



# Basi di Dati

Informatica ICA (LC)

Giacomo Boracchi

[giacomo.boracchi@polimi.it](mailto:giacomo.boracchi@polimi.it)

21 Novembre 2017

POLITECNICO DI MILANO



Le presenti slide sono tratte dalle slide del libro di testo  
*"Basi di dati", Atzeni, Ceri, Fraternali, Paraboschi, Torlone,*  
*McGraw-Hill 4a edizione, 2014*

**Credits Pierluigi Plebani**



Facciamo subito una distinzione:

- **informazione**: notizia, dato o elemento che consente di avere **conoscenza più o meno esatta di fatti, situazioni, modi di essere.**
- **dato**: ciò che è immediatamente **precedente alla conoscenza, prima di ogni elaborazione;** (in informatica) **elementi di informazione costituiti da simboli che debbono essere elaborati**

(definizioni dal Vocabolario della lingua italiana 1987)



## Sistema Informatico/Informativo

Nello svolgimento di ogni attività è necessario:

- disporre di informazioni
- gestire le informazioni

**Sistema Informativo:** è il sistema complessivo che organizza e gestisce le informazioni necessarie per lo svolgimento di un'attività.

**Sistema Informatico:** è la parte automatizzata del sistema informativo.

I sistemi informativi esistono quindi da prima dell'avvento dei calcolatori (banche/uffici anagrafe esistono da secoli). Al giorno d'oggi, molti sistemi informativi sono anche informatici



**Nelle attività umane**, le informazioni vengono gestite in forme diverse:

- idee informali
- linguaggio naturale (scritto o parlato, formale o colloquiale, in varie lingue)
- disegni, grafici, schemi
- numeri e codici

e su vari supporti

- mente umana, carta, dispositivi elettronici



Nelle attività standardizzate dei **sistemi informativi** complessi, sono state introdotte col tempo **forme di organizzazione e codifica delle informazioni**

Ad esempio, nei servizi anagrafici si è iniziato con registrazioni discorsive e poi si è passati a registrare:

- nome e cognome
- estremi anagrafici
- codice fiscale



I sistemi informatici si occupano della

- Raccolta, acquisizione
- Archiviazione, conservazione
- Elaborazione, trasformazione, produzione
- Distribuzione, comunicazione, scambio

dei dati e dell'informazione



I dati di per se non hanno alcun significato, ma una volta elaborati ed interpretati possono fornire informazioni, conoscenza.

Nei sistemi informatici,

- tipicamente i dati sono l'input, le informazioni l'output
- dati ed informazioni vengono immagazzinati allo stesso modo ma rappresentano comunque due entità distinte.



Un esempio:

Mario 2753

su un foglio di carta sono due **dati** e non significano molto

Se il foglio di carta viene fornito in risposta alla domanda “*A chi mi devo rivolgere per abilitare il mio badge aziendale; qual è il suo numero di telefono interno?*”, allora i dati possono essere interpretati per fornire **informazione** e arricchire la conoscenza



## Il valore dei dati

La rappresentazione precisa di forme più ricche di informazione e conoscenza è difficile

I dati tipicamente sono *più stabili* delle applicazioni

I dati costituiscono spesso una risorsa strategica, perché più stabili nel tempo di altre componenti (processi, tecnologie, ruoli umani):

- ad esempio, i dati delle banche o delle anagrafi



### Accezione generica, metodologica

- **Insieme organizzato di dati** di interesse per un'applicazione (o più in generale per il supporto allo svolgimento delle attività di un'organizzazione es. azienda, ufficio, persona)

### Accezione specifica, metodologica e tecnologica

- **insieme di dati gestito** da un Data Base Management System (DBMS)



# Data Base Management System (DBMS)

Sistema che gestisce collezioni di dati:

- **Grandi:**
  - dimensioni (molto) maggiori della memoria centrale dei sistemi di calcolo utilizzati
  - il limite deve essere solo quello fisico dei dispositivi
- **Condivise:**
  - Applicazioni ed utenti diversi devono poter accedere secondo opportune modalità a dati comuni
  - La condivisione riduce la ridondanza
  - Garantisce accesso simultaneo mediante il *controllo di concorrenza* per evitare inconsistenze
- **Persistenti:**
  - hanno un tempo di vita indipendente dalle singole esecuzioni dei programmi che le utilizzano



## Data Base Management System (DBMS)

Il DBMS inoltre garantisce:

- *Affidabilità*: il contenuto della base dati viene preservato anche in caso di malfunzionamenti
- *Privatezza*: ogni utente/applicazione è abilitato a svolgere solo specifiche operazioni (accesso regolato da autorizzazioni)
- *Efficienza*: svolgono operazioni utilizzando un numero limitato di risorse (tra cui spazio su disco e tempo di attesa)
- *Efficacia*: rendono produttive le attività degli utenti

**N.B** il DBMS a sua volta utilizza files per immagazzinare i dati (ma offre le caratteristiche sopra). I files del DBMS sono organizzati in maniera specifica.



Molti software DBMS disponibili sul mercato;

Esempi:

- Access
- DB2
- Oracle
- Informix
- Sybase
- SQLServer
- MySql



# Modello Concettuale di una Base Dati

Progettazione



Insieme di concetti utilizzati per **organizzare i dati** di interesse e **descrivere la loro struttura** in modo che sia fruibile da un elaboratore

### **Modello Concettuale:**

- Descrive concetti del mondo reale più che i dati
- Sono indipendenti da come i dati verranno rappresentati nel DBMS
- Si utilizzano per la progettazione della base dati

### **Modelli Logici:**

- Supportano una descrizione dei dati che può essere **elaborata dal sistema** (DBMS)
- Vengono **mappati facilmente** sulle strutture fisiche di memorizzazione



Insieme di concetti utilizzati per **organizzare i dati** di interesse e **descrivere la loro struttura** in modo che sia fruibile da un elaboratore

### Modello Concettuale:

oggi

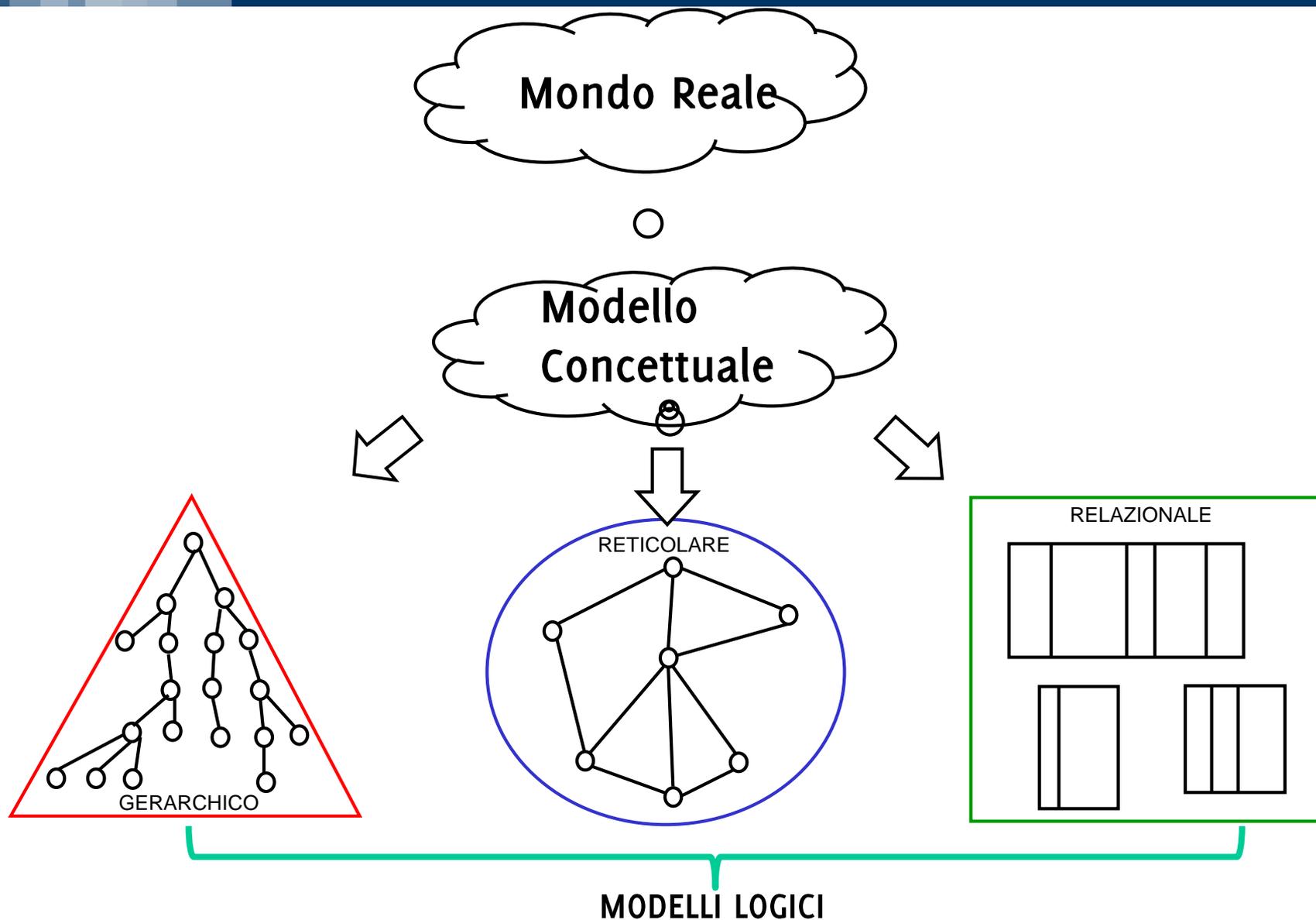
- Descrive concetti del mondo reale più che i dati
- Sono indipendenti da come i dati verranno rappresentati nel DBMS
- Si utilizzano per la progettazione della base dati

### Modelli Logici:

- Supportano una descrizione dei dati che può essere **elaborata dal sistema** (DBMS)
- Vengono **mappati facilmente** sulle strutture fisiche di memorizzazione



# Modelli dei dati





## Il progetto della base di dati

Si inserisce nel ciclo di vita del sistema informativo

Comprendente in generale le seguenti attività:

1. Raccolta ed analisi dei requisiti
2. **Progettazione di schemi del DB** e delle applicazioni
3. Implementazione
4. Validazione e collaudo
5. Funzionamento

Le basi di dati sono solo una componente del sistema informativo

Ci concentriamo sulla parte specifica di questo corso: **la progettazione degli schemi per la base di dati**



## 1. Prerequisiti della progettazione

L'**analisi dei requisiti** è condotta dalla una specifica figura professionale (analista) tramite interviste dell'utente (che deve essere per quanto possibile parte attiva della definizione dei requisiti)

Produce una descrizione che esprimiamo tramite un **testo riassuntivo** (tipicamente ambiguo).

Nella realtà si aggiungono anche:

- Descrizioni terminologiche, glossari, raccolte informali di definizioni
- Descrizioni astratte dei programmi (dataflow diagrams) e del loro uso da parte degli utenti (use case diagrams).



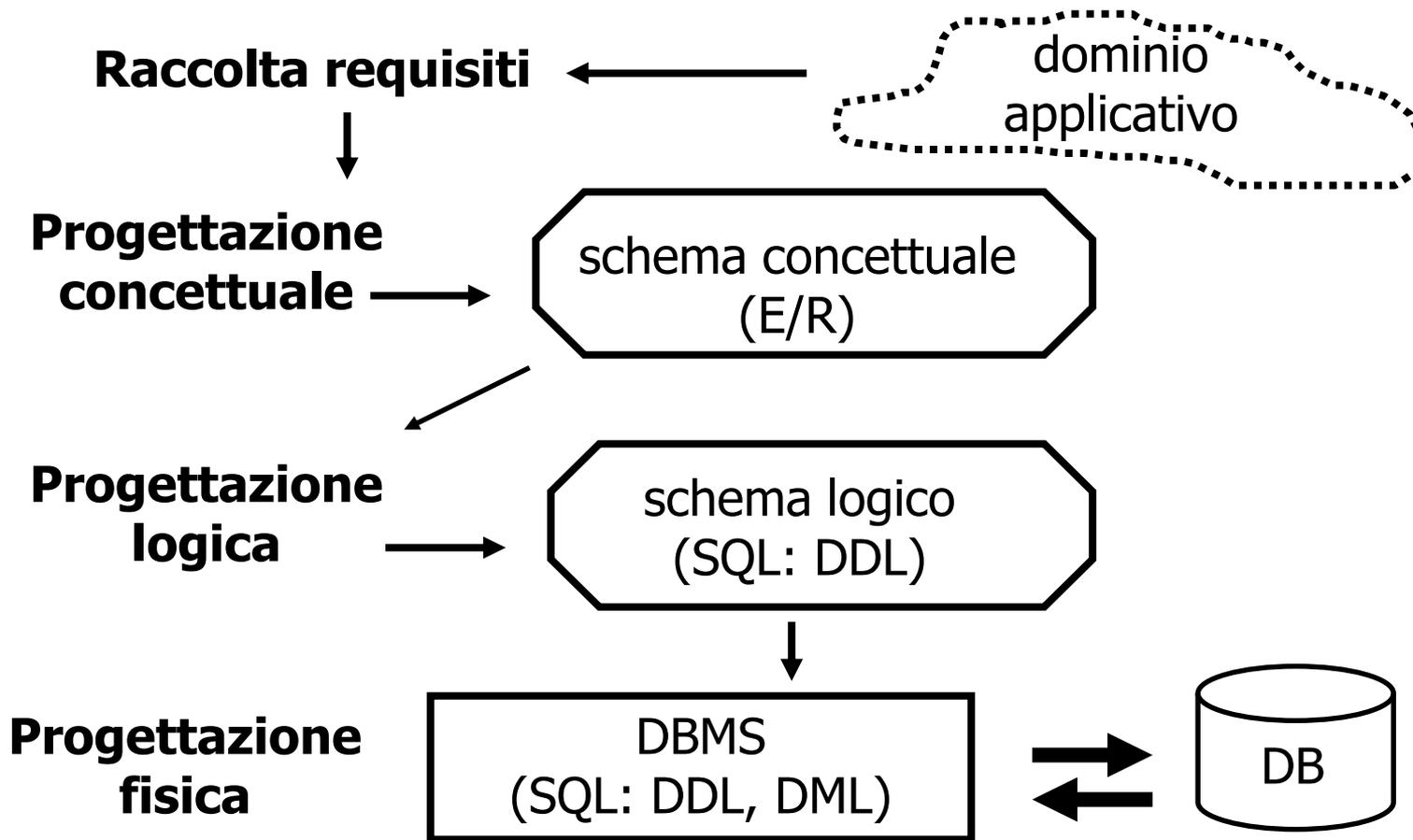
## 2. Fasi della progettazione

Si distinguono le seguenti fasi:

