

Normalizzazione

Università degli Studi del Sannio
Facoltà di Ingegneria
Corso di Laurea in Ingegneria Informatica

Corso di Basi di Dati
Anno Accademico 2005/2006
docente: ing. Corrado Aaron Visaggio

email: visaggio@unisannio.it

ricevimento: mercoledì 11.00-13.00.

Corrado Aaron Visaggio

1

Ridondanze e Anomalie

Impiegato	Stipendio	Progetto	Bilancio	Funzione
Rossi	20	Marte	2	Tecnico
Verdi	35	Giove	15	progettista
Verdi	35	Venere	15	progettista
Neri	55	Venere	15	direttore
Neri	55	Giove	15	consulente
Neri	55	Marte	2	consulente
Mori	48	Marte	2	direttore
Mori	48	Venere	15	progettista
Bianchi	48	Venere	15	progettista
Bianchi	48	Giove	15	direttore

▪ La chiave è Impiegato, Progetto

Lo stipendio dipende dall'impiegato

Il bilancio dipende dal progetto

Corrado Aaron Visaggio

2



Problemi

Ridondanze (Duplicazione della stessa informazione) il valore dello stipendio di ogni impiegato è ripetuto in tutte le tuple relative ad esso.



Anomalia di aggiornamento: se lo stipendio di un impiegato varia è necessario modificare il valore in tutte le tuple ove è presente l'impiegato.

Anomalia di cancellazione: se un impiegato non partecipa ad alcun progetto, ma non lascia l'azienda, si perde l'informazione sul suo stipendio.

Anomalia di inserimento: se si assume un nuovo impiegato, non si possono tracciare le sue informazioni se non prende parte ad un progetto.



In presenza di ridondanza

Alcuni dati devono essere ripetuti in diverse tuple, **senza aggiungere nuova informazione.**

L'**aggiornamento** di alcune informazioni deve essere ridonato anche laddove non occorre: anche se il linguaggio evita allo sviluppatore di aggiornare tutte le tuple, il sistema comunque lo fa!

La **cancellazione** di una tupla, perché l'insieme dei concetti non vale più, cancella sottoinsieme di concetti che hanno valore.

L'**inserimento** di informazioni relative ad un sottoinsieme di concetti non è possibile se non si inserisce l'intera tupla di concetti.



Dipendenze funzionali...

Una **dipendenza funzionale** è un vincolo di integrità che descrive legami di tipo funzionale tra gli attributi di una relazione.

Il valore dell'attributo *Impiegato* determina il valore dell'attributo *Stipendio*: *esiste una funzione che associa ad ogni elemento del dominio dell'attributo *Impiegato* un solo elemento del dominio dell'attributo *Stipendio**. (Vale lo stesso per *Progetto* e *Bilancio*).

Data una relazione r su uno schema $R(x)$ e due sottoinsiemi di attributi non vuoti Y e Z di X , diremo che esiste su r una dipendenza funzionale tra Y e Z se, per ogni coppia di tuple t_1, t_2 di r aventi gli stessi valori sugli attributi di Y , risulta che t_1, t_2 hanno gli stessi valori anche su Z .



...Dipendenze funzionali

Impiegato → *Stipendio*

Progetto → *Bilancio*

Una dipendenza funzionale $Y \rightarrow Z$ è **non banale** se nessun attributo di Z compare tra gli attributi di Y . Si può rendere non banale una dipendenza selezionando l'intersezione tra il primo ed il secondo membro di una relazione banale.

La chiave K di una relazione ha sicuramente una relazione funzionale con qualsiasi altro attributo della relazione: il vincolo di dipendenza funzionale generalizza il vincolo di chiave: una dipendenza funzionale $Y \rightarrow Z$ degenera nel vincolo di chiave in uno schema $R(X)$ se l'unione di Y e Z è pari ad X . Y è super chiave per lo schema $R(X)$.

