

Presentazione del corso

corso di basi di dati e laboratorio

Prof. Alfio Ferrara

Anno Accademico 2020/2021

1 Presentazione

1.1 Informazioni utili

Docenti

Corso di Basi di Dati

Prof. Alfio Ferrara

alfio.ferrara@unimi.it

Ricevimento: Venerdì, 11.00-13.00, S7012 (7 piano)

Dipartimento di Informatica, via Celoria 18, 20133 Milano

Modulo di Laboratorio

Prof. Stefano Montanelli

stefano.montanelli@unimi.it

Ricevimento: Giovedì, 11.00-12.00, S7015 (7 piano)

Dipartimento di Informatica, via Celoria 18, 20133 Milano

Prof. Valerio Bellandi

valerio.bellandi@unimi.it

Dipartimento di Informatica, via Celoria 18, 20133 Milano

Uso dell'email

- In nessun caso, si svolge attività di ricevimento studenti via email. L'email va utilizzata solo per motivate ragioni e in assenza di informazioni adeguate sul sito web del corso.

- In generale, nel mandare un'email è necessario usare **sempre e solo** l'indirizzo di posta elettronica di Ateneo, specificare il corso a cui si fa riferimento, e indicare il proprio nome e cognome.

Sito web del corso

<https://aferrarabd.ariel.ctu.unimi.it>

- Il sito web del corso fornisce informazioni e avvisi **aggiornati e certificati** dai docenti
- Per gli studenti registrati, il sito dà inoltre accesso al materiale didattico aggiuntivo

Orari e luogo delle lezioni

- Corso
 - Mercoledì 13.30-16.30 Aula Magna ALBERTO BERTONI (Informatica - Celoria 18)
 - Giovedì 13.30-16.30 aula G14 (Via Golgi)
- Laboratorio
 - TURNO A, cognomi A-K: Martedì 10.30-14.30 - Delta Lab (Informatica - Celoria 18)
 - TURNO B, cognomi L-Z: Lunedì 10.30-14.30 - Omega Lab (Informatica - Celoria 18)

1.2 Programma e organizzazione del corso

Struttura generale

- Concetti e architettura di un sistema di basi di dati
- Modello relazionale, vincoli, normalizzazione
- Modellazione dei dati, modello ER e nozioni di progettazione
- Progettazione logica
- Algebra relazionale
- SQL
- Organizzazione fisica dei dati e indici
- Sicurezza e controllo dell'accesso
- Transazioni (concetti generali)

Struttura del laboratorio

- Il DBMS PostgreSQL
 - creazione e manipolazione di schemi
 - gestione di utenti e ruoli
 - firewall degli accessi (hba.conf)
 - dump di basi di dati
 - linguaggio procedurale (PLpgSQL)
- Programmazione web con PHP
 - architettura client/server
 - protocollo HTTP
 - passaggio di parametri GET/POST
 - cookie/sessioni
 - interazione con i DBMS

- Esercitazioni

Materiale didattico

Il corso è costruito a partire da due testi di riferimento:

- R. Elmasri, S.B. Navathe Sistemi di basi di dati - Fondamenti (6 ed.) edizione italiana a cura di Silvana Castano Pearson-Addison Wesley, 2011.
- P. Atzeni, S. Ceri, S. Paraboschi, R. Torlone Basi di dati - Modelli e linguaggi di interrogazione (2 ed.) McGraw-Hill, 2006.

Altri testi utili

- S. Castano, M. Fugini, G. Martella, P. Samarati Database Security Addison-Wesley, 1995.
- Materiali didattici, dispense e materiali di laboratorio (sul sito del corso a seguito dell'iscrizione).

Come seguire e come studiare

- La frequenza regolare del corso è fortemente consigliata
- I materiali didattici sono integrativi ma non sostitutivi dei testi
- L'obiettivo è imparare le basi di dati, non passare l'esame; quest'ultimo avvenimento è solo la conseguenza del primo
- Usare i testi come riferimento per l'approfondimento e la piena comprensione degli argomenti trattati a lezione

1.3 Modalità d'esame

Organizzazione delle prove d'esame

L'esame consiste in una prova scritta (obbligatoria), una prova di laboratorio (obbligatoria) e una discussione orale (solo se necessaria secondo il regolamento che segue).

