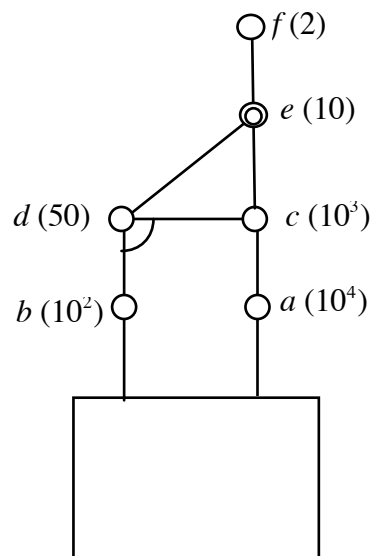


1) E' dato il seguente schema logico relazionale relativo a una società finanziaria italiana:

```
CLIENTI(codCliente, nome, cognome, dataNascita, sesso, città, regione, nazione)
PRESTITI(codPrestito, codCliente:CLIENTI, dataRichiesta, dataApprovazione,
         importoTot, agente:AGENTI, tassoInteresse, numeroRate)
AGENTI(codAgente, nome, cognome, sesso, sede:SEDI)
SEDE(codSede, nomeSede, via, civico, città, regione)
RATE(numProgressivo, codPrestito, importo, dataPrevista)
PAGAMENTI((numProgressivo, codPrestito):RATE, dataEffettiva, importoEffettivo)
```

Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (14/32 punti).

2) Date le gerarchie DFM in figura (tra parentesi le cardinalità di ciascun attributo), effettuarne la progettazione logica. Sapendo che ciascun attributo occupa 40 byte, mentre un surrogato ne richiede 4, stimare l'occupazione complessiva di memoria delle dimension table ed eventuali bridge table (11/32 punti).



3) Illustrare brevemente le tre principali soluzioni per la progettazione logica delle gerarchie temporali (slowly-changing dimension), con particolare riferimento agli scenari temporali che ciascuna di esse supporta (4/32 punti).

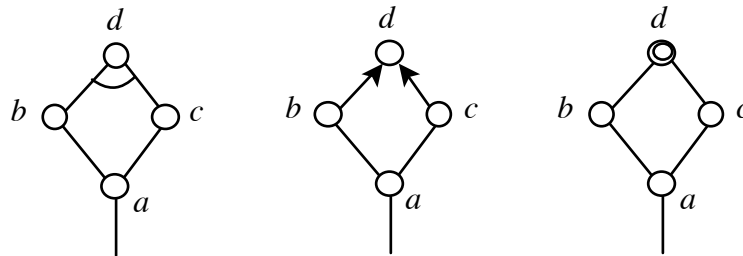
4) Descrivere sinteticamente le tre architetture “data mart indipendenti”, “data mart bus” e “hub-and-spoke”, indicando pregi e difetti di ciascuna di esse (3/32 punti).

1) E' dato il seguente schema logico relazionale relativo a una catena internazionale di hotel:

```
HOTEL(codHotel, nome, indirizzo, cittàHotel, nazione, nomeDirettore)
TIPI_CAMERE(tipoCamera, descrizione, fumatoriS/N, numeroPersone)
DISPONIBILITA'(codHotel:HOTEL, tipoCamera:TIPI_CAMERE, numeroCamereTotali, prezzo)
CLIENTI(codCliente, login, nome, cognome, sesso, dataNascita, cittàResidenza,
        numeroCartaCredito)
PRENOTAZIONI(codCliente:CLIENTI, codHotel:HOTEL, daData, aData,
             tipoCamera:TIPI_CAMERE, numeroCamere, statoPrenotazione)
```

Si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM. Si disegni poi lo schema logico per un'implementazione ROLAP, ipotizzando che sia vantaggioso materializzare due viste con group-by {numeroCamereTotali, cittàResidenza} e {meseInizioSoggiorno, cittàHotel} (15/32 punti).

2) Evidenziare la differenza tra i tre seguenti costrutti DFM elencando le dipendenze funzionali che ciascuno di essi rappresenta (9/32 punti).



3) Spiegare, eventualmente con l'ausilio di un diagramma, come si effettua il caricamento delle dimension table e delle fact table (4/32 punti).

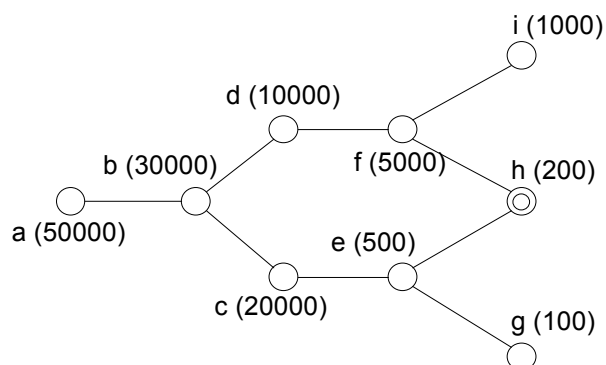
4) Illustrare la struttura e il funzionamento di un bitmap index, evidenziandone in particolare i pro e i contro rispetto a un B<sup>+</sup>-tree (4/32 punti).

1) E' dato il seguente schema logico relazionale relativo al database delle Terme Quisisana:

```
PAZIENTI(codPaziente, nome, cognome, dataNascita, sesso, città, regione, nazione)
TERAPIE(codTerapia, descrizione, areaSpecialistica:AREE)
AREE(codArea, descrizione)
VISITE(codVisita, codPaziente:PAZIENTI, codMedico:MEDICI, dataVisita, diagnosi)
PRESCRIZIONI(codTerapia:TERAPIE, dataPrescrizione, numeroRipetizioni,
             codVisita:VISITE)
EROGAZIONI(codErog, (codVisita, codTerapia):PRESCRIZIONI, dataErogazione)
MEDICI(codMedico, nome, cognome, areaSpecialistica:AREE)
```

(durante una visita il medico può fare una o più prescrizioni, ma solamente per terapie pertinenti all'area in cui è specializzato). Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (13/32 punti).

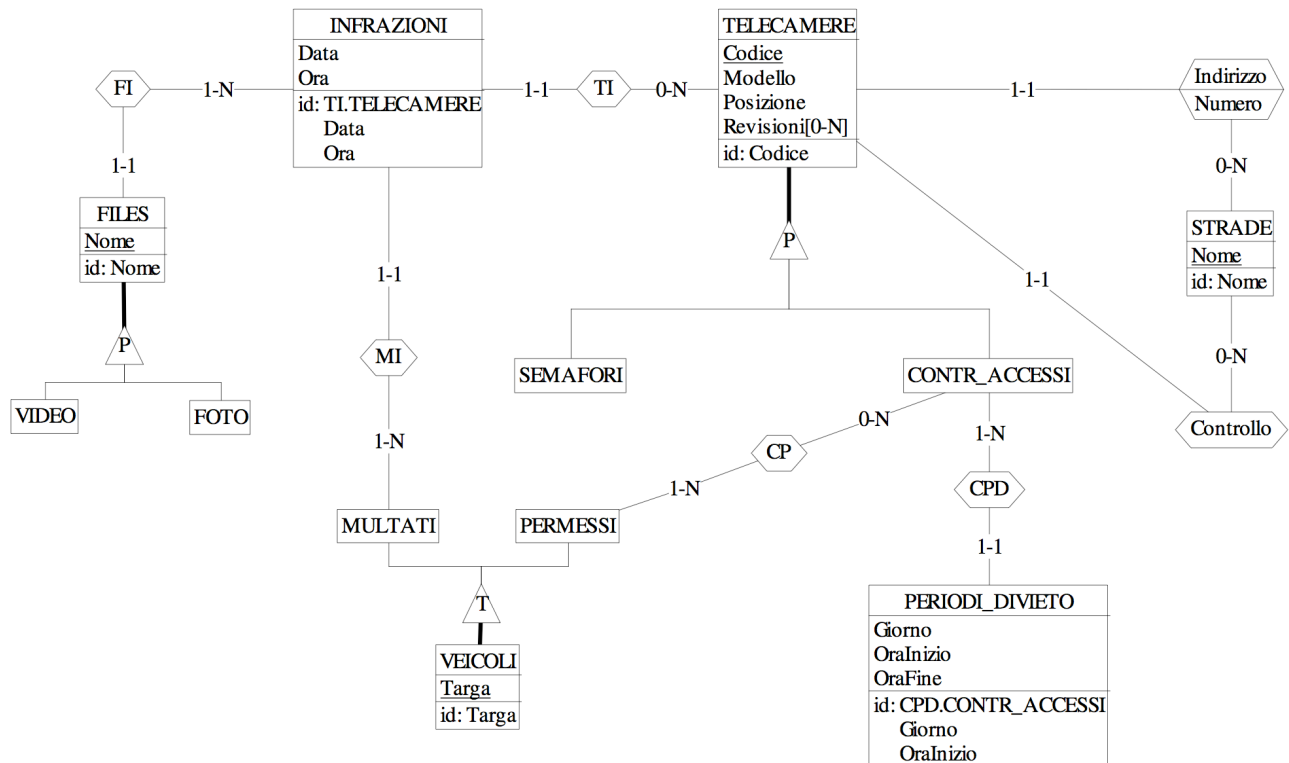
2) Dato il frammento di gerarchia in figura (tra parentesi le cardinalità di dominio degli attributi), determinare il punto di snowflaking più vantaggioso dal punto di vista dello spazio di memoria occupato. Disegnare poi lo schema logico risultante e calcolarne l'occupazione di memoria supponendo che tutti gli attributi abbiano lunghezza pari a 50B (10/32 punti).



