

## Basi di Dati – Corso B – Appello: 6 luglio 2018

Cognome, Nome \_\_\_\_\_

Matricola \_\_\_\_\_

### Domanda 1.

Si consideri la seguente base di dati (chiamata “Incarichi”) per la gestione degli incarichi da Primo Ministro della Repubblica Italiana:

PRIMOMINISTRO(Cognome, Nome, PartitoPolitico\*, AnnoNascita)

LEGISLATURA(Numero, DataInizio, DataFine\*)

INCARICO(CognomePM, NomePM, NumeroL, DataPreincarico, DataGiuramento\*, Durata\*)

La LEGISLATURA, della durata tipica di cinque anni, è caratterizzata da un numero, da una data di inizio e una data di fine legislatura.

Nella relazione INCARICO, CognomePM e NomePM costituiscono una chiave esterna su PRIMOMINISTRO, mentre NumeroL è in vincolo di integrità referenziale con LEGISLATURA. DataPreincarico è la data in cui il Primo Ministro riceve l’incarico (eventualmente esplorativo) da parte del Presidente della Repubblica. Se il preincarico (esplorativo) va a buon fine (cioè, il Primo Ministro individua una squadra di ministri e una maggioranza in parlamento), DataGiuramento contiene la data di effettiva formazione (corrispondente al giorno del giuramento) del governo presieduto dal Primo Ministro. Durata contiene invece la eventuale durata del mandato. Gli attributi asteriscati possono essere nulli. Quando PartitoPolitico è nullo, il Primo Ministro è un tecnico (ovvero, non appartenente ad alcun partito politico).

Con riferimento alla base dati “Incarichi” si richiede di rispondere alla seguente domanda in SQL:

**Individuare la legislatura con il maggior numero di preincarichi a tecnici non andati a buon fine.**

### Soluzione 1

```
WITH num_incarichi AS (  
    SELECT NumeroL, COUNT(*) AS Num  
    FROM incarico JOIN primoministro ON (CognomePM=Cognome AND NomePM=Nome)  
    WHERE PartitoPolitico IS NULL AND DataGiuramento IS NULL  
    GROUP BY NumeroL)  
SELECT NumeroL  
FROM num_incarichi  
WHERE Num >= ALL (SELECT Num FROM num_incarichi);
```

Altre soluzioni sono possibili.

### Domanda 2.

Con riferimento alla base dati “Incarichi” della Domanda 2, esprimere, sia in **algebra relazionale** che in **calcolo relazionale** su tuple con dichiarazione di range, l’interrogazione:

**Indicare cognome, nome e anno di nascita del primo ministro che ha avuto l’incarico più lungo.**

### Soluzione 2

Calcolo relazionale

$$\{p.Cognome, p.Nome, p.AnnoNascita \mid p(PRIMOMINISTRO) \mid \\ \exists i(INCARICO)(i.CognomePM=p.Cognome \wedge i.NomePM=p.Nome \wedge \\ \neg \exists i'(INCARICO)(i'.Durata > i.Durata))\}$$

Algebra relazionale:

$$\Pi_{\text{Cognome, Nome, AnnoNascita}} (\text{PRIMOMINISTRO}) \bowtie_{\Theta 2} (\sigma_{\text{Durata IS NOT NULL}} (\text{INCARICO}) - (\text{INCARICO} \bowtie_{\Theta 1} \rho_{I \leftarrow \text{INCARICO}} (\text{INCARICO})))$$

Dove:

- $\Theta 1$  :  $\text{INCARICO.Durata} < I.Durata$
- $\Theta 2$  :  $\text{Cognome} = \text{CognomePM} \wedge \text{Nome} = \text{NomePM}$

