

Base de Datos (75.15 / 75.28 / 95.05)

Evaluación Integradora - 24 de agosto de 2021

FECHA 20211C4						Padrón: 100029
Proc.	B-	Proc.	B	CyT	B	Apellido: del Mazo
NoSQL	Reg	NoSQL	B-			Nombre: Federico
Nota: 7 (siete)						<input type="checkbox"/> Aprobado <input type="checkbox"/> Insuficiente

1. (*Procesamiento de Consultas*) La Peluquería Canina PECAS está recibiendo la visita de un destacado peluquero proveniente de Australia, famoso por introducir una nueva técnica en el corte de *Fox Terriers*. Con este motivo quiere contactar telefónicamente a aquellas de personas registradas en la peluquería y que tienen un Fox Terrier en su familia, para invitarlas a aprovechar la ocasión y que el pequeño reciba un corte de pelo único.

Aldo, el empleado de la peluquería, accede a las siguientes dos tablas con información sobre los clientes de la misma y los perros que habitan con cada uno:

- **Mascotas**(nombre, DNI_familiar, raza, fecha_ultimo_corte)
- **Personas**(DNI, apellido, nombre, teléfono)

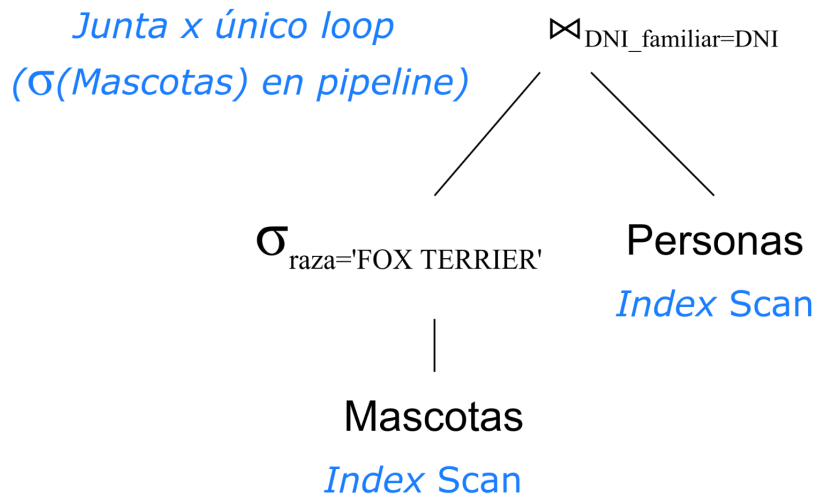
Adicionalmente, se dispone de dos índices para realizar la consulta: un índice de clustering por raza en la tabla *Mascotas*, y un índice primario por DNI en la tabla *Personas*.

A continuación se muestra el plan de ejecución a realizar para obtener los datos requeridos:

Reservorio de caracteres. Cálculo de tuplas: \forall \exists \neq

Álgebra relacional: π_{xxx} , σ_{xxx} , ρ_{xxx} , \leftarrow , x , \cup , $-$, \cap , \bowtie_{xxx} , \div , \wedge , \vee , \neq

Tip: Presionando <CTRL> + ',' puede convertir texto a subíndice.



Se pide:

- Estime el costo de la consulta en términos de cantidad de bloques transferidos desde/hacia disco.
- Estime la cardinalidad del resultado en términos de cantidad de tuplas.

Considere utilizar la siguiente información de catálogo para sus estimaciones:

MASCOTAS	PERSONAS
$n(\text{Mascotas}) = 3000$	$n(\text{Personas}) = 2000$
$B(\text{Mascotas}) = 300$	$B(\text{Personas}) = 200$
$H(I(\text{raza}, \text{Mascotas})) = 2$	$H(I(\text{DNI}, \text{Personas})) = 2$
$V(\text{raza}, \text{Mascotas}) = 50$	

Respuesta:

El index scan por raza

$\Rightarrow H(I) + \text{cota sup de } N/(V \cdot F)$

$\Rightarrow \text{La } F \text{ es } N/B = 3000/300 = 10$

$\Rightarrow V \cdot F = 50 \cdot 10 = 500$

$\Rightarrow 2 + (3000 / 500) = 2 + 6$

La junta

$\Rightarrow B(\text{FoxTerrier}) (\text{pipelineado}) + n(\text{FoxTerrier}) \cdot (H(I(\text{Personas})) + 1)$

$\Rightarrow 0 + (60 \cdot 3)$

Total $\Rightarrow (2 + 6) + (180)$

Reservorio de caracteres. Cálculo de tuplas: $\forall \exists \# \neq$

Álgebra relacional: $\pi_{xxx}, \sigma_{xxx}, \rho_{xxx}, \leftarrow, x, \cup, -, \cap, \bowtie_{xxx}, \div, \wedge, \vee, \neq$

Tip: Presionando <CTRL> + ', puede convertir texto a subíndice.

