



Analisi e gestione delle banche dati

- Introduzione alle basi di dati

Barbara Re

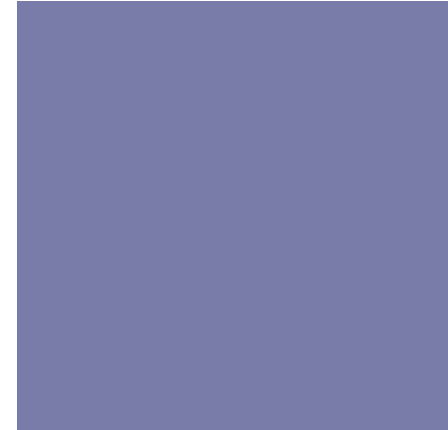
+ Agenda

- Sistemi informativi, informazioni e dati
- Basi di dati e sistemi di gestione delle basi di dati
- Modelli di dati
- Linguaggi ed utenti delle basi di dati
- Vantaggi e svantaggi dei DBMS





Sistemi informativi, informazioni e dati



+ Sistema

- Un sistema è rappresentabile da un'organizzazione che svolge una **funzione** (e ha un obiettivo), fornendo opportune **uscite** (rilevabili dall'esterno) in corrispondenza a determinati **ingressi** (condizioni esterne che influiscono su di esso) presenti o passati
- Esempi
 - naturali (organismi viventi, ecosistemi)
 - artificiali (macchine, fabbriche, circuiti)
 - sociali (aziende, enti pubblici, comunità locali o nazionali)

+ Sistemi e sottosistemi

- Ogni sistema ha di solito una **struttura**, cioè è articolato in **sottosistemi**, che interagiscono fra loro e con il mondo esterno
- Ogni sottosistema ha una funzione specifica (e obiettivi propri)
- Gli ingressi e le uscite dei sottosistemi possono essere ingressi e uscite del sistema complessivo oppure semplicemente flussi interni
- In un sistema vengono svolti processi, attività rivolte verso l'esterno o verso l'interno, nell'ambito di sottosistemi o trasversalmente

+ Sistema azienda (o amministrazione pubblica)

- obiettivi, i fini istituzionali
- struttura, l'insieme delle parti e sottosistemi
- processi, le attività

+ Risorse e Processi

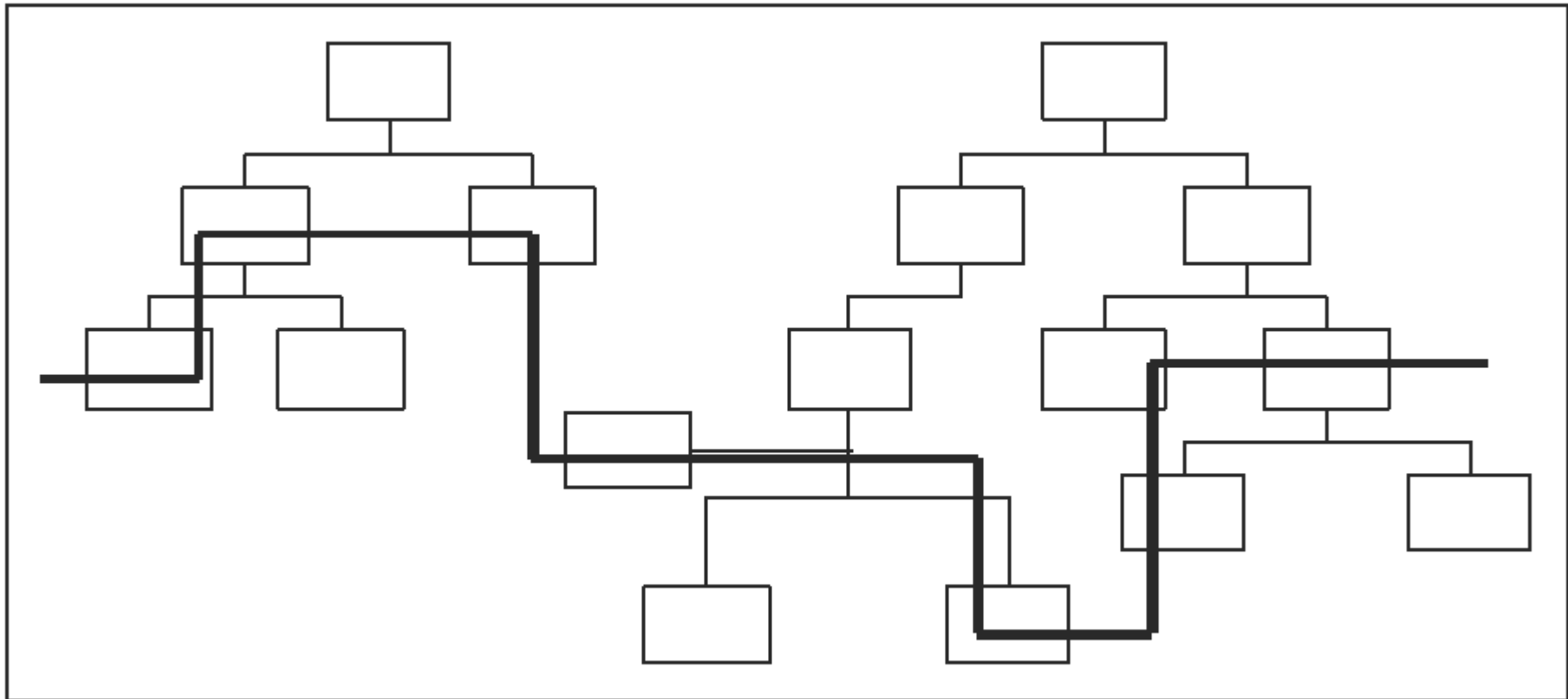
■ Risorsa

- E' tutto ciò con cui l'organizzazione opera, sia materiale che immateriale, per perseguire i suoi obiettivi
 - persone
 - denaro
 - materiali
 - **informazioni**

■ Processo

- È l'insieme delle attività (sequenze di decisioni e azioni) che l'organizzazione nel suo complesso svolge per gestire il ciclo di vita di una risorsa o di un gruppo omogeneo di risorse (vedremo meglio fra poco)

+ Processi e strutture organizzative



+ Gestione delle informazioni

- 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

+ Gestione delle informazioni

- 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
 - nome e cognome
 - estremi anagrafici
 - codice fiscale

+ Informazioni e dati

- Nei sistemi informativi ed informatici (ma non solo), le **informazioni** vengono rappresentate in modo essenziale, spartano: attraverso i **dati**

SISTEMA INFORMATIVO <-> SISTEMA INFORMATICO

+ Informazioni e dati

(definizioni dal Vocabolario della lingua italiana 1987)

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 presente alla conoscenza, prima di ogni elaborazione; (in informatica) elementi di informazione costituiti da simboli che debbono essere elaborati

+ Dati e informazioni

- Un esempio:

Mario 275

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 il problema X; qual è il suo numero di telefono?”, allora i dati possono essere interpretati per fornire **informazione** e arricchire la conoscenza

+ Perché i dati?

- La rappresentazione precisa di forme più ricche di informazione e conoscenza è difficile
- I dati costituiscono spesso una risorsa strategica, perché più stabili nel tempo di altre componenti (processi, tecnologie, ruoli umani) ne sono un esempio i dati delle banche o delle anagrafi

+ Dati Informazione Conoscenza

■ DATO

- è un rozzo numero o fatto con nessun significato per il ricevente

■ INFORMAZIONE

- è il dato contestualizzato che ha significato per il ricevente

■ CONOSCENZA

- è l'informazione che è stata autenticata e assunta essere vera accumulata ed integrata dal ricevente