Prova scritta E' composta da una parte A con domande di teoria e da una parte B con esercizi. La prova scritta è valutata in trentesimi (voto S).

Prova di laboratorio E' composta da una parte A con domande di teoria e da una parte B con esercizi. La prova di laboratorio è valutata in trentesimi (voto P).

Discussione orale Dopo aver sostenuto la prova scritta e la prova di laboratorio, lo studente riceve un voto in trentesimi corrispondente alla media pesata delle due prove sostenute così calcolata:

$$voto = \frac{2S + P}{3}$$

Se si è conseguito un voto di almeno 25/30 in entrambe le prove ($S \geq 25$ e $P \geq 25$): la discussione orale è possibile. Se non si sostiene la discussione orale, il voto complessivo della prova scritta e del progetto di laboratorio è automaticamente confermato. Se NON si è conseguito un voto di almeno 25/30 in ciascuna delle due prove ($S < 25$ o $P < 25$) la discussione orale non è possibile. Il voto complessivo della prova scritta e del progetto di laboratorio è automaticamente confermato.

Valutazione: esempi

- $S = 30, P = 18, V = 26$, orale impossibile $\rightarrow F = 26$
- $S = 18, P = 30, V = 22$, orale impossibile $\rightarrow F = 22$
- $S = 25, P = 30, V = 27$, orale possibile $\rightarrow F \geq 27$
- $S = 30, P = 25, V = 29$, orale possibile $\rightarrow F \geq 29$

- L'eventuale insufficienza di una delle due prove implica necessariamente l'insufficienza dell'intero voto.

La prova scritta

- La prova scritta consta di una parte di domande chiuse di teoria (Parte A) e una parte di esercizi (Parte B)
- Il tempo complessivo per la prova è di 2 ore
- La Parte A è svolta per prima con un tempo massimo di 30 minuti. Nel caso la Parte A sia consegnata prima del 30 minuti previsti, il tempo risparmiato rimane a disposizione per lo svolgimento della Parte B.
- L'accesso alla Parte B è possibile solo se si è risposto correttamente a almeno 3/5 delle domande della Parte A (in genere 3 su 5).

La prova di laboratorio

Le prove di laboratorio si svolgono nei mesi di febbraio, aprile, luglio, settembre. Le date sono pubblicate sul sito del corso appena possibile nella sezione avvisi e seguono indicativamente di qualche giorno la prova scritta. Le iscrizioni sono aperte nelle 3 settimane antecedenti ciascuna prova e avvengono mediante una procedura web dedicata che sarà comunicata in seguito. Le iscrizioni alla prova di laboratorio NON avvengono via SIFA.

2 Introduzione alle basi di dati

2.1 Definizione e contesto

Definizione di basi di dati

Base di dati

Una base di dati è una collezione di dati correlati. Per **dati** intendiamo fatti noti che possono essere memorizzati e che hanno un significato implicito.

- Esempi di dati sono nomi, indirizzi, numeri di telefono, quantità, quali ad esempio temperature, altezze.
- Esempi di basi di dati sono rubriche del telefono, agende, cataloghi delle biblioteche, sia digitali sia cartacei.

Caratteristiche delle basi di dati

Nonostante in questo senso quasi ogni collezione di dati possa considerarsi una base di dati, in genere si assume che una base di dati abbia alcune caratteristiche distintive.

- Rappresenta un aspetto limitato della realtà, riferito a un sottoinsieme dei dati disponibili denominato **universo del discorso**.
- Costituisce un insieme di dati logicamente coerenti e dotati di un significato intrinseco; una collezione casuale di dati non è una base di dati.
- Una base di dati è progettata, costruita e popolata per uno scopo specifico, rivolto a un gruppo di utenti determinato e a alcune specifiche applicazioni a cui gli utenti sono interessati.

Correlazione e informazione

In una base di dati i dati sono **correlati**. Correlare i dati ha lo scopo primario di fornirne un'**interpretazione** e trasformarli pertanto in **informazione**.

