

BASES de DATOS

conceptos preliminares

Bases de datos No Relacionales

Instituto de Computación, Facultad de Ingeniería, UdelAR

2023

Contenido

- Introducción
- Conceptos Generales:
 - Dato, Información y Conocimiento
 - ¿Qué es un sistema?
- Sistemas de Información
 - Funciones
 - ¿Cómo se desarrollan?

Introducción

- El valor de los datos:
 - Los datos se han convertido en uno de los **activos más importantes de las organizaciones**
 - Son la materia prima para toma de decisiones operativas y estratégicas
 - Son lo que guía a procesos de todo tipo
- Distintos tipos de organizaciones necesitan los datos
 - Compañías de seguros, bancos, financieras, fábricas, proveedores, distribuidores. Sus actividades están guiadas por los datos.

Algunas definiciones

■ Datos

- Un parámetro o hecho, un número, una afirmación, una imagen
- Representan algo en el mundo real
- Son la materia prima para la producción de información

■ Información

- Datos con un significado en cierto contexto
- Datos relacionados
- Datos luego de su manipulación

■ Conocimiento

- Experiencia e información acumulada

El precio de la
harina es de \$50
por kg

El precio de la
harina subió un 5%
en el último año

Cuando el precio de la
harina sube un 5%, en un
mes el precio del pan y
sus derivados crece en un
10%

Sistema de Información

- Conjunto de componentes que interactúan con el objetivo de almacenar, recuperar y procesar datos e información para crear nueva información.
- Los componentes de un SI son software y hardware, pero es fundamental el rol de las personas.
- Ejemplo:
 - Primeros Sistemas de Información: Censos (de personas y/o bienes) Babilonios año 3800 a.c. !!!!

Sistema de Información

- Los **Sistemas de Información Informáticos** (SI) utilizan tecnologías informáticas para realizar algunas de sus tareas.
- Cumplen con 3 funciones principales:
 - **Memoria**: mantienen una representación del estado de cierto dominio
 - **Informativa**: proveen información acerca del estado de cierto dominio
 - **Activa**: realizan acciones que cambian el estado de cierto dominio



Supongamos un supermercado que basa sus sistemas de información en un conjunto de archivos y planillas de cálculo.

Algunos problemas



- La dirección de un proveedor se repite 2500 veces. ¿Qué se obtiene?
 - **Redundancia de datos NO controlada**
- La dirección de un proveedor se repite 2500 veces y es necesario actualizarla, entonces hay que cambiarla cada vez que se repite. Si no se actualizan todos los datos al mismo tiempo. ¿Qué se obtiene como resultado?
 - **Inconsistencia de datos**
- Los empleados acceden al mismo tiempo al archivo de datos del sector ventas modifican el precio de un producto al mismo tiempo, y el precio final no es el correcto. ¿Qué NO tiene el Supermercado?
 - **Control de concurrencia**

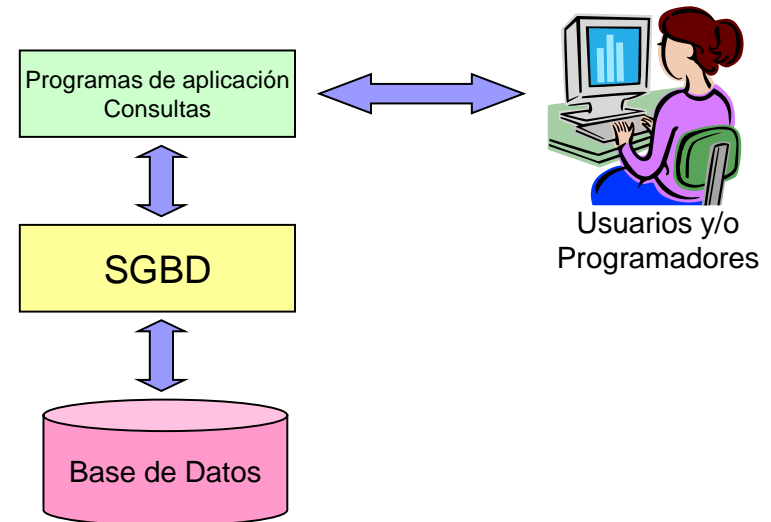
Más problemas



- La gerente necesita obtener una lista de todos los productos con los precios de venta y los precios de compra, y cada sector (de ventas y de compras), utiliza identificadores diferentes para cada producto. ¿Cuáles son los resultados en este caso?
 - **Aislamiento de los datos**
- Un empleado ve los archivos, se entera de las promociones y se las cuenta al supermercado de la competencia. ¿Con qué NO cuenta el Supermercado?
 - **Restricción de acceso NO autorizado**
- Una empleada trabaja en la actualización de los precios de venta y hay un apagón y no sabe si se pierden los datos. Además, no cuenta con una copia de los datos que le permitiría recuperar la información perdida. ¿Qué debía tener en cuenta el Supermercado?
 - **Respaldo y recuperación de los datos**

