



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE  
 ESCUELA DE INGENIERÍA  
 DEPARTAMENTO DE CIENCIA DE LA COMPUTACIÓN  
 PROCESAMIENTO DE DATOS MASIVOS - IIC2440 – 2024-1

IIC2440 — Procesamiento de Datos Masivos — 1' 2024

## Tarea – 1

A continuación están las tablas que ideamos para poder solventar lo que pedido:

1. usuarios(id, nombre, correo, edad)
2. frutas(id, nombre, descripcion, precio\_unitario)
3. despacho(id, id\_compra\_usuario, id\_direccion, fecha\_despacho)
4. direccion(id, id\_usuario, calle, numero, region, codigo\_postal, adicional)
5. compra(id\_compra\_usuario, id\_fruta, cantidad)
6. compra\_usuario(id, id\_usuario, fecha\_compra)

Así se verían las tablas en cuestión junto con su Esquema [Haz clic aquí para ver el Esquema:](#)

1. usuarios(id, nombre, correo, edad)

Columna	Tipo de Datos	Restricciones
<u>id</u>	INTEGER	not null
nombre	TEXT	
correo	TEXT	
edad	INTEGER	

2. frutas(id, nombre, descripcion, precio\_unitario)

Columna	Tipo de Datos	Restricciones
<u>id</u>	INTEGER	not null
nombre	TEXT	
descripcion	TEXT	
precio_unitario	NUMERIC	

3. **despacho**(id, id\_compra, id\_direccion, fecha\_despacho)

Columna	Tipo de Datos	Restricciones
<u>id</u>	INTEGER	not null
id_compra_usuario	INTEGER	
id_direccion	INTEGER	
fecha_despacho	DATE	

4. **direccion**(id, id\_usuario, calle, numero, region, codigo\_postal, adicional)

Columna	Tipo de Datos	Restricciones
<u>id</u>	INTEGER	not null
id_usuario	INTEGER	
calle	TEXT	
numero	INTEGER	
region	TEXT	
codigo_postal	TEXT	
adicional	TEXT	

5. **compra**(id\_compra, id\_fruta, cantidad)

Columna	Tipo de Datos	Restricciones
id_compra_usuario	INTEGER	not null
id_fruta	INTEGER	
cantidad	INTEGER	

6. **compra\_usuario**(id, id\_usuario, fecha\_compra)

Columna	Tipo de Datos	Restricciones
<u>id</u>	INTEGER	not null
id_usuario	INTEGER	
fecha_compra	DATE	

## **Tipo de indices por Tabla:**

1. Para la tabla usuarios el tipo de indice que usariamos sería B+Tree Clustered en el "id" para facilitar la busqueda en igualdad y rangos de "id". También, para facilitar las busquedas por edades, incluiriamos un B+Tree Unclustered para cada edad asi tener acceso directo a los punteros de las personas que se requiere obtener información.
2. En el caso de la tabla frutas para el id usaremos un B+tree Clustered para facilitar. Por otro lado, tambien agregariamos un B+tree Unclustered para el precio\_unitario ya que si necesitamos obtener rangos o igualdades de precios será más facil.
3. En la tabla Despacho, a lo igual que en las demas tablas para el id usaremos un B+Tree Clustered. cabe recalcar que, para buscar con mejor eficiencia se agregaria una indexacion multicolumnar de las columnas (id\_compra\_usuario, id\_direccion). También, el agregar un indice a la fecha\_despacho nos facilitaria las busquedas por lo que usariamos un B+Tree Unclustered.
4. Para esta tabla de direcciones en el apartado de id se usara un B+Tree Clustered y en el caso de los id de los usuarios(id\_usuario) tambien se usara B+tree Unclustered (asumiendo que una persona puede tener más de una direccion registrada).
5. En la tabla de compra realizaremos una indexacion del estilo multiindex para las columnas de id\_fruta e id\_compra\_usuario obteniendo esta tupla (id\_fruta, id\_compra\_usuario) aunque tambien es una buena idea realizar la otra dirección para optimizar la busqueda (id\_compra\_usuario, id\_fruta)
6. Finalmente, para la tabla de compra\_usuario tendremos la indexación del id mediante un B+Tree Clustered y un B+Tree Unclustered para la columna id\_usuario. También, el agregar un indice a la fecha\_compra nos facilitaria las busquedas por lo que usariamos un B+Tree Unclustered.

