

Progettazione concettuale

La costruzione di uno schema concettuale deve tenere conto di alcune **proprietà generali** che ne determinano la **qualità**:

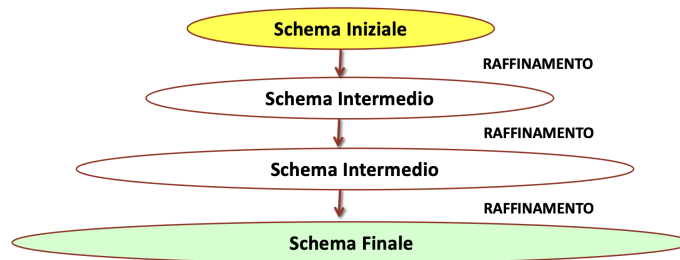
- **Correttezza**: utilizzo corretto sintattico e semantico dei costrutti del modello E/R
- **Completezza**: rappresentazione di tutti i dati d'interesse descritti nel documento di specifica

Per garantire tali proprietà vengono utilizzate diverse **metodologie di progettazione concettuale**. Alcune di queste possono essere:

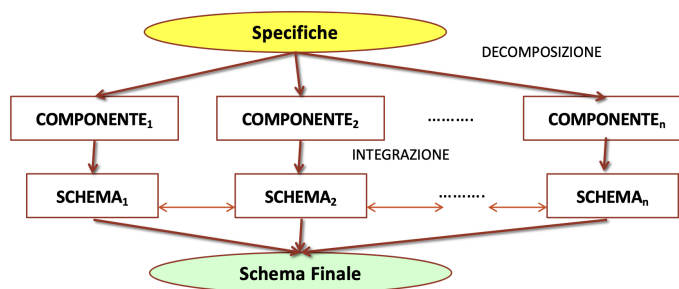
Strategie di progettazione

Il documento di specifica potrebbe essere **molto complesso** e **denso di contenuti**. Per costruire il modello E/R si possono utilizzare diverse **strategie**:

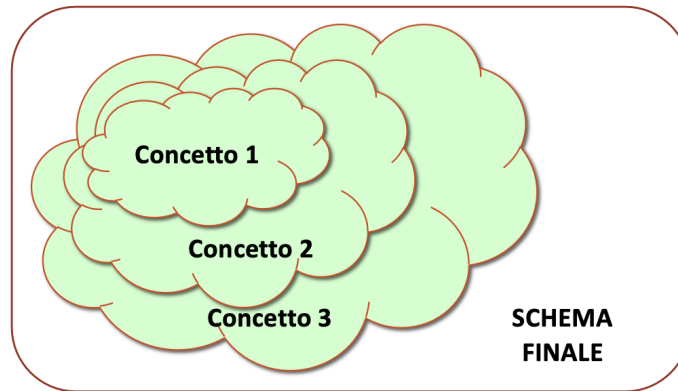
- **Strategia top-down**: lo schema concettuale viene ottenuto attraverso una **serie di raffinamenti successivi a partire da uno schema iniziale** molto astratto



- **Strategia bottom-up**: le **specifiche iniziali** sono suddivise in **componenti** via via più piccole, ed in un secondo momento i **vari schemi** sono **integrati** tra loro



- **Strategia inside-out**: si **individuano una serie di concetti importanti** e poi si procede a **partire da questi verso concetti correlati**, con un'estensione a macchia d'olio. E' un caso particolare della strategia *bottom-up*



- **Strategia mista:** si utilizza una combinazione delle strategie precedenti. Si suddivide quindi in diversi step:
 1. Si individuano i **concetti principali** o più citati
 2. Si realizza uno **schema scheletro**
 3. **Si decompone** lo schema
 4. **Si raffina** lo schema, si espande e si integra

In molti casi pratici di una certa complessità, la strategia mista rappresenta la scelta migliore