+ Dati Informazione Conoscenza: esempio

■ DATO

- 10

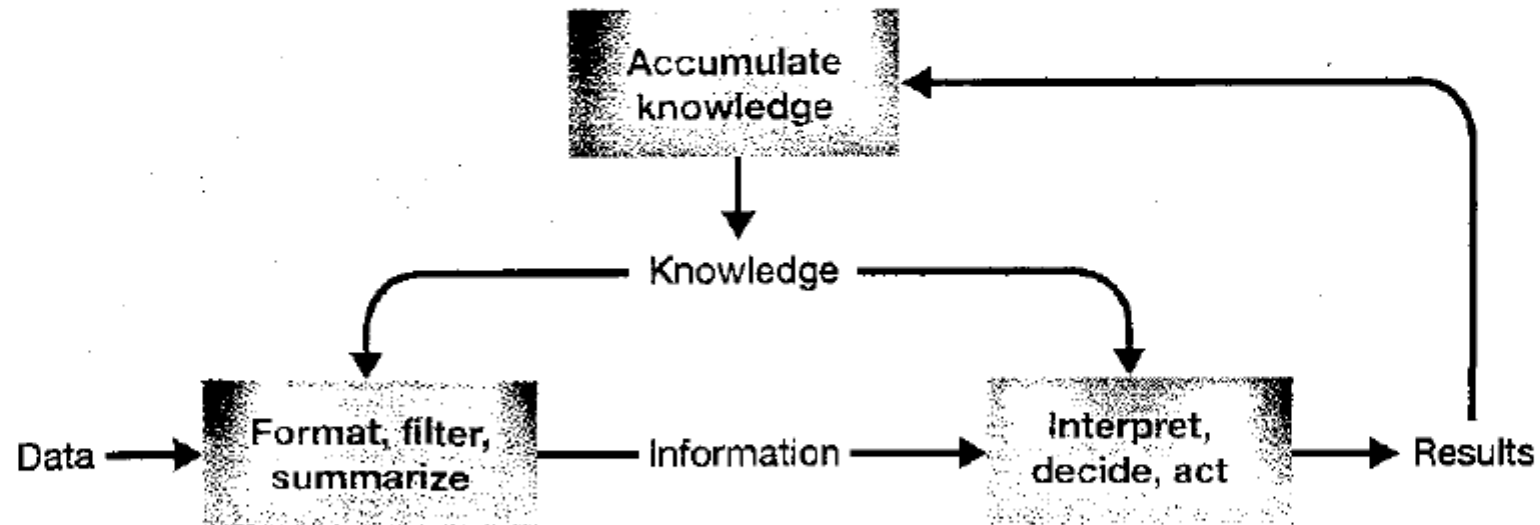
■ INFORMAZIONE

- 10 gradi a Roma il 21 Agosto

■ CONOSCENZA

- Con 10 gradi a Roma il 21 Agosto probabilmente c'è qualcosa che non va o il termostato è rotto o qualche serio problema atmosferico va affrontato

+ Relazione tra dati – informazioni e conoscenza



+ Sistemi informativi e sistemi informatici: una premessa

- Che cos'è l'informatica?
 - Scienza del trattamento razionale, specialmente per mezzo di macchine automatiche, dell'informazione, considerata come supporto alla conoscenza umana e alla comunicazione (Académie Française)
- L'informatica ha due anime
 - **tecnologica**: i calcolatori elettronici e i sistemi che li utilizzano;
 - **metodologica**: i metodi per la soluzione di problemi e la gestione delle informazioni.

+ Sistema informativo

- Componente (sottosistema) di una organizzazione che gestisce (acquisisce, elabora, conserva, produce) le informazioni di interesse (cioè utilizzate per il perseguimento degli scopi dell'organizzazione stessa)
- Gestisce la risorsa "informazione" (e le informazioni sulle altre risorse)

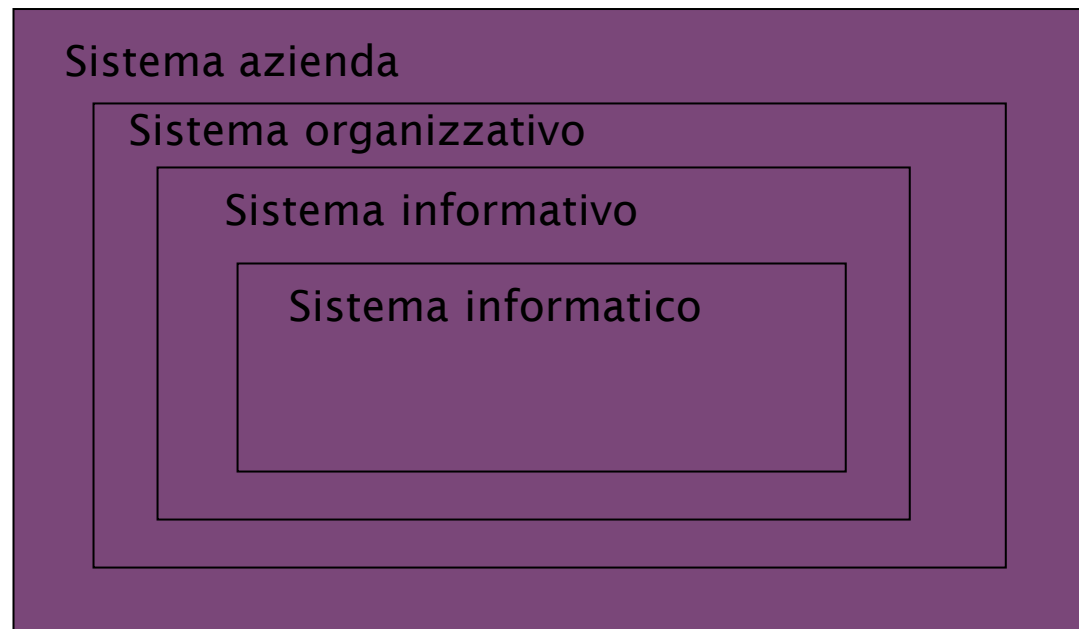
+ Funzioni di un sistema informativo

- Gestione delle informazioni
 - raccolta, acquisizione delle informazioni
 - archiviazione, conservazione delle informazioni
 - elaborazione, trasformazione, correlazione delle informazioni
 - distribuzione, scambio di informazioni
- Il ciclo dell'informazione
 - l'informazione deve soprattutto "fluire" non essere "rigenerata" ogni volta

+ Sistemi informativi e automazione

- Il concetto di “sistema informativo” è indipendente da qualsiasi automatizzazione: esistono organizzazioni la cui ragion d'essere è la gestione di informazioni (p. es. servizi anagrafici e banche) e che operano da secoli
- La porzione di sistema informativo che viene gestita in modo automatico mediante tecnologie informatiche prende il nome di **sistema informatico**

+ Sistema Informatico ed informativo visione d'insieme



+ Sistema informatico

- I sistemi informatici hanno il compito di raccogliere, organizzare e conservare le informazioni. Essi garantiscono che tali informazioni vengano conservate in modo permanente su dispositivi per la loro memorizzazione, aggiornate per riflettere rapidamente le loro variazioni e rese accessibili alle richieste degli utenti
- Idealmente il sistema informatico dovrebbe coincidere con quello informativo
- Purtroppo però i processi informativi gestiti in modo automatico costituiscono solo una porzione di gran parte dei sistemi informativi
- Dal momento che un sistema informatico è di supporto ai processi di manipolazione automatica dell'informazione è opportuno chiedersi che cosa vuol dire "informazione" e come gestirla

+ Finalità dell'automazione

- Anche prima di essere automatizzati, molti sistemi informativi si sono evoluti verso una razionalizzazione e standardizzazione delle procedure e dell'organizzazione delle informazioni
- Incremento di produttività (come rapporto fra prodotti e costi, in termini monetari e temporali) soprattutto per i processi informativi
- Più esplicitamente:
 - automazione di attività manuali (con economizzazione di forza lavoro)
 - riduzione di funzioni parassite, quali acquisizione ripetuta di informazione e copiatura di documenti (con economizzazione e riduzione di errori)
- Ma non solo ...

