

## Esercizio 2 della parte prima – traccia di soluzione

Per distinguere la risposta dal testo della domanda, rappresentiamo quest'ultimo con caratteri piccoli.

In questa prima settimana state imparando i concetti fondamentali delle Basi di dati, concetti che poi approfondiremo durante l'anno. I due concetti più importanti sono quelli di **schema** e **istanza**. Nella tabella seguente

Professori

Id-Professore	Cognome	Nazionalità
34	Xu	Cina
45	Smith	USA
78	Rossi	Italia

lo schema è composto dalla parte seguente

Professori

Id-Professore	Cognome	NAZIONALITÀ
---------------	---------	-------------

e la istanza da

34	Xu	Cina
45	Smith	USA
78	Rossi	Italia

Diremo nel seguito che la istanza rappresenta elementi **osservabili** della realtà (observables), mentre lo schema rappresenta **classi di osservabili**. Ad esempio nella base di dati sopra, gli osservabili sono 3 professori, e lo schema è formato da una relazione con nome Professori e tre attributi. La relazione Professori rappresenta una classe di 3 osservabili.

**Adesso vi propongo di progettare una base di dati per la sola parte costituita dallo schema, senza istanza.** E' quello che accade spesso quando si progetta una base di dati: si progetta lo schema e poi si popola la base di dati con gli osservabili, la cui classe corrisponde allo schema.

Esercizio 2 della parte introduzione - Lo schema da progettare rappresenta la diffusione del coronavirus in Lombardia, a partire dal primo giorno in cui si è manifestato un caso di positività. Per ogni giorno e per ogni provincia della Lombardia (Lodi, Milano, ecc.) vogliamo rappresentare il codice e nome della provincia, il numero di nuovi positivi rispetto al giorno precedente, il numero totale di positivi, il numero di ricoverati in quel giorno.

Progettate lo schema, che chiameremo **Schema1**, formato da una sola relazione, cercando di dare nomi comprensibili agli attributi, non vi preoccupate se sono lunghi, l'importante è che non lo siano troppo, trovate un punto di equilibrio.

Traccia di risposta – evidenziamo anzitutto in **bold** le parole che ci pare corrispondano a attributi, cioè alle colonne della tabella. Ricordiamo che in una relazione le "colonne" sono quelle a cui sono associati valori.

Lo schema da progettare rappresenta la diffusione del coronavirus in Lombardia, a partire dal primo giorno in cui si è manifestato un caso di positività. Per ogni **giorno** e per ogni **provincia della Lombardia** (Lodi, Milano, ecc.) vogliamo rappresentare il **codice e nome della provincia**, il **numero di nuovi positivi** rispetto al giorno precedente, il **numero totale di positivi**, il **numero di ricoverati in quel giorno**. Quindi lo schema della relazione può essere

Giorno	Cod.Provincia	NomeProvincia	Nuovi positivi	Totale positivi	Ricoverati
--------	---------------	---------------	----------------	-----------------	------------

Osserviamo lo schema. C'è qualcosa su cui abbiamo dubbi? Forse il concetto di giorno...

Come esprimiamo i giorni? Ci sono almeno due possibilità

1. Con un unico valore, come ad esempio 13/2 oppure 13/febbraio.
2. Con due valori:

13	febbraio
----	----------

Quale delle due soluzioni scegliamo? In questi casi il criterio guida è la chiarezza, la comprensibilità. Diciamo che è un po' più chiara la soluzione 2, ma anche la 1 andrebbe bene.... Quindi lo schema diventa

Giorno del mese	Mese	Cod Provincia	Nome Provincia	Nuovi positivi	Totale positivi	Ricoverati
-----------------	------	---------------	----------------	----------------	-----------------	------------

Ci siamo scordati qualcosa? Certo! Il nome della relazione.