Pattern di progettazione

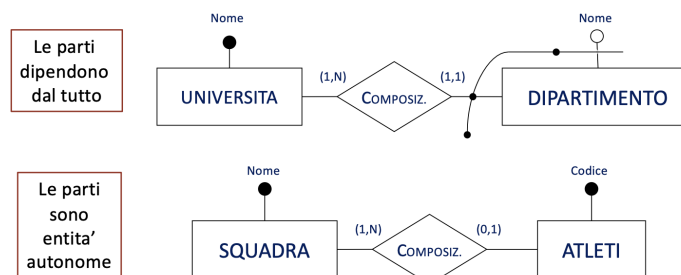
Si parte dal concetto che **non esiste una rappresentazione univoca delle specifiche**, quindi è meglio attenersi alle **Regole Concettuali (RC)** del diagramma E/R. Le tre *regole concettuali* sono:

- **RC1:** se un concetto ha proprietà significative e descrive oggetti con esistenza autonoma, si utilizzano **Entità**
- **RC2:** se un concetto correla due o più Entità, si utilizzano **Relazioni**
- **RC3:** se un concetto è un caso particolare dell'altro, si utilizzano **Generalizzazioni**

Esistono molti **pattern**, ovvero **soluzioni di problemi ricorrenti**, usati nella progettazione concettuale. Se ne osservino alcuni.

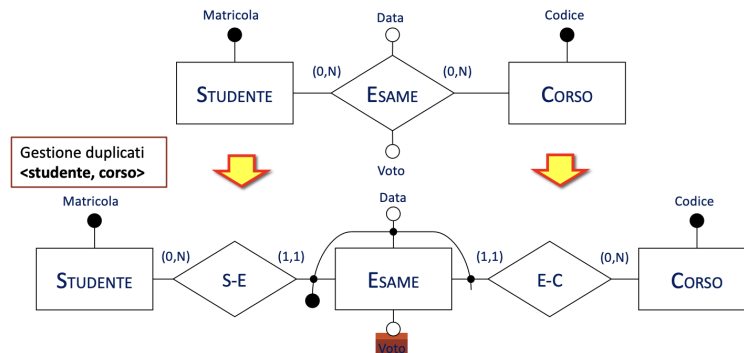
Pattern₁

Concetti di tipo "**parte di**" attraverso l'utilizzo di **relazioni uno a molti**.



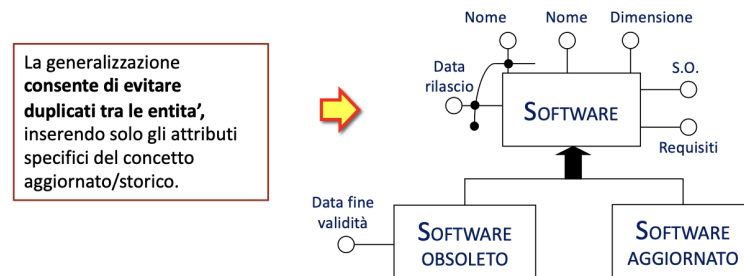
Pattern₂

Introduzione di **nuove entità in relazioni uno a molti** per la gestione dei duplicati.



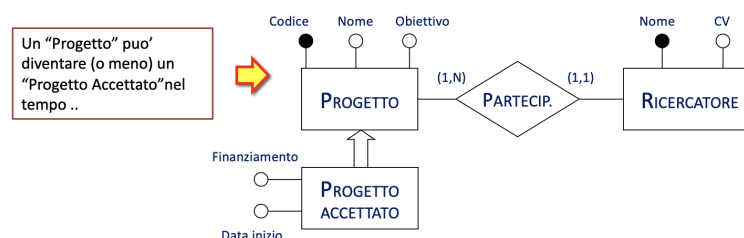
Pattern₃

Utilizzo di **generalizzazioni** per tenere traccia della **storia di un concetto**, ossia della sua istanza attuale e di quelle pregresse.



Pattern₄

Utilizzo di **generalizzazioni** per tenere traccia dell'**evoluzione nel tempo di un certo concetto**, ossia la creazione di nuove istanze diverse dal concetto originario.