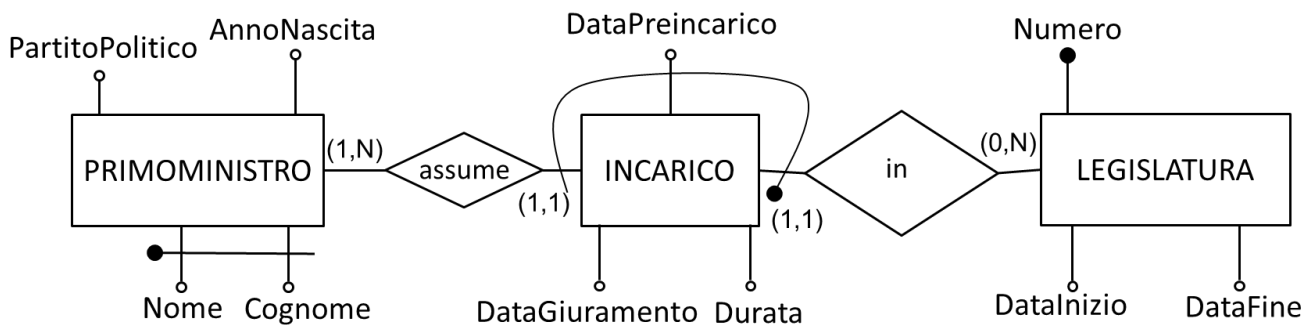
### Domanda 3.

Progettare uno schema concettuale ER che specifichi la realtà considerata nella base dati "Incarichi" e insieme tenga conto che ciascun primo ministro può aver ricevuto più incarichi militando in partiti diversi. In tal modo si potrà avere anche l'informazione di quali partiti abbiano avuto un preincarico e, in caso, un incarico (è sufficiente presentare lo schema concettuale ER e le eventuali regole aziendali, **motivando le scelte**).

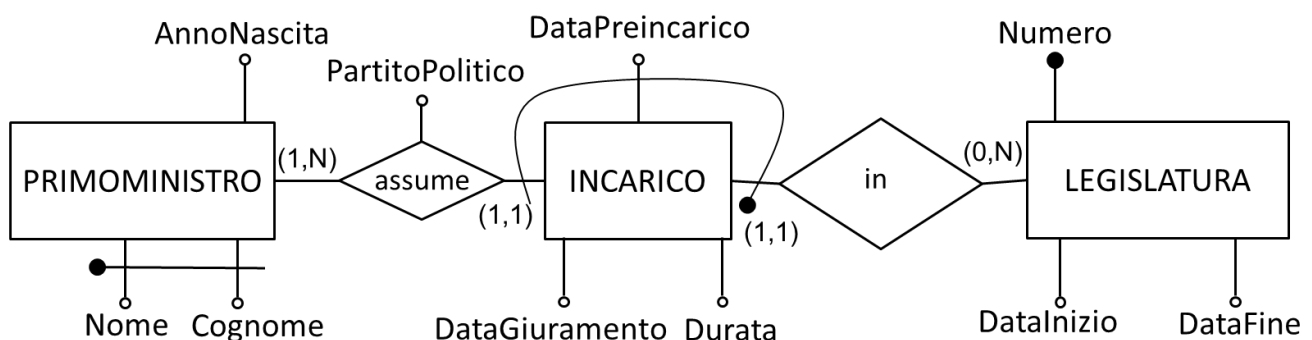
Non si chiede di considerare situazioni complesse quali cambiamento di partito e ritorno al medesimo o, addirittura di cambio partito tra preincarico e incarico. Più preincarichi o incarichi in una legislatura sono da considerare possibili.

### Soluzione 3

Lo schema ER della domanda 1 è il seguente



Per soddisfare i requisiti, la soluzione più semplice è quella di “spostare” l’attributo PartitoPolitico sull’associazione “assume”:



Ma ovviamente, altre soluzioni più complesse sono possibili.

Regole aziendali:

R1: Se  $\text{Legislatura.DataFine}$  IS NOT NULL, allora  $\text{Legislatura.DataFine} > \text{Legislatura.DataInizio}$

R2: Se  $\text{Incarico.DataGiuramento}$  IS NOT NULL, allora  $\text{Incarico.DataGiuramento} > \text{Incarico.DataPreincarico}$

R3: Gli attributi PartitoPolitico, DataGiuramento, Durata e DataFine possono essere NULL

#### Domanda 4.

- A. Esprimere la condizione necessaria e sufficiente affinché una scomposizione sia senza perdita di informazione.
- B. Dire se la seguente relazione (con le sue dipendenze funzionali) è in 3FN e, se non lo è, effettuare la scomposizione in 3FN **esplicitando tutti i passaggi**:

$R(A,C,E,G,I,M,O,Q)$  con  $F=\{CEG \rightarrow IM, CE \rightarrow IMQ, M \rightarrow GEO, CE \rightarrow O\}$

#### Soluzione 4

A. Si vedano gli appunti/slide.

B. Per prima cosa è necessario identificare la o le chiavi della relazione. L'attributo A non compare in nessuna dipendenza funzionale, quindi farà sicuramente parte della chiave. Degli altri attributi, I, O e Q compaiono sempre e solo a destra. Possiamo provare quindi, ad esempio, con CE:

$ACE^+ = \{ACEIMQGEO\}$  (usando nell'ordine  $CE \rightarrow IMQ$  e  $M \rightarrow GEO$ ).

I sottoinsiemi non sono chiavi ( $AC^+ = \{AC\}$  e  $AE^+ = \{AE\}$ ) quindi  $ACE^+$  è chiave.

**ATTENZIONE:** anche ACM è chiave, infatti:

$ACM^+ = \{ACMGEOIQ\}$  (usando nell'ordine  $M \rightarrow GEO$  e  $CE \rightarrow IMQ$ ).

e il sottoinsieme stretto AM non è chiave, infatti:

$AM^+ = \{AMGEO\}$ .

Di conseguenza, nessuna delle dipendenze funzionali è del tipo 1, 2 o 3, e la relazione non è in 3FN.

Per normalizzare, calcoliamo prima la copertura minimale dell'insieme F. Portiamo l'insieme in forma canonica:

$F' = \{$   
     $CEG \rightarrow I,$   
     $CEG \rightarrow M,$   
     $CE \rightarrow I,$   
     $CE \rightarrow M,$   
     $CE \rightarrow Q,$   
     $M \rightarrow G,$   
     $M \rightarrow E,$   
     $M \rightarrow O,$   
     $CE \rightarrow O$   
     $\}$

Identifichiamo gli attributi estranei. Nelle due dipendenze  $CEG \rightarrow I$  e  $CEG \rightarrow M$ , G è estraneo, perché abbiamo  $CE \rightarrow I$  e  $CE \rightarrow M$ . Non esistono altri attributi estranei, quindi:

$F' = \{$   
     $CE \rightarrow I,$   
     $CE \rightarrow M,$   
     $CE \rightarrow Q,$   
     $M \rightarrow G,$   
     $M \rightarrow E,$   
     $M \rightarrow O,$   
     $CE \rightarrow O$   
     $\}$

Eliminiamo le dipendenze ridondanti:  $CE \rightarrow O$  è ridondante in quanto deducibile per transitività da  $CE \rightarrow M$  ed  $M \rightarrow O$ . Non ci sono altre ridondanze. Quindi:

$F' = \{$   
 $CE \rightarrow I,$   
 $CE \rightarrow M,$   
 $CE \rightarrow Q,$   
 $M \rightarrow G,$   
 $M \rightarrow E,$   
 $M \rightarrow O,$   
 $\}$

Accorpiamo le dipendenze con lo stesso antecedente:

$F' = \{$   
 $CE \rightarrow IMQ,$   
 $M \rightarrow GEO$   
 $\}$

E scomponiamo la relazione R in:

$R1(\underline{CE}IMQ)$

$R2(\underline{M}GEO)$

Le chiavi ACE o ACM non sono contenute in nessuna delle relazioni, quindi va aggiunta una nuova relazione R3, per esempio:

$R3(\underline{ACE}).$

### Domanda 5.

Con riferimento allo schema logico della Domada 1, considerare due indici secondari **IPM** ed **IINC** (entrambi con struttura B+Albero) definiti, rispettivamente, sull'attributo AnnoNascita di PRIMOMINISTRO e sulla coppia di attributi CognomePM e NomePM di INCARICO. Si considerino inoltre i seguenti dati quantitativi:

$N_{foglie}(IPM) = 9$

$CARD(PRIMOMINISTRO) = 60$

$N_{page}(PRIMOMINISTRO) = 4$

$MIN(AnnoNascita, PRIMOMINISTRO) = 1880$

$MAX(AnnoNascita, PRIMOMINISTRO) = 1970$

$N_{foglie}(IINC) = 5$

$CARD(INCARICO) = 100$

$N_{page}(INCARICO) = 4$

$VAL(CognomePM, INCARICO) = 10$

$VAL(NomePM, INCARICO) = 10$

Calcolare la stima del costo della seguente interrogazione (già logicamente ottimizzata) eseguita sfruttando tutti gli indici a disposizione (fornire sia le formule risolutive che i risultati numerici). Si tenga presente che una tupla di PRIMOMINISTRO occupa circa 250 Byte.