Prima, Seconda e Terza Forma Normale

Una relazione in cui ogni dominio è **atomico**, si dice che è in prima forma normale, **1st Normal Form (1NF)**.

relazione in 2^a f.n.: *se è in 1^a f.n. e se tutti i suoi attributi non primi dipendono funzionalmente e completamente dalla chiave;*

BDR in 2^a f.n.: *se lo è ogni sua relazione;*

relazione in 3^a f.n.: *se è in 2^a f.n. e se ogni suo attributo non primo non dipende transitivamente dalla chiave;*

BDR in 3^a f.n.: *se lo è ogni sua relazione.*

Normalizzazione in Prima

CF	Nome	Cognome	Data Di Nascita	Indirizzo
1234	Gio'	Vanni	01/01/01	Via Taglia la Zazzera n°20, Bari
2345	Anna	Lacapa	29/02/04	Via Taglia la Zazzera n°18, Bari
3456	Natale	di Pasqua	24/12/00	Piazza Santo Stefano n°1, Salerno
4567	Marco	Antonio	15/03/03	Via Salaria n°4, Ostia (Roma)



Id	Den	Valore	Numero Civico	Città	Provincia
1	Via	Taglia la Zazzera	20	Bari	Null
2	Via	Taglia la Zazzera	18	Bari	Null
3	Piazza	Santo Stefano	1	Salerno	Null
4	Via	Salaria	4	Ostia	Roma



Normalizzazione in Terza

CF	Nome	Cognome	Progetto	Sede
1234	Miles	Davis	Fiato	Roma
2345	Stefano	Bollani	Piano	Bari
3456	Angela	McCluskey	Chitarra	Benevento
4567	Chet	Baker	Fiato	Roma
5678	Ellis	Peterson	Chitarra	Benevento



CF	Nome	Cognome	Progetto
1234	Miles	Davis	Fiato
2345	Stefano	Bollani	Piano
3456	Angela	McCluskey	Chitarra
4567	Chet	Baker	Fiato
5678	Ellis	Peterson	Chitarra

Progetto	Sede
Fiato	Roma
Piano	Bari
Chitarra	Benevento



Impiegato	Stipendio	Progetto	Bilancio	Funzione
Rossi	20	Marte	2	Tecnico
Verdi	35	Giove	15	progettista
Verdi	35	Venere	15	progettista
Neri	55	Venere	15	direttore
Neri	55	Giove	15	consulente
Neri	55	Marte	2	consulente
Mori	48	Marte	2	direttore
Mori	48	Venere	15	progettista
Bianchi	48	Venere	15	progettista
Bianchi	48	Giove	15	direttore



Forma Normale di Boyce and Codd

Le dipendenze

Impiegato → Stipendio
 Progetto → Bilancio

sono causa di anomalie;

Non lo è la dipendenza

Impiegato Progetto → Funzione

La differenza risiede nel fatto che Impiegato Progetto è una superchiave della relazione.

Si conclude che le anomalie sono generate dalle dipendenze funzionali $x \rightarrow y$ che permettono la presenza di più tuple fra loro uguali sugli attributi in x , tali cioè che x non contenga una chiave.

Una relazione r è in forma normale di Boyce e Codd se per ogni dipendenza funzionale non banale $X \rightarrow Y$ definita su di essa, X contiene una chiave K di r , cioè X è superchiave per r .



Decomposizione...

La normalizzazione consente di sostituire una relazione che non è in forma normale di B&C in due o più relazione che invece lo sono.

Impiegato	Stipendio
Rossi	20
Verdi	35
Neri	55
Mori	48
Bianchi	48

Progetto	Bilancio
Marte	2
Giove	15
Venere	15

Impiegato	Progetto	Funzione
Rossi	Marte	Tecnico
Verdi	Giove	progettista
Verdi	Venere	progettista
Neri	Venere	direttore
Neri	Giove	consulente
Neri	Marte	consulente
Mori	Marte	direttore
Mori	Venere	progettista
Bianchi	Venere	progettista
Bianchi	Giove	direttore

In molti casi pratici si possono determinare tante scomposizioni, quante sono le dipendenze funzionali. Ma **questa tecnica non è sempre utilizzabile...**



Decomposizione senza perdite...

Impiegato	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano

Questa relazione ha le seguenti dipendenze funzionali

Impiegato → Sede

Progetto → Sede

Procediamo in accordo alla forma normale di Boyce e Codd e ricostruiamo la tabella originale con un join.

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

Corrado Aaron Visaggio

Progetto	Sede
Marte	Roma
Giove	Milano
Venere	Milano
Saturno	Milano

...Decomposizione senza perdite...

Impiegato	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano
Verdi	Saturno	Milano
Neri	Giove	Milano



Problema: la relazione ottenuta ha tutte le tuple della relazione di origine più altre due.

