

Introduzione alle basi di dati

Matteo Gorgone
Università degli Studi di Messina

Che cos'è una base di dati

In ogni modello di organizzazione della vita dell'uomo, dal più complesso al più semplice, vengono trattate **informazioni** che sono risorse preziosissime, grazie alle quali la stessa organizzazione può sopravvivere.

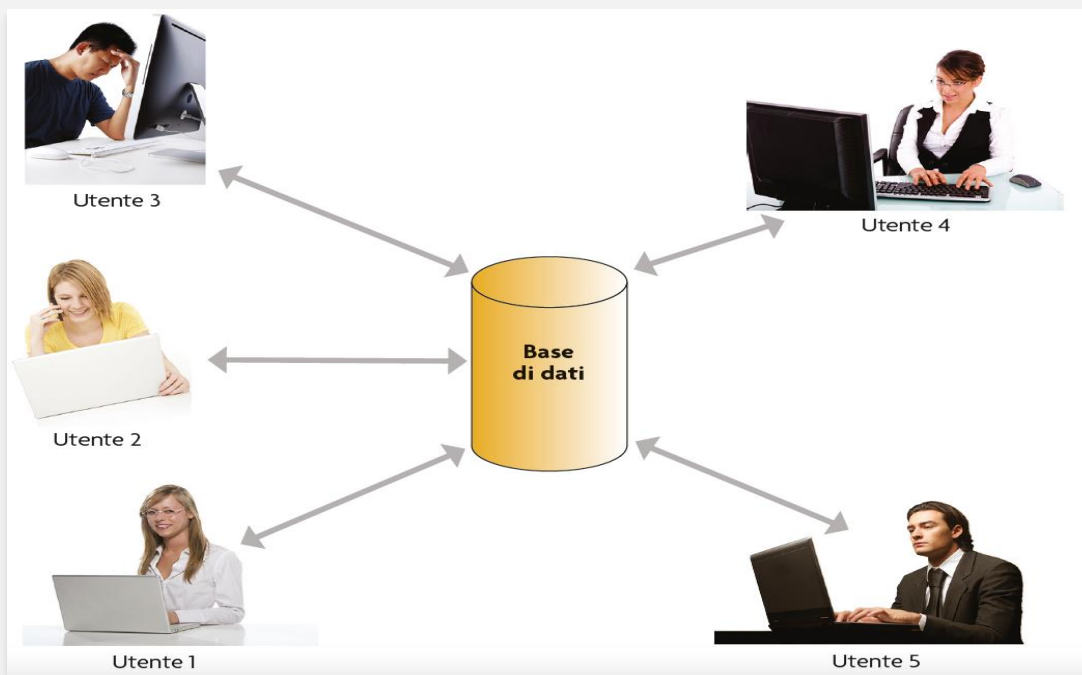
In campo informatico, la teoria delle **basi di dati** studia come organizzare al meglio grandi quantità di informazioni, per poterle gestire in modo:

- | | | | |
|---|---|--|--|
| <ul style="list-style-type: none">• semplice, in quanto le applicazioni devono essere facilmente fruibili in applicazioni differenti e da parte di differenti utenti | <ul style="list-style-type: none">• efficiente, perché l'utilizzo delle risorse deve essere ottimizzato:<ul style="list-style-type: none">– “in tempo” (efficiente utilizzo del microprocessore)– “in spazio” (efficiente utilizzo della memoria) | <ul style="list-style-type: none">• efficace, nel senso che le informazioni devono essere rappresentative della realtà che si vuole analizzare (ad esempio, un'età non può essere negativa) | <ul style="list-style-type: none">• sicuro, in quanto le operazioni sui dati sono permesse solo a soggetti identificabili e autorizzati |
|---|---|--|--|

Che cos'è una base di dati

Una **base di dati** (in inglese **database**) può essere considerata come una raccolta di dati logicamente correlati, utilizzata per modellare una realtà.

I dati sono memorizzati su un supporto di memoria di massa e sono progettati per essere fruiti in maniera ottimizzata da differenti applicazioni e utenti diversi.



Che cos'è una base di dati

Una base di dati, per poter essere definita tale, deve essere:

- **sicura**: non si deve danneggiare a causa di eventi accidentali e interventi non autorizzati;
- **integra**: le operazioni effettuate da utenti autorizzati non devono provocare una perdita di consistenza dei dati;
- **consistente**: i dati devono essere significativi e utilizzabili nelle applicazioni per cui la base di dati è stata progettata;
- **condivisibile**: applicazioni e utenti diversi devono poter accedere ai dati comuni;
- **persistente**: deve avere un tempo di vita non limitato a quello delle singole esecuzioni dei programmi che la utilizzano;
- **scalabile**: deve mantenere intatte le proprie performance all'aumentare della quantità di dati e deve essere in grado di scalare con l'aggiunta di nuove funzionalità.

