



2. DISEÑO DE BASES DISTRIBUIDAS

Índice

2. DISEÑO DE BASES DISTRIBUIDAS	1
Índice	1
2.1. Estrategias de diseño	1
2.1.1. Proceso Top-Down	1
2.1.2. Proceso Bottom-Up.	1
2.1.3. Diseño de una BDD con proceso Top-Down	1
2.2. Estrategias de fragmentación	2
2.2.1. Tipos de fragmentación	2
2.2.2. Reglas de fragmentación	3
Complejidad	3
Reconstrucción	4
Exclusión	4
2.2.3. Alternativas para decidir la ubicación de cada fragmento.	4
2.2.4. Caso de estudio para ilustrar el concepto de fragmentación:	5
Ejercicio en clase 1	5
2.3. Fragmentación Horizontal	5
2.3.1. Fragmentación horizontal primaria	5
2.3.2. Predicado Simple	5
Ejemplo	6
2.3.3. Estrategia simple para fragmentar	6
Ejercicio en clase 2	6
Ejercicio en clase 3	7
Ejemplo resuelto	8
2.3.4. Fragmentación horizontal derivada	11
Ejemplo	11
2.3.4.1. Semi Join	13
Ejercicio en clase 4	14
2.3.4.2. Fórmula para realizar fragmentación horizontal derivada.	14
Ejemplo	14
Ejercicio en clase 4	15
Ejercicio en clase 5	16
2.4. Fragmentación Vertical	18
Beneficios	19

Retos	19
Ejercicio en clase 6	19
2.5. Fragmentación Híbrida	19
2.5.1. Expresiones de reconstrucción	20
Ejercicio en clase 7	20
Ejercicio en clase 8	21
Ejercicios de repaso - Parcial 1	21

2.1. Estrategias de diseño

Existen 2 principales estrategias para diseñar una BDD.

2.1.1. Proceso Top-Down

- Esta técnica es más apropiada para DDBMS homogéneos.
- Empleada para diseñar BDD desde cero.



2.1.2. Proceso Bottom-Up.

- Esta técnica es más apropiada para Multi bases de datos.
- Las bases de datos ya existen y se decide conectarlas para resolver tareas comunes.



2.1.3. Diseño de una BDD con proceso Top-Down

La metodología para realizar el diseño e implementación de una BDD es similar a la de una BD centralizada excepto por la etapa llamada **Diseño de distribución**

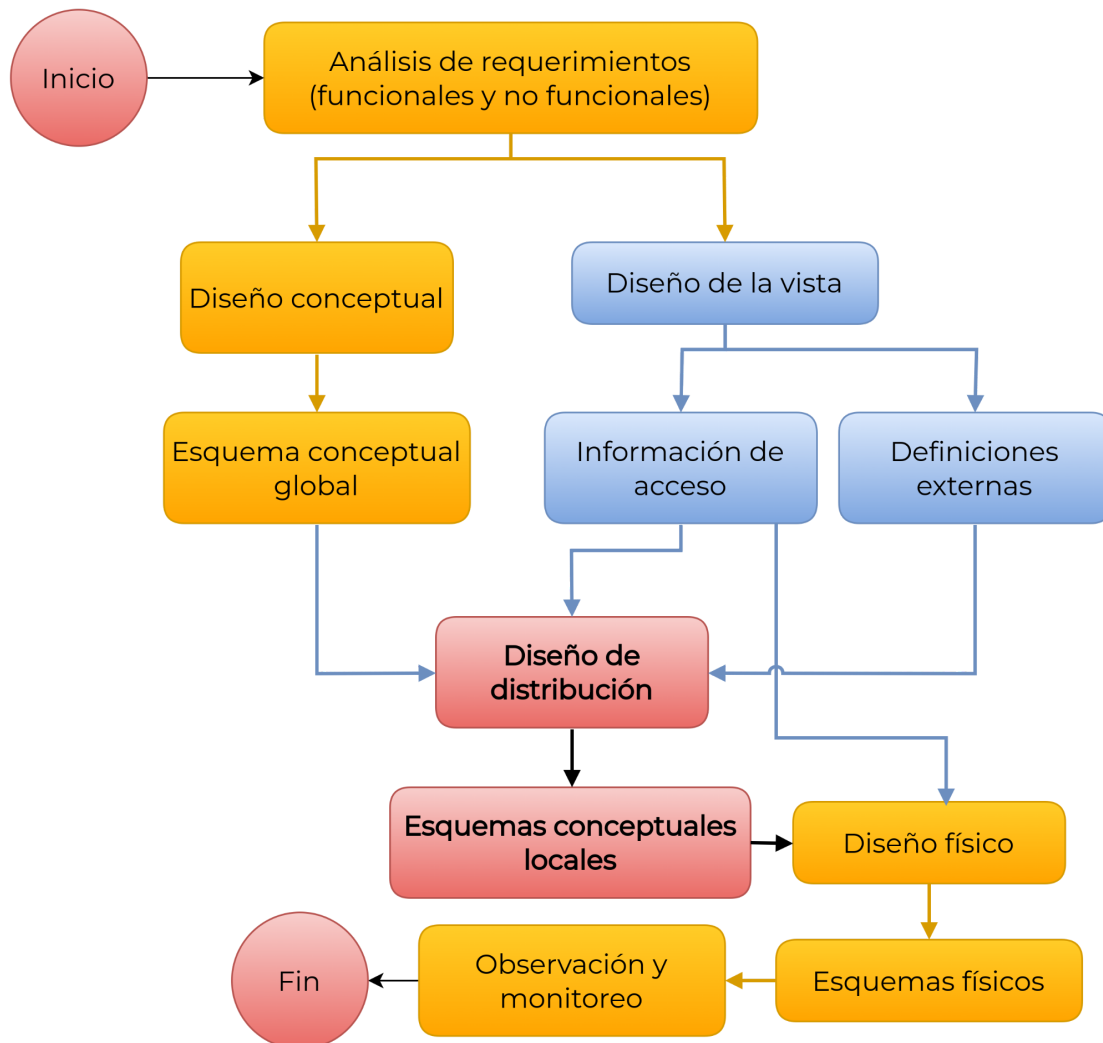
Análisis de requerimientos

- Se definen las necesidades en términos de:
 - Desempeño
 - Confiabilidad
 - Disponibilidad
 - Costos
 - Escalabilidad.
- Diseño de la vista
 - Se definen las interfaces para los usuarios finales
 - Diseño conceptual
 - Similar al diseño conceptual en un DBMS centralizado.
- *Diseño de distribución*
 - En este paso se definen los esquemas conceptuales locales (LCS) en términos de fragmentos.
 - Por lo anterior, el diseño de distribución se enfoca en 2 actividades:
 - Fragmentación

■ Asignación.