## **Plan de Consulta:**

1. Para la realizacion de esta consulta como tenemos indexadas los id, realizar la busqueda mediante id es rapida y obtenemos toda la información de la fruta directamente.
2. En esta consulta, como tenemos la tabla de compra\_usuario ya tenemos los usuarios indexados seria directo obtener todas las compras que ha realizado, y con esto obtenemos los id necesarios para poder consultar a las demas tablas, en el caso de la direccion la podemos obtener rapidamente de la tabla despacho junto con la direccion. Ahora, para el apartado del calculo del total usaremos la tabla de compras la cual mediante la multiindexación podemos encontrar todas las frutas compradas por el usuarios por cada compra y su cantidad respectiva, luego con el id de la fruta obtenemos el precio unitario.
3. como ya tenemos la id de la compra del usuario, podemos acudir a la tabla compra la cual gracias a su indexacion multicolumnar de (id\_compra\_usuario, id\_fruta) obtenemos

todos los ide de las frutas que se realizo en la compra y con esto obtenemos la informacion del precio unitario de la fruta gracias al id y de la tabla compra ya sabemos la cantidad de fruta que compró.

4. Primero filtramos la tabla compra\_usuario por fecha para asi obtener los id de las compras realizadas en ese tiempo. Como ya tenemos los id de compra podemos acceder a las frutas que se realizaron en esa compra gracias a el multiindex de la tabla compra y finalmente filtramos por el id fruta.

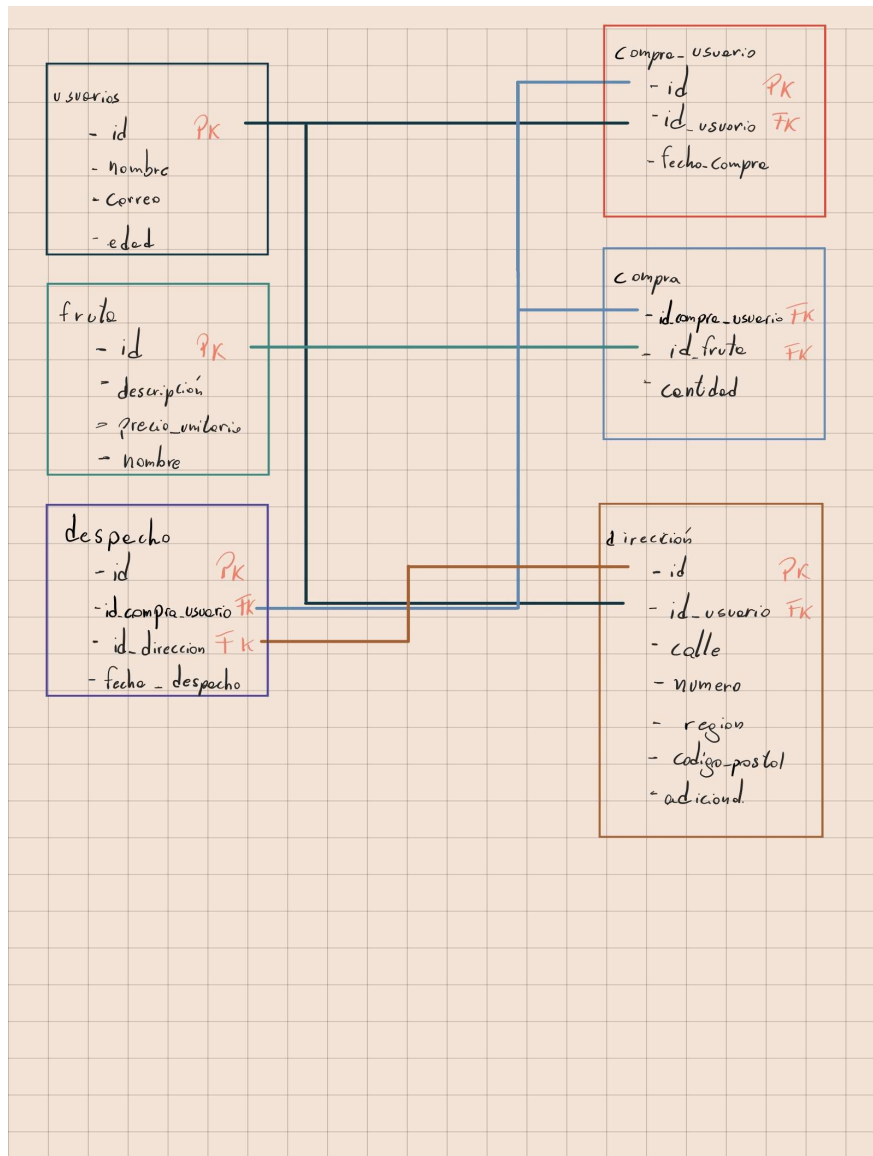


Figure 1: Esquema de las tablas de la parte 1

## Esquema de las tablas de la parte 2

1. usuarios(id, id\_usuario, region)
2. frutas(id\_frutas, fruta, precio)
3. compra\_user(id\_usuario, id\_compra, fecha, region)
4. compra\_fruta(id\_compra, id\_fruta, cantidad)

1. **usuarios**(id, id\_usuario, region)