Il sistema informativo

Le informazioni necessarie a un'organizzazione sono gestite da un sistema informativo.

Un **sistema informativo** è un insieme organizzato di strumenti, procedure manuali, norme organizzative, risorse umane e materiali, orientato alla gestione delle informazioni rilevanti per un'organizzazione.

La gestione è intesa come raccolta, archiviazione, elaborazione e scambio di informazioni necessarie alle attività operative, di gestione, di preparazione, controllo e valutazione dell'organizzazione.

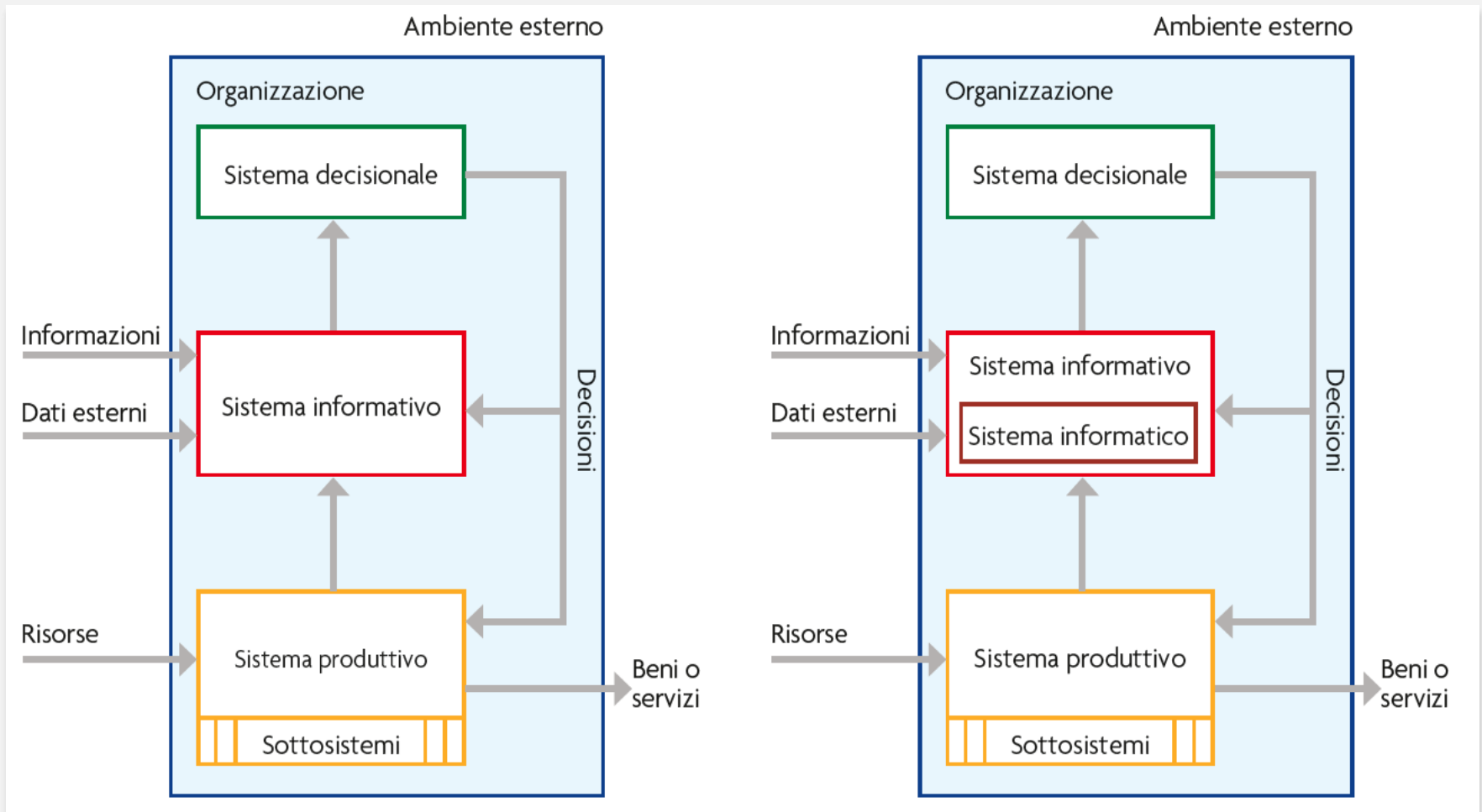
Il sistema informatico

La parte del sistema informativo di un'organizzazione che può essere automatizzata è chiamata sistema informatico.

