

Progettazione di Basi di Dati

1 Problema della progettazione

Progettare una base di dati significa *definirne la struttura e il contenuto*.

La progettazione è spesso un'operazione molto complicata, perché la base di dati può contenere numerose informazioni, che devono essere correttamente collegate tra loro. Per affrontare questo problema, si è definito un approccio sistematico, che suddivide l'attività di progettazione in una serie di sotto-fasi con diversi livelli di astrazione.

2 Fasi della progettazione

Le fasi della progettazione di una base di dati sono:

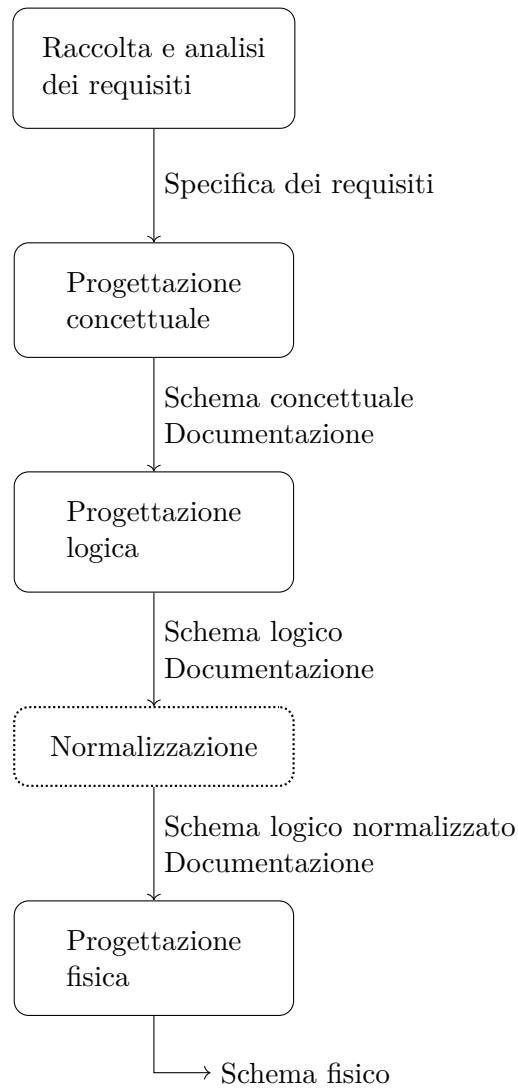
1. **raccolta e analisi dei requisiti;**
2. **progettazione concettuale:** modella i concetti, indipendentemente dal tipo di DBMS / schema logico usato;
3. **progettazione logica:** trasforma lo schema concettuale in uno schema logico, in base al modello usato dal DBMS adottato;
4. **normalizzazione;**
5. **progettazione fisica.**

Ogni fase si basa su un **modello**, che permette di rappresentare la base di dati a un determinato livello di astrazione in modo standardizzato. Inoltre, un modello formale permette anche di verificare determinate proprietà della base di dati progettata.

Le rappresentazioni della base di dati nei vari modelli, prodotte dalle diverse fasi della progettazione, devono sempre essere corredate da una documentazione, utile per:

- capire successivamente le scelte progettuali adottate;
- specificare eventuali vincoli che non possono essere espressi con un modello.

Infine, ogni fase della progettazione prevede opportune verifiche di qualità.



3 Raccolta e analisi dei requisiti

In questa fase, vengono definite le caratteristiche della base di dati e del dominio. Essa è solitamente condotta in maniera informale, qualitativa, mediante delle interviste con gli utenti e l'analisi di basi di dati esistenti, normativa, ambiente operativo, ecc.

L'output è un *documento di specifica dei requisiti*, che è scritto in linguaggio naturale, quindi è facile da comprendere, ma ci possono essere delle ambiguità.

3.1 Tipologie di requisiti

I requisiti si suddividono in:

- requisiti **informativi**: le caratteristiche e tipologie dei dati;
- requisiti **sulle operazioni**: il **carico di lavoro**, cioè che tipo di operazioni (ad alto livello) dovranno essere svolte, e quanto spesso saranno richieste;
- requisiti **sui vincoli di integrità e di autorizzazione**: le proprietà di correttezza e protezione da assicurare ai dati;
- requisiti **sulla popolosità della base di dati**: il **volume dei dati**, cioè quanti dati ci si aspetta di avere.

Tutte queste informazioni impattano diverse fasi della progettazione.

4 Progettazione concettuale

La progettazione concettuale viene introdotta per affrontare la progettazione della base di dati a un livello di astrazione più alto, con una visione libera dai limiti di dello schema logico, che verranno invece affrontati in seguito.

Questa fase usa un opportuno **modello concettuale** per definire lo **schema concettuale** della base di dati, cioè una descrizione formale e ad alto livello del suo contenuto. In particolare, esso definisce

- gli attori principali
- come essi sono collegati (relazioni)
- i vari vincoli
- ecc.

in modo indipendente dal modello logico del DBMS che verrà utilizzato.

Oltre allo schema concettuale, viene prodotta in output anche della documentazione, contenente tutto ciò che il modello concettuale non permette di rappresentare:

- le scelte progettuali;
- alcuni tipi di vincoli (ad esempio, di autorizzazione);
- ecc.

5 Progettazione logica

La progettazione logica consiste nella traduzione dallo schema concettuale a uno **schema logico**, definito in base al **modello logico** del DBMS adottato. Ad esempio, se si utilizza un DBMS relazionale, l'output di questa fase è un insieme di relazioni (corredato, come sempre, da un'opportuna documentazione).

Durante la progettazione logica, vengono anche considerati gli aspetti di:

- integrità;
- sicurezza (autorizzazioni);
- efficienza.

6 Normalizzazione

Dopo la progettazione logica, si usano appositi strumenti formali per verificare la qualità dello schema logico prodotto. Tale attività può portare a una ristrutturazione dello schema.

Nel caso di basi di dati relazionali, questa fase prende il nome di *normalizzazione*, e verifica, in particolare, che non ci siano dati ridondanti, mediante un'analisi basata sui fondamenti matematici del modello relazionale.

7 Progettazione fisica

In questa fase, vengono effettuate alcune scelte riguardo alla memorizzazione fisica dei dati (ad esempio, gli indici). L'output è lo **schema fisico**, che descrive le strutture di memorizzazione e accesso ai dati.