● *Diseño físico*

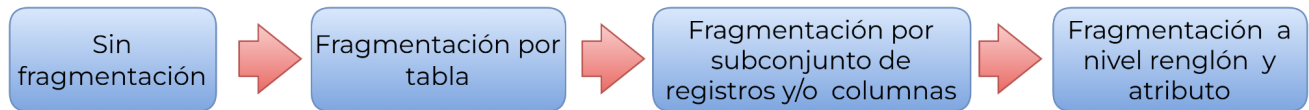
- Similar al diseño físico de una BD centralizada, pero enfocado a cada uno de los sitios.
- En este diseño se definen las estructuras físicas de almacenamiento físico de los datos. Ejemplo: Tablespace, grupos de discos, etc.
- Este diseño puede variar por sitio para aprovechar al máximo sus capacidades particulares.



2.2. Estrategias de fragmentación

Preguntas a responder:

- ¿Qué tanto hay que fragmentar?
- ¿Cómo fragmentar?
- ¿Cómo asignar los fragmentos a cada sitio?



Considerar que: Las aplicaciones generalmente consultan subconjuntos de datos de una o varias tablas.

Si se decide fragmentar por relación (tabla):

- Si la tabla no se replica, la cantidad de accesos remotos puede aumentar, además de no contar con tolerancia a fallas.
- Si la tabla se replica, se producen problemas de desempeño para actualizar adicional al espacio en disco requerido.

Si se decide fragmentar por subconjunto:

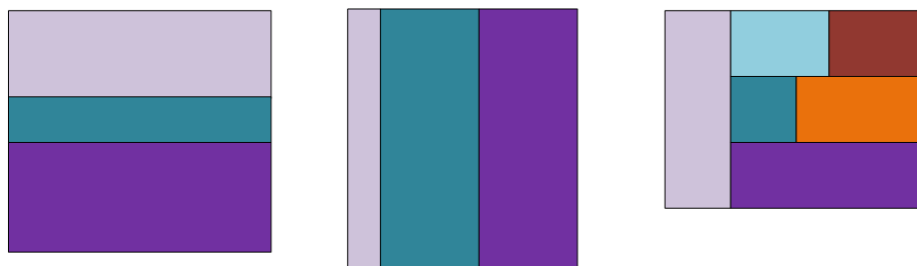
- Al tener datos divididos, permite la ejecución concurrente de transacciones (operaciones DML) .
- Ejecución paralela de consultas (Intra query)
- Complejidad a considerar: Control de la integridad de los datos.

2.2.1. Tipos de fragmentación

Responde a la pregunta 2.

Considerando el razonamiento anterior en cuanto a los niveles de fragmentación, la estrategia comúnmente utilizada es **fragmentación por subconjunto**. Existen 3 variantes:

- Fragmentación Horizontal
- Fragmentación Vertical
- Fragmentación Mixta o híbrida



2.2.2. Reglas de fragmentación

Permiten identificar si un diseño de fragmentación es correcto.

Completitud

Si una relación R es descompuesta en fragmentos $F_R = \{R_1, R_2, \dots, R_n\}$ cada dato debe ser encontrado en alguno de los fragmentos. Ningún dato puede quedar sin ser ubicado en algún R_i

Reconstrucción

Si una relación R es descompuesta en fragmentos F_R , Debe ser posible definir un operador relacional ∇ que permita realizar su reconstrucción. $R = \nabla R_i, \forall R_i \in F_R$

Exclusión

Si una relación R es horizontalmente descompuesta en fragmentos F_R Cada registro debe ubicarse en 1 y solo un fragmento, es decir, los datos contenidos en los fragmentos deben ser mutuamente excluyentes (Esta regla no aplica para fragmentación vertical).

2.2.3. Alternativas para decidir la ubicación de cada fragmento.

Responde a la pregunta 3

Aspectos a considerar:

- Niveles de replicación.
 - A mayor nivel, disminuyen los accesos remotos, pero aumenta la complejidad para manejar concurrencia y sincronización de operaciones DML.
 - Conviene replicar datos que cambian con muy poca frecuencia.
- Ubicación de las aplicaciones que accederán a los datos.
- Características del hardware y software que exista en cada sitio.
- Características de las interfaces de comunicación entre sitios.
- Evitar en la medida de lo posible, operaciones **join** entre fragmentos ubicados en nodos diferentes.
- Cada nodo debería contener únicamente los datos que necesita.

2.2.4. Caso de estudio para ilustrar el concepto de fragmentación:

Una empresa decide distribuir los datos de sus empleados y los datos de los proyectos en los que trabajan. Se almacena nombre, apellido paterno, apellido materno y el puesto asignado al empleado. La empresa cuenta con un catálogo de puestos que contiene el nombre del puesto y su salario mensual.

Para cada uno de los proyectos que se desarrollan en la empresa se almacena el nombre del proyecto, el presupuesto asignado y la ubicación donde se va a desarrollar el proyecto (clave del país, no se requiere crear un catálogo). En un proyecto pueden participar varios empleados.

Para cada uno de los proyectos en los que participa un empleado se almacena la duración en horas que el empleado estará trabajando en el proyecto. Finalmente, para cada uno de los proyectos en los que participa un empleado, se almacena el rol que desempeña (no se requiere crear catálogo).



Ejercicio en clase 1

Generar el modelo relacional del caso de estudio con notación Crow's foot.

2.3. Fragmentación Horizontal

- Considera la formación de fragmentos por subconjunto de registros.
- Cada subconjunto formará un fragmento.
- Existen 2 variantes:
 - Fragmentación horizontal primaria
 - Fragmentación horizontal derivada.

2.3.1. Fragmentación horizontal primaria

Se define empleando una operación de **selección** aplicada a las relaciones (tablas) del esquema global de la base de datos.

