

Basi di dati I — 2 febbraio 2021 — Prova breve — Compito A
Tempo a disposizione: un'ora

Cognome: _____ Nome: _____ Matricola: _____

Basi di dati I — 2 febbraio 2021 — Compito A

Domanda 1 (35% per la prova breve e 20% per la prova lunga)

Considerare la relazione seguente

CF	NomePaz	CT	Data	Esito	CodASL	ASL	SedeASL	IDReg	NomeReg
RSSMRI...	M. Rossi	1	12/12/20	Pos	1	RM 1	Via Po	LAZ	Lazio
RSSMRI...	M. Rossi	2	12/01/21	Neg	1	RM 1	Via Po	LAZ	Lazio
RSSMNL...	M. Rossi	3	15/12/20	Dub	1	RM 1	Via Po	LAZ	Lazio
RSSMNL...	M. Rossi	4	15/12/20	Neg	1	RM 1	Via Po	LAZ	Lazio
BRNLGU...	L. Bruni	5	12/01/21	Pos	3	LT	Via Italia	LAZ	Lazio
BNILCA ...	L. Bini	6	12/12/20	Neg	1	Ma 1	Via Roma	MAR	Marche
VRDPRI ...	P. Verdi	7	15/12/20	Pos	1	Mo 1	Corso ...	MOL	Molise

La relazione mostra (in forma non normalizzata) i dati relativi ad un insieme di tamponi eseguiti su pazienti, secondo le seguenti specifiche:

- ogni tampone viene eseguito su un paziente; ogni tampone ha un codice (CT), univoco, una data e un esito
- ogni paziente ha un codice fiscale e un nome e appartiene ad una ASL
- ogni ASL si trova in una regione, ha un codice (CodASL) che la identifica nell'ambito della regione e ha una sede
- ogni regione ha un identificatore (IDReg) e un nome

Con riferimento alle specifiche e ai dati nell'esempio

1. mostrare le dipendenze funzionali rilevate (limitarsi a quelle che hanno a primo membro ID, codici e numeri) e la chiave della relazione

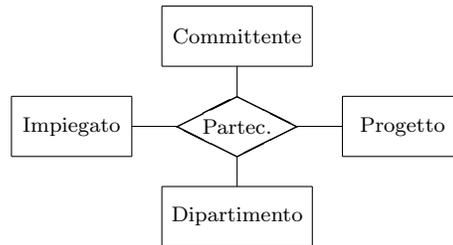
2. mostrare uno schema concettuale per la realtà di interesse (rispettando le specifiche per i vari codici e senza aggiungere altri attributi)

Basi di dati I — 2 febbraio 2021 — Compito A

3. mostrare una buona decomposizione della relazione originaria che soddisfi la BCNF (mostrare le tabelle, con i dati, indicando la chiave di ciascuna con la sottolineatura)

Domanda 2 (30% per la prova breve)

A seguito di una prima, superficiale analisi di una realtà di interesse, è stato prodotto lo schema in figura (gli impiegati di un'azienda di servizi, che lavorano nei dipartimenti nei quali è organizzata l'azienda, partecipano a progetti commissionati da committenti esterni):



Modificare lo schema (decomponendo la relationship e aggiungendo ulteriori entità, se necessario) tenendo conto delle seguenti specifiche:

- ogni impiegato afferisce ad uno e un solo dipartimento;
- ogni progetto ha zero o più committenti e ogni committente uno e un solo progetto
- ogni impiegato lavora per zero o più progetti (e quindi per tutti e soli i relativi committenti)
- ogni dipartimento ha come committenti e progetti quelli dei propri impiegati

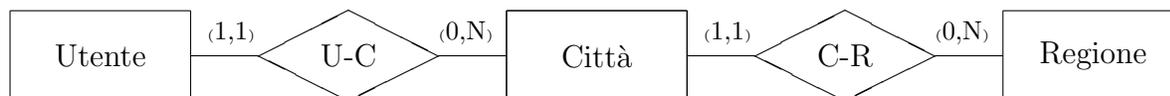
Si noti che sicuramente ognuna delle entità ha attributi, che non vengono mostrati e non vanno inseriti nella soluzione.

Ripetere quanto fatto sopra con riferimento alle seguenti specifiche:

- ogni impiegato afferisce ad uno e un solo dipartimento;
- ogni progetto ha uno e un solo committente e ogni committente uno o più progetti
- ogni dipartimento ha zero o più committenti e segue tutti e soli i relativi progetti; ogni committente interagisce con uno e un solo dipartimento
- ogni impiegato lavora per tutti e soli i progetti dei committenti del proprio dipartimento

Domanda 3 (35% per la prova breve e 20% per la prova lunga)

Considerare lo schema concettuale seguente:



Valutare la convenienza dell'introduzione di una relationship U-R fra Utente e Regione. Tale relationship è ridondante in quanto derivabile dalla concatenazione delle relationship esistenti. Considerare un carico applicativo che includa come operazioni principali le seguenti:

1. inserimento di una nuova occorrenza della relationship U-C, dati i codici identificativi di Utente e Città, assumendo per semplicità che le occorrenze di Utente e Città siano già nella base di dati, con frequenza $f_1 = 100$
2. ricerca della Regione di un Utente (di cui è dato il codice), con frequenza $f_2 = 10.000$
3. ricerca della Città di un Utente (di cui è dato il codice), con frequenza $f_3 = 1.000$

Considerare i costi delle letture e scritture delle entità, ignorando quelli delle relationship e considerare il costo delle scritture triplo di quello delle letture.

Basi di dati I — 2 febbraio 2021 — Prova breve — Compito B
Tempo a disposizione: un'ora

Cognome: _____ Nome: _____ Matricola: _____

Basi di dati I — 2 febbraio 2021 — Compito B

Domanda 1 (35% per la prova breve e 20% per la prova lunga)

Considerare la relazione seguente

CF	NomePaz	CT	Data	Esito	CodASL	ASL	SedeASL	IDReg	NomeReg
RSSMRI...	M. Rossi	1	12/12/20	Pos	1	RM 1	Via Po	LAZ	Lazio
RSSMRI...	M. Rossi	2	12/01/21	Neg	1	RM 1	Via Po	LAZ	Lazio
RSSMNL...	M. Rossi	1	15/12/20	Dub	1	RM 1	Via Po	LAZ	Lazio
RSSMNL...	M. Rossi	2	15/12/20	Neg	1	RM 1	Via Po	LAZ	Lazio
BRNLGU...	L. Bruni	1	12/01/21	Pos	2	LT	Via Italia	LAZ	Lazio
BNILCA ...	L. Bini	1	12/12/20	Neg	3	Ma 1	Via Roma	MAR	Marche
VRDPRI ...	P. Verdi	1	15/12/20	Pos	4	Mo 1	Corso ...	MOL	Molise

La relazione mostra (in forma non normalizzata) i dati relativi ad un insieme di tamponi eseguiti su pazienti, secondo le seguenti specifiche:

- ogni tampone viene eseguito su un paziente; ogni tampone ha un codice (CT), univoco fra i tamponi di quel paziente, una data e un esito
- ogni paziente ha un codice fiscale e un nome e appartiene ad una ASL
- ogni ASL si trova in una regione, ha un codice (CodASL) che la identifica in modo univoco e ha una sede
- ogni regione ha un identificatore (IDReg) e un nome

Con riferimento alle specifiche e ai dati nell'esempio

1. mostrare le dipendenze funzionali rilevate (limitarsi a quelle che hanno a primo membro ID, codici e numeri) e la chiave della relazione

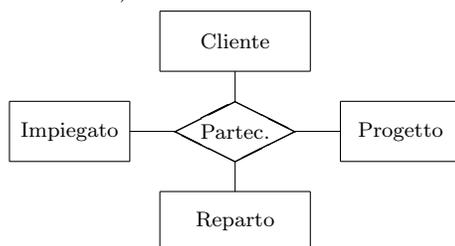
2. mostrare uno schema concettuale per la realtà di interesse (rispettando le specifiche per i vari codici e senza aggiungere altri attributi)

Basi di dati I — 2 febbraio 2021 — Compito B

3. mostrare una buona decomposizione della relazione originaria che soddisfi la BCNF (mostrare le tabelle, con i dati, indicando la chiave di ciascuna con la sottolineatura)

Domanda 2 (30% per la prova breve)

A seguito di una prima, superficiale analisi di una realtà di interesse, è stato prodotto lo schema in figura (gli impiegati di un'azienda di servizi, che lavorano nei reparti nei quali è organizzata l'azienda, partecipano a progetti commissionati da clienti esterni):



Modificare lo schema (decomponendo la relationship e aggiungendo ulteriori entità, se necessario) tenendo conto delle seguenti specifiche:

- ogni impiegato afferisce ad uno e un solo reparto;
- ogni impiegato partecipa ad un solo progetto (che non è necessariamente legato ad un reparto, cioè impiegati di reparti diversi possono lavorare sul medesimo progetto);
- ogni progetto ha un solo cliente e ogni impiegato interagisce solo con il cliente del proprio progetto.

