



A.A. 2020-2021
CdL Informatica Triennale

Prof. Alfredo Pulvirenti (A-L)
Prof. Salvatore Alaimo (M-Z)

- Prof. Alfredo Pulvirenti
 - Ufficio: Terzo Blocco, stanza n. 35
 - Tel. 095-7383087
 - e-mail: apulvirenti@dmf.unict.it
 - Homepage: <http://www.dmf.unict.it/~apulvirenti/>
- Prof. Salvatore Alaimo
 - Ufficio: Terzo Blocco, stanza n. 35
 - Tel. 095-7383087
 - e-mail: alaimos@dmf.unict.it
 - Homepage: <http://www.alaimos.com/>

-
- 9 CFU (72h)
 - Prerequisiti
 - Programmazione 2

Informazioni

- Orario lezioni
 - Lun/Mer/Ven 10:00-12.00 (A-L)
 - Lun/Mer 10:00-12.00; Ven 8:00-10:00(M-Z)
- Ricevimento:
 - Giovedì 9.30-11.30 su Teams
- Durante il corso sarà usato STUDIUM (e Teams) come canale di comunicazione ufficiale;
- Materiale delle lezioni su STUDIUM (e Teams)
<http://studium.unict.it/>

Programma

- Introduzione alle basi di dati:
 - Generalità sui DBMS
 - Modelli dei dati e indipendenza fisica
 - Linguaggi e utenti delle basi di dati
- Il modello Relazionale dei dati
 - Relazioni, attributi, istanze di relazione, n-uple
 - Vincoli di integrità, chiavi, chiavi esterne
- Algebra relazionale
 - Operatori fondamentali e derivati
 - Algoritmo per l'ottimizzazione delle query

Programma

- Il linguaggio SQL
 - Il linguaggio di definizione dei dati (DDL):
 - Definizione di tabelle, domini, indici.
 - Specifica di semplici vincoli di integrità
 - Il linguaggio di interrogazione (DML):
 - Operatori di join-selezione-proiezione, operatori insiemistici
 - Operatori di raggruppamento.
 - Interrogazioni nidificate e correlate
 - Query ricorsive.
 - Il linguaggio di manipolazione dei dati (DML):
 - Inserimento, eliminazione e modifica di record.
 - Viste
 - Controllo dell'accesso ad una base di dati
 - Grant/Revoke
 - Basi di dati attive
 - Trigger

Programma

- Progettazione delle basi di dati:
 - Progettazione concettuale
 - Progettazione logica
- Normalizzazione delle basi di dati:
 - Anomalie
 - Dipendenze Funzionali
 - Decomposizioni di Schemi
 - Forme Normali: di Boyce-Codd e 3NF
- Organizzazione fisica e gestione delle interrogazioni
 - Indici primari e secondari, Strutture ad albero
 - Gestione delle interrogazioni: esecuzione ed ottimizzazioni

Programma

- Sviluppo di applicazioni
 - Stored procedure
 - Linguaggi host: Php
 - Oggetti persistenti
- Gestione delle transazioni
 - Controllore dell'affidabilità
 - Log e gestione dei guasti
 - Controllo concorrenza, lock a due fasi
- Cenni sui NoSQL database
 - Diversificazione dei sistemi
 - Modelli dei dati nei sistemi NoSQL
 - Gestione delle Transazioni, Map-Reduce
- Basi di dati per XML
 - Definizione di dati semistrutturati in XML
 - Xpath, XLS, Xquery

Programma

- DBMS:
 - MySQL
 - NoSQL database
 - Neo4j

Bibliografia

- Libri di testo:
 - Atzeni, Ceri, Fraternali, Paraboschi, Torlone, *Basi di Dati* (V edizione), McGraw-Hill.
 - Albano, Ghelli, Orsini, *Fondamenti di basi di dati*, Zanichelli.
- Opzionale:
 - Garcia-Molina, Ullman, Widom, *Database Systems: The Complete Book*, Prentice Hall