Cardinalidad de la primera Selección (raza)

$$\implies 3000 / 50 = 60$$

Cardinalidad de la junta por DNI

$$\implies N(\text{FoxTerrier}) * N(\text{personas}) / \max(\text{Var}(\text{raza}), \text{Var}(\text{DNI}))$$

$$\implies 60 * 2000 / \max(50, 2000)$$

$$\implies 60 * 2000 / 2000$$

$$\implies 60$$

2. (*Procesamiento de Consultas*) Se discute oralmente sobre variaciones del Ejercicio 1.
3. (*Concurrencia y Transacciones*) Se discute oralmente sobre mecanismos de control de concurrencia, serializabilidad y recuperabilidad.
4. (*NoSQL*) El restorán vegano "Mr Bison" realizó una encuesta online en el que se pidió a sus clientes que calificaran los platos de su restorán con un valor entre 1 y 5, y además indicaran el barrio donde vivían y cómo conocieron al restorán. Los datos recogidos tienen la siguiente estructura:

```
{
  "_id": ObjectId("407ca8372c83b6b192d12512"),
  "barrio": "Palermo",
  "comoConocio": "Instagram",
  "puntajes": [
    { "plato": "Chorizo de Lechuga",
      "puntuación": 4
    },
    { "plato": "Solomillo de Seitán",
      "puntuación": 2
    },
    { "plato": "Risotto de Hongos",
      "puntuación": 5
    }
  ]
}
{
  "_id": ObjectId("207ca8372c83b6b092d12512"),
  "barrio": "Belgrano",
  "comoConocio": "Vecino",
```

Reservorio de caracteres. Cálculo de tuplas: $\forall \exists \nexists \neq$

Álgebra relacional: $\pi_{xxx}, \sigma_{xxx}, \rho_{xxx}, \leftarrow, x, \cup, -, \cap, \bowtie_{xxx}, \div, \wedge, \vee, \neq$

Tip: Presionando <CTRL> + ', puede convertir texto a subíndice.

```
"puntajes": [
  { "plato": "Risotto de Hongos",
    "puntuación": 4
  },
  { "plato": "Solomillo de Seitán",
    "puntuación": 1
  },
  { "plato": "Panceta de maní",
    "puntuación": 2
  }
]
```

- Escriba una consulta en MongoDB que encuentre la puntuación promedio recibida por cada plato, y la distribución de barrios de los que provienen sus votantes, y luego ordene los resultados en forma descendente por puntuación promedio. El resultado deberá tener la siguiente estructura:

```
{
  "plato": "Solomillo de Seitán",
  "puntuación_promedio": 3.3,
  "barrios_votantes": [
    { barrio: "Belgrano",
      cantidad_votantes: 2 },
    { barrio: "Villa Crespo",
      cantidad_votantes: 4 }
  ]
},
...
```

A continuación le ofrecemos un dataset en caso de que quiera probar su consulta en MongoDB: <https://mongoplayground.net/p/W7eM5b65K7K>.

Ayudas:

- Le recomendamos comenzar por contabilizar la cantidad de votantes de cada plato en cada barrio.
- Puede resultarle útil la función \$push de MongoDB para agregar elementos en un vector.

Respuesta:

```
db.collection.aggregate([
  {
    "$unwind": "$puntajes"
  },
  {
```

Reservorio de caracteres. Cálculo de tuplas: $\forall \exists \neq$

Álgebra relacional: $\pi_{xxx}, \sigma_{xxx}, \rho_{xxx}, \leftarrow, x, \cup, -, \cap, \bowtie_{xxx}, \div, \wedge, \vee, \neq$

Tip: Presionando <CTRL> + ', puede convertir texto a subíndice.

```
"$group": {
  _id: "$puntajes.plato",
  puntuacion_promedio: {
    "$avg": "$puntajes.puntuación"
  },
  barrios: {
    "$push": "$barrio"
  },
},
{
  "$sort": {
    puntuacion_promedio: -1
  }
}
])
```

Cada día, los dueños del restaurante toman un barrio en consideración (por ejemplo, ayer tomaron “La Boca”, y analizan cuántos clientes de ese barrio hay, cuál es la forma en que los clientes de ese barrio conocieron al restaurante. Suponiendo que la colección está shardeada en varios nodos, ¿cómo cree que debería organizarse el sharding para que el procesamiento se pueda paralelizar lo mejor posible?

Respuesta:

Quiero que todos mis nodos laburen en mi consulta:

- Si shardeo por barrio: solo tengo un nodo por barrio -> solo me labura un nodo
- Si shardeo por como nos conocieron: me laburan todos los nodos cuando pido un solo barrio

Tiene más sentido shardear por cómo nos conocieron si eso es lo único que voy a analizar.

Reservorio de caracteres. Cálculo de tuplas: $\forall \exists \neq$

Álgebra relacional: $\pi_{xxx}, \sigma_{xxx}, \rho_{xxx}, \leftarrow, x, \cup, -, \cap, \bowtie_{xxx}, \div, \wedge, \vee, \neq$

Tip: Presionando <CTRL> + ' puede convertir texto a subíndice.