Data una relazione r su un insieme di attributi X , se X_1 e X_2 sono due sottoinsiemi di X la cui unione sia pari a X stesso, allora il join delle due relazioni ottenute per proiezione da r su X_1 ed X_2 , rispettivamente, è una relazione che contiene tutte le tuple di r , più eventualmente altre ("spurie").



...Decomposizione senza perdite

Sia r una relazione su X e siano X_1 ed X_2 sottoinsiemi di X tali che $X_1 \cup X_2 = X$. Inoltre sia $X_0 = X_1 \cap X_2$. Se r soddisfa la dipendenza funzionale $X_0 \rightarrow X_1$ oppure la dipendenza funzionale $X_0 \rightarrow X_2$, allora r si decompone senza perdita su X_1 ed X_2 .

*r si **decompone senza perdita** su due relazioni se l'insieme degli attributi comuni alle due relazioni è chiave per almeno una delle relazioni decomposte.*

La condizione è sufficiente ma non necessaria: esistono relazioni che non soddisfano nessuna delle due relazioni e non riportano perdite nelle decomposizioni.



Conservazione delle dipendenze

Per decomporre senza perdita, utilizziamo le due relazioni:

Impiegato \rightarrow Sede

Impiegato \rightarrow Progetto

ove Impiegato è chiave della prima relazione.

Problema: Inseriamo la tupla che specifica la partecipazione di Neri che opera a Milano al progetto Marte. Nella tabella originaria violazione di Progetto \rightarrow Sede. Tale violazione non si rileva sulle relazioni decomposte: i due attributi sono, infatti, stati separati.

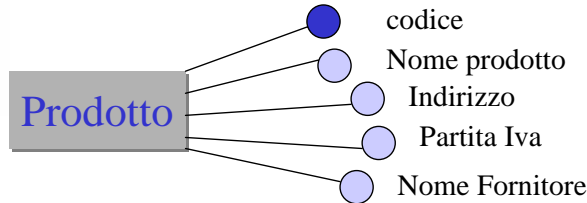
Conservazione delle dipendenze: *In ogni decomposizione, ciascuna delle dipendenze funzionali dello schema originario dovrebbe coinvolgere attributi che compaiono tutti insieme in uno degli schemi decomposti.*

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

Impiegato	Progetto
Rossi	Marte
Verdi	Giove
Verdi	Venere
Neri	Saturno
Neri	Venere

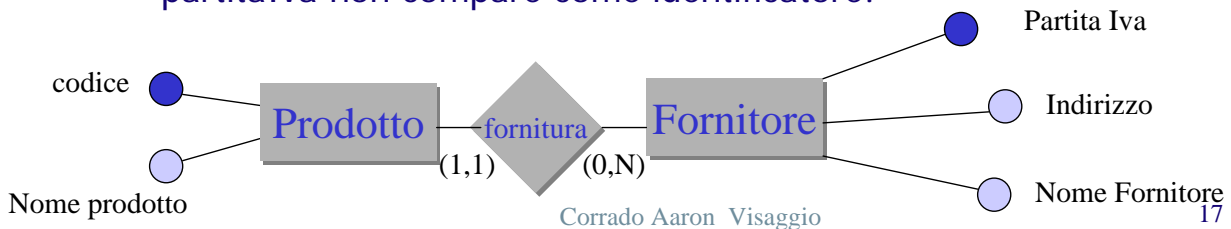
Normalizzazione su entità

Bisogna considerare le dipendenze funzionali che sussistono tra gli attributi dell'entità e verificare che ciascuna di esse abbia come primo membro l'identificatore.



partitaIva → NomeFornitore

partitaIva non compare come identificatore!



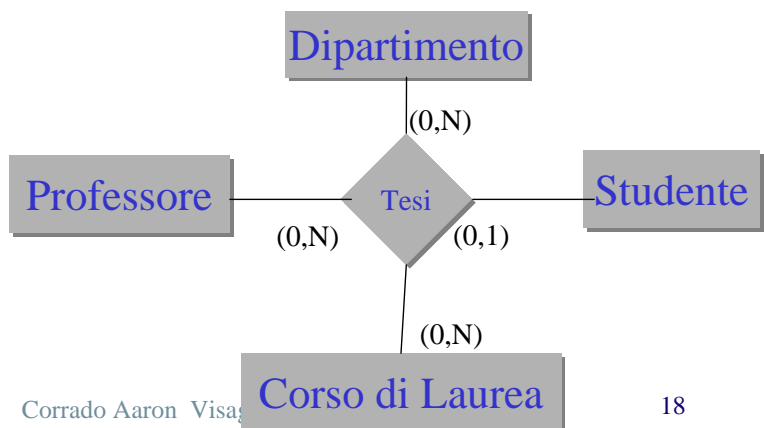
Corrado Aaron Visaggio

Verifiche di Normalizzazioni...

