



Estructuras de Almacenamiento de Datos

Cursada 2010

Facultad de Ciencias Exactas Universidad Nac. Centro de la Pcia. de Bs. As.



Introducción a la Materia

Objetivo:

- Realizar una introducción al diseño y organización de los datos en memoria secundaria.
- Se espera básicamente que:
 - Se adquieran conocimientos del modelado conceptual de datos y su importante rol dentro del ciclo de desarrollo de software.
 - Se adquieran conceptos teóricos necesarios para estructurar y acceder eficientemente a grandes volúmenes de datos.

Introducción

Dos enfoques esencialmente diferentes para organizar los datos

→ Estructuras de Almacenamiento específicas

→ Sistemas de Archivos (SA) (enfoque convencional)

Un **Sistema de Archivos** es un conjunto de tipo de datos abstractos que son implementados para el almacenamiento, la organización jerárquica, la manipulación, el acceso, el direccionamiento y la recuperación de datos; esto tiene mucho en común con la tecnología de bases de datos.

→ **Bases de Datos (BD)**

→ Sistemas de Archivos + DBMS (enfoque actual)

(DBMS - DataBase Management System o SGBD - Sistema de Gestión de Bases de Datos)



Cronología

Sistemas de archivos (<'70) →

→ BD Red y Jerárquicas (<'70)

→ BD Relacionales - BDR ('70 –'90)

→ BD Orientadas a Objetos BDOO y
BD Objeto-Relacionales BDOR ('90
actualmente)

Introducción

- Las primeras aplicaciones
 - manejo de datos simples (sueldos, stock, etc)
- Actualmente
 - sistemas de archivos para grandes volúmenes de datos
 - para algunos contextos son casi obsoletos pero ...
- ... su estudio brinda las siguientes VENTAJAS:
 - Perspectiva histórica → evolución
 - Conocimiento de problemas y fallas que ocurren en SA
 - evitarlos en BD
 - Conocer las características de SA (simples)
 - facilitan el entendimiento de BD (complejas)
 - Migrar SA a BD exige conocer limitaciones y posibilidades
 - Comprensión de la forma de almacenamiento de los datos para entender cómo los administra el DBMS
 - No todas las aplicaciones requieren el uso de una base de datos (aunque la disponibilidad de una de ellas induce su utilización, a pesar de que su mantenimiento, configuración, etc. a veces dificulta más de lo que aporta)

Sistemas de Archivos

- Operaciones básicas sobre archivos: modificaciones, altas, bajas, consultas → programas en lenguaje específico
- Crecimiento del sistema de archivos (incorpora más archivos y nuevas relaciones entre los existentes)
 - nuevos programas para su administración:
 - Creación de archivos
 - Población de archivos
 - Listado total o parcial del contenido
 - Borrado o modificación de datos
 - Búsqueda de datos dispersos
 - Etc...
- Datos dispersos (administrados independientemente)
 - cómo sincronizar el procesamiento de datos en distintos archivos?



Sistemas de Archivos

- Dificultades en la elaboración de aplicaciones que permitan salvaguardar la información (contraseñas, zonas de exclusión, confidencialidad, ...)
→ medidas de seguridad y protección insuficientes !
- Dificultades para compartir los datos → noción de propiedad de los datos
→ duplicación de datos y múltiples problemas asociados con la redundancia
- Dificultades para manejar cambios en la estructura de los archivos
→ propagación de las actualizaciones a todo el software involucrado !

PROBLEMA BÁSICO: DEPENDENCIA ESTRUCTURAL

Bases de Datos

En SA (resumen):

- Redundancia de los datos →
 - Carencias expresivas
 - Inconsistencia
 - Anomalías de inserción, borrado y actualización
- Formatos de archivos incompatibles

En BD:

- **Problemas resueltos**
 - definición de los datos fuera de los programas de aplicación
 - control y acceso a los datos independiente de los programas de aplicación

SOLUCIÓN BÁSICA: INDEPENDENCIA DE LOS DATOS

Datos - Información

- Inteligencia del “negocio” o Business Intelligence (BI) se define como los procesos, tecnologías y herramientas necesarias para convertir los datos en información, la información en conocimiento, y el conocimiento en los planes que direccionan las acciones de un negocio rentable.