Si noti che sicuramente ognuna delle entità ha attributi, che non vengono mostrati e non vanno inseriti nella soluzione.

Ripetere quanto fatto sopra con riferimento alle seguenti specifiche:

- ogni impiegato afferisce ad uno e un solo reparto;
- ogni progetto ha uno e un solo cliente
- ogni impiegato lavora per zero o più clienti e potenzialmente per tutti i progetti di quei clienti (e per nessun altro progetto)
- ogni reparto ha come clienti e progetti quelli dei propri impiegati

Domanda 3 (35% per la prova breve e 20% per la prova lunga)

Considerare lo schema concettuale seguente:



Valutare la convenienza dell'introduzione di una relationship U-R fra Cliente e Regione. Tale relationship è ridondante in quanto derivabile dalla concatenazione delle relationship esistenti. Considerare un carico applicativo che includa come operazioni principali le seguenti:

1. inserimento di una nuova occorrenza della relationship C-C, dati i codici identificativi di Cliente e Città, assumendo per semplicità che le occorrenze di Cliente e Città siano già nella base di dati, con frequenza $f_1 = 100$
2. ricerca della Regione di un Cliente (di cui è dato il codice), con frequenza $f_2 = 10.000$
3. ricerca della Città di un Cliente (di cui è dato il codice), con frequenza $f_3 = 1.000$

Considerare i costi delle letture e scritture delle entità, ignorando quelli delle relationship e considerare il costo delle scritture triplo di quello delle letture.

Basi di dati I — 2 febbraio 2021 — Prova lunga — Compito E
Tempo a disposizione: un'ora e quarantacinque minuti

Cognome: _____ Nome: _____ Matricola: _____

Basi di dati I — 2 febbraio 2021 — Compito E

Domanda 1 (35% per la prova breve e 20% per la prova lunga)

Considerare la relazione seguente

CF	NomePaz	CT	Data	Esito	CodASL	ASL	SedeASL	IDReg	NomeReg
RSSMRI...	M. Rossi	1	12/12/20	Pos	1	RM 1	Via Po	LAZ	Lazio
RSSMRI...	M. Rossi	2	12/01/21	Neg	1	RM 1	Via Po	LAZ	Lazio
RSSMNL...	M. Rossi	3	15/12/20	Dub	1	RM 1	Via Po	LAZ	Lazio
RSSMNL...	M. Rossi	4	15/12/20	Neg	1	RM 1	Via Po	LAZ	Lazio
BRNLGU...	L. Bruni	5	12/01/21	Pos	3	LT	Via Italia	LAZ	Lazio
BNILCA ...	L. Bini	6	12/12/20	Neg	1	Ma 1	Via Roma	MAR	Marche
VRDPRI ...	P. Verdi	7	15/12/20	Pos	1	Mo 1	Corso ...	MOL	Molise

La relazione mostra (in forma non normalizzata) i dati relativi ad un insieme di tamponi eseguiti su pazienti, secondo le seguenti specifiche:

- ogni tampone viene eseguito su un paziente; ogni tampone ha un codice (CT), univoco, una data e un esito
- ogni paziente ha un codice fiscale e un nome e appartiene ad una ASL
- ogni ASL si trova in una regione, ha un codice (CodASL) che la identifica nell'ambito della regione e ha una sede
- ogni regione ha un identificatore (IDReg) e un nome

Con riferimento alle specifiche e ai dati nell'esempio

1. mostrare le dipendenze funzionali rilevate (limitarsi a quelle che hanno a primo membro ID, codici e numeri) e la chiave della relazione

2. mostrare uno schema concettuale per la realtà di interesse (rispettando le specifiche per i vari codici e senza aggiungere altri attributi)

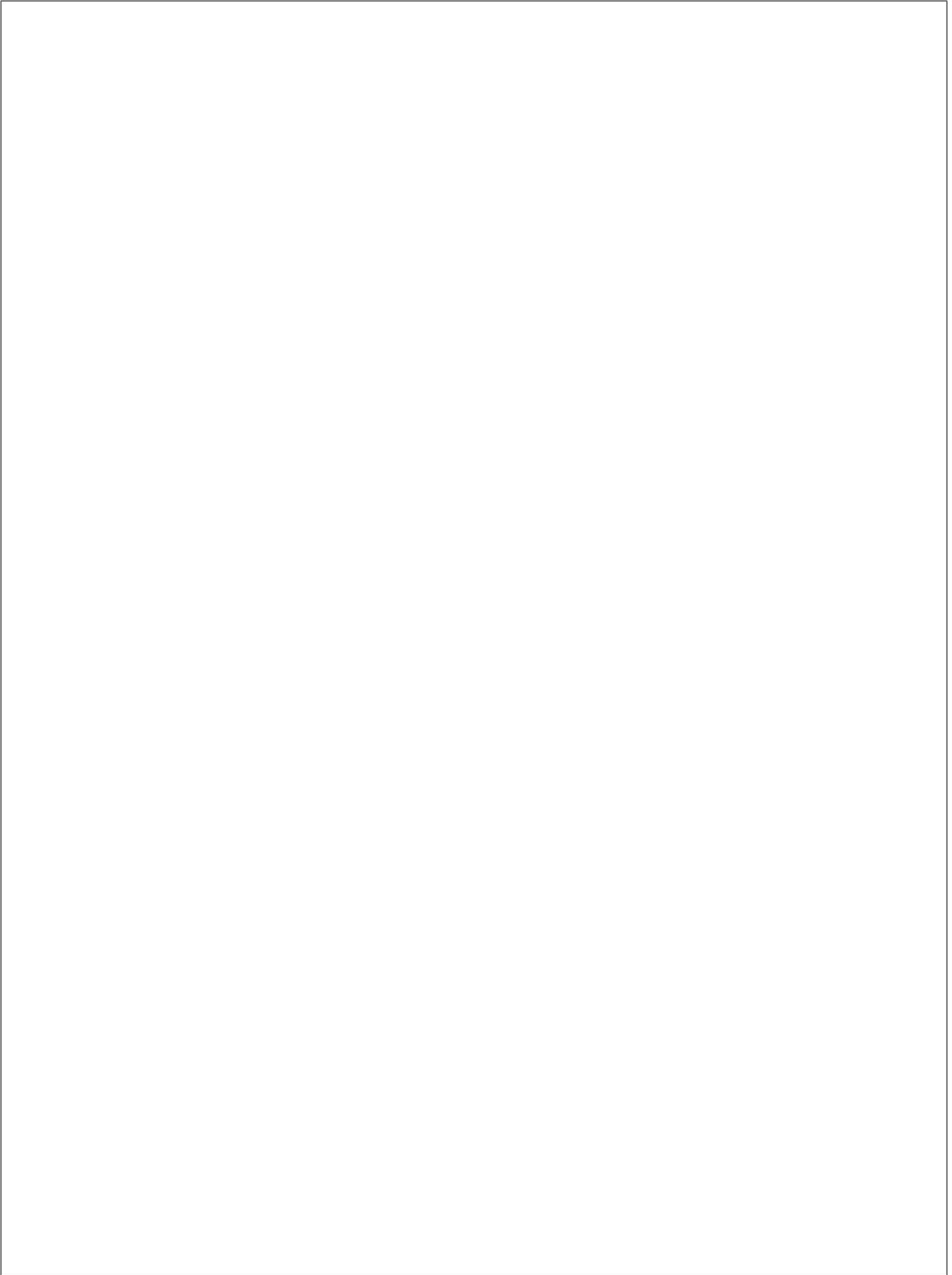
Basi di dati I — 2 febbraio 2021 — Compito E

3. mostrare una buona decomposizione della relazione originaria che soddisfi la BCNF (mostrare le tabelle, con i dati, indicando la chiave di ciascuna con la sottolineatura)

Domanda 2 (30% per la prova completa)

Mostrare, nel riquadro alla pagine seguente, uno schema concettuale per una realtà relativa ad un insieme di ristoranti di una regione, secondo le specifiche seguenti (molto semplificate rispetto alla realtà, ma con alcuni dettagli che vanno specificati come richiesto).