3) Illustrare il paradigma OLAP per l'analisi dei dati multidimensionali (4/32 punti).

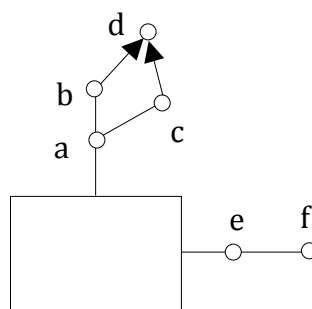
4) Descrivere la tecnica di materializzazione delle viste spiegandone pro e contro, e discutere in particolare il ruolo del reticolo multidimensionale (5/32 punti).

1) Il sistema il cui database è modellato in figura ha lo scopo di rilevare infrazioni al codice stradale, controllando mediante telecamere semafori e strade ad accesso limitato. Per le telecamere dedite al controllo degli accessi vengono specificati, per ogni giorno della settimana, i periodi (uno o più) di inizio e fine del divieto di accesso, oltre a un elenco di targhe esentate dal divieto.



Dopo avere effettuato la ricognizione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (13/32 punti).

2) Dato lo schema di fatto in figura, se ne disegni il reticolo multidimensionale completo (8/32 punti).



3) Spiegare l'utilizzo del costrutto di gerarchia ricorsiva nel DFM, descrivendone le possibili implementazioni su piattaforma ROLAP (6/32 punti).

4) Illustrare la struttura di uno star index, spiegandone pro e contro (5/32 punti).

**1)** E' dato il seguente schema logico relazionale relativo al database BigData:

CONNESSIONI(SNA\_ID\_1:SOCIAL\_NETWORK\_ACCOUNT, SNA\_ID\_2:SOCIAL\_NETWORK\_ACCOUNT,  
RTC\_Codice:REF\_TIPI\_CONNESSIONE)

MESSAGGI(MSG\_ID, RTM\_Codice:REF\_TIPI\_MESSAGGIO, SNA\_ID:SOCIAL\_NETWORK\_ACCOUNT,  
MSG\_Data\_Spedizione, MSG\_Corpo\_Messaggio, MSG\_Foto,  
MSG\_Posizione\_Geografica, MSG\_altri\_Dettagli)

TAG(TAG\_ID, MSG\_ID:MESSAGGI, SNA\_ID:SOCIAL\_NETWORK\_ACCOUNT)

PERSONE(PRS\_ID, PRS\_Nome, PRS\_Secondo\_Nome, PRS\_Cognome, PRS\_Genere,  
PRS\_Data\_Di\_Nascita, PRS\_altri\_Dettagli)

REF\_TIPI\_CONNESSIONE(RTC\_Codice, RTC\_Descrizione)

REF\_TIPI\_MESSAGGIO(RTM\_Codice, RTM\_Descrizione)

SOCIAL\_NETWORK(SNK\_Codice, SNK\_nome, SNK\_Dettagli)

SOCIAL\_NETWORK\_ACCOUNT(SNA\_ID, PRS\_ID:PERSONE, SNK\_Codice:SOCIAL\_NETWORK,  
SNA\_Name, SNA\_Data\_Registrazione, SNA\_Data\_Cancellazione)

(es: RTC\_Descrizione: *Google Circles, Tweet Following, ecc*; SNK\_Nome: *Facebook, Google+, LinkedIn, ecc.*). Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

**2)** E' dato uno schema di fatto NOLEGGIO che, tra le altre, include quattro dimensioni degeneri sesso (1 byte), fasciaEtà (4 byte), fasciaRischio (4 byte), e nazionalità (20 byte) che descrivono il conducente del mezzo noleggiato. La progettazione ha evidenziato 5 fasce d'età di interesse, 15 fasce di rischio e 60 nazionalità. Determinare il numero di eventi primari di NOLEGGIO oltre il quale la soluzione di progettazione logica basata su una junk dimension (che combina tutte e quattro le dimensioni degeneri) occupa uno spazio inferiore rispetto alla normale soluzione basata su dimension table. Considerare che le chiavi surrogate occupano 4 byte. (10/32 punti).

**3)** Definire i concetti di data warehouse, data mart e operational data store (5/32 punti).

**4)** Illustrare il ruolo delle procedure di ETL nel processo di data warehousing e dettagliarne le singole fasi (5/32 punti).

1) E' dato il seguente schema logico relazionale relativo ad una società adetta alla raccolta differenziata dei rifiuti:

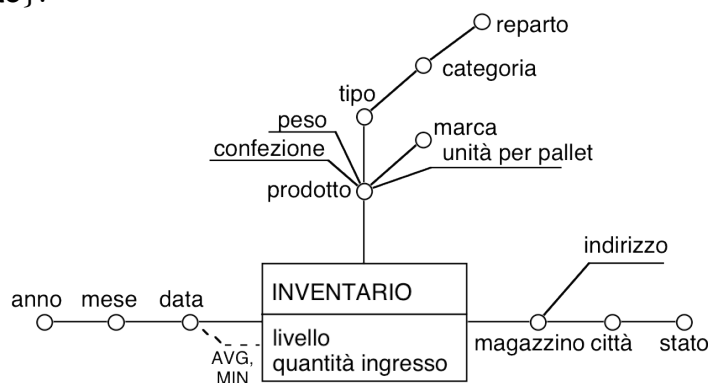
```

CATEGORIE(codCategoria, descrizione)
CONFERIMENTO(codConferimento, codPersona: PERSONE, codStazione: STAZIONI_ECOLOGICHE,
              dataConferimento, numBollettataIA)
PERSONE(codPersona, nome, cognome, telefono, dataNascita, sesso, città, regione)
PESATE(codConferimento: CONFERIMENTI, codRifiuto: RIFIUTI, peso, scontoMaturato)
RIFIUTI(codRifiuto, descrizione, tassoRiciclabilità, categoria: CATEGORIE)
STAZIONI_ECOLOGICHE(codStazione, nome, via, civico, città, regione)
    
```

(si intenda per “pesata” l'utilizzo dell'impianto di pesatura per misurare la quantità di un rifiuto al suo conferimento). Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (11/32 punti).

2) Dato lo schema di fatto in figura, se ne effettui la progettazione logica supponendo che:

- con riferimento all'associazione tra prodotti e tipi, gli utenti siano interessati agli scenari temporali di attualizzazione (oggi per ieri) e verità storica (oggi o ieri);
- con riferimento all'associazione tra categorie e reparti, gli utenti siano interessati al solo scenario di attualizzazione;
- con riferimento a tutte le altre associazioni, gli utenti siano interessati al solo scenario di verità storica;
- il progettista abbia deciso di creare due viste materializzate, una aggregata per {mese, stato}, una per {prodotto, stato};
- si elenchino i group-by delle interrogazioni che possono essere risolte sulla vista aggregata per {mese, stato}.

