



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Corso di Laurea in Informatica

Esame di Basi di Dati

Esame del 1 Luglio 2022

Regole dell'esame:

- Non è possibile utilizzare alcun materiale, né appunti, né il libro.
- **Inserire le risposte nei riquadri che seguono i testi degli esercizi**
- L'uso della matita per mostrare la soluzione degli esercizi è fatta a rischio e pericolo degli studenti. Nessuna rivendicazione verrà accettata in caso parte della soluzione si cancelli, per via delle caratteristiche transitorie dei tratti a matita.
- Il massimo possibile è 32 punti
- Occorre avere acquisito almeno metà dei punti dell'Esercizio 3 (cioè 3.5 punti).
- **Coloro che siano sorpresi a copiare o a far copiare:**
 - Dovranno lasciare l'aula dell'esame
 - Riceveranno 0 punti "di ufficio"
 - Dovranno saltare l'appello successivo

Nome: _____ Num. Matric.: _____

Esercizio 1: Diagramma ER (8 punti)

Si richiede di produrre lo schema concettuale Entità-Relazione di un'applicazione relativa a ricercatori biologi e i loro esperimenti. Disegnare il diagramma ER nel riquadro della pagina che segue.

Di ogni ricercatore interessa il codice fiscale (identificativo), la data di nascita, la nazione di nascita, gli esperimenti effettuati ed i laboratori in cui ha lavorato (almeno uno), con data di inizio e data di fine rapporto (ovviamente, quest'ultima informazione manca per il laboratorio in cui il ricercatore lavora attualmente) e stipendio medio ottenuto nel periodo del rapporto stesso.

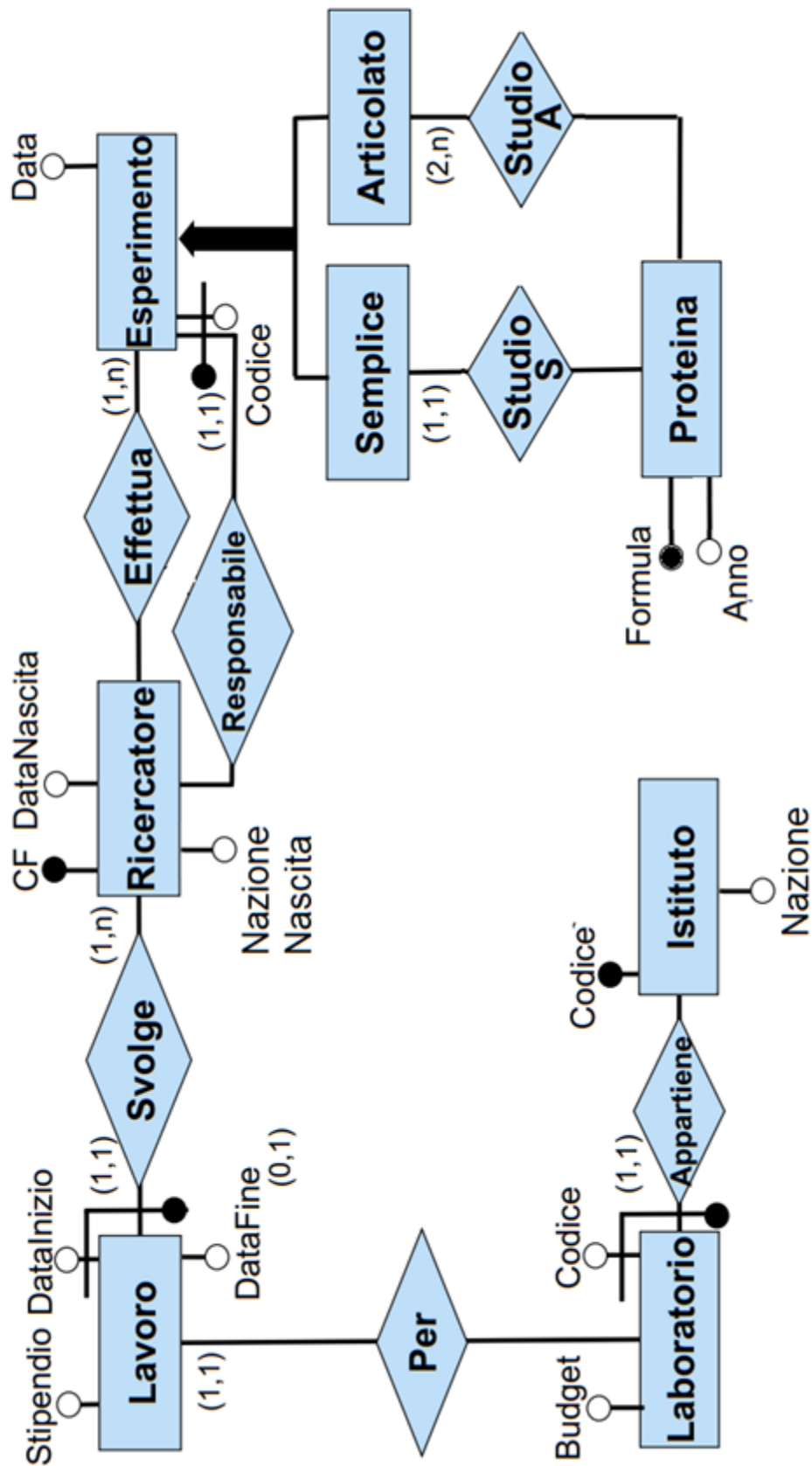
Si noti che un ricercatore non può lavorare contemporaneamente in più di un laboratorio e che non ci sono limiti al numero di rapporti che un ricercatore può avere con i singoli laboratori. Di ogni laboratorio interessa l'istituto a cui appartiene, il codice (unico nell'ambito dell'istituto a cui appartiene) ed il budget attuale. Di ogni istituto interessa il codice (identificativo) e la nazione in cui si trova.