Dada una relación R , sus fragmentos horizontales R_i están dados por la expresión:

$$R_i = \sigma_{F_i}(R), 1 \leq i \leq w,$$

F_i Se le conoce como fórmula de selección, o predicado de fragmentación.

	C1	C2	C3
T1			
T2			F1
T3			
T4			F2
T5			
T6			
T7			F3
T8			

F_i Puede ser un predicado formado por un predicado simple, o por un conjunto de predicados simples llamado mini término. No confundir, Mini término representa a varios predicados simples, no solo a uno.

2.3.2. Predicado Simple

Dada una relación $R(A_1, A_2, \dots, A_n)$, donde A_i es un atributo definido sobre un Dominio D_i , un predicado simple p_j , definido en R tiene la forma:

$$p_i: A_i \theta Valor$$

Donde: $\theta \in \{=, <, \neq, \leq, >, \geq\}$, $valor \in D_i$

Ejemplo

pue.nombre="mantenimiento"

A_0 : pue.nombre

θ : =

valor: "mantenimiento"

2.3.3. Estrategia simple para fragmentar

- Para determinar el número de fragmentos, es necesario conocer los predicados simples que emplean las aplicaciones que acceden a los datos, pero ¿Cuántos predicados simples pueden existir en una aplicación?
- Definitivamente es imposible o impráctico considerar el 100% de los predicados. La siguiente regla puede ser útil:

Regla 80/20: El 20% de las consultas de usuario más activas representan el 80% del total de acceso a disco.



Ejercicio en clase 2

Realizar las actividades del siguiente ejercicio. La respuesta no se incluye en estas notas, se sugiere agregar la respuesta en sus apuntes

Del resultado del análisis, se obtuvo que las consultas a la relación AE se comportan de la siguiente manera:

La empresa decide distribuir los datos de la tabla en 2 grupos.

- El primer grupo de se ubicará en su sitio 1 y estará formado por todos aquellos registros que cumplan con todas las siguientes condiciones:
 - Se consultan los registros que tienen rol de **admin** o **supervisor**.
 - Su duración en el proyecto asignado está entre 1 y 3 años.
 - El segundo grupo lo formarán todos los demás registros y estarán ubicados en el sitio 2.
- A. Generar la lista de predicados simples que son de interés para las condiciones anteriores.
- B. Determinar los N fragmentos con sus respectivos predicados.
- C. Determinar la expresión que permita realizar la reconstrucción de la Tabla.



Ejercicio en clase 3

Realizar las actividades del siguiente ejercicio. La respuesta no se incluye en estas notas, se sugiere agregar la respuesta en sus apuntes

Considere el siguiente catálogo de autos de una empresa distribuidora.

AUTO

 AUTO_ID	NUMERIC(10,0)	NOT NULL
 MARCA	VARCHAR(40)	NOT NULL
 MODELO	VARCHAR(40)	NOT NULL
 FECHA_CONSTRUCCION	DATE	NOT NULL
 LUGAR_CONSTRUCCION	VARCHAR(5)	NOT NULL
 CLAVE_AGENCIA	VARCHAR(4)	NOT NULL

La empresa ha decidido distribuir los datos en sus 5 sitios:

- S1 es un sitio que se ha designado para almacenar datos de autos antiguos que prácticamente ya no se modifican.
- S2 es un sitio que se encuentra al norte de la ciudad y administra los datos de los autos que se encuentran en las agencias con claves cuyo primer carácter pertenece al rango [A-H].
- S3 es un sitio que se encuentra en el centro de la ciudad y administra los datos de los autos que se encuentran en las agencias con claves cuyo primer carácter pertenece al rango [I-R]
- S4 es un sitio que se encuentra en la planta de construcción PL-A y administra los datos de los autos que se encuentran en las agencias con claves cuyo primer carácter pertenece al rango [S-Z] y que fueron construidos en la planta PL-A.
- S5 es un sitio que se encuentra en la planta de construcción PL-B y también administra los datos de los autos que se encuentran en las agencias con claves contenidas en el rango [S-Z] y que fueron construidos en la planta PL-B.
- La empresa distribuidora solo cuenta con 2 plantas de construcción PL-A y PL-B

El orden en el que se decide realizar la distribución de los datos es el siguiente:

- Orden 1: Todos los datos de los autos cuya fecha de fabricación sea menor a marzo del 2010, serán almacenados en S1.
- Orden 2: Los datos de los autos restantes se almacenan en los sitios S2, S3, S4, y S5 con base a los criterios

- Construir el esquema de fragmentación en términos de álgebra relacional. Los predicados deben ser expresiones SQL válidas. Se recomienda emplear las funciones `to_char`, `substr`.
- Definir la expresión de reconstrucción en términos de álgebra relacional.

A nivel general se recomienda familiarizarse con las siguientes funciones ya que son ampliamente utilizadas para formar predicados complejos:

- `to_char`
- `to_date`
- `nvl`
- `nvl2`
- `substr`
- `instr`

- **decode**
- **upper**
- **lower**

Para mayores detalles en cuanto a las especificaciones de estas funciones, realizar la siguiente búsqueda: *Single-Row Function Oracle XX* (sustituir XX con la versión correcta).

Ejemplo resuelto

Considere la siguiente tabla de una base de datos centralizada de una universidad que se desea distribuir empleando los diferentes sitios con los que cuenta

PROFESOR		
 PROFESOR_ID	NUMERIC(10,0)	NOT NULL
 NOMBRE	VARCHAR(40)	NOT NULL
 AP_PATERNO	VARCHAR(40)	NOT NULL
 AP_MATERNO	VARCHAR(40)	NULL
 TIPO	CHAR(1)	NULL

Los profesores se clasifican con base en su tipo: Investigadores (I), Directivos (D) y Administrativos (A). Cuando un profesor es de nuevo ingreso no se le asigna su tipo hasta después del primer año (observar que el campo tipo es **NULL**).

- A. Una primera propuesta de fragmentación es considerar únicamente el tipo de profesor. Generar el esquema de fragmentación indicando las expresiones en términos de álgebra relacional y el tipo de fragmentación utilizado.

Para este caso se tendrán 4 fragmentos:

PROF ₁
PROF ₂
PROF ₃
PROF ₄

Las expresiones que definen a cada fragmento se expresan de la siguiente forma:

$$PROF_1 = \sigma_{tipo='I'}(PROF)$$

$$PROF_2 = \sigma_{tipo='D'}(PROF)$$

$$PROF_3 = \sigma_{tipo='A'}(PROF)$$

$$PROF_4 = \sigma_{tipo \text{ is null}}(PROF)$$

Notar que los predicados de cada selección debe ser una expresión booleana válida en términos de SQL.