- Dati non correlati: Roma, 41.89, 12.48
- Dati correlati: Roma → Posizione: Latitudine 41.89, Longitudine 12.48

2.2 Considerazioni generali sull'uso delle basi di dati

Basi di dati VS gestione applicativa dei dati

- Ogni applicazione prevede un sistema di gestione dei dati, tipicamente in memoria e su file definendo di volta in volta le strutture dati necessarie e l'organizzazione dei file.
- Utilizzare invece una base di dati per la gestione dei dati in ausilio alle applicazioni è una scelta che implica alcune caratteristiche specifiche nel modo in cui i dati vengono trattati rispetto alle applicazioni.

Caratteristiche delle basi di dati

- **Autodescrizione**: contrariamente ai dati memorizzati dalle applicazioni, una base di dati non contiene solo i dati, ma anche una descrizione completa della sua struttura e dei suoi vincoli, attraverso un **catalogo**.
- **Indipendenza tra programmi e dati**: la struttura di memorizzazione dei dati è astratta e coerente al **modello dei dati** usato dalla base di dati. In tutti i casi essa è del tutto indipendente dalle applicazioni.

- **Viste multiple dei dati:** a partire dagli stessi dati, una base di dati può fornire un accesso a sottoinsiemi e a forme di organizzazione diversi dei dati a diversi utenti e applicazioni, per mezzo di **viste virtuali**.
- **Condivisione e controllo della concorrenza:** una base di dati garantisce che utenti e applicazioni diverse possano gestire in modo concorrente gli stessi dati, garantendo al tempo stesso la loro coerenza e correttezza, per mezzo di **transazioni**.

3 Sistemi di gestione delle basi di dati

3.1 Modelli dei dati

Definizione di modello dei dati

Organizzare i dati in una base di dati richiede la definizione di un modello logico che detti le regole di organizzazione dei dati stessi.

Modello dei dati

Insieme di **costrutti**, **formalismi di rappresentazione**, e **operazioni di base** necessari alla realizzazione della descrizione astratta e indipendente dalla memorizzazione fisica dei dati, ai vincoli che ne garantiscono coerenza e correttezza, e alla loro gestione

Il modello relazionale

- Esistono molti modelli dei dati, fra i quali i modelli reticolari, gerarchici, a oggetti.
- Il modello dei dati più diffuso (e più utilizzato dai principali sistemi di gestione di basi di dati) è il **modello relazionale**, oggetto di questo corso.

Le nozioni di schema e istanza

- Una caratteristica fondamentale di molti modelli dei dati, compreso il modello relazionale, è di operare una netta distinzione fra la descrizione dei dati, detta **schema**, e i dati stessi, detti **istanza**.
- Lo schema è un modello logico. Definisce cioè la struttura dei dati e i vincoli che i dati devono rispettare all'interno della base di dati. L'istanza è un particolare insieme di dati organizzato in modo da rispettare lo schema.
- Ne consegue che lo schema è fisso e (in genere) non subisce variazioni nel tempo, mentre l'istanza è continuamente aggiornata con inserimenti, cancellazioni e cambiamenti dei dati.

- L'istanza di una base di dati in un particolare istante di tempo è detto **stato della base di dati**.
- Si fa talvolta riferimento allo schema con il termine **intensione** e all'istanza col termine **estensione** della base di dati.

Esempi diversi

SCHEMA | **ISTANZA**

Cognome	Rossi
Nome	Antonia
nato il	14/05/1985
(atto n. P. S.)	
a.	
Cittadinanza	Italia
Residenza	Italia
Via	Casilina 320
Stato civile	Sposata
Professione	Insegnante
CONNOTATI E CONTRASSEGNI SALIENTI	
Statura	1,68
Capelli	Biondi
Occhi	Azzurri
Segni particolari	

SCHEMA

Studenti	Matricola	Cognome	Nome	Nascita
	276545	Rossi	Maria	23/04/1968
	276545	Neri	Anna	23/04/1972
	788854	Verdi	Fabio	12/02/1972