- Scritto
 - Esercizi + domande sulla teoria
- Progetto
 - Implementazione di una base di dati equipaggiata con una opzionale interfaccia web.
 - **Presentare una relazione dettagliata sulla progettazione completa del database.**
 - Il progetto viene richiesto tramite email e viene assegnato durante il ricevimento.
 - Validità un anno.
 - Max 1 persona per progetto.
 - I progetti devono essere consegnati dopo il superamento dell'esame scritto.
 - Il docente comunicherà una o più date (all'incirca un mese dopo lo scritto) per la presentazione del progetto.
 - La relazione va consegnata tramite email almeno 72 ore prima della data fissata per la presentazione
 - Altre comunicazioni...STUDIUM

- Modalità d'esame:
 - Per chi segue:
 - 2 prove in itinere (della durata di 2h ciascuna) + progetto finale;
 - Esame della durata di 3h sui contenuti del corso (aperto a tutti).

Metodo di studio

- Altamente consigliato seguire interamente il corso
- Studio individuale, con riflessione approfondita sui concetti;
- Svolgere esercizi;
- Sviluppo di progetti e esercitazioni pratiche anche con l'uso di DBMS (es. MySQL)

Introduzione alle basi di dati

Prof. Alfredo Pulvirenti (A-L)

Prof. Salvatore Alaimo (M-Z)

(Atzeni-Ceri Capitolo 1)

Base di Dati: definizione

- Un insieme organizzato di dati, che esistono e si evolvono nel tempo, disponibili in una certa struttura (impresa, banca, ospedale, ...) per lo svolgimento della propria attività.

Dove sono

- Le basi di dati al giorno d'oggi sono essenziali in ogni tipologia di attività.
- Esempi:
 - Usate per mantenere “record” interni ad una struttura;
 - Per offrire servizi attraverso il World-Wide-Web;
 - Per supportare diversi altri processi (commerciali);
 - Alla base di ricerche scientifiche, utilizzate per memorizzare e rappresentare dati (collezionati in una qualche maniera).

Sistema organizzativo e sistema informativo

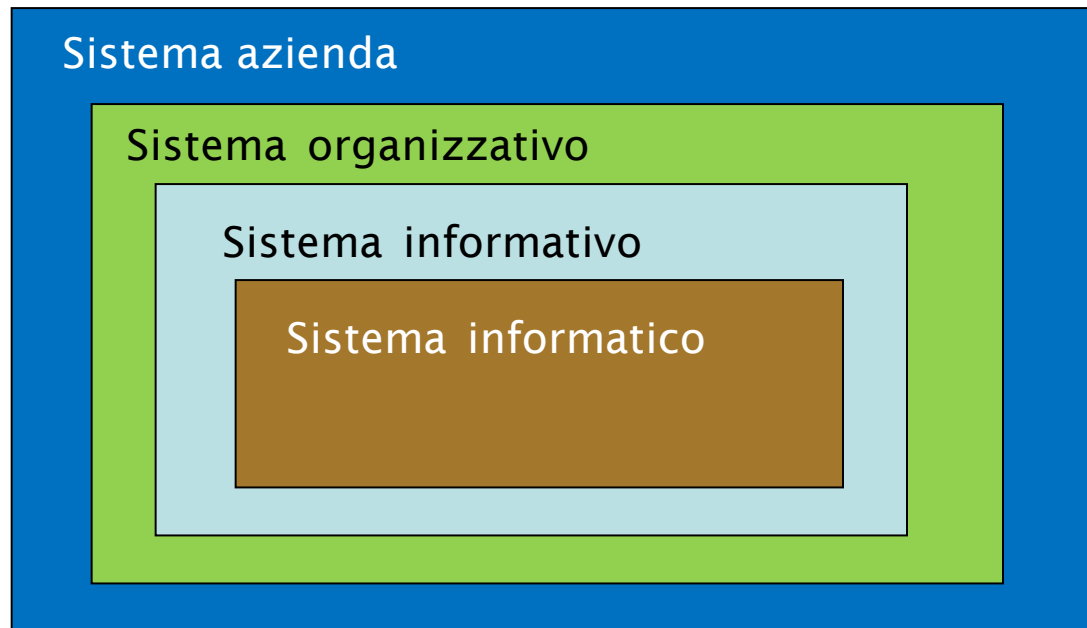
- Insieme di **risorse** (persone, denaro, materiali e beni, **informazioni**) e **regole** per lo svolgimento coordinato delle attività al fine del perseguimento degli scopi;
- Il sistema informativo è parte del sistema organizzativo (eventualmente non esplicitato nella struttura), il sistema informativo esegue/gestisce processi informativi (cioè i processi che coinvolgono informazioni).

