

DATABASE

Struttura e operazioni principali

Database (relazionali)

- Una **base di dati** o **database (DB)** serve a memorizzare dati in modo strutturato
 - Qualcosa di più che rappresentare i dati in maniera ordinata come in Excel
- I DB sono formati da **tabelle che contengono i dati**
- Esistono diversi tipi di DB: noi studieremo i **database relazionali**
 - come dice il nome, i **DB relazionali descrivono le relazioni** tra i dati contenuti nelle tabelle

Copyright © 2006 Pearson Education. All rights reserved.

Tabelle

- **Una tabella è un insieme di dati correlati tra loro**
- Pensiamo alle tabelle che si possono creare con uno spreadsheet (Excel)
- **Problema:** in una tabella di Excel si può inserire qualunque tipo di dato, anche se non ha senso
 - Es. Nella colonna Età posso inserire 300

Copyright © 2006 Pearson Education. All rights reserved.

Vantaggi del DB

- In una tabella di un DB relazionale si possono esprimere **vincoli sui/tra i dati**
 - Es. i vincoli **limitano il tipo di dati** che possono essere inseriti nel DB ($0 < \text{Età} < 100$)
- I DB rendono possibile eseguire operazioni molto più complesse e potenti
- In un DB è necessario specificare la **struttura di ogni tabella** e il **tipo di dati** contenuti in ogni colonna
 - Consente di verificare che i dati inseriti siano appropriati

Copyright © 2006 Pearson Education. All rights reserved.

Entità

- Ma cosa vogliamo memorizzare nelle tabelle?
Le tabelle di un DB relazionale memorizzano “**entità**”
 - Concetto volutamente generico
- Un’entità è una qualsiasi “cosa” possa essere identificata da un numero prefissato di caratteristiche (**attributi**)
 - Gli attributi hanno un **nome e un valore**
 - I nomi rappresentano le **colonne**
 - I valori rappresentano i **dati memorizzati nella tabella**

Copyright © 2006 Pearson Education. All rights reserved.

Definizione delle tabelle

- Per definire una tabella si specifica:
 - Il **nome generico dell’insieme delle entità** contenute (ovvero il nome della tabella stessa)
 - Gli **attributi** che descriveranno le entità
 - Il **tipo di valore** che si può inserire in ogni attributo (**dominio**)
- Fatto questo la tabella è creata
 - **NOTA:** la tabella inizialmente è **vuota**

Copyright © 2006 Pearson Education. All rights reserved.

Esempi

Balena => entità con diverse caratteristiche






<p>Nome <u>GRIGIA</u> Cibo Crostacei Peso 36.000 kg Genere <i>Eschrichtius</i> Specie <i>robustus</i> Lunghezza 13,7-14 m Figura </p>	<p>Nome <u>ORCA</u> Peso 5.400 kg Genere <i>Orcinus</i> Specie <i>orca</i> Cibo Salmoni, Foche Lunghezza 6,7-8,2 m Figura </p>	<p>Genere <i>Balaenoptera</i> Specie <i>musculus</i> Cibo Krill Nome <u>BLU</u> Lunghezza 23-24,5 m Peso 100.000 Kg Figura </p>
<p>Lunghezza 15-18 m Nome <u>CAPODOGLIO</u> Cibo Calamari Peso 40.000 Kg Genere <i>Physeter</i> Specie <i>macrocephalus</i> Figura </p>	<p>Nome <u>MEGATTERA</u> Cibo Krill Peso 36.000 Kg Genere <i>Megaptera</i> Specie <i>novaeangliae</i> Lunghezza 13-15 m Figura </p>	

Figura 12.3 Alcune entità di tipo balena.

Copyright © 2006 Pearson Education. All rights reserved.

12-7

Tabella Balene

Dominio

Balene		
nome	Character, 15	Nome comune
genere	Character, 15	Prima parte del nome scientifico
specie	Character, 15	Seconda parte del nome scientifico
cibo	Character, 25	Fonte primaria di cibo
peso	Integer	Peso medio di un adulto
lunghezza	Character, 20	Lunghezza tipica di un adulto
figura	GIF	Immagine di un esemplare adulto
Chiave primaria: nome		

Figura 12.4 La specifica della tabella Balene.