- la progettazione concettuale
- la progettazione logica
- la progettazione fisica

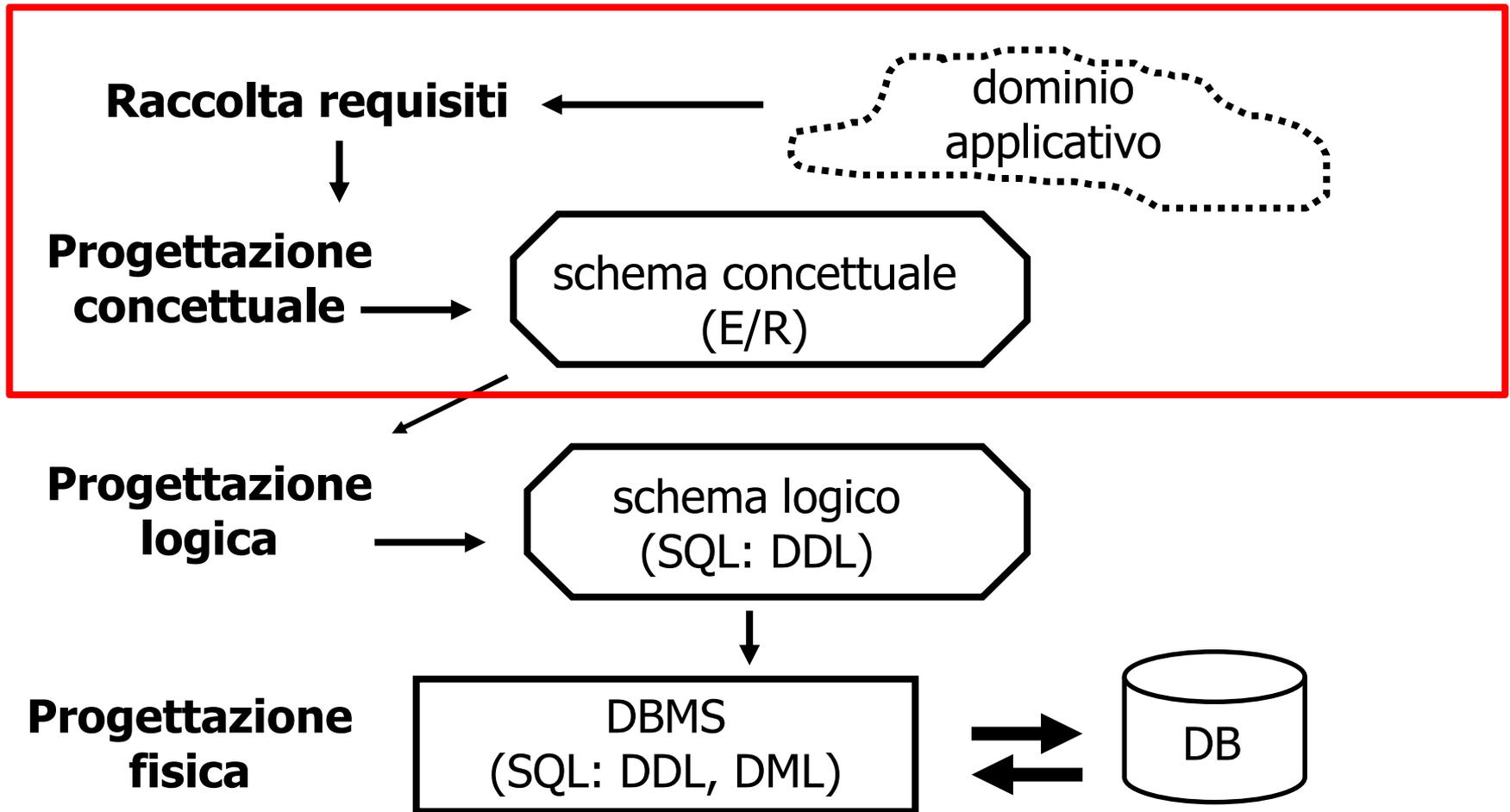


# Processo di design





# Processo di design





## La Progettazione Concettuale

Ha per scopo tradurre il risultato dell'analisi dei requisiti in una **descrizione formale** che dovrà essere indipendente dal DBMS

La descrizione formale è espressa tramite uno **SCHEMA CONCETTUALE**, costruito utilizzando un **MODELLO CONCETTUALE DEI DATI**

Il progettista nel produrre lo schema concettuale deve rappresentare il **contenuto informativo** della base dati

**Si guarda il cosa rappresentare, non il come**



## La Progettazione Logica

Ha per scopo tradurre lo SCHEMA CONCETTUALE in uno SCHEMA LOGICO, scelto all'interno dei modelli logici dei dati:

- Gerarchico, Reticolare, Relazionale, Orientato ad oggetti, XML, NoSQL

Lo schema logico è dipendente dal DBMS ma non dallo specifico prodotto

I dati vengono rappresentati indipendentemente dai dettagli fisici ma in maniera concreta, vicina alla macchina

**Si guarda il come rappresentare i dati, il cosa è già stato definito**



## La Progettazione Fisica

Ha per scopo produrre un PROGETTO FISICO della base dei dati, cioè un progetto che ottenga prestazioni ottimali tramite scelta e dimensionamento di strutture fisiche di accesso.

Il progetto fisico viene eseguito in modo differente su ciascun prodotto.



## Dipendenze da MODELLO e DBMS

	<b>Dipende dal MODELLO</b>	<b>Dipende dal DBMS</b>
<b>Progetto concettuale</b>	NO	NO
<b>Progetto logico</b>	SI	NO
<b>Progetto fisico</b>	SI	SI



Il modello **Entity-Relationship** (E-R), (P.P.Chen 1976) si è affermato come standard industriale di buona parte delle metodologie e degli strumenti per il progetto concettuale di basi di dati

Attenzione: Relationship = Associazione (*anche se poi si dice informalmente “relazione”*)



## Elementi fondamentali

**Entità:** *classe di oggetti* (e.g., fatti, cose, persone) che *esiste* di per sé nel dominio applicativo, della quale si vogliono registrare fatti specifici e che può essere chiaramente identificata in modo da poterla distinguere dalle altre

**Relazione:** fatto che descrive un'azione o una situazione e che stabilisce legami tra entità (associa, mette in relazione)

**Proprietà** (dette anche attributi): sono fatti che descrivono le caratteristiche delle entità e delle relazioni. Le proprietà assumono valori.



Rappresenta una **classe di oggetti o fatti** (es., automobili, impiegati, studenti, città, conti correnti, corsi universitari, corse degli autobus )

Rappresenta **oggetti rilevanti** per l'applicazione che **esistono autonomamente**

Ogni entità è caratterizzato da un nome

Un oggetto della classe rappresentata dall'entità è **un'occorrenza (o istanza) dell'entità** (es., Milano è un'istanza di CITTA', il mio conto in banca è un'istanza di CONTO, il corso di Informatica è un'istanza di CORSO)



# Simbolo grafico per rappresentare entità

**nome  
dell' entità**

esempio:

**studente**



## Relazione (o Associazione)

Rappresenta un **legame logico, di interesse per l'applicazione, tra entità** (es, l'associazione tra l'entità Auto e l'entità Persone).

Ogni istanza di una associazione è **una ennupla tra istanze di entità differenti** (es., legame tra un automobile targata 313 e Paperino, il suo proprietario)

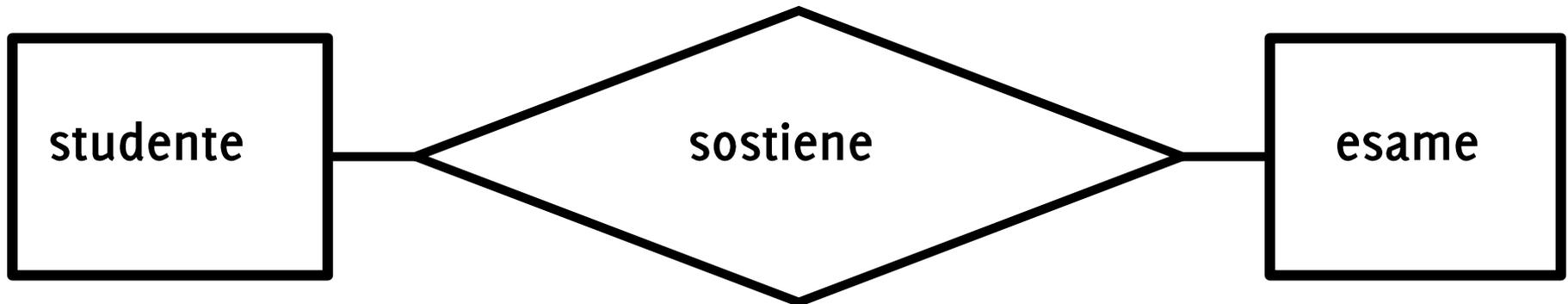
Ogni associazione è caratterizzata da un nome (es., «possiede»)



## Simbolo grafico per rappresentare associazioni

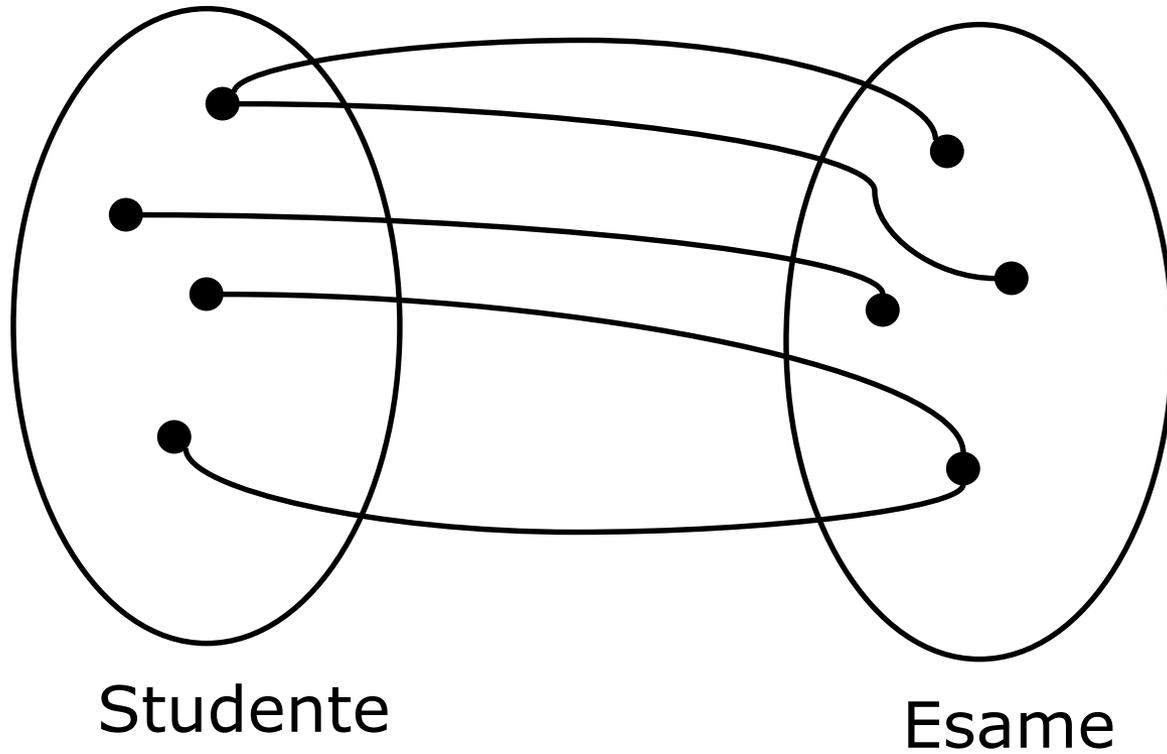


esempio:





# L'associazione tra occorrenze delle entità



Studente

Esame



Rappresentano **caratteristiche delle entità e delle associazioni di interesse per l'applicazione**

Ogni attributo è **caratterizzato da un nome e da un dominio di valori ammissibili**