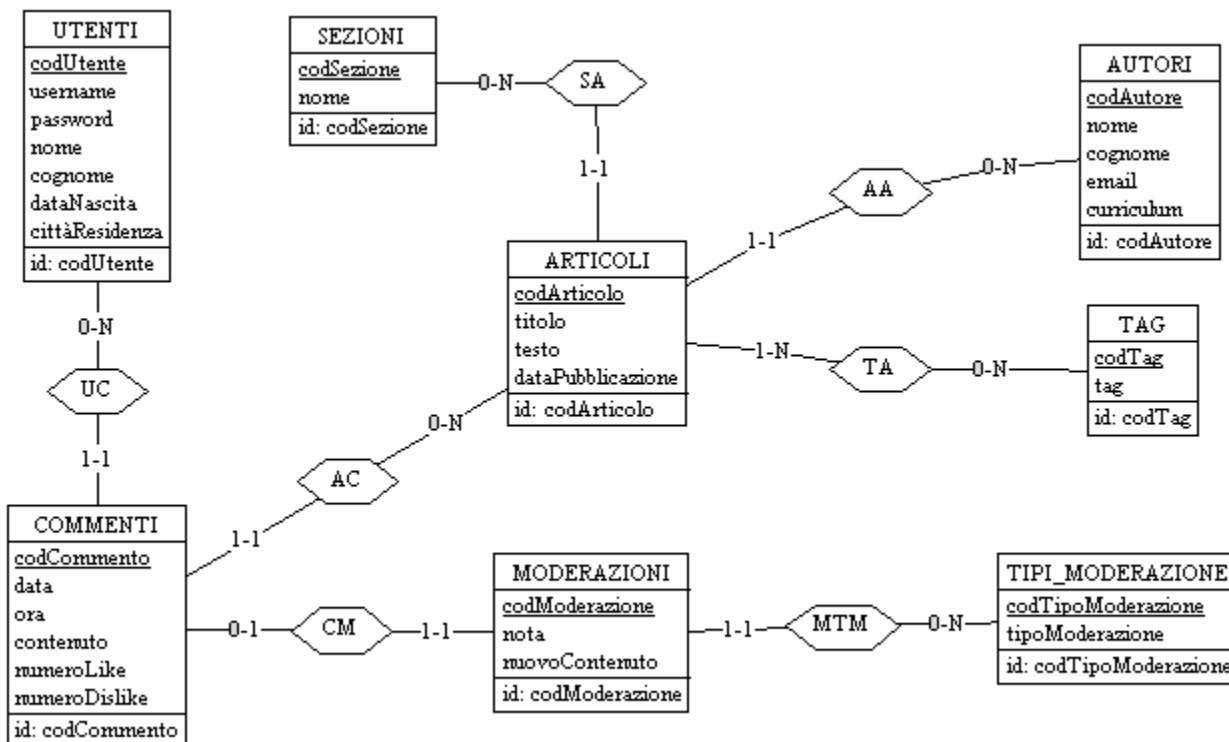


(9/32 punti)

3) Elencare e descrivere brevemente le tecniche per l'estrazione incrementale dei dati dalle sorgenti (6/32 punti).

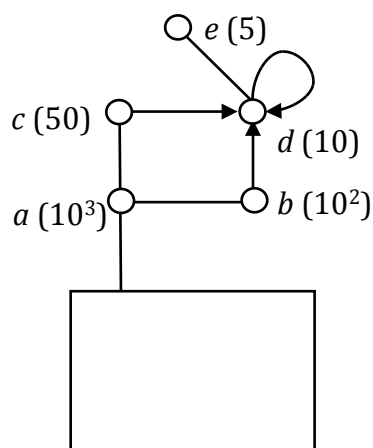
4) Illustrare le differenze tra i due principali approcci alla progettazione di data mart, ossia *data-driven* (guidato dai dati) e *requirement-driven* (guidato dai requisiti), discutendo brevemente per ciascuno i vantaggi, gli svantaggi e l'applicabilità (6/32 punti).

1) E' dato il seguente schema concettuale relativo al portale internet di un quotidiano nazionale:



(es: tipoModerazione: *linguaggioOffensivo*, *contenutoViolento*, *spam*, ecc.). Dopo avere effettuato la ricognizione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (11/32 punti).

2) Date le gerarchie DFM in figura (tra parentesi le cardinalità di ciascun attributo), effettuarne la progettazione logica. Sapendo che ciascun attributo occupa 40 byte, mentre un surrogato ne richiede 4, stimare l'occupazione complessiva di memoria delle tabelle coinvolte (9/32 punti).



3) Spiegare le principali differenze tra le piattaforme *ROLAP*, *MOLAP* e *HOLAP* (6/32 punti).

4) Definire la *business intelligence* e descrivere brevemente i livelli della piramide (6/32 punti).

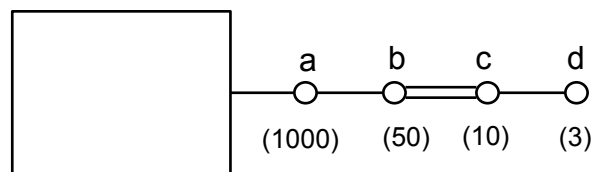
1) E' dato il seguente schema logico relazionale relativo ad un servizio di streaming di serie televisive:

```

ABBONAMENTI (codAbbonamento, codTipoAbbonamento:TIPI_ABBONAMENTO, dataInizio,
              dataFine)
DISPOSITIVI (codDispositivo, codUtente:UTENTI, codTipoDispositivo:TIPI_DISPOSITIVO,
              indirizzoIP, marca)
EPISODI (codEpisodio, codTelefilm:TELEFILM, titolo, trama, dataUscita)
STREAM (codStream, codEpisodio:EPISODI, codDispositivo:DISPOSITIVI,
         codAbbonamento:ABBONAMENTI, dataInizio, oraInizio, qualitàTrasmissione,
         giudizio)
TELEFILM (codTelefilm, titolo, genere)
TIPI_ABBONAMENTO (codTipoAbbonamento, tipoAbbonamento, durata, importo)
TIPI_DISPOSITIVO (codTipoDispositivo, tipoDispositivo, categoria)
UTENTI (codUtente, username, nome, cognome, dataNascita, sesso, nazione)
    
```

(qualitàTrasmissione e giudizio sono valori numerici da 1 a 10). Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (11/32 punti).

2) Dato lo schema di fatto in figura, se ne effettui la progettazione logica e si stimi l'occupazione di memoria delle tabelle risultanti, supponendo che ogni valore di  $b$  sia in media legato, attraverso l'arco multiplo, a 3 valori di  $c$  (9/32 punti).



3) Spiegare le principali differenze tra i carichi di lavoro di tipo OLAP e OLTP (6/32 punti).

4) Illustrare il ruolo e l'utilità dell'ODS (Operational Data Store o database riconciliato) all'interno di un'architettura a 3 livelli (6/32 punti).



1) E' dato il seguente schema logico relazionale relativo ad una compagnia di trasporto ferroviario

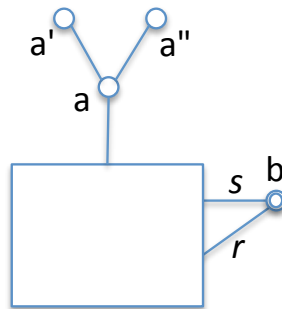
```

BIGLIETTI (codBiglietto, codCliente:CLIENTI, dataAcquisto, prezzoTotale)
CLIENTI (codCliente, nome, cognome, idDocumento, dataNascita, cittaNascita)
CORSE (codTreno:TRENI, data, status)
TIPI_TRENO (codTipoTreno, tipoTreno, capienza)
TRATTE (codTratta, codStazionePartenza:STAZIONI, codStazioneArrivo:STAZIONI,
        distanzaKm, durata, tariffa)
TRATTE_BIGLIETTI (codBiglietto:BIGLIETTI,
                   (codTreno, data, codTratta):TRATTE_CORSE, prezzo)
TRATTE_CORSE ((codTreno, data):CORSE, codTratta:TRATTE, sequenza)
TRENI (codTreno, codTipoTreno:TIPI_TRENO)
STAZIONI (codStazione, nome, paese, citta, regione)

```

(esempi di valori di status: *inOrario*, *inRitardo*, *cancellato*, ecc.). Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

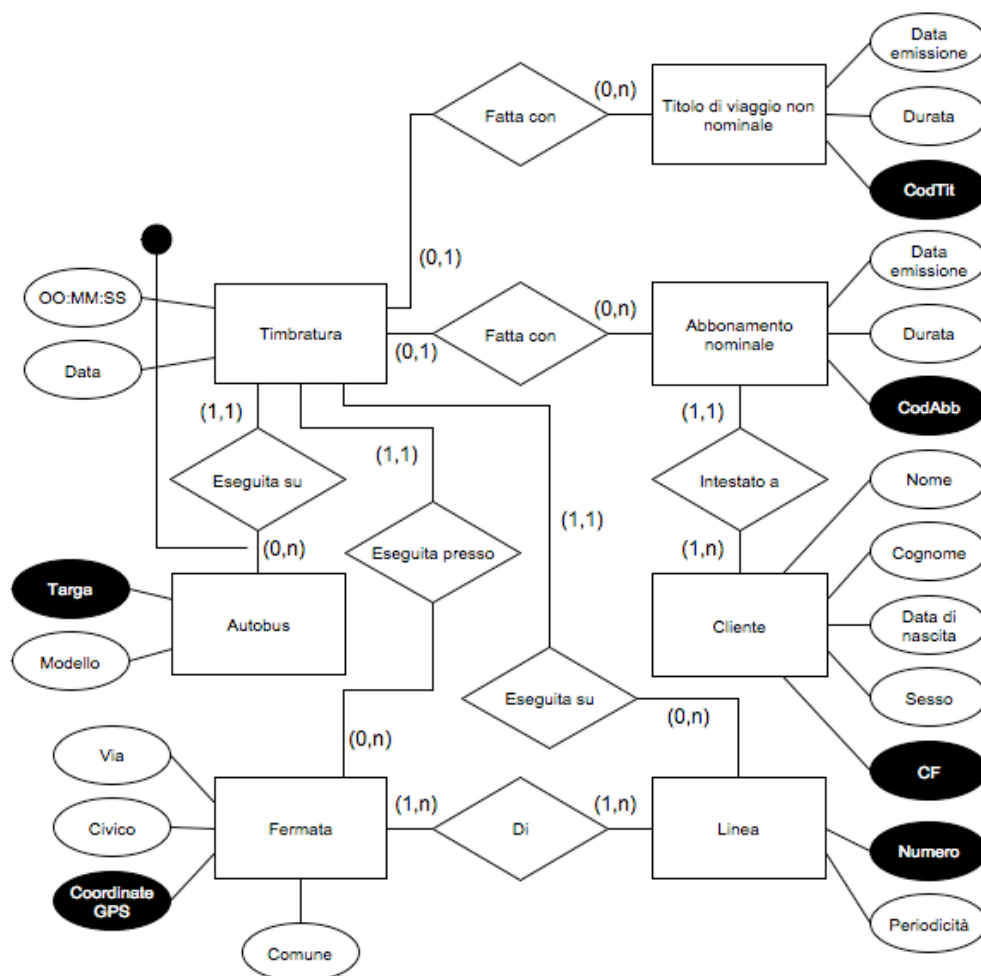
2) Dato lo schema di fatto in figura, se ne disegni il reticolo multidimensionale completo (7/32 punti).