Il sistema informatico è l'insieme degli strumenti informatici utilizzati per il trattamento automatico delle informazioni, rappresentate mediante dati digitali, al fine di agevolare le funzioni del suo sistema informativo.

Il sistema informatico è un sistema informativo che si avvale di tecnologie informatiche.

Il sistema informatico



Dati e informazioni

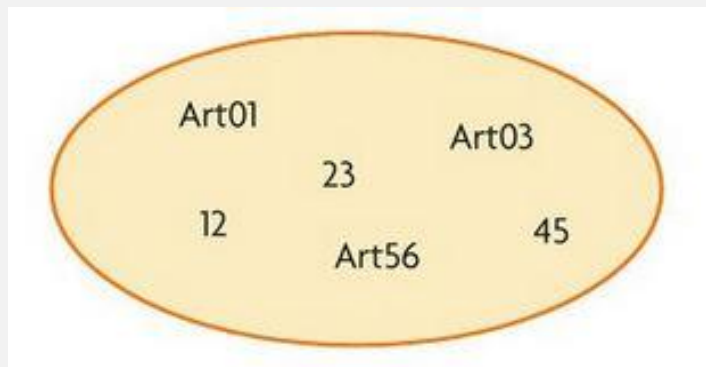
In informatica un **dato** è la descrizione elementare di una cosa, di un avvenimento.

Lo scopo dei dati è quello di codificare in vari modi i fatti (anche ipotetici) ritenuti importanti nell'ambito di un'organizzazione.

L'**informazione** è l'incremento di conoscenza che può essere acquisita (inferita) dai dati.

I dati sono utilizzabili (danno informazioni) se si attribuisce loro una **chiave di interpretazione** che consente di comprendere il loro **significato** (semantica), cioè i fatti che essi codificano.

Dati e informazioni



ARTICOLI IN MAGAZZINO	
Codice articolo	Quantità
Art01	23
Art03	12
Art56	45

Schemi e istanze

Chiameremo **schema** o **intensione** la chiave di interpretazione dei dati, ovvero il significato (la semantica) che si attribuisce al dato per ricavare l'informazione da esso completata.

Chiameremo **istanza di uno schema** o **estensione** l'insieme dei valori assunti da uno schema in un certo istante di tempo.

Il **significato intensionale** (o schema) è il contenuto informativo dei dati: la ricostruzione dei fatti (o significato completo dei dati) nasce dunque dall'interpretazione dei dati.

Il **significato estensionale** (o istanza) dei dati si riferisce ai valori che può assumere uno schema in un certo istante.

Schemi e istanze

Chiamiamo **categoria** un gruppo di dati aventi la stessa chiave di interpretazione, ovvero lo stesso schema.

Una **occorrenza** di una base di dati è l'insieme delle istanze delle categorie in un determinato istante di tempo.

The diagram illustrates a table structure with annotations. The table has a blue header row and three data rows. The header row is labeled 'ARTICOLI IN MAGAZZINO' and is identified as the 'Categoria'. The first two columns are 'Codice articolo' and 'Quantità', which together form the 'Schema'. The 'Quantità' column is specifically labeled as an 'Attributo'. The second data row, containing 'Art03' and '12', is highlighted with a red dashed border and labeled as an 'Istanza'. The value '45' in the third data row is labeled as a 'Valore'.

ARTICOLI IN MAGAZZINO	
Codice articolo	Quantità
Art01	23
Art03	12
Art56	45

Il modello dei dati

La più diffusa classificazione delle basi di dati si basa sul modello di dati che essa supporta.

Un **modello di dati** è un insieme di concetti e di costrutti utilizzati per organizzare i dati di interesse e descriverne la struttura e la dinamica (associazioni e vincoli che devono rispettare).

Nella teoria delle basi di dati, i modelli si distinguono in:

- **modelli concettuali**, che permettono di rappresentare i concetti (dati) in modo indipendente da ogni sistema, cercando di descrivere i concetti del mondo reale (come il **modello ER**).
- **modelli logici**, che consentono una specifica rappresentazione dei dati (tabelle, alberi, grafi, etc).