NOTA: la specifica è scritta riga per riga, ma ogni attributo indica una colonna della tabella Balene

Copyright © 2006 Pearson Education. All rights reserved.

12-8

Esempi di domini degli attributi

- **Testo:** combinazione di caratteri/numeri con dimensione massima fissata
- **Data:** numero memorizzato con il formato GG/MM/AAAA oppure GG mese AAAA
- **Valuta:** numero che esprime una valuta (cifre decimali fissate)
- **Booleano:** valore che ammette due soli valori
 - Vero e falso
- **Numerico:**
 - Numero intero o reale con limiti minimo e massimo
- **Oggetto:** figura, pagina Web, link ipertestuale

Copyright © 2006 Pearson Education. All rights reserved.

Istanza di tabella Balene

- Inserendo le entità nella tabella vuota si ottiene una **istanza** di tabella
- **Istanza = insieme specifico di righe**
- Se togliessimo la riga di Orca avremmo ***un'istanza differente***

Nome	Genere	Specie	Cibo	Peso	Lunghezza
Grigia	Eschrichtius	robustus	Crostacei	36 mila	13,7-14 m
Orca	Orcinus	orca	Salmone	6 mila	6,7-8,2 m
Blu	Balaenopetra	musculus	Krill	100 mila	23-34.5 m
Capodoglio	Physeter	macrocephalus	Calamari	40 mila	15-18 m
Megattera	Megaptera	novaeangliae	krill	36 mila	13-15 m

Copyright © 2006 Pearson Education. All rights reserved.

Riassumendo...

- Una tabella è inizialmente vuota: ha un nome e intestazioni di colonna (i nomi degli attributi) ma le righe sono vuote
- Ogni riga inserita (detta anche tupla) rappresenta un'entità
- Un'**istanza** (di tabella) è una particolare tabella con un insieme specifico di righe
 - Se togliamo o aggiungiamo righe abbiamo istanze differenti della tabella

Copyright © 2006 Pearson Education. All rights reserved.

12-11

Strutture, contenuti e metadati

- Possiamo parlare di tabelle vuote e istanze perché i DB **separano l'informazione dalla struttura dell'informazione stessa**, ossia i **dati dai metadati**
- I **metadati** forniscono informazioni su come sono organizzati i dati nelle tabelle (la loro struttura)
- I **metadati** di una tabella includono almeno il suo nome, i nomi degli attributi, il dominio di ogni attributo e la *chiave primaria*

Copyright © 2006 Pearson Education. All rights reserved.

12-12

Altre proprietà

- Gli attributi sono elencati nella definizione in un ordine specifico (del resto è impossibile fare altrimenti) ma in generale l'**ordine degli attributi non ha importanza**
- **L'ordine stesso delle righe** è ininfluente: una tabella con le righe ordinate alfabeticamente su una colonna e la stessa tabella con le righe mescolate **sono la stessa istanza di tabella**

Copyright © 2006 Pearson Education. All rights reserved.

Entità e Unicità

- **Unicità delle entità**
 - una tabella di database **non può avere righe duplicate**
 - Se due entità hanno tutti gli attributi uguali, sono a tutti gli effetti la stessa entità
 - alcune caratteristiche possono essere uguali, ma non tutte
 - dev'esserci almeno una caratteristica (o un insieme di esse) distintiva per ogni entità
 - **chiave della tabella**

Copyright © 2006 Pearson Education. All rights reserved.

12-14

Chiavi delle tabelle

- **Chiavi**

- qualsiasi attributo o insieme di attributi per cui le entità in una tabella sono tutte diverse si chiama **chiave candidata** (ES. “Nome” oppure “Genere E Specie”)
- ne scegliamo una e la chiamiamo **chiave primaria**
- **NOTA:** una chiave dev’essere in grado di distinguere tutte le possibili entità, non solo quelle presenti nel particolare stato corrente della tabella
- se nessuna combinazione di attributi si qualifica chiave candidata, occorre assegnare a ogni entità un **codice numerico distinto (ID)**
 - pensiamo ad esempio ai numeri di matricola assegnati agli studenti dall’università

Copyright © 2006 Pearson Education. All rights reserved.

