Esercizio 15. Si considerino lo schema di relazione $R(A, B, C, D, E, F)$ e l'insieme di dipendenze associato: $G = \{A \to B, C \to AD, AF \to EC\}$.

(10.1) Si determinino le chiavi candidate di R.

(10.2) Si stabilisca se R e' in 3NF. Qualora non lo sia, si definisca una decomposizione di R in 3NF che conservi le dipendenze date.

1) F non compare nella parte destra di alcuna DF, dunque appartiene ad ogni chiave candidata.
   Dunque, A, C, AF e
   AF^+ = AFBDEC = R
   CF^+ = CFADEC = R
   F^+ = F
   e includendo tutto sia AF che CF sono le uniche chiavi candidate.

2) L'unica relazione che non rispetta la 3NF è A → B con A che fa parte di superchiave ma B non è membro di chiave. Calcolando una copertura minimale (facciamo in modo di avere a sinistra un solo attributo):

$$G' = \{A \to B, C \to A, C \to D, AF \to E, AF \to C\}$$

da cui otteniamo la decomposizione: $R1 = (AB), R2 = (CAD), R3 = (AFEC)$.

Esercizio 16. Si considerino lo schema di relazione $R(A, B, C, D, E, G)$ e l'insieme di dipendenze associato: $F = \{E \to D, C \to B, CE \to G, B \to A\}$.

(11.1) Si stabilisca se R e' in 3NF. Qualora non lo sia, si definisca una decomposizione di R in 3NF che conservi le dipendenze date.

(11.2) Se R non e' in BCNF, determinare una decomposizione losslessjoin di R in BCNF.

(11.1) Si stabilisca se R e' in 3NF. Qualora non lo sia, si definisca una decomposizione di R in 3NF che conservi le dipendenze date.

- Soluzione. Determiniamo innanzitutto le chiavi della relazione. C ed E non compaiono a destra in alcuna DF, quindi devono appartenere ad ogni chiave. Si ha:
  - $CE^+ = CEDBGA = R$
  - $C^+ = CBA$
  - $E^+ = ED$

  Dunque CE e' l'unica chiave candidata di R.
  R non e' in $3NF$: La DF $E \to D$ e' tale che E non e' una superchiave e D non e' un attributo primo. Applichiamo dunque ad R l'algoritmo

  visto a lezione per definire una decomposizione 3NF che preserva le dipendenze.
  - Passo 1. Mediante l'algoritmo apposito, ci assicuriamo che F sia una copertura minimale (lo e').
  - Passo 2. Da $F = \{E \to D, C \to B, CE \to G, B \to A\}$ otteniamo la decomposizione: $R1 = (ED), R2 = (CB), R3 = (CEG), R4 = (BA)$
  - Passo 3. $R3$ contiene una chiave di R e dunque la decomposizione rimane invariata.

(11.2) Se R non e' in BCNF, determinare una decomposizione losslessjoin di R in BCNF.

- Soluzione. La scomposizione e' in BCNF. Infatti:
  - DE,BC,AB sono relazioni binarie e dunque in BCNF
  - R3=CEG rispetta la BCNF poiche' per ogni $X \subseteq \{C, E, G\}$, si ha: X^+ contiene tutti gli attributi di $R3$ oppure X^+ non include attributi di $R3 \setminus X$.

Esercitazione 5: Normalizzazione

Si inseriscono i cenni fondamentali:

Terza Forma Normale

Una relazione R con chiavi $K_1, \dots, K_n$ è in Terza Forma Normale se:

Per ogni dipendenza funzionale non banale $X \rightarrow Y$, almeno una delle seguenti condizioni sono valide:

- X è superchiave (BCNF)
- ogni attributo in Y è contenuto in almeno una tra le chiavi $K_1, \dots, K_n$.