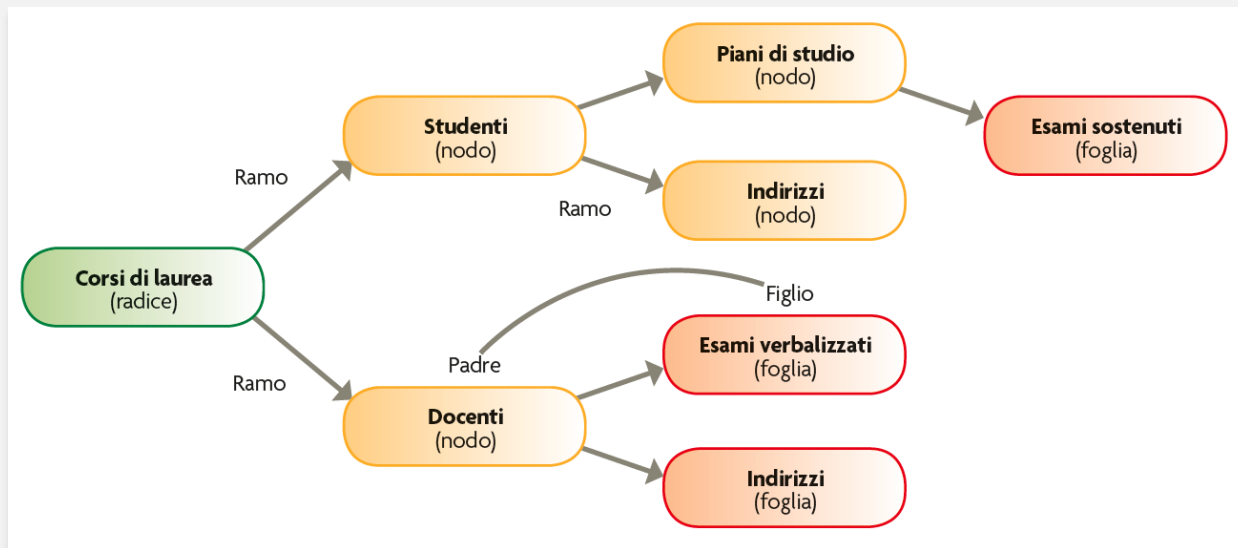
I modelli logici

Modello gerarchico

I dati sono organizzati in record connessi tra loro secondo strutture ad albero. Ogni record del database deve avere un unico padre.

Possono esserci più record, su alberi diversi, che rappresentino la stessa informazione, anche se ciò può creare vari problemi come la ridondanza dei dati e continui controlli di consistenza.

Approccio di tipo ***procedurale***.

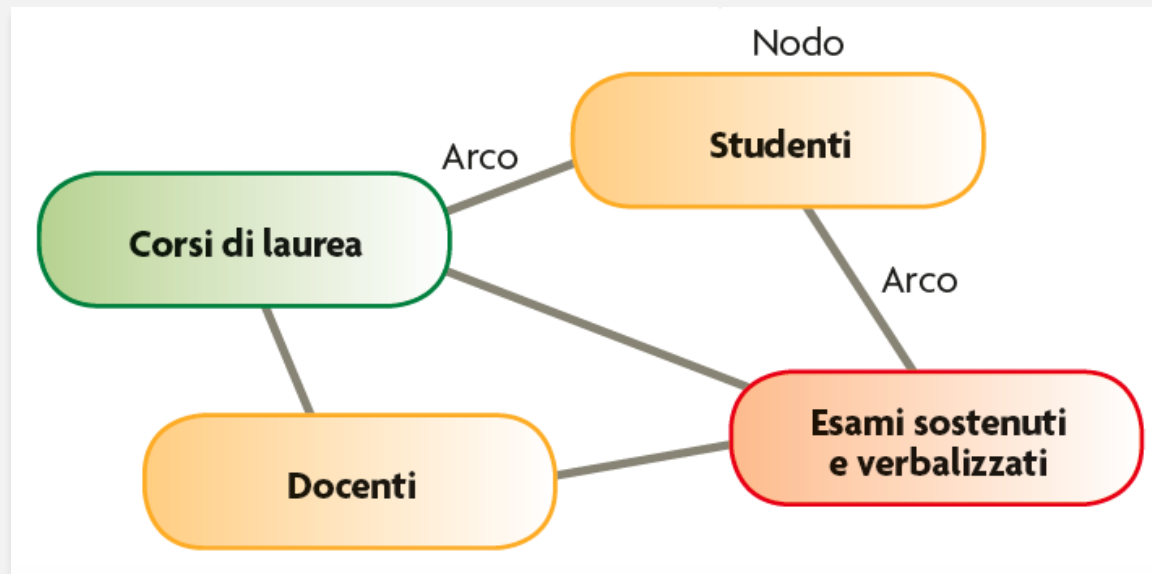


I modelli logici

Modello reticolare (modello *CodasyI*, 1973)

I record sono legati tra loro con strutture ad anello che permettono all'utente finale di accedere ai dati in maniera più semplice, senza i vincoli rigidi della struttura gerarchica. Ogni nodo può essere un punto di partenza per raggiungere un campo.

Approccio di tipo *procedurale*.