Diseño de una estructura de datos

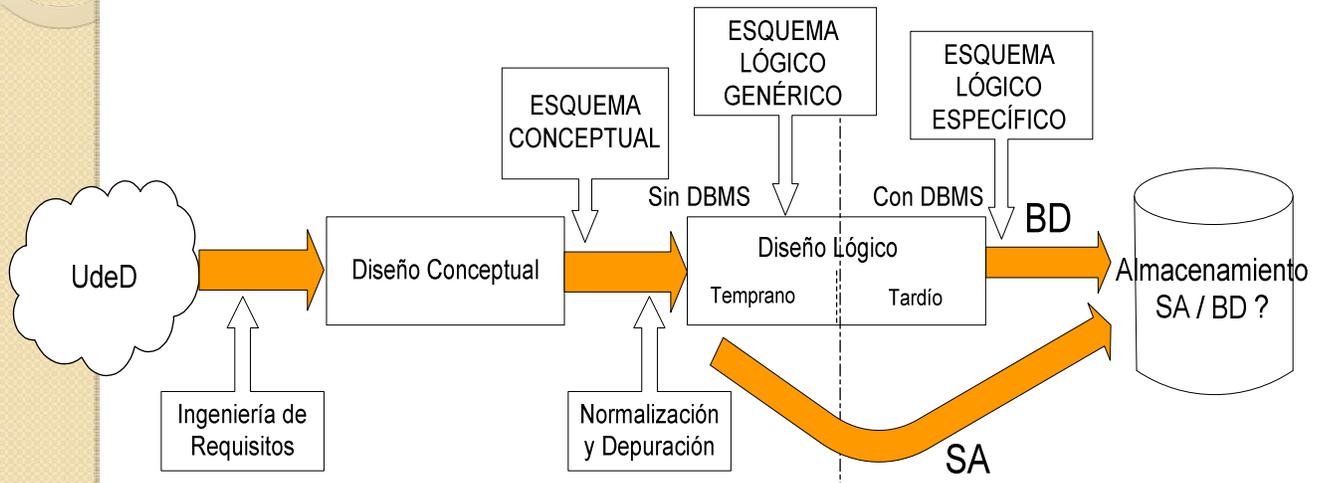
- Se puede dividir en 5 niveles:
 - *identificación de los datos de la realidad*
 - *abstracción de datos*
 - *identificación de relaciones entre datos*
 - *representación de datos*
 - *codificación*



Diseño de un Modelo de Datos

- Involucra los tres primeros pasos, si bien en esta materia comenzaremos a partir del 2do. paso.
- Encontrar la parte del mundo real que es necesario modelar para construir el sistema de información (Ingeniería de Requisitos)
- **Cómo lograr la representación de un sistema del 'mundo real' de manera que pueda ser manejado en un 'mundo informático' (o sea mediante un SA ó una BD)?**
- **Construir un modelo de datos**
 - Formado por un conjunto de conceptos útiles para definir la estructura de los datos a representar y relaciones entre ellos
 - Proporcionan cierto nivel de abstracción de la realidad

Ciclo de Diseño



Ciclo de Diseño

Análisis de los Requisitos:

Datos y requisitos del modelo de datos

Diseño Conceptual:

- *Modelo Conceptual de Datos (MCD)*
- *Para describirlo → objetos del mundo real*
 - + *vínculos semánticos entre ellos*
 - + *descripción de ambos*
- *Metodologías más difundidas:*
 - Entidades y Relaciones –MER*
 - Unified Modeling Language – UML*
 - Object-Role Modeling – ORM*
 - otras*

Ciclo de Diseño (cont.)

Diseño Lógico (DL)

DL Temprano: diseño de de los datos sobre una plataforma específica (transformación del modelo de datos a una especificación algebraica relacional)

Refinamiento del Esquema: analizar y depurar la colección de relaciones → buscar una representación normalizada → **Normalización**

DL Tardío:

MCD 

Diseño Físico: se especifican las estructuras de almacenamiento de datos internas y la organización de los archivos de *índices*, *archivos de datos en un SA* ó *en una BD*

→ *construcción de un sistema de archivos e índices*

Etapas de Modelado Conceptual



Análisis de requisitos:

- estudiar las reglas de la organización mediante entrevistas a usuarios, etc.
- determinar qué modelar
- obtener un esquema descriptivo de la realidad
- puede requerir realimentación y refinamientos



Clasificación de los Modelos de Datos

- Los distintos modelos propuestos se pueden clasificar según el tipo de conceptos que ofrecen para describir la estructura de la base de datos.
 - Modelos conceptuales (de alto nivel): proporcionan conceptos cercanos a cómo el usuario percibe los datos
 - Modelos de implementación (nivel intermedio): sus conceptos pueden ser entendidos por usuarios finales aunque representan la organización de datos en computadora
 - Modelos físicos (de bajo nivel): ofrecen conceptos para describir detalles de almacenamiento.



Modelos de Datos Conceptuales

- Se destaca el Modelo de Entidades y Relaciones (Modelo E/R - MER), propuesto por Chen en dos artículos ya históricos, 1976 y 1977.
- Según Chen, “El Modelo E/R puede ser usado como una base para una vista unificada de los datos”, adoptando “el enfoque más natural del mundo real que consiste en entidades y relaciones (interrelaciones)”.
- Posteriormente otros autores lo han extendido con importantes aportes, lo que ha dado lugar a una familia de Modelos de Datos.



Modelo de Entidades y Relaciones (MER)

- El MER ha tenido una gran difusión en la comunidad informática dedicada a las bases de datos → ha sido el modelo más extendido en las herramientas CASE de ayuda al diseño de BD.
- Es simple y poderoso para modelar abstracciones del mundo real
- Es una representación gráfica de un modelo de datos, fácilmente traducible a un esquema de BD:
Reglas de transformación → esquema relacional
- Se ha convertido en un estándar 'de facto', incluso muchas herramientas de diseño de BD utilizan sus conceptos



Conceptos Básicos del MER

En el Modelo E/R, tal como fue propuesto por Chen, se distinguen los siguientes elementos estáticos:

- Entidad (entity)
- Relación (o interrelación) (relationship)
- Dominio (domain)
- Atributo (attribute)

Representación de elementos básicos del MER

Diagrama de Entidades y Relaciones

Entidades



nombre entidad

Atributos



Nombre atributo (con variaciones en la gráfica de acuerdo a su tipo)

Relaciones entre entidades



nombre relación



Elementos Básicos del MER: ENTIDADES

Definiciones de Entidad:

- “Cualquier objeto (real o abstracto) que existe en la realidad y acerca del cual queremos almacenar información en la base de datos”.
- “Algo con realidad objetiva que existe o puede ser pensado”.
- “Una persona, lugar, cosa, concepto o suceso, real o abstracto, de interés para la empresa”.