Diffusione del coronavirus in Lombardia

Giorno del mese	Mese	Cod Provincia	Nome Provincia	Nuovi positivi	Totale positivi	Ricoverati
-----------------	------	---------------	----------------	----------------	-----------------	------------

Oppure

Coronavirus in Lombardia

Giorno del mese	Mese	Cod Provincia	Nome Provincia	Nuovi positivi	Totale positivi	Ricoverati
-----------------	------	---------------	----------------	----------------	-----------------	------------

Siamo soddisfatti? Beh, abbiamo scelto nomi degli attributi comprensibili e non troppo lunghi, va bene così. Magari Cod potrebbe diventare Codice, ma lasciamo così

Seconda domanda - Adesso provate ad "affettare" la relazione in tre relazioni, rappresentando nella prima l'evoluzione dei nuovi positivi rispetto al giorno precedente, nella seconda il numero totale di positivi, nella terza il numero di ricoverati in quel giorno. Chiamiamo lo schema **Schema2**. Attenzione! Il nuovo schema deve avere **contenuto informativo** complessivo identico al precedente. Cosa intendo per contenuto informativo? Devo affidarmi alla intuizione, perchè nonostante il fatto che le basi di dati esistano da 60 anni, non è ancora chiaro cosa si intenda per contenuto informativo di una base di dati. Non vi meravigliate, l'informatica è una scienza relativamente giovane... Diciamo che deve valere la seguente proprietà: partendo dal nuovo schema, e da una sua istanza, deve sempre essere possibile ricostruire il precedente schema costituito da una sola relazione, e la sua istanza. Notate che in questa definizione operativa di avere lo stesso contenuto informativo ho dovuto reintrodurre le istanze.

Scegliamo prima i nomi degli attributi, poi penseremo ai nomi delle relazioni

Schema2

Relazione 1

Giorno del mese	Mese	Cod Provincia	Nome Provincia	Nuovi positivi
-----------------	------	---------------	----------------	----------------

Relazione 2

Giorno del mese	Mese	Cod Provincia	Nome Provincia	Totale positivi
-----------------	------	---------------	----------------	-----------------

Relazione 3

Giorno del mese	Mese	Cod Provincia	Nome Provincia	Ricoverati
-----------------	------	---------------	----------------	------------

Ora scegliamo i nomi delle relazioni. La nostra fantasia non è infinita.... Possibili nomi sono per le tre relazioni

1. Nuovi positivi del coronavirus in Lombardia
2. Totale positivi del coronavirus in Lombardia
3. Ricoverati per coronavirus in Lombardia

Attenzione, devo condividere con voi una scelta molto importante: in tutte e tre le relazioni abbiamo riportato gli attributi giorno del mese, mese e provincia; perché? Voi come avevate scelto gli attributi nelle relazioni?

E' facile convincerci che abbiamo riprodotto i tre attributi nelle tre relazioni perche' altrimenti non sapremmo come ricomporre lo schema di partenza. Se non siete convinti/e, pensiamo a quest'altra soluzione

Relazione 1

Nuovi positivi
----------------

Relazione 2

Giorno del mese	Mese	Nome Provincia	Totale positivi
-----------------	------	----------------	-----------------

Relazione 3

Cod. Provincia	Ricoverati
----------------	------------

Come diavolo facciamo a riprodurre la relazione di partenza se non sappiamo, ad esempio, i nuovi positivi a quale giorno del mese, mese e provincia si riferiscono?

Osserviamo ancora un attimo lo schema. Rappresenta informazioni rilevanti per descrivere il fenomeno, ma non ne rappresenta altre, per esempio i guariti. Ricordiamoci sempre che uno schema relazionale è una rappresentazione del mondo, e come tale è sempre estremamente imperfetta!

Ora dobbiamo affrontare la domanda sul contenuto informativo. Si può ricostruire la relazione di partenza? Certo, è un po' difficile rispondere alla domanda senza rappresentare una possibile istanza, cosa che potete fare voi.

Ma proviamo a ragionare, come si dice, intensionalmente, cioè ragionando solo sullo schema. Una volta definito un giorno e una provincia, vi chiedo: possono essere associati ad essi due valori diversi, ad esempio, di Ricoverati? Certamente no. Quindi, siccome giorno e provincia sono rappresentati in tutti e tre le relazioni, ricomponendo le tre relazioni si avrà una sola riga per un certo giorno e provincia. Il contenuto informativo è lo stesso di prima, la istanza sarà la stessa di prima.

