



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Corso di Laurea in Informatica

Esame di Basi di Dati

Esame del 14 Giugno 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 (6.5 punti)

*Si richiede di produrre lo schema concettuale Entità-Relazione di un'applicazione relativa ai concorsi banditi da istituzioni pubbliche. **Disegnare il diagramma ER nel riquadro della pagina che segue.***

Si richiede di progettare lo schema concettuale Entità-Relazione di un'applicazione relativa ad una compagnia operante nel campo della cybersecurity, che raccoglie dati su malware, antimalware ed utenti.

Di ogni malware¹ interessa il codice (identificatore), l'anno in cui è apparso ed il nome dei sistemi operativi (almeno uno) in cui può operare. Di ogni antimalware interessa il codice (identificativo), il costo di sviluppo ed i malware (almeno uno) che esso è potenzialmente in grado di rilevare.

Di ogni utente interessa il codice fiscale (identificativo), l'anno di registrazione e gli episodi di infezione di cui è stato vittima, ciascuno con le seguenti caratteristiche: (i) il malware coinvolto, (ii) il danno monetario subito e (iii) la data in cui l'infezione è avvenuta. Si noti che un utente può infettarsi dello stesso malware, ma in giorni diversi.

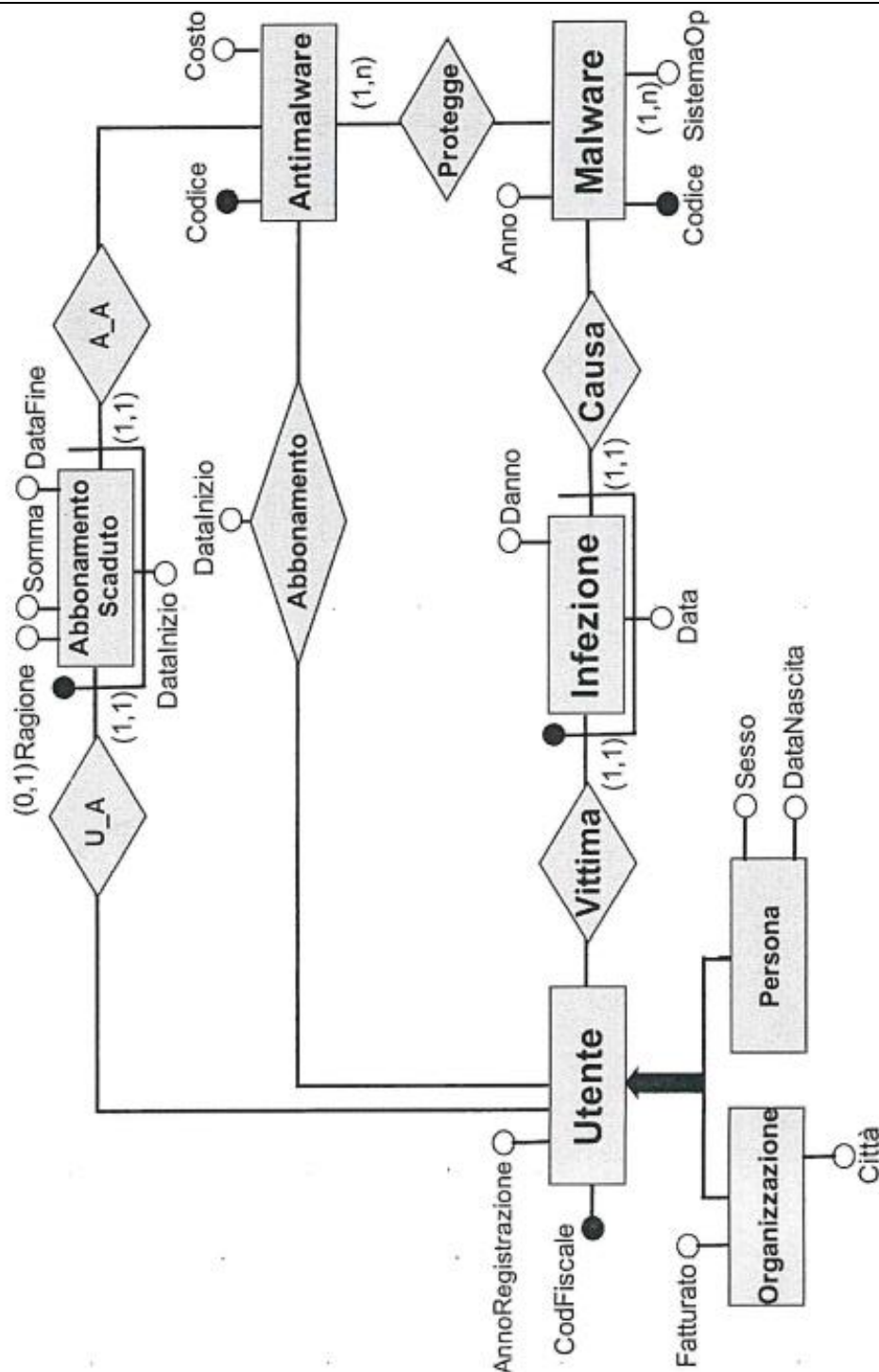
Di ogni utente interessa anche sapere quali sono stati gli antimalware di cui si è dotato nel tempo mediante abbonamento. In particolare:

- riguardo agli antimalware ai quali l'utente è attualmente abbonato interessa sapere la data di inizio dell'abbonamento.
- riguardo agli abbonamenti scaduti di antimalware, interessa la somma complessivamente pagata dall'utente per l'abbonamento, la data di inizio e la data di fine dello stesso e la ragione della disdetta (se tale ragione è nota).

Per utente nel nostro contesto si intende o una persona o una organizzazione di interesse per la compagnia di cybersecurity. Di ogni persona interessa il sesso, la data di nascita. Di ogni organizzazione interessa il fatturato e la città in cui è ubicata la sede legale

¹ Un malware è un programma maligno il cui scopo è danneggiare i computer in cui riesce ad insinuarsi

Soluzione Esercizio 1

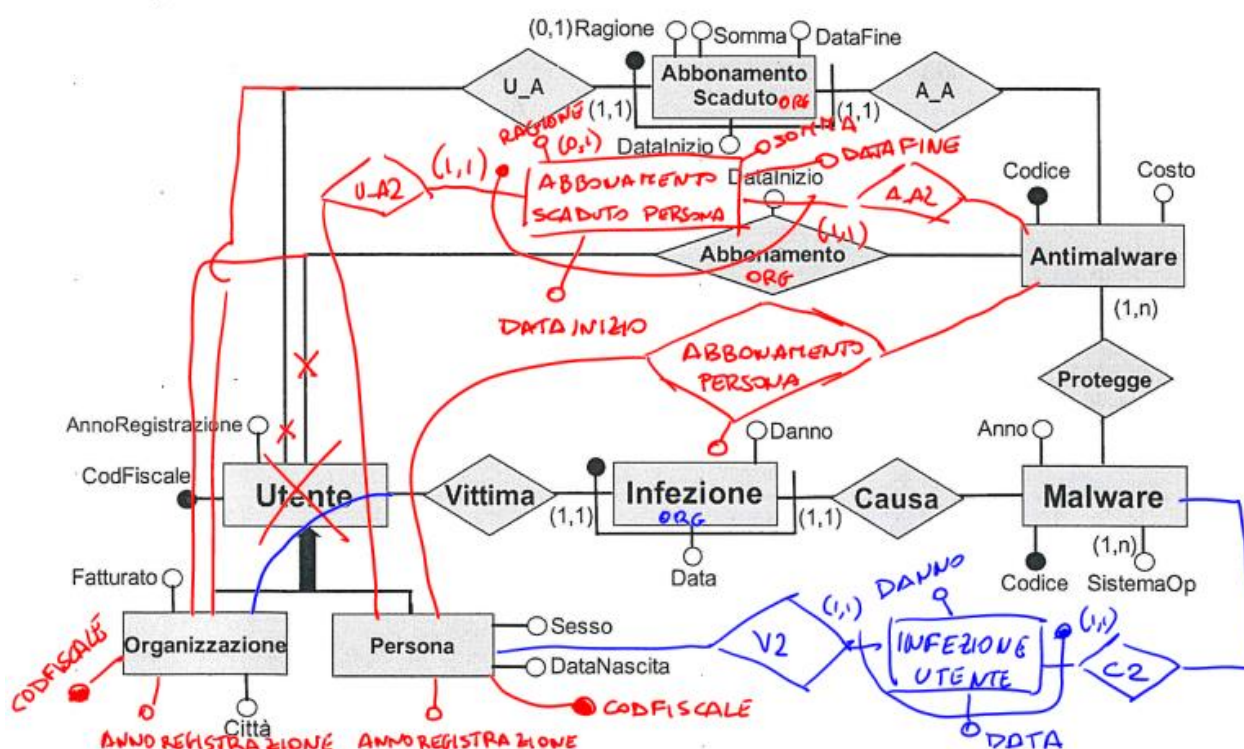


Le cardinalità (0,n) omesse per leggibilità. Si osservi come Infezione e Abbonamento Scaduto siano implementate come entità e non relazione. Se fossero relazioni, **(1)** non sarebbe possibile per un utente di essere infettato dallo stesso malware in date diverse (si noti come Data sia parte dell'identificatore), **(2)** non sarebbe possibile per un utente di avere avuto abbonamenti in passato per lo stesso malware ma con data inizio diversa.

Esercizio 2: Progettazione Logica (7.5 punti)

A partire dal Diagramma ER dell'Esercizio 1, produrre uno schema relazionale del database nel riquadro sottostante. Indicare i vincoli di chiave e gli attributi che ammettono valori nulli. **Occorre velocizzare la seguente operazione di join senza introdurre valori nulli: se si accede ad un utente, si accede anche alle informazioni sull'organizzazione o sulla persona.** Illustrare come ristrutturare l'ER per essere traducibile in uno schema relazionale.²

Soluzione Esercizio 2 (Parziale)



