el objetivo de sortear un premio. Escriba una consulta que permita obtener este resultado, con un formato como el mostrado en el siguiente documento de ejemplo, en que la persona se identifica con su DNI:

```
1  { "_id": {
2    "nro.dni": 38291440,
3    "apell.nombre": "AUZQUI, BELÉN"
4  },
5    "listado.viajes": [
6      {
7        "fecha": Date("2020-01-18 11:30"),
8        "cod.vuelo": "AE2840",
9        "recorrido": [ "EZE", "PAN", "BOS" ]
10      },
11      {
12        "fecha": Date("2020-01-23 18:20"),
13        "cod.vuelo": "AE2931",
14        "recorrido": [ "BOS", "EZE" ]
15      }
16    ],
17    "kilómetros.totales": 7702
18  }
```

Para recordar parte de la sintaxis de *MongoDB* puede ayudarse con la siguiente consulta que encuentra para cada código de aeropuerto el listado de comandantes que alguna vez pasaron por el mismo, indicando el código de aeropuerto y el listado:

---

```
[{ $unwind:    "$recorrido" },
  { $group:    { _id: "$recorrido",
                 listado_comandantes: { $addToSet: "$comandante" } } } ]
```

---

5. (*Recuperación*) Un sistema informático almacena su base de datos en un servidor, utilizando un RAID de dos discos totalmente espejados para la base en sí y un disco SSD para el *log*. El SGBD que gestiona la base utiliza el algoritmo de recuperación UNDO. Cada 48 horas se realiza un backup del disco RAID a un disco en otro edificio.

Considere las siguientes fallas, asumiendo que con cada una de ellas se reinicia el sistema pero que no se produce un daño mayor que el indicado específicamente en el ítem. Para cada falla indique si, al reiniciarse, el sistema *(i) podrá volver a un estado consistente y recuperará todas las transacciones que hicieron commit; (ii) podrá volver a un estado consistente pero podría perder algunas de las transacciones committeadas; (iii) podría no volver a un estado consistente*. Justifique brevemente sus respuestas.

- a) Destrucción de uno de los discos del RAID.
- b) Destrucción del disco SSD.
- c) Caída del sistema por una falla temporal de suministro eléctrico.
- d) Destrucción completa del RAID.
- e) Incendio con destrucción total en el edificio en que se encuentra el servidor.

6. (*Data Warehousing*) Indique si las siguientes afirmaciones sobre *data warehouses* son verdaderas o falsas, justificando su respuesta.

- a) Un cubo de datos OLAP es una matriz multidimensional.
- b) Los procesos ETL (Extract-Transform-Load) tienen como finalidad resolver consultas analíticas sobre el *Data Warehouse*.
- c) Uno de los objetivos de OLAP es procesar grandes volúmenes de transacciones en forma concurrente.
- d) El diagrama de estrella (*star scheme*) utilizado en el modelado conceptual de data warehouses conecta distintas tablas de hechos entre sí a través de una tabla dimensional central.
- e) El estándar SQL posee extensiones vinculadas con OLAP.
- f) La operación de *slicing* realiza una selección algebraica en una de las dimensiones.