Per verificare il soddisfacimento della terza forma normale è necessario individuare le dipendenze funzionali che sussistono nell'ambito dell'associazione in esame.

Ogni associazione binaria è in terza forma normale e in B&C, per cui verifichiamo la normalizzazione sulle associazioni non binarie.

- Studente → Corso di Laurea
- Studente → professore
- Professore → Dipartimento

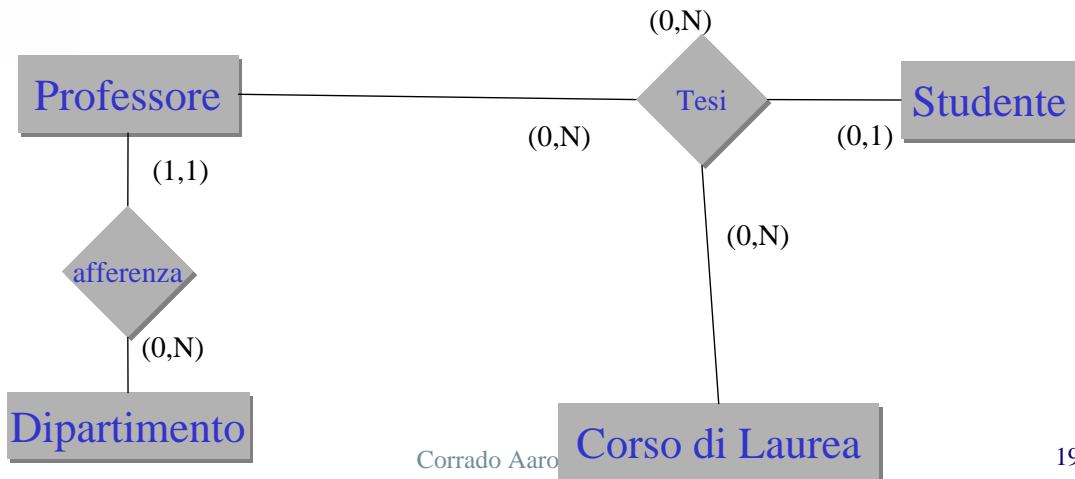


Corrado Aaron Visaggio

... Verifiche di Normalizzazioni

Lo studente è chiave: dato uno studente riesco ad ottenere tutte le relazioni: la terza relazione causa una violazione della terza forma normale.

Decomponiamo la relazione:

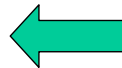


Terza Forma Normale

Non è sempre possibile raggiungere la forma normale di Boyce & Codd, conservando le dipendenze



Dirigente -> Sede
 Progetto Sede-> Dirigente



La relazione non è in forma normale di B&C perchè: nella dip. Funzionale dirigente-> Sede, il primo membro non è superchiave della relazione. Inoltre, la seconda dip funzionale non è decomponibile in quanto coinvolge tutti gli elementi della tavola.

Dirigente	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Marte	Milano
Neri	Saturno	Milano
Neri	Venere	Milano



Terza forma normale...

Si ricorre ad una condizione meno restrittiva:

Una relazione r è in terza forma normale se, per ogni dipendenza funzionale non banale $X \rightarrow Y$ almeno una delle seguenti condizioni è verificata:

- X contiene una chiave K di r
- Ogni attributo di Y è contenuto in almeno una chiave di r .

La relazione dell'esempio è in terza forma normale, infatti:

- Progetto Sede \rightarrow Dirigente ha al primo membro una chiave della relazione
- Dirigente \rightarrow Sede ha un attributo al secondo membro che fa parte della chiave della relazione [Progetto Sede].

Vi è una ridondanza:

Ogni volta che un dirigente compare in una tupla, è ripetuta la sede di lavoro in cui opera



...Terza forma normale

La ridondanza è tollerata dalla terza forma normale, dal momento che non vi è una decomposizione che mantenga le dipendenze.