12-15

Un altro semplice esempio

Tabella CANTANTI

Nome	Cognome	Comune di nascita	Data di nascita
Luciano	Ligabue	Correggio	13 marzo 1960
Francesco	Guccini	Modena	14 giugno 1940
Lucio	Dalla	Bologna	4 marzo 1943
Francesco	Baccini	Genova	4 ottobre 1960
Fabrizio	De André	Genova	18 febbraio 1940
Vasco	Rossi	Zocca	7 febbraio 1952

Proviamo a crearla sul DB Access di Microsoft

Copyright © 2006 Pearson Education. All rights reserved.

Proviamo a creare due tabelle

Nome	Cognome	Comune di nascita
Luciano	Ligabue	Correggio
Francesco	Guccini	Modena
Francesco	Baccini	Genova

Comune	Data di nascita
Correggio	13 marzo 1960
Modena	14 giugno 1940
Genova	4 ottobre 1960

Copyright © 2006 Pearson Education. All rights reserved.

Problema

- Immaginiamo di dover costruire la base di dati relativa a solventi chimici con le seguenti informazioni
 - Nome prodotto
 - Categoria del prodotto
 - Formula
 - Peso
 - Viscosità
 - Elenco delle aziende produttrici
- Come strutturare la base di dati?
- Soluzione: Si definisce una “mega” tabella che contiene, per ogni solvente tutte le informazioni
- ***Quali sono gli inconvenienti?***

Copyright © 2006 Pearson Education. All rights reserved.

Soluzione

Nome	Categoria	Formula	Peso atomico	Viscosità	Produttore 1	Produttore 2
benzene	alcohol	C6H6	78,12	0,6028	Coastal Refining & Marketing, Inc	Exxon Chemical Americas
pentane	aliphatic hydrocarbon	C5H12	72,17	0,225	International Commodities Export Co	Phillips 66 Company, Div Phillips Petroleum
hexane	aliphatic hydrocarbon	C6H14	86,2	0,294	Phillips 66 Company, Div Phillips Petroleum	The Humphrey Chemical Co, Inc, Cambrex Fine Chemicals Group
...

- Ogni solvente può avere un numero variabile di produttori (quante colonne mettere?)
- Più solventi appartengono alla stessa categoria (dati ripetuti)

Copyright © 2006 Pearson Education. All rights reserved.

Una soluzione migliore

- “Raggruppiamo l'informazione”
- Creiamo una tabella per gruppi di informazioni fra loro logicamente correlati
 - Un gruppo di informazioni riguarda i solventi e le loro proprietà
 - Un secondo tipo di informazioni riguarda le categorie
 - Un terzo tipo riguarda le aziende produttrici
- In un secondo momento “collegheremo” opportunamente i gruppi (es., solventi e produttori)

Copyright © 2006 Pearson Education. All rights reserved.

Descrizione delle singole entità

Solventi

Nome	Categoria	Formula	Peso atomico	Viscosità
benzene	1	C ₆ H ₆	78,12	0,6028
pentane	2	C ₅ H ₁₂	72,17	0,225
hexane	2	C ₆ H ₁₄	86,2	0,294
...

Categorie

Id categoria	Nome
1	alcohol
2	aliphatic hydrocarbon
3

Copyright © 2006 Pearson Education. All rights reserved.

Descrizione delle singole entità (2)

Produttori

Id produttore	Nome
1	Coastal Refining & Marketing, Inc
2	Exxon Chemical Americas
3	International Commodities Export Co
4	Phillips 66 Company, Div Phillips Petroleum
5	The Humphrey Chemical Co, Inc, Cambrex Fine Chemicals Group
6	...

Copyright © 2006 Pearson Education. All rights reserved.

Collegare l'informazione

- Colleghiamo i solventi ai produttori con una tabella a due colonne

Id produttore	Solvente
1	benzene
2	benzene
3	pentane
4	pentane
4	exane
5	exane
6	...

Copyright © 2006 Pearson Education. All rights reserved.