B. Generar la expresión de reconstrucción que permita recuperar todos los datos de la tabla original.

Para este caso, la expresión de reconstrucción en una fragmentación horizontal primaria corresponde a la unión de todos sus fragmentos.

$$PROF = PROF_1 \cup PROF_2 \cup PROF_3 \cup PROF_4$$

C. Considerando la siguiente tabla de datos, generar los N fragmentos (**PROF_1**, **PROF_2**, etc.) con base a los criterios de fragmentación.

PROFESOR_ID	NOMBRE	AP_PATERNO	AP_MATERNO	TIPO
1	JUAN	LOMA	KIM	I
2	LALO	PAEZ	MONTES	D
3	LUIS	GIL		A
4	PEPE	PEREZ	BAEZ	I
5	HUGO	RUIZ	LUNA	D
6	MARA	RUIZ		
7	EVA	GIL	CARSO	

Al aplicar las expresiones de fragmentación a la tabla original, los datos quedarán distribuidos de la siguiente manera:

PROF_1

PROFESOR_ID	NOMBRE	AP_PATERNO	AP_MATERNO	TIPO
1	JUAN	LOMA	KIM	I
4	PEPE	PEREZ	BAEZ	I

PROF_2

PROFESOR_ID	NOMBRE	AP_PATERNO	AP_MATERNO	TIPO
2	LALO	PAEZ	MONTES	D
5	HUGO	RUIZ	LUNA	D

PROF_3

PROFESOR_ID	NOMBRE	AP_PATERNO	AP_MATERNO	TIPO
3	LUIS	GIL		A

PROF_4

PROFESOR_ID	NOMBRE	AP_PATERNO	AP_MATERNO	TIPO
6	MARA	RUIZ		

7	EVA	GIL	CARSO	
---	-----	-----	-------	--

Suponga que tiempo después las reglas de negocio cambian. Ahora un profesor puede tener varios tipos a la vez. Por ejemplo, el profesor JUAN puede ser I y D. El diseño global de la BD quedaría definido de la siguiente manera:

PROFESOR		
 PROFESOR_ID	NUMERIC(10,0)	NOT NULL
 NOMBRE	VARCHAR(40)	NOT NULL
 AP_PATERNO	VARCHAR(40)	NOT NULL
 AP_MATERNO	VARCHAR(40)	NULL
 ES_INVESTIGADOR	BIT	NOT NULL
 ES_ADMIN	BIT	NOT NULL
 ES_DIRECTIVO	BIT	NOT NULL

Se decide reubicar a los profesores de la siguiente manera:

- Profesores de nuevo ingreso (sin roles asignados), se reubicarán al sitio 1 en Guadalajara (GDL).
- Profesores con un solo rol asignado se reubicarán al sitio 2 en Monterrey (MTY)
- Profesores con más de un rol se reubicarán al sitio 3 en la Cd de México (CDMX)

D. Proponer un nuevo esquema de fragmentación que pudiera resolver el problema asegurándose de cumplir con las reglas de fragmentación. Los predicados de cada fragmento deben ser expresiones SQL válidas. Recomendación: emplear la función **decode** de Oracle.

$$PROF_1 = \sigma_{p1}(PROF)$$

$p1$:

$$(decode(es - investigador, true, 1, false, 0) + \\ decode(es - directivo, true, 1, false, 0) + \\ decode(es - admin, true, 1, false, 0)) = 0$$

$$PROF_2 = \sigma_{p2}(PROF)$$

$p2$:

$$(decode(es - investigador, true, 1, false, 0) + \\ decode(es - directivo, true, 1, false, 0) + \\ decode(es - admin, true, 1, false, 0)) = 1$$

$$PROF_3 = \sigma_{p3}(PROF)$$

$p3$:

$$(decode(es - investigador, true, 1, false, 0) + \\ decode(es - directivo, true, 1, false, 0) + \\ decode(es - admin, true, 1, false, 0)) > 1$$

El uso de la función **decode** permite simplificar las expresiones booleanas. Las columnas **es_investigador**, **es_directivo** y **es_admin** regresan valores booleanos. La función **decode** se comporta de la siguiente manera:

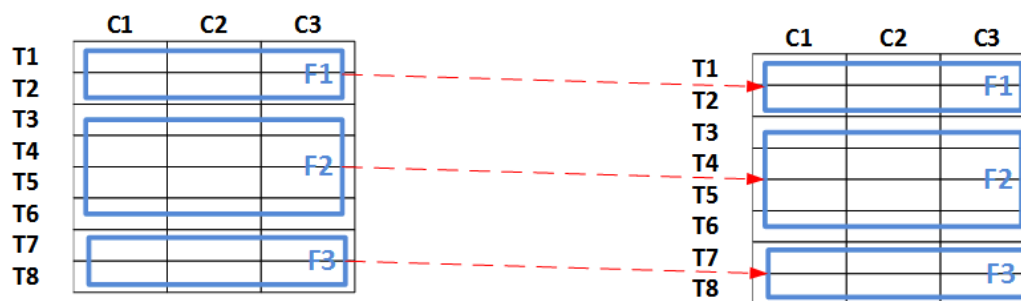
Se compara el valor de la columna con el segundo parámetro. Si son iguales, regresa el tercer parámetro, de lo contrario, la columna se compara con el cuarto parámetro, si son iguales, se regresa el quinto parámetro. Nota que este comportamiento es similar a un **switch**:

- **es_investigador == true** => regresa 1
- **es_investigador == false** => regresa 0

La misma estrategia se aplica con los otros 2 atributos. Al final se suman los valores y se comparan con el valor esperado.

2.3.4. Fragmentación horizontal derivada

Se define a partir de atributos que se encuentran en una relación R diferente a la relación S que se va a fragmentar. Se desea fragmentar a la relación “hija” (la que contiene a la FK) con base a los predicados aplicados sobre los atributos de la tabla “padre”.



Ejemplo

La empresa decide distribuir los datos de los empleados en 2 sitios con base en su sueldo.