Di ogni esperimento interessano i ricercatori che lo hanno effettuato, tra questi quello che ne è responsabile, il codice (unico per il ricercatore responsabile) e la data in cui si è svolto.

Esistono esattamente due tipi di esperimenti: semplici ed articolati. Ogni esperimento semplice studia singole proteine. Ogni esperimento articolato studia l'interazione tra almeno due proteine.

Ogni proteina è caratterizzata dall'anno in cui è stata scoperta, e dalla sua formula chimica, univoca per una data proteina.

Soluzione Esercizio 1



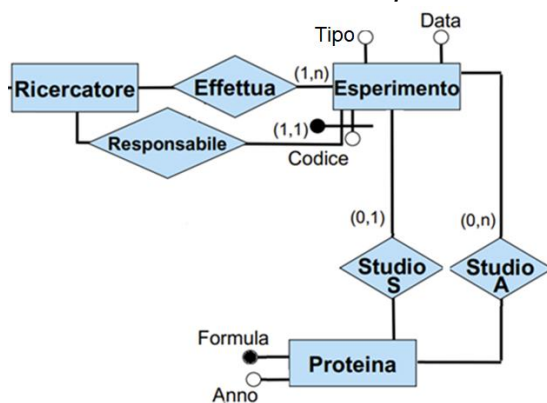
Cardinalità (0,n) sono omesse

Esercizio 2: Progettazione Logica (7 punti)

A partire dallo ER dell'Esercizio 1, produrre uno schema relazionale del database nel riquadro sottostante che minimizza il numero di tabelle. Indicare i vincoli e gli attributi che ammettono valori nulli. Illustrare come ristrutturare il diagramma ER per essere direttamente traducibile in uno schema relazionale.¹

Soluzione Esercizio 2

La fase della ristrutturazione si focalizza sulla eliminazione dei costrutti ER non esprimibili in uno schema relazionale, come attributi multi-valore e generalizzazioni. Nello specifico, occorre eliminare una generalizzazione. Siccome l'obiettivo è minimizzare il numero delle tabelle, le entità *Semplice* e *Articolato* sono accorpate nel padre, con le relazioni



collegate. Per differenziare le istanze di un esperimento semplice da quelle di esperimenti articolati, si aggiunge attributo *Tipo*, che può idealmente assumere due valori: *Semplice* e *Articolato*.