Sistemi informativi e automazione

- Il concetto di “sistema informativo” è indipendente da qualsiasi automazione:
 - esistono organizzazioni la cui ragion d’essere è la gestione di informazioni (p. es. servizi anagrafici e banche) e che operano da secoli.

Sistemi informativi, informazioni e dati

- porzione automatizzata del sistema informativo:
 - la parte del sistema informativo che gestisce informazioni con tecnologia informatica



Gestione delle informazioni/1

- Aspetto fondamentale:
 - Razionalizzazione e standardizzazione dell'organizzazione delle informazioni e delle procedure.
 - Anche prima di essere informatizzati, molti sistemi informativi si sono evoluti in questa direzione;
 - Esempio: gli uffici anagrafe fino a pochi anni fa

Gestione delle informazioni/2

- In qualsiasi attività le informazioni vengono gestite (registrate e scambiate) in diversi modi:
 - idee informali;
 - linguaggio naturale (scritto o parlato, formale o colloquiale, in una lingua o in un'altra);
 - disegni, grafici, schemi;
 - numeri e codici.
- e su vari supporti
 - memoria umana, carta, dispositivi elettronici.

Informazioni e dati

- Nei sistemi informatici (e non solo), le informazioni vengono rappresentate in modo essenziale e spartano attraverso i **dati**
- Dal Vocabolario della lingua italiana (1987) distinguiamo:
 - **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

- I dati hanno bisogno di essere interpretati

Esempio

su un foglio di carta sono due dati: ‘Mario’ ‘2075’

Se il foglio di carta viene fornito in risposta alla domanda

- “A chi mi devo rivolgere per il problema X; qual è il suo interno?”
- 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).

- La potenza delle basi di dati deriva da un bagaglio di conoscenze e tecnologie che sono state sviluppate in diverse decadi che hanno dato luogo a software specializzati chiamati:

**Sistema di gestione di basi di dati
DataBase Management System, o DBMS**

Basi di dati/2

- Un DBMS è uno strumento particolarmente potente per la **creazione e la gestione efficiente ed efficace di grandi quantità di dati.**

DataBase Management System — DBMS

- Sistema (**prodotto software**) in grado di gestire collezioni di **dati che siano** (anche):
 - **grandi** (di dimensioni (molto) maggiori della memoria centrale dei sistemi di calcolo utilizzati)
 - **persistenti** (con un periodo di vita indipendente dalle singole esecuzioni dei programmi che le utilizzano)
 - **condivise** (utilizzate da applicazioni diverse)

DataBase Management System — DBMS

- **Ambiente di programmazione:**
 - Un DBMS consente all'utente o ad un'applicazione di accedere e modificare i dati attraverso un potente linguaggio di interrogazione.
 - garantendo **affidabilità** (resistenza a malfunzionamenti hardware e software) e **privatezza** (con una disciplina e un controllo degli accessi).
 - Come ogni prodotto informatico, un DBMS deve essere **efficiente** (utilizzando al meglio le risorse di spazio e tempo del sistema) ed **efficace** (rendendo produttive le attività dei suoi utilizzatori).