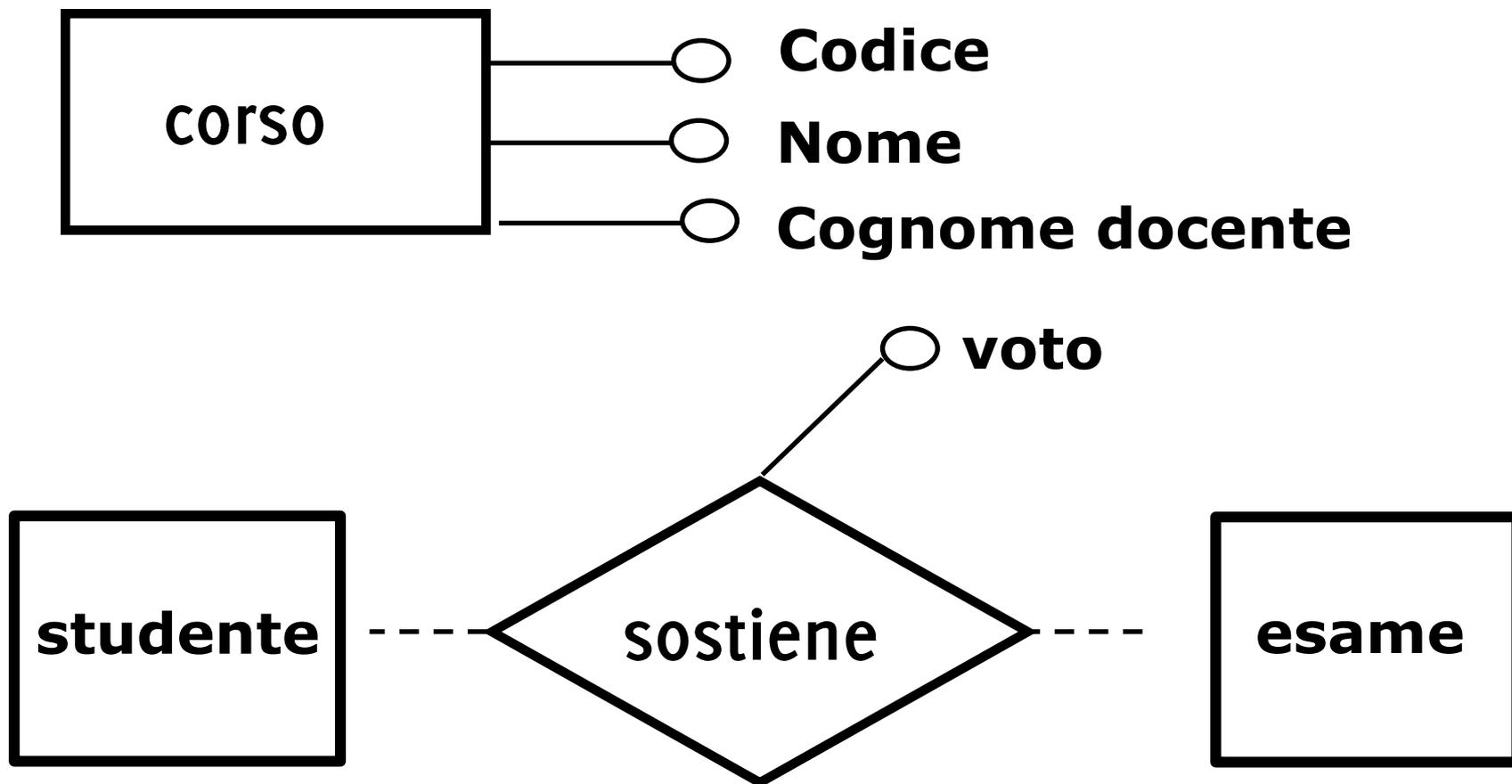
Esempio:

- Entità: persona
- Attributi: nome, cognome, via, civico, città, cap

E' possibile raggruppare attributi della stessa entità (o associazione) che sono affini in un unico attributo composto (es, residenza che raggruppa via, civico, città, cap)



## Simbolo grafico per rappresentare attributi





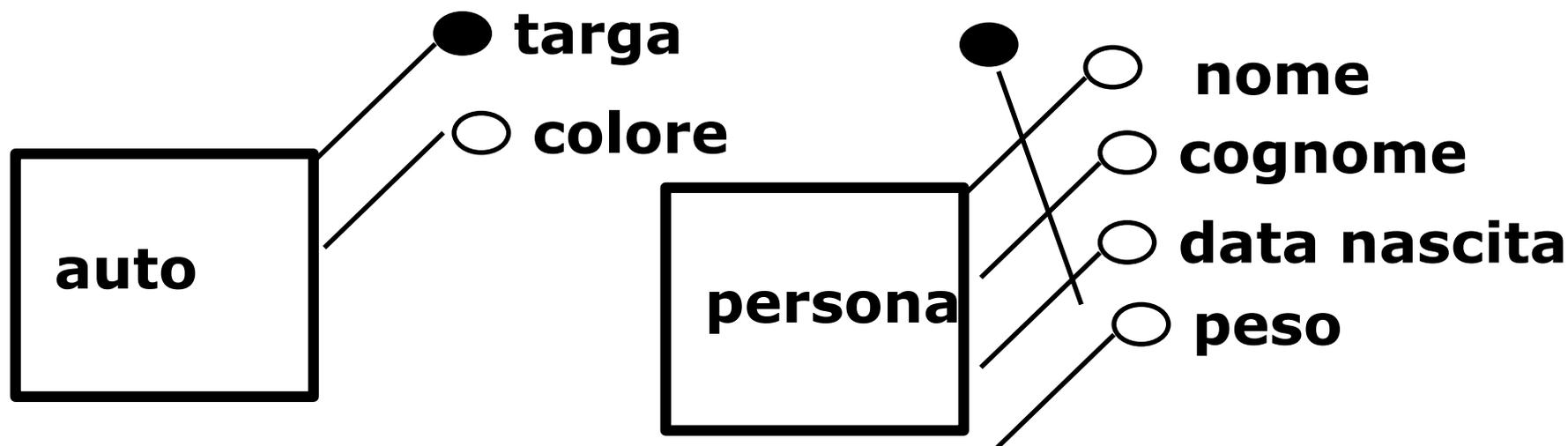
## Identificatori di Entità

Per ciascuna entità è possibile specificare quali **attributi identificano in maniera univoca le istanze delle entità**

Tali attributi sono detti *chiavi*

Non possono quindi esistere due istanze di questa entità aventi lo stesso valore dell'attributo chiave

Una chiave può anche essere una combinazione di attributi





## Esempio di testo analisi requisiti

Sviluppare il modello concettuale a partire dal seguente testo fornito da un'analisi dei requisiti

### Domanda 5

Si vuole progettare un sistema informativo per una società di distribuzione di farmaci per le farmacie comunali della città di Lecco. Il trasporto dei farmaci alle farmacie avviene tramite furgoni. Ogni furgone può servire una o più farmacie. La distribuzione dei farmaci può avvenire a diverse ore del giorno (anche alla stessa farmacia).

Ogni farmacia è caratterizzata da un nome, da un codice, da un indirizzo, da un direttore e dal numero di farmacisti occupati.

Ogni furgone è guidato da un conducente ed è caratterizzato da una cilindrata, una marca e dalla targa.

Progettare la base dati a livello concettuale e logico evidenziando le chiavi primarie.



## Linee guida per il progetto

Se il concetto è significativo per il contesto applicativo: entità

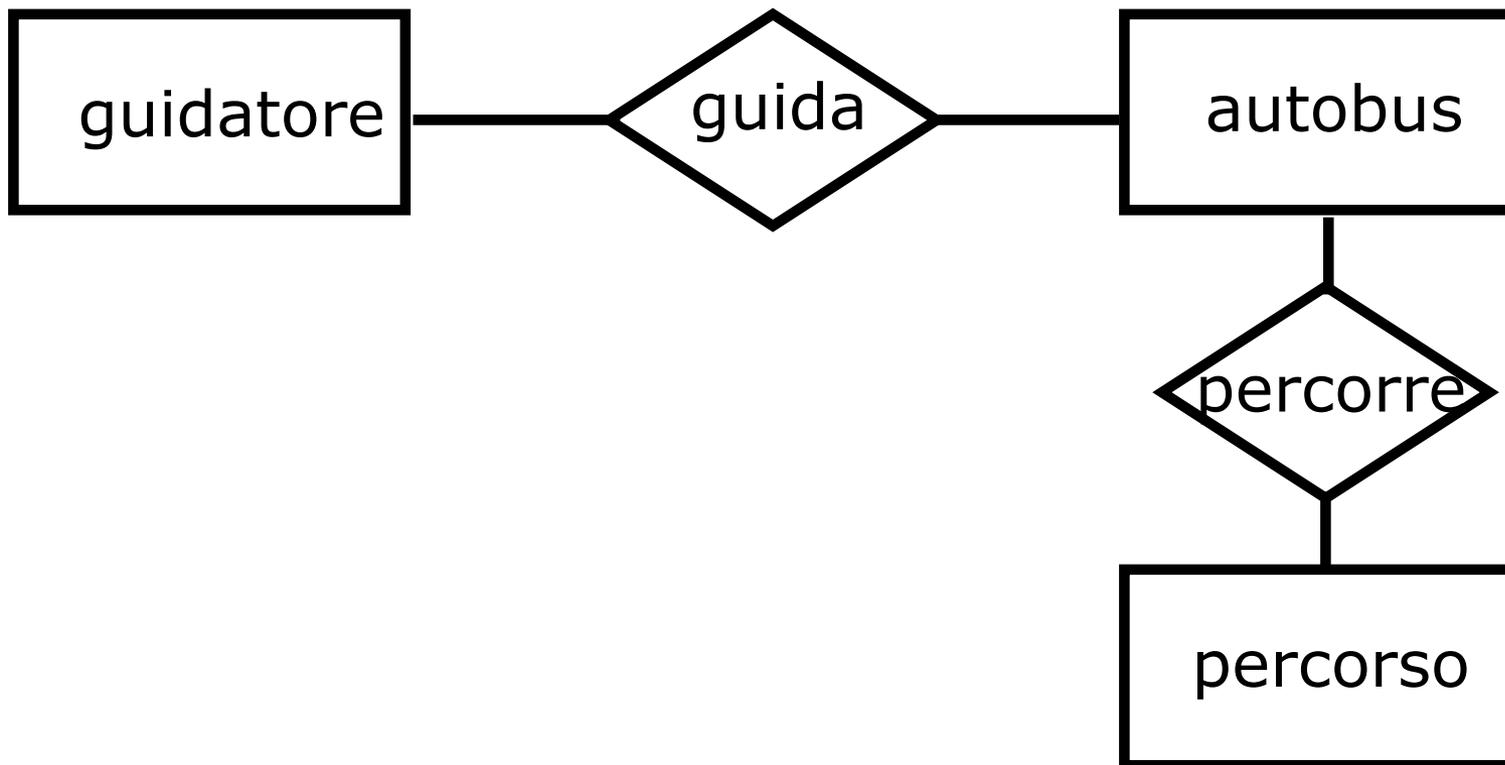
Se il concetto è descrivibile tramite un dato elementare: attributo

Se il concetto definisce un legame tra entità: associazione

La corrispondenza tra oggetti e fatti del mondo reale ed entità, associazioni e attributi non è assoluta ma dipende dal contesto