Nell'accorpamento, le relazioni *StudioS* e *StudioA* hanno cardinalità minima uguale a 0, perché, se l'esperimento è semplice, si popolerà la prima, altrimenti, se è articolato, si popolerà la seconda. La figura a lato mostra la ristrutturazione del diagramma ER per la parte rilevante, con alcuni attributi omessi per

leggibilità. Sebbene non in linea con la metodologia, è anche ammissibile accorpare *StudioS* e *StudioA* in una sola relazione *Studio* con cardinalità $(1,n)$ dal lato di *Esperimento*.

Lo schema logico è quindi come segue, dove l'asterico segue gli attributi che ammettono valori nulli:

PROTEINA(Formula,Anno)

RICERCATORE(CF,DataNascita,NazioneNascita)

ESPERIMENTO(Codice,Responsabile>Data,Tipo,Proteina*)

Esperimento.Responsabile→Ricercatore.CF

Esperimento.Proteina→Proteina.Formula

EFFETTUA(Ricercatore,CodiceEsp,ResponsabEsp)

Effettua.Ricercatore→Ricercatore.CF

Effettua.(CodiceEsp,ResponsabEsp)→Esperimento.(Codice,Responsabile)

STUDIOA(Proteina,CodiceEsp,ResponsabEsp)

Effettua.(CodiceEsp,ResponsabEsp)→Esperimento.(Codice,Responsabile)

Effettua.Proteina→Proteina.Formula

ISTITUTO(Codice, Nazione)

LABORATORIO(Codice,Istituto,Budget)

Laboratorio.Istituto→Istituto.Codice

LAVORO(Ricercatore,DataInizio,LabCodice,LabIstituto,Stipendio,DataFine*)

Lavoro.Ricercatore→Ricercatore.CF

Lavoro.(LabCodice,LabIstituto)→Laboratorio.(Codice,Istituto)

¹ Allo scopo di mostrare la ristrutturazione del diagramma ER, è possibile semplicemente mostrare i cambiamenti apportati al diagramma nel riquadro soluzione dell'Esercizio 1, utilizzando una penna di diverso colore (non rossa!)

Esercizio 3: Algebra Relazionale & SQL (7 punti)

Si consideri la seguente base di dati con le relazioni:

- **Diplomato(codice, voto, codScuola, anno)**, che per ogni studente diplomato memorizza il codice, il voto conseguito alla maturità, l'anno di conseguimento e la scuola in cui l'ha conseguita.
- **Scuola(codScuola, città)** che memorizza codice e città delle scuole.

A. Restituire le città in cui nessuna scuola ha avuto diplomati con 100 nel 2021 (2 punti).²

CITTA_CON_100_NEL_2021=

$\pi_{città} (\sigma_{Voto=100 \text{ AND } anno=2021} (Diplomato \bowtie Scuola))$

$\pi_{città} (SCUOLA) - CITTA_CON_100_NEL_2021$

² Si assuma che l'operatore di join $A \bowtie B$ **senza condizioni** mantenga le tuple di $A \times B$ con valori uguali su attributi uguali (join naturale). Se **una condizione C è specificata**, \bowtie_C mantiene le tuple di $A \times B$ per cui la condizione C è vera.

Nome: _____ Num. Matric.: _____

B. Scrivere una query in Standard SQL che, per ogni città con almeno 100 diplomati nel 2021, restituisce il voto medio nel 2021. (2.5 punti).

```
SELECT CITTA, AVG(VOTO)
FROM DIPLOMATO D, SCUOLA S
WHERE D.CODSCUOLA=S.CODSCUOLA
      AND ANNO=2021
GROUP BY CITTA
HAVING COUNT(*)>=100
```

C. Nel riquadro, scrivere una query in Standard SQL che restituisce la città con la scuola con più diplomati con 100 (2.5 punti)

```
CREATE VIEW PIU(CODSCUOLA,NUM-100) AS
SELECT CODSCUOLA, COUNT(*)
FROM DIPLOMATO D
WHERE VOTO=100
GROUP BY CODSCUOLA;

SELECT CITTA
FROM PIU P JOIN SCUOLA S ON
P.CODSCUOLA=S.CODSCUOLA
WHERE P.NUM-100 =
      (SELECT MAX(NUM-100) FROM PIU)
```

Esercizio 4: Transazioni (5 punti)

Si consideri il seguente schedule:

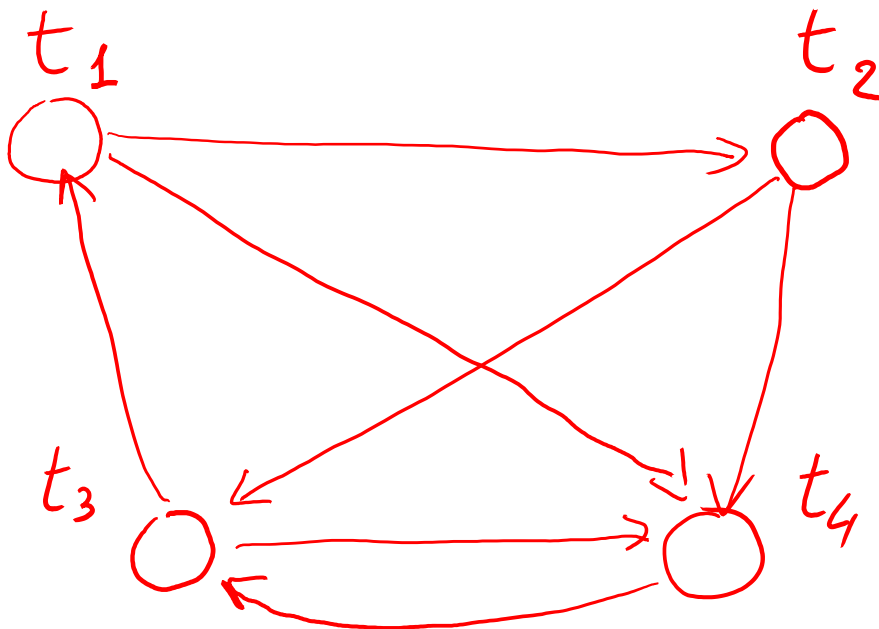
$S = r_2(v) \ r_1(x) \ w_4(v) \ w_3(y) \ w_2(x) \ r_1(y) \ w_4(y) \ w_3(v).$

Dire se S è conflict-serializable o no.

Se la risposta è sì, mostrare uno schedule seriale conflict-equivalente a S .

Se la risposta è no, dire quale transazione causa il problema e come è possibile risolvere il problema.

Il grafo dei conflitti è come seguente:



È chiaramente possibile identificare diversi cicli (t_1, t_2 e t_3) oppure (t_3 e t_4). Quindi, lo schedule non è conflict-serializzabile.

La transazione t_3 è la transazione che causa i problemi. Il problema può essere risolto posticipando la scrittura di y da parte di t_3 dopo tutte le operazioni delle altre transazioni, ma comunque prima di $w_3(v)$:

$r_2(v) \ r_1(x) \ w_4(v) \ w_2(x) \ r_1(y) \ w_4(y) \ w_3(y) \ w_3(v)$

Nome: _____ Num. Matric.: _____

Esercizio 5: Quiz (5 punti)

Rispondere alle seguenti domande, sottolineando quale risposta è corretta (solo una è corretta).

Domanda 1 (1.5 Punti)

Data la query `SELECT * FROM S WHERE A=4 ORDER BY B` sulla relazione $S(A, B, C)$. Quale dei seguenti indici in genere assicura le migliori performance in termini di velocità dell'esecuzione della query?

1. Indice B+Tree sulla coppia (B,A)
2. Indice B+Tree su B
3. Indice B+Tree su A
4. Indice B+Tree sulla coppia (A,B)

Domanda 2 (1.5 Punti)

Sia data la relazione $R(A,B,C,D)$ con dipendenze funzionali $\{A \rightarrow B, AB \rightarrow C, D \rightarrow AC\}$. Indicare se AD è chiave o superchiave.

1. AD è chiave e superchiave
2. AD è chiave ma non è superchiave
3. AD non è chiave e ma è superchiave
4. AD non è chiave, né superchiave.

Domanda 3 (2 Punti)

Data una relazione $R(\underline{A}, B, C, D)$ con dipendenze funzionali $\{A \rightarrow B, AB \rightarrow C, D \rightarrow AC\}$

Quale delle seguenti affermazioni è vera?

1. R è in BCNF **ma non** in 3NF
2. R **non** è in BCNF **ma** è in 3NF
3. R è in BCNF **ed** è in 3NF
4. R **non** è in BCNF e non è 3NF