DBMS relazionali

- Prodotti software (complessi) disponibili sul mercato; esempi:
 - DB2
 - Oracle
 - SQLServer
 - MySQL
 - PostgreSQL
 - ...

DB-Engines Ranking

The DB-Engines Ranking ranks database management systems according to their popularity. The ranking is updated monthly.

Read more about the [method](#) of calculating the scores.



283 systems in ranking, October 2015

Rank	DBMS	Database Model	Score		
			Oct 2015	Sep 2015	Oct 2014
1.	Oracle	Relational DBMS	1466.95	+3.58	-4.95
2.	MySQL	Relational DBMS	1278.96	+1.21	+15.99
3.	Microsoft SQL Server	Relational DBMS	1123.23	+25.40	-96.37
4.	MongoDB	Document store	293.27	-7.30	+52.86
5.	PostgreSQL	Relational DBMS	282.13	-4.05	+24.41
6.	DB2	Relational DBMS	206.81	-2.33	-0.86
7.	Microsoft Access	Relational DBMS	141.83	-4.17	+0.19
8.	Cassandra	Wide column store	129.01	+1.41	+43.30
9.	SQLite	Relational DBMS	102.67	-4.99	+7.71
10.	Redis	Key-value store	98.80	-1.86	+19.42
11.	SAP Adaptive Server	Relational DBMS	85.64	-0.88	-1.15
12.	Solr	Search engine	79.07	-2.87	-0.89
13.	Teradata	Relational DBMS	73.44	-0.83	+6.09
14.	Elasticsearch	Search engine	70.23	-1.32	+27.41
15.	HBase	Wide column store	57.24	-1.79	+10.14
16.	Hive	Relational DBMS	53.56	+0.03	+18.78

Rank			DBMS	Database Model	Score		
Oct 2020	Sep 2020	Oct 2019			Oct 2020	Sep 2020	Oct 2019
1.	1.	1.	Oracle	Relational, Multi-model	1368.77	-0.59	+12.89
2.	2.	2.	MySQL	Relational, Multi-model	1256.38	-7.87	-26.69
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	1043.12	-19.64	-51.60
4.	4.	4.	PostgreSQL	Relational, Multi-model	542.40	+0.12	+58.49
5.	5.	5.	MongoDB	Document, Multi-model	448.02	+1.54	+35.93
6.	6.	6.	IBM Db2	Relational, Multi-model	161.90	+0.66	-8.87
7.	8.	7.	Elasticsearch	Search engine, Multi-model	153.84	+3.35	+3.67
8.	7.	8.	Redis	Key-value, Multi-model	153.28	+1.43	+10.37
9.	9.	11.	SQLite	Relational	125.43	-1.25	+2.80
10.	10.	10.	Cassandra	Wide column	119.10	-0.08	-4.12
11.	11.	9.	Microsoft Access	Relational	118.25	-0.20	-12.93
12.	12.	13.	MariaDB	Relational, Multi-model	91.77	+0.16	+5.00
13.	13.	12.	Splunk	Search engine	89.40	+1.51	+2.57
14.	14.	15.	Teradata	Relational, Multi-model	75.79	-0.61	-2.95
15.	15.	14.	Hive	Relational	69.55	-1.62	-15.19
16.	16.	16.	Amazon DynamoDB	Multi-model	68.41	+2.23	+8.24
17.	17.	25.	Microsoft Azure SQL Database	Relational, Multi-model	64.40	+3.95	+36.89
18.	18.	19.	SAP Adaptive Server	Relational	55.16	+1.15	-0.67
19.	19.	20.	SAP HANA	Relational, Multi-model	54.24	+1.38	-1.11
20.	20.	17.	Solr	Search engine	52.48	+0.86	-5.09
21.	21.	22.	Neo4j	Graph	51.34	+0.71	+1.87

