

# **BASI DI DATI**

2018/2019

Prof. Alberto Belussi

*Enrico Martini*

## **IL SISTEMA INFORMATIVO**

Insieme delle attività umane e dei dispositivi di memorizzazione ed elaborazione che organizza e gestisce l'informazione di interesse per un'organizzazione di dimensioni qualsiasi.

**DATO:** elemento di conoscenza di base, costituito da simboli che devono essere elaborati.

**INFORMAZIONE:** interpretazione dei dati che permette di ottenere conoscenza più o meno esatta di fatti e situazioni.

**BASE DI DATI:** collezione di dati utilizzati per rappresentare con tecnologia informatica le informazioni di interesse per un sistema informativo.

### **Perché non salviamo i dati su file?**

- Scarsa efficienza a causa dell'accesso sequenziale;
- Possibile ridondanza nei dati, con conseguente possibile inconsistenza;
- Accesso concorrente ai file;
- Progettazione dei dati replicata per ogni programma.

## **DBMS (Data Base Management System)**

Sistema che gestisce su memoria secondaria collezioni di grandi dati, condivise e persistenti. Assicura affidabilità, privacy e accesso efficiente.

Vi si interagisce attraverso diversi linguaggi:

- DDL: linguaggio per la definizione dei dati;
- DML: linguaggio per interrogazione e aggiornamento dati;

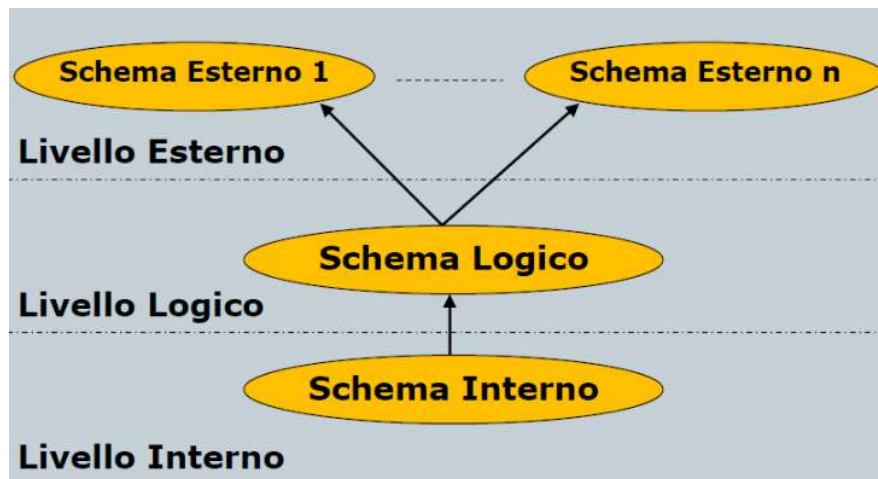
**Modello di dati:** insieme dei costrutti forniti dal DBMS per descrivere struttura e proprietà dell'informazione contenuta in una base di dati.

**Costrutti:** definiscono le strutture dati che conterranno le informazioni della base di dati e specificano le proprietà che dovranno soddisfare le istanze di informazione che saranno contenute nelle strutture dati.

**Schema di una base di dati:** descrizione della struttura e delle proprietà di una specifica base di dati utilizzando i costrutti del modello di dati.

**Istanza di una base di dati:** costituita dai valori effettivi che in un certo istante popolano le strutture dati della base di dati.

## ARCHITETTURA DI UN DBMS



**Schema esterno:** descrive una porzione dello schema logico di interesse per uno specifico utente o applicazione.

**Schema logico:** rappresentazione della struttura e delle proprietà della base di dati attraverso i costrutti del modello dei dati del DBMS