Copertura ridotta

- Un insieme di dipendenze F è una copertura ridotta:
  - **non ridondante** se non esiste dipendenza f ∈ F tale che F − {f} implica f;
  - **ridotto** se
    - **non ridondante** se non esiste dipendenza f ∈ F tale che F − {f} implica f;
    - non esiste un insieme F' equivalente a F ottenuto eliminando attributi dai primi membri di una o più dipendenze di F.
- Esempio (parte in rosso rimovibile):
  - {A → B; AB → C; A → C} è ridondante;
  - {A → B; AB → C} non è ridondante né ridotto;
  - {A → B; A → C} è ridotto

I passi per calcolare la copertura ridotta di una relazione sono i seguenti:

1. Sostituzione dell'insieme dato con quello equivalente che ha tutti i secondi membri costituiti da singoli attributi;
2. Per ogni dipendenza verifica dell'esistenza di attributi eliminabili dal primo membro;
3. Eliminazione delle dipendenze ridondanti.

Attenzione. Di seguito alcune osservazioni utili sull'algoritmo e sulla sua esecuzione.
Questo nasce dall'aver fatto tanti esercizi e aver capito cosa, effettivamente, vada fatto.

1) Il primo punto significa scomporre a destra (quindi, dipendenze funzionali che a sinistra possono avere più di un attributo e a destra basta averne uno: easy).

2) Il secondo punto indica che possiamo eliminare attributi dal primo membro e significa che:

   a. Raggiungiamo già un attributo del primo membro in qualche modo sempre attraverso uno degli attributi del primo membro (ad esempio: AD → C avrà D ridondante se ho già A → D)
   b. Alternativamente, si vede se l'attributo a secondo membro (a destra della freccia) della dipendenza funzionale sia contenuto nella chiusura di un attributo a sinistra (primo membro) meno l'attributo che si pensa essere ridondante.
   Primo esempio:
   AD → C   A → C
   Allora (questa volta ragioniamo con A)
   D è ridondante se C fosse già contenuto nella chiusura di A A^+ = {D, C} AD^+ = {A, D, C}
   Allora, eliminiamo D; in questo caso avremmo una doppia dipendenza A → C e ne togliamo una.

   Secondo esempio:
   AD → C   D → C
   Allora (questa volta ragioniamo con D)
   A è ridondante se C fosse contenuto già nella chiusura di D D^+ = {D, C} AD^+ = {A, D, C}
   Allora, eliminiamo A; in questo caso avremmo una doppia dipendenza D → C e ne togliamo una.

3) Infine, consideriamo che eliminare le dipendenze ridondanti significhi vedere quelle che si raggiungono o sono raggiunte transitivamente e si elimina quella doppia (teniamo "il giro più lungo e non "il più corto").
   Se avessi ad esempio:
   $A \rightarrow D$ $\quad$ $D \rightarrow C$ $\quad$ $A \rightarrow C$
   Toglierò $A \rightarrow C$, in quanto posso già raggiungerlo tramite $A \rightarrow D$, $D \rightarrow C$ (giro più lungo, per come la vedo e spiego io) e togliamo appunto $A \rightarrow C$ (giro più corto).

Esercizio 1
Data la relazione R(A,B,C,D) con dipendenze funzionali $\{C \rightarrow D, C \rightarrow A, B \rightarrow C\}$.

1) Mostrare tutte le chiavi di R e motivare perché ognuna è chiave
2) Dire quali dipendenze violano la BNCF spiegandone la ragione
3) Decomporre in BNCF

1) Partiamo dalle chiusure:
   $C^+ = \{C, D, A\}$
   $B^+ = \{B, C\}$
   Per proprietà transitiva sappiamo che vale anche $B^+ = \{B, C, D, A\}$ e similmente $B \rightarrow C \rightarrow A$.
   Siamo quindi certi che la ch. transitiva di B contenga tutti gli attributi.