Basi di dati e condivisione delle informazioni

- Una base di dati e' una risorsa **integrata**, **condivisa** fra le varie applicazioni
- Conseguenze
 - Attivita' diverse su dati in parte condivisi:
 - meccanismi di **autorizzazione**
 - Attivita' multi-utente su dati condivisi:
 - controllo della **concorrenza**
- Problemi
 - **Ridondanza**: informazioni ripetute
 - Rischio di **incoerenza**: le versioni possono non coincidere

Descrizione dei dati

- **Modello dei dati**, formalismo (matematico) composto da due parti:
 - Una **notazione** per **descrivere** i dati;
 - Un insieme di **operatori** per **manipolare** tali dati.
- 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

Descrizione dei dati nei DBMS

- Descrizioni e rappresentazioni dei dati a livelli diversi
 - Permettono l'**indipendenza dei dati** dalla rappresentazione fisica:
 - i programmi fanno riferimento alla struttura a livello più alto (logica), e le rappresentazioni sottostanti possono essere modificate senza necessità di modifica dei programmi.

Due tipi (principali) di modelli

- **modelli logici**: utilizzati nei DBMS esistenti per l'organizzazione dei dati
 - Si chiama logico per sottolineare il fatto che le strutture utilizzate da questi modelli, pure essendo astratte, riflettono una particolare organizzazione
 - utilizzati dai programmi
 - indipendenti dalle strutture fisiche
- esempi: **relazionale**, reticolare, gerarchico, a oggetti

Due tipi (principali) di modelli

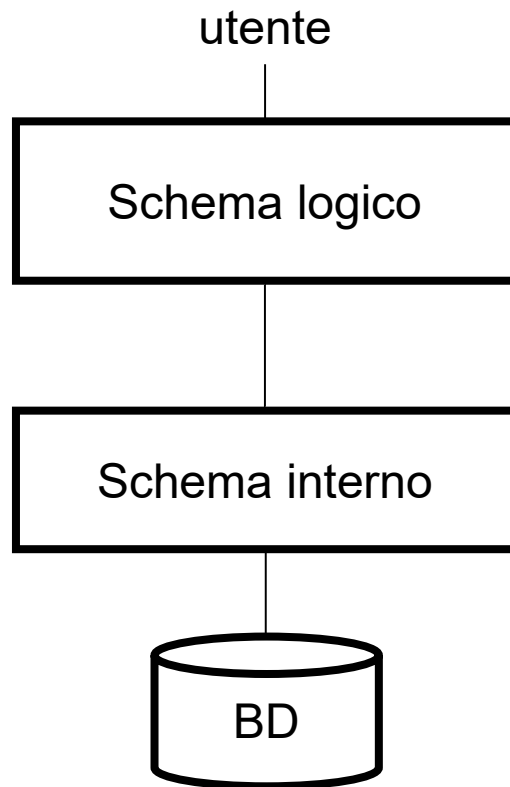
- **modelli concettuali**: permettono di rappresentare i dati in modo indipendente da ogni sistema
 - cercano di descrivere i concetti del mondo reale piuttosto che i dati utili a rappresentarli
 - sono utilizzati nelle fasi preliminari di progettazione
- il più noto è il modello **Entità-Relazione**

Esempio

- **Modello relazionale:**
 - prevede il costruttore **relazione**, che permette di definire insiemi di record omogenei;
 - Usa un insieme di nomi (detti attributi) per descrivere i dati, che individuano le colonne delle tabelle.