- Sitio 1: empleados con sueldo $\leq 25,000$
- Sitio 2: empleados con sueldo $> 25,000$

A. Considerando la siguiente muestra de datos, construir 2 nuevas tablas que contengan los datos en cada sitio:

EMPLEADO

empleado_id	nombre	clave_puesto
1	JUAN	DR
2	EVA	JD
3	MARIO	OP
4	LILI	DR
5	LUIS	JD

empleado_id	nombre	clave_puesto
6	PEDRO	AT
7	IVAN	OP
8	IRMA	AT

PUESTO

sueldo	nombre	clave puesto
48000	DIRECTOR	DR
33000	JEFE DEPTO.	JD
15000	OPERATIVO	OP
18000	AUXILIAR TÉCNICO	AT
9000	BECARIO	BE

De la muestra anterior, se puede deducir que los datos en cada sitio serán:

EMP1

empleado_id	nombre	clave_puesto
3	MARIO	OP
6	PEDRO	AT
7	IVAN	OP
8	IRMA	AT

EMP2

empleado_id	nombre	clave_puesto
1	JUAN	DR
2	EVA	JD
4	LILI	DR
5	LUIS	JD

- Para determinar los registros anteriores se debe realizar un join entre ambas tablas:

$$EMP_1 = \pi_{emp-id, nombre, clave-puesto} (EMP \bowtie_{clave-puesto} \sigma_{sueldo \leq 25000} (PU))$$

$$EMP_2 = \pi_{emp-id, nombre, clave-puesto} (EMP \bowtie_{clave-puesto} \sigma_{sueldo > 25000} (PU))$$

- Las expresiones anteriores tienen un problema de desempeño en especial si las relaciones EMP Y PU se encuentran en nodos diferentes:

Caso 1: PU (PUESTO) se transmite a un sitio

- PU debe transmitirse a otro sitio para poder realizar el join.

- El resultado final no requiere incluir los datos de PU que solo se requiere conocer los registros de EMP que tienen correspondencia con PU₁ y PU₂.
- Para optimizar se aplican las operaciones de selección y proyección sobre PU₁ y PU₂ para minimizar costos de transmisión:

$$EMP_1 = \pi_{emp-id, nombre, clave-puesto}(EMP \bowtie_{clave-puesto} (\pi_{clave-puesto} (\sigma_{sueldo \leq 25000}(PU))))$$

$$EMP_2 = \pi_{emp-id, nombre, clave-puesto}(EMP \bowtie_{clave-puesto} (\pi_{clave-puesto} (\sigma_{sueldo > 25000}(PU))))$$

- Observar que el resultado está formado por subconjuntos de EMP.
- Cada subconjunto contiene registros que tienen correspondencia o hacen "join" con la tabla PU.

Caso 2: EMP se transmite a un sitio.

- De forma similar hay problemas de desempeño ya que se tiene que transmitir el 100% de los registros para poder determinar cada subconjunto en el otro sitio.
- Para estas situaciones se emplea el concepto de **semi join** (\bowtie_F).

2.3.4.1. Semi Join

El Semijoin de una relación R con atributos A aplicado con una relación S con atributos B, corresponde al subconjunto de tuplas de R que participan en el join de R con S se denota por $R \bowtie_F S$. Es decir:

$$R \bowtie_F S = \pi_A(R \bowtie_F S) = \pi_A(R) \bowtie_F \pi_{A \cap B}(S)$$

De lo anterior, en una fragmentación horizontal derivada, cada fragmento está formado por registros que tienen correspondencia o hacen JOIN con cada uno de los fragmentos de la tabla padre.



Ejercicio en clase 4

Realizar las actividades del siguiente ejercicio. La respuesta no se incluye en estas notas, se sugiere agregar la respuesta en sus apuntes

Empleando el concepto de semi-join generar 2 consultas SQL que determinen los registros obtenidos anteriormente.

2.3.4.2. Fórmula para realizar fragmentación horizontal derivada.

Con base al análisis anterior, la fragmentación horizontal derivada de una relación R está formada por un conjunto de fragmentos R_i definidos por:

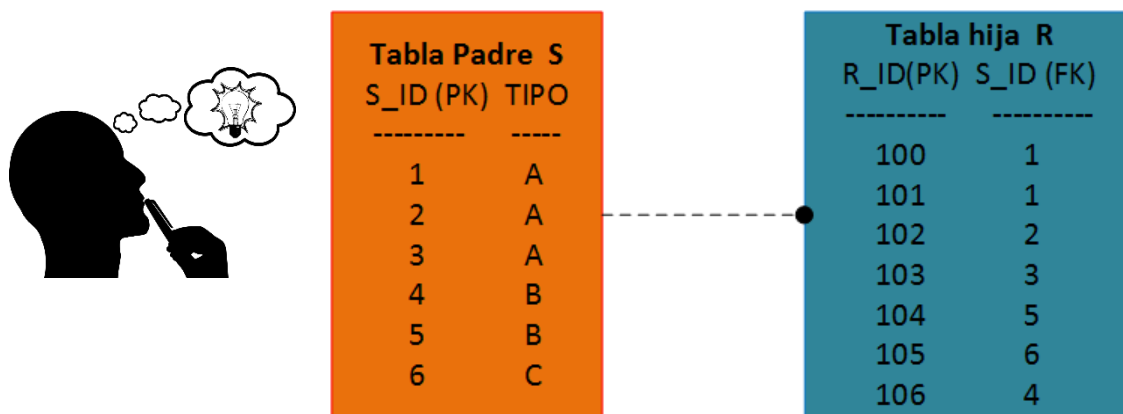
$$R_i = R \bowtie S_i, 1 \leq i \leq w$$

Donde w Es el número máximo de fragmentos, R es relación (tabla) hija que se desea fragmentar, S es la relación (tabla) padre.

El efecto que produce esta operación es filtrar los registros que contendrán los fragmentos R_i , los cuales corresponden a todos aquellos registros que tienen correspondencia con los registros del fragmento S_i .