Columna	Tipo de Datos	Restricciones
<u>id</u>	INTEGER	not null
id_usuario	INTEGER	
region	TEXT	

2. **frutas**(id\_frutas, fruta, precio)

Columna	Tipo de Datos	Restricciones
<u>id frutas</u>	INTEGER	not null
fruta	TEXT	
precio	NUMERIC	

3. **compra**(id\_usuario, id\_compra, fecha, region)

Columna	Tipo de Datos	Restricciones
<u>id_usuario</u>	INTEGER	not null
<u>id compra</u>	INTEGER	not null
fecha	DATE	
region	TEXT	

4. **compra\_fruta**(id\_compra, id\_fruta, cantidad)

Columna	Tipo de Datos	Restricciones
<u>id_compra</u>	INTEGER	not null
<u>id_fruta</u>	INTEGER	not null
cantidad	INTEGER	

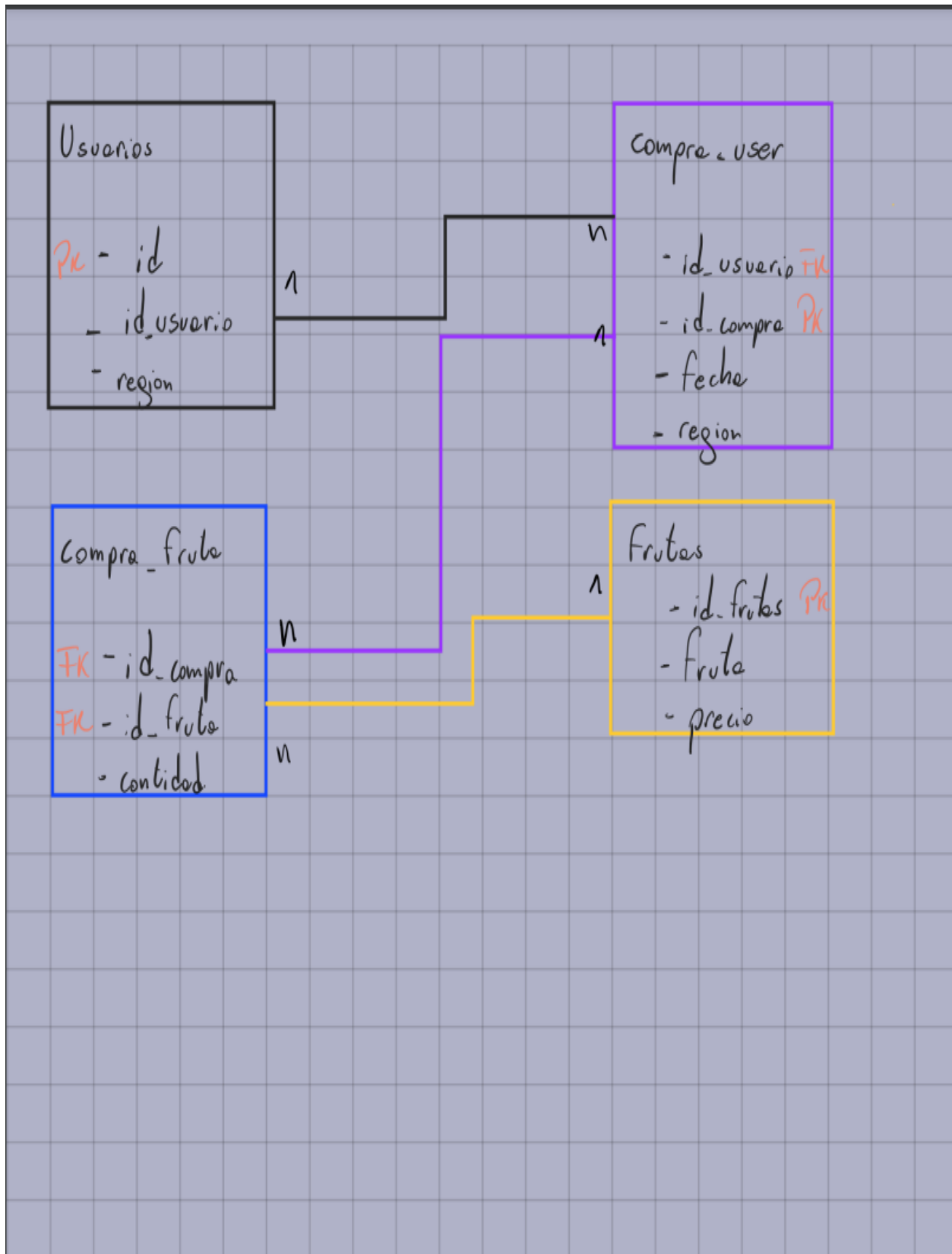


Figure 2: Esquema de las tablas de la parte 2