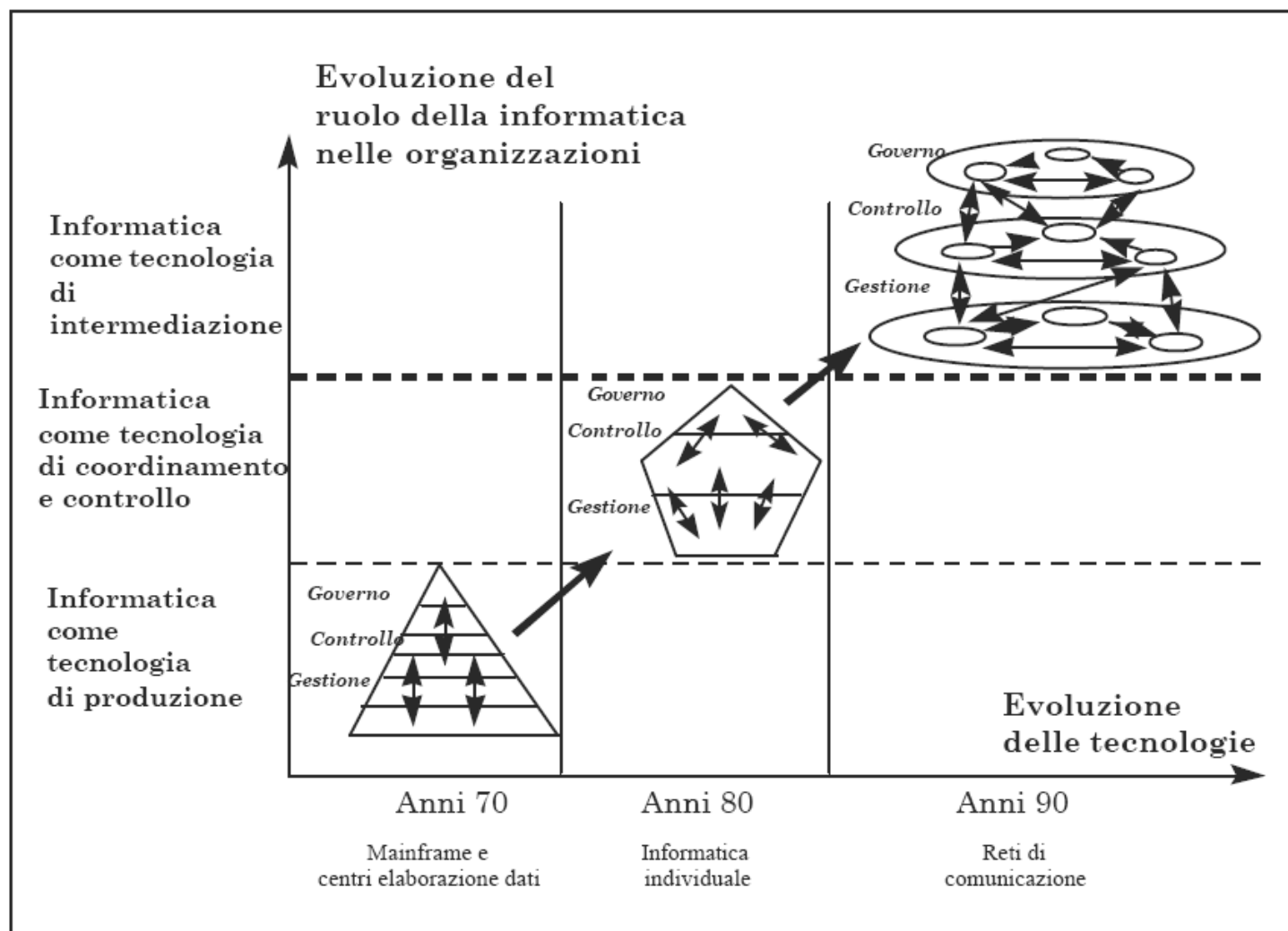
+ L'obiettivo tradizionale della tecnologia

- automatizzare le attività banali, per liberare risorse ai fini dello svolgimento di attività più qualificate
- ma anche:
 - inventare nuovi servizi, riorganizzare le attività

+ (Sotto-)sistemi informatici: una classificazione

- sistemi settoriali (per i processi operativi):
 - *“transaction processing systems”*
- sistemi integrati (anche per i processi gestionali):
 - *“management information systems”*
- sistemi di supporto ai processi direzionali:
 - *“decision support systems”*

+ Evoluzione dei sistemi informatici

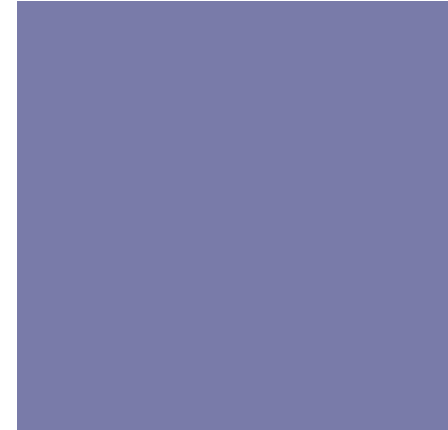


+ Una riflessione importante

- Un sistema “fortemente integrato” o un “mosaico” di componenti?
- I sistemi aperti: metodologie e tecnologie per l'integrazione morbida
- Le reti: l'infrastruttura fondamentale per l'integrazione



Basi di dati e sistemi di gestione delle basi di dati



+ Base di dati

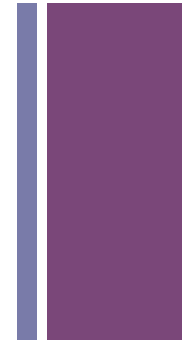
(accezione generica, metodologica)

- Insieme organizzato di dati utilizzati per il supporto allo svolgimento delle attività di un organizzazione (azienda, ufficio, persona)

(accezione specifica, metodologica e tecnologica)

- insieme di dati gestito da un DBMS

+ DB e DBMS



- DB (Database = Base di Dati): collezione di dati logicamente correlati di interesse per il Sistema Informativo
- DBMS (Database Management System = Sistema di gestione della Base di Dati): componente software che interagisce con la Base di Dati e con i programmi applicativi degli utenti

+ Sistema di gestione di basi di dati DataBase Management System (DBMS)

- Sistema che gestisce collezioni di dati:

- grandi
- persistenti
- condivise

garantendo

- privatezza
- affidabilità
- efficienza
- efficacia

+ Le basi di dati sono ... grandi

- dimensioni (molto) maggiori della memoria centrale dei sistemi di calcolo utilizzati
- il limite deve essere solo quello fisico dei dispositivi

+ Le basi di dati sono ... persistenti

- hanno un tempo di vita indipendente dalle singole esecuzioni dei programmi che le utilizzano