La terza forma normale ha:

- Il vantaggio di essere sempre ottenibile, contrariamente alla forma normale di B&C
- Lo svantaggio di lasciar filtrare, comunque, delle ridondanze.



Es 1...

Prodotto	Componente	Tipo	Quantità	Prezzo-C	Fornitore	Prezzo-P
Libreria	Legno	Noce	50	10000	Forrest	400000
Libreria	Bulloni	B212	1000	100	Bolt & Son	400000
Libreria	Vetro	Cristal	3	5000	Clean	400000
Scaffale	Legno	Mogano	50	15000	Forrest	300000
Scaffale	Bulloni	B212	2000	100	Bolt & Son	300000
Scaffale	Bulloni	B412	300	250	Bolt & Son	300000
Scrivania	Legno	Noce	20	8000	Wood	250000
Scrivania	Maniglie	B212	10	2000	Bolt & Son	250000
Tavolo	Legno	Noce	40	10000	Forrest	200000

Corrado Aaron Visaggio

23



...Es1...

Descriviamo le ridondanze presenti ed individuare una decomposizione della relazione che non presenti tali ridondanze

Fornire l'istanza dello schema così ottenuto, corrispondente all'istanza originale.

È possibile ricostruire l'istanza originale a partire da tale istanza?



Prodotto	Componente	Tipo	Quantità	Prezzo-C	Fornitore	Prezzo-P
Libreria	Legno	Noce	50	10000	Forrest	400000
Libreria	Bulloni	B212	1000	100	Bolt & Son	400000
Libreria	Vetro	Cristal	3	5000	Clean	400000
Scaffale	Legno	Mogano	50	15000	Forrest	300000
Scaffale	Bulloni	B212	2000	100	Bolt & Son	300000
Scaffale	Bulloni	B412	300	250	Bolt & Son	300000
Scrivania	Legno	Noce	20	8000	Wood	250000
Scrivania	Maniglie	B212	10	2000	Bolt & Son	250000
Tavolo	Legno	Noce	40	10000	Forrest	200000

Corrado Aaron Visaggio

25



Prodotto	Componente	Tipo	Quantità	Prezzo-C	Fornitore	Prezzo-P
Libreria	Legno	Noce	50	10000	Forrest	400000
Libreria	Bulloni	B212	1000	100	Bolt & Son	400000
Libreria	Vetro	Cristal	3	5000	Clean	400000
Scaffale	Legno	Mogano	50	15000	Forrest	300000
Scaffale	Bulloni	B212	2000	100	Bolt & Son	300000
Scaffale	Bulloni	B412	300	250	Bolt & Son	300000
Scrivania	Legno	Noce	20	8000	Wood	250000
Scrivania	Maniglie	B212	10	2000	Bolt & Son	250000
Tavolo	Legno	Noce	40	10000	Forrest	200000

Corrado Aaron Visaggio

26



Prodotto	Componente	Tipo	Quantità	Prezzo-C	Fornitore	Prezzo-P
Libreria	Legno	Noce	50	10000	Forrest	400000
Libreria	Bulloni	B212	1000	100	Bolt & Son	400000
Libreria	Vetro	Cristal	3	5000	Clean	400000
Scaffale	Legno	Mogano	50	15000	Forrest	300000
Scaffale	Bulloni	B212	2000	100	Bolt & Son	300000
Scaffale	Bulloni	B412	300	250	Bolt & Son	300000
Scrivania	Legno	Noce	20	8000	Wood	250000
Scrivania	Maniglie	B212	10	2000	Bolt & Son	250000
Tavolo	Legno	Noce	40	10000	Forrest	200000

Corrado Aaron Visaggio

27



...Es1...

Riconduciamoci alla forma normale di Boyce and Codd

Prezzo-P -> Prodotto

Tipo, Componente, Fornitore -> Prezzo-C

Componente, Tipo, Prezzo-C -> Fornitore



R1 (Prodotto, Prezzo-P)

R2 (Tipo, Componente, Fornitore, Prezzo-C)

R3 (Prodotto, Prezzo-C, Componente, Tipo, Fornitore, Quantità)

Corrado Aaron Visaggio

28

Es 2...