Analisi di prestazione

Una volta realizzato il modello E/R, è importante analizzarne l'**efficienza dal punto di vista prestazionale**. Alcuni indici di prestazione possono essere:

- **Costo operativo:** numero di entità/associazioni mediamente visitate per implementare una certa operazione sui dati
- **Occupazione di memoria:** spazio di memoria necessario per memorizzare i dati

Per poter stimare correttamente l'**efficienza prestazionale** di uno schema E/R, abbiamo necessità di **informazioni aggiuntive**.

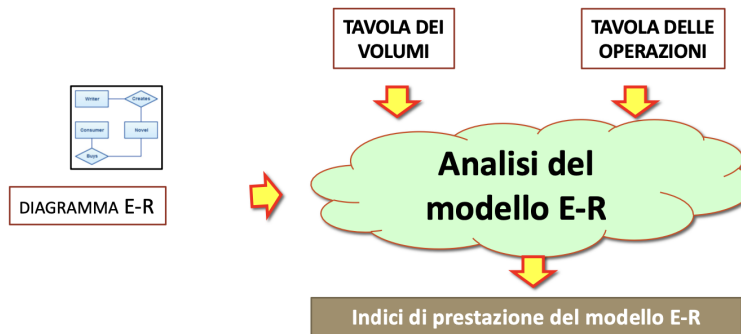


Tavola dei volumi

La **tavola dei volumi** fornisce una **stima del numero di occorrenze entità/relazioni** presenti nel modello E/R. Si osservi un esempio:

Concetto	Tipo	Volume
Progetto	E	100
Release	E	1000
Dipendente	E	500
Tecnico	E	200
Sviluppatore	E	100
Programmatore	E	70
Analista	E	30

Stima del **numero medio** di occorrenze di una entità

Concetto	Tipo	Volume
Direzione	R	100
Partecipazione	R	2000
Versioni	R	1000

Stima del **numero medio** di occorrenze di una relazione

In questo caso, le **assunzioni** possibili sono due: ogni progetto ha in media 10 *release*, e ad ogni progetto lavorano in media 20 *dipendenti*.

Tavola delle operazioni

La **tavola delle operazioni** definisce l'**insieme delle operazioni** che devono essere implementate. Inoltre definisce la **tipologia** delle operazioni (**interattive/batch**) e la **frequenza** delle operazioni.

Le informazioni riguardanti le due tavole vengono fornite durante la **raccolta ed analisi dei requisiti**.

Si osservi un esempio di tavola delle operazioni:

Operazione	Tipo	Frequenza
Operazione ₁	I	10 volte/giorno
Operazione ₂	I	100 volte/giorno
Operazione ₃	B	5 volte/giorno

Operazione₁: assegnare un dipendente ad un progetto

Operazione₂: visualizzare tutti i dati di un progetto, delle release associate e del direttore

*Operazione*₃: per ciascun progetto, visualizzare tutti i dati dei dipendenti associati

Si definisce ora il concetto di **costo di una operazione**. Data un'operazione O di tipo T , definiamo il suo **costo** $c(O_T)$ come:

$$c(O_T) = f(O_T) \cdot w_T \cdot (\alpha \cdot NCwrite + NCread)$$

Si osservi ora il glossario dei seguenti simboli:

$f(O_T)$: frequenza dell'operazione

$NCread$: numero di **accessi in lettura** a componenti dello schema

$NCwrite$: numero di **accessi in scrittura** a componenti dello schema

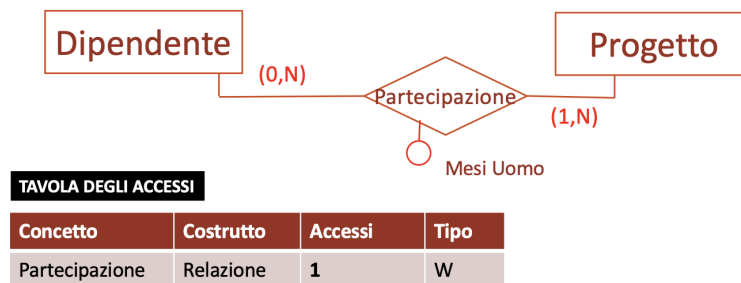
w_T : **peso** dell'operazione

α : **coefficiente moltiplicativo** delle operazioni in scrittura

Tornando all'esempio delle operazioni visto in precedenza, calcoliamo il costo delle tre operazioni:

*Operazione*₁

Assegnare un dipendente ad un progetto. La sua frequenza è di **10 volte al giorno**.