Nome	Matricola	Indirizzo	Telefono
Mario Rossi	123456	Via Etnea 18	777777
Maria Bianchi	234567	Via Roma 2	888888
Giovanni Verdi	345678	Via Etnea 18	999999
Enzo Gialli	456789	Via Catania 3	444444

Architettura di un DBMS (semplificata)



- **schema logico**: descrizione della base di dati nel modello logico (ad esempio, la struttura della tabella)
- **schema fisico**: rappresentazione dello schema logico per mezzo di strutture memorizzazione (file)

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)

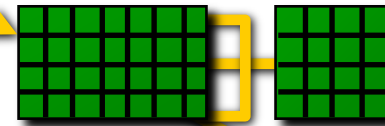
Modello Dati



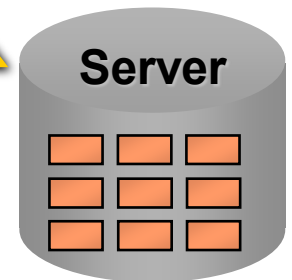
Utente



Modello concettuale

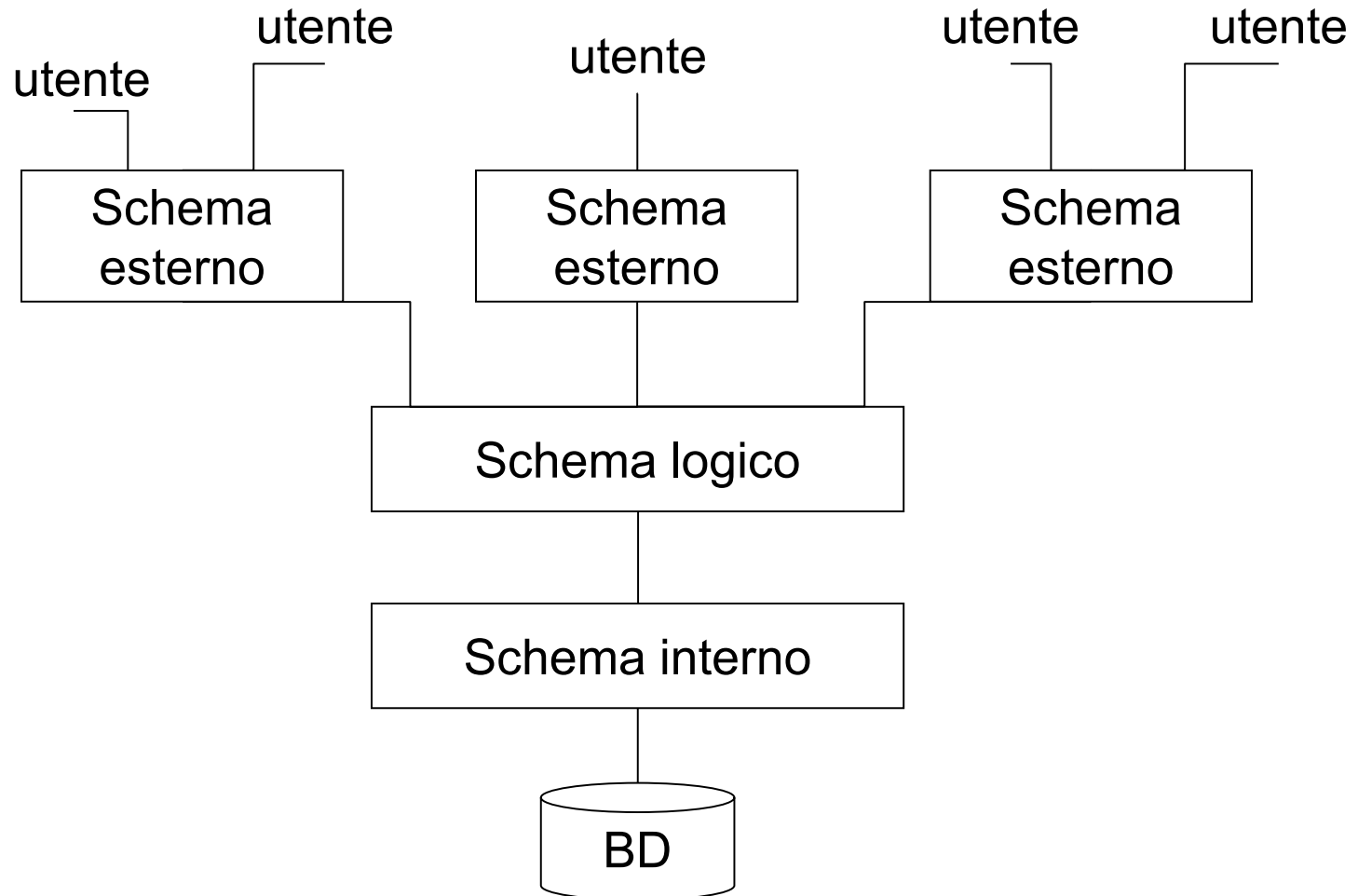


Modello logico



**Schema interno
Tabelle sul disco**

Architettura standard(ANSI/SPARC) a tre livelli per DBMS



Architettura ANSI/SPARC: schemi

- **schema logico:** descrizione dell'intera base di dati nel modello logico “principale” del DBMS
- **schema 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)

Una Vista

Corsi

Corso	Docente	Aula
Basi di dati	Rossi	4
Algoritmi	Neri	2
Reti	Bruni	1
Analisi dati	Rossi	G

Aule

Nome	Edificio	Piano
4	DMI	Terra
1	DMI	Terra
G	DAU	Primo

CorsiSedi

Corso	Aula	Edificio	Piano
Basi di dati	4	DMI	Terra
Reti	1	DMI	Terra
Analisi dati	G	DAU	Primo

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

Linguaggi per le basi di dati

- Nei DBMS distinguiamo

Data-Definition Language (DDL):

- consente all'utente di creare nuovi database e di specificarne i loro **schemi** (logici, esterni, fisici) la loro strutturazione logica

• ***Data-Manipulation Language (DML):***

