

Segundo Parcial de Fundamentos de Bases de Datos

Noviembre 2019

Duración: 3 horas

Presentar la resolución del parcial:

- Con las hojas numeradas y escritas de un solo lado. Comience cada ejercicio en una hoja nueva.
- Con la cantidad de hojas entregadas en la primer hoja.
- Con cédula de identidad y nombre en cada hoja.
- **Escrita a lápiz y en forma prolija.**

Ejercicio 1 (10 puntos)

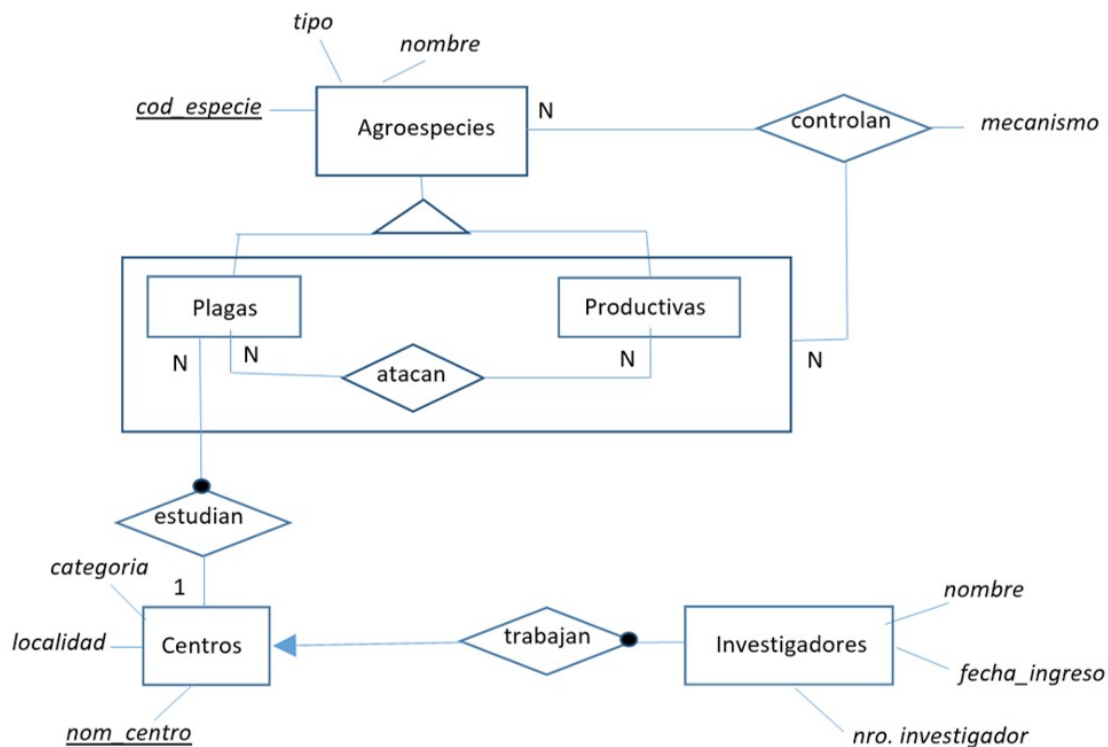
El siguiente Modelo Entidad-Relación es el modelo conceptual de una base de datos de agroespecies.

Las agroespecies son un conjunto de seres vivos que conviven en un ecosistema agrícola, las cuales pueden ser plagas o especies productivas. Cada agroespecie se identifica a través de un código de especie.

Las plagas atacan a una o varias especies productivas, y éstas pueden ser atacadas por una o varias plagas. A su vez, algunas agroespecies (de cualquier tipo) se utilizan para el control biológico de algunas plagas cuando atacan a determinadas especies productivas. Interesa saber el mecanismo por el cual se realiza este control biológico.

Cada plaga es estudiada por un único centro de investigación en control de plagas, que se identifican por su nombre, y cada centro se dedica al estudio de una o más plagas.

En cada centro de control de plagas trabajan un conjunto de investigadores, que se identifican por un número de investigador dentro de cada centro.



RNE: No hay restricciones no estructurales.

Se pide:

Hacer el pasaje de MER a Modelo Relacional, especificando las tablas, sus claves y las dependencias de inclusión.