ISTANZA

3.2 Architettura e indipendenza dei dati

Database Management System

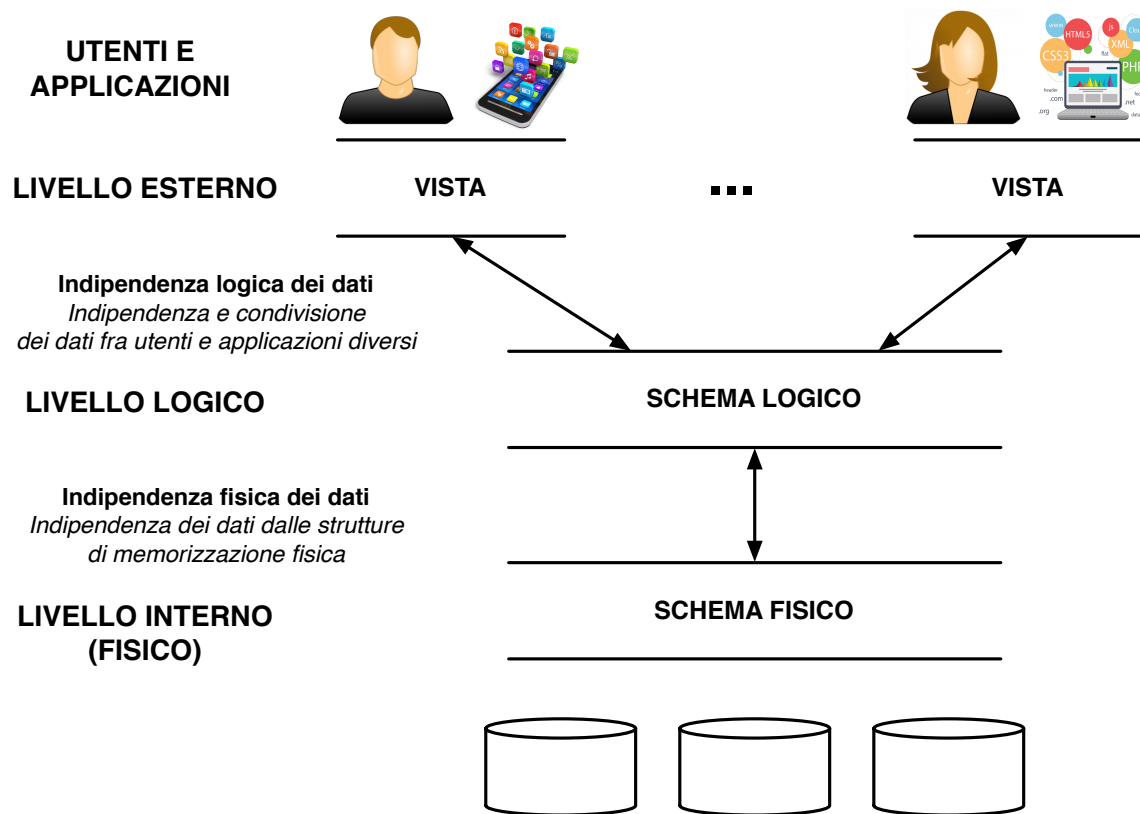
Il modello dei dati e tutte le caratteristiche derivanti dall'uso di una base di dati sono realizzate in uno insieme di strumenti software che prende il nome di **DBMS (DataBase Management System)**.

DBMS - DataBase Management System

Insieme di programmi che consente la creazione a manutenzione di basi di dati. In particolare, consente di **definire**, **costruire**, **manipolare**, e , **condividere** basi di dati per vari utenti e applicazioni.

Architettura a tre livelli

L'architettura concettuale di una base di dati è caratterizzata da tre livelli, che corrispondono a due diversi passaggi d'astrazione e indipendenza dei dati:



Categorie di linguaggi

- **DDL:** Data Definition Language → istruzioni per la creazione e manutenzione dello schema
- **DML:** Data Manipulation Language → istruzioni per la manutenzione e interrogazione dei dati (istanza)

Architettura generale di un DBMS