Ejemplo

Considerar a las 2 relaciones globales R y S en la que existe una relación 1:M



- Suponer que la tabla padre S se distribuye de forma horizontal primaria con base a los valores del campo tipo para formar 3 fragmentos.

$$S_1 = \sigma_{tipo=A}(S)$$

$$S_2 = \sigma_{tipo=B}(S)$$

$$S_3 = \sigma_{tipo=C}(S)$$

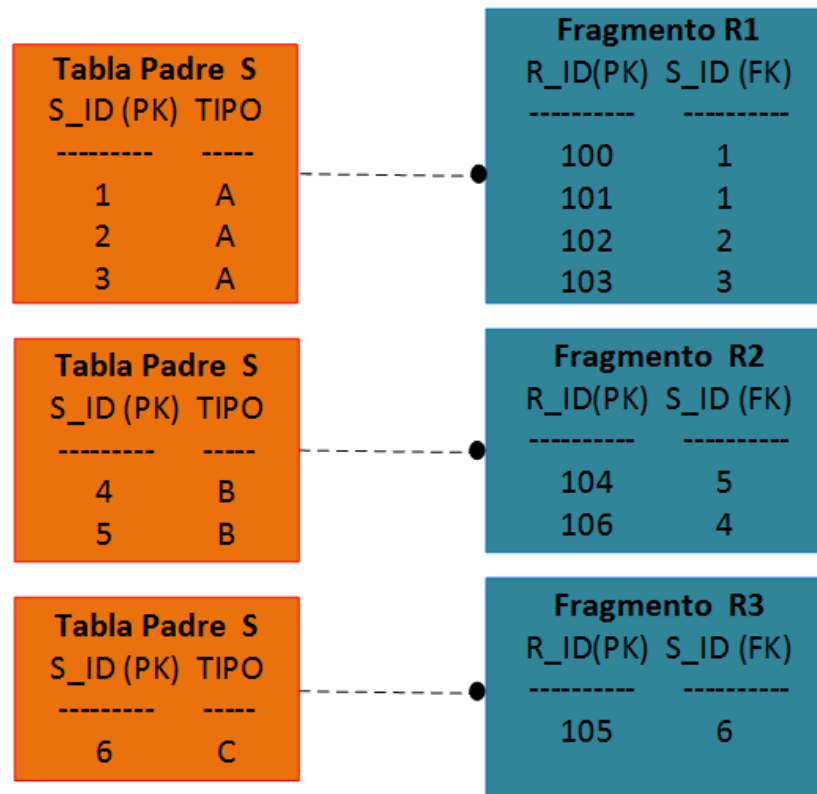
Tabla Padre S	Tabla Padre S	Tabla Padre S
S_ID (PK) TIPO	S_ID (PK) TIPO	S_ID (PK) TIPO
-----	-----	-----
1 A	4 B	6 C
2 A	5 B	
3 A		

- Si R se fragmenta empleando el mismo criterio, los registros de cada fragmento en R deberán corresponder con los registros de S a través de la relación $PK \rightarrow FK$ que existe entre ambas.
- Otra forma de describir esta relación es: Cada fragmento de R contendrá todas aquellas tuplas que *hacen join* (concepto de semi-join) con las tuplas de S .

$$R_1 = R \bowtie_{s-id} S_1$$

$$R_2 = R \bowtie_{s-id} S_2$$

$$R_3 = R \bowtie_{s-id} S_3$$



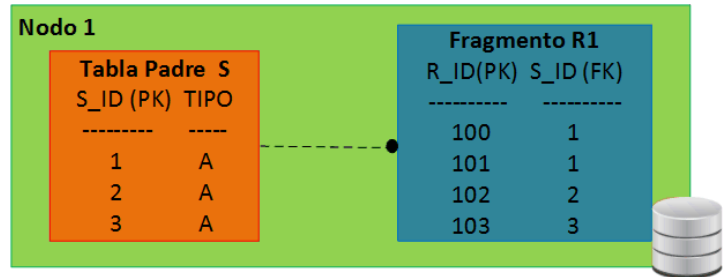
¿Dónde ubicar estos fragmentos?

- Típicamente tanto el fragmento padre como el fragmento hijo se deben ubicar en el mismo sitio.
- Lo anterior permite que la FK entre el fragmento padre y el fragmento hijo se conserve. De esta forma será posible realizar operaciones Join de forma **local y paralela**.

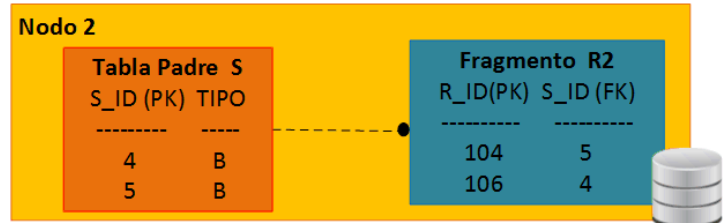


```
select *
from S
join R on
s.s_id = r.s_id;
```

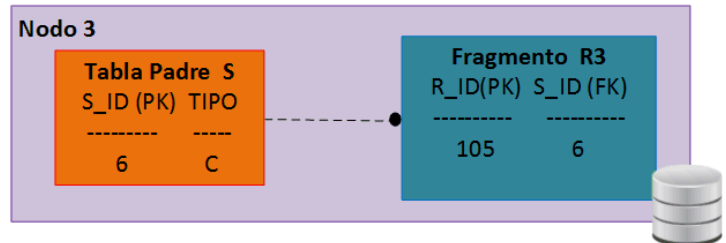
```
select *
from S1
join R1 on
s.s_id = r.s_id;
```



```
select *
from S2
join R2 on
s.s_id = r.s_id;
```



```
select *
from S3
join R3 on
s.s_id = r.s_id;
```



Ejercicio en clase 4

Realizar las actividades del siguiente ejercicio. La respuesta no se incluye en estas notas, se sugiere agregar la respuesta en sus apuntes

La empresa ha decidido distribuir su BD en 5 sitios considerando las siguientes reglas:

- Los proyectos se administran con base a su ubicación: MX, BR y JAP. Existe un nodo en cada país. Por otro lado, los empleados se administran con base a su sueldo. Existen 2 aplicaciones, cada una instalada en un sitio (su ubicación es diferente a los 3 anteriores).
 - Aplicación 1: Empleados con sueldo $\leq 20,000$
 - Aplicación 2: Sueldo $>20,000$

- Determinar el esquema de fragmentación para empleado (EMP), puesto (PU) Y proyecto (PR)
- Determinar las posibles opciones que existen para fragmentar **asignacion_empleado** (AE).
- Para cada opción proponer un esquema de asignación empleado para ello un modelo relacional.