**Schema interno:** rappresentazione della base di dati per mezzo delle strutture fisiche di memorizzazione.

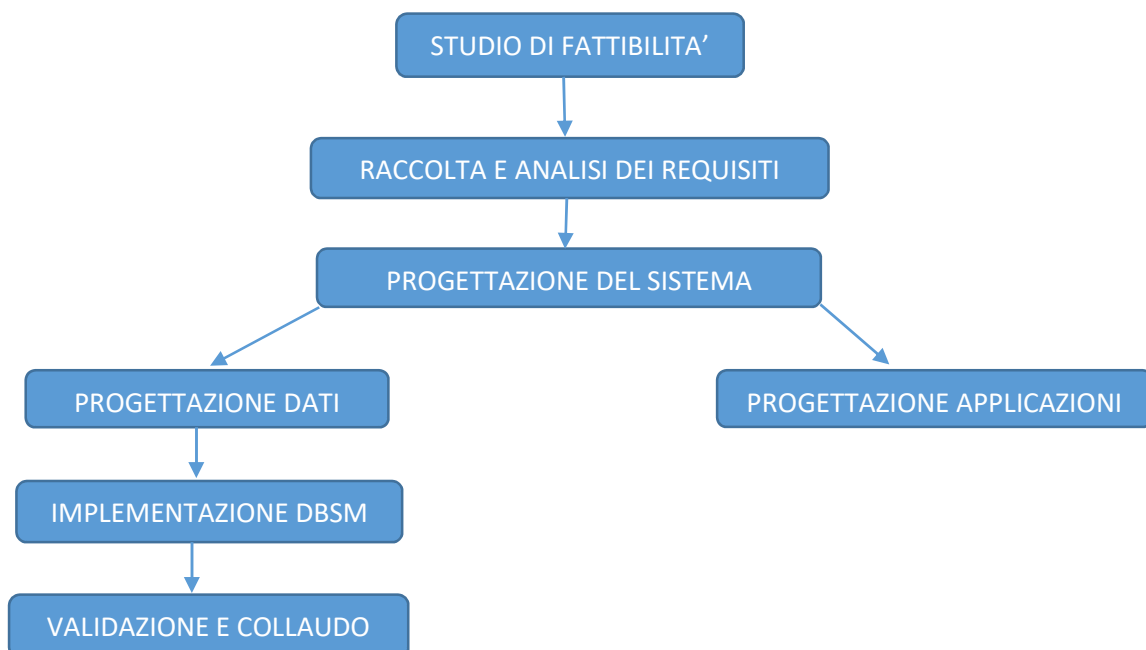
### ***Indipendenza FISICA***

Lo schema logico è indipendente dallo schema interno, perciò eventuali variazioni delle strutture fisiche non impattano sullo schema logico e quindi sulle applicazioni.

### ***Indipendenza LOGICA***

Gli schemi esterni della base di dati sono indipendenti dallo schema logico, perciò ogni variazione dello schema logico non impatta sugli schemi esterni e quindi sulle applicazioni.

## PROGETTAZIONE DI UNA BASE DI DATI



## **METODOLOGIA DI PROGETTAZIONE DEI DATI**

- 1) **PROGETTAZIONE CONCETTUALE:** rappresentare il contenuto informativo della base di dati in modo formale ma indipendente dall'implementazione e dalle operazioni;
- 2) **PROGETTAZIONE LOGICA:** tradurre lo schema concettuale in schema logico aderente al modello dei dati del DBMS scelto. Si tiene conto delle operazioni più frequenti che le applicazioni eseguiranno;
- 3) **PROGETTAZIONE FISICA:** completare lo schema logico con i parametri relativi alla memorizzazione fisica dei dati e con gli opportuni indici per garantire un accesso efficiente ai dati.

## **SCHEMA CONCETTUALE**

Documento formale che rappresenta il contenuto della base di dati in modo indipendente dall'implementazione.

Non è solo un prodotto intermedio del processo di progettazione della base di dati, ma costituisce una porzione del risultato finale. È una descrizione ad alto livello del contenuto della base di dati.

## PROGETTAZIONE DIGITALE – MODELLO ENTITA' RELAZIONALE (ER)

### 1) ENTITA'

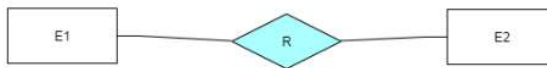
Classe di oggetti che hanno proprietà comuni, esistenza autonoma e identificazione univoca.

Un istanza di un entità è un oggetto che appartiene alla classe rappresentata da E. Si indica con I(E) l'insieme delle istanze di E che esistono nella base di dati in un certo istante.



### 2) RELAZIONE

Legame logico tra due o più entità. Possono esistere relazioni binarie, ternarie, ennarie. La relazione può essere anche ricorsiva. Non è possibile rappresentare la stessa enupla più volte.



