

Progetto Basi di Dati

Paolo Addis
Samuele Poz
Tristano Munini

RENDERE CARINO

Indice

1	Progettazione Concettuale	1
1.1	Raccolta ed Analisi dei Requisiti	1
1.2	Progettazione del Modello E-R	2
1.2.1	Le Entità	2
1.2.2	Le Relazioni	3
2	Progettazione Logica	6
2.1	Ristrutturazione del Modello E-R	6
2.1.1	Semplificazione dei Concetti	6
2.2	Analisi delle Ridondanze	6
2.3	Traduzione da ER a Relazionale	9
3	Progettazione Fisica	10
3.1	Nuovi Indici	10
4	Definizione della Base di Dati in SQL	11
4.1	Definizione delle Tabelle	11
4.2	Definizione di Query Significative	11
5	Analisi Dati	12
5.1	Popolamento della Base di Dati	12
5.2	Analisi Dati in R	12
6	Portale Web	13
6.1	Interfaccia Grafica	13
6.2	Comunicare con pgAdmin	13

1 Progettazione Concettuale

1.1 Raccolta ed Analisi dei Requisiti

Si vuole modellare il seguente insieme di informazioni riguardanti un sistema per la gestione delle diagnosi e delle terapie dei pazienti ricoverati in un dato ospedale.

- Di ogni ricovero, il sistema deve memorizzare il codice univoco, il nome della divisione ospedaliera (Cardiologia, Reumatologia, Ortopedia, ...), il paziente ricoverato, le date di inizio e fine del ricovero e il motivo principale del ricovero.
- Di ogni paziente, il sistema deve memorizzare il codice sanitario (univoco), il cognome, il nome, la data di nascita, il luogo di nascita e la provincia di residenza. Per i pazienti residenti fuori regione, vengono memorizzati anche il nome della ULSS e la regione di appartenenza.
- Di ogni diagnosi effettuata durante il ricovero del paziente, sono memorizzati la patologia diagnosticata, col suo codice ICD10 (classificazione internazionale delle patologie) e l'indicazione della sua gravità (grave: si/no), la data e il nome e cognome del medico che ha effettuato la diagnosi.
- Nella base di dati si tiene traccia delle terapie prescritte ai pazienti durante il ricovero. Di ogni terapia, si memorizzano il farmaco prescritto, la dose giornaliera, le date di inizio e di fine della prescrizione, la modalità di somministrazione ed il medico che ha prescritto la terapia.
- Di ogni farmaco sono memorizzati il nome commerciale (univoco), l'azienda produttrice, il nome e la quantità dei principi attivi contenuti e la dose giornaliera raccomandata.
- Si tiene, infine, traccia delle diagnosi che hanno motivato le terapie. In particolare, ogni terapia è prescritta al fine di curare una o più patologie diagnosticate. Può capitare anche che una nuova patologia (registrata come nuova diagnosi) sia causata, come effetto collaterale, da una terapia precedentemente prescritta. Tale legame causa-effetto va registrato nella base di dati.

Tramite un portale web dovrà essere possibile accedere a tale base di dati ed aggiungere o rimuovere pazienti, ricoveri, diagnosi con relative terapie e farmaci.

1.2 Progettazione del Modello E-R

Inizialmente l'attenzione è stata concentrata sulle entità PAZIENTE e RICOVERO in quanto ritenute fondamentali: si può dire che in assenza di un'entità che rappresenti i pazienti, l'intera base di dati perderebbe significato. Infatti per ogni ricovero o diagnosi deve essere possibile risalire al soggetto che è stato ricoverato o a cui è stata effettuata una diagnosi, altrimenti non si potrebbero utilizzare correttamente le informazioni raccolte in passato. RICOVERO, invece, segna i momenti in cui il paziente raggiunge e lascia la struttura ospedaliera, inoltre permette di indicare in modo univoco le visite e le cure alle quali è stato sottoposto durante questo periodo. Le due entità sono in relazione tra di loro attraverso una VIENE-RICOVERATO che intuitivamente collega una entità PAZIENTE a molte entità RICOVERO. È stata poi costruita la tripla FARMACO, SOMMINISTRATO-DURANTE e TERAPIA composta rispettivamente da: un'entità che rappresenta il farmaco con i relativi attributi; una relazione che permette di indicare che farmaco viene usato durante una terapia; infine un'entità che raccoglie i dettagli della terapia, ad esempio la dose giornaliera di farmaco da assumere. L'entità DIAGNOSI e le rimanenti relazioni, in particolare CAUSA-EFFETTO, verranno analizzate successivamente.