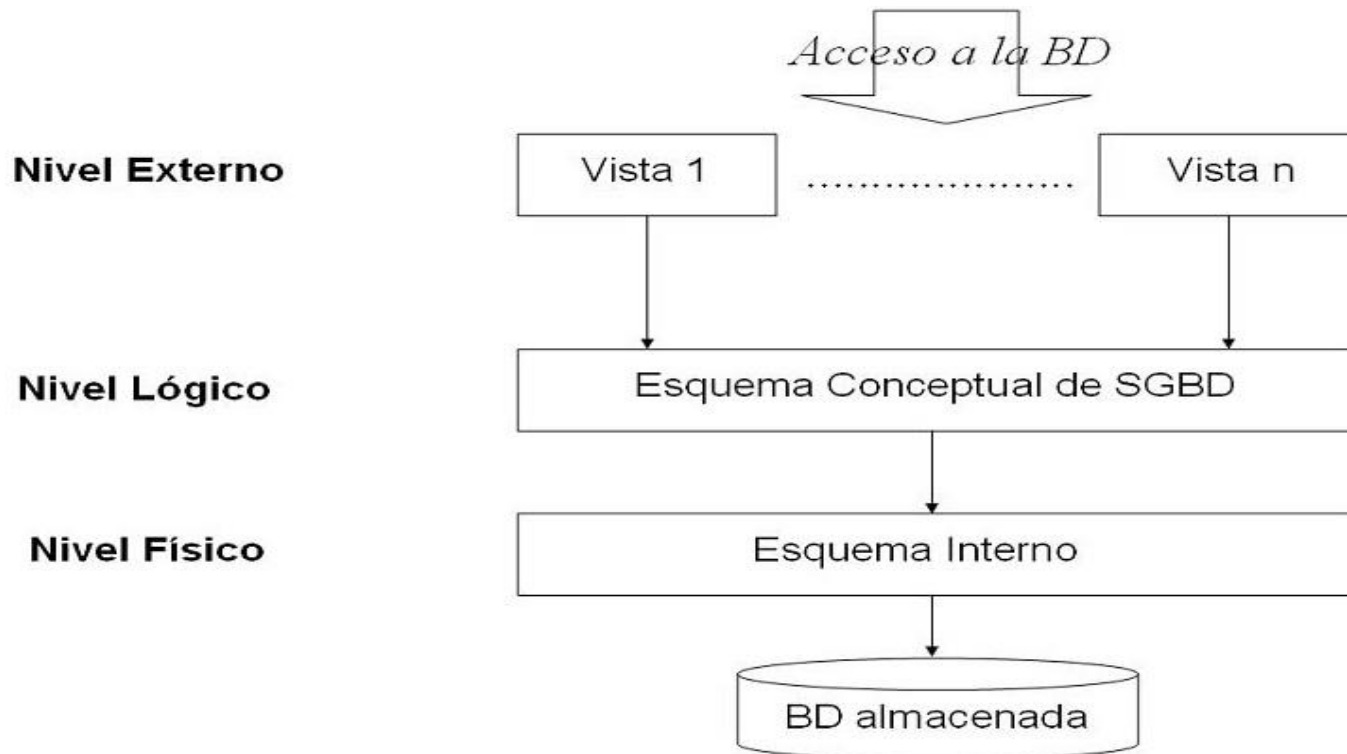
¿Cómo podría el Supermercado evitar todos estos problemas?

- Utilizando un SISTEMAS DE GESTIÓN DE BASES DE DATOS (SGBD)
- Un SGBD (o DBMS) es un software encargado de realizar la comunicación entre las personas, las bases de datos y otros softwares que utilizan los datos guardados en la Base de datos. Permite:
 - definir
 - construir y
 - manipular la base de datos.
- Un SISTEMA DE BASES DE DATOS
 - la Base de Datos (BD)
 - el Sistema de Gestión de Base de Datos
 - otros programas que consultan a la BD



Sistema de Base de Datos

- 3 niveles en el trabajo con BD:



Sistema de Base de Datos

■ Vistas

- Una BD tiene muchos usuarios (aplicaciones y/o personas), y cada uno de los cuales necesita información diferente.

■ Nivel externo

- Es el responsable de proveer las diferentes vistas a los diferentes usuarios.