I modelli logici

Modello relazionale (*Codd, 1970*)

È basato sul concetto di insiemi di record (tabelle) e le relazioni tra le informazioni derivano dalla corrispondenza di alcuni campi di record appartenenti a tabelle diverse.

Approccio di tipo ***dichiarativo***.

Modello a oggetti

Nasce come evoluzione del modello relazionale, estendendo alle basi di dati il concetto di programmazione a oggetti.

Lo sviluppatore del database gestisce tutti gli aspetti del database stesso, incluse le operazioni che manipolano i dati dall'interno del software di programmazione del database object oriented.

La progettazione di una base di dati

La **metodologia di progettazione** di una base di dati è un insieme di attività tra loro collegate, prodotti intermedi e finali di tali attività, criteri di verifica di qualità di tali fasi e prodotti, volti a realizzare una base di dati a partire da un insieme di specifiche che formalizzano le esigenze dell'utente.

Progettare una base di dati significa progettare la **struttura dei dati** e le **applicazioni**.

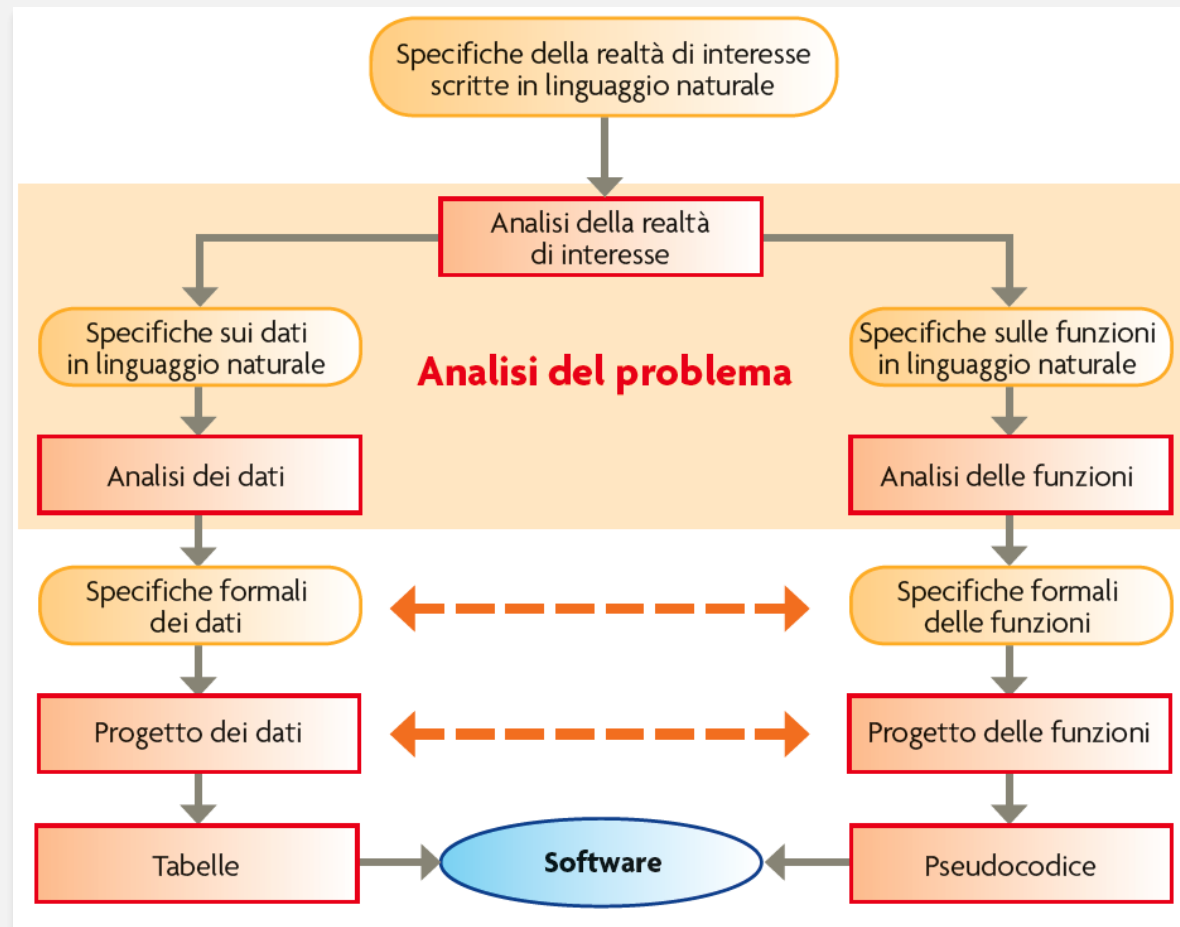
La progettazione della struttura dei dati è l'attività fondamentale, in quanto richiede di specificare un modello della realtà di interesse (**universo del discorso**) quanto più possibile fedele.