1.2.1 Le Entità

- PAZIENTE: dai requisiti è possibile individuare le seguenti voci significative: *codice sanitario, cognome, nome, data di nascita, luogo di nascita e provincia di residenza* del paziente. Per ognuna di esse è stato creato un attributo. Il codice fiscale (d'ora in poi detto anche CF) è stato selezionato come chiave primaria in quanto, come definito nei requisiti, univoco¹. Per salvare gli attributi esclusivi per i pazienti fuori regione riguardanti il nome della ULSS e la regione di appartenenza, è stata creata l'entità PAZIENTE-FUORI-REGIONE, specializzazione dell'entità precedente. Poiché solo una frazione dell'insieme dei pazienti sarà di questo tipo la specializzazione è parziale.
- RICOVERO: dai requisiti è possibile individuare le seguenti voci significative: *codice ricovero, divisione ospedaliera, paziente ricoverato, data inizio, data fine, motivo* del ricovero. Per ognuna di esse è stato creato un attributo, fatta eccezione per *paziente ricoverato*. Quest'ultima informazione verrà memorizzata attraverso la relazione VIENE-RICOVERATO, descritta in seguito. Il *codice ricovero* è stato selezionato come chiave primaria dell'entità in quanto univoco.
- DIAGNOSI: dai requisiti è possibile individuare le seguenti voci significative: *patologia diagnosticata, codice patologia, gravità patologia, data, medico*. Per ognuna di esse è stato creato un attributo. In particolare

¹Proprietà fondamentale di un attributo per la definizione di esso come chiave primaria.

patologia è un attributo composto che si suddivide in *codice* e *gravità*. Si assume che la data comprenda anche l'ora in cui è stata effettuata la diagnosi. DIAGNOSI è un'entità debole in quanto la chiave primaria è definita attraverso l'attributo *data* e la relazione EFFETTUATA-DURANTE. La chiave è stata stabilita sfruttando l'assunzione precedente riguardo la data in modo tale che per ogni ricovero non sia possibile associare più diagnosi effettuate nello stesso istante.

- FARMACO: dai requisiti è possibile individuare le seguenti voci significative: *nome commerciale*, *azienda produttrice*, *nome principi attivi*, *quantità principi attivi*, *dose giornaliera raccomandata*. Per ognuna di esse è stato creato un attributo. In particolare *nome principi attivi* e *quantità principi attivi* sono attributi multivalore e la chiave primaria è rappresentata da *nome commerciale* in quanto univoco.
- TERAPIA: dai requisiti è possibile individuare le seguenti voci significative: *farmaco prescritto*, *dose giornaliera*, *modalità di somministrazione*, *medico prescrivente*. Sono stati definiti gli attributi *dose giornaliera* e *modalità somministrazione* mentre l'informazione relativa alle altre voci verrà considerata in altre entità e relazioni. In particolare il medico prescrivente, le date di inizio e fine della terapia verranno gestite nella relazione MOTIVA-TERAPIA, in quanto la terapia è intesa come metodo per la somministrazione di un farmaco al fine di curare una patologia. La chiave primaria è composta dagli attributi di terapia e dalla relazione SOMMINISTRATO-DURANTE che rende l'entità un'entità debole e consente di mantenere i dati relativi al farmaco prescritto.

1.2.2 Le Relazioni

- VIENE-RICOVERATO: un tipo di relazione Uno-A-Molti tra PAZIENTE e RICOVERO. Entrambe le partecipazioni sono totali.
- EFFETTUATA-DURANTE: un tipo di relazione Uno-A-Molti tra DIAGNOSI e RICOVERO. La partecipazione di RICOVERO è parziale mentre quella di DIAGNOSI è totale.
- SOMMINISTRATO-DURANTE: un tipo di relazione Uno-A-Molti tra FARMACO e TERAPIA. La partecipazione di FARMACO è parziale mentre quella di TERAPIA è totale.
- MOTIVA-TERAPIA: una relazione ternaria in cui partecipano le entità TERAPIA e DIAGNOSI in una meccanica di causa-effetto. Possiede i seguenti attributi *medico prescrivente*, *data inizio*, *data fine*. L'entità DIAGNOSI partecipa con cardinalità (0,1) mentre TERAPIA partecipa con cardinalità (1,N).