Ejercicio 2 (20 puntos)

Considere un esquema de relación $R(A,B,C,D,E,G)$ con el siguiente conjunto de dependencias funcionales:

$$F = \{ AB \rightarrow DE, DA \rightarrow C, B \rightarrow G, DE \rightarrow B \}$$

y las siguientes dependencias multivaluadas embebidas: $\{ C \twoheadrightarrow A \mid D, G \twoheadrightarrow A \mid E \}$

- Determine todas las claves. Justifique su respuesta.
- Determine en qué forma normal está R .
- Obtener una descomposición en BCNF de R siguiendo el algoritmo visto en el curso. Explique por qué motivo aplica cada paso. (O justifique).
- Determinar si se perdieron dependencias funcionales.
- Considerar las dependencias multivaluadas embebidas y llevar a 4NF con el algoritmo visto en el curso. Justifique.

Ejercicio 3 (15 puntos)

El siguiente es parte del esquema de una base de datos de gestión de un laboratorio de estudios clínicos.

PACIENTES (idPaciente, nomPaciente, institucionMedica)

Representa información de los pacientes. Identificador, nombre e institución médica a la cual pertenece.

ESTUDIOS (idEstudio, nomEstudio, indicaciones)

Representa información respecto a los diferentes estudios que efectúa el laboratorio. Identificador, nombre del estudio e indicaciones.

ESTUDIOS_PACIENTES (idPaciente, idEstudio, fecha, resultado, entregado)

Representa la información sobre los estudios realizados a cada paciente. Identificador del paciente, identificador del estudio, fecha en la cual se realizó el estudio, resultado del mismo y si ya fue entregado al paciente (con valor TRUE si fue entregado y FALSE en caso contrario).

Además, se conoce la siguiente información sobre estos datos:

	Tamaño	Atributos
PACIENTES	12000	InstituciónMedica tiene 200 valores distintos, distribuidos uniformemente.
ESTUDIOS	550	
ESTUDIOS_PACIENTES	450000	Un 98% de los estudios ya fueron entregados al paciente.

Considere la siguiente consulta sobre el esquema dado:

```
SELECT EP.idPaciente, P.nomPaciente, EP.resultado  
FROM Estudios_Pacientes EP, Pacientes P, Estudios E  
WHERE EP.idPaciente = P.idPaciente AND  
    EP.idEstudio = E.idEstudio AND  
    P.institucionMedica = 'CASMU' AND  
    EP.entregado = TRUE
```

Se pide:

- Detallar el árbol canónico del plan lógico para la consulta.
- Aplicar las heurísticas para optimización llegando al plan lógico optimizado. Explique cada uno de los pasos ejecutados.
- ¿Qué índices le parece útil tener para este esquema y esta consulta? Para cada uno, decir qué tipo de índice sería y sobre qué atributo/s se definiría.
- Considerando las respuestas de las partes a), b) y c), dar un plan físico que le parezca adecuado.

Ejercicio 4 (15 puntos)

Dadas las siguientes transacciones:

T1: r1(X) w1(X) r1(Y) c1

T2: r2(X) r2(Y) w2(Y) c2

y las siguientes historias:

H1: r1(X) w1(X) r2(X) r2(Y) r1(Y) c1 w2(Y) c2

H2: r1(X) r2(X) r2(Y) w1(X) w2(Y) c2 r1(Y) c1

a) Decir cuál de esas historias sería mejor, considerando si su ejecución es equivalente a una ejecución serial y la robustez de las historias para los casos de fallas y/o abortos. Explicitar las propiedades que verifique siguiendo el siguiente formato y justificando:

	Propiedad1	Propiedadn
H1	si/no	si/no
H2	si/no	si/no

b) Ahora considere las siguientes transacciones con locks y unlocks de los ítems de datos.

T1: rl1(X) r1(X) wl1(X) w1(X) rl1(Y) u1(X) r1(Y) u1(Y) c1

T2: rl2(X) r2(X) wl2(X) w2(X) rl2(Y) r2(Y) u2(X) u2(Y) c2

T3: rl3(X) r3(X) u3(X) rl3(Y) r3(Y) u3(Y) wl3(Y) w3(Y) u3(Y) c3

b.1) Para cada transacción decir si cumple el protocolo 2PL, justificando.

b.2) Dar un historia H1 que ejecute T1 y T2, y una historia H2 que ejecute T1 y T3. Ambas historias deben ser posibles en el sistema y deben entrelazar operaciones de lectura y/o escritura de las transacciones que ejecutan.

b.3) Decir si las historias H1 y H2 de la parte anterior son serializables, justificando.