- Ogni ristorante ha codice, nome e indirizzo; gli indirizzi sono “standardizzati”, nel senso che la base di dati contiene tutti i possibili indirizzi della regione, ciascuno con via, numero civico, comune (con codice e nome) e provincia (con codice e nome)
- Per ogni ristorante (ad esempio per contrastare le infiltrazioni della criminalità organizzata) si rappresentano i proprietari, con codice fiscale, nome, cognome, data di nascita, quota di possesso del ristorante (ad esempio Mario Rossi può possedere il 30% del ristorante Buon Gusto) e parentele eventualmente esistenti tra proprietari (anche di ristoranti diversi), con tipo di parentela (ad esempio Mario Rossi è cugino di Aldo Verdi e marito di Anna Pini). Le parentele sono rappresentate con codice e descrizione (ad esempio “A” per fratello/sorella; “B” per genitore/figlio)
- Ogni ristorante ha un menu, composto da un insieme di piatti offerti (ad es. “spaghetti alla carbonara,” “agnello al forno”). I piatti sono caratterizzati da un codice (unico per piatto per tutti i ristoranti) e un nome. Ogni piatto ha un costo, che varia e dipende dal ristorante, mentre il nome del piatto è identico per tutti i ristoranti.
- Ogni piatto ha un insieme di ingredienti, che sono descritti ciascuno da un codice e un nome (unici per tutti i ristoranti).
- Si vuole tenere traccia delle possibili allergie o intolleranze alimentari, indicando, per ciascun ingrediente quali sono le intolleranze o allergie cui è associato (ad esempio, la panna è associata all’intolleranza al lattosio). Nota bene, il tutto è molto semplificato e si desidera solo rappresentare intolleranze/allergie (con codice e nome) e associarle agli ingredienti.



Basi di dati I — 2 febbraio 2021 — Compito E

Domanda 3 (30%) Considerare una base di dati sul seguente schema

- PROVINCE(Sigla, Nome, Abitanti, Regione); contiene dati su tutte le province d'Italia, una ennupla per provincia, con nome, abitanti e regione di appartenenza
- TAMPONI(Provincia, Data, NumeroTamponi, NumeriPositivi), con vincolo di integrità referenziale fra Provincia e la chiave Sigla di PROVINCE; ogni ennupla contiene dati sui tamponi effettuati in un giorno in una certa provincia; i dati sono acquisiti gradualmente e quindi può succedere che in un certo momento i dati relativi ad una certa giornata possano essere disponibili solo per una parte delle province

Con riferimento a tale schema

1. Formulare in algebra relazionale l'interrogazione che trova le province (basta indicare la sigla) per le quali non è disponibile l'informazione sui tamponi del 31/01/2021

2. Formulare in algebra relazionale l'interrogazione che trova le regioni per le quali l'informazione sui tamponi del 31/01/2021 è disponibile per tutte le province

3. Formulare in SQL l'interrogazione che, per ogni provincia, mostra: la sigla, il nome, il numero di tamponi effettuati e il numero di positivi per il giorno 25/01/2021

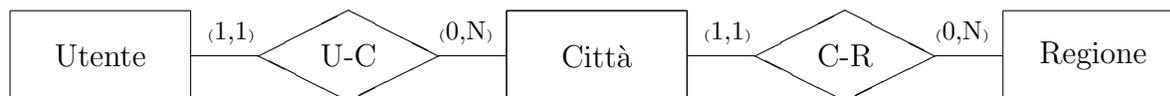
Basi di dati I — 2 febbraio 2021 — Compito E

4. Formulare in SQL l'interrogazione che trova, per ciascuna regione, il numero totale di tamponi effettuati, il numero totale di tamponi positivi e la percentuale di positivi (dove la percentuale è pari al totale dei positivi moltiplicato per cento diviso il totale di tamponi effettuati)

5. Formulare in SQL l'interrogazione che trova le regioni che hanno una percentuale di tamponi positivi, come calcolata in risposta alla domanda precedente, superiore alla percentuale media nazionale (cioè, se la media nazionale è il 27%, l'interrogazione deve restituire le regioni che hanno una percentuale superiore al 27%)

Domanda 4 (35% per la prova breve e 20% per la prova lunga)

Considerare lo schema concettuale seguente:



Valutare la convenienza dell'introduzione di una relationship U-R fra Utente e Regione. Tale relationship è ridondante in quanto derivabile dalla concatenazione delle relationship esistenti. Considerare un carico applicativo che includa come operazioni principali le seguenti:

1. inserimento di una nuova occorrenza della relationship U-C, dati i codici identificativi di Utente e Città, assumendo per semplicità che le occorrenze di Utente e Città siano già nella base di dati, con frequenza $f_1 = 10.000$
2. ricerca della Regione di un Utente (di cui è dato il codice), con frequenza $f_2 = 100$
3. ricerca della Città di un Utente (di cui è dato il codice), con frequenza $f_3 = 1.000$

Considerare i costi delle letture e scritture delle entità, ignorando quelli delle relationship e considerare il costo delle scritture triplo di quello delle letture.

Basi di dati I — 2 febbraio 2021 — Prova breve — Compito A

Possibili soluzioni

Tempo a disposizione: un'ora

Cognome: _____ **Nome:** _____ **Matricola:** _____

Basi di dati I — 2 febbraio 2021 — Compito A

Domanda 1 (35% per la prova breve e 20% per la prova lunga)

Considerare la relazione seguente

CF	NomePaz	CT	Data	Esito	CodASL	ASL	SedeASL	IDReg	NomeReg
RSSMRI...	M. Rossi	1	12/12/20	Pos	1	RM 1	Via Po	LAZ	Lazio
RSSMRI...	M. Rossi	2	12/01/21	Neg	1	RM 1	Via Po	LAZ	Lazio
RSSMNL...	M. Rossi	3	15/12/20	Dub	1	RM 1	Via Po	LAZ	Lazio
RSSMNL...	M. Rossi	4	15/12/20	Neg	1	RM 1	Via Po	LAZ	Lazio
BRNLGU...	L. Bruni	5	12/01/21	Pos	3	LT	Via Italia	LAZ	Lazio
BNILCA ...	L. Bini	6	12/12/20	Neg	1	Ma 1	Via Roma	MAR	Marche
VRDPRI ...	P. Verdi	7	15/12/20	Pos	1	Mo 1	Corso ...	MOL	Molise

La relazione mostra (in forma non normalizzata) i dati relativi ad un insieme di tamponi eseguiti su pazienti, secondo le seguenti specifiche:

- ogni tampone viene eseguito su un paziente; ogni tampone ha un codice (CT), univoco, una data e un esito
- ogni paziente ha un codice fiscale e un nome e appartiene ad una ASL
- ogni ASL si trova in una regione, ha un codice (CodASL) che la identifica nell'ambito della regione e ha una sede
- ogni regione ha un identificatore (IDReg) e un nome

Con riferimento alle specifiche e ai dati nell'esempio

1. mostrare le dipendenze funzionali rilevate (limitarsi a quelle che hanno a primo membro ID, codici e numeri) e la chiave della relazione

Chiave:

CT

Dipendenze funzionali:

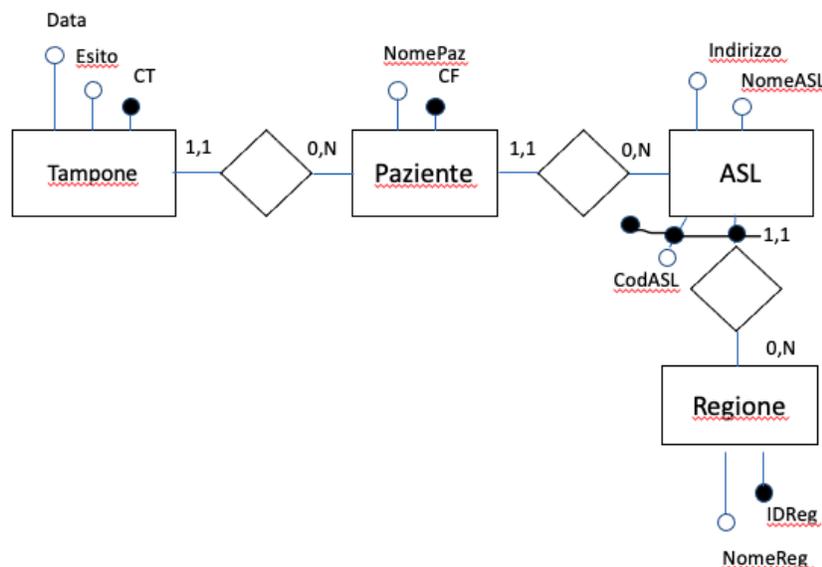
CF → NomePaz, CodASL, IDReg

CT → CF, Data, Esito

CodASL, IDReg → ASL, SedeASL

IDReg → NomeReg

2. mostrare uno schema concettuale per la realtà di interesse (rispettando le specifiche per i vari codici e senza aggiungere altri attributi)



Basi di dati I — 2 febbraio 2021 — Compito A

3. mostrare una buona decomposizione della relazione originaria che soddisfi la BCNF (mostrare le tabelle, con i dati, indicando la chiave di ciascuna con la sottolineatura)

<u>CF</u>	NomePaz	CodASL	IDReg
RSSMRI...	M. Rossi	1	LAZ
RSSMNL...	M. Rossi	1	LAZ
BRNLGU...	L. Bruni	3	LAZ
BNILCA ...	L. Bini	1	MAR
VRDPRI ...	P. Verdi	1	MOL

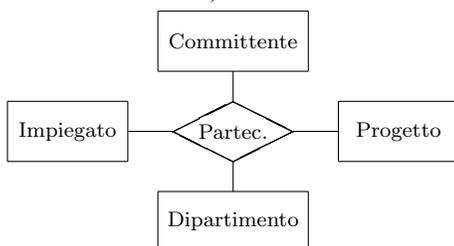
CF	<u>CT</u>	Data	Esito
RSSMRI...	1	12/12/20	Pos
RSSMRI...	2	12/01/21	Neg
RSSMNL...	3	15/12/20	Dub
RSSMNL...	4	15/12/20	Neg
BRNLGU...	5	12/01/21	Pos
BNILCA ...	6	12/12/20	Neg
VRDPRI ...	7	15/12/20	Pos

<u>CodASL</u>	ASL	SedeASL	<u>IDReg</u>
1	RM 1	Via Po	LAZ
3	LT	Via Italia	LAZ
1	Ma 1	Via Roma	MAR
1	Mo 1	Corso ...	MOL

<u>IDReg</u>	NomeReg
LAZ	Lazio
MAR	Marche
MOL	Molise

Domanda 2 (30% per la prova breve)

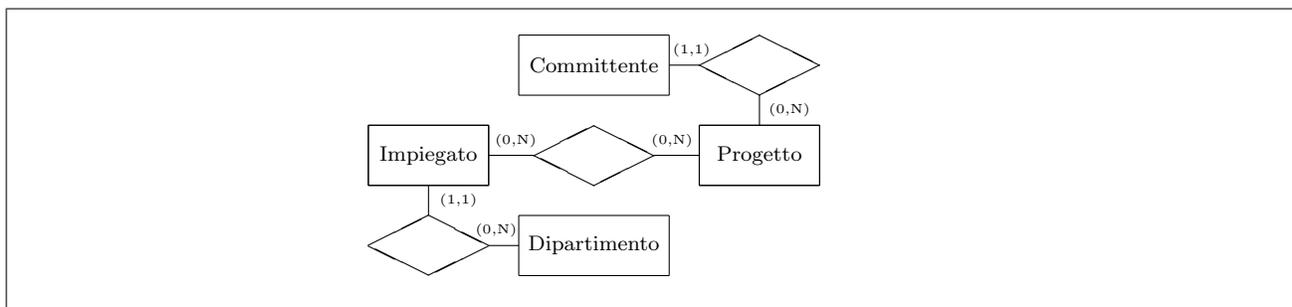
A seguito di una prima, superficiale analisi di una realtà di interesse, è stato prodotto lo schema in figura (gli impiegati di un'azienda di servizi, che lavorano nei dipartimenti nei quali è organizzata l'azienda, partecipano a progetti commissionati da committenti esterni):



Modificare lo schema (decomponendo la relationship e aggiungendo ulteriori entità, se necessario) tenendo conto delle seguenti specifiche:

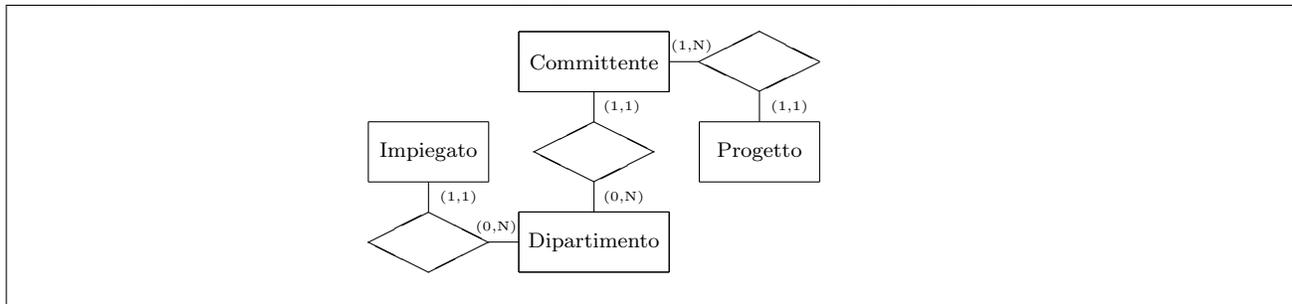
- ogni impiegato afferisce ad uno e un solo dipartimento;
- ogni progetto ha zero o più committenti e ogni committente uno e un solo progetto
- ogni impiegato lavora per zero o più progetti (e quindi per tutti e soli i relativi committenti)
- ogni dipartimento ha come committenti e progetti quelli dei propri impiegati

Si noti che sicuramente ognuna delle entità ha attributi, che non vengono mostrati e non vanno inseriti nella soluzione.



Ripetere quanto fatto sopra con riferimento alle seguenti specifiche:

- ogni impiegato afferisce ad uno e un solo dipartimento;
- ogni progetto ha uno e un solo committente e ogni committente uno o più progetti
- ogni dipartimento ha zero o più committenti e segue tutti e soli i relativi progetti; ogni committente interagisce con uno e un solo dipartimento
- ogni impiegato lavora per tutti e soli i progetti dei committenti del proprio dipartimento



Domanda 3 (35% per la prova breve e 20% per la prova lunga)

Considerare lo schema concettuale seguente:



Valutare la convenienza dell'introduzione di una relationship U-R fra Utente e Regione. Tale relationship è ridondante in quanto derivabile dalla concatenazione delle relationship esistenti. Considerare un carico applicativo che includa come operazioni principali le seguenti:

1. inserimento di una nuova occorrenza della relationship U-C, dati i codici identificativi di Utente e Città, assumendo per semplicità che le occorrenze di Utente e Città siano già nella base di dati, con frequenza $f_1 = 100$
2. ricerca della Regione di un Utente (di cui è dato il codice), con frequenza $f_2 = 10.000$
3. ricerca della Città di un Utente (di cui è dato il codice), con frequenza $f_3 = 1.000$

Considerare i costi delle letture e scritture delle entità, ignorando quelli delle relationship e considerare il costo delle scritture triplo di quello delle letture.