3) Spiegare i concetti di base del *modello multidimensionale classico* (5/32 punti).

4) Illustrare il costrutto DFM di *arco multiplo*, spiegando in particolare i problemi che esso induce sull'aggregazione e come può essere tradotto in un'implementazione ROLAP (8/32 punti).

1) E' dato il seguente schema E/R relativo al servizio di trasporto cittadino:



Dopo avere effettuato la ricognizione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) E' dato uno schema di fatto **ORDINE** che, tra le altre, include tre dimensioni degeneri **statoOrdine** (2 byte), **tipoSpedizione** (4 byte) e **resoY/N** (1 byte). L'analisi ha evidenziato 4 stati ordine e 10 tipi di spedizione. Determinare il numero di eventi primari di **ORDINE** oltre il quale la soluzione di progettazione logica basata su una junk dimension (che combina tutte e tre le dimensioni degeneri) risulta conveniente dal punto di vista dello spazio occupato rispetto alle soluzioni che non ne fanno uso. Considerare che le chiavi surrogate occupano 4 byte. (8/32 punti).

3) Spiegare la differenza tra *misure di flusso*, di *livello* e *unitarie*, con particolare riferimento agli aspetti di additività (6/32 punti).

4) Illustrare il costrutto DFM di *attributo cross-dimensionale*, spiegando in particolare i problemi che esso induce sull'aggregazione e come può essere tradotto in un'implementazione ROLAP (6/32 punti).

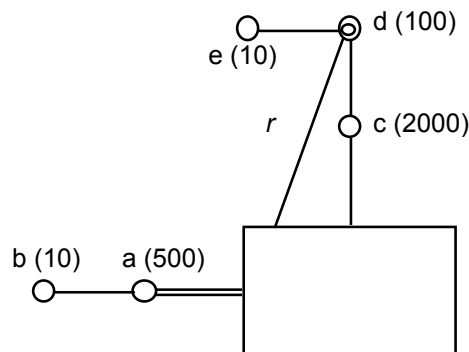
1) E' dato il seguente schema logico relazionale relativo al database di un rivenditore di componenti elettronici:

```

ACQUISTO(codAcquisto, data, totale, codCliente:CLIENTE)
CLIENTE(codCliente, nome, cognome, dataNascita, città, regione, nazione)
DETTAGLIO_ACQUISTO(codAcquisto:ACQUISTO, codProdotto:PRODOTTO, quantità)
PRODOTTO(codProdotto, nome, durataGaranzia, prezzo, iva, codProduttore:PRODUTTORE,
          codCategoria:CATEGORIA)
CATEGORIA(codCategoria, nome)
PRODUTTORE(codProduttore, nome, città, nazione, telefono)
RECENSIONE(codRecensione, testo, voto, codCliente:UTENTE, codProdotto:PRODOTTO)
SOSTITUZIONE((codAcquisto, codProdotto):DETTAGLIO_ACQUISTO, numProgressivo, data,
              codTipoProblema:TIPO_PROBLEMA)
TIPO_PROBLEMA(codTipoProblema, tipoProblema)
UTENTE(codCliente:CLIENTE, email, password, dataIscrizione)
    
```

Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) Date le gerarchie DFM in figura (tra parentesi le cardinalità di ciascun attributo), effettuarne la progettazione logica. Sapendo che ciascun attributo occupa 50 byte (mentre un surrogato ne richiede 4), che il numero di eventi primari è pari a  $10^5$  e che ciascun evento primario è legato in media a 5 valori di *a*, stimare l'occupazione complessiva di memoria delle tabelle coinvolte (8/32 punti).



3) Spiegare la differenza tra operatori di aggregazione distributivi, algebrici e olistici (6/32 punti).

4) Illustrare il costrutto DFM di gerarchia incompleta, spiegando in particolare i pro e contro delle varie soluzioni di bilanciamento (6/32 punti).

1) E' dato il seguente schema logico relazionale relativo al database di una società che si occupa di scommesse calcistiche:

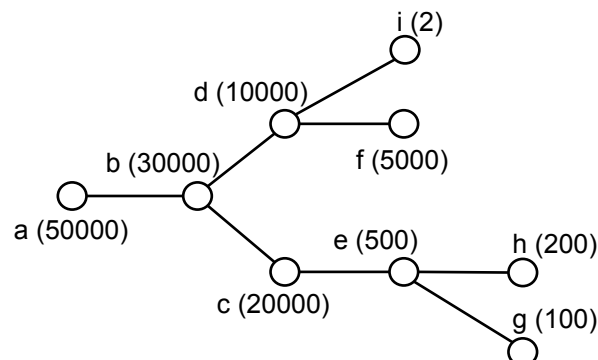
```

CAMPIONATO(codCampionato,nomeCampionato)
CLIENTE(codCliente,nome,cognome,sexo,dataNascita,idDocumento,tipDocumento)
GIOCATO(codGiocato,codCliente:CLIENTE,data,importo)
DETTAGLIO_GIOCATO(codGiocato:GIOCATO,(codTipoScommessa,codPartita):QUOTAZIONE)
PARTITA(codPartita,codCampionato:CAMPIONATO,dataSvolgimento,descrizione)
QUOTAZIONE(codTipoScommessa:TIPO_SCOMMESSA,codPartita:PARTITA,esito,quotazione)
TIPO_SCOMMESSA(codTipoScommessa,tipoScommessa)

```

(quotazione è un numero decimale compreso tra 1 e 20; esito è un valore booleano, 0 o 1; tipoScommessa: "risultato finale", "risultato parziale", "almeno un gol segnato", ecc., nomeCampionato: "Serie A", "Serie B", ecc.). Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP.

2) Dato il frammento di gerarchia in figura (tra parentesi le cardinalità di dominio degli attributi), determinare euristicamente i **due punti di snowflaking** potenzialmente più vantaggiosi dal punto di vista dello spazio di memoria occupato. Per ciascuno dei punti di snowflaking individuati disegnare poi lo schema logico risultante e, supponendo che tutti gli attributi abbiano la stessa lunghezza, determinare tra i due schemi ottenuti quello che occupa meno spazio (12/32 punti).



3) Si elenchino e si definiscano almeno tre fattori che caratterizzano la *qualità dei dati* in un data warehouse (5/32 punti).

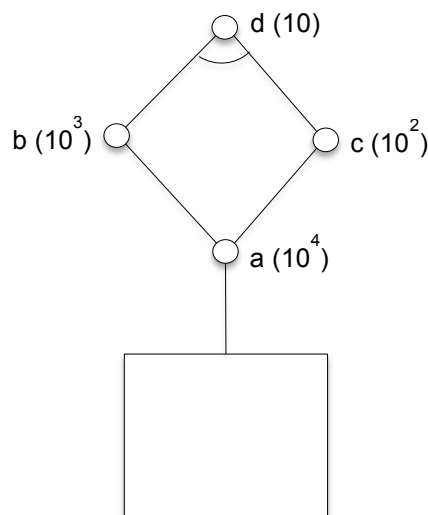
4) Illustrare la struttura e il funzionamento di un *join index* (5/32 punti).