Le specifiche della realtà di interesse vengono trasformate in **specifiche formali** dopo una prima fase di analisi.

Fase di analisi: determina cosa il programma deve fare.

Fase di progettazione: come il documento di specifiche formali viene trasformato in programma.

La progettazione di una base di dati



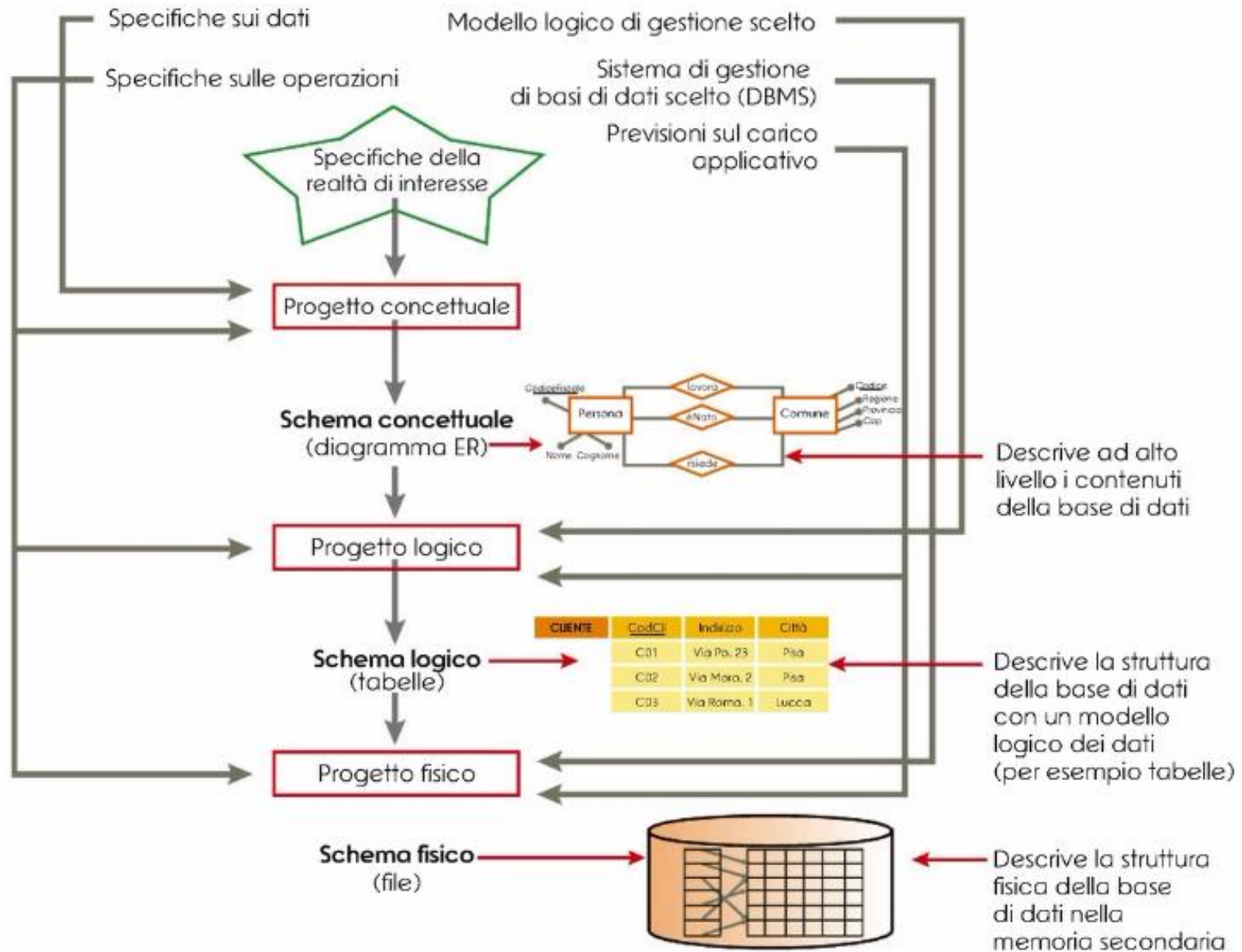
La progettazione di una base di dati

Al termine della fase di analisi delle funzioni è indispensabile verificare che:

- per ogni funzione definita nello schema delle funzioni siano rappresentate tutte le tipologie di dati da essa manipolate;
- per ogni aspetto riguardante i dati, all'interno delle specifiche formali relative, siano definite tutte le funzioni che su quell'aspetto operano.