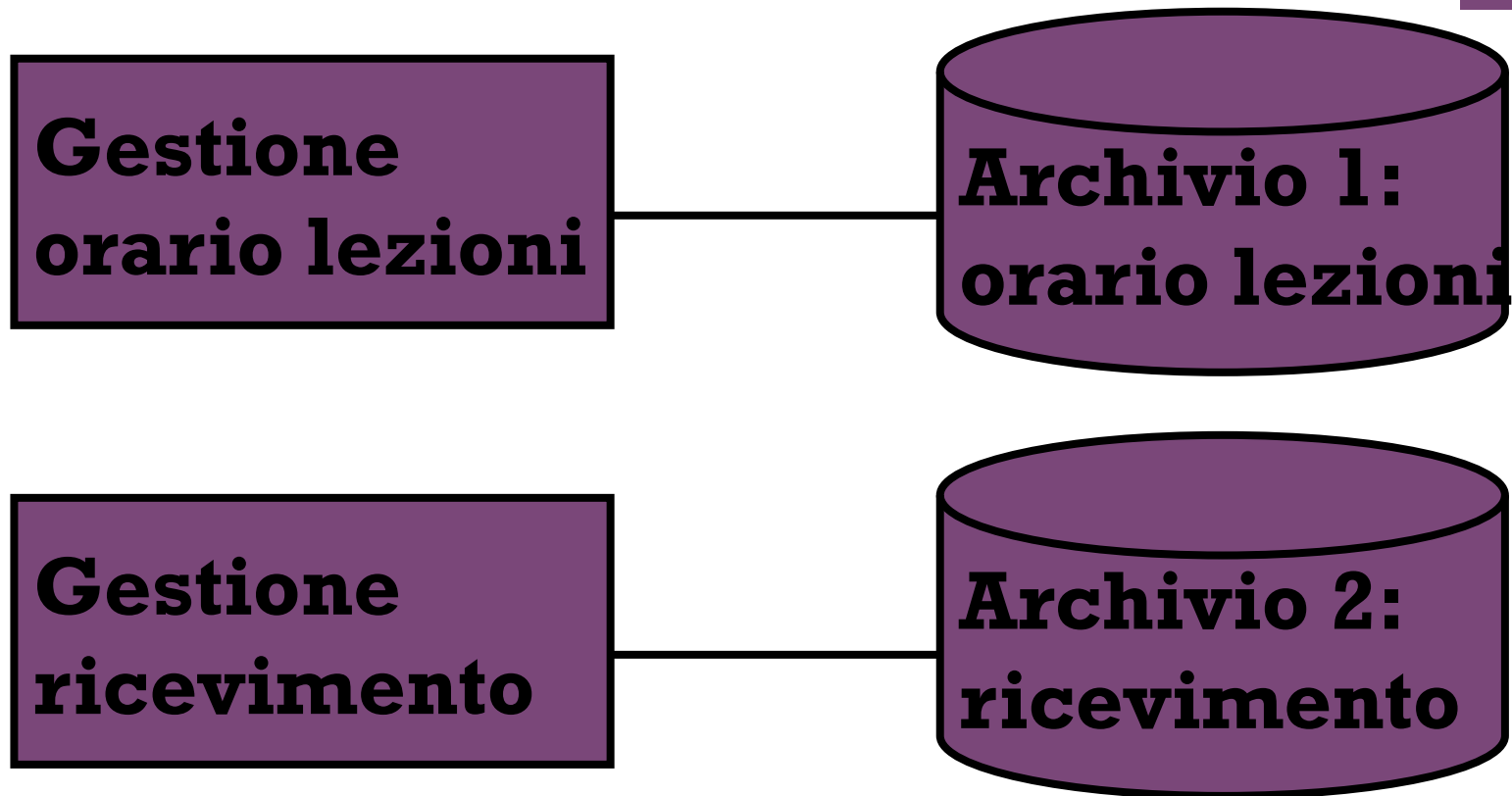
+ Le basi di dati sono ... condivise

- Ogni organizzazione (specie se grande) è divisa in settori o comunque svolge diverse attività
- Ciascun settore/attività ha un (sotto)sistema informativo (non necessariamente disgiunto)

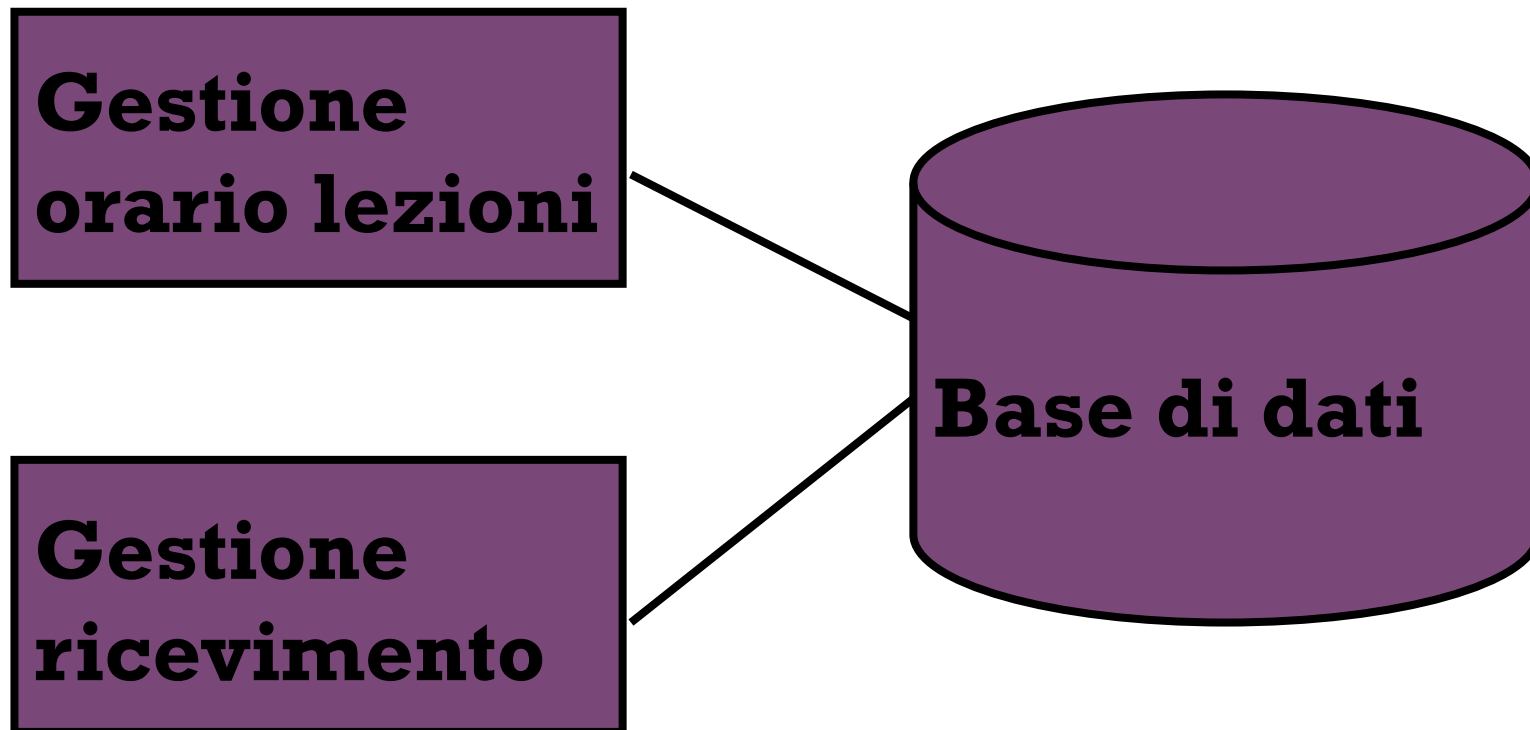
+ Problemi

- Ridondanza:
 - informazioni ripetute
- Rischio di **incoerenza**:
 - le versioni possono non coincidere

+ Archivi e basi di dati



+ Archivi e basi di dati



+ Le basi di dati sono condivise

- Una base di dati e' una risorsa **integrata, condivisa** fra applicazioni
- conseguenze
 - Attività diverse su dati condivisi:
 - meccanismi di **autorizzazione**
 - Accessi di più utenti ai dati condivisi:
 - controllo della **concorrenza**

+ I DBMS garantiscono ... privacy

- Si possono definire meccanismi di autorizzazione
 - l'utente A è autorizzato a leggere tutti i dati e a modificare quelli sul ricevimento
 - l'utente B è autorizzato a leggere i X e a modificare Y

+ I DBMS garantiscono... affidabilità

- Affidabilità (per le basi di dati):
 - resistenza a malfunzionamenti hardware e software
- Una base di dati è una risorsa pregiata e quindi deve essere conservata a lungo termine
- Tecnica fondamentale:
 - gestione delle transazioni

+ Transazione

- Insieme di operazioni da considerare indivisibile ("atomico"), corretto anche in presenza di concorrenza e con effetti definitivi

+ Le transazioni sono ... atomiche

- La sequenza di operazioni sulla base di dati viene eseguita per intero o per niente:
 - trasferimento di fondi da un conto A ad un conto B: o si fanno il prelevamento da A e il versamento su B o nessuno dei due

+ Le transazioni sono ... concorrenti

- L'effetto di transazioni concorrenti deve essere coerente (ad esempio "equivalente" all'esecuzione separata)
 - se due assegni emessi sullo stesso conto corrente vengono incassati contemporaneamente si deve evitare di trascurarne uno

+ I risultati delle transazioni sono permanenti

- La conclusione positiva di una transazione corrisponde ad un impegno (in inglese **commit**) a mantenere traccia del risultato in modo definitivo, anche in presenza di guasti e di esecuzione concorrente

+ I DBMS debbono essere...efficienti

- Cercano di utilizzando al meglio le risorse di spazio di memoria (principale e secondaria) e tempo (di esecuzione e di risposta)
- I DBMS, con tante funzioni, rischiano l'inefficienza e per questo ci sono grandi investimenti e competizione
- L'efficienza è anche il risultato della qualità delle applicazioni

+ I DBMS debbono essere...efficaci

- Cercano di rendere produttive le attività dei loro utilizzatori, offrendo funzionalità articolate, potenti e flessibili:
 - il corso è in buona parte dedicato ad illustrare come i DBMS perseguono l'efficacia

+ DBMS vs file system

- La gestione di insiemi di dati grandi e persistenti è possibile anche attraverso sistemi più semplici — gli ordinari **file system** dei sistemi operativi
- I file system prevedono forme rudimentali di condivisione: "tutto o niente"
- I DBMS estendono le funzionalità dei file system, fornendo più servizi ed in maniera integrata