1) E' dato il seguente schema logico relazionale relativo a un'agenzia di taxi:

```
TAXI(numTaxi, modello, marca, targa, dataAcquisto, numeroPosti,
      codConducente:CONDUCENTI)
CONDUCENTI(codConducente, nome, cognome, dataNascita)
PRENOTAZIONI(progressivoPrenotazione, data, HHMMSS, daIndirizzo:INDIRIZZI, daCivico,
              cognomeCliente, cellulareCliente, codOperatore:OPERATORI)
OPERATORI(codOperatore, nome, cognome)
CORSE((progressivoPrenotazione, data):PRENOTAZIONI, numTaxi:TAXI,
       aIndirizzo:INDIRIZZI, aCivico, oraPartenza, oraArrivo, durata, kmPercorsi, costo)
INDIRIZZI(codIndirizzo, via, quartiere, comune, regione)
```

(targa è chiave alternativa per TAXI). Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) Data la gerarchia DFM in figura (tra parentesi le cardinalità di ciascun attributo), effettuarne la progettazione logica **con snowflaking sull'attributo b**. Sapendo che ciascun attributo occupa 50 byte (mentre un surrogato ne richiede 4), stimare l'occupazione complessiva di memoria delle tabelle coinvolte. Spiegare poi come deve essere modificato lo schema se l'utente desidera attuare uno scenario temporale di retrodatazione (ieri per oggi) sull'arco tra a e c (8/32 punti).



3) Spiegare attraverso quali architetture di data warehouse e con quali accorgimenti metodologici è possibile ottenere l'*integrabilità dei data mart* sviluppati (6/32 punti).

4) Si discutano pro e contro dei tre principali approcci all'analisi dei dati multidimensionali: *reportistica statica*, *OLAP* e *reportistica semi-statica* (6/32 punti).

1) E' dato il seguente schema logico relazionale relativo a una catena di negozi di compravendita di oggetti usati:

```

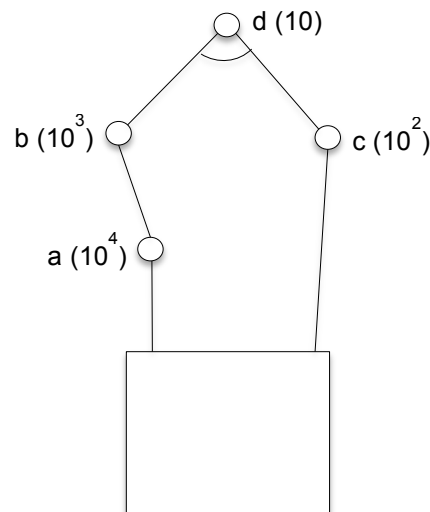
NEGOZI(codNegozio, via, civico, città, provincia, regione)
CLIENTI(codCliente, nome, cognome, dataNascita, documento, telefono, indirizzo)
ACQUISTI(codAcquisto, codCliente:CLIENTI, data, codNegozio:NEGOZI)
OGGETTI(codOggetto, codAcquisto:ACQUISTI, descrizioneTestuale, tipo:TIPI,
        prezzoDiAcquisto, prezzoDiVendita)
TIPI(codTipo, nomeTipo, codCategoria:CATEGORIE)
CATEGORIE(codCategoria, nomeCategoria)
VENDITE(codVendita, codNegozio:NEGOZI, data, codCliente:CLIENTI, importoTotale)
RIGHEVENDITE(codVendita:VENDETE, codOggetto:OGGETTI, importo)

```

in cui un cliente registrato può vendere un insieme di oggetti a uno dei negozi (tabella ACQUISTI) e un cliente, registrato o meno, può acquistare oggetti in uno dei negozi (tabella VENDITE). Si tenga presente che gli oggetti non vengono mai spostati da un negozio all'altro, per cui un oggetto può essere venduto solo presso il negozio da cui è stato acquistato.

Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) Data la gerarchia DFM in figura (tra parentesi le cardinalità di ciascun attributo), ricavarne due soluzioni di progettazione logica: una in cui usa una bridge table senza surrogati, una in cui si usa una bridge table con surrogati (pur senza effettuare snowflaking). Sapendo che ciascun attributo occupa 100 byte (mentre un surrogato ne richiede 4), stimare l'occupazione complessiva di memoria delle tabelle coinvolte nelle due soluzioni (8/32 punti).



3) Si illustri, anche attraverso un esempio, la differenza tra *pulitura* e *trasformazione* dei dati (6/32 punti).

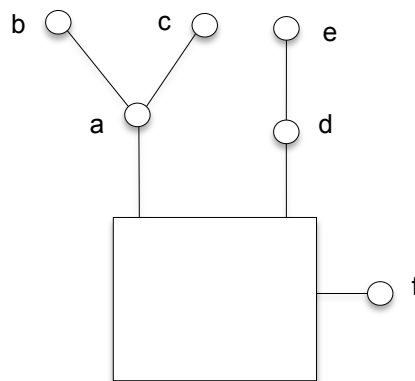
4) Si elenchino i fattori che incidono sulla *scelta dell'architettura* in un progetto di data warehouse (6/32 punti).

1) E' dato il seguente schema logico relazionale relativo ad una catena di fast-food:

RISTORANTE(codRistorante, nome, numCoperti, numDipendenti, indirizzo, città)  
 CASSA(codRistorante:RISTORANTE, progressivo)  
 TURNO(codTurno, (codRistorante, progressivo):CASSA, codDipendente:DIPENDENTE, dataOraInizio, dataOraFine)  
 DIPENDENTE(codDipendente, nome, cognome, tipoContratto)  
 SCONTRINO(codScontrino, (codRistorante, progressivo):CASSA, codTessera:CLIENTE, codDipendente:DIPENDENTE, dataOra, importoTotale)  
 CLIENTE(codTessera, nome, cognome, codiceFiscale, città, puntiAccumulati)  
 DETTAGLIO\_SCONTRINO(codScontrino:SCONTRINO, codProdotto:PRODOTTO, importo)  
 PRODOTTO(codProdotto, nome, nomeAbbreviato, prezzo, categoria)

(in fase di pagamento, l'utilizzo di una tessera non è obbligatorio). Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) Dato lo schema di fatto in figura, si elenchino le viste candidate alla materializzazione per un carico di lavoro comprendente i soli group-by **ae** e **cf** (8/32 punti).



3) Si spieghi cosa intende, nella definizione classica di data warehouse, per “*contenitore di dati orientato ai soggetti e non volatile*” (6/32 punti).

4) Si elenchino i costrutti DFM attraverso i quali è possibile creare dei cicli nelle gerarchie, illustrandone le differenze semantiche (6/32 punti).

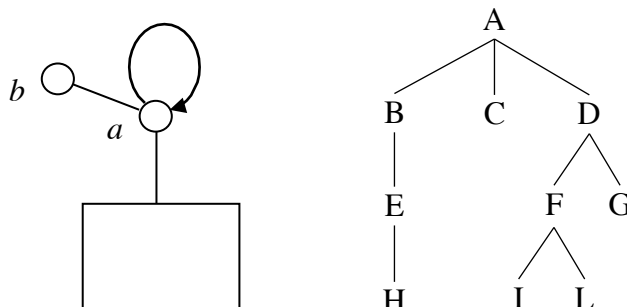
1) È dato il seguente schema logico relazionale relativo a un sito specializzato in liste nozze, in cui uno degli sposi crea una lista di oggetti desiderati (priorità è un numero intero compreso tra 1 (più importante) e 5 (meno importante)) scomponendone il valore in quote di importo prefissato. Un invitato (anche non registrato al sito) può versare una somma per effettuare un regalo acquistando quote di oggetti (percentualeSulTotale è il rapporto tra importo di DETTAGLIO\_REGALO e valore di OGGETTO) e/o sotto forma di donazione libera.

```

DONAZIONE_LIBERA(codLista:LISTA,codRegalo:REGALO,importo,data)
LISTA(codLista,nome,sposo:UTENTE,url,dataApertura,dataChiusura)
CATEGORIA(codCategoria,nome)
OGGETTO(codOggetto,codLista:LISTA,nome,valore,numQuote,valorePerQuota,priorità,
        codCategoria:CATEGORIA)
DETTAGLIO_REGALO(codRegalo:REGALO,codOggetto:OGGETTO,numQuote,importo,
                 percentualeSulTotale)
REGALO(codRegalo,invitato:UTENTE,nominativi,dedica,data,importoTotale)
UTENTE(codUtente,email,nome,cognome,dataNascita,città,regione,nazione)
    
```

Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) È dato lo schema di fatto in figura (a sinistra). Si individuino **due** possibili soluzioni per il progetto logico e se ne valuti l'occupazione di memoria, sapendo che *a* e *b* occupano 50 Byte, mentre le chiavi surrogate e gli attributi numerici occupano 4 Byte. Per determinare la cardinalità delle tabelle si faccia riferimento alla rappresentazione della ricorsione riportata nella parte destra di figura, in cui A..L sono i valori assunti dall'attributo *a* (8/32 punti).



3) Discutere pro e contro degli approcci *top-down* e *bottom-up* alla progettazione di data warehouse (6/32 punti).

4) Si spieghi il ruolo della *tabella di look-up* nell'ETL, con particolare riferimento all'alimentazione di una dimension table gestita come slowly-changing dimension di tipo 1 (6/32 punti).