2) $X \rightarrow Y$ è BNCF se X è superchiave
   $C \rightarrow D$ e $C \rightarrow A$ violano ma $B \rightarrow C$ no perché B è superchiave

3) Dunque, si formeranno:

$R(A,\underline{B},C,D)$
/ \
~~$R_1(\underline{C},D)$~~ $\quad$ $R_2(A,B,C)$ $\quad$ (tolgo la parte sinistra)
$R_4(\underline{C},A,D)$ $\quad$ / \
$R_2(\underline{C},A)$ $\quad$ $R_3(\underline{B},C)$

Dato che C è chiave, la decomposizione riporta tutti gli attributi a seguito di join.
Si otterranno quindi come relazioni:
$R_1(\underline{C}, D)$ $\quad$ $R_2(\underline{C}, A)$ $\quad$ $R_3(\underline{B}, C)$

Esercizio 2
Considerare uno schema di relazione R(E,N,L,C,S,D,M,P,A) con le seguenti dipendenze funzionali:
$E \rightarrow NS$
$NL \rightarrow EMD$
$EN \rightarrow LCD$
$C \rightarrow S$
$D \rightarrow M$
$M \rightarrow D$
$EPD \rightarrow A$
$NLCP \rightarrow A$
Calcolare una **copertura ridotta** della relazione data e decomporre la relazione in **terza forma normale**.
Passi della copertura:

1) Sostituzione dell'insieme dato con quello equivalente che ha tutti i secondi membri costituiti da singoli attributi

$E \rightarrow NS$
$NL \rightarrow EMD$
$EN \rightarrow LCD$
$C \rightarrow S$

D → M
M → D
EPD → A
NLCP → A

Risultato:
E → S
E → N
NL → E
NL → M
NL → D
EN → L
EN → C
EN → D
C → S
D → M
M → D
EPD → A
NLCP → A

2) Per ogni dipendenza verifica dell'esistenza di attributi eliminabili dal primo membro (si consiglia di guardare direttamente a destra delle frecce, verificando di controllare subito gli attributi raggiunti per dipendenza transitiva e/o due volte da parte di relazioni)

Le prime cinque dipendenze sono a posto, in quanto sono tutte univoche.

- EN → L presenta una ridondanza; si ha infatti E → N ma EN → L e dunque E può andare direttamente ad L
  Quindi EN → L diventa E → L
- EN → C similmente presenta la stessa ridondanza, avendo EN → C ed E → N, dunque E può andare direttamente a C
  Quindi EN → C diventa E → C
- EN → D similmente presenta la stessa ridondanza, avendo EN → D ed E → D, dunque E può andare direttamente a D. Quindi EN → D diventa E → D

Similmente sono a posto anche C → S, D → M, M → D (queste ultime due, anche se si richiamano tra loro, non costituiscono ridondanza).
Però abbiamo:

- EPD → A che presenta E → D (ex EN → D); P non compare da nessuna parte e va solo in A, dunque ci interessa mantenerlo. In particolare, EN → D ma anche EPD → A, dunque raggiungeremmo D due volte. Ecco quindi che l'eliminazione si ha su D, portando ad avere EP → A
- NLCP → A, avendo NL → E ed E → C, dunque andiamo a togliere C in quanto già raggiunto da NL. Dunque, NLCP → A diventa NLP → A.

Risultato finale:

E → S	E → S	
E → N	E → N	
NL → E	NL → E	
NL → M	NL → M	
NL → D	NL → D	
EN → L	E → L	(EN → L, E → N)
EN → C	E → C	(EN → C, E → N)
EN → D	E → D	(EN → D, E → N)
C → S	C → S	
D → M	D → M	
M → D	M → D	
EPD → A	EP → A	(EPD → A , E → D)
NLCP → A	NLP → A	(NLCP → A, NL→E, E→C

3) Eliminazione delle <u>dipendenze ridondanti</u>

Partendo dallo schema sopra:

- E → S risulta essere ridondante in quanto E → C e C → S
- E → N non è ridondante
- NL → E non è ridondante
- NL → M è ridondante perché NL → D e D → M, oltre che M → D
- E → L non è ridondante
- E → C non è ridondante
- E → D non è ridondante
- C → S non è ridondante
- D → M non è ridondante
- M → D non è ridondante
- EP → A è ridondante perché NL → E ed NLP → A
- NLP → A non è ridondante quindi

Risultato:

E → N
NL → E
E → L
E → C
E → D
C → S
D → M
M → D
NLP → A

Abbiamo la copertura ridotta; tuttavia ora occorre individuare le chiavi partendo dalla copertura ridotta.
Si vanno quindi ad eseguire le chiusure di tutti i membri:
E^+ = {E,N,L,C,D,M,S}
NL^+ = {E,N,L,D,C,S,M}
C^+ = {C,S}
D^+ = {D,M}
NLP^+ = {A,E,P,N,L,E,C,D,M,S}

Per capire chi è chiave dobbiamo capire l'insieme con più attributi; questo chiaramente è NLP.
Esso viene scelto in quanto contiene anche {A,P}, evidentemente non presenti in E ed NL (peraltro sono la stessa chiusura).
Tuttavia, se decido di includere P nell'insieme di E, automaticamente avrò anche A; quindi anche EP sarebbe chiave.

Chiavi:
NLP^+ = {N,L,P,A,E,D,C,S,M}
EP^+ = {E,N,L,D,C,S,M,P,A}

Ora dobbiamo, partendo dalle chiavi NLP - EP, applicare lo *schema della terza forma normale*:

Dati uno schema R(U) e un insieme di dipendenze F su U, con chiavi $K_1, \ldots, K_n$

1. Viene calcolata una copertura ridotta G di F
2. G viene partizionato in sottoinsiemi tali che due dipendenze funzionali X→ A e Y → B sono insieme se $X_G^+ = Y_G^+$
3. Viene costruita una relazione per ogni sotto-insieme
4. Se esistono due relazioni S(X) e T(Y) con X ⊆ Y, S viene eliminata
5. Se, per qualche *i*, non esiste una relazione S(X) con K_i ⊆ X, viene aggiunta una relazione T(K_i)

Prendiamo il passo 2:
G è partizionato in sottoinsiemi tali che due DF X→ A e Y → B sono insieme se $X_G{}^+ = Y_G{}^+$

Come si diceva prima, in questo caso si parla di E ed NL con le chiusure coincidenti, con il resto che rimane uguale:

E^+ = { E, N, L, C, D, S, M }

CHIUSURE COINCIDONO

NL^+ = { E, N, L, C, D, S, M }

C^+ = { C, S }

D^+ = { D, M } ; M^+= { D, M }

NLP^+ = { E, N, L, C, D, C, S, M, A }

Prendiamo il passo 3:
Viene costruita una relazione per ogni sottoinsieme
R_1 (E,N,L,C,D)
R_2 (C,S)
R_3 (D,M)
R_4 (N,L,P,A)

Prendiamo il passo 4:
Se esistono due relazioni S(X) e T(Y) con X ⊆ Y, S viene eliminata

Non succede nulla, resta tutto invariato, come si vede a fianco.

E → N
E → L
E → C
E → D
NL → E
R_1 (E, N, L, C, D)

C → S
R_2 (C, S)

D → M
M → D
R_3 (D,M)

NLP → A
R_4 (N, L, P, A)

Non accade

Andiamo al passo 5:
Se, per qualche i, non esiste una relazione S(X) con K_i ⊆ X, viene aggiunta una relazione T(K_i)

Siccome tra tutte le relazioni, non ne esiste una che comprenda EP (chiave), allora si aggiunge proprio R_5(E,P).

Dunque tra tutte le relazioni, si struttura:
R_1 ($\underline{E}$, $\underline{N,L}$, C,D) con E chiave ed NL chiave esterna
R_2 ($\underline{C}$, S)
R_3 ($\underline{D}$, $\underline{M}$) con D chiave ed M chiave esterna

R_4 (<u>N,L,P</u>, A) con NLP chiave
R_5 (<u>E,P</u>) con EP chiave

<u>Esercizio 3</u>
Dato lo schema R(A, B, C, D, E, F) con dipendenze:
CE → A, C → D, A → B, D → BE, B → F, AD → CF

1) Trovare la copertura ridotta G
2) Trovare tutte le chiavi
3) Dire se ci sono e quali dipendenze violano la 3NF
4) Normalizzare lo schema in 3NF
5) Lo schema normalizzato al punto 4 è anche in BCNF?