+ DBMS vs file system (2)

- Nei programmi tradizionali che accedono a file, ogni programma contiene una descrizione della struttura del file stesso, con i conseguenti rischi di incoerenza fra le descrizioni (ripetute in ciascun programma) e i file stessi
- Nei DBMS, esiste una porzione della base di dati (il **catalogo** o **dizionario**) che contiene una descrizione centralizzata dei dati, che può essere utilizzata dai vari programmi

+ Descrizioni dei dati nei DBMS

- Rappresentazioni dei dati a livelli diversi
 - permettono l'indipendenza dei dati dalla rappresentazione fisica:
 - i programmi fanno riferimento alla struttura a livello più alto, e le rappresentazioni sottostanti possono essere modificate senza necessità di modifica dei programmi
- Precisiamo attraverso il concetto di
 - modello dei dati

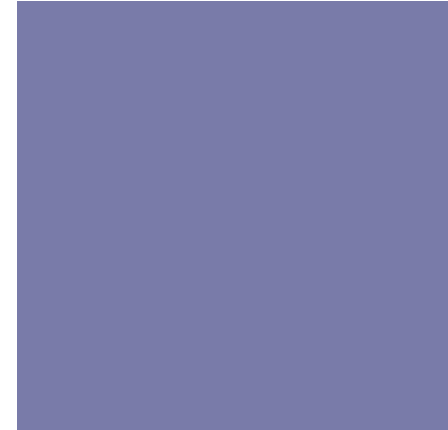
+ DBMS

51

- Prodotti software (complessi) disponibili sul mercato; esempi:
 - Access
 - DB2
 - Oracle
 - Informix
 - Sybase
 - SQLServer



Modelli di dati



+ Modello dei dati

- Insieme di costrutti utilizzati per organizzare i dati di interesse e descriverne la dinamica
- Componente fondamentale: **meccanismi di strutturazione** (o **costruttori di tipo**)
- Come nei linguaggi di programmazione esistono meccanismi che permettono di definire nuovi tipi, così ogni modello dei dati prevede alcuni costruttori
- Esempio: il **modello relazionale** prevede il costruttore **relazione**, che permette di definire insiemi di record omogenei

+ Due tipi (principali) di modelli

- modelli concettuali
- modelli logici

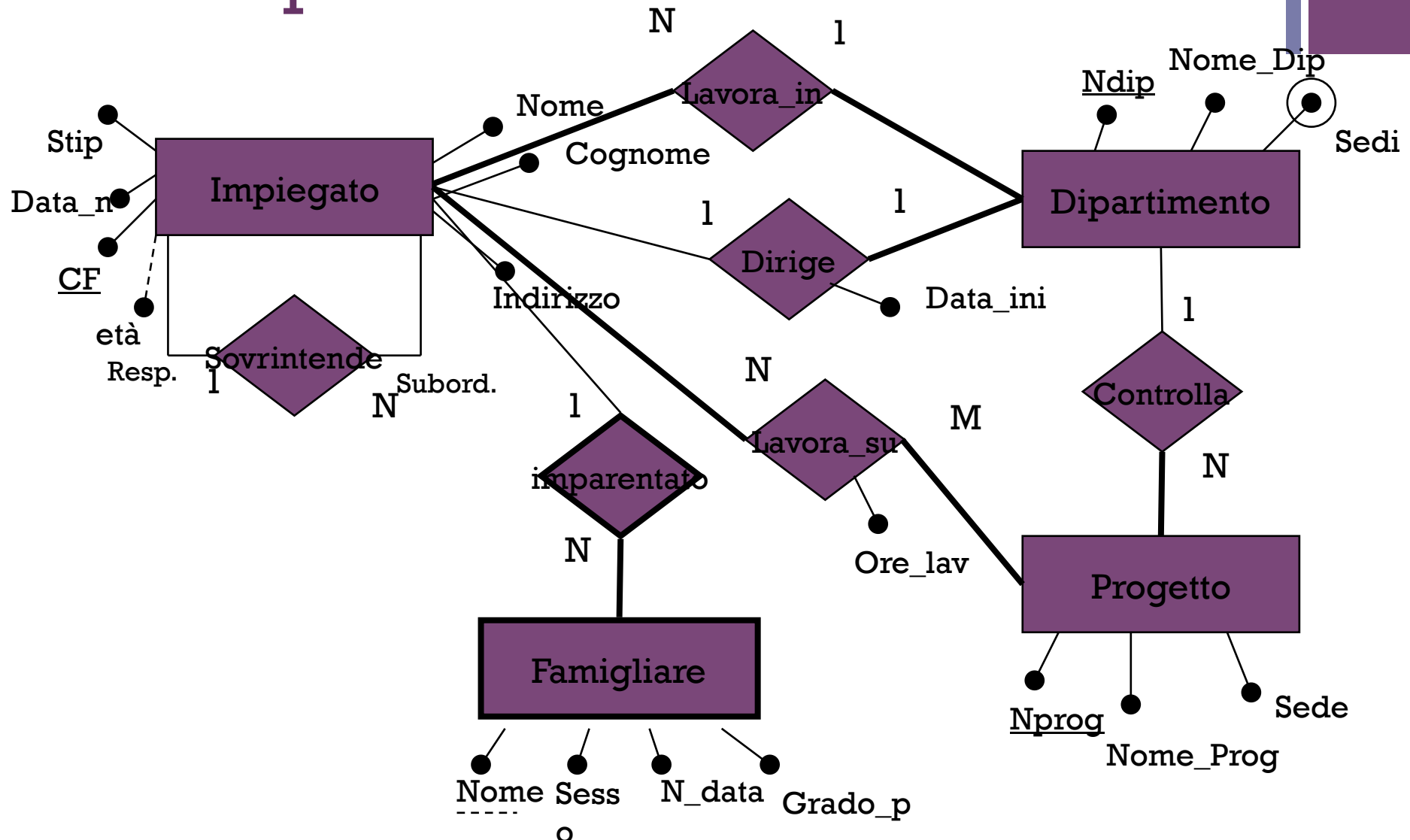
+ Modelli concettuali

- Permettono di rappresentare i dati in modo indipendente da ogni sistema
 - cercano di descrivere i concetti del mondo reale
 - sono utilizzati nelle fasi preliminari di progettazione
- Il più diffuso è il **Modello Entity-Relationship**

+ Modello Entity-Relationship

- Nel contesto della progettazione dei database, il modello entity-relationship (anche detto modello entità-relazione, modello entità-associazione o modello E-R) è un modello per la rappresentazione concettuale dei dati ad un alto livello di astrazione
- Viene spesso utilizzato nella prima fase della progettazione di una base di dati in cui è necessario tradurre le informazioni risultanti dall'analisi di un determinato dominio in uno schema concettuale
- Il modello E-R si basa su un insieme di concetti molto vicini alla realtà di interesse: quindi facilmente intuibili dai progettisti (e in genere considerati sufficientemente comprensibili e significativi anche per i non-tecnici), ma non implementabili sugli elaboratori
- Pur essendo orientato alla progettazione di basi di dati, il modello prescinde dai criteri specifici di organizzazione fisica dei dati persistenti nei sistemi informatici

+ Modello Entity-Relationship: esempio



+ Modelli logici

- Adottati nei DBMS esistenti per l'organizzazione dei dati
 - utilizzati dai programmi
 - indipendenti dalle strutture fisiche
- Alcuni modelli
 - Gerarchico
 - Reticolare
 - Relazionale
- I modelli reticolare e gerarchico rispecchiano (in qualche modo) l'architettura fisica dei dati
- Il relazionale e' basato solo sul valore dei dati

+ Modello Relazionale

- Introdotto da Codd nel 1970
- Implementa bene il concetto di indipendenza dei dati
- Oggi e' di gran lunga il modello dei dati più diffuso
- Implementato nei DBMS commerciali dal 1981
- E' basato sul concetto matematico di **RELAZIONE**