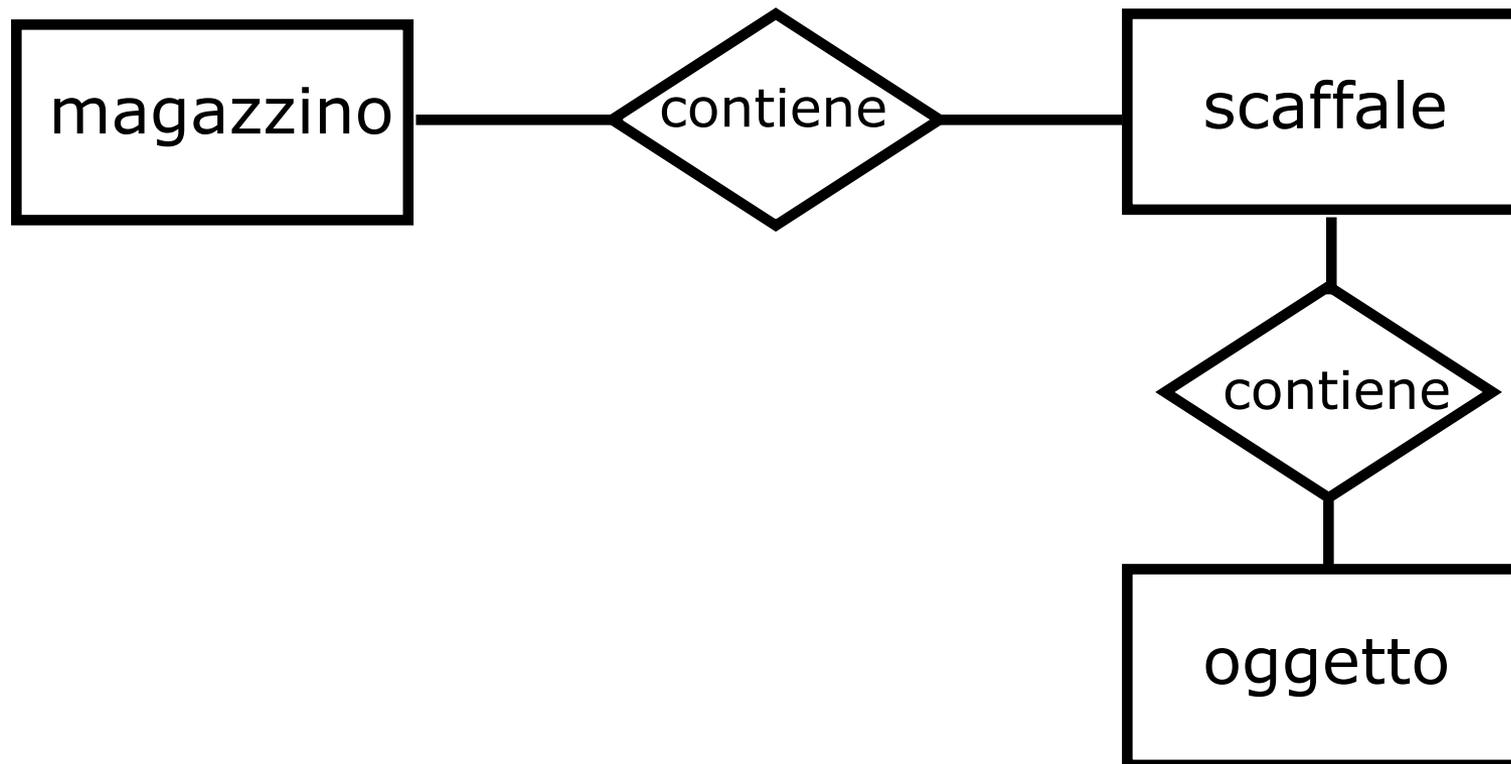


## Esempio: gestione viaggi



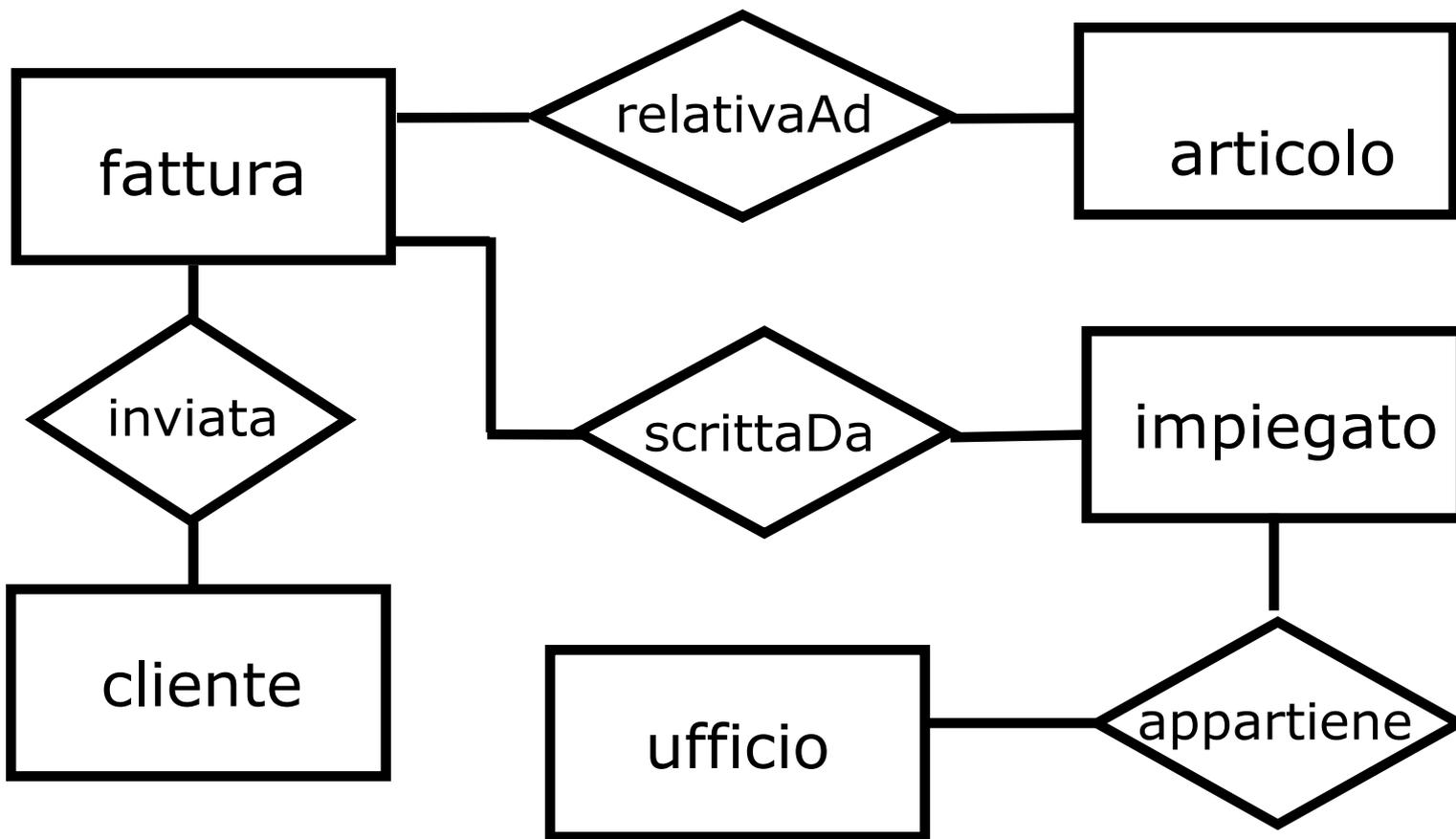


## Esempio: gestione magazzino



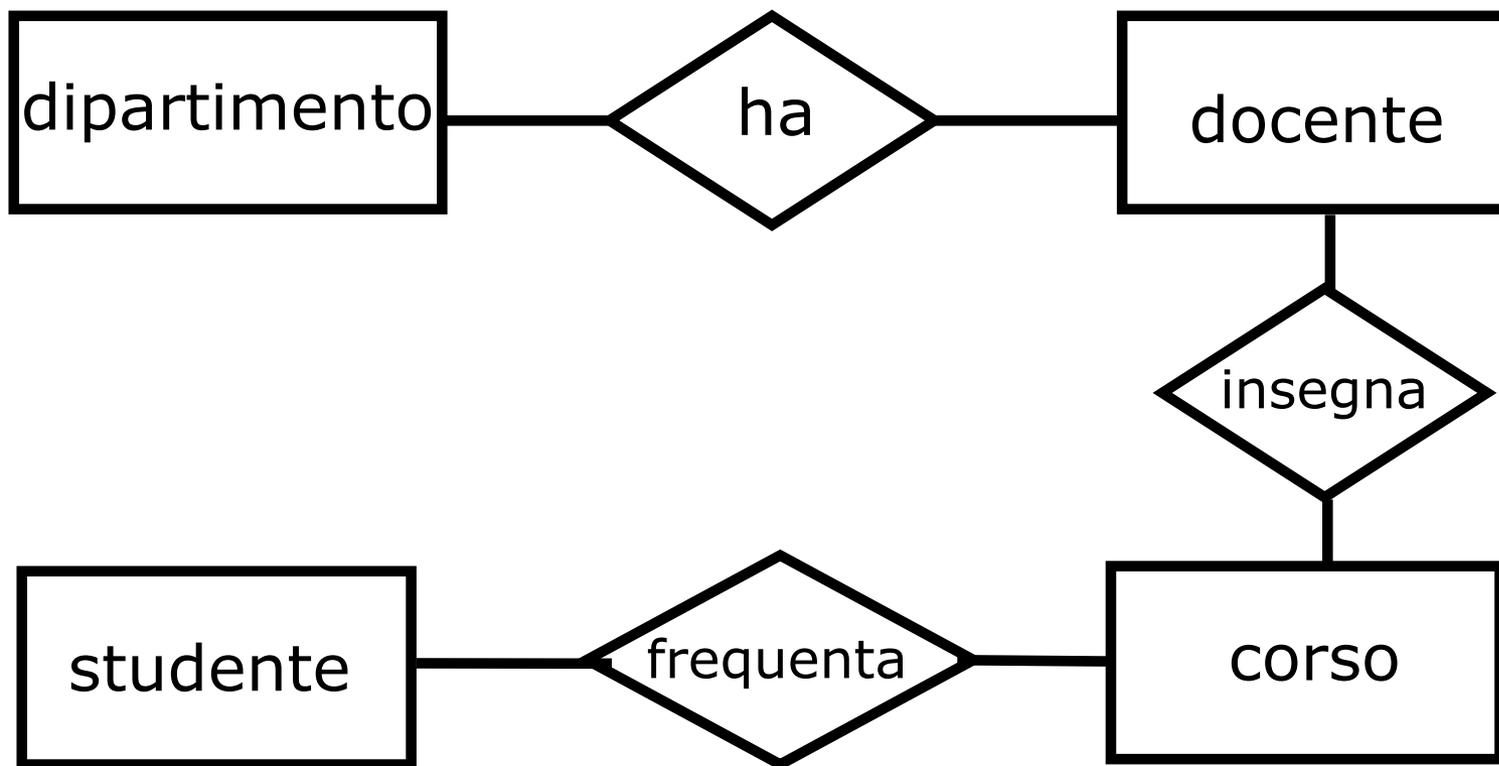


## Esempio: gestione fatture





## Esempio: università





## Cardinalità delle associazioni

Per cardinalità si intende **un vincolo sul numero di istanze di associazione** cui ciascuna istanza di entità deve partecipare.

È una coppia (MIN-CARD, MAX-CARD)

MIN-CARD = 0 (opzionale)

= 1 (obbligatoria)

MAX-CARD = 1 (uno)

= N (molti)

In base alla sola cardinalità massima si hanno associazioni uno-uno, uno-molti, molti-molti

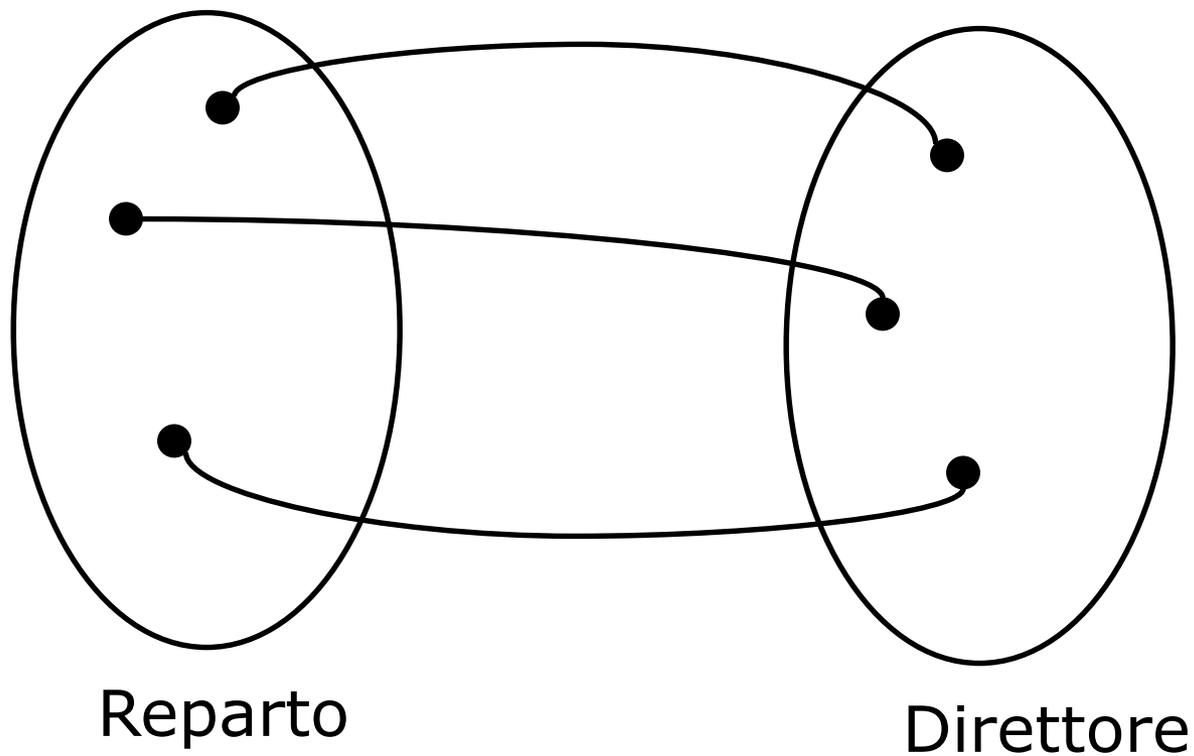
E' possibile assegnare cardinalità (minima-massima) anche agli attributi



- un reparto deve essere diretto da uno e un solo direttore (1,1)
- un direttore deve dirigere uno ed un solo reparto (1,1)



## Relazione tra istanze (ass 1:1)



Reparto

Direttore



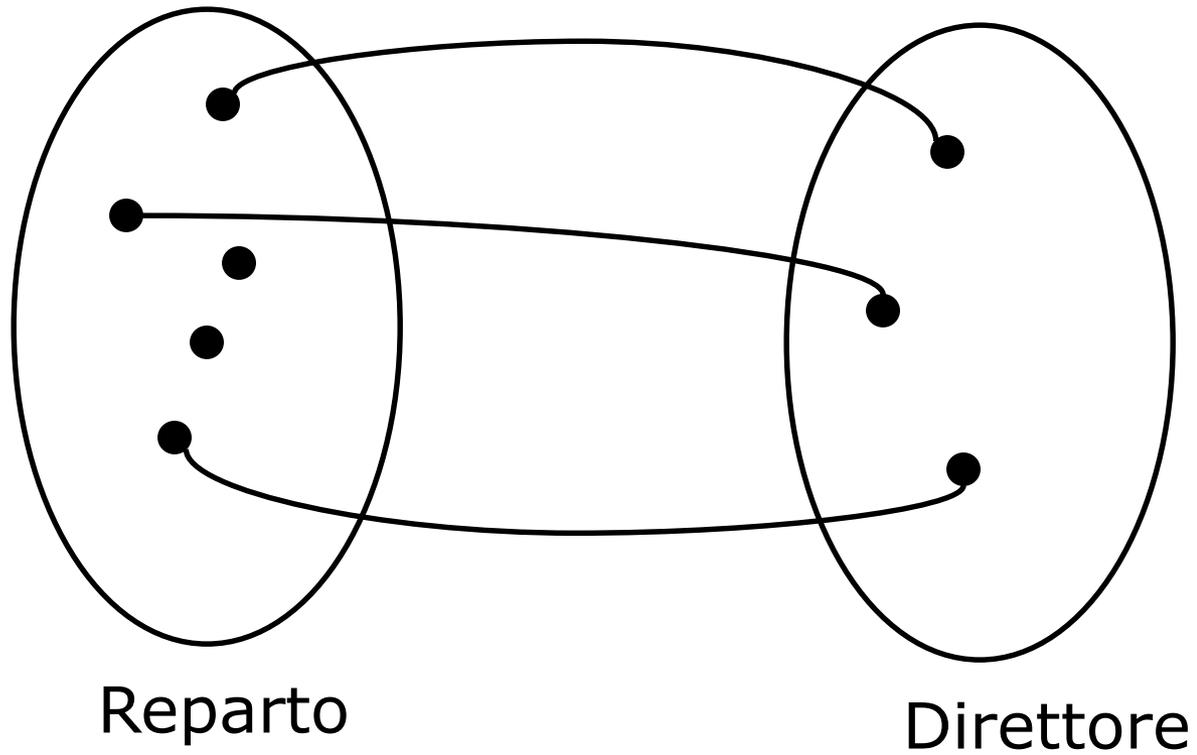
## Associazione 1:1 con opzionalità



- un reparto può essere diretto da un solo direttore (0,1)
- un direttore deve dirigere uno ed un solo reparto (1,1)