Impiegato (Nome, Livello, Stipendio)

Nome -> Livello, Stipendio

Livello -> Stipendio

1. R1 (Nome, Stipendio) R2 (Livello, Stipendio)
2. R1 (Nome, Livello) R2 (Livello, Stipendio)

...Es 2

La soluzione migliore è:

1. R1 (Nome, Livello) R2 (Livello, Stipendio)

Infatti la soluzione

1. R1 (Nome, Stipendio) R2 (Livello, Stipendio)

Produrrebbe delle tuple spurie... (infatti Stipendio, che è l'unico attributo in comune non è chiave per nessuna delle due relazioni)



Es 3

Codice avviamento postale (città, via, Cap)

Città, via->Cap

Cap -> Città

Cap non è chiave della relazione!

R1 (Via,Cap) R2(Città, Cap), ma la dipendenza Città, via->Cap non è catturata da nessuna relazione: possiamo introdurre vie con lo stesso nome, appartenenti a città diverse ma che hanno cap differente.

Introduciamo, allora l'attributo settore....

Via, Città -> Settore

Città, Settore -> CAP

CAP -> Città, Settore

R1 (Via, città, Settore) R2 (CAP, Città, Settore)

Perché non va bene R2 (CAP, Città, Settore)?



Es 4

DEGENTE (Cofice-Fiscale, Nome-Degente, Cognome-Degente, **Giorno, Reparto, Capo-Reparto, Stanza, Infermiere-di-Stanza**)

- Quali sono le dipendenze?
- E' in BCNF? Se no, portarla in BCNF.

Codice-Fiscale -> Nome-degente, Cognome-degente

Reparto -> Capo-Reparto

Stanza -> Reparto

Stanza -> Infermiere-di-stanza

Codice-Fiscale, Data -> Stanza

DEGENTE (Codice-Fiscale, Nome-degente, Cognome-degente)

REPARTO (Reparto, Capo-Reparto)

STANZA (Stanza, Reparto, Infermiere-di-Stanza)

DEGENZA (Codice-Fiscale, Data, Stanza)

Es 5...

Un'agenzia bancaria vuole ricordare i diversi dati relativi ai clienti.

- Ogni cliente può avere più conti correnti
- Su ogni conto i clienti possono effettuare diverse transazioni.

CLIENTE (Codice-Cliente, Nome, Cognome, Indirizzo, Città, Telefono, Numero-Conto, Saldo, Numero-transazione, Tipo, Data, Ammontare)

- Tipo ha come dominio {versamento, prelievo}
- Ammontare fa riferimento all'ammontare della singola transazione
- Ogni conto fa riferimento all'ammontare della singola transazione.
- Ogni conto fa riferimento ad un solo cliente ed ogni transazione fa riferimento ad un solo conto.
- Verificare se lo schema è normalizzato

...Es5...

CLIENTE (Codice-Cliente, Nome, Cognome, Indirizzo, Città, Indirizzo, Telefono)

CONTO(Codice-Cliente, Numero-conto, Saldo, Numero-Transazione, Tipo, Data, Ammontare)

E' in BCNF?

No, infatti in conto sono riportati codice-cliente e saldo.

...Es 5

- Ammontare fa riferimento all'ammontare della singola transazione
- Ogni conto fa riferimento all'ammontare della singola transazione.
- Numero-Transazione -> Numero-Conto, Tipo, Data, Ammontare
- Ogni conto fa riferimento ad un solo cliente ed ogni transazione fa riferimento ad un solo conto.
- Numero-Conto -> Codice-Cliente, Saldo
- Codice-Cliente -> Nome, Cognome, Indirizzo, Città, Telefono

TRansazione(Numero-Transazione, Numero-Conto, Tipo, Data, Ammontare)
CONTO(Numero-Conto, Codice-Cliente, Saldo)
CLIENTE(Codice-Cliente , Nome, Cognome, Indirizzo, Città, Telefono)

Es6

VOLO-AEREO (Codice-Volo, Aeroporto-Partenza, Città-Partenza, Ora-Partenza, Aeroporto-Arrivo, Città-Arrivo, Ora-Arrivo)

VOLO (Codice-Volo, Aeroporto-Partenza, Ora-Partenza, Aeroporto-Arrivo, Ora-Arrivo)

AEROPORTO-ARRIVO (Aeroporto-Arrivo, Città-Arrivo)

AEROPORTO-PARTENZA (Aeroporto-Partenza, Città-Partenza)

VOLO (Codice-Volo, Aeroporto-Partenza, Ora-Partenza, Aeroporto-Arrivo, Ora-Arrivo)

AEROPORTO(Aeroporto, Città)