Ejercicio en clase 5

Realizar las actividades del siguiente ejercicio. La respuesta no se incluye en estas notas, se sugiere agregar la respuesta en sus apuntes

Considere el siguiente modelo relacional de una red social que ofrece servicios de streaming para archivos de audio. La empresa decide distribuir sus datos debido al crecimiento de usuarios. Los requerimientos de distribución son:

Para los Usuarios:

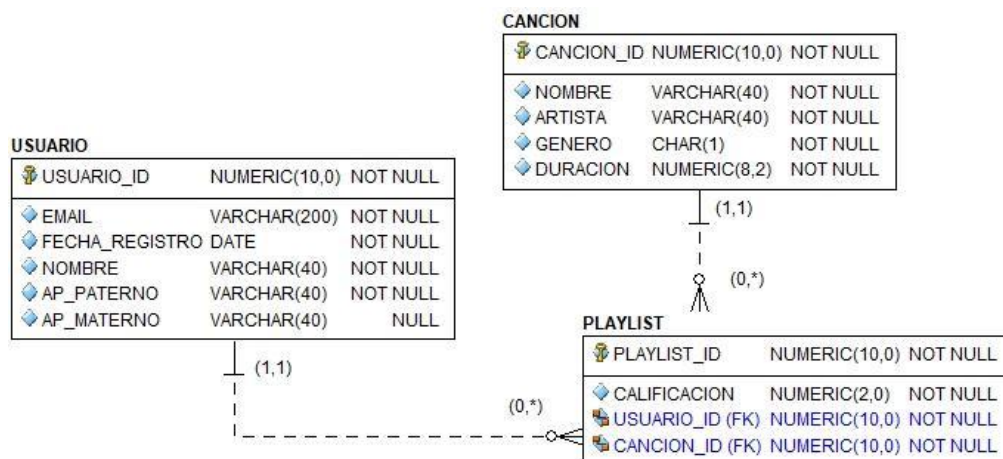
- Se decide crear 2 grupos de usuarios: El grupo 1 estará formado por todos aquellos usuarios registrados hasta el 31/12/2010. El resto de los usuarios se ubicarán en el grupo 2.
- Para cada uno de los grupos anteriores se considera un criterio de distribución adicional: Usuarios cuyo correo electrónico pertenece al dominio social.music.com y usuarios que pertenezcan a otro dominio.
- Cada fragmento estará en un sitio.

Para las canciones:

- El catálogo de canciones se distribuye en 2 grupos: canciones que tengan una duración de hasta 180 segundos se ubicarán en un sitio donde el enlace de red con velocidad media-alta. El resto de las canciones se ubicarán en un sitio donde el enlace de red es de alta calidad.
- Estos 2 sitios son independientes a los sitios donde se encuentran los datos de los usuarios.

Para el playlist:

- Se consideran 2 criterios:
 - Cuando el usuario agrega una canción a su playlist, este le asigna una calificación entre 1 y 10. Se ha decidido separar a todos los registros que tengan calificación 1,2,3 o 4 y el resto en otro grupo.
 - El segundo criterio se refiere al desempeño. Los datos de **PLAYLIST** y los datos de **USUARIO** se consultan con alta frecuencia. Se desea minimizar los accesos remotos para este tipo de consultas. Pueden existir más de un fragmento en el mismo sitio.



- Genere el esquema de fragmentación para las tablas globales. Los predicados deben ser expresiones SQL válidas.
- Determinar el número de sitios que se requieren para distribuir a los fragmentos obtenidos en el punto anterior. Para cada sitio S_1, \dots, S_n , sin indicar la lista de fragmentos que contendrá.
- Genere un modelo relacional que muestre los fragmentos que existirán únicamente para el sitio donde se encontrarán las canciones con calificación 10 que le pertenecen a un usuario registrado el 01/01/2009 y cuyo email es rocko@social.music.com. Indicar en el diagrama las restricciones de referencia que pueden conservarse.
- Genere un modelo relacional que muestre los fragmentos que existirán en el sitio donde se ubican las canciones con una duración de al menos 5 min.

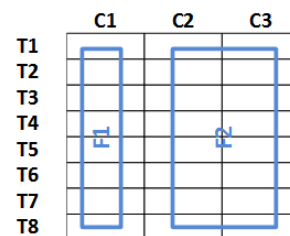
Para ambos modelos relacionales, emplear notación Crow's foot e indicar cardinalidad.

2.4. Fragmentación Vertical

La fragmentación vertical de una relación R produce fragmentos R_1, R_2, \dots, R_n donde cada fragmento contiene un subconjunto de los atributos de R .

La fragmentación vertical se define a partir de una operación de proyección en álgebra relacional:

$$\pi_{A_1, A_2, \dots, A_n}(R)$$



Beneficios

- Contar con relaciones más pequeñas implica menos operaciones I/O
- Permite almacenar los fragmentos en medios distintos, por ejemplo: sub-relaciones identificadas como las más activas pueden manejarse en esquemas de memoria más eficientes: caché.
- Seguridad. Columnas que requieren cifrado pueden almacenarse en medios especializados que cuenten con el software adecuado para cifrado.
- Manejo de datos binarios en medios de almacenamiento especializados.

Retos

- La fragmentación vertical puede resultar más complicada que la horizontal.
- En la fragmentación vertical una tabla con m atributos diferentes a la PK el número posible de fragmentos está dado por $B(m)$, el cual es el n -enésimo número Bell, es decir, para valores grandes de m , $B(m) \approx mm$.
- Lo anterior complica la determinación de alguna estrategia óptima. En su lugar se emplea la siguiente heurística: Identificar conjuntos de atributos que normalmente se consultan juntos: medida de afinidad.
- Un punto importante es que la PK debe replicarse en todos los fragmentos para poder reconstruir la relación.
- La regla de exclusión no aplica para la fragmentación vertical.



Ejercicio en clase 6

Realizar las actividades del siguiente ejercicio. La respuesta no se incluye en estas notas, se sugiere agregar la respuesta en sus apuntes

Considerar la relación **cliente** y las características de los siguientes 3 sitios:

- Sitio 1: Servidor con amplia capacidad de almacenamiento y software especializado para procesar imágenes.
- Sitio 2: Servidor principal de captura de datos generales.
- Sitio 3: Servidor de seguridad empleado para almacenar los datos considerados como “delicados” o “confidenciales”.