Operazioni sulle tabelle

- Un DB è una collezione di tabelle il cui uso principale è la ricerca di informazioni
 - Gli utenti interrogano il database attraverso **query** e il database recupera e presenta le informazioni che soddisfano la richiesta
- Le query vengono poste:
 - Con **interfaccia grafica**, se l'ambiente di uso (il cosiddetto client) la prevede
 - Con **interfaccia testuale**: linguaggio di interrogazione SQL (Structured Query Language)

Copyright © 2006 Pearson Education. All rights reserved.

12-24

Tabelle come risultato di operazioni

- Le operazioni sono applicate alle tabelle **per ottenere altre tabelle**
 - ES Seleziona dalla tabella dei CANTANTI quelli che hanno nome Francesco
- Se la risposta è univoca l'istanza di tabella risultante avrà una sola riga, ma in generale non sarà così
- Può anche essere che non ci sia alcuna risposta → tabella vuota come risultato

Copyright © 2006 Pearson Education. All rights reserved.

Operazioni sulle tabelle

Le operazioni fondamentali applicabili alle tabelle sono le seguenti:

- **selezione**
- **proiezione**
- **unione**
- **differenza**
- **intersezione**
- **prodotto**



Join: in realtà ottenibile come combinazione delle altre, ma è così potente e comodo che solitamente viene fornito come operatore separato

Copyright © 2006 Pearson Education. All rights reserved.

Operazione di selezione

- Prende alcune righe di una tabella per crearne un'altra (tabella risultato)
 - Si deve specificare la tabella da cui prelevare le righe e il *test* di selezione
 - *Sintassi: SELECT * FROM Tabella WHERE test;*
 - Il *Test* è applicato a ogni riga per decidere se includerla o no nella tabella risultato
 - Il *Test* usa nomi degli attributi, costanti numeriche o testuali e operatori relazionali <, <=, =, !=, >, >=
 - Se il *Test* è verificato, la riga è aggiunta nella tabella risultato; altrimenti è ignorata
 - Il *Test* può fare uso di operatori booleani

Copyright © 2006 Pearson Education. All rights reserved.

Esempi

- `SELECT * FROM Cantanti WHERE Nome="Francesco";`
 - Restituisce:

Francesco	Guccini	Modena	14 giugno 1940
Francesco	Baccini	Genova	4 ottobre 1960
- `SELECT * FROM Cantanti WHERE Nome="Francesco" AND Data di nascita >="1 gennaio 1950";` → **Sfrutta operatori booleani**
 - Restituisce:

Francesco	Baccini	Genova	4 ottobre 1960
-----------	---------	--------	----------------

NOTA: Il risultato è costruito copiando entità, la tabella originale non viene minimamente toccata

Copyright © 2006 Pearson Education. All rights reserved.

Operazione di proiezione

- Prende alcune colonne di una tabella esistente e ne crea un'altra (tabella risultato)
 - Si deve specificare il nome della tabella esistente e le colonne (il nome dei campi) da includere nella nuova tabella risultato
 - *Sintassi:*
 - `SELECT Lista_di_campi From Tabella`
 - La nuova tabella avrà il numero di colonne specificato e lo **stesso numero di righe di quella originale, a meno che...**
 - ... la nuova tabella elimini un campo chiave: in questo caso, eventuali righe duplicate nella nuova tabella saranno eliminate*

Copyright © 2006 Pearson Education. All rights reserved.

Esempi

- `SELECT Nome,Cognome FROM Cantanti;`
 - Restituisce:

Luciano	Ligabue
Francesco	Guccini
Lucio	Dalla
Francesco	Baccini
Fabrizio	De André
Vasco	Rossi
- `SELECT * FROM Cantanti;`
 - Restituisce l'intera tabella

Ora è chiaro l'asterisco visto nella operazione di selezione

Copyright © 2006 Pearson Education. All rights reserved.

Selezione + proiezione

- `SELECT Cognome FROM Cantanti WHERE Nome="Francesco";`
 - Restituisce:
 - Guccini
 - Baccini

- `SELECT Cognome, Comune di nascita FROM Cantanti WHERE Comune di nascita="Genova";`
 - Restituisce:
 - Baccini Genova
 - De André Genova

Copyright © 2006 Pearson Education. All rights reserved.