Nella costruzione della relazione MOTIVA-TERAPIA si è fatta particolare attenzione all'ultimo punto dell'analisi dei requisiti in cui si specifica che una terapia può essere associata a due tipi di diagnosi: quelle che motivano la prescrizione della terapia e quelle che registrano una nuova patologia presentatasi come effetto collaterale dovuto alla terapia stessa. Per creare il modello il più attinente possibile ai requisiti si è pensato di

Ricordando che la terapia descrive soltanto il metodo da applicare per curare una patologia e non le date di inizio e fine ed del medico prescrivente. Le informazioni riguardo la durata della terapia (dunque la data di inizio e quella di fine) e il medico prescrivente sono relative alla coppia DIAGNOSI,TERAPIA. Per questo motivo non può essere che l'entità TERAPIA, in quanto rappresentate di un modello astratto, abbia tali attributi. Per assurdo se così fosse applicando un'astessa terapia a due diagnosi distinte bisognerebbe aggiungere alla chiave primaria le date contraddicendo la definizione data.

La progettazione di tale relazione è stata suddivisa in due fasi:

La prima, illustrata in *figura 1*, riguarda la relazione che intercorre tra diagnosi e terapia come cura.

La seconda registra il legame di causa-effetto tra la terapia applicata e la nuova patologia diagnosticata.

Osservando la *figura 1* si può notare che la relazione illustrata tiene traccia delle diagnosi che hanno motivato le terapie.

Ad esempio: Nel caso in cui la patologia diagnosticata sia una infezione batterica alle vie aeree, la terapia prescritta potrebbe essere "farmaco antibiotico, da assumere 2 volte al giorno per via orale".

In particolare la terapia è prescritta al fine di curare almeno una o più patologie diagnosticate

Il legame di causa descritto nel testo può essere modellato attraverso relazione...

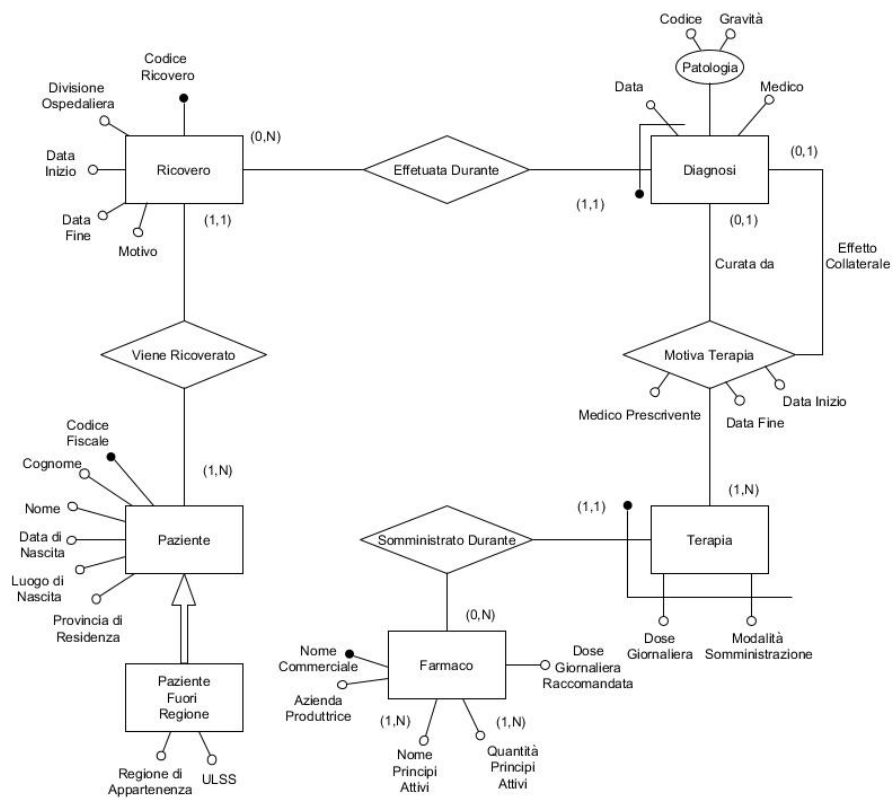


Figura 1: Lo schema Entità-Relazioni completo

2 Progettazione Logica

2.1 Ristrutturazione del Modello E-R

2.1.1 Semplificazione dei Concetti

È utile semplificare strutture come specializzazioni, attributi composti e attributi multivalore perché renderà più facile la traduzione del modello Entità-Relazioni in quello Relazionale.