## Relazione tra istanze (ass 1:1 opz)



Reparto

Direttore



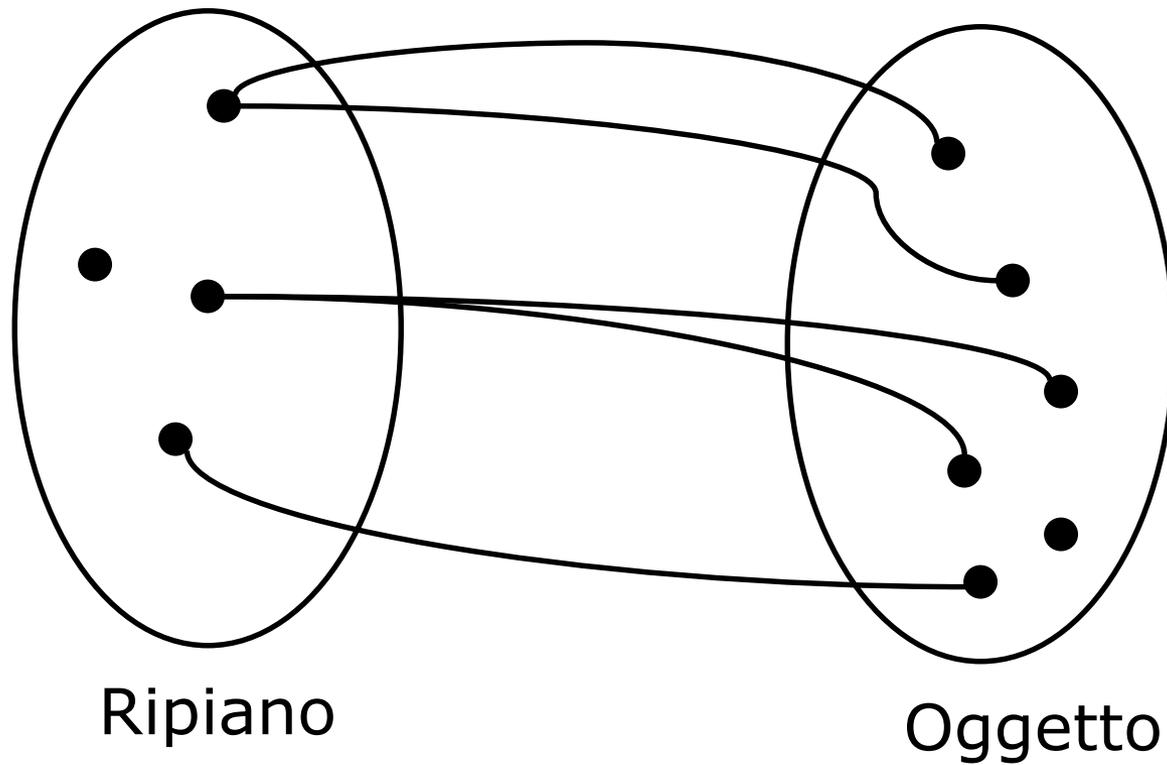
## Associazione 1:N



- un ripiano può contenere molti oggetti (0,n)
- un oggetto può essere contenuto al più su un ripiano (0,1)



## Relazione tra istanze (ass 1:N)





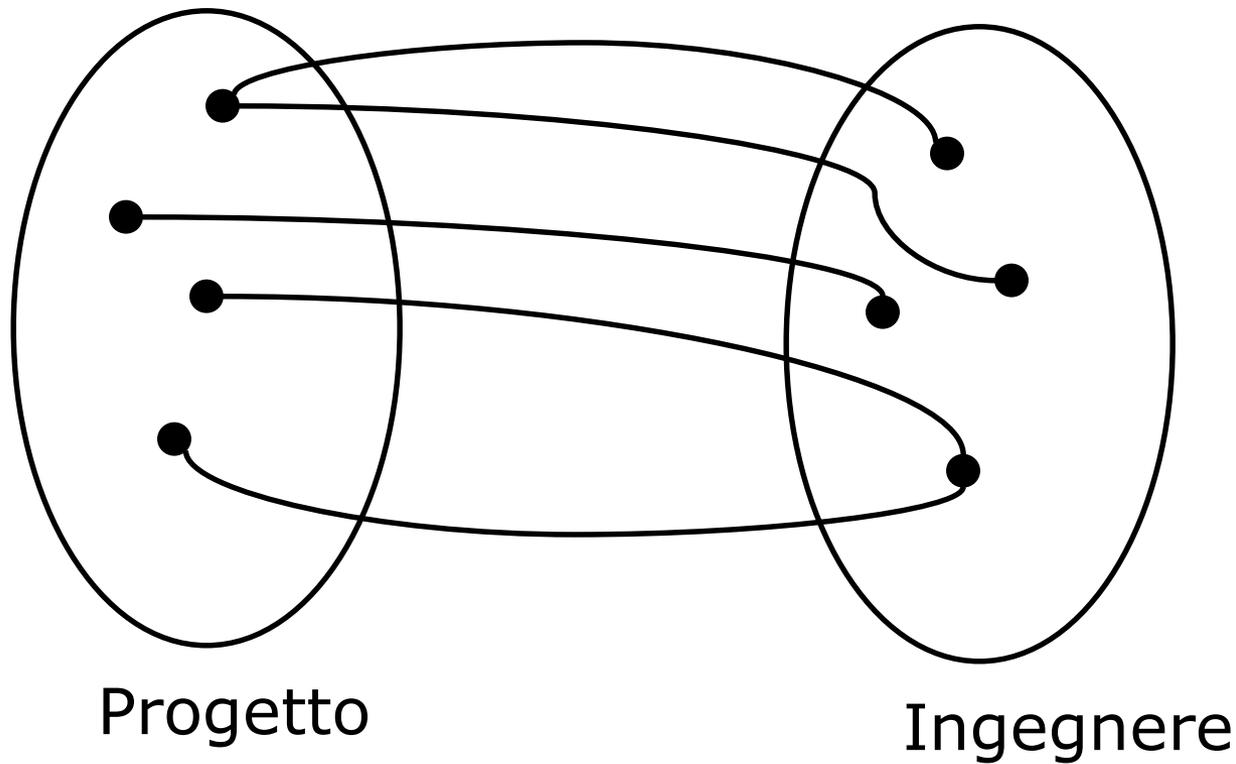
## Associazione N:M



- un progetto può essere fatto da molti ingegneri (0,n),
- un ingegnere deve partecipare ad uno o più progetti (1,m)



## Relazione tra istanze (ass N:M)

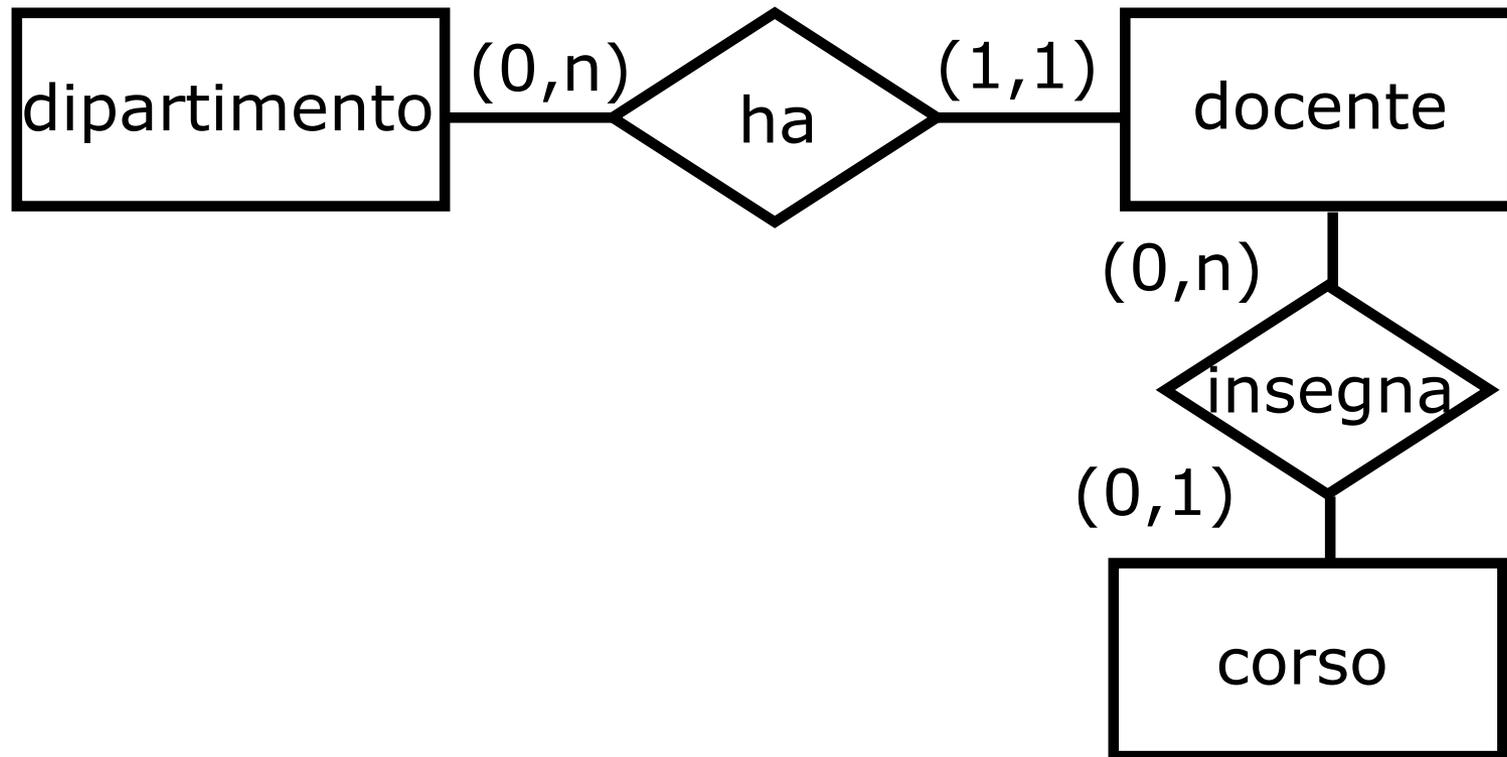


Progetto

Ingegnere

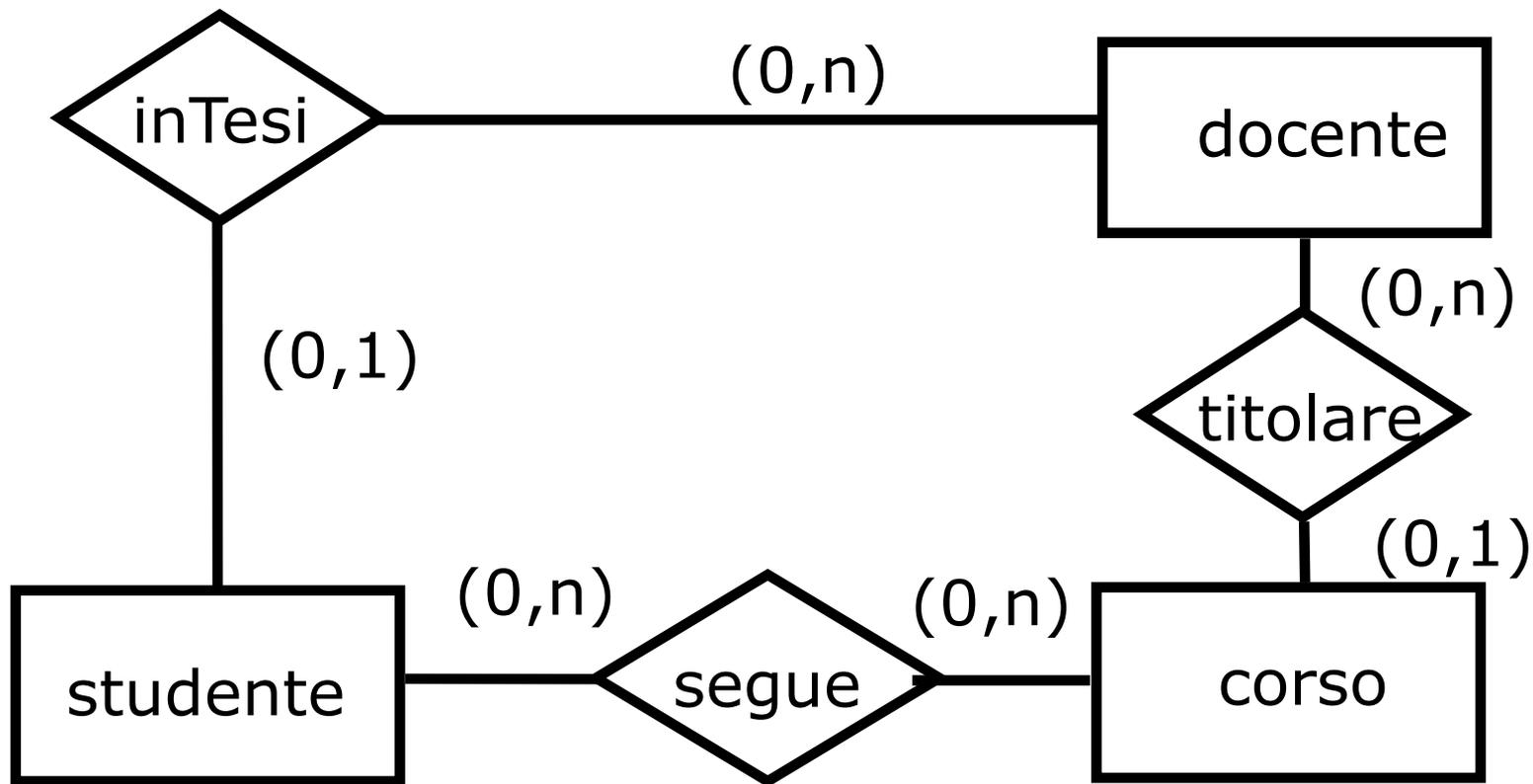


## Esempio: università





## Esempio: università





#### **Domanda 4**

Un entomologo vuole realizzare una base dati per gestire gli insetti autoctoni della provincia di Lecco. Ogni insetto è identificato dall'ordine, dalla specie, dal peso medio e dalla lunghezza media dell'individuo adulto. L'entomologo, nelle sue campagne di ricerca svolte in certe date programmate, trova, in una certa località (nome località, quota, punto GPS), un nuovo insetto e aggiorna la sua base di dati. Ogni individuo rinvenuto viene anche pesato e misurato (peso, altezza) e valutato (giovane/adulto, sano/nonsano). Si inseriscano tutti quei campi che consentono una gestione ottimizzata della base dati.