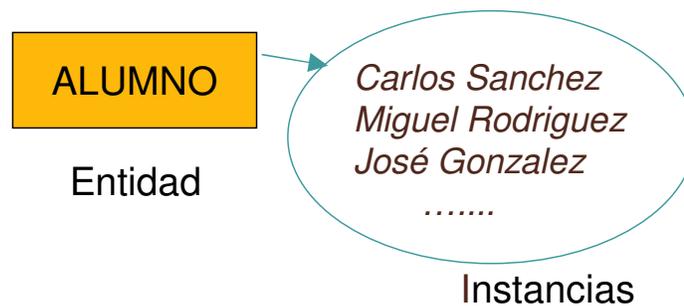
Entidades (cont.)

Teniendo en cuenta la abstracción de las características comunes de un conjunto de ejemplares debemos distinguir entre:

- Conjunto Entidad o Entidad: El tipo de entidades o estructura genérica que describe un conjunto de entidades aplicando dicha abstracción.
- Ejemplo: todos los Alumnos.
 - Todas las entidades de un conjunto tienen el mismo conjunto de atributos.
 - Cada conjunto de entidades tiene una clave (K).
 - Cada atributo tiene un dominio de definición.
- Las entidades o ejemplares o instancias de ese tipo de entidad; por tanto, el tipo de entidad es el resultado de la clasificación de un conjunto de entidades. Objeto del mundo real, distinguible entre otros. Se describe por medio de sus propios atributos. Ejemplo: un alumno en particular, Carlos Sanchez

Entidades (cont.)

- Puede tener existencia física (Ej: un automóvil, un empleado) o ser un objeto conceptual (Ej: una categoría, una materia)
- Los objetos individuales son instancias de la entidad



Entidades (cont.)

Existen dos categorías de tipos de entidades:

- **Regulares o fuertes**, que son aquellas cuyos ejemplares tienen existencia por sí mismos (como ALUMNO y LIBRO), y
- **Débiles**, en las cuales la **existencia e identificación** de un ejemplar dependen de un cierto ejemplar de otro tipo de entidad, por ejemplo, la existencia e identificación de un EJEMPLAR depende de un LIBRO, (otro ejemplo: los postes de alumbrado en las rotondas de nuestras rutas).

Entidades (cont.)

...problema para identificarlas

- Uno de los problemas que existirán en el diseño E/R es la decisión de si un determinado objeto o concepto se modela como una entidad o no.
- Por ejemplo, el color es habitualmente una propiedad de una entidad (como es el caso del color de un coche), pero en una fábrica de pinturas probablemente sería apropiado modelar el color como una entidad con sus propias propiedades.

Para precisar el concepto de entidad → tres reglas generales que debe cumplir una entidad:

- tiene que tener existencia propia,
- cada ejemplar de un tipo de entidad debe poder distinguirse de los demás,
- todos los ejemplares de un tipo de entidad deben tener las mismas propiedades.

Entidades (cont.)

Pero ...

- La primera de estas reglas no es aplicable a las entidades débiles.
- La segunda supone la obligación de un identificador que permita distinguir los distintos ejemplares de un tipo de entidad, lo que tampoco es universalmente aceptado (ni por algunos autores, ni por los modelos, ni por los productos).
- La tercera es relativa: ¿exactamente las mismas?, ¿las mismas entre las que nos interesan?, ...

Elementos Básicos del MER: ATRIBUTOS

- Son los datos relativos a una entidad o relación
- Es una característica (adjetivo) que puede:
 - Identificar y/o
 - Relacionar y/o
 - Describir
- Cada atributo tiene asociado un **Dominio de definición** (entero, cadena de caracteres, fechas, etc.) y puede tomar un cierto valor dentro del dominio
- Un atributo puede ser:
 - simple (*ej. Nombre, nro.de documento, etc.*)
 - compuesto: se pueden dividir en componentes más pequeños, su valor es la concatenación de los valores de los atributos simples (*ej. Dirección: calle, número, piso, dpto.*)
 - univaluado (*ej. Edad*)
 - multivaluado: puede tener un conjunto de valores para una misma instancia (*ej. Teléfono, pueden ser varios números*)