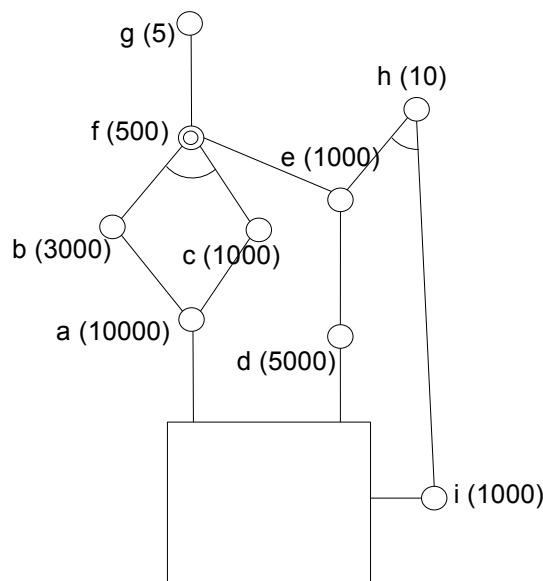
1) E' dato il seguente schema logico relazionale per la registrazione delle votazioni espresse dai parlamentari italiani sugli atti normativi proposti:

```

TIPO_LEGGE (codTipoLegge, tipoLegge)
LEGGE (codLegge, anno, titolo, dataProposta, primoFirmatario: PARLAMENTARE,
        codTipoLegge: TIPO_LEGGE)
ITERAZIONE ( (codLegge, anno): LEGGE, numIterazione, testo, numModificheEffettuate,
              dataVotazione)
VOTO ( (codLegge, anno, numIterazione): ITERAZIONE, codParlamentare: PARLAMENTARE, voto)
PARLAMENTARE (codParlamentare, nome, cognome, dataNascita, sesso, camera)
GRUPPO_POLITICO (codGruppo, nomeGruppo)
APPARTENENZA (codParlamentare: PARLAMENTARE, codGruppo: GRUPPO_POLITICO, dataInizio,
              dataFine)
    
```

(esempio: voto: *favorevole, contrario, astenuto*; tipoLegge: *"disegno di legge", "decreto legge", ecc.*; camera: *"Camera dei Deputati", "Senato della Repubblica"*; il primoFirmatario può essere omesso se la legge non è di iniziativa parlamentare). Si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti)

2) Date le gerarchie DFM in figura (tra parentesi le cardinalità di ciascun attributo), effettuarne la progettazione logica. Sapendo che ciascun attributo occupa 50 byte, mentre un surrogato ne richiede 4, stimare l'occupazione complessiva di memoria delle tabelle coinvolte (8/32 punti).



3) Spiegare le differenze concettuali tra i due costrutti del DFM che permettono di rappresentare gerarchie con istanze a lunghezza variabile (6/32 punti).

4) Illustrare i criteri qualitativi che il progettista può seguire per la scelta degli indici in uno star schema (6/32 punti).

**1)** E' dato il seguente schema logico relazionale relativo al database di gestione di intercettazioni telefoniche da parte della polizia:

```
AGENTI (codAgente, nome, cognome, idTesserino, baseOperativa)
AZIENDE (codAzienda, nomeAzienda, indirizzo, città, segnalata)
PERSONE_INTERCETTATE (codIntercettazione:INTERCETTAZIONI, codPersona:PERSONE)
INDAGINI (codIndagine, nomeInCodice, dataInizio, dataFine, responsabile:AGENTI,
          personaIndagata:PERSONE, aziendaIndagata:AZIENDA)
INTERCETTAZIONI (codIntercettazione, data, durata)
PERSONE (codPersona, nome, cognome, codAzienda:AZIENDA, segnalata)
UTILIZZO_INTERCETTAZIONI (codIntercettazione:INTERCETTAZIONI, codIndagine:INDAGINI)
```

(si consideri che: *segnalata* è un valore booleano; l'azienda in cui lavora una persona è indicata solo se conosciuta; l'indagato di un'indagine può essere un'azienda o una persona, ma non entrambe; in un'intercettazione possono essere coinvolte anche più di due persone). Dopo avere effettuato la ricognizione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

**2)** Uno schema di fatto (con cardinalità di eventi primari pari a 1 milione) include quattro dimensioni degeneri: **a** (2 byte), **b** (50 byte), **c** (4 byte) e **d** (10 byte), con cardinalità rispettivamente 4, 200000, 10 e 100. Individuare la soluzione di progettazione logica più conveniente dal punto di vista dello spazio occupato; a tale scopo si valutino anche (seguendo un criterio euristico) soluzioni con junk table che includano un sottoinsieme delle dimensioni degeneri. Considerare che le chiavi surrogate occupano 4 byte. (8/32 punti).

**3)** Illustrare a grandi linee la procedura per il *caricamento delle dimension table e delle fact table*, spiegando in particolare il ruolo delle look-up table (6/32 punti).

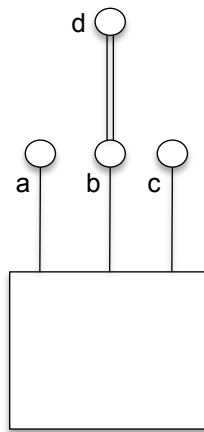
**4)** Si discutano pro e contro delle diverse tecniche per l'*estrazione incrementale* dei dati dalle sorgenti (6/32 punti).

1) E' dato il seguente schema logico relazionale relativo al database della Serie A Lega Basket:

```
GIOCATORI (codGiocatore, nome, cognome, dataNascita, cittaNascita, altezza, peso)
GIORNATE (codGiornata, giornata, stagione)
INGAGGI (codGiocatore:GIOCATORI, dataInizio, dataFine, codSquadra:SQUADRE)
PARTITE (codPartita, codSquadraCasa:SQUADRE, codSquadraOspite:SQUADRE,
        codGiornata:GIORNATE, data, ora, arena, risultato)
PUNTI (codPunto, codSquadra:SQUADRE, codGiocatore:GIOCATORI, codPartita:PARTITE,
       codTipoPunto:TIPI_PUNTO, tempoDiGioco)
SQUADRE (codSquadra, nome, città, presidente, sponsor)
TIPI_PUNTO (codTipoPunto, tipoPunto, valore)
```

Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) Dato lo schema di fatto in figura, se ne disegni il reticolo multidimensionale completo (8/32 punti).



3) Spiegare le differenze tra i tre principali *scenari temporali* utilizzabili per interrogare le gerarchie dinamiche e descrivere le soluzioni di progettazione logica che li supportano (6/32 punti).

4) Spiegare in che modo, dato un reticolo multidimensionale e un carico di lavoro, si possono determinare le *viste candidate alla materializzazione* (6/32 punti).

1) E' dato il seguente schema logico relazionale relativo al database di una compagnia aerea:

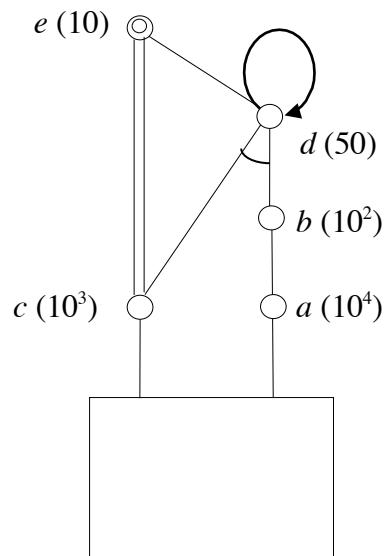
```

ACQUISTO (codAcquisto, data, importoTotale, codAcquirente: PERSONA)
AEROPORTO (codAeroporto, nome, città)
BIGLIETTO (codBiglietto, (data, codTratta): VOLO, codAcquisto: ACQUISTO,
            codPasseggero: PERSONA, importoVolo, importoTasse, numeroBagagli, priorità)
CARTA_IMBARCO (codCartaImbarco, codBiglietto: BIGLIETTO, posto)
PERSONA (codPersona, nome, cognome, dataNascita, cittàResidenza, idDocumento,
          tipoDocumento, telefono, email)
TIPO_AEROPLANO (codTipoAeroplano, nome, numeroPosti)
TRATTA (codTratta, codAeroportoPartenza: AEROPORTO, codAeroportoArrivo: AEROPORTO,
        oraPartenza, oraArrivo)
VOLO (data, codTratta: TRATTA, tipoAeroplano: TIPO_AEROPLANO, ritardoInPartenza,
      ritardoInArrivo, cancellato)

```

(cancellato e priorità sono campi booleani; la carta d'imbarco di un biglietto viene generata al momento del check-in). Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (11/32 punti).