Progettare il modello E-R della base dati.

valute  
(70)



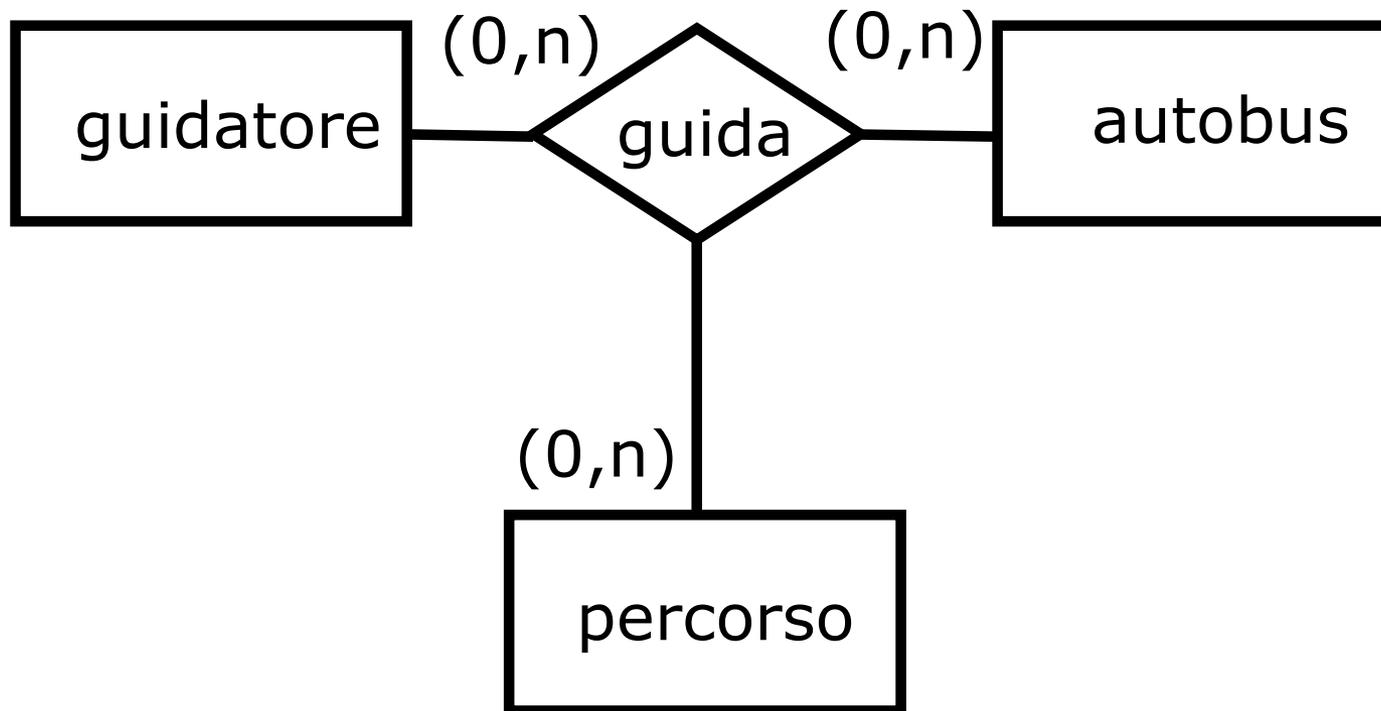
### **Domanda 5**

Si vuole progettare un sistema informativo per le iscrizioni ad un meeting di atletica leggera. Il programma del meeting prevede un certo numero di gare. Per ogni gara il sistema deve memorizzare un codice identificativo, un giudice di gara, un nome (ad es. "100 metri", "200 metri", "salto in alto", "salto in lungo"), il record europeo ed il record mondiale. Gli atleti (nome, cognome, data di nascita) possono iscriversi ad una o più gare del meeting. Ogni atleta che si iscrive ad una gara deve specificare se vuole partecipare "in classifica" o "fuori classifica". Progettare la base dati a livello concettuale e logico evidenziando le chiavi primarie.



## Esempio: gestione autobus

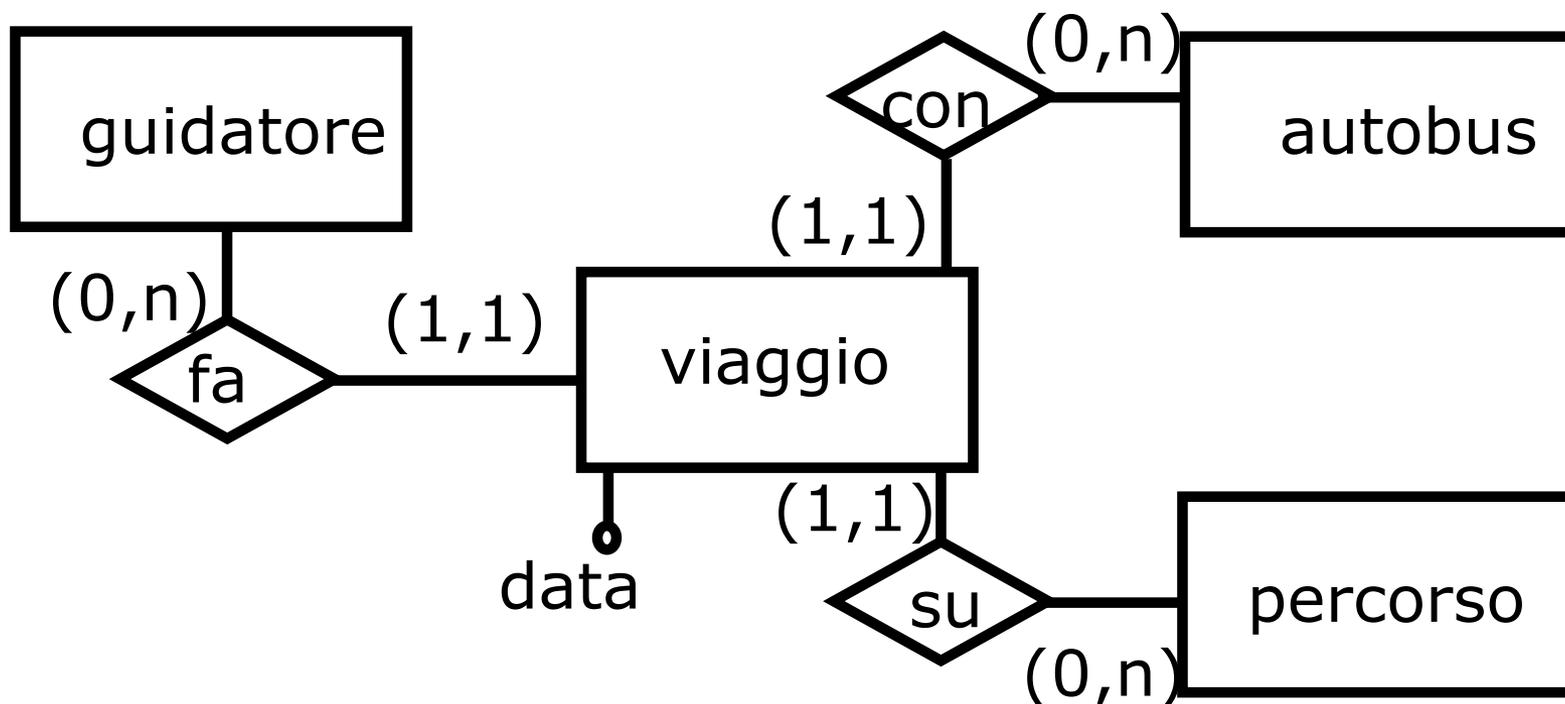
Come EVITARE l'utilizzo di una relazione ternaria





## Esempio: gestione autobus

Soluzione senza relazione ternaria





## Un esempio di progetto concettuale

“Si vuole realizzare un archivio per la gestione di una libreria. I principali requisiti espressi dall’utente sono la possibilità di memorizzare i dati relativi ai libri (titolo, autore, casa editrice, anno, argomento, copie disponibili). Si devono poter cercare i libri a partire dal titolo, dall’autore, dall’argomento e dall’editore. [...]”



## Individuazione entità in gioco

### Libro

- Titolo
- Anno
- Copie disponibili
- Autore
- Casa editrice
- Argomento

### Libro

- Titolo
- Anno
- Copie disponibili

### Autore

### Casa editrice

### Argomento



## Perché tutte queste entità?

### Ma non ci conviene fare un'unica tabella?

No, per i seguenti motivi:

#### Integrità dei dati

- un autore dovrebbe avere lo stesso nome per ogni suo libro; eventuali modifiche non dovrebbero essere apportate manualmente ad ogni libro

#### Occupazione di memoria

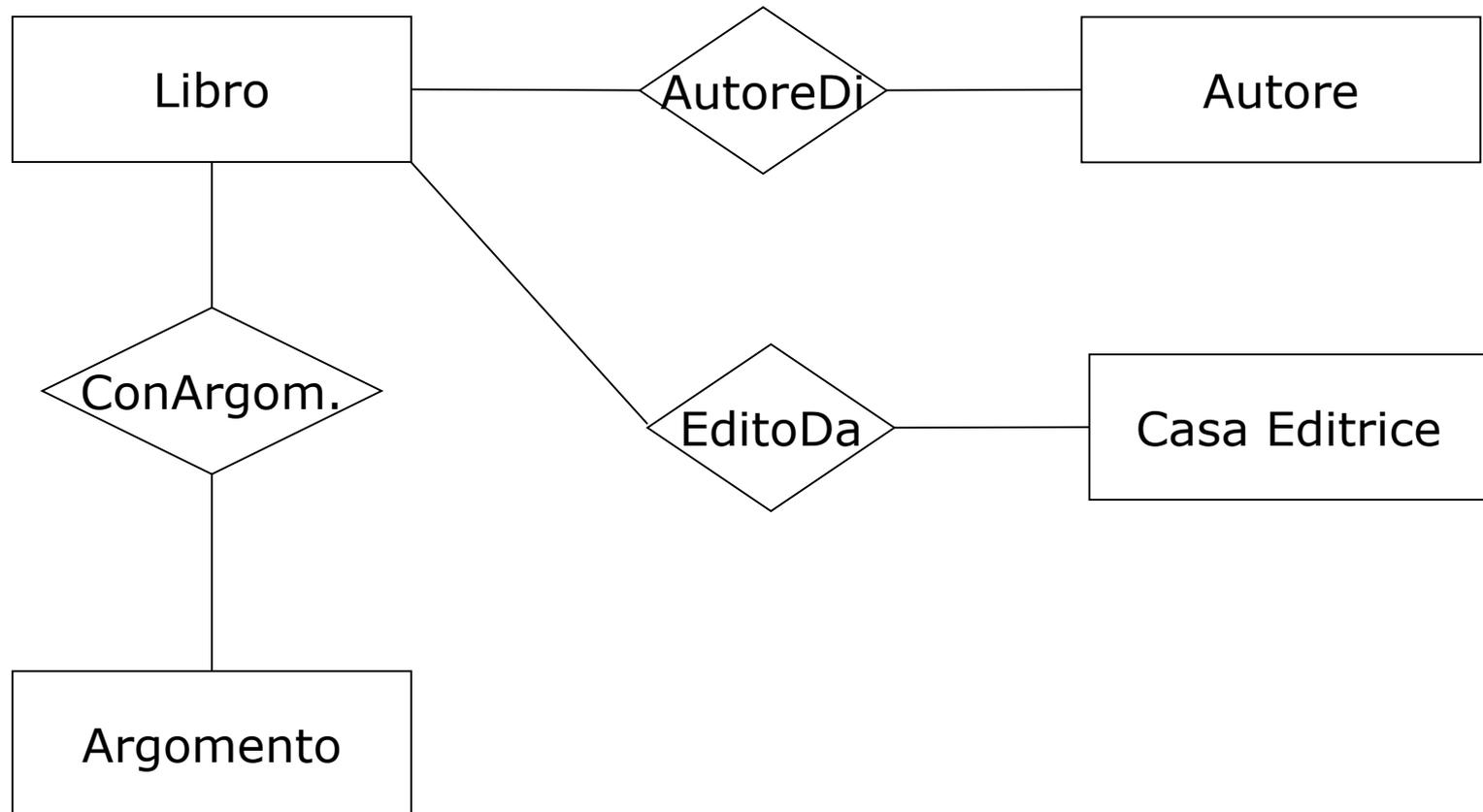
- ripetere nome, cognome ed eventualmente altri dati per ogni libro scritto da un autore è sconveniente

#### Problemi in inserimento

- Inserire ogni volta i dati completi di un autore rende molto probabile un errore