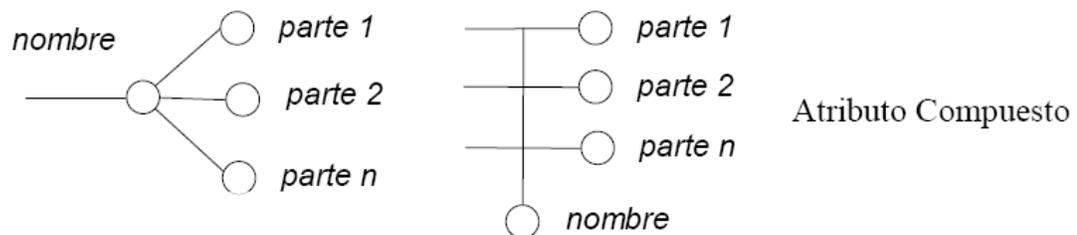
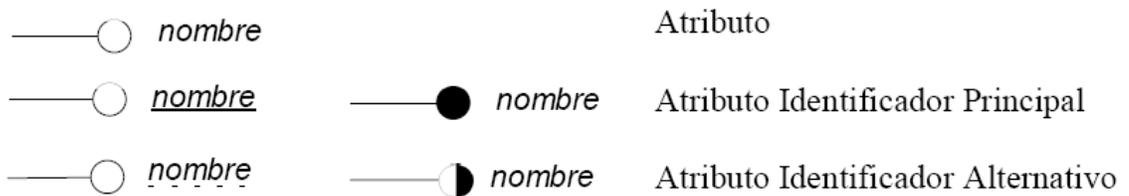
Atributos (cont.)

- Todas las instancias de una entidad se describen mediante el mismo conjunto de atributos ... (3a. propiedad de entidades)
- (Casi) Siempre hay un atributo cuyo valor siempre es distinto para cada instancia particular de una entidad (o relación) → **atributo clave o identificador** (ej. *Número de libreta de un Alumno, patente de un Automóvil*) ... (2a. propiedad de entidades)
- Algunas entidades pueden tener más un atributo clave → **clave alternativa** (ej. *documento del Alumno*) ... (2a. propiedad de entidades)
- Los atributos de una entidad pueden ser **obligatorios** (*deben tener un valor*) u **opcionales** (*pueden ser nulos, es decir no contener valor*)

Atributos (cont.)

Los atributos se colocan junto a las entidades.

- Con simbología específica según su tipo.
- Hay otras formas de notación aceptables.



Atributos (cont.)

$\overset{D}{\text{-----}} \circ \text{ nombre}$ Atributo derivado (*D contiene la fórmula de derivación*)

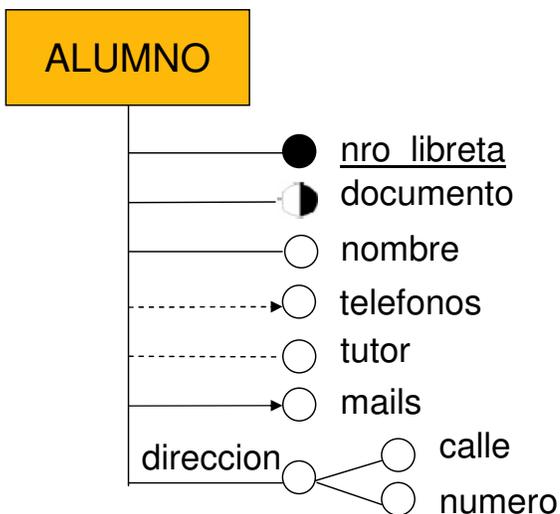
$\text{-----} \circ \text{ nombre (1,1)}$ $\text{-----} \circ \text{ nombre}$ Atributo obligatorio y univaluado

$\text{-----} \circ \text{ nombre (0,1)}$ $\text{-----} \circ \text{ nombre}$ Atributo opcional y univaluado

$\text{-----} \circ \text{ nombre (1,n)}$ $\text{-----} \rangle \circ \text{ nombre}$ Atributo obligatorio y multivaluado

$\text{-----} \circ \text{ nombre (0,n)}$ $\text{-----} \rangle \circ \text{ nombre}$ Atributo opcional y multivaluado

Atributos (cont.)



Asumiremos que:
Cada conjunto de entidades debe tener al menos una clave dado que existe la necesidad de identificar cada ejemplar dentro del conjunto



Elementos Básicos del MER: RELACIONES O INTERRELACIONES

Se entiende por relación (interrelación) una asociación, vinculación o correspondencia entre entidades.

Igual que en el caso de las entidades, distinguiremos entre:

- **Conjunto Relación:** *el tipo de relación o estructura genérica que describe un conjunto de relaciones, y*
- **Cada relación,** *o instancia de relación es decir, cada uno de los ejemplares concretos.*
- CURSA es un tipo de relación que vincula los tipos de entidad ALUMNO y MATERIA; un ejemplar del tipo de relación CURSA es la vinculación entre el alumno “Carlos Sánchez” y el curso “Estructuras de Datos” dado que satisface la frase “Carlos Sanchez cursa la materia Estructuras de Datos”.

Relaciones (cont.)