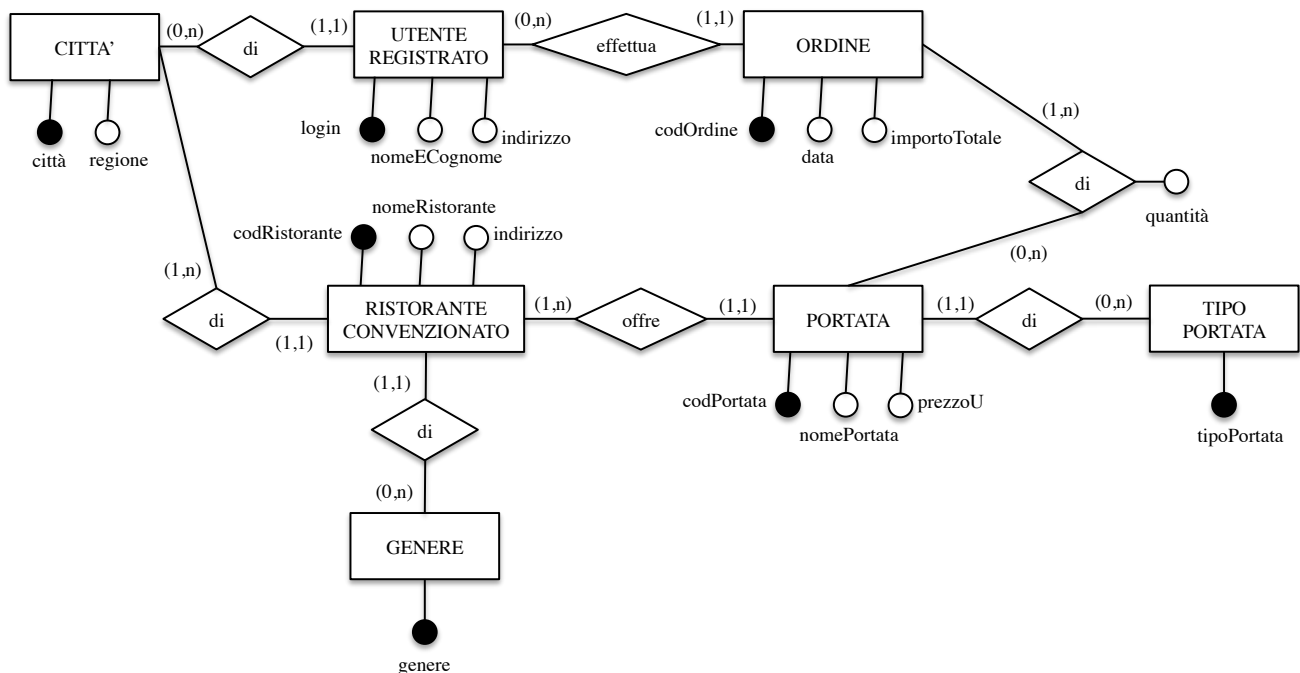
2) Date le gerarchie DFM in figura (tra parentesi le cardinalità di ciascun attributo), effettuarne la progettazione logica. Sapendo che ciascun attributo occupa 40 byte, mentre un surrogato ne richiede 4, stimare l'occupazione complessiva di memoria delle dimension table ed eventuali bridge table. Si consideri che ogni istanza di *c* è legata, in media, a 2 istanze di *e* (9/32 punti).



3) Illustrare, anche con l'aiuto di un esempio, il ruolo delle *misure di supporto* nell'utilizzo degli operatori di aggregazione algebrici (6/32 punti).

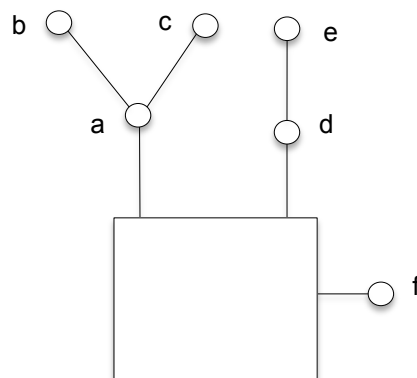
4) Spiegare in che modo uno *scenario temporale di retrodatazione* può essere implementato in una piattaforma ROLAP (6/32 punti).

1) E' dato il seguente schema concettuale E/R di un'applicazione per gestire la prenotazione e la consegna di cibo tramite app mobile.



(un cliente può effettuare ordini solo nella sua città; in ciascun ordine fatto, le portate ordinate sono tutte offerte dallo stesso ristorante). Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (11/32 punti).

2) Dato lo schema di fatto in figura, si elenchino le viste candidate alla materializzazione per un carico di lavoro comprendente i soli group-by bd, cd e ae (9/32 punti).



3) Illustrare il costrutto di *gerarchia condivisa* nel DFM, descrivendone le possibili implementazioni su piattaforma ROLAP (6/32 punti).

4) Illustrare il ruolo delle procedure di *ETL* nel processo di data warehousing e dettagliarne le singole fasi (6/32 punti).

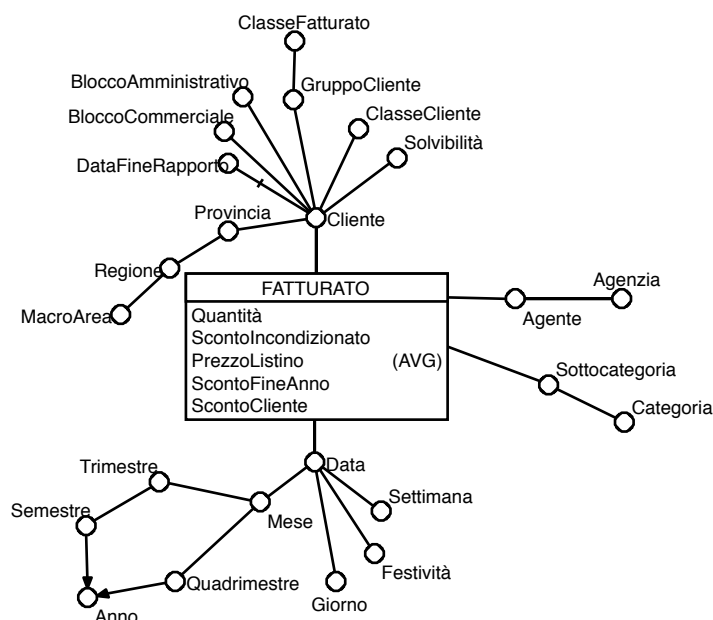
1) E' dato il seguente schema logico relazionale relativo ai dati raccolti dalle macchine per il fitness di una catena di palestre:

```
MACCHINE (codMacchina, modello:MODELLI, codPalestra:PALESTRE)
MODELLI (modello, tipologia, serie, cardioY/N)
PALESTRE (codPalestra, nome, località, metriQuadri, convenzionataY/N)
ISCRITTI (codIscritto, nome, cognome, età, cittàNascita, dataIscrizione, peso)
ABBONAMENTI (codIscritto:ISCRITTI, dataInizioAbbonamento, durata)
ALLENAMENTI (codMacchina:MACCHINE, codIscritto:ISCRITTI, data, OOMMSS, durata,
pesoUtilizzato, velocità)
```

(l'attributo pesoUtilizzato in ALLENAMENTI si riferisce all'eventuale peso usato; sia pesoUtilizzato sia velocità sono opzionali). Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (11/32 punti).

2) Dato lo schema di fatto in figura, se ne effettui la progettazione logica supponendo che:

- con riferimento all'associazione tra gruppo cliente e classe fatturato, gli utenti siano interessati agli scenari temporali di attualizzazione (oggi per ieri) e verità storica (oggi o ieri);
- con riferimento all'associazione tra agenti e agenzie, gli utenti siano interessati al solo scenario di attualizzazione;
- con riferimento a tutte le altre associazioni, gli utenti siano interessati al solo scenario di verità storica;
- il progettista abbia deciso di creare due viste materializzate, una aggregata per {Mese, Regione}, una per {BloccoAmministrativo, Mese}.