$(\sigma_{AnnoNascita > 1940}(PRIMOMINISTRO)) \bowtie_{Nome=NomePM \wedge Cognome=CognomePM} INCARICO$

### Soluzione 5.

Il costo della selezione è dato da:  $C(\sigma) = CI + CD$

Il fattore di selettività della selezione è:

$$fp = (\text{MAX}(\text{AnnoNascita}, \text{PRIMOMINISTRO}) - 1940) / (\text{MAX}(\text{AnnoNascita}, \text{PRIMOMINISTRO}) - \text{MIN}(\text{AnnoNascita}, \text{PRIMOMINISTRO})) = (1970 - 1940) / (1970 - 1880) = 1/3.$$

$$CI = \lceil fp \cdot N_{\text{foglie}}(\text{IPM}) \rceil = \lceil 1/3 \cdot 9 \rceil = 3$$

$$CD = \min\{\text{Ereg}, N_{\text{page}}(\text{PRIMOMINISTRO})\} = \min\{\lceil fp \cdot \text{CARD}(\text{PRIMOMINISTRO}) \rceil, N_{\text{page}}(\text{PRIMOMINISTRO})\}$$

$$CD = \min\{\lceil 1/3 \cdot 60 \rceil, 4\} = \min\{20, 4\} = 4$$

$$C(\sigma) = CI + CD = 3 + 4 = 7$$

Il costo del join (nested loop con indice) è dato da:  $C(\bowtie) = N_{\text{page}}(\sigma) + \text{CARD}(\sigma) \cdot (CI + CD)$

$$\text{CARD}(\sigma) = fp \cdot \text{CARD}(\text{PRIMOMINISTRO}) = 1/3 \cdot 60 = 20$$

$N_{\text{page}}(\sigma) = 1$  (una pagina occupa 4KB/8KB, una tupla di PRIMOMINISTRO occupa 250B quindi sono sufficienti due pagine da 4KB, o una pagina da 8KB, per contenere le 20 tuple).

Calcoliamo ora il costo del join, partendo dal fattore di selettività del predicato “Nome=NomePM  $\wedge$  Cognome=CognomePM”:

$fp = fp1 \cdot fp2$ , dove:

$$fp1 = 1/\text{VAL}(\text{CognomePM}, \text{INCARICO}) = 1/10$$

$$fp2 = 1/\text{VAL}(\text{NomePM}, \text{INCARICO}) = 1/10$$

$$\text{Quindi, } fp = 1/10 \cdot 1/10 = 1/100$$

$$CI = \lceil fp \cdot N_{\text{foglie}}(\text{IINC}) \rceil = \lceil 1/100 \cdot 5 \rceil = 1$$

$$CD = \min\{\text{Ereg}, N_{\text{page}}(\text{INCARICO})\} = \min\{\lceil fp \cdot \text{CARD}(\text{INCARICO}) \rceil, N_{\text{page}}(\text{INCARICO})\}$$

$$CD = \min\{\lceil 1/100 \cdot 100 \rceil, 4\} = \min\{1, 4\} = 1$$

$$\text{Quindi: } C(\bowtie) = 2 + 20 \cdot (1 + 1) = 42$$

Il costo totale è quindi dato da:  $C_{\text{tot}} = C(\sigma) + C(\bowtie) = 7 + 42 = 49$

### Domanda 6.

- A. Dire cos'è e quali problemi risolve il protocollo **2PL (due fasi) stretto**.
- B. Dire in cosa consiste la tecnica del **dump/restore** (ripresa a freddo) e quando si rende necessaria.

### Soluzione 6.

Si vedano gli appunti/slide.