Operazioni di unione, differenza e intersezione

- L'operazione di unione ($Ris = Tab1 + Tab2$)
 - Combina due tabelle (**che hanno gli stessi attributi**) considerando le righe di entrambe
 - Elimina eventuali duplicati

- L'operazione di differenza ($Ris = Tab1 - Tab2$)
 - Si effettua tra tabelle **che hanno gli stessi attributi**
 - Rimuove dalla prima tabella tutte le righe contenute anche nella seconda tabella

- L'operazione di intersezione ($Ris = Tab1 \cap Tab2$)
 - Combina tabelle **che hanno gli stessi attributi**
 - Considera solo le righe comuni, cioè presenti in entrambe le tabelle

Copyright © 2006 Pearson Education. All rights reserved.

12-32

Operazione di prodotto

- L'operazione di prodotto ($Ris = Tab1 \times Tab2$)
 - Si effettua **tra tabelle con attributi diversi**
 - Crea una **super-tabella** che ha tutti gli attributi di entrambe le tabelle originali
 - Combina tutte le righe → ogni riga della prima tabella viene concatenata con ciascuna riga della seconda tabella
- Pensiamo alla definizione di prodotto cartesiano
 - Dati due insiemi A e B, il loro prodotto cartesiano $A \times B$ è l'insieme di tutte le possibili coppie ordinate $(x; y)$, dove il primo elemento x appartiene ad A, mentre il secondo elemento y appartiene a B

Copyright © 2006 Pearson Education. All rights reserved.

Esempio di prodotto

- Consideriamo le due tabelle (diciamo c1 e c2)

Nome	Cognome	Comune di nascita	Comune	Data di nascita
Luciano	Ligabue	Correggio	Correggio	13 marzo 1960
Francesco	Guccini	Modena	Modena	14 giugno 1940
Francesco	Baccini	Genova	Genova	4 ottobre 1960

- Il prodotto $c1 \times c2$ produrrebbe la seguente tabella

Nome	Cognome	Comune di nascita	Comune	Data di nascita
Luciano	Ligabue	Correggio	Correggio	13 marzo 1960
Luciano	Ligabue	Correggio	Modena	14 giugno 1940
Luciano	Ligabue	Correggio	Genova	4 ottobre 1960
Francesco	Guccini	Modena	Correggio	13 marzo 1960
Francesco	Guccini	Modena	Modena	14 giugno 1940
Francesco	Guccini	Modena	Genova	4 ottobre 1960
Francesco	Baccini	Genova	Correggio	13 marzo 1960
...

Copyright © 2006 Pearson Education. All rights reserved.

Utilità dell'operazione prodotto

- Operazione apparentemente alquanto bizzarra, poiché creando tutte le combinazioni mette insieme righe che non c'entrano
- Nella maggior parte dei casi è usata per creare una grande tabella che poi viene sfoltita con altre operazioni di selezione, proiezione e differenza
→ si tiene alla fine solo l'informazione significativa

Copyright © 2006 Pearson Education. All rights reserved.

Esempio

Amico	Nazionalità
Isabella	Argentina
Brian	Sudafrica
Wen	Cina
Clare	Canada

**Tabella
Amici
(4 righe)**

Nome	Capitale	Lat	NS	Lon	EW	Interesse
Irlanda	Dublino	52	N	7	W	Storico
Israele	Gerusalemme	32	N	35	E	Storico
Italia	Roma	42	N	12	E	Arte
Giamaica	Kingston	18	N	77	W	Spiagge
...

**Tabella
Nazioni
(230
righe)**

Copyright © 2006 Pearson Education. All rights reserved.

Ipotesi

- Alcuni degli amici si sono offerti volontari per fare da tutor a studenti che si stanno preparando alle Olimpiadi di Geografia
- Ogni amico ha acconsentito ad aiutare per quanto riguarda la propria “parte” del mondo, cioè il quarto di pianeta da dove proviene
 - Isabella → Argentina → SW
 - Brian → Sudafrica → SE
 - Wen → Cina → NE
 - Clare → Canada → NW
- Dovete creare la **tabella Master** dei responsabili per ogni nazione



Nome	Amico
Cile	Isabella
Giappone	Wen
...	...

Copyright © 2006 Pearson Education. All rights reserved.