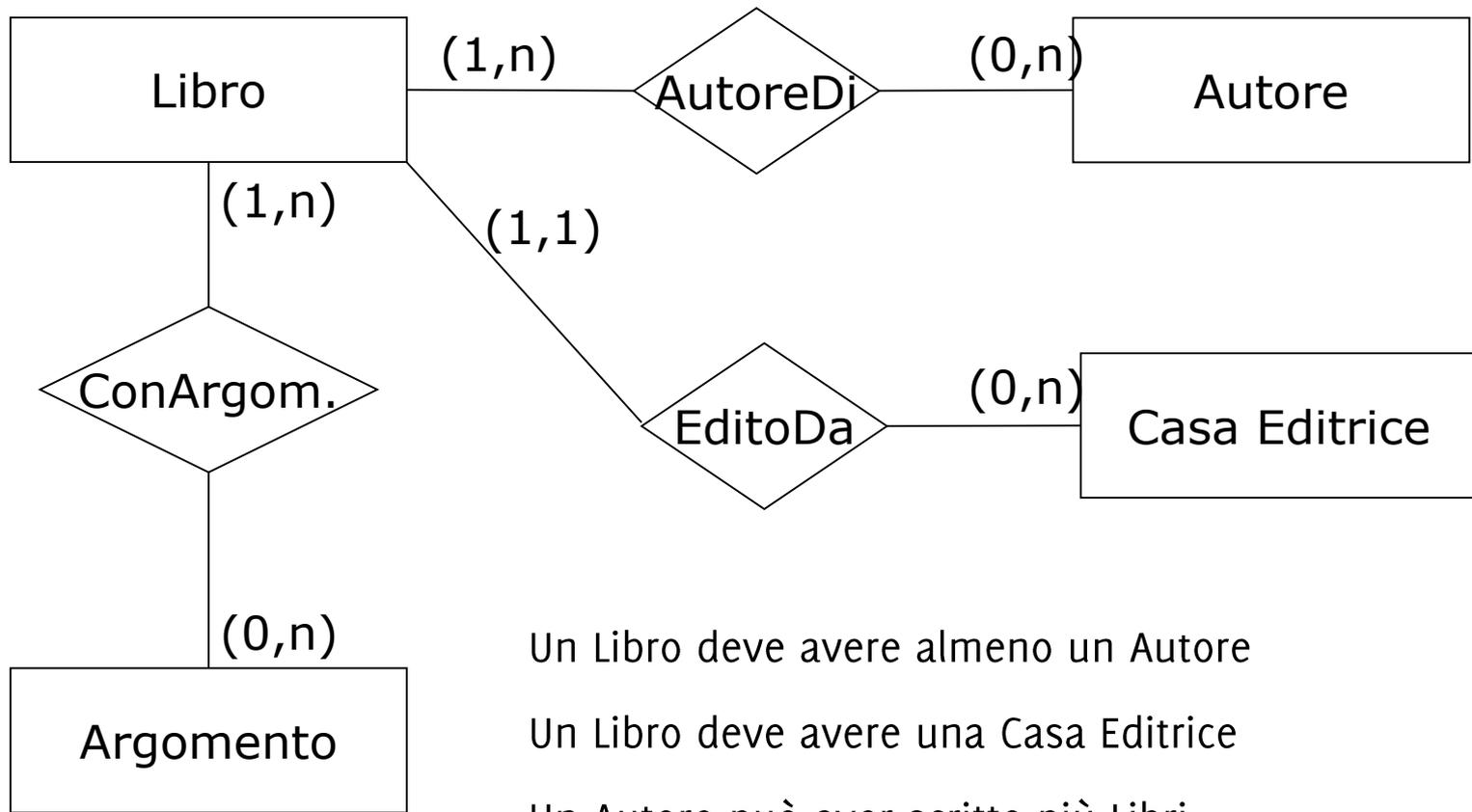


## Individuazione relazioni fra entità





## Cardinalità delle relazioni



Un Libro deve avere almeno un Autore

Un Libro deve avere una Casa Editrice

Un Autore può aver scritto più Libri

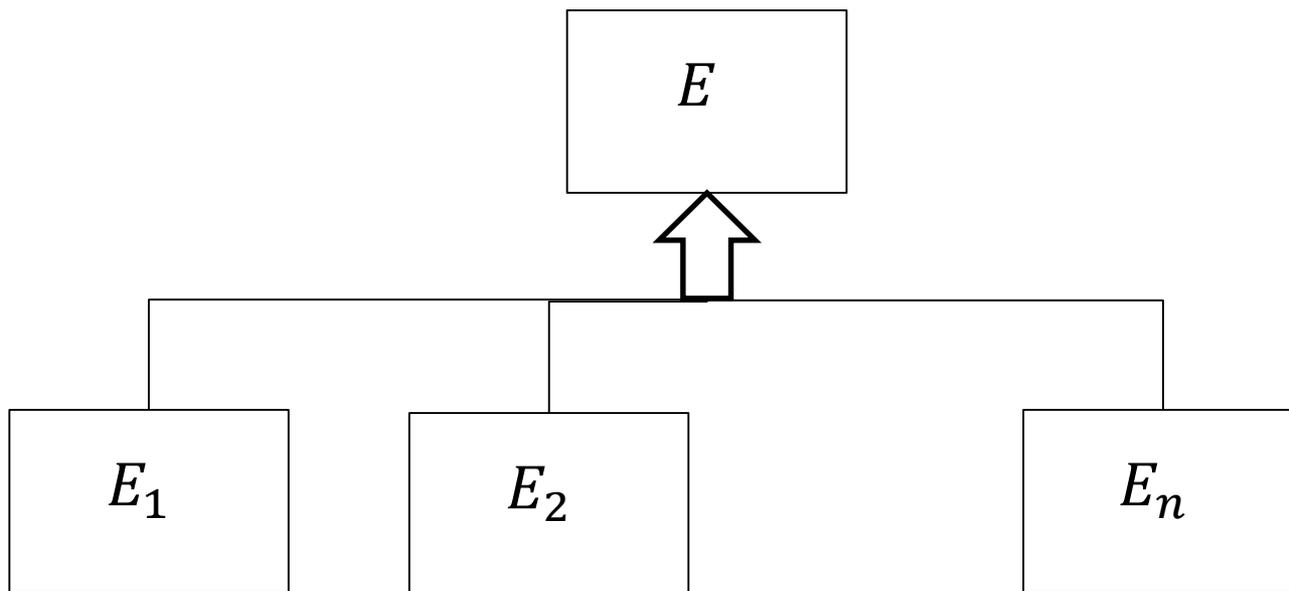


## Generalizzazioni

Le generalizzazioni sono legami logici tra:

- un'entità  $E$  (detta genitore)
- una o più entità  $E_1, E_2, \dots, E_n$  (dette entità figlie)

e di cui  $E$  è **più generale** nel senso che **comprende**  $E_1, E_2, \dots, E_n$  come casi particolari

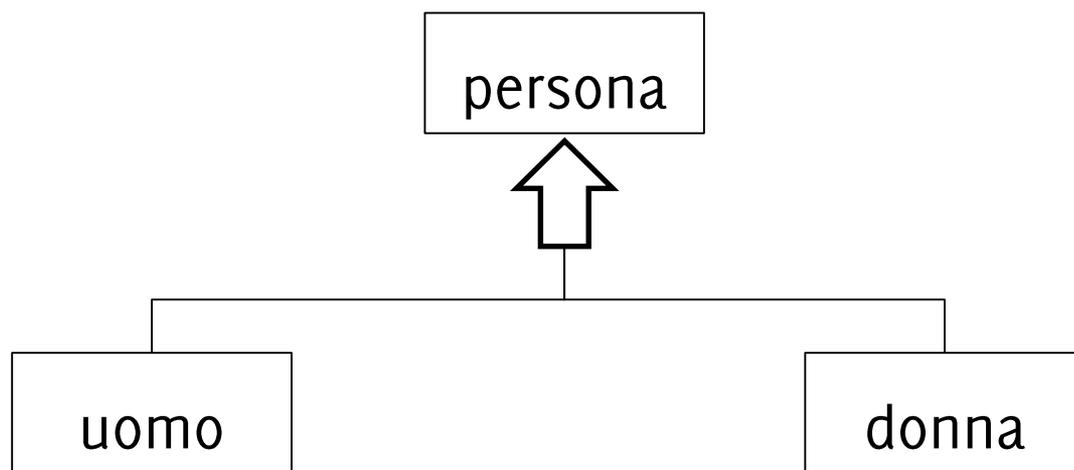




## Generalizzazioni, proprietà

Le generalizzazioni hanno le seguenti proprietà:

- Ogni entità figlia è un'occorrenza dell'entità genitore

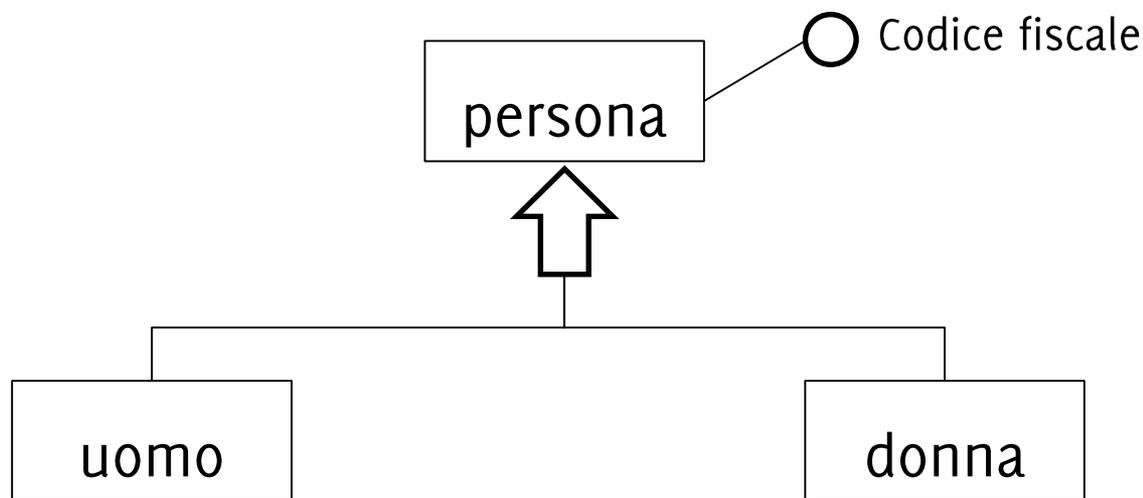




## Generalizzazioni, proprietà

Le generalizzazioni hanno le seguenti proprietà:

- Ogni entità figlia è un'occorrenza dell'entità genitore
- Ogni attributo del padre è attributo delle sue entità figlie

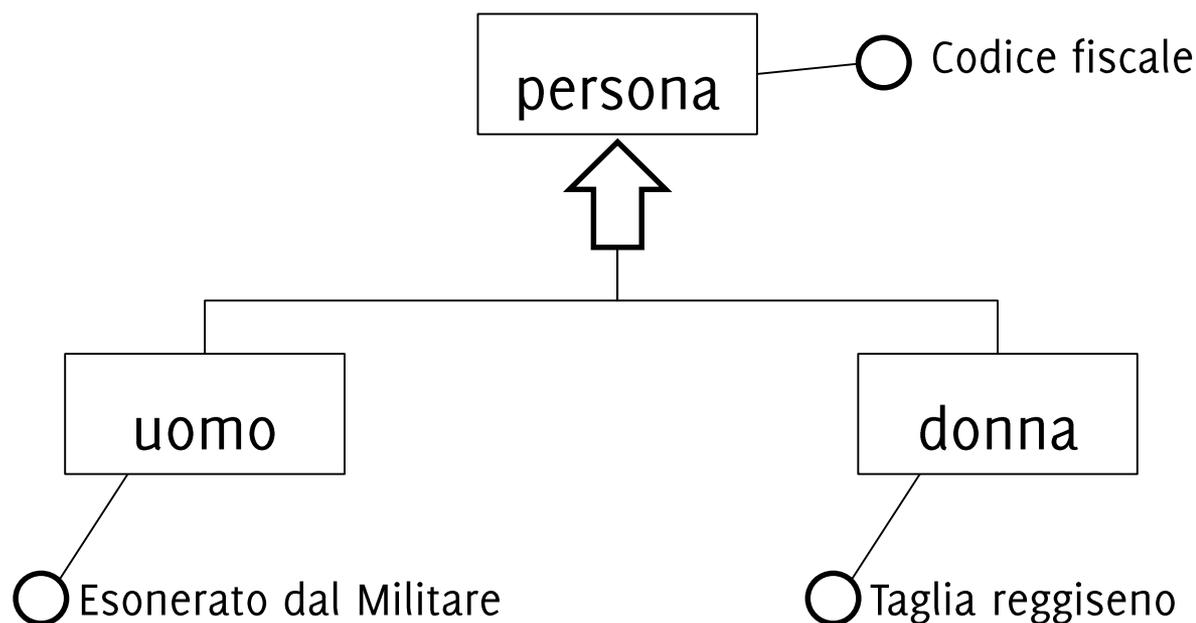




## Generalizzazioni, proprietà

Le generalizzazioni hanno le seguenti proprietà:

- Ogni entità figlia è un'occorrenza dell'entità genitore
- Ogni attributo del padre è attributo delle sue entità figlie
- Non vale il viceversa, attributi dei figli possono non essere del padre

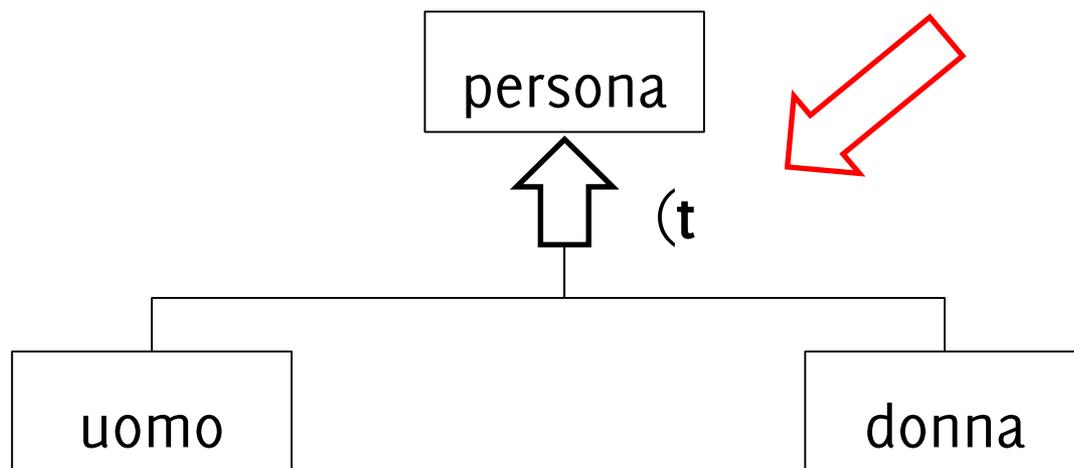




## Generalizzazioni: totali vs parziali

### Generalizzazioni: totali vs parziali

- **Totale:** ogni istanza dell'entità genitore è occorrenza di una o più delle sue entità figlie