Elementos de un tipo de relación

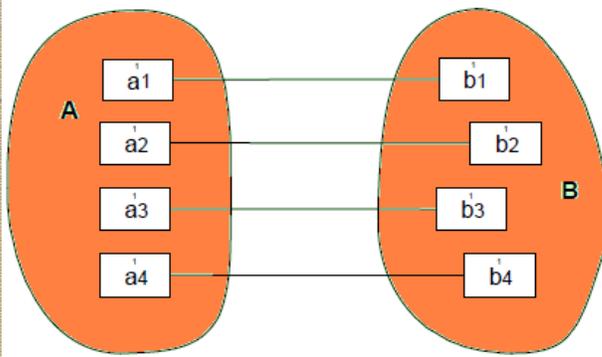
- Nombre (único en el esquema)
- Grado u orden (número de tipos de entidades participantes)
- Tipo de correspondencia o cardinalidad o multiplicidad (1 a 1, 1 a muchos, muchos a muchos, ...)

Subrayados los términos más usuales

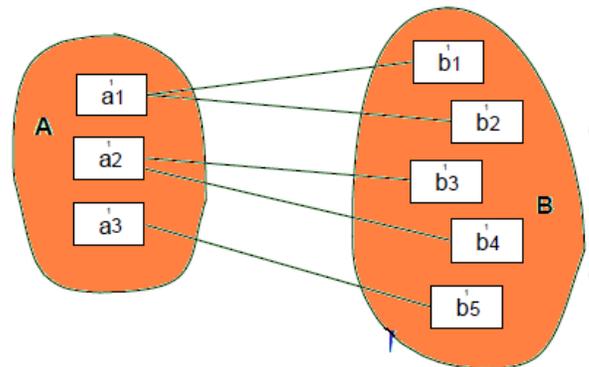
Relaciones - Grado

- Una interrelación de orden n (n -aria) R relaciona n conjuntos de entidades $E_1 \dots E_n$.
- Cada tupla en R involucra las entidades $e_1 \in E_1, \dots, e_n \in E_n$
- Entonces si:
 - $n=1$ la relación se denomina unaria;
 - $n=2$ la relación se denomina binaria;
 - $n=3$ la relación se denomina ternaria;
 - ...

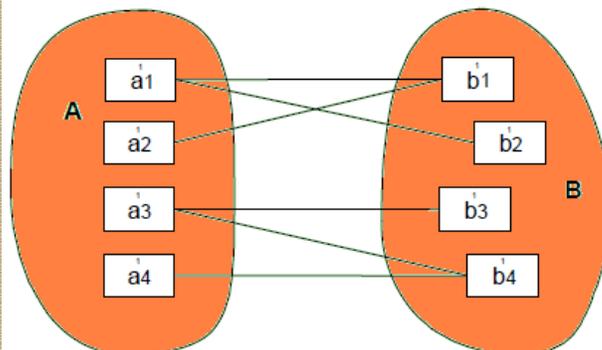
Relaciones – Tipos de correspondencia



Uno a uno – 1 : 1



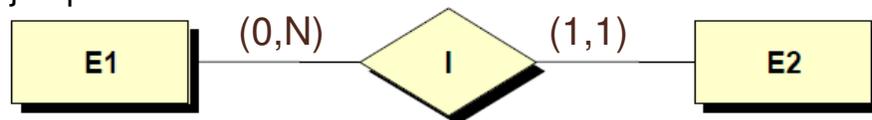
Uno a muchos – 1 : N



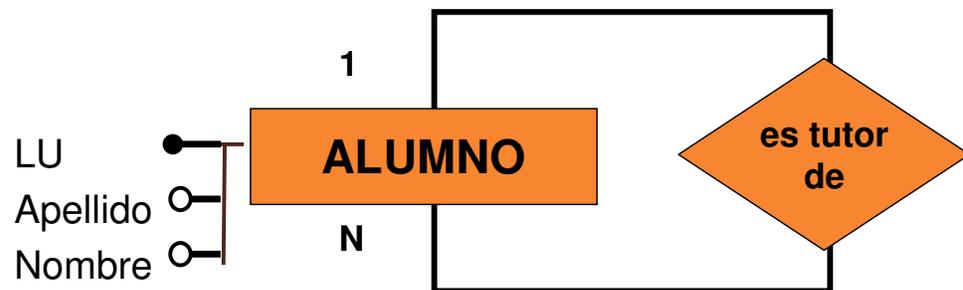
Muchos a muchos – N : N

Relaciones – Cardinalidades

- Es el nro. máximo y mínimo de ejemplares de una entidad que pueden estar relacionadas con ejemplares de otra u otras entidades.
- Esta información se coloca sobre los vínculos (líneas), en el ejemplo encerrado entre paréntesis.
- Lectura Look-Across o Chen-Style: se lee sobre la línea de la 'entidad destino'
- La **cardinalidad máxima** representa el máximo número de ejemplares de una entidad con los que se puede relacionar otra entidad: **al menos 1, como máximo N** (muchos o varios, es variable)
- La **cardinalidad mínima** también representa información valiosa:
 - Un ejemplar de una entidad **puede** estar relacionado con otro: **Cardinalidad mínima 0**
 - Un ejemplar de una entidad **debe** estar relacionado al menos con un ejemplar **Cardinalidad mínima 1**



Relaciones Unarias, reflexivas o recursivas

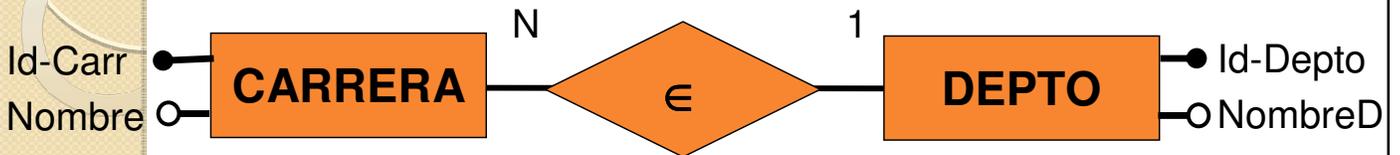


Cual es la semántica de esta relación?