Ricordando che quest'ultimo non permette di rappresentare generalizzazioni sarà necessario rimuovere tutte le specializzazioni dal nostro schema E-R. Si può notare che la specializzazione PAZIENTE-FUORI-REGIONE può essere facilmente rappresentata aggiungendo due attributi all'entità PAZIENTE. I nuovi attributi, *ULSS* e *Regione di Appartenenza*, in PAZIENTE saranno facoltativi poiché la specializzazione è parziale. La rimozione delle generalizzazioni non è sempre così immediata: se PAZIENTE-FUORI-REGIONE avesse avuto una relazione in più rispetto al genitore si sarebbero dovute valutare varie opzioni. Per esempio la creazione di una nuova relazione per simboleggiare il legame tra genitore e figlio, lasciando invariati gli attributi. Oppure la duplicazione nel figlio di tutti gli attributi e di tutte le relazioni del genitore.

Anche nel caso degli attributi composti la semplificazione è immediata, infatti nello schema l'unico attributo composto è *Patologia* in DIAGNOSI. Può essere trasformato in due nuovi attributi *Codice Patologia* e *Gravità Patologia* che contengono esattamente la stessa informazione.

Bisogna prestare più attenzione agli attributi multivalore: in FARMACO sono presenti *Nome Principi Attivi* e *Quantità Principi Attivi* entrambi obbligatori e multivalore. L'unica semplificazione attuabile è creare una nuova entità chiamata PRINCIPIO-ATTIVO e metterla in relazione Molti-A-Molti con FARMACO attraverso la nuova relazione CONTIENE. PRINCIPIO-ATTIVO avrà un solo attributo, scelto anche come chiave primaria, chiamato *Nome*, mentre alla relazione CONTIENE verrà aggiunto l'attributo *Quantità*. In questo modo saranno modellate esattamente le stesse informazioni precedentemente modellate dagli attributi multivalore. Rimozione e scelta delle chiavi primarie

2.2 Analisi delle Ridondanze

Consideriamo il ciclo PAZIENTE, RICOVERO, DIAGNOSI dato dalle relazioni VIENE RICOVERATO, EFFETTUATA DURANTE, EFFETTUATA A.

Procediamo a valutare se è conveniente mantenere la relazione EFFETTUATA A tra PAZIENTE e DIAGNOSI in funzione della frequenza delle operazioni di inserimento nuove diagnosi e stampa dello storico diagnosi di un paziente.