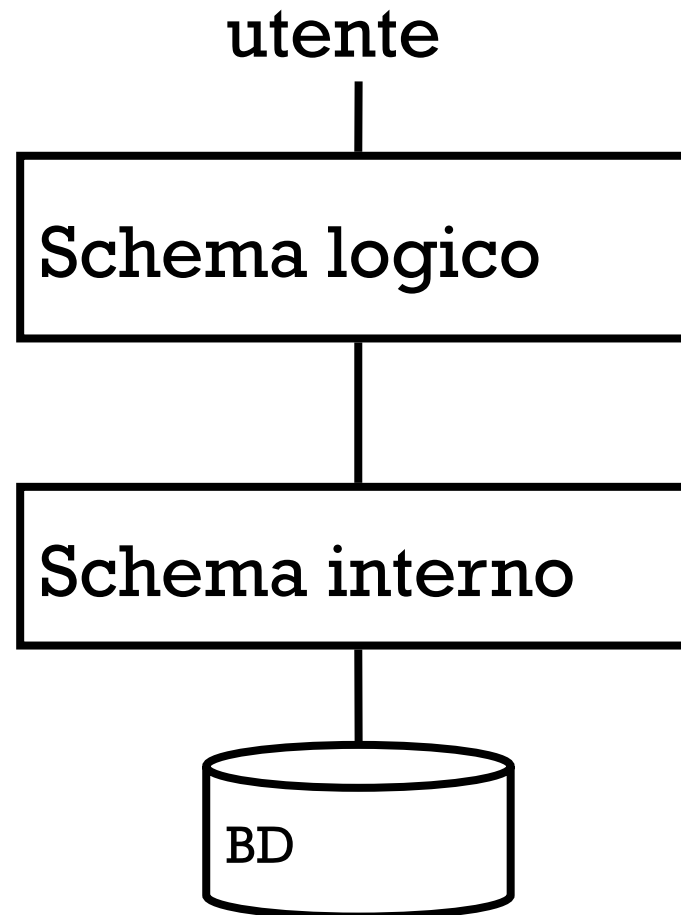


+ Relazioni

NOME	MATRICOLA	INDIRIZZO	TELEFONO
Mario Rossi	123456	Via Etnea 18	777777
Maria Bianchi	234567	Via Roma 2	888888
Giovanni Verdi	345678	Via Catania 3	999999

- Una **Relazione** n-aria è un sottoinsieme del prodotto cartesiano $D_1 \times D_2 \times \dots \times D_n$ di insiemi di valori di tipo elementare detti **domini**.
- Ogni elemento $\langle d_1, d_2, \dots, d_n \rangle$ si chiama n-upla (o tupla)

+ Architettura (semplificata) di un DBMS



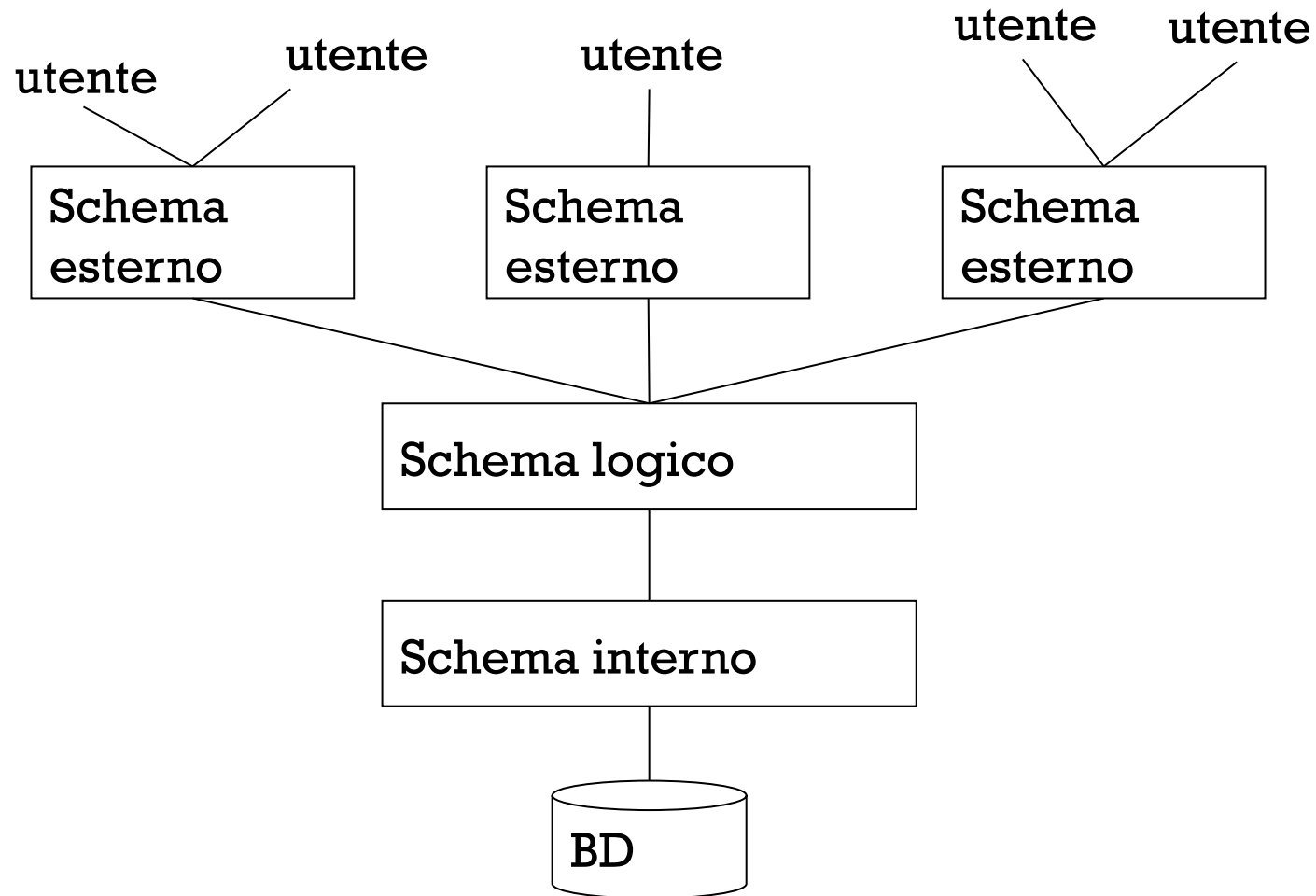
+ Architettura semplificata di un DBMS: schemi

- **schema logico**: descrizione della base di dati nel modello logico (ad esempio, la struttura della tabella)
- **schema interno (o fisico)**: rappresentazione dello schema logico per mezzo di strutture memorizzazione (file; ad esempio, record con puntatori, ordinati in un certo modo)

+ Indipendenza dei dati

- Il livello logico è indipendente da quello fisico:
 - una tabella è utilizzata nello stesso modo qualunque sia la sua realizzazione fisica (che può anche cambiare nel tempo)
- Perciò in questo corso vedremo solo il livello logico e non quello fisico

+ Architettura standard a tre livelli per DBMS



+ Architettura ANSI/SPARC: schemi

Schema logico: descrizione dell'intera base di dati nel modello logico “principale” del DBMS

Schema interno (o fisico): rappresentazione dello schema logico per mezzo di strutture fisiche di memorizzazione

schema esterno: descrizione di parte della base di dati in un modello logico (“viste” parziali, derivate, anche in modelli diversi)

+ Indipendenza dei dati

- conseguenza della articolazione in livelli
- l'accesso avviene solo tramite il livello esterno (che può coincidere con il livello logico)
- due forme:
 - indipendenza fisica
 - indipendenza logica

+ Indipendenza fisica

- il livello logico e quello esterno sono indipendenti da quello fisico
 - una relazione è utilizzata nello stesso modo qualunque sia la sua realizzazione fisica
 - la realizzazione fisica può cambiare senza che debbano essere modificati i programmi