Cada alumno **puede** tener **un** tutor

Cada alumno **puede** ser tutor **de otros** alumnos

Relaciones Binarias (1:N)



Cual es la semántica de esta relación?

Cada carrera pertenece a un único departamento

Cada departamento tiene muchas carreras

<u>Id-Carr</u>	Nombre	<u>Id-Depto</u>
1	Ing. en Sist.	11
2	Lic. Fisica	22
3	Anal. Prog	11

UN

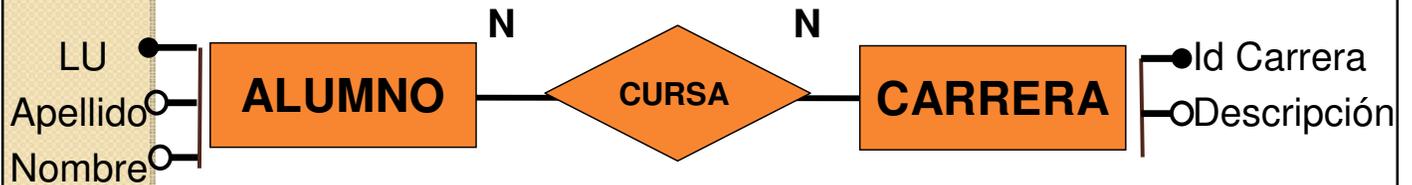
<u>Id-Depto</u>	NombreD
11	Comp.y SS.
22	Física
33	Matem.

Relaciones Binarias (N : N)

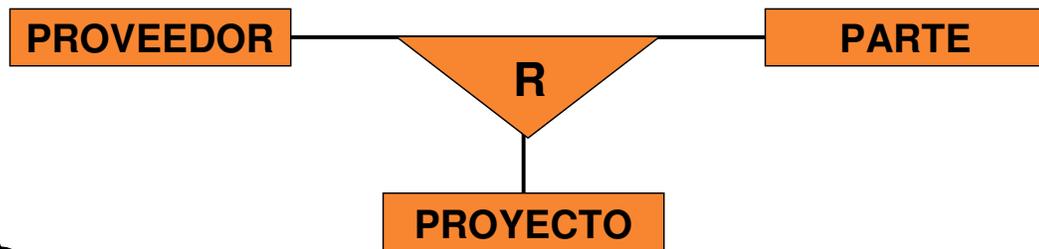
<u>LU</u>	Nombre	Apellido
111	Carlos	Sanchez
222	Miguel	Rodriguez
333	José	Gonzalez

<u>IdCarrera</u>	NombreD
A	Ing.en SS.
B	Lic. Física
C	Lic.Matem.

<u>LU</u>	<u>IdCarrera</u>
111	A
222	B
222	C



Relaciones Ternarias



Cual es la semántica de esta relación?

¿Cual es el nro. máximo de ejemplares de cada entidad que está vinculado con un par de ejemplares de la otra entidad?

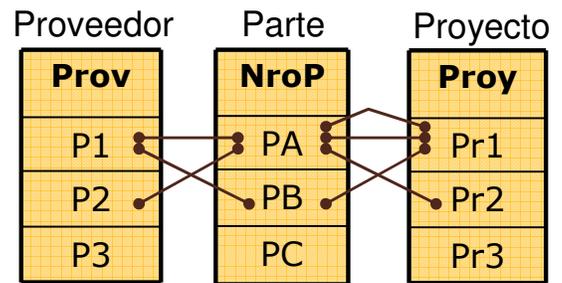
Por ejemplo: un proveedor, ¿cuantas partes provee para cada proyecto (una o muchas)?

Alternativas Diferentes: N:N:N 1:1:N 1:N:N 1:1:1

Relaciones Ternarias

Cardinalidad N:N:N

- Un proveedor provee cada parte a **muchos** proyectos.
- Un proveedor provee para cada proyecto **muchas** partes.
- Una parte, para cada proyecto es provista por **muchos** proveedores.



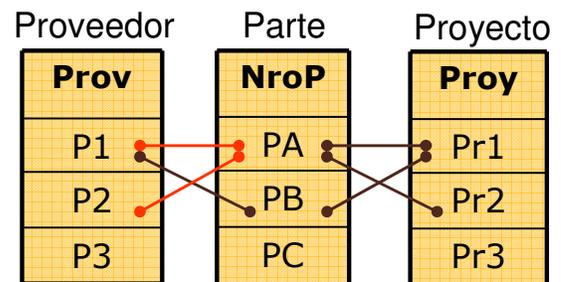
R

Prov	NroP	Proy
P1	PA	Pr1
P1	PB	Pr1
P1	PA	Pr2
P2	PA	Pr1

Relaciones Ternarias

Cardinalidad 1:N:N

- Un proveedor provee cada parte a **muchos** proyectos.
- Un proveedor a cada proyecto le provee **muchas** partes.
- Cada parte provista a un proyecto lo hace **sólo UN** proveedor.