■ Nivel lógico

- Tiene una visión más unificada de todos los datos. En este nivel se describen los datos y las relaciones entre ellos.

■ Nivel físico

- Es este nivel quien realmente accede a los datos y es el que determina cómo y dónde se almacenan los datos.

Modelo de datos: definición

- Son lenguajes usados para especificar y manipular BDs
- Un Modelo de Datos permite expresar:
 - **Estructuras:** Elementos de los problemas.
 - Ejemplo: CURSOS(nro_curso, nombre, horas).
 - **Restricciones:** Reglas que deben cumplir los datos para que la base sea considerada válida.
 - Ejemplo: (NO NULL c NO NULL CURSOS) (c.horas < 120)
 - **Operaciones:** Insertar, borrar y consultar la BD.
 - Ejemplo: Insert into CURSOS (1911,"FBD",90)

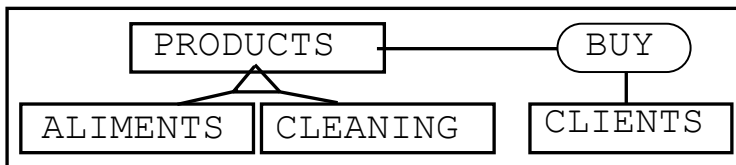
Modelo de Datos: Clasificación

Según el nivel de abstracción:

- Conceptuales
 - Representan la realidad independientemente de cualquier implementación de BD
 - Usado en etapa de Análisis
- Lógicos
 - Implementados en DBMSs
 - Usado en etapas de Diseño e Implementación
- Físicos
- Implementación de estructuras de datos
- Ejemplo: Árboles B, Hash.

Modelo de Datos: Aplicación

↓
Diseño Conceptual



Esquema Conceptual
(Ejemplo: Entidad-Relación)

↓
Diseño Lógico

```
CREATE TABLE PRODUCTS (...);  
CREATE TABLE CLIENTS (...);
```

Esquema Lógico
(Ejemplo: Relacional, Orientado-a-objetos)

Esquema de una Base de Datos

- Describe qué datos hay en la base, cómo se relacionan esos datos entre sí y qué restricciones de integridad deben cumplir
 - Estructuras + Restricciones
- Por ejemplo:
 - CURSOS (nro_curso, nombre, horas).
 - ESTUDIANTES (CI, nombre, fecha_nacimiento).
 - TOMA_CURSO (nro_curso, CI).
- Muy estables

Instancia de una Base de Datos

- Conjunto de datos almacenados en una base.
- Es el valor de base en un instante de tiempo.
 - Si respetan todos las restricciones, se considera que la instancia es correcta.
- Muy volátiles.
- Una instancia es un CONJUNTO DE ELEMENTOS.

Lenguajes e Interfases en Ambientes BD

- ¿Cómo interactuar con un SGBD? Existen lenguajes que me permiten:
 - Definir el esquema (crear, borrar, modificar)
 - Data Definition Language (DDL)
 - Manipular las instancias (crear, borrar, modificar)
 - Data Manipulation Language (DML)
 - Hacer consultas
 - Query Language (QL)
 - Crear vistas
 - View Definition Language (VDL)
- SQL (**Structured Query Language**) lenguaje que permite hacer todo esto sobre los modelos de datos relacionales (RDBMS)

EL MODELO RELACIONAL

Bases de datos No Relacionales

Instituto de Computación, Facultad de Ingeniería, UdelaR

2023

Contenido

- Conceptos generales
 - Ejemplos
- Restricciones de Integridad
- Operaciones de Modificación
 - Insert
 - Delete
 - Update

Conceptos generales

- Es un Modelo de Datos Lógico.
 - Se usa como Modelo implementado por DBMS.
- Creado por Codd en 1970.
 - Se comenzó con una definición teórica.
 - Se proponía un modelo con fuertes elementos matemáticos para BDs.
- Actualmente: uno de los modelos lógicos predominantes.