Seconda domanda - Ora torniamo allo Schema1. Supponete di voler aggiungere i comuni, con codici e nomi, In altre parole, ora vogliamo esprimere il numero dei nuovi positivi ecc. per ogni comune di ogni provincia della Lombardia. Progettate questo nuovo schema, che chiamiamo **Schema3**. Ho una serie di domande, **cercate di rispondere a una domanda per volta, prima di passare alla successiva**.

1. Ricordando le conclusioni del primo esercizio che vi ho proposto ieri l'altro, quale è il numero **giusto** di relazioni di questo schema?

Eh dipende da cosa si intende per giusto! Se però torniamo alle presentazioni power point con voce della prima parte, e alla soluzione dell'esercizio 1, *giusto* significa che non ci sono dati rappresentati più volte, e che gli aggiornamenti si fanno in una sola riga di una sola tabella. Pensando a come si deve modificare lo schema 1, cominciamo a intuire che se mettiamo Comune e Provincia insieme in una relazione (con i loro codici), beh allora un certo comune e una certa provincia sono rappresentati tante volte quanti sono i giorni! Quindi dobbiamo tenere separati comune e provincia, rappresentando nella relazione di partenza solo il comune e non la provincia. Quindi le relazioni sono 2.

2. Come è fatto lo Schema3?

Così

### Coronavirus in Lombardia

Giorno del mese	Mese	Cod Comune	Nome comune	Nuovi positivi	Totale positivi	Ricoverati
-----------------	------	------------	-------------	----------------	-----------------	------------

### Comuni e Province

Cod Comune	Nome comune	Cod Provincia	Nome Provincia
------------	-------------	---------------	----------------

Coloro tra di voi che hanno fatto l'istituto tecnico con ramo informatica o liceo con informatica forse stanno "storcendo la bocca", e forse stanno dicendo: eh, ma nella seconda relazione, ad esempio, il nome provincia compare tante volte quanti sono i comuni. E' vero, avete ragione. Bisognerebbe creare altre due relazioni con i soli attributi di comune e di provincia; ma non facciamo le cose troppo complicate, se no vi stufate di questa materia, e avreste ragione. Va bene anche così.

Supponete ora che lo Schema3 sia composto da due relazioni, nella prima delle quali rappresentate comuni e province, e nella seconda comuni, giorni e dati sulla epidemia. Chiamiamo **Schema32** il nuovo schema.

3. Come è fatto lo **Schema32**?

Mi sono accorto che questa domanda coincideva in pratica con la seconda, scusate.... Ho già risposto poco fa.

4. Intuitivamente, il contenuto informativo dello Schema3 è uguale, inferiore, o superiore a quello dello Schema1? Fate appello alla intuizione....

Beh, qui abbiamo rappresentato la stessa informazione di prima, e molta di più perchè ora i dati sulla diffusione sono per ogni comune; quindi il contenuto informativo è maggiore.

5. Supponete che intendiamo inserire nello **Schema1** i valori relativi a dieci giorni; quante saranno le righe della istanza dell'unica relazione dello Schema1? Se non avete a mente alcuni dati del problema, cercate su Google...

Questa domanda è interessante perché vi obbliga a cercare una informazione sul Web. Basta impostare le parole chiave su Google "province lombardia" per vedere che sono 11, quindi le righe dello schema 1 sono  $10 \times 11 = 110$ .

6. Supponete che intendiamo inserire nello **Schema32** i valori relativi a dieci giorni; quante saranno le righe delle istanze delle due relazioni dello Schema32? Se non avete a mente alcuni dati del problema, cercate su Google...

Ora dobbiamo capire quanti sono i comuni della lombardia, e vediamo che sono 1507, quindi la relazione Coronavirus in Lombardia ha  $1507 \times 10 = 15070$  righe e la relazione Comuni e Province ha 1507 righe. Tra l'altro provate ad accedere all'indirizzo

[www.comuniverso.it/index.cfm?Comuni\\_della\\_Regione\\_Lombardia&menu=203](http://www.comuniverso.it/index.cfm?Comuni_della_Regione_Lombardia&menu=203),

notate qualcosa di trano nella relazione con i 1507 comuni? Provate a pensarci, domani la soluzione.

Abbiamo finito. Ho cercato di simulare il modo in cui avete ragionato o avreste dovuto ragionare per risolvere questo secondo esercizio, se volete fare la stessa operazione con voi stessi.....