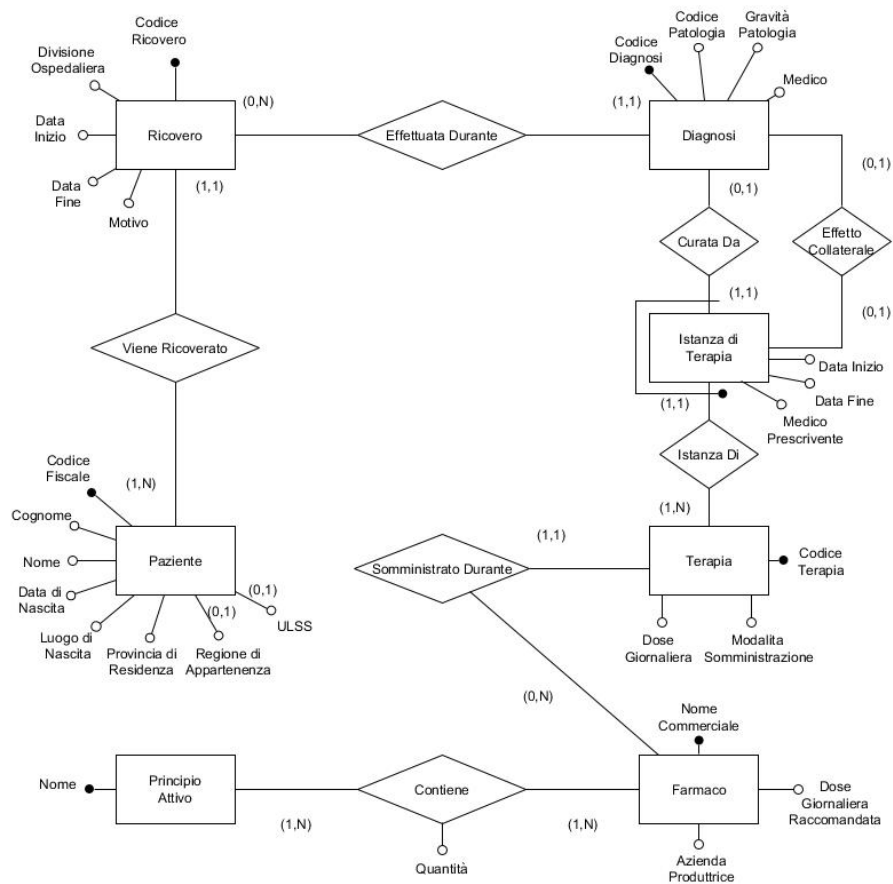


Figura 2: Lo schema Entità-Relazioni completo

Riportiamo i volumi stimati durante la fase dei requisiti in Tabella 2.2 e la frequenza delle operazioni interessate in Tabella 2.2.

Concept	Type	Volume
Paziente	E	100000
Ricovero	E	300000
Diagnosi	E	1200000
Terapia	E	30000
Istanza di Terapia	E	1200000
Farmaco	E	10000
Viene Ricoverato	R	300000
Effettuata Durante	R	1200000
Effettuta A	R	1200000
Curata Da	R	1200000
Istanza Di	R	1200000
Effetto Collaterale	R	50000
Somministrato Durante	R	10000

Tabella 1: Tabella dei volumi

Concept	Type	Volume
Paziente	E	100000
Ricovero	E	300000
Diagnosi	E	1200000
Viene Ricoverato	R	300000
Effettuata Durante	R	1200000
Effettuata A	R	1200000

Tabella 2: Tabella dei volumi ridotta

Operation	Type	Frequency
Inserimento Diagnosi	I	700/giorno
Stampa Storico Diagnosi	I	1400/giorno

Tabella 3: Tabella delle operazioni ridotta

Nel caso con ridondanza per l'operazione di inserimento di una nuova diagnosi sono necessari 1 accesso in scrittura per l'entità DIAGNOSI e 1 accesso in scrittura per entrambe le relazioni EFFETTUATA DURANTE e EFFETTUA-TA A. Per l'operazione di stampa delle diagnosi di un paziente sono necessari 12 accessi in lettura per la relazione EFFETTUATA A e 12 accessi in lettura per l'entità DIAGNOSI.

Considerando il costo di 1 accesso in scrittura come 2 accessi in lettura, calcoliamo il numero di accessi complessivo giornaliero per le 2 operazioni nel caso con ridondanza:

$$3W \times 700 + 24R \times 1400 \approx 37800R \quad (1)$$

Concept	Type	Access	Type
Inserimento Diagnosi			
Diagnosi	E	1	W
Effettutata Durante	R	1	W
Effettuata A	R	1	W
Stampa Storico Diagnosi			
Effettuata A	R	12	R
Diagnosi	E	12	R

Tabella 4: Tabella dei costi, caso con ridondanza

Nel caso senza ridondanza per l'operazione di inserimento di una nuova diagnosi sono necessari 1 accesso in scrittura per l'entità DIAGNOSI e 1 accesso in scrittura per la relazione EFFETTUTATA DURANTE. Per l'operazione di stampa delle diagnosi di un paziente sono necessari 3 accessi in lettura per la relazione VIENE RICOVERATO, 12 accessi in lettura per la relazione EFFETTUTATA DURANTE e 12 accessi in lettura per l'entità DIAGNOSI.

Considerando il costo di 1 accesso in scrittura come 2 accessi in lettura, calcoliamo il numero di accessi complessivo giornaliero per le 2 operazioni nel caso senza ridondanza:

$$2W \times 700 + 27R \times 1400 \approx 40600R \quad (2)$$

Concept	Type	Access	Type
Inserimento Diagnosi			
Diagnosi	E	1	W
Effettutata Durante	R	1	W
Stampa Storico Diagnosi			
Viene Ricoverato	R	3	R
Effettutata Durante	R	12	R
Diagnosi	E	12	R

Tabella 5: Tabella dei costi, caso senza ridondanza

Concludiamo che è conveniente mantenere la ridondanza visto il costo minore delle operazioni giornaliere.

2.3 Traduzione da ER a Relazionale

3 Progettazione Fisica

3.1 Nuovi Indici

4 Definizione della Base di Dati in SQL

4.1 Definizione delle Tabelle

4.2 Definizione di Query Significative

5 Analisi Dati

5.1 Popolamento della Base di Dati

5.2 Analisi Dati in R

6 Portale Web

6.1 Interfaccia Grafica

6.2 Comunicare con pgAdmin