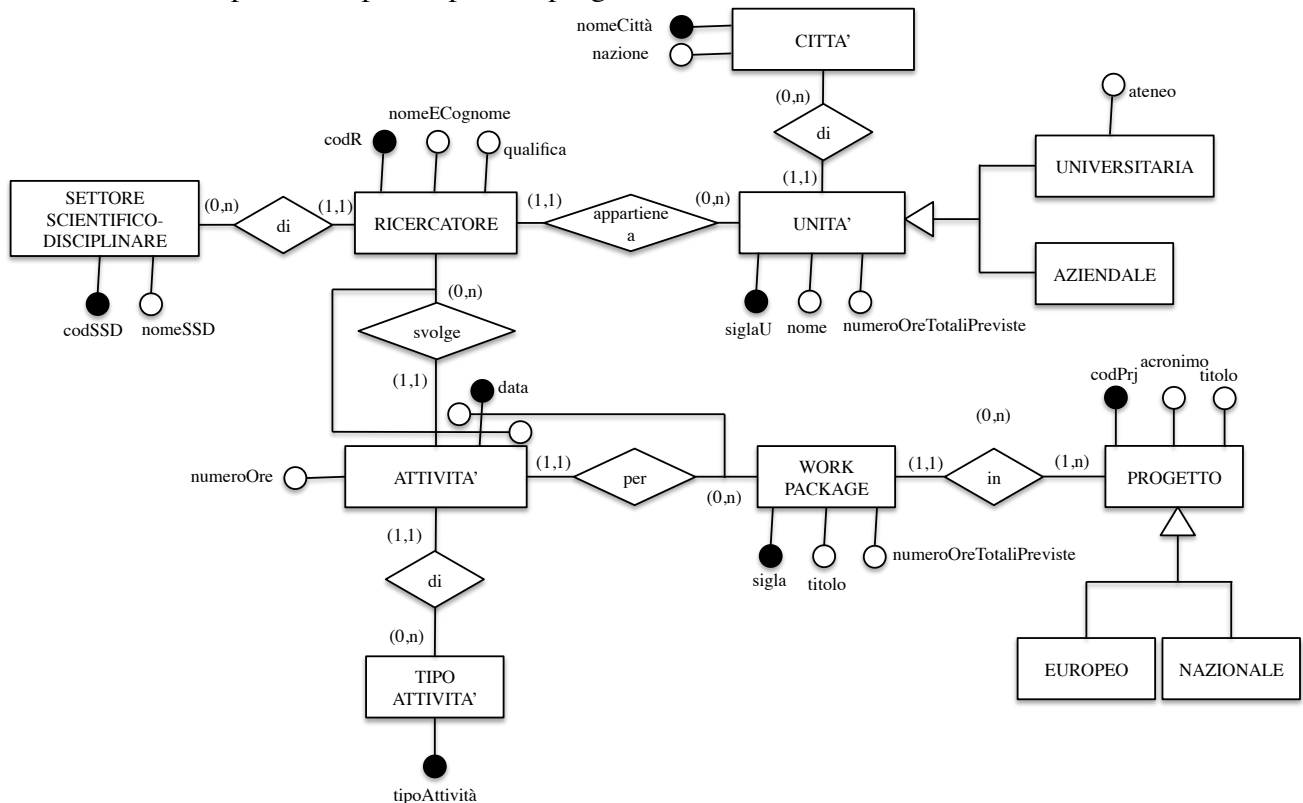


(9/32 punti)

3) Spiegare in che modo le architetture “data mart bus” e “hub-and-spoke” permettono di ricostruire una visione “enterprise” integrata dei dati (6/32 punti).

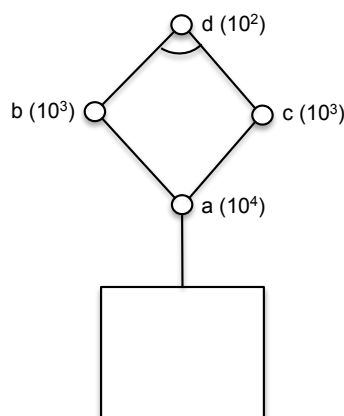
4) Cos'è il group-by set di un'interrogazione OLAP e in che modo è implicitamente rappresentato nel DFM? (6/32 punti).

1) E' dato il seguente schema concettuale E/R di un'applicazione per gestire la rendicontazione delle attività svolte dal personale partecipante a progetti di ricerca.



Dopo avere effettuato la ricognizione e l'eventuale normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) Data la gerarchia DFM in figura (tra parentesi le cardinalità di ciascun attributo), effettuarne la progettazione logica. Sapendo che ciascun attributo occupa 20 byte (mentre un surrogato ne richiede 4), stimare l'occupazione complessiva di memoria delle tabelle coinvolte (8/32 punti).



3) Spiegare punti in comune e differenze tra i costrutti di *gerarchia incompleta* e *gerarchia ricorsiva* nel DFM (6/32 punti).

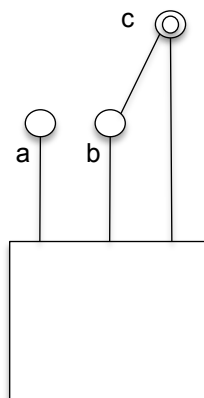
4) Illustrare la struttura di uno *star index*, spiegandone pro e contro (6/32 punti).

1) E' dato il seguente schema logico relazionale utilizzato in locale dall'applicazione *Uotsap* per gestire l'invio di messaggi tra utenti:

```
MEDIA (nomeFile, tipo, dimensione)
UTENTE (numeroTel, nickName, dataOraUltimaConnessione, foto:MEDIA, info)
GRUPPO (nomeGruppo, dataCreazione)
UTENTE_IN_GRUPPO (numeroTel:UTENTE, nomeGruppo:GRUPPO)
MESSAGGIO_INVIATO (codMI, aNumeroTel:UTENTE, aGruppo:GRUPPO, dataOra, status, testo)
MEDIA_INVIATO (nomeFile:MEDIA, codMI:MESSAGGIO_INVIATO)
MESSAGGIO_RICEVUTO (codMR, daNumeroTel:UTENTE, daGruppo:GRUPPO, dataOra, testo)
MEDIA_RICEVUTO (nomeFile:MEDIA, codMR:MESSAGGIO_RICEVUTO)
```

(l'attributo *status* può valere "inviato", "ricevuto" o "letto", mentre *tipo* vale "audio", "video" o "immagine"; un messaggio riguarda o un singolo utente o un gruppo). Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) Dato lo schema di fatto in figura, se ne disegni il reticolo multidimensionale completo (8/32 punti).



3) Cosa si intende per *architettura federata* in ambito data warehouse? (6/32 punti).

4) Descrivere la tecnica di materializzazione delle viste spiegandone pro e contro, e discutere in particolare il ruolo del reticolo multidimensionale (6/32 punti).



**1)** E' dato il seguente schema logico relazionale relativo al traffico ferroviario:

```
CONVOGLIO (idConvoglio, modello, numVagoni, capacità, dataAcquisto)
TRENO (codTreno, daStazione:STAZIONE, aStazione:STAZIONE, oraPartenza, oraArrivo,
       festivoYN)
STAZIONE (idStazione, nome, comune, provincia, regione)
VIAGGIO (codTreno:TRENO, idConvoglio:CONVOGLIO, data, manovratore:PERSONALE,
         controllore:PERSONALE, ritardoPartenza, ritardoArrivo)
PERSONALE (codP, nome, cognome, dataNascita, sesso, qualifica)
OBLITERAZIONE (data, OOMMSS, idObliteratrice, (codTreno, data):VIAGGIO)
```

Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

**2)** Uno schema di fatto (con cardinalità di eventi primari pari a 500k) include tre dimensioni degeneri: **a** (1 byte), **b** (100 byte) e **c** (10 byte), con cardinalità rispettivamente 20, 1000 e 100. Individuare la soluzione di progettazione logica più conveniente dal punto di vista dello spazio occupato; a tale scopo si valutino anche (seguendo un criterio euristico) soluzioni con junk table che includano un sottoinsieme delle dimensioni degeneri. Considerare che le chiavi surrogate occupano 4 byte (8/32 punti).

**3)** Illustrare le differenze tra i costrutti di *gerarchia condivisa* e *convergenza* (6/32 punti).

**4)** Spiegare le principali differenze tra le piattaforme *ROLAP* e *MOLAP* (6/32 punti).

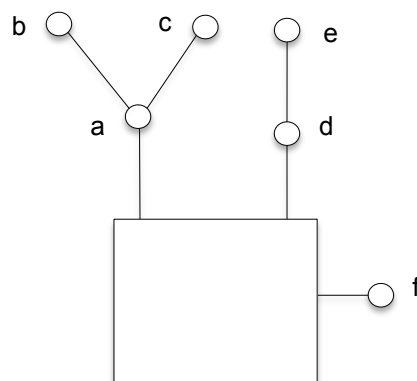
1) E' dato il seguente schema logico relazionale relativo al pubblicazioni scientifiche:

```
FORUM(codForum, rank)
RIVISTE(codRiv:FORUM, nomeRiv, casaEditrice)
CONFERENZE(codConf:FORUM, nomeConf, città, nazione, dataConferenza)
ARTICOLI(codArt, pubblicatoIn:FORUM, annoPubblicazione)
AUTORI(codAutore, nome, cognome, dataNascita, dipartimento:DIPARTIMENTI,
        impactFactor)
DIPARTIMENTI(codDip, nomeDip, settoreScientifico, area)
SCRITTO_DA(codArt, codAut)
```

Un articolo viene pubblicato in un forum, che può essere o una rivista o una conferenza, ed è caratterizzato da un rank (A, B, C, D) che ne descrive la rilevanza scientifica. L'impact factor è un indice numerico intero che misura la produttività scientifica complessiva di un autore. Un settore scientifico appartiene esattamente a un'area.

Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