Per ottimizzare l'accesso congiunto alle informazioni di un utente e dell'organizzazione o della persona, occorre incorporare *Utente* in *Organizzazione* e *Persona*. Questo richiede di duplicare le entità *AbbonamentoScaduto*, e *Infezione*, e la relazione *Abbonamento*, come mostrato in figura. I cambiamenti vengono intenzionalmente mostrati con una penna per illustrare come sia possibile mostrare la ristrutturazione in sede di esame, cioè sul diagramma originario usando una penna di diverso colore (non rossa, anche se qui viene mostrato in rosso per maggiore leggibilità!).

Si noti che la soluzione di incorporare le entità figlie nel padre *Utente* avrebbe anche l'effetto di ottimizzare l'accesso congiunto, ma causerebbe l'introduzione di parecchi valori nulli.

² 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 per le votazioni locali in diverse città italiane:

- **RISULTATO-PARTITO** (Città, Partito, Anno, Percentuale)
- **CANDIDATURA** (Candidato, Anno, Partito, Città, NumVoti)

Una tupla (PD, Padova, 2022, 30) in RISULTATO-PARTITO indica che il partito PD ha preso il 30% dei voti a Padova nel 2022. Una tupla (Giordani,2022,PD,Padova,3000) in CANDIDATURA indica che Giordani ha preso 3000 voti a Padova nel 2022 con il PD.

- A. Restituire il nome del candidato che si sono presentati una sola volta (cioè in un solo anno) per il PD (2 punti).³

$C1 = \sigma_{Partito='PD'} (CANDIDATURA)$

$C2 = C1$

$DUE_CANDIDATURE =$

$\pi_{C1.CANDIDATO}$

$(C1 \bowtie_{C1.CANDIDATO=C2.CANDIDATO \text{ AND } C1.PARTITO \neq C2.PARTITO} C2)$

$(\pi_{C1.CANDIDATO} (C1)) \setminus DUE_CANDIDATURE$

³ 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. Nel riquadro, scrivere una query in Standard SQL che restituisce, per ogni partito che ha ottenuto almeno il 10% a Padova nel 2022, il numero medio di voti tra tutti i candidati del 2022. (2.5 punti). ⁴

```
SELECT PARTITO, AVG(NUMVOTI)
FROM CANDIDATURA C
WHERE ANNO=2022 AND CITTA='Padova'
AND PARTITO IN
    (SELECT PARTITO FROM RISULTATO-PARTITO
     WHERE PERCENTUALE>=10 AND ANNO=2022
     AND CITTA='Padova')
GROUP BY PARTITO
```