1) La copertura ridotta è data da:

C →A, C→D, A→B, D→B, D→E, B→F, AD→C

2) In merito alle chiavi, si vedono le chiusure

C^+ = {A, B, C, D, E, F}
AD^+ = {A, B, C, D, E, F}
Dunque, queste sono AD e C

3) Per essere in 3NF si ha che per ogni FD X → Y si abbia che:
   - X contiene chiave K di r
   - ogni attributo di Y è contenuto in almeno una chiave di r

   Dunque, per ogni dipendenza:
   - C → A non viola la dipendenza perché C è chiave
   - C → D non viola la dipendenza perché C è chiave
   - A → B viola perché A non è super chiave e B non è presente nella chiave
   - D → B viola perché D non è super chiave e B non è presente nella chiave
   - D → E viola perché D non è super chiave ed E non è presente nella chiave
   - B → F viola perché B non è super chiave ed F non è presente nella chiave
   - AD → C non viola perché AD è chiave

   Si conclude che lo schema non sia in 3FN

4) Per normalizzare lo schema partiamo dalle chiusure:
   C^+ = {A, B, C, D, E, F}
   A^+ = {A, B, F}
   D^+ = {D, B, E, F}
   B^+ = {B, F}
   AD^+ = {A, B, C, D, E, F}

Come si discuteva prima, le chiusure di C e AD coincidono e fanno parte della stessa partizione.
Ciò comporta la creazione delle successive relazioni:

R_1 = {A, C, D}	con chiavi C, AD
R_2 = {A, B}	con chiave A
R_3 = {B, E, D}	con chiave D
R_4 = {B, F}	con chiave B

5) Dato che tutte le dipendenze sono parte di chiave, non violano la BNCF.
Nel dettaglio:

C→A, C→D, AD→C	R1 (C, A, D)	chiavi C, AD
A→B,	R2 (A, B)	chiave A
D→B, D→E,	R3 (B, E, D)	chiave D
B→F,	R4 (B, F)	chiave B

Tutte le dipendenze funzionali non violano BCNF:
- C→A e C→D si applicano su R1 dove C è chiave
- AD→C si applica su R1 dove AD è chiave.
- A→B si applica su R2 dove A è chiave
- …

Concludiamo dunque con:

R (A, B, C, D, E, F, **G**)

con:

AF→BE, EF→BCD , A→F, B→C

1. Trovare copertura ridotta

   A→B, A→E, EF→B, EF→D, A→F, B→C

2. Trovare tutte le chiavi

   A^+ contiene tutti gli attributi, inoltre G è già una chiave

   Quindi le chiavi sono: A, G

3. Dire se ci sono e quali dipendenze violano 3NF

4. Normalizzare in 3NF

2) Come si vede la chiave dalle chiusure:
A^+ = {A,B,C,D,E,F}
EF^+ = {B,D}
A è chiave di sicuro e l'esercizio ci dà G come se fosse tale

3) $A \rightarrow B$, non viola 3NF
$A \rightarrow E$, non viola 3NF
$EF \rightarrow B$, viola (EF non super chiave e B non presente in chiave)
$EF \rightarrow D$, viola (EF non super chiave e D non presente in chiave)
$A \rightarrow F$, non viola 3NF
$B \rightarrow C$ viola (B non super chiave e C non presente in chiave)

Con chiavi: A, G

Non è in 3NF

4) Abbiamo già la copertura ridotta e si ha un partizionamento di questo tipo:
{A→B, A→E, A→F}, {EF→B, EF→D}, {B→C}
Si costruiscono le relazioni per ogni sottoinsieme
R_1 (A,B,E,F)
R_2 (E,F,B,D)
R_3 (B, C)

Se esistono due relazioni S(X) e T(Y) con $X \subseteq Y$, S viene eliminata.
Ciò non accade e rimane tutto uguale

Se, per qualche i, non esiste una relazione S(X) con $K_i \subseteq X$, viene aggiunta una relazione $T(K_i)$
Nessuna relazione contiene G e viene aggiunta una relazione:

R_1(A, B, E, F), R_2(E, F, B, D), R_3(B, C), **R_4(G)**