A. Generar una propuesta de esquema de fragmentación

B. Generar la expresión de reconstrucción.

2.5. Fragmentación Híbrida

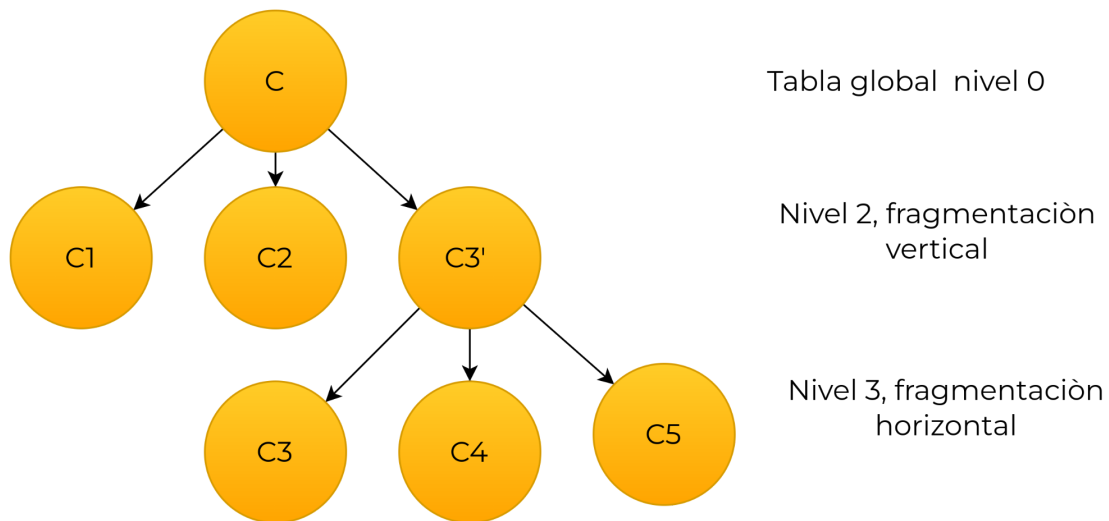
En la mayoría de los casos prácticos, un esquema de fragmentación horizontal o vertical no es suficiente. La fragmentación híbrida consiste de una fragmentación horizontal seguida de una vertical o viceversa.

$$\sigma_p(\pi_{A_1, \dots, A_n}(R)) \text{ o también } \pi_{A_1, \dots, A_n}(\sigma_p(R))$$

	C1	C2	C3
T1			
T2			
T3			F3
T4			
T5	F1	F2	F4
T6			
T7			
T8			F5

2.5.1. Expresiones de reconstrucción

- Teóricamente en una fragmentación híbrida se pueden formar jerarquías con N niveles de fragmentación.
- A mayor profundidad, aumenta la complejidad para establecer la expresión de distribución.
- Para simplificar este proceso, es posible construir un árbol que represente a la jerarquía.
- Cada nodo del árbol representa a un fragmento o a una expresión intermedia.
- El nodo raíz representará al primer nivel de fragmentación, se agregan tantos niveles como niveles de fragmentación existan.
- La generación de la expresión de reconstrucción se inicia de abajo hacia arriba, es decir, de las hojas del árbol hacia el nodo raíz.
- Por cada nodo padre se genera una expresión de reconstrucción que depende del tipo de fragmentación aplicada en dicho nivel.
- El resultado de esta expresión se emplea para construir la expresión del nivel superior, y así sucesivamente hasta llegar al nodo raíz.



Ejercicio en clase 7

Realizar las actividades del siguiente ejercicio. La respuesta no se incluye en estas notas, se sugiere agregar la respuesta en sus apuntes

Suponer la relación cliente mostrada en el siguiente modelo relacional. Se requiere realizar un proceso de distribución con base a los siguientes criterios:

1. Se cuenta con servidor dedicado con grandes capacidades de almacenamiento y procesamiento de imágenes ubicado en un sitio S1. Se ha decidido almacenar todas las fotos de los clientes en este sitio.
2. Por seguridad, todos los números de tarjeta de los clientes serán almacenados de forma independiente en un servidor que cuenta con herramientas de cifrado y diversas herramientas de seguridad, ubicado en S2.
3. Para los datos restantes se ha decidido generar 3 grupos de empleados con base a los últimos 2 dígitos del número de seguridad. Los rangos son [00 – 39], [40 - 79], [80-99]. Se requiere generar un fragmento para cada grupo.

CLIENTE		
CLIENTE_ID	NUMERIC(10,0)	NOT NULL
NOMBRE	VARCHAR(40)	NOT NULL
AP_PAT	VARCHAR(40)	NOT NULL
AP_MAT	VARCHAR(40)	NOT NULL
FOTO	VARBINARY/BLOB(max)	NOT NULL
NUM_TARJETA	VARCHAR(16)	NOT NULL
NUM_SEGURIDAD	VARCHAR(3)	NOT NULL

- A. Generar el esquema de fragmentación en términos de álgebra relacional.
- B. Generar una expresión en términos de álgebra relacional que permita realizar su reconstrucción.
- C. Generar la expresión de reconstrucción del punto anterior en SQL. Emplear sintaxis estándar y sintaxis anterior. Asumir la existencia de transparencia de localización.

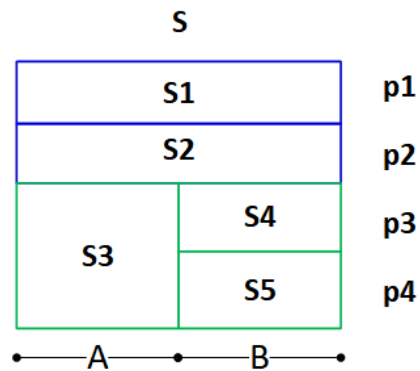


Ejercicio en clase 8

Realizar las actividades del siguiente ejercicio. La respuesta no se incluye en estas notas, se sugiere agregar la respuesta en sus apuntes

Considerar el siguiente esquema de fragmentación de una relación "S" mostrado en la figura.

- A. Generar el esquema de fragmentación en términos de álgebra relacional.
- B. Generar una expresión en términos de álgebra relacional que permita realizar su reconstrucción.



Ejercicios de repaso - Parcial 1

Realizar los ejercicios de repaso para los temas 1 y 2 indicados en el documento

[ejercicios-repaso.pdf](#)