PROGETTAZIONE		
Progettazione concettuale	Progettazione logica	Progettazione fisica
<ul style="list-style-type: none">• Il suo scopo è costruire e definire una rappresentazione corretta e completa della realtà di interesse.• L'input di questa fase è il documento delle specifiche formali.• L'output è uno schema concettuale, cioè una rappresentazione astratta e il più possibile formale della realtà; un esempio di output, riferendoci solo ai dati (e non alle funzioni), è il diagramma ER che vedremo nella prossima unità.	<ul style="list-style-type: none">• Il suo scopo è quello di trasformare questa rappresentazione ancora astratta in uno schema logico, ovvero in una rappresentazione efficiente rispetto alle strutture del sistema di gestione che si intende utilizzare; un esempio è una descrizione relazionale tramite tabelle del diagramma ER.• L'input di questa fase è il diagramma ER della fase di progettazione concettuale.• L'output di questa fase è uno schema logico riassumibile con relazioni rappresentate tramite tabelle.	<ul style="list-style-type: none">• Il suo scopo è quello di implementare lo schema logico definendo tutti gli aspetti fisici di memorizzazione e rappresentazione in memoria di massa.• L'input di questa fase sono le tabelle della fase di progettazione logica.• L'output di questa fase è l'implementazione in memoria di massa di tali tabelle.

La progettazione di una base di dati



Pre – Database

Problemi di:

1. **Ridondanza** dei dati:

- Costo addizionale legato alla generazione e memorizzazione dei singoli dati;
- Complessità delle operazioni necessarie per aggiornare tutti gli archivi;
- Impossibilità di avere tutti gli archivi aggiornati allo stesso livello.

2. **Incongruenza** dei dati;

3. **Inconsistenza** dei dati;

4. **Concorrenza** di software.

II DBMS

Un **DBMS** (*Data Base Management System*) è un insieme di strumenti software che, sulla base delle specifiche dell'utente, è in grado di gestire dati strutturati che sono tanti, importanti, condivisi, sia interrogati che aggiornati.

Provvede inoltre alla generazione dello schema, alla ricerca e all'aggiornamento dei dati.

Nel caso specifico dei database relazionali si parla di **RDBMS** (*Relational Data Base Management System*).

Il DBMS rappresenta un'interfaccia tra gli sviluppatori, gli utenti del database e il sistema di elaborazione.

II DBMS: le caratteristiche

Le caratteristiche di un DBMS sono:

- gestione della base di dati;
- persistenza e consistenza dei dati;
- privacy e sicurezza dei dati;
- integrità dei dati;
- supporto alle transazioni (una **transazione** è una sequenza di operazioni effettuate su una base di dati, che può concludersi con un successo o un insuccesso);
- gestione del dizionario dei dati (il dizionario contiene i **metadati**, ossia le informazioni che descrivono gli oggetti della base di dati).