Conceptos generales

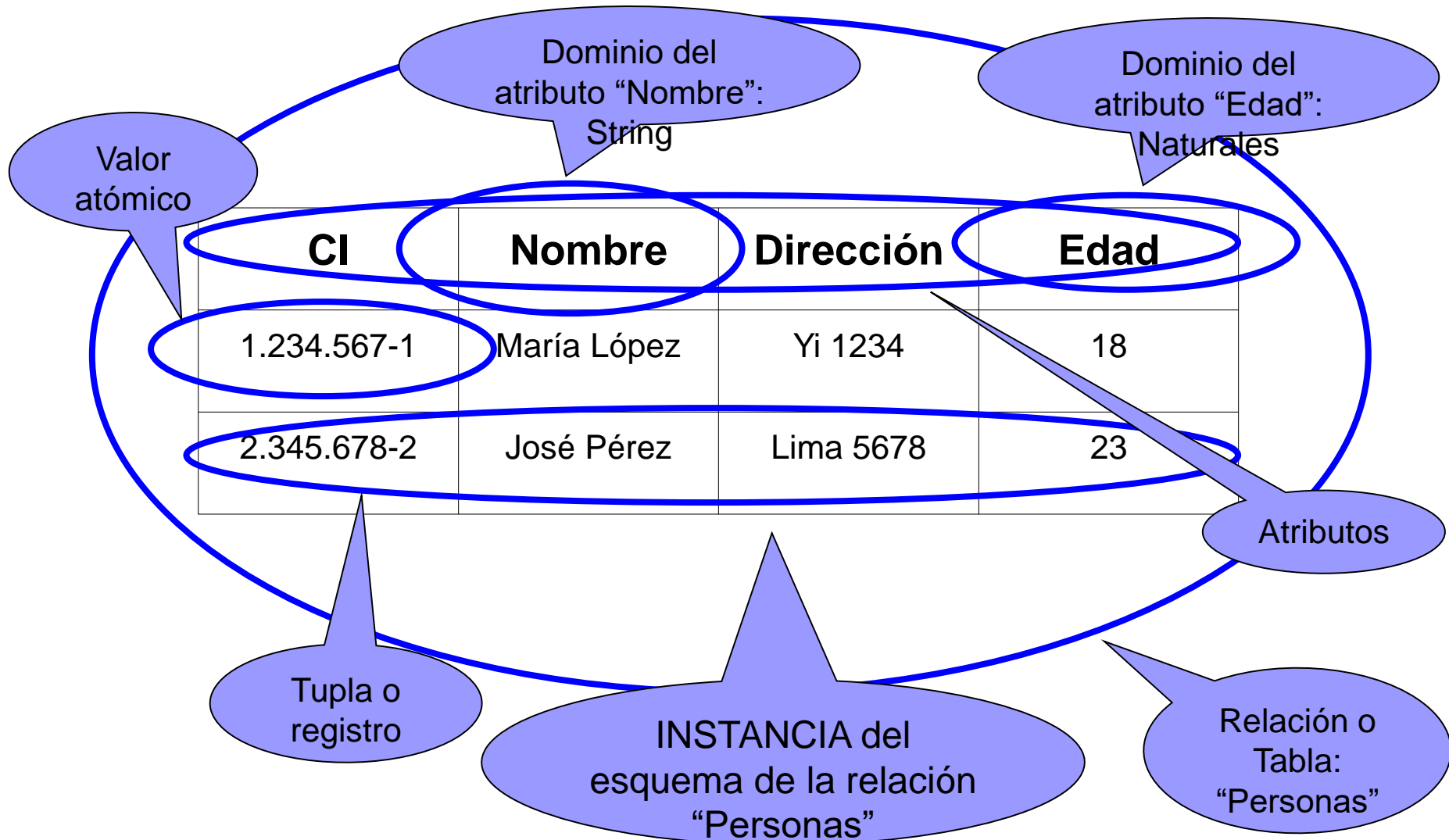
- Vision Informal del Modelo
 - Las estructuras consisten en TABLAS:
 - las columnas corresponden a ATRIBUTOS de tipo atómico.
 - las filas corresponden a registros de datos.
 - Las operaciones están fundamentalmente orientadas a manejo de TABLAS como conjuntos de registros.
 - Es un modelo de datos
 - extremadamente simple
 - potente para la mayor parte de las aplicaciones de BDs

Agreguemos formalidad

- Dominio D
 - Es un conjunto de valores atómicos
- Esquema de relación $R(A_1, \dots, A_n)$
 - R es el *nombre de relación*
 - A_1, \dots, A_n son los *atributos* con dominios D_1, \dots, D_n
- Relación $r(R)$
 - Es una instancia de un esquema de relación R
 - Consiste en un conjunto de t-uplas (o **tuplas**)
 - $r = \{ \langle a_1, \dots, a_n \rangle, \langle b_1, \dots, b_n \rangle, \langle c_1, \dots, c_n \rangle, \dots \}$
 - También puede interpretarse a r como:
 - $r(R) \subseteq (D_1 \times \dots \times D_n)$

⊆

Conceptos generales. Ejemplos



Conceptos generales. Ejemplos

- Esquema de relación:
 - FABRICANTES(#fabricante, nombre, direccion)
- Dominio de los atributos:
 - nombre y dirección son del tipo STRING
 - #fabricante es un NUMERO
- Una instancia de la relación FABRICANTES:

#fabricante	Nombre	Dir
12	Ana Pérez	Avda. A 2134
24	Luis Rodriguez	Calle B 1445

- Una tupla de la relación FABRICANTES:
 - <“12”, “Ana Pérez”, “Avda. A 2134”>

Conceptos generales

- Esquema de BD Relacional o Esquema Relacional:
 - Conjunto de esquemas de relación
- Ejemplo de esquema de BD relacional:

FABRICANTES(#fabricante, nombre, direccion)

PRODUCTOS(#producto, descripcion)

VENTAS(#producto, #fabricante, precio)

Conceptos generales. Ejemplos

- Instancia de BD Relacional

#fabricante	nombre	direccion
1	Juan	Avda. A 1234
2	Pedro	Calle C 987
3	María	Avda. B 2345
4	Ana	Calle C 987
5	Pepe	Calle C 900
6	Juan	Avda. A 1230

FABRICANTES

#producto	descripcion
P1	Jabón
P2	Pintura
P3	Pintura
P4	Detergente
P5	Limpiador

PRODUCTOS

#producto	#fabricante	precio
P1	1	200
P2	1	150
P2	4	140
P3	2	120
P3	4	140
P5	6	300

VENTAS

Conceptos generales

- Características de Relaciones
 - Son conjuntos de tuplas que:
 - No están ordenados
 - No hay repetidos
 - Valores de Atributos en tuplas:
 - Son valores atómicos (indivisibles)
 - Propiedad: *primera forma normal* (1NF)

Restricciones de Integridad

- Las restricciones de integridad (RI) se especifican junto con el esquema de la base de datos
- El responsable de que se cumplan es el SGBD
- Restricciones de Dominios:
 - Restricciones de tipo en los Di
 - El asociar un dominio a cada atributo restringe el conjunto de valores que puede tomar ese atributo
- Ejemplo:
 - FUNCIONARIO(CI, Nombre, Dir, Edad).
 - CI: number(9);
 - Nombre, Dir: String;
 - Edad: number(2); Edad > 18;

Restricciones de Integridad

■ Superclave y Clave:

- Dado $R(A_1, \dots, A_n)$, se dice que $X \subseteq \{A_1, \dots, A_n\}$ es **superclave** en un esquema R , si no puede existir ninguna $r(R)$ tal que tenga dos tuplas con valores iguales de X ($t[X] = t'[X]$)
- Una **clave** es una *superclave* que no contiene propiamente una *superclave* (o sea minimal)
- Ejemplo: #fabricante es clave en FABRICANTES

Restricciones de Integridad

- Integridad Referencial y Foreign Keys (FK):
 - Dado R , un conjunto de atributos X es una FK de R si:
 - Los atributos de X coinciden en dominio con los de una clave Y de S
 - Los valores de X en tuplas de $r(R)$ (para toda r) corresponden a valores de Y en la relación $s(S)$
 - Ejemplo:
 - `VENTAS.#fabricante FK FABRICANTES.#fabricante`

Restricciones de Integridad

- Una BD se considera válida si todas las relaciones r satisfacen las RIs.
 - Todas las **instancias** actuales de todas las relaciones declaradas en el esquema relacional satisfacen todas las RIs.
- Las Restricciones de Integridad surgen de la observación de la realidad, NO de la observación de relaciones.
- Las RI se definen a nivel del ESQUEMA RELACIÓN, NO a nivel de instancia.
- Las RI son violadas por las instancias, NO por los esquemas de relación.

Operaciones de Modificación

- Sea el esquema relacional $R(A,B,C)$ y la relación $r(R)$
- Insert
 - **insert** $\langle a,b,c \rangle$ **into** R
 - Incluye la tupla $\langle a,b,c \rangle$ en la relación r
 - Las tuplas insertadas deben cumplir las RI
- Delete
 - Delete from R where $A="a"$
 - Borra de las tuplas de r cuyo valor para A es "a"
 - Borrar tuplas puede generar violaciones a RI. ¿En qué casos ?

Operaciones de Modificación

- Sea el esquema relacional $R(A,B,C)$ y la relación $r(R)$
- Update
 - **update R set A = "a1" where B = "b"**
 - modifica las tuplas de r cuyo valor de B es "b", colocando "a1" como valor de A
 - Actualizar tuplas puede generar violaciones a RI. ¿En qué casos ?

Referencias

- “Fundamentals of Database Systems 3th Ed.”, Elmasri & Navathe, Addison Wesley
- *Curso Fundamentos de Bases de Datos, INCO*, <http://www.fing.edu.uy/inco/cursos/bdatos/>