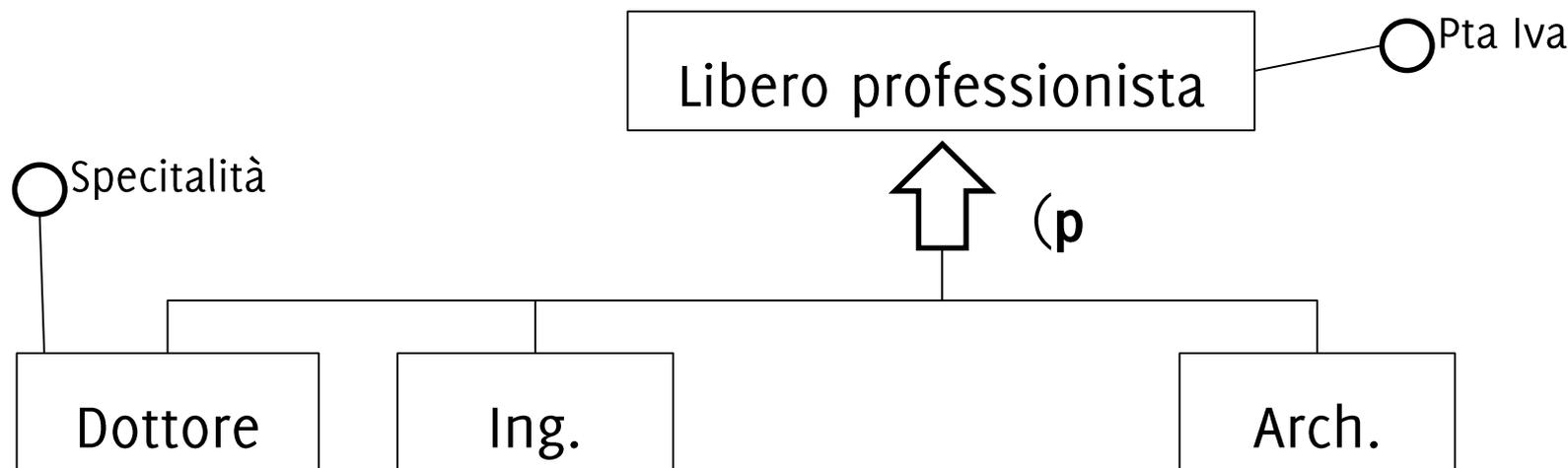




## Generalizzazioni: totali vs parziali

### Generalizzazioni: **totali** vs **parziali**

- **Totale:** ogni istanza dell'entità genitore è occorrenza di una o più delle sue entità figlie
- **Parziale:** esistono istanze dell'entità padre che non rientrano in nessuna entità figlia

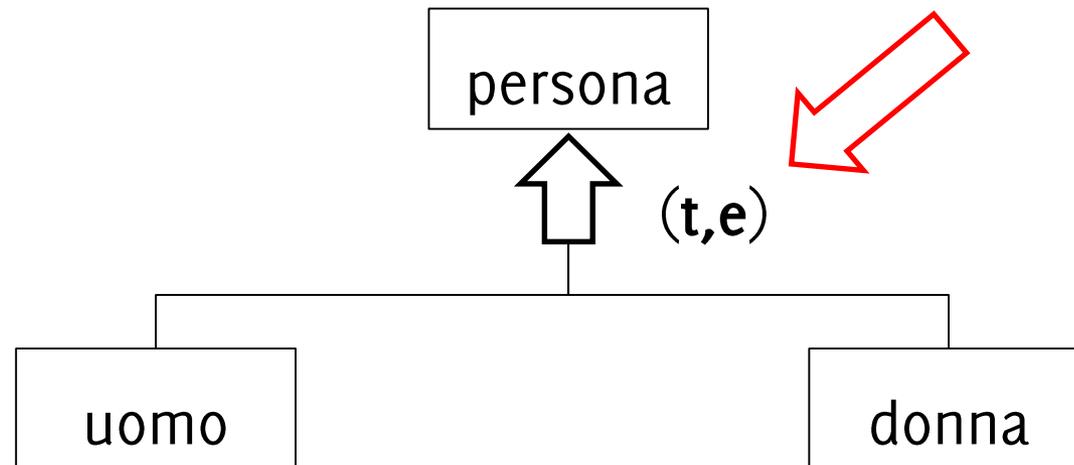




## Generalizzazioni: esclusiva vs sovrapposta

### Generalizzazioni: esclusiva vs sovrapposta

- **Esclusiva:** ogni istanza dell'entità genitore è istanza di al più un'entità figlia

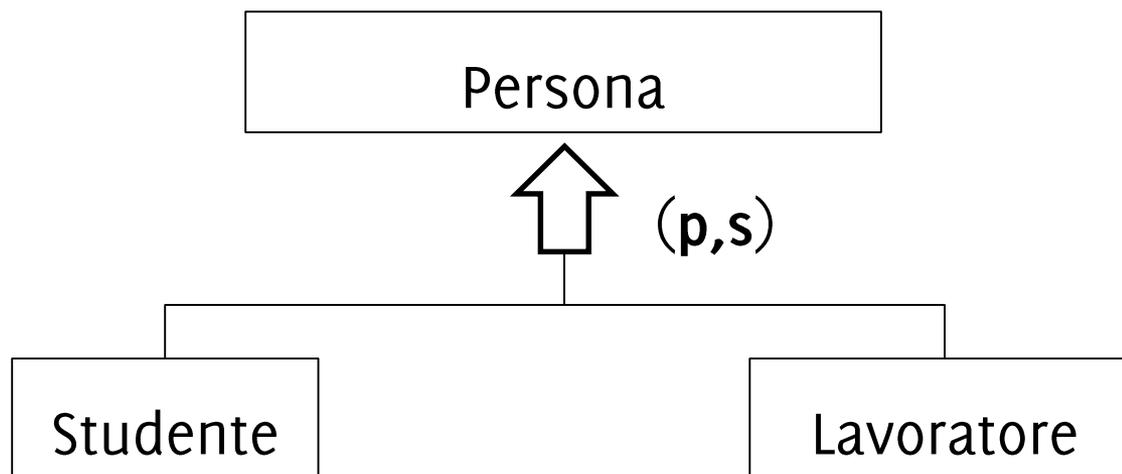




## Generalizzazioni: esclusiva vs sovrapposta

### Generalizzazioni: esclusiva vs sovrapposta

- **Esclusiva:** ogni istanza dell'entità genitore è istanza di al più un'entità figlia
- **Sovrapposta:** esistono istanze dell'entità padre che sono istanze di più entità figlie

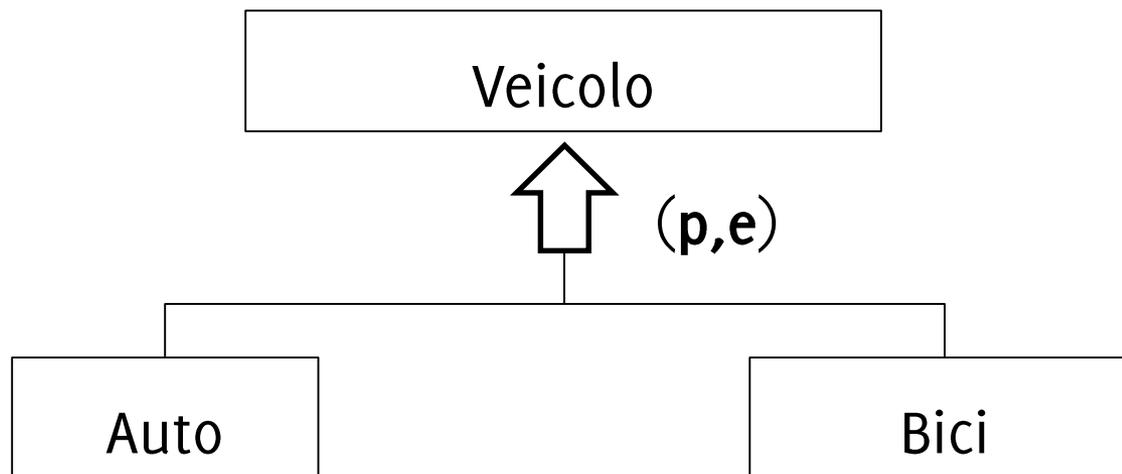




## Generalizzazioni: esclusiva vs sovrapposta

### Generalizzazioni: esclusiva vs sovrapposta

- **Esclusiva:** ogni istanza dell'entità genitore è istanza di al più un'entità figlia
- **Sovrapposta:** esistono istanze dell'entità padre che sono istanze di più entità figlie

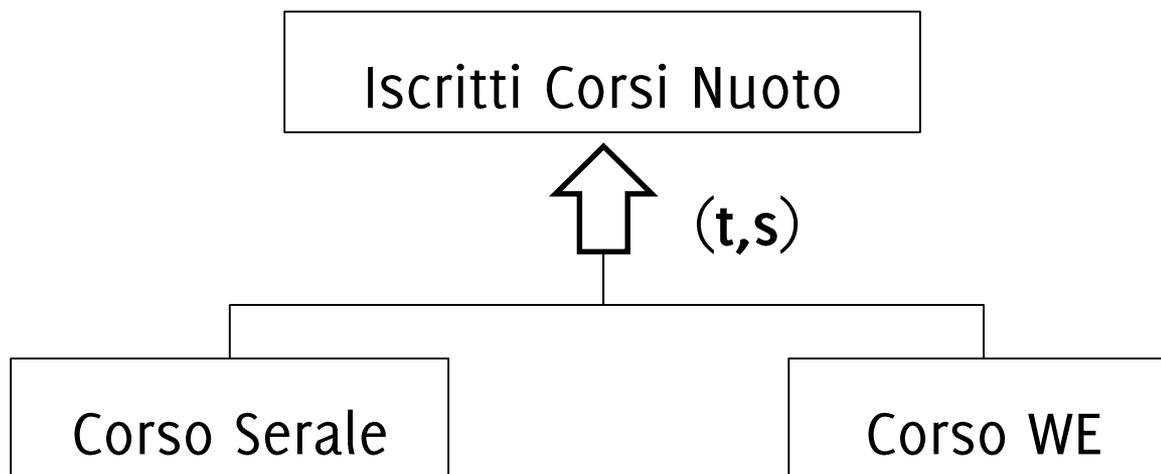




## Generalizzazioni: esclusiva vs sovrapposta

### Generalizzazioni: esclusiva vs sovrapposta

- **Esclusiva:** ogni istanza dell'entità genitore è istanza di al più un'entità figlia
- **Sovrapposta:** esistono istanze dell'entità padre che sono istanze di più entità figlie





## Esempio TDE Febbraio 2016 (Recupero)

Si progetti una base dati atta a descrivere gli esami del sangue che vengono svolti in una clinica privata. L'anagrafica dei clienti della clinica viene registrata nel sistema ed i clienti sono identificati da un univoco codice cliente. La clinica svolge un numero ben definito di tipologie di esami del sangue. Ogni esame è identificato dal nome (es. Trigliceridi), e da un intervallo di normalità (valore massimo, valore minimo), da un costo e da un codice identificativo. Quando un cliente fa un esame viene registrata la data in cui questo avviene ed i rispettivi valori. Tra i clienti vi è particolare attenzione per le donne incinta e per queste viene tenuta traccia della data attesa del parto, del numero di figli che già ha. La clinica esegue, solo per le donne incinta, esami particolari (e.g. ecografia morfologica) di cui si registra, viene dispensato dal SSN, se è un'esame a rischio, e se può essere fatto a digiuno. Per ogni esame svolto da una donna incinta si tiene traccia, oltre che del valore, del mese di gestazione. Infine, tutti i referti degli esiti vengono tenuti in una cartella clinica riposta in un magazzino pieno di scaffali numerati e con riferimenti di posizione. La cartella clinica e tutti i referti di un paziente possono stare su un solo scaffale



## Esempio TDE Febbraio 2016 (Recupero)

Progettare lo schema **Entità-relazione** per progettare la base dati del sistema sopra descritto (si inseriscano brevi commenti solo se necessari per giustificare alcune scelte progettuali)

Progettare lo **schema relazionale** della base dati. (si inseriscano brevi commenti solo se necessari per giustificare alcune scelte progettuali)



## ESERCIZIO 1

Realizzare uno schema ER relativo alla seguente descrizione: Si ha una società che gestisce un servizio di trasporto con autobus. L'azienda possiede un certo numero di autobus, di diversi modelli, che vengono utilizzati su diversi percorsi (un autobus può eseguire corse su percorsi diversi, così come gli autisti e i controllori possono utilizzare diversi autobus e diversi percorsi). Il personale dell'azienda è costituito da autisti e controllori (che svolgono servizio su determinate corse), addetti alla vendita (che risiedono in alcune fermate) e addetti alla amministrazione ed alla manutenzione che lavorano alla sede centrale. Di ogni dipendente bisogna memorizzare le solite informazioni anagrafiche. Di ogni percorso si devono memorizzare il luogo di partenza, di arrivo e le fermate intermedie. I mezzi sono conservati durante la notte in depositi, ogni autobus è associato a un particolare deposito. Ogni autista è associato a un deposito, e la prima e l'ultima corsa della giornata verranno effettuate sempre su un autobus del deposito a cui è assegnato, ma le altre corse possono essere svolte anche su mezzi facenti capo ad altri depositi.