### 3) ATTRIBUTO

Proprietà elementare di un entità o relazione. Ogni attributo associa ad ogni istanza di entità o relazione, UNO e UNO SOLO valore appartenente al dominio dei valori ammissibili.

attributo a dell'entità E



attributo b della relazione R



### 4) IDENTIFICATORE DI IDENTITA'

Insieme di proprietà che identificano univocamente le istanze di entità. Si dice interno se contiene solo attributi, esterno se contiene anche relazioni.

entità E con identificatore a



entità E con identificatore a,b



entità E con identificatore a,R



## 5) VINCOLI DI CARDINALITA'

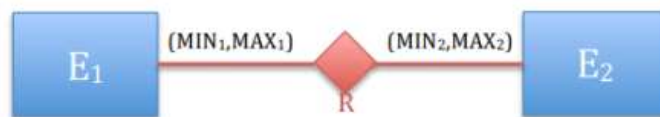
Numero minimo e numero massimo di occorrenze a cui un'istanza di E può partecipare in una relazione.

Valori minimi:

- 0: la partecipazione è opzionale;
- 1: la partecipazione è obbligatoria;
- >1: partecipazione minima 1.

Valori massimi:

- 1: la partecipazione massima è di 1 istanza;
- n: la partecipazione di quell'istanza è infinita;
- >1: per ogni istanza, devono essere presenti.



## 6) ATTRIBUTO OPZIONALE / MULTIVALEORE

Specificano un vincolo di cardinalità sui valori che l'attributo può assumere.

Valori possibili:

- (0,1): attributo opzionale;
- (1,N): attributo multivalore;
- (0,N): attributo opzionale multivalore.



## 7) GENERALIZZAZIONE

Legame logico (simile a una ereditarietà tra classi) tra un'entità padre E e n (con  $n > 0$ ) entità figlie E1, ..., En, dove E rappresenta una classe di oggetti più generale rispetto alle classi di oggetti rappresentate dalle entità E1, ..., En.

- Ogni istanza di un'entità figlia Ei è anche istanza dell'entità padre E.
- Ogni proprietà (attributi, identificatori e relazioni) dell'entità padre E è anche proprietà di ogni istanza delle entità figlie E1, ..., En

Una generalizzazione si dice totale se ogni istanza dell'entità padre E è anche istanza di almeno un'entità figlia Ei, altrimenti si dice parziale. Una generalizzazione si dice esclusiva se ogni istanza dell'entità padre E è istanza al più di un'entità figlia Ei, altrimenti si dice sovrapposta. Il tipo di generalizzazione si indica con la stenografia seguente: - totale e esclusiva: (t,e) - totale e sovrapposta: (t,s) - parziale e esclusiva: (p,e) - parziale e sovrapposta: (p,s) posta vicino alla freccia che indica la generalizzazione.

## MODELLO RELAZIONALE

**DOMINI DI BASE:** Domini da cui si scelgono i valori delle proprietà delle istanze di informazione da rappresentare (caratteri, stringhe, interi, decimali, ...).

**COSTRUTTO RELAZIONALE:** una relazione può essere vista come una tabella, un contenitore dati la cui struttura è caratterizzata da una lista di colonne.

### RELAZIONE come insieme di ennuple

Dati  $n$  insiemi di valori (domini)  $D_1, \dots, D_n$  con  $n > 0$  e indicato con  $D_1 \times \dots \times D_n$  il loro prodotto cartesiano:

$$D_1 \times \dots \times D_n = \left\{ (v_1, \dots, v_n) \mid v_1 \in D_1 \wedge \dots \wedge v_n \in D_n \right\}$$

una relazione  $\rho$  di grado  $n$  è un qualsiasi sottoinsieme di  $D_1 \times \dots \times D_n$ :

$$\rho \subseteq D_1 \times \dots \times D_n$$

dove:

$(v_1, \dots, v_n)$  è una ennupla della relazione

$|\rho|$  è la cardinalità della relazione (numero di ennuple)

Dalla definizione si può dedurre che:

- Non è definito alcun ordinamento sulle ennuple di una relazione;
- Non sono ammessi duplicati di una ennupla;
- I valori nelle ennuple sono ordinati.

Per accedere ai valori di una ennupla  $t$ , si utilizza la notazione  $t[i]$ .