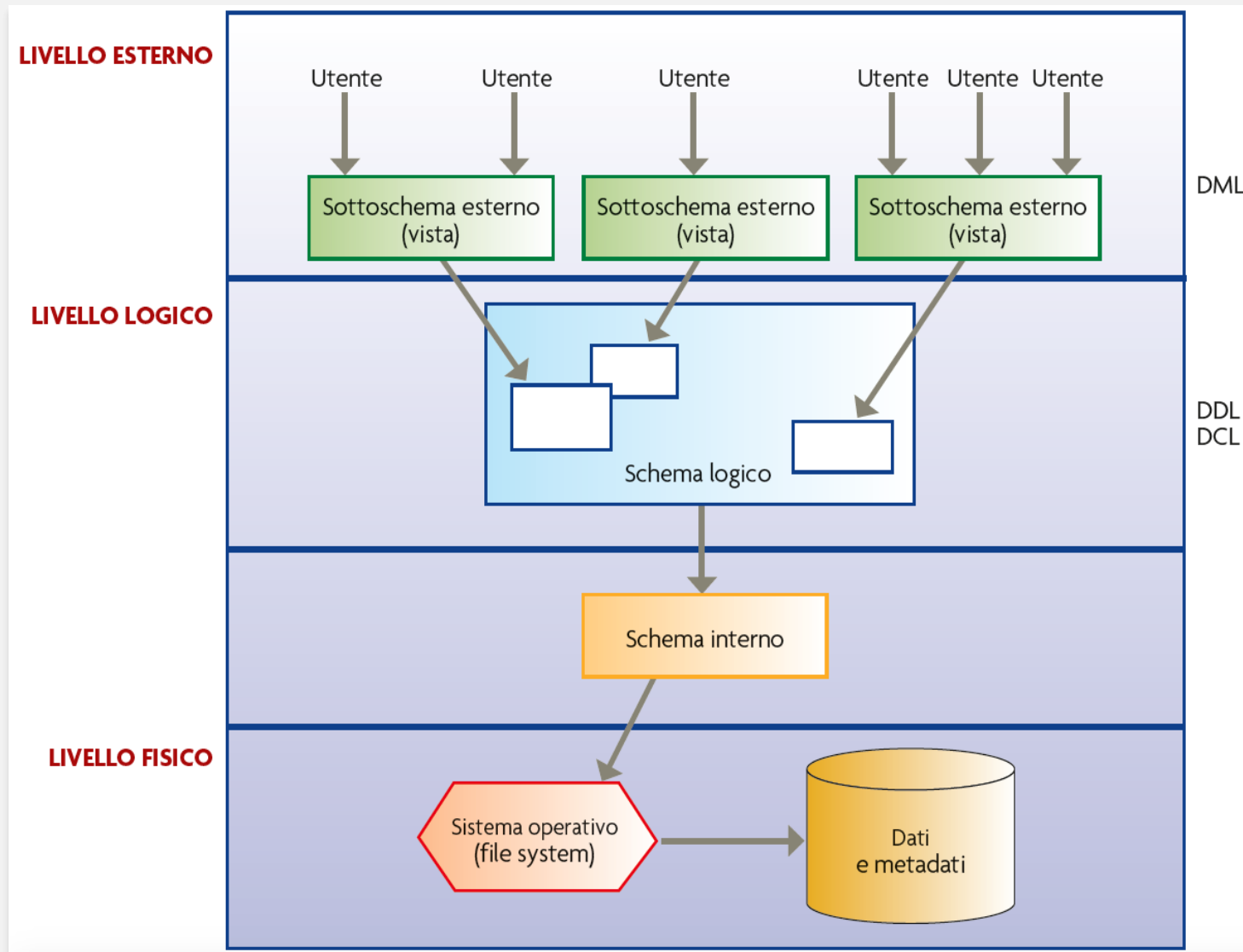
Livelli di astrazione di un DBMS

Il DBMS ha, tra le sue funzioni, quella di mantenere un **modello astratto dei dati** che consenta all'utente di concepire le informazioni secondo strutture riferite direttamente alle entità reali che i dati sono preposti a rappresentare.

Gli utenti di un DBMS vedono una macchina astratta (**macchina DBMS**) che consente loro di compiere specifiche operazioni per descrivere e manipolare i dati contenuti nel database, utilizzando appositi linguaggi.

I linguaggi di interrogazione del database mediante *query* (interrogazioni) e i generatori di *report* permettono agli utenti di interrogare in maniera interattiva il database e di analizzarne i dati.

Livelli di astrazione di un DBMS



Livelli di astrazione di un DBMS

Il **livello esterno** si riferisce al modo in cui ciascun utente può vedere gli stessi dati, che mantengono l'organizzazione fisica e concettuale precedentemente descritte, ma vengono messi a disposizione secondo il formato richiesto. Il linguaggio DML (Data Manipulation Language) permette di interrogare il database e di inserire, cancellare e modificare dati.

Il **livello concettuale o logico** riguarda la struttura logica assunta dai dati registrati, quindi il loro schema astratto. Il linguaggio DDL (Data Definition Language) descrive categorie di dati. Il linguaggio DCL (Data Control Language) permette il controllo degli accessi degli utenti.

Il **livello fisico** riguarda l'effettiva memorizzazione dei dati, organizzati in file, record e strutture di accesso. Il linguaggio DMCL (Device Media Control Language) organizza i dati.

Livelli di astrazione di un DBMS

Riassumendo:

<ul style="list-style-type: none">• il livello fisico riguarda l'effettiva memorizzazione dei dati, organizzati in file, record e strutture di accesso;	<ul style="list-style-type: none">• il livello concettuale riguarda la struttura logica assunta dai dati registrati, quindi il loro schema astratto;	<ul style="list-style-type: none">• il livello esterno si riferisce al modo in cui ciascun utente può vedere gli stessi dati, che mantengono l'organizzazione fisica e concettuale precedentemente descritte, ma vengono messi a disposizione secondo il formato richiesto.
--	---	--

Con l'organizzazione a tre livelli, un DBMS permette di ottenere **indipendenza fisica e logica** dei dati.

Indipendenza fisica = possibilità di modificare l'organizzazione fisica dei dati senza modificare lo schema logico e i programmi applicativi che agiscono sul database.

Indipendenza logica = possibilità di modificare lo schema logico del database senza dover modificare i programmi applicativi non interessati alla modifica.