+ Indipendenza logica

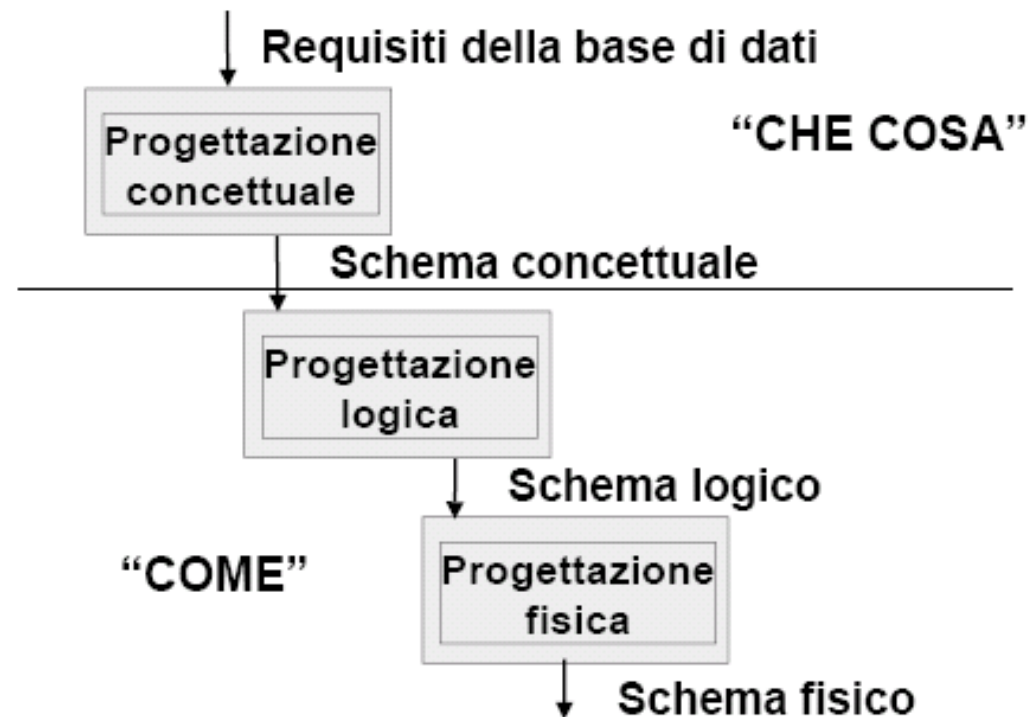
- il livello esterno è indipendente da quello logico
- aggiunte o modifiche alle viste non richiedono modifiche al livello logico
- modifiche allo schema logico che lascino inalterato lo schema esterno sono trasparenti

+ Fasi di progettazione base di dati

- Si basa sul principio semplice ma efficace

di separazione netta tra decisioni relative a

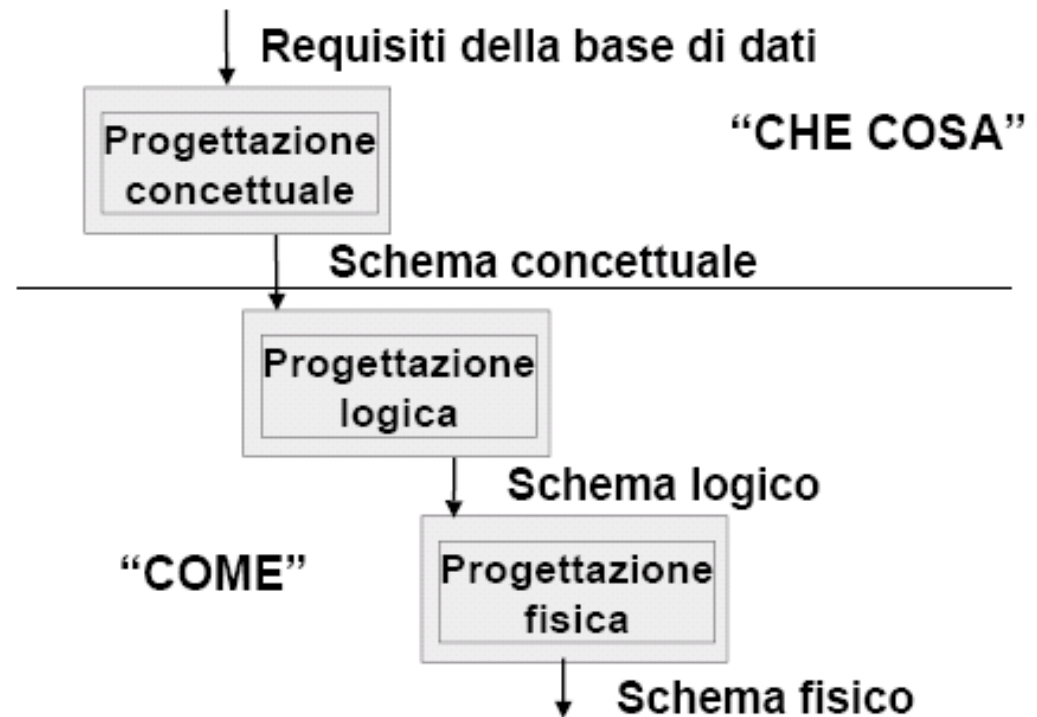
- che cosa rappresentare
- come farlo



+ Output delle fasi di progettazione

- I prodotti della varie fasi sono schemi di alcuni modelli di dati

- Schema concettuale
- Schema logico
- Schema fisico



+ Schemi concettuale - logico - fisico

■ Schema concettuale

- Fornisce una rappresentazione della base di dati di alto livello
- Utile per scopi documentativi

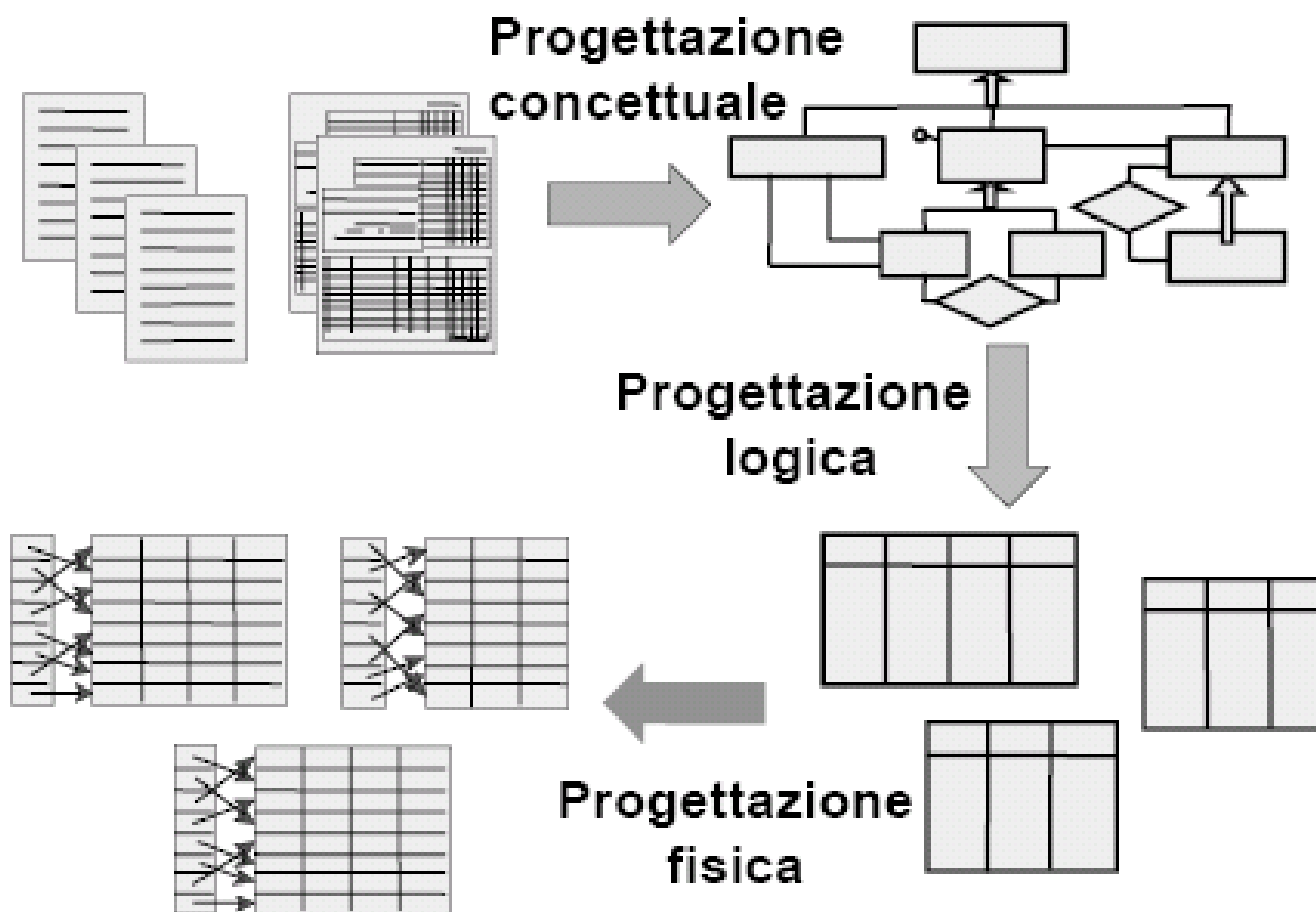
■ Schema logico

- Fornisce una descrizione concreta del contenuto della base di dati
- Prescinde dagli aspetti implementativi
- Utile per le operazioni di interrogazione ed aggiornamento