Creazione tabella Master

Super = Nazioni x Amici

Assign = (SELECT * FROM Super WHERE NS='S'
AND EW='W' AND Amico='Isabella')

+ (SELECT * FROM Super WHERE NS='S' AND
EW='E' AND Amico='Brian')

+ (SELECT * FROM Super WHERE NS='N' AND
EW='E' AND Amico='Wen')

+ (SELECT * FROM Super WHERE NS='N' AND
EW='W' AND Amico='Clare')

Master= SELECT Nome, Amico FROM Assign

Copyright © 2006 Pearson Education. All rights reserved.

Tabella Super

- Super è il prodotto di Nazioni e Amici
 - 9 colonne (7 di Nazioni e 2 di Amici)
 - 920 righe
 - Per ogni paese, c'è una riga per ciascun amico

Nome	Capitale	Lat	NS	Lon	EW	Interesse	Amico	Nazionalità
Irlanda	Dublino	52	N	7	W	Storico	Isabella	Argentina
Irlanda	Dublino	52	N	7	W	Storico	Brian	Sudafrica
...
Italia	Roma	42	N	12	E	Arte	Isabella	Argentina
Italia	Roma	42	N	12	E	Arte	Brian	Sudafrica
...

Copyright © 2006 Pearson Education. All rights reserved.

Creazione tabella Master

- Assign opera un'unione su 4 tabelle, ognuna ottenuta da selezione su Super
 - La prima select conserva solo le nazioni che hanno il nome di Isabella e si trovano nell'emisfero sud-occidentale SW, la seconda solo le nazioni che hanno il nome di Brian e si trovano nell'emisfero sud-orientale SE, etc ...
 - La tabella risultante ha 230 righe – una per ogni nazione con assegnato il nome di un amico
- Master fa una proiezione su assign tenendo solo il nome della nazione e dell'amico

Nome	Amico
Cile	Isabella
Giappone	Wen
...	...

Copyright © 2006 Pearson Education. All rights reserved.

L'operazione di join

- Operazione molto potente
- Combina due tabelle, come il prodotto, **ma non produce necessariamente tutte le combinazioni di righe**
 - le due tabelle devono avere un attributo che contiene lo stesso tipo di dato
 - la nuova tabella concatena solo le righe in cui c'è una **corrispondenza** su quell'attributo
 - sintassi (una delle molte possibili):

```
SELECT Attributi FROM Tabella1 INNER JOIN
Tabella2 ON Corrispondenza;
```

Copyright © 2006 Pearson Education. All rights reserved.

12-41

L'operazione di join (cont.)

- **Corrispondenza** rappresenta il test di uguaglianza che riguarda l'attributo dello stesso tipo presente in entrambe le tabelle (*attributo su cui si fa il join*)
- Se l'attributo contengono lo stesso identico dato in entrambe le tabelle, la concatenazione delle due righe corrispondenti viene aggiunta nella tabella risultato

NOTA: Per riferire attributi di tabelle diverse, si usa la notazione Tabella.attributo

Copyright © 2006 Pearson Education. All rights reserved.

12-42

Esempio

- Consideriamo le due tabelle (c1 e c2)

Nome	Cognome	Comune di nascita
Luciano	Ligabue	Correggio
Francesco	Guccini	Modena
Francesco	Baccini	Genova

Comune	Data di nascita
Correggio	13 marzo 1960
Modena	14 giugno 1940
Genova	4 ottobre 1960

- `Select c1.Nome,c1.Cognome,c1.Comune di nascita, c2.Data di nascita FROM c1 INNER JOIN c2 ON c1.Comune di nascita = c2.Comune`

Copyright © 2006 Pearson Education. All rights reserved.

Risultato

- `Select c1.Nome,c1.Cognome,c1.Comune di nascita, c2.Data di nascita FROM c1 INNER JOIN c2 ON c1.Comune di nascita = c2.Comune`
- Produrrebbe

Nome	Cognome	Comune di nascita	Data di nascita
Luciano	Ligabue	Correggio	13 marzo 1960
Francesco	Guccini	Modena	14 giugno 1940
Francesco	Baccini	Genova	4 ottobre 1960

Copyright © 2006 Pearson Education. All rights reserved.