### RELAZIONE come insieme di tuple

Sia  $X$  un insieme di nomi e sia  $\Delta$  l'insieme di tutti i domini di base ammessi dal modello. Si definisce la funzione:

$$\text{DOM}: X \rightarrow \Delta$$

Che associa ad ogni nome  $A$  di  $X$  un dominio  $\text{DOM}(A)$  di  $\Delta$ . I nomi di  $X$  si dicono **attributi**.

Una tupla  $t$  su  $X$  è una funzione:  $t: X \rightarrow \bigcup_{A \in X} \text{DOM}(A)$

dove:  $t[A] = v \in \text{DOM}(A)$

**Una relazione su  $X$  è un insieme di tuple su  $X$ , dove  $X$  è l'insieme di attributi della relazione.**

- Una relazione non può contenere tuple duplicate;
- I domini per gli attributi possono essere solo domini di base, non sono ammessi altri domini;
- Una base di dati relazionale è costituita da più relazioni.

## IL PROBLEMA DELLA RIDONDANZA

Se in una relazione si uniscono concetti disomogenei e con esistenza autonoma, per le istanze di informazione che sono coinvolte in più associazioni si produce una ripetizione inutile di valori in diverse tuple.

La presenza di ridondanza produce:

- Anomalia di aggiornamento: per aggiornare il valore di un attributo si è obbligati a modificarlo su più tuple;
- Anomalia di inserimento: per inserire una nuova istanza di un concetto è necessario inserire valori al momento sconosciuti per gli attributi non disponibili;
- Anomalia di cancellazione: per cancellare un'istanza di un concetto è necessario cancellare valori ancora validi oppure inserire valori nulli per gli attributi da cancellare.

Un metodo per eliminare la ridondanza consiste nel decomporre la relazione unica in più tabelle, seguendo delle regole per non perdere informazione.

Per conservare il legame logico che legherà le tabelle appena composte, è necessario replicare parte degli attributi scegliendo tra questi quelli che hanno la proprietà di identificare univocamente il concetto verso il quale si vuole generare il legame.

Il modello relazionale è VALUE-BASED:

- Totalmente indipendente dalla rappresentazione fisica;
- I legami logici tra tuple si rappresentano con la replicazione di attributi;
- È facile trasferire i dati da una base di dati all'altra;
- È rappresentato solo ciò che è rilevante per l'applicazione.

### Terminologia:

- Schema di una relazione: è costituito dal nome della relazione e da un insieme di nomi per i suoi attributi:
  - $R(A_1, \dots, A_n)$  oppure  $R(A_1:D_1, \dots, A_n:D_n)$  dove  $DOM(A_i) = D_i$
- Schema di una base di dati: è un insieme di schemi di relazione:
  - $S = \{R_1(A_{1,1}, \dots, A_{1,n_1}), \dots, R_m(A_{m,1}, \dots, A_{m,n_m})\}$  dove  $R_1 \neq \dots \neq R_m$
- Istanza di una relazione di schema  $R(A_1, \dots, A_n)$  con  $X = \{A_1, \dots, A_n\}$ : è un insieme  $r$  di tuple su  $X$ ;
- Istanza di una base di dati di schema  $S = \{R_1(A_{1,1}, \dots, A_{1,n_1}), \dots, R_m(A_{m,1}, \dots, A_{m,n_m})\}$ : è un insieme di istanze di relazioni  $db = \{r_1, \dots, r_m\}$  dove ogni  $r_i$  è un'istanza della relazione di schema  $R_i (A_{i,1}, \dots, A_{i,n_i})$ .



## INFORMAZIONE INCOMPLETA

Non sempre in una base di dati reale esistono i valori per tutti gli attributi di una tupla.

- Il valore di un attributo è inesistente se non esiste per questa tupla un valore per quell'attributo.
- Il valore di un attributo è sconosciuto se esiste un valore per l'attributo di questa tupla ma non è noto alla base di dati.

Per gestire queste situazioni si ricorre al valore speciale: **valore nullo**. Gli attributi di una tupla possono assumere un valore del dominio oppure un valore nullo.

Def. Tupla su X con valori nulli

Una tupla su X è una funzione:

$$t : X \rightarrow \{NULL\} \cup \left( \bigcup_{A \in X} DOM(A) \right)$$

dove:

$$t[A] = v \in DOM(A) \vee t[A] = NULL$$

- La presenza di valori nulli è accettabile solo in alcuni attributi;
- Non è accettabile avere tuple di soli valori nulli;
- Negli attributi replicati per rappresentare legami tra tuple l'utilizzo di valori nulli può rendere inutilizzabile l'informazione.