Cenni sulla soluzione:

la decisione deve essere basata sul confronto fra il costo complessivo nei due casi (i) introduzione della relationship ridondante e (ii) mantenimento dello schema dato, senza relationship ridondante.

In dettagli il costo va calcolato come somma dei costi delle singole operazioni, ciascuna moltiplicata per la relativa frequenza:

$$c_1 \times f_1 + c_2 \times f_2 + c_2 \times f_3$$

Le frequenze f_i sono date e sono le stesse nei due casi, mentre i costi unitari differiscono per le prime due operazioni (non per la terza, che non è influenzata dalla eventuale ridondanza).

I costi unitari sono calcolabili solo in modo approssimato e schematizzato, quindi non esiste per essi una risposta univoca. Comunque:

- per l'operazione 1 il costo con ridondanza è superiore a quello senza ridondanza (perché per inserire il valore di U-R è necessario leggere la città associata all'utente)
- per l'operazione 2, il costo con ridondanza è inferiore a quello senza ridondanza (perché non serve leggere la città)

Le due differenze sono paragonabili (differenza di una o due unità in ciascuno dei casi), per cui la scelta dipende dalle frequenze f_1 e f_2 . Nei compiti con valore di f_1 molto maggiore di f_2 , la ridondanza non conviene, negli altri conviene

Basi di dati I — 2 febbraio 2021 — Prova breve — Compito B

Possibili soluzioni

Tempo a disposizione: un'ora

Cognome: _____ Nome: _____ Matricola: _____

Basi di dati I — 2 febbraio 2021 — Compito B

Domanda 1 (35% per la prova breve e 20% per la prova lunga)

Considerare la relazione seguente

CF	NomePaz	CT	Data	Esito	CodASL	ASL	SedeASL	IDReg	NomeReg
RSSMRI...	M. Rossi	1	12/12/20	Pos	1	RM 1	Via Po	LAZ	Lazio
RSSMRI...	M. Rossi	2	12/01/21	Neg	1	RM 1	Via Po	LAZ	Lazio
RSSMNL...	M. Rossi	1	15/12/20	Dub	1	RM 1	Via Po	LAZ	Lazio
RSSMNL...	M. Rossi	2	15/12/20	Neg	1	RM 1	Via Po	LAZ	Lazio
BRNLGU...	L. Bruni	1	12/01/21	Pos	2	LT	Via Italia	LAZ	Lazio
BNILCA ...	L. Bini	1	12/12/20	Neg	3	Ma 1	Via Roma	MAR	Marche
VRDPRI ...	P. Verdi	1	15/12/20	Pos	4	Mo 1	Corso ...	MOL	Molise

La relazione mostra (in forma non normalizzata) i dati relativi ad un insieme di tamponi eseguiti su pazienti, secondo le seguenti specifiche:

- ogni tampone viene eseguito su un paziente; ogni tampone ha un codice (CT), univoco fra i tamponi di quel paziente, una data e un esito
- ogni paziente ha un codice fiscale e un nome e appartiene ad una ASL
- ogni ASL si trova in una regione, ha un codice (CodASL) che la identifica in modo univoco e ha una sede
- ogni regione ha un identificatore (IDReg) e un nome

Con riferimento alle specifiche e ai dati nell'esempio

1. mostrare le dipendenze funzionali rilevate (limitarsi a quelle che hanno a primo membro ID, codici e numeri) e la chiave della relazione

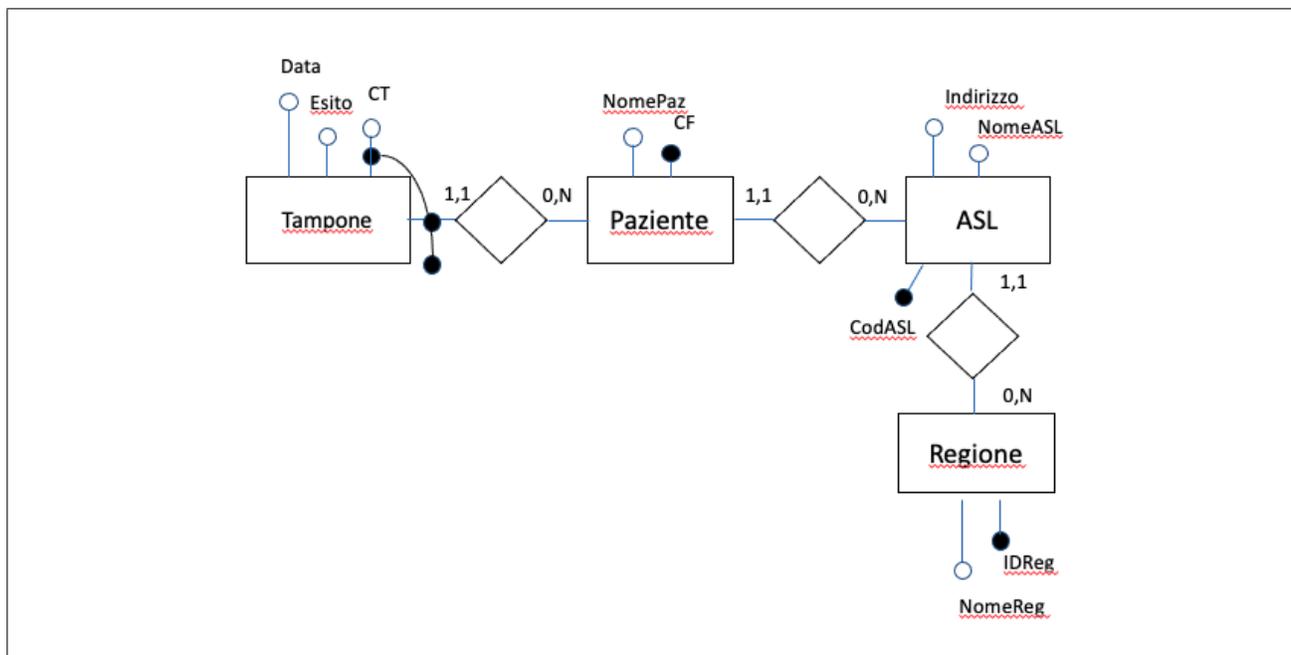
Chiave:

CF, CT

Dipendenze funzionali:

CF → NomePaz, CodASL
 CF, CT → Data, Esito
 CodASL → ASL, SedeASL, IDReg
 IDReg → NomeReg

2. mostrare uno schema concettuale per la realtà di interesse (rispettando le specifiche per i vari codici e senza aggiungere altri attributi)



Basi di dati I — 2 febbraio 2021 — Compito B

3. mostrare una buona decomposizione della relazione originaria che soddisfi la BCNF (mostrare le tabelle, con i dati, indicando la chiave di ciascuna con la sottolineatura)

<u>CF</u>	NomePaz	CodASL
RSSMRI...	M. Rossi	1
RSSMNL...	M. Rossi	1
BRNLGU...	L. Bruni	3
BNILCA ...	L. Bini	1
VRDPRI ...	P. Verdi	1

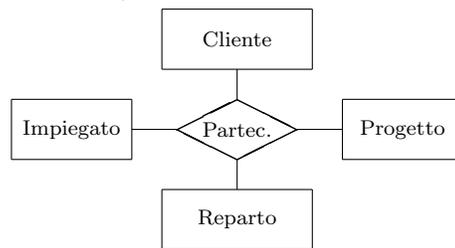
<u>CF</u>	<u>CT</u>	Data	Esito
RSSMRI...	1	12/12/20	Pos
RSSMRI...	2	12/01/21	Neg
RSSMNL...	3	15/12/20	Dub
RSSMNL...	4	15/12/20	Neg
BRNLGU...	5	12/01/21	Pos
BNILCA ...	6	12/12/20	Neg
VRDPRI ...	7	15/12/20	Pos

<u>CodASL</u>	ASL	SedeASL	IDReg
1	RM 1	Via Po	LAZ
3	LT	Via Italia	LAZ
1	Ma 1	Via Roma	MAR
1	Mo 1	Corso ...	MOL

<u>IDReg</u>	NomeReg
LAZ	Lazio
MAR	Marche
MOL	Molise

Domanda 2 (30% per la prova breve)