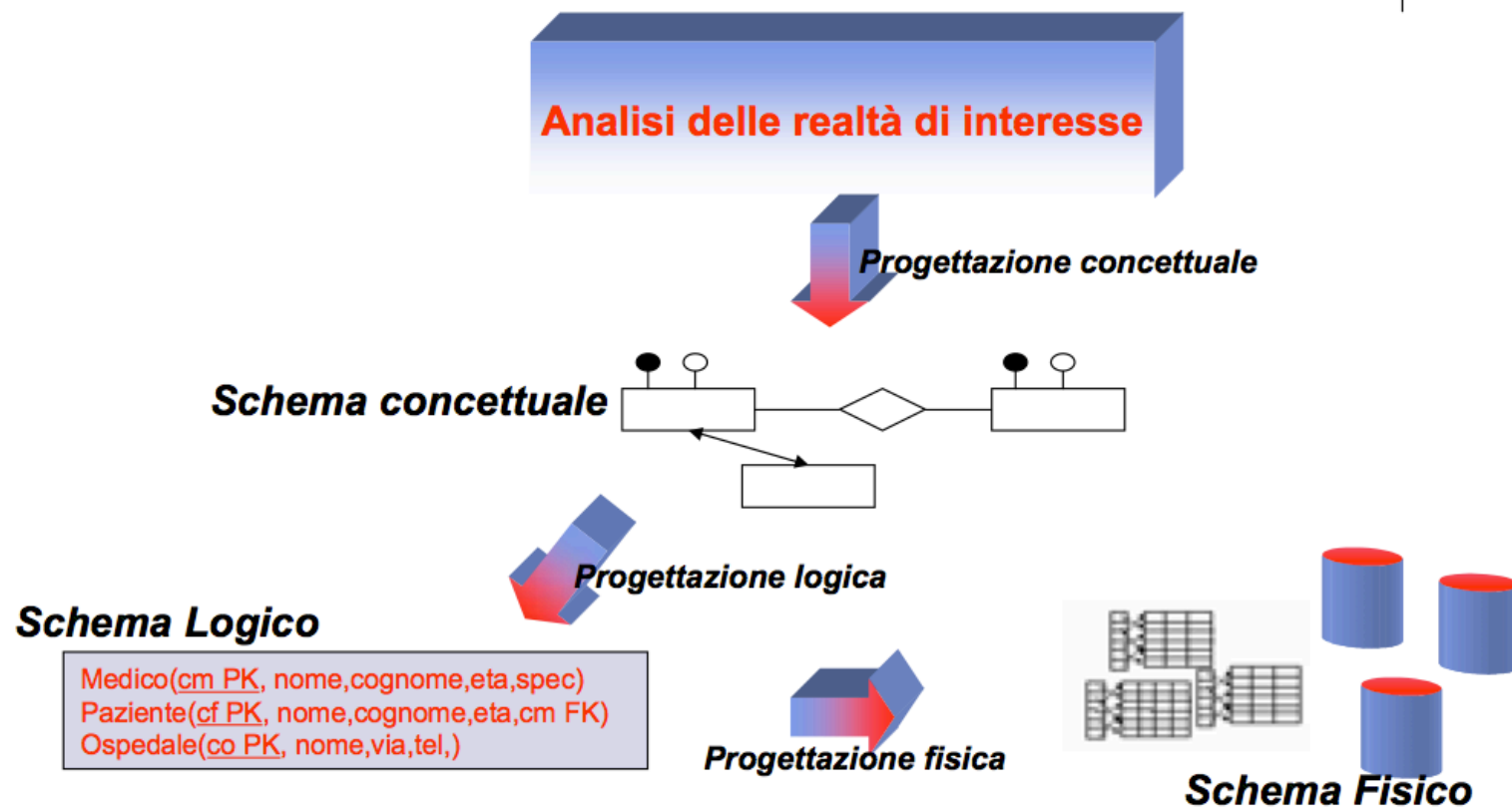
■ Schema fisico

- Dipende dallo specifico sistema di gestione di basi di dati
- Si basa sui criteri di organizzazione fisica dei dati nel sistema

+ Prodotti delle varie fasi del



+ Visione d'insieme





Linguaggi ed utenti delle basi di dati

+ Personaggi e interpreti

- progettisti e realizzatori di DBMS
- progettisti della base di dati e amministratori della base di dati (DBA)
- progettisti e programmatori di applicazioni
- utenti
 - utenti **finali** (terminalisti): eseguono applicazioni predefinite (transazioni)
 - utenti **casuali**: eseguono operazioni non previste a priori, usando linguaggi interattivi

+ Database administrator (DBA)

- Persona o gruppo di persone responsabile del controllo centralizzato e della gestione del sistema, delle prestazioni, dell'affidabilità, delle autorizzazioni
- Le funzioni del DBA includono quelle di progettazione, anche se in progetti complessi ci possono essere distinzioni

+ Transazioni (per l'utente)

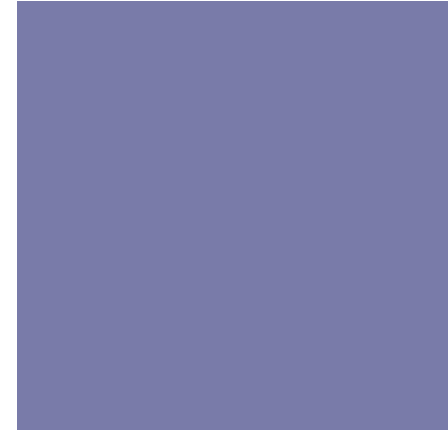
- Programmi che realizzano attività frequenti e predefinite, con poche eccezioni, previste a priori.
- Esempi:
 - versamento presso uno sportello bancario
 - emissione di certificato anagrafico
 - dichiarazione presso l'ufficio di stato civile
 - prenotazione aerea
- Le transazioni sono di solito realizzate in linguaggio ospite (tradizionale o ad hoc)

+ Transazioni, due accezioni

- Per l'utente:
 - programma a disposizione, da eseguire per realizzare una funzione di interesse
- Per il sistema:
 - sequenza indivisibile di operazioni (cfr. affidabilità)



Vantaggi e svantaggi dei DBMS

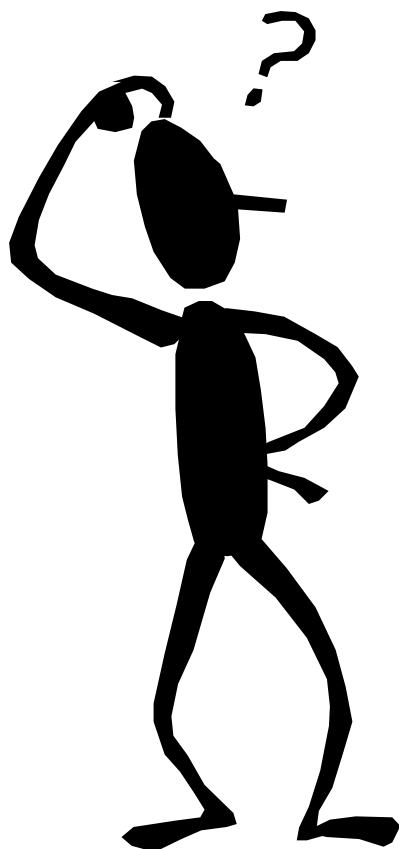


+ Vantaggi dei DBMS

- Dati come risorsa comune, base di dati come modello della realtà
- Gestione centralizzata con possibilità di standardizzazione ed “economia di scala”
- Disponibilità di servizi integrati
- Riduzione di ridondanze e inconsistenze
- Indipendenza dei dati (favorisce lo sviluppo e la manutenzione delle applicazioni)

+ Svantaggi dei DBMS

- costo dei prodotti e della transizione verso di essi
- non scorporabilità delle funzionalità (con riduzione di efficienza)



Ci sono domande?