R

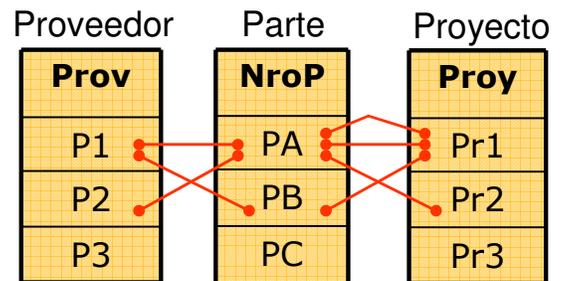
Prov	NroP	Proy
P1	PA	Pr1
P1	PB	Pr1
P1	PA	Pr2
P2	PA	Pr1

Este par de tuplas no corresponde, la relación es 1:N:N: sólo una de las tuplas es correcta

Relaciones Ternarias

Cardinalidad 1:1:N

- Un proveedor provee cada parte a **muchos** proyectos.
- Un proveedor a cada proyecto le provee **UNA** única parte.
- Cada parte provista a un ?? proyecto lo hace **sólo UN** proveedor.



R

Prov	NroP	Proy
P1	PA	Pr1
P1	PB	Pr1
P1	PA	Pr2
P2	PA	Pr1

Hay tuplas que no corresponden, la relación, sólo una de cada par es correcta...

Relaciones Ternarias

Cardinalidad 1:1:1

- Un proveedor provee cada parte a **UN solo** proyecto.
- Un proveedor a cada proyecto le provee **UNA única** parte.
- Cada parte es provista a un proyecto por **sólo UN** proveedor.

Proveedor

Prov
P1
P2
P3

Parte

NroP
PA
PB
PC

Proyecto

Proy
Pr1
Pr2
Pr3

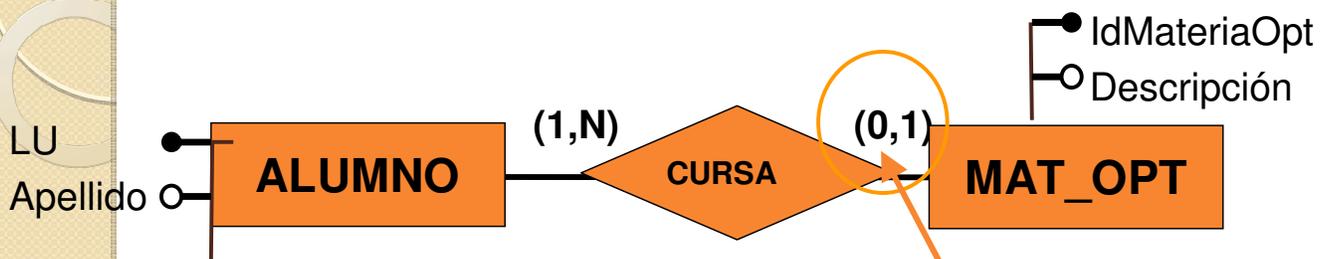
R

Prov	NroP	Proy
P1	PA	Pr1
P1	PB	Pr1
P1	PA	Pr2
P2	PA	Pr1

Algunas de estas tuplas no corresponden, la relación es 1:1:1

Relaciones Opcionales

Cardinalidades mínimas



Cual es la semántica de esta relación?

Un alumno puede estar **NO ESTAR** cursando ninguna optativa !!

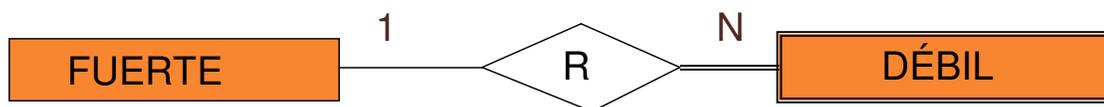
<u>LU</u>	Apellido	IdMateriaOpt
22334	Date, José	11
23456	Ka, Luis	22
34567	Pla, Juan	nulo