- da agli utenti la possibilità di interrogare e modificare **istanze** di basi di dati

Query su una tabella

- Vorrei conoscere l'indirizzo e il telefono di Giovanni Verdi

Nome	Matricola	Indirizzo	Telefono
Mario Rossi	123456	Via Etnea 18	777777
Maria Bianchi	234567	Via Roma 2	888888
Giovanni Verdi	345678	Via Etnea 18	999999
Enzo Gialli	456789	Via Catania 3	444444

Indirizzo	Telefono
Via Etnea 18	999999

Query su due tabelle

- Quali esami ha superato Mario Rossi?

Corso	Matricola	Voto
Programmazione 1	345678	27
Architettura	123456	30
Matematica discreta	234567	19
Basi di Dati	345678	28

Nome	Matricola	Indirizzo	Telefono
Mario Rossi	123456	Via Etnea 18	777777
Maria Bianchi	234567	Via Roma 2	888888
Giovanni Verdi	345678	Via Etnea 18	999999
Enzo Gialli	456789	Via Catania 3	444444

Corso
Architettura

Query su più tabelle

Corso	Matricola	Voto
Programmazione 1	345678	27
Architettura	123456	30
Matematica discreta	345678	19
Basi di Dati	345678	28

Nome	Matricola	Indirizzo	Telefono
Mario Rossi	123456	Via Etnea 18	777777
Maria Bianchi	234567	Via Roma 2	888888
Giovanni Verdi	345678	Via Etnea 18	999999
Enzo Gialli	456789	Via Catania 3	444444

Corso	Professore
Architettura	Barbanera
Programmaizone 1	Cincotti
Matematica discreta	Milici
Basi di dati	Pulvirenti

Quali professori hanno dato più di 24 a Giovanni Verdi e in quali corsi?

Corso	Professore
Programmazione 1	Cincotti
Basi di dati	Pulvirenti

Vantaggi e svantaggi DBMS

- Pro
 - 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)
- Contro
 - costo dei prodotti (a volte) e della transizione
 - architetture complesse a fronte di requisiti minimi comuni