I parametri in questo caso sono:

$f(O_T)$: 10

$NCread$: 0

$NCwrite$: 1

w_T : 0.5

α : 2

Ottenendo un costo di

$$c(Operazione1) = 10 \cdot 0.5 \cdot (2 \cdot 1 + 0) = 10$$

Operazione₂

Visualizzare tutti i dati di un progetto, delle release associate e del direttore. La frequenza è di **100 volte al giorno**.

TAVOLA DEGLI ACCESSI			
Concetto	Costrutto	Accessi	Tipo
Progetto	Entita'	1	L
Versioni	Relazione	10	L
Release	Entita'	10	L
Direzione	Relazione	1	L
Dipendente	Entita'	1	L

I parametri in questo caso sono:

$$f(O_T): 100$$

$$NC_{read}: 23$$

$$NC_{write}: 0$$

$$w_T: 0.5$$

$$\alpha: 2$$

Ottenendo un costo di

$$c(Operazione2) = 100 \cdot 0.5 \cdot (2 \cdot 0 + 23) = 1150$$

Operazione₃

Per ciascun progetto, visualizzare tutti i dati dei dipendenti associati. La frequenza è di **5 volte al giorno**.

TAVOLA DEGLI ACCESSI			
Concetto	Costrutto	Accessi	Tipo
Progetto	Entita'	100	R
Partecipazione	Relazione	2000	R
Dipendente	Entita'	2000	R

I parametri in questo caso sono:

$$f(O_T): 5$$

$$NC_{read}: 4100$$

$$NC_{write}: 0$$

$$w_T: 0.5$$

$$\alpha: 2$$

Ottenendo un costo di

$$c(\text{Operazione3}) = 5 \cdot 0.5 \cdot (2 \cdot 0 + 4100) = 10250$$

Si osservi ora come calcolare il **costo dello schema** completo. Dato uno schema S , ed un'insieme di operazioni sui dati $O1, O2, \dots, On$, con costi $c(O1), c(O2), c(On)$, il **costo dello schema** è definito come:

$$c(S) = \sum_{i=1}^n c(Oi)$$

Nell'esempio precedente si ottiene $(10 + 1150 + 10250) = 11410$.

L'obiettivo del progettista è quello di **determinare lo schema E/R di costo minimo**. Conoscendo la tavola dei volumi, il tipo di ciascun attributo e la sua dimensione del disco, è possibile stimare l'**occupazione di memoria dello schema**.

$$M(S) = \sum_{entita'e} V(e) \cdot size(e) + \sum_{relazione r} V(r) \cdot size(r)$$

Dove i parametri indicano:

$V(e), size(e)$: tabella dei volumi e dimensione in termini di occupazione di memoria dell'entità e

$V(r), size(r)$: tabella dei volumi e dimensione in termini di occupazione di memoria della relazione r

Si osservi un esempio: come stimare l'occupazione di memoria dell'entità *Dipendente*.

Nome	Anno Assunzione
●	○
Dipendente	

Concetto	Tipo	Volume
Dipendente	E	500

Attributo	Tipo	Dimensione
Nome	Stringa (max 8 caratteri)	1B*8=8B
AnnoAss.	Intero	4B

La **Memoria Occupata** per l'entità *Dipendente* è uguale a : $500 * (8B + 4B) = 6000B$.

In pratica, si cerca di determinare il miglior **trade-off tra occupazione di memoria e costo delle operazioni dello schema**.

Gli indici di prestazione di un diagramma E/R sono forniti come input alla fase di **progettazione logica**, e sono utilizzati per la **traduzione dal modello concettuale** e per l'**analisi delle ridondanze**.