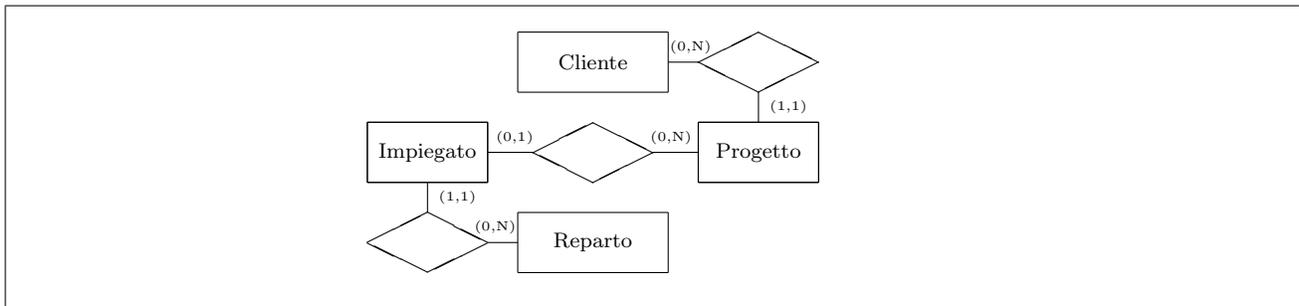
A seguito di una prima, superficiale analisi di una realtà di interesse, è stato prodotto lo schema in figura (gli impiegati di un'azienda di servizi, che lavorano nei reparti nei quali è organizzata l'azienda, partecipano a progetti commissionati da clienti esterni):



Modificare lo schema (decomponendo la relationship e aggiungendo ulteriori entità, se necessario) tenendo conto delle seguenti specifiche:

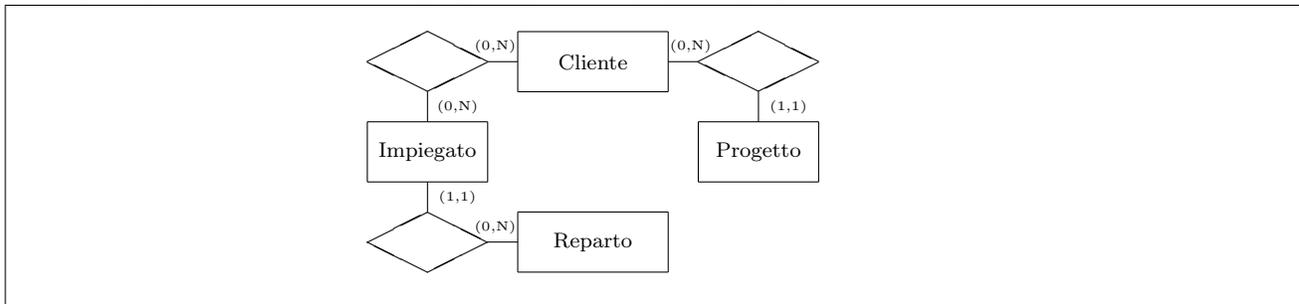
- ogni impiegato afferisce ad uno e un solo reparto;
- ogni impiegato partecipa ad un solo progetto (che non è necessariamente legato ad un reparto, cioè impiegati di reparti diversi possono lavorare sul medesimo progetto);
- ogni progetto ha un solo cliente e ogni impiegato interagisce solo con il cliente del proprio progetto.

Si noti che sicuramente ognuna delle entità ha attributi, che non vengono mostrati e non vanno inseriti nella soluzione.



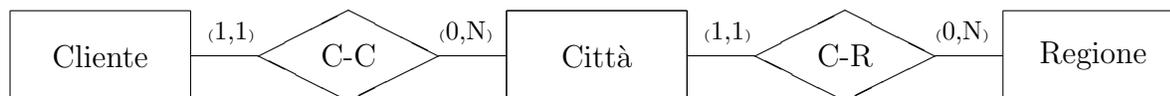
Ripetere quanto fatto sopra con riferimento alle seguenti specifiche:

- ogni impiegato afferisce ad uno e un solo reparto;
- ogni progetto ha uno e un solo cliente
- ogni impiegato lavora per zero o più clienti e potenzialmente per tutti i progetti di quei clienti (e per nessun altro progetto)
- ogni reparto ha come clienti e progetti quelli dei propri impiegati



Domanda 3 (35% per la prova breve e 20% per la prova lunga)

Considerare lo schema concettuale seguente:



Valutare la convenienza dell'introduzione di una relationship U-R fra Cliente e Regione. Tale relationship è ridondante in quanto derivabile dalla concatenazione delle relationship esistenti. Considerare un carico applicativo che includa come operazioni principali le seguenti:

1. inserimento di una nuova occorrenza della relationship C-C, dati i codici identificativi di Cliente e Città, assumendo per semplicità che le occorrenze di Cliente e Città siano già nella base di dati, con frequenza $f_1 = 100$
2. ricerca della Regione di un Cliente (di cui è dato il codice), con frequenza $f_2 = 10.000$
3. ricerca della Città di un Cliente (di cui è dato il codice), con frequenza $f_3 = 1.000$

Considerare i costi delle letture e scritture delle entità, ignorando quelli delle relationship e considerare il costo delle scritture triplo di quello delle letture.

Cenni sulla soluzione:

la decisione deve essere basata sul confronto fra il costo complessivo nei due casi (i) introduzione della relationship ridondante e (ii) mantenimento dello schema dato, senza relationship ridondante.

In dettagli il costo va calcolato come somma dei costi delle singole operazioni, ciascuna moltiplicata per la relativa frequenza:

$$c_1 \times f_1 + c_2 \times f_2 + c_2 \times f_3$$

Le frequenze f_i sono date e sono le stesse nei due casi, mentre i costi unitari differiscono per le prime due operazioni (non per la terza, che non è influenzata dalla eventuale ridondanza).

I costi unitari sono calcolabili solo in modo approssimato e schematizzato, quindi non esiste per essi una risposta univoca. Comunque:

- per l'operazione 1 il costo con ridondanza è superiore a quello senza ridondanza (perché per inserire il valore di U-R è necessario leggere la città associata all'utente)
- per l'operazione 2, il costo con ridondanza è inferiore a quello senza ridondanza (perché non serve leggere la città)

Le due differenze sono paragonabili (differenza di una o due unità in ciascuno dei casi), per cui la scelta dipende dalle frequenze f_1 e f_2 . Nei compiti con valore di f_1 molto maggiore di f_2 , la ridondanza non conviene, negli altri conviene

Basi di dati I — 2 febbraio 2021 — Prova lunga — Compito E

Possibili soluzioni

Tempo a disposizione: un'ora e quarantacinque minuti

Cognome: _____ Nome: _____ Matricola: _____

Basi di dati I — 2 febbraio 2021 — Compito E

Domanda 1 (35% per la prova breve e 20% per la prova lunga)

Considerare la relazione seguente

CF	NomePaz	CT	Data	Esito	CodASL	ASL	SedeASL	IDReg	NomeReg
RSSMRI...	M. Rossi	1	12/12/20	Pos	1	RM 1	Via Po	LAZ	Lazio
RSSMRI...	M. Rossi	2	12/01/21	Neg	1	RM 1	Via Po	LAZ	Lazio
RSSMNL...	M. Rossi	3	15/12/20	Dub	1	RM 1	Via Po	LAZ	Lazio
RSSMNL...	M. Rossi	4	15/12/20	Neg	1	RM 1	Via Po	LAZ	Lazio
BRNLGU...	L. Bruni	5	12/01/21	Pos	3	LT	Via Italia	LAZ	Lazio
BNILCA ...	L. Bini	6	12/12/20	Neg	1	Ma 1	Via Roma	MAR	Marche
VRDPRI ...	P. Verdi	7	15/12/20	Pos	1	Mo 1	Corso ...	MOL	Molise

La relazione mostra (in forma non normalizzata) i dati relativi ad un insieme di tamponi eseguiti su pazienti, secondo le seguenti specifiche:

- ogni tampone viene eseguito su un paziente; ogni tampone ha un codice (CT), univoco, una data e un esito
- ogni paziente ha un codice fiscale e un nome e appartiene ad una ASL
- ogni ASL si trova in una regione, ha un codice (CodASL) che la identifica nell'ambito della regione e ha una sede
- ogni regione ha un identificatore (IDReg) e un nome