## VINCOLI DI INTEGRITA'

Il vincolo di integrità è una condizione espressa da un predicato che deve sempre essere soddisfatta da ogni istanza della base di dati.

Spesso è necessario introdurre restrizione sulla popolazione di una base di dati in quanto non tutte le possibili istanze sono corrette rispetto al sistema informativo considerato.

Esistono varie categorie di vincoli:

### 1) VINCOLI DI DOMINIO

Impongono una restrizione sul dominio dell'attributo di una relazione.

Es:  $\forall t \in \text{TRENO} : t[\text{OraPart}] \in \{0, 1, \dots, 23\}$

### 2) VINCOLI DI TUPLA

Impongono una restrizione alla combinazione di valori che una tupla della relazione può assumere indipendentemente dalle altre tuple.

Es:  $\forall t \in \text{TRENO} : t[\text{Numero}] > 5000 \Rightarrow t[\text{Categoria}] = \text{'regionale'}$

### 3) VINCOLI INTRARELAZIONALI

Impongono una restrizione al contenuto di una relazione e specificano una condizione che ogni tupla della relazione deve soddisfare rispetto alle altre tuple della medesima relazione. Una sottocategoria importante include i vincoli di chiave:

- **SUPERCHIAVE**

Data una relazione di schema  $R(X)$ , un insieme di attributi  $K$ , sottoinsieme di  $X$ , è SUPERCHIAVE per  $R(X)$  se, per ogni istanza  $r$  di  $R(X)$  vale la seguente condizione:

$$\forall t, t' \in r : t \neq t' \Rightarrow t[K] \neq t'[K]$$

dove:

$$t[K] \neq t'[K] \equiv \exists A_i \in K : t[A_i] \neq t'[A_i]$$

- **CHIAVE CANDIDATA**

Data una relazione di schema  $R(X)$ , un insieme di attributi  $K$ , sottoinsieme di  $X$ , è CHIAVE CANDIDATA (o CHIAVE) per  $R(X)$ , se  $K$  è superchiave per  $R(X)$  e vale la seguente condizione:

$$\neg \exists K' \subset K : K' \text{ è superchiave per } R(X)$$

- **CHIAVE PRIMARIA**

Data una relazione di schema  $R(X)$  la sua CHIAVE PRIMARIA è la chiave candidata scelta per identificare le tuple della relazione.

Una chiave primaria  $K$  ha le seguenti caratteristiche:

- Non contiene mai valori nulli
- Su  $K$  il sistema genera una struttura d'accesso ai dati (o indice) per supportare le interrogazioni.

#### 4) VINCOLI INTERRELAZIONALI

Impongono una restrizione al contenuto di una relazione e specificano una condizione che ogni tupla della relazione deve soddisfare rispetto alle tuple di altre relazioni della base di dati. Una sottocategoria importante di tali vincoli include i vincoli di integrità referenziale.

**Def. Vincolo di integrità referenziale**

Un vincolo di integrità referenziale tra un insieme di attributi  $Y=\{A_1, \dots, A_p\}$  di  $R_1$  e un insieme di attributi  $K=\{K_1, \dots, K_p\}$ , chiave primaria di un'altra relazione  $R_2$ , è soddisfatto se, per ogni istanza  $r_1$  di  $R_1$  e per ogni istanza  $r_2$  di  $R_2$  vale la seguente condizione:

$$\forall t \in r_1 : \exists s \in r_2 : \forall i \in \{1, \dots, p\} : t[A_i] = s[K_i]$$

**Def. Vincolo di integrità referenziale in presenza di valori nulli (legame opzionale)**

Un vincolo di integrità referenziale tra un insieme di attributi  $Y=\{A_1, \dots, A_p\}$  di  $R_1$  e un insieme di attributi  $K=\{K_1, \dots, K_p\}$ , chiave primaria di un'altra relazione  $R_2$ , è soddisfatto se, per ogni istanza  $r_1$  di  $R_1$  e per ogni istanza  $r_2$  di  $R_2$  vale la seguente condizione:

$$\forall t \in r_1 : \exists s \in r_2 : (\forall i \in \{1, \dots, p\} : t[A_i] = s[K_i]) \vee (\exists i \in \{1, \dots, p\} : t[A_i] = NULL)$$