⊆

<u>IdMateria Opt</u>	Descripcion
11	Ing.en Sist.
22	Lic. Mat.
33	Analista

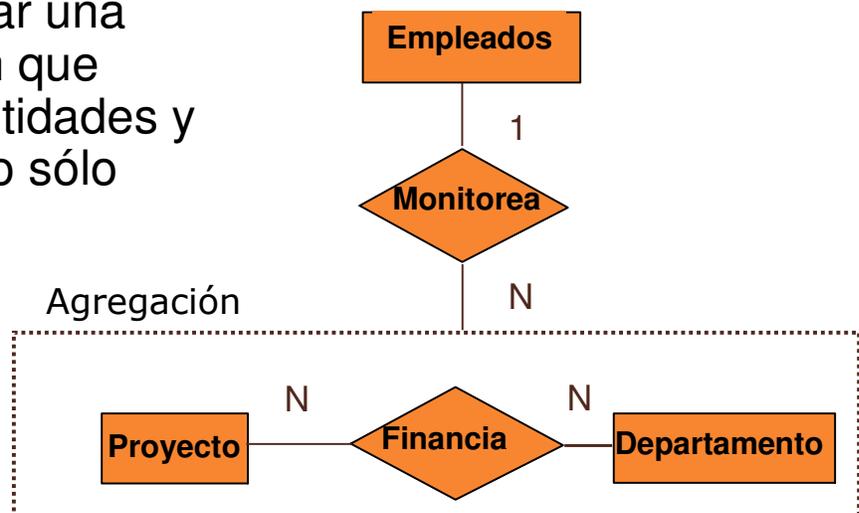
Entidades Débiles vs. Entidades Fuertes

- Una *entidad débil* puede ser unívocamente identificada **sólo** considerando la clave de otra entidad propietaria.
- Propietaria y débil están vinculadas por una relación 1:N.



Agregación

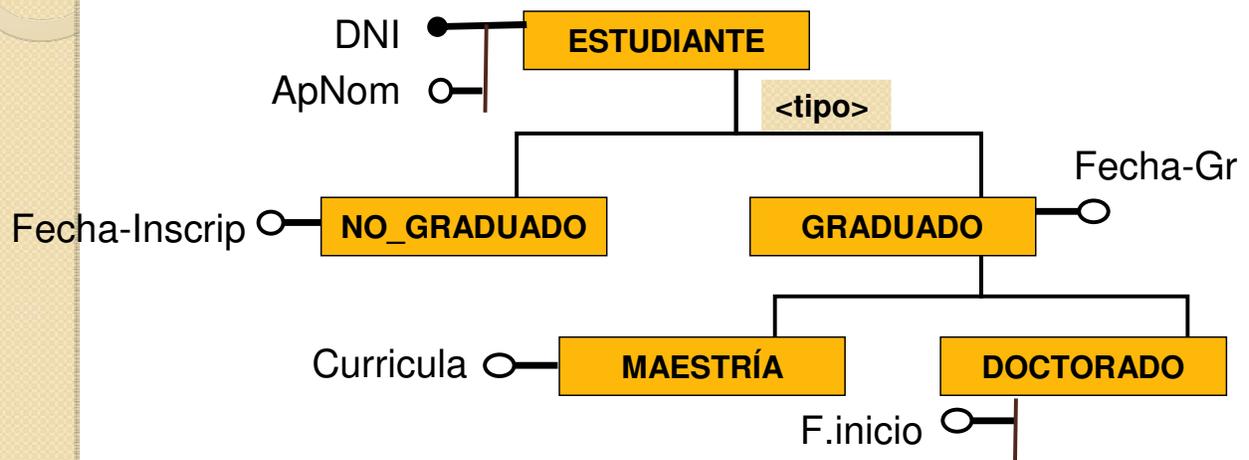
- Para modelar una interrelación que involucra entidades y relaciones, o sólo relaciones.



La interrelación se trata como si fuese otra entidad con el objetivo de su participación en otras relaciones.

Jerarquías

Subtipos y Supertipos



- Exclusivas y compartidas
- De participación total y parcial

Construcción del MER

No existen reglas que indiquen cómo construir un modelo de datos, sólo principios generales a aplicar junto al criterio del diseñador experimentado

- Interpretar las frases expresadas en lenguaje natural (en el relevamiento), identificando cuáles son las *entidades* (datos) y cuales las *relaciones* (entre los datos) en la organización. Chen propuso las siguientes heurísticas:
 - En general un sustantivo es una entidad, aunque también podría ser un atributo (Ej: “los ALUMNOS cursan MATERIAS”)
 - Un verbo o frase verbal puede indicar una relación entre entidades (Ej: “los alumnos CURSAN materias”). Asociaciones entre los datos.
- Que información acerca de las entidades y relaciones deberían registrarse? → determinar los *atributos* (de entidades y relaciones)

Decisiones de diseño

- Cuándo un concepto debería modelarse como entidad? O como atributo?
- Cuándo un concepto debería modelarse como una entidad? O una relación?

DEPENDEN DE LOS REQUERIMIENTOS DEL CLIENTE !!

- Cómo definir correctamente el grado de una relación? Es binaria? Es ternaria o de orden >3 ? Son varias o es una?

ANÁLISIS DE TIEMPOS Y ACTORES

- Es equivalente una relación ternaria a dos o más binarias entre las entidades involucradas?
- Es equivalente una relación ternaria a una agregación?

NO SIEMPRE !!

→ EXPERIENCIA Y PRÁCTICA SON INDISPENSABLES PARA RESOLVER ESTOS INTERROGANTES ←

... continuará ...



Sumario: Diseño Conceptual

- *El diseño conceptual se guía por el análisis de los requisitos*
 - Permite obtener una descripción de alto nivel de los datos que finalmente serán almacenados
- MER es uno de los modelos más populares para el diseño conceptual.
 - Construcciones expresivas, cercanas a la forma en que los usuarios piensan acerca de sus aplicaciones.
 - Numerosas herramientas CASE (que generalmente siguen el modelo binario).
 - En actual evolución, mismo estilo que MER: UML
- Mecanismos de derivación casi automática desde el esquema conceptual al esquema lógico genérico.