Con riferimento alle specifiche e ai dati nell'esempio

1. mostrare le dipendenze funzionali rilevate (limitarsi a quelle che hanno a primo membro ID, codici e numeri) e la chiave della relazione

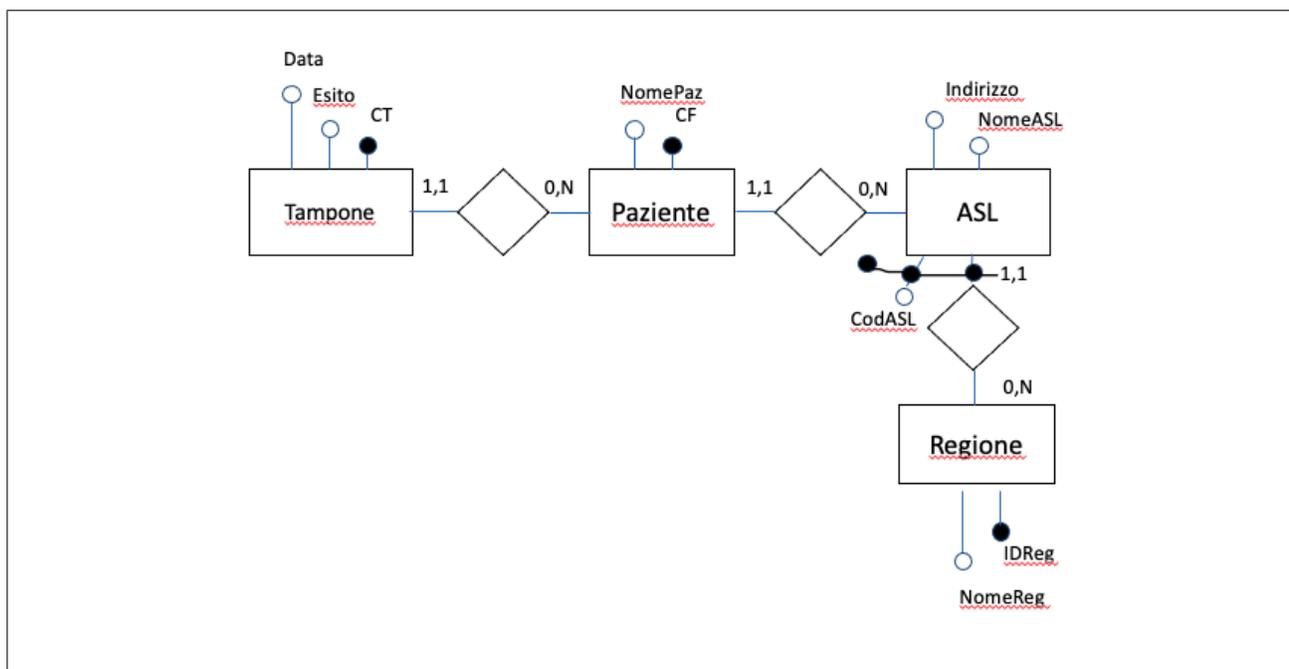
Chiave:

CT

Dipendenze funzionali:

CF → NomePaz, CodASL, IDReg
 CT → CF, Data, Esito
 CodASL, IDReg → ASL, SedeASL
 IDReg → NomeReg

2. mostrare uno schema concettuale per la realtà di interesse (rispettando le specifiche per i vari codici e senza aggiungere altri attributi)



Basi di dati I — 2 febbraio 2021 — Compito E

3. mostrare una buona decomposizione della relazione originaria che soddisfi la BCNF (mostrare le tabelle, con i dati, indicando la chiave di ciascuna con la sottolineatura)

<u>CF</u>	NomePaz	CodASL	IDReg
RSSMRI...	M. Rossi	1	LAZ
RSSMNL...	M. Rossi	1	LAZ
BRNLGU...	L. Bruni	3	LAZ
BNILCA ...	L. Bini	1	MAR
VRDPRI ...	P. Verdi	1	MOL

CF	<u>CT</u>	Data	Esito
RSSMRI...	1	12/12/20	Pos
RSSMRI...	2	12/01/21	Neg
RSSMNL...	3	15/12/20	Dub
RSSMNL...	4	15/12/20	Neg
BRNLGU...	5	12/01/21	Pos
BNILCA ...	6	12/12/20	Neg
VRDPRI ...	7	15/12/20	Pos

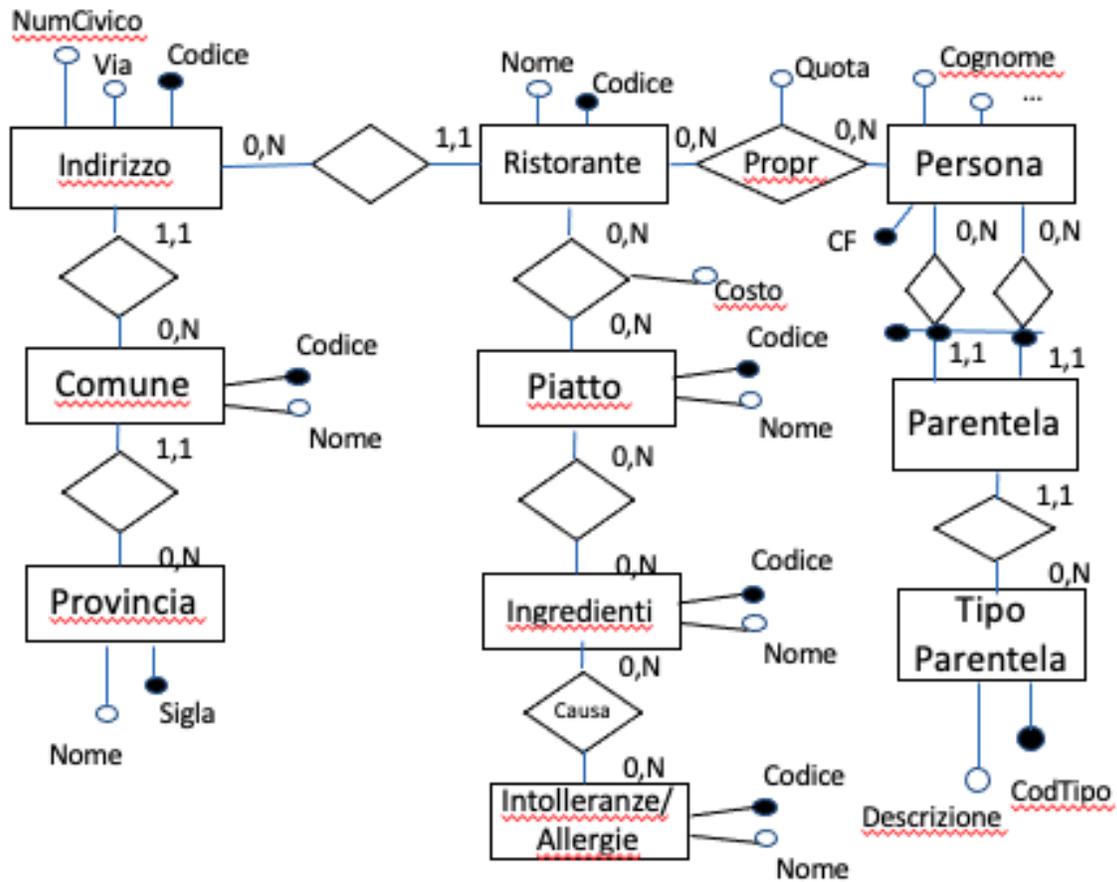
<u>CodASL</u>	ASL	SedeASL	<u>IDReg</u>
1	RM 1	Via Po	LAZ
3	LT	Via Italia	LAZ
1	Ma 1	Via Roma	MAR
1	Mo 1	Corso ...	MOL

<u>IDReg</u>	NomeReg
LAZ	Lazio
MAR	Marche
MOL	Molise

Domanda 2 (30% per la prova completa)

Mostrare, nel riquadro alla pagine seguente, uno schema concettuale per una realtà relativa ad un insieme di ristoranti di una regione, secondo le specifiche seguenti (molto semplificate rispetto alla realtà, ma con alcuni dettagli che vanno specificati come richiesto).

