

Parcial Bases de Datos: Consultas y Ecosistema

17 de Noviembre 2022

Nombre: Evan R. P. 2020

Ejercicio 1a		Ejercicio 3a	
Ejercicio 1b		Ejercicio 3b	
Ejercicio 2a		Ejercicio 3c	
Ejercicio 2b		Ejercicio 4	
		Ejercicio 5	

Ejercicio 1: Responder:

a) Sea la consulta que se va a evaluar:

$$\Pi_{\text{name, courseID}} (\sigma_{\text{dept_name} = \text{'Computer Science'}} (\text{instructor} \bowtie \text{teaches}))$$

Asumir que *instructor* y *teaches* tienen como atributo en común *legajo* y están ordenadas por ese atributo. Hay 500 instructores en total y 200 materias (i.e. valores de *courseID*) y *teaches* tiene 800 tuplas. Asumir que todo docente enseña al menos una materia. En *teaches* no hay índice. Se pide:

1. Estimar tamaño de resultado intermedio para la reunión natural en cantidad de tuplas.
2. ¿Qué operador físico conviene usar para la reunión natural? Justificar.
3. ¿Qué operador físico se usa para la selección? Justificar.

b) Considerar la optimización heurística siguiente:

1. Realizar selección tempranamente
2. Realizar proyección tempranamente
3. Hacer selección más restrictiva
4. Hacer operaciones de reunión usando reunión profunda a la izquierda.
5. Ciertas selecciones pueden ser combinadas con producto cartesiano para tornar las operaciones en una reunión (natural o selectiva)

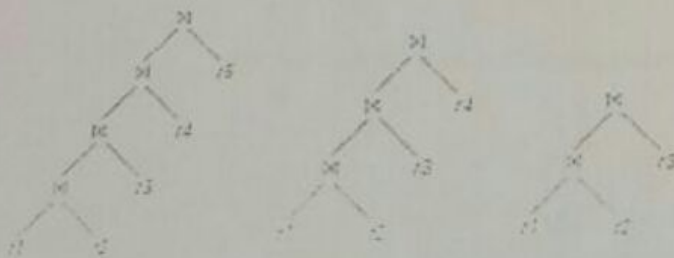
Aplicar esta optimización heurística a la siguiente consulta:

$$\Pi_{\text{legajo, número}} (\sigma_{\text{número} = \text{'optimización de consultas'}} (\text{curso} \bowtie (\text{profe} \bowtie \text{teléfono})))$$

Donde *curso* tiene *nombre* y *legajo* de docente a cargo (y tiene más atributos), *profe* tiene *legajo* como clave primaria (tiene más atributos), *teléfono* tiene *legajo* de su dueño y *número* entre sus atributos (tiene solo esos atributos).

Aplicar paso a paso las reglas de optimización heurística mostrando como la expresión de consulta va a cambiando luego de la aplicación de cada regla. Aplicar esas reglas lo más posible.

Ejemplos de reunión en profundidad a la izquierda:



Ejercicio 2: Sea el esquema universal $R = (A, B, C, D, E)$ y para ese problema del mundo real hallamos las dependencias funcionales: $F = \{AB \rightarrow C, BD \rightarrow E, C \rightarrow D, E \rightarrow A\}$.

Se pide:

- Aplicar el primer paso de descomposición del algoritmo de normalización en FNBC para R y F . Justificarlo.
- Supongamos que para este problema del mundo real encontramos la dependencia funcional $A \rightarrow B$. ¿Conviene agregarla a F ? Justifique su respuesta con un razonamiento paso a paso.

Ejercicio 3: En la siguiente porción de una ontología geográfica la relación entre los nodos hermanos, es decir, entre los que se encuentran en el mismo nivel, es disjunta, y la descomposición es exhaustiva, es decir, todos los elementos del nodo madre se encuentran en alguno de sus nodos hija.

- Proponga una traducción de esta porción de la ontología a un modelo relacional (diagrama entidad relación o esquema)
- Haga la traducción a tablas del modelo relacional.
- Para representar los objetos de la clase "Aguas Marinas", describa qué capacidades expresivas adicionales tiene una base de datos orientada a objetos a diferencia de las capacidades expresivas que ofrece una base de datos relacional sin la extensión de orientación a objetos.

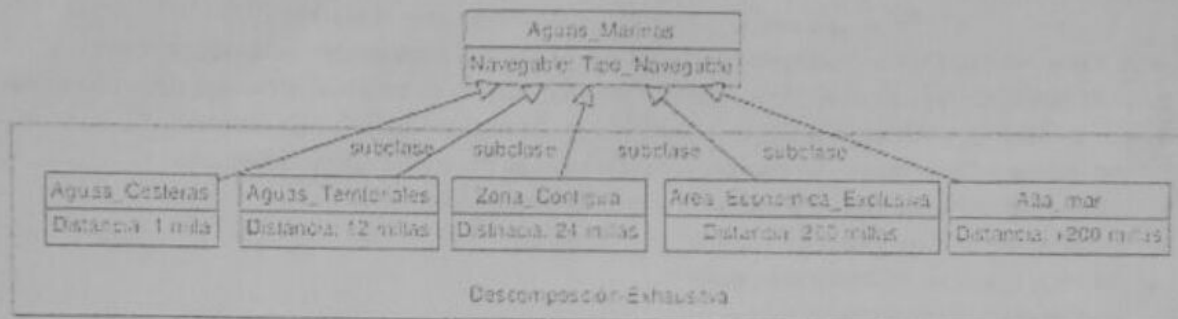


Fig. 2. Ejemplo de Descomposición-Exhaustiva

Ejercicio 4: En el contexto de un ecosistema de información en una institución en la que se ha desplegado un almacén de datos (*data warehouse*), describa la funcionalidad de una componente ETL, glosando el siguiente texto:

"Extract, transform, load (ETL) is a three-phase process where data is extracted, transformed (cleaned, sanitized, scrubbed) and loaded into an output data container. The data can be collated from one or more sources and it can also be outputted to one or more destinations. ETL processing is typically executed using software applications but it can also be done manually by system operators. ETL software typically automates the entire process and can be run manually or on reoccurring schedules either as single jobs or aggregated into a batch of jobs."

"A properly designed ETL system extracts data from source systems and enforces data type and data validity standards and ensures it conforms structurally to the requirements of the output."

Ejercicio 5: Explique dos ventajas y dos desventajas de una base de datos distribuida.