- C. Nel riquadro, scrivere una query in Standard SQL che, per ogni città con votazioni nel 2022, restituisce il partito che ha ottenuto più voti (2.5 punti)

```
CREATE VIEW TOP-PERCENTUALE-
CITTA(CITTA,TOP) AS
    SELECT CITTA, MAX(PERCENTUALE)
    FROM RISULTATO-PARTITO
    WHERE ANNO=2022
    GROUP BY CITTA
SELECT CITTA, PARTITO
FROM RISULTATO-PARTITO R,
    TOP-PERCENTUALE-CITTA T
WHERE R.CITTA=T.CITTA
```

⁴ Vengono quindi restituite coppie del tipo (PD,50) ad indicare che i candidati del PD hanno ottenuto in media 50 voti.

Esercizio 4: Normalizzazione (7.5 punti)

Sia data la relazione $R(A,B,C,D,E,F)$ e l'insieme di dipendenze associato $F=\{C\rightarrow CABDEF, AD\rightarrow E, BE\rightarrow C, B\rightarrow A\}$:

- Trovare la/e chiave/i di R , motivando la risposta.
- Indicare quali dipendenze violano la 3NF, motivando la risposta
- Calcolare la copertura ridotta di F . Aggiungere solo la soluzione.
- Effettuare una decomposizione in 3NF.

Punto a

Si considerano le seguenti chiusure

$C^+=\{C,A,B,D,E,F\}$

$AD^+=\{A,D,E\}$

$BE^+=\{C,A,B,D,E,F\}$

$B^+=\{B,A\}$

Quindi, C e BE sono chiavi della relazione.

Punto b

$B\rightarrow A$ viola la 3NF perché B non è superchiave e A non appartiene a nessuna delle chiavi.

Punto c

La copertura ridotta è $G=\{C\rightarrow B, C\rightarrow D, C\rightarrow F, BE\rightarrow C, B\rightarrow A, AD\rightarrow E\}$

Punto d

1. G è partizionato in sottoinsiemi tali che dip. funz. $X \rightarrow A$ e $Y \rightarrow B$ sono insieme se $X^+ = Y^+$

$\{C\rightarrow B, C\rightarrow D, C\rightarrow F, BE\rightarrow C\}, \{B\rightarrow A\}, \{AD\rightarrow E\}$

2. Viene costruita una relazione per ogni sottoinsieme:

$R_1(C,B,D,F,E), R_2(B,A), R_3(AD,E)$

3. Se esistono due relazioni $A(X)$ and $T(Y)$ con $X \subseteq Y$, A viene eliminata:

Non accade

4. Se esiste una chiave K per quale non esiste una relazione che contiene tutti gli attributi di K , viene aggiunta una relazione $T(K)$:

Non accade

5. Indicare le chiavi delle relazioni ottenute dalla normalizzazione

$R_1(C,B,D,F,E)$ con chiave C (aggiungendo anche chiave BE , si ottiene una decomposizione BCNF).

$R_2(B,A)$ con chiave B

$R_3(A,D,E)$ con chiave AD

Nome: _____ Num. Matric.: _____

Esercizio 5: Quiz (3.5 punti)

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

Domanda 1 (1.5 Punti)

Sia data la seguente porzione di log fino al guasto: CK(T5,T6), B(T7), U(T7,O6,B6,A6), U(T6, O3, B7, A7), B(T8), I(T8,O5,A5), C(T8), A(T5). Quali transazioni richiedono il REDO?

1. T5,T6,T8
2. T5,T6,T7,T8
3. T8
4. T7,T8

Domanda 3 (2 Punti)

Si consideri le relazioni R(A, B, C, D) e la seguente query

SELECT MIN(B) FROM R WHERE A=10

Quale dei seguenti indici garantisce l'efficienza massima?

1. Indice Hash sulla coppia (B,A)
2. Indice Hash sulla coppia (A,B)
3. Indice B-TREE sulla coppia (A,B)
4. Indice B-TREE sulla coppia (B,A)