- Ogni ristorante ha codice, nome e indirizzo; gli indirizzi sono “standardizzati”, nel senso che la base di dati contiene tutti i possibili indirizzi della regione, ciascuno con via, numero civico, comune (con codice e nome) e provincia (con codice e nome)
- Per ogni ristorante (ad esempio per contrastare le infiltrazioni della criminalità organizzata) si rappresentano i proprietari, con codice fiscale, nome, cognome, data di nascita, quota di possesso del ristorante (ad esempio Mario Rossi può possedere il 30% del ristorante Buon Gusto) e parentele eventualmente esistenti tra proprietari (anche di ristoranti diversi), con tipo di parentela (ad esempio Mario Rossi è cugino di Aldo Verdi e marito di Anna Pini). Le parentele sono rappresentate con codice e descrizione (ad esempio “A” per fratello/sorella; “B” per genitore/figlio)
- Ogni ristorante ha un menu, composto da un insieme di piatti offerti (ad es. “spaghetti alla carbonara,” “agnello al forno”). I piatti sono caratterizzati da un codice (unico per piatto per tutti i ristoranti) e un nome. Ogni piatto ha un costo, che varia e dipende dal ristorante, mentre il nome del piatto è identico per tutti i ristoranti.
- Ogni piatto ha un insieme di ingredienti, che sono descritti ciascuno da un codice e un nome (unici per tutti i ristoranti).
- Si vuole tenere traccia delle possibili allergie o intolleranze alimentari, indicando, per ciascun ingrediente quali sono le intolleranze o allergie cui è associato (ad esempio, la panna è associata all’intolleranza al lattosio). Nota bene, il tutto è molto semplificato e si desidera solo rappresentare intolleranze/allergie (con codice e nome) e associarle agli ingredienti.



Basi di dati I — 2 febbraio 2021 — Compito E

Domanda 3 (30%) Considerare una base di dati sul seguente schema

- PROVINCE(Sigla, Nome, Abitanti, Regione); contiene dati su tutte le province d'Italia, una ennupla per provincia, con nome, abitanti e regione di appartenenza
- TAMPONI(Provincia, Data, NumeroTamponi, NumeriPositivi), con vincolo di integrità referenziale fra Provincia e la chiave Sigla di PROVINCE; ogni ennupla contiene dati sui tamponi effettuati in un giorno in una certa provincia; i dati sono acquisiti gradualmente e quindi può succedere che in un certo momento i dati relativi ad una certa giornata possano essere disponibili solo per una parte delle province

Con riferimento a tale schema

1. Formulare in algebra relazionale l'interrogazione che trova le province (basta indicare la sigla) per le quali non è disponibile l'informazione sui tamponi del 31/01/2021

```
 $\pi_{\text{Sigla}}(\text{Province}) - \pi_{\text{Sigla}}(\text{Province} \bowtie_{\text{Sigla} = \text{Provincia}} \sigma_{\text{Data} = '31/01/2021'}(\text{Tamponi}))$ 
```

2. Formulare in algebra relazionale l'interrogazione che trova le regioni per le quali l'informazione sui tamponi del 31/01/2021 è disponibile per tutte le province

```
 $\pi_{\text{Regione}}(\text{Province})$   
-  
 $\underline{(\pi_{\text{Regione}}(\pi_{\text{Sigla}, \text{Regione}}(\text{Province}))$   
-  
 $\pi_{\text{Sigla}, \text{Regione}}(\text{Province} \bowtie_{\text{Sigla} = \text{Provincia}} \sigma_{\text{Data} = '31/01/2021'}(\text{Tamponi})))}$ 
```

3. Formulare in SQL l'interrogazione che, per ogni provincia, mostra: la sigla, il nome, il numero di tamponi effettuati e il numero di positivi per il giorno 25/01/2021

```
SELECT Sigla, Nome, NumeroTamponi, NumeroPositivi  
FROM Province JOIN Tamponi ON Sigla = Provincia  
WHERE Data = '2021-01-25'
```

Basi di dati I — 2 febbraio 2021 — Compito E

4. Formulare in SQL l'interrogazione che trova, per ciascuna regione, il numero totale di tamponi effettuati, il numero totale di tamponi positivi e la percentuale di positivi (dove la percentuale è pari al totale dei positivi moltiplicato per cento diviso il totale di tamponi effettuati)

soluzione per i compiti E e G, gli altri sono simili; l'indicazione del tipo dei risultati ("`::Numeric(10,2)`") non era rilevante ai fini della valutazione

```
SELECT Regione, SUM(NumeroTamponi) AS TotaleTamponi,
       SUM(NumeroPositivi) AS TotalePositivi,
       (SUM(NumeroPositivi)::Numeric(10,2)/SUM(NumeroTamponi)*100)::Numeric(5,2) AS PercPositivi
FROM Province JOIN Tamponi ON Sigla = Provincia
GROUP BY Regione;
```

5. Formulare in SQL l'interrogazione che trova le regioni che hanno una percentuale di tamponi positivi, come calcolata in risposta alla domanda precedente, superiore alla percentuale media nazionale (cioè, se la media nazionale è il 27%, l'interrogazione deve restituire le regioni che hanno una percentuale superiore al 27%)

```
CREATE VIEW DatiPerRegione AS
SELECT Regione, SUM(NumeroTamponi) AS TotaleTamponi,
       SUM(NumeroPositivi) AS TotalePositivi,
       (SUM(NumeroPositivi)::Numeric(10,2)/SUM(NumeroTamponi)*100)::Numeric(5,2) AS PercPositivi
FROM Province JOIN Tamponi ON Sigla = Provincia
GROUP BY Regione;

CREATE VIEW MediaNazionale AS
SELECT AVG(PercPositivi)::Numeric(5,2) AS MediaPercPositivi
FROM DatiPerRegione;

SELECT *
FROM DatiPerRegione, MediaNazionale
WHERE PercPositivi > MediaPercPositivi;
```

Domanda 4 (35% per la prova breve e 20% per la prova lunga)

Considerare lo schema concettuale seguente:



Valutare la convenienza dell'introduzione di una relationship U-R fra Utente e Regione. Tale relationship è ridondante in quanto derivabile dalla concatenazione delle relationship esistenti. Considerare un carico applicativo che includa come operazioni principali le seguenti:

1. inserimento di una nuova occorrenza della relationship U-C, dati i codici identificativi di Utente e Città, assumendo per semplicità che le occorrenze di Utente e Città siano già nella base di dati, con frequenza $f_1 = 10.000$
2. ricerca della Regione di un Utente (di cui è dato il codice), con frequenza $f_2 = 100$
3. ricerca della Città di un Utente (di cui è dato il codice), con frequenza $f_3 = 1.000$

Considerare i costi delle letture e scritture delle entità, ignorando quelli delle relationship e considerare il costo delle scritture triplo di quello delle letture.

Cenni sulla soluzione:

la decisione deve essere basata sul confronto fra il costo complessivo nei due casi (i) introduzione della relationship ridondante e (ii) mantenimento dello schema dato, senza relationship ridondante.

In dettagli il costo va calcolato come somma dei costi delle singole operazioni, ciascuna moltiplicata per la relativa frequenza:

$$c_1 \times f_1 + c_2 \times f_2 + c_2 \times f_3$$

Le frequenze f_i sono date e sono le stesse nei due casi, mentre i costi unitari differiscono per le prime due operazioni (non per la terza, che non è influenzata dalla eventuale ridondanza).

I costi unitari sono calcolabili solo in modo approssimato e schematizzato, quindi non esiste per essi una risposta univoca. Comunque:

- per l'operazione 1 il costo con ridondanza è superiore a quello senza ridondanza (perché per inserire il valore di U-R è necessario leggere la città associata all'utente)
- per l'operazione 2, il costo con ridondanza è inferiore a quello senza ridondanza (perché non serve leggere la città)

Le due differenze sono paragonabili (differenza di una o due unità in ciascuno dei casi), per cui la scelta dipende dalle frequenze f_1 e f_2 . Nei compiti con valore di f_1 molto maggiore di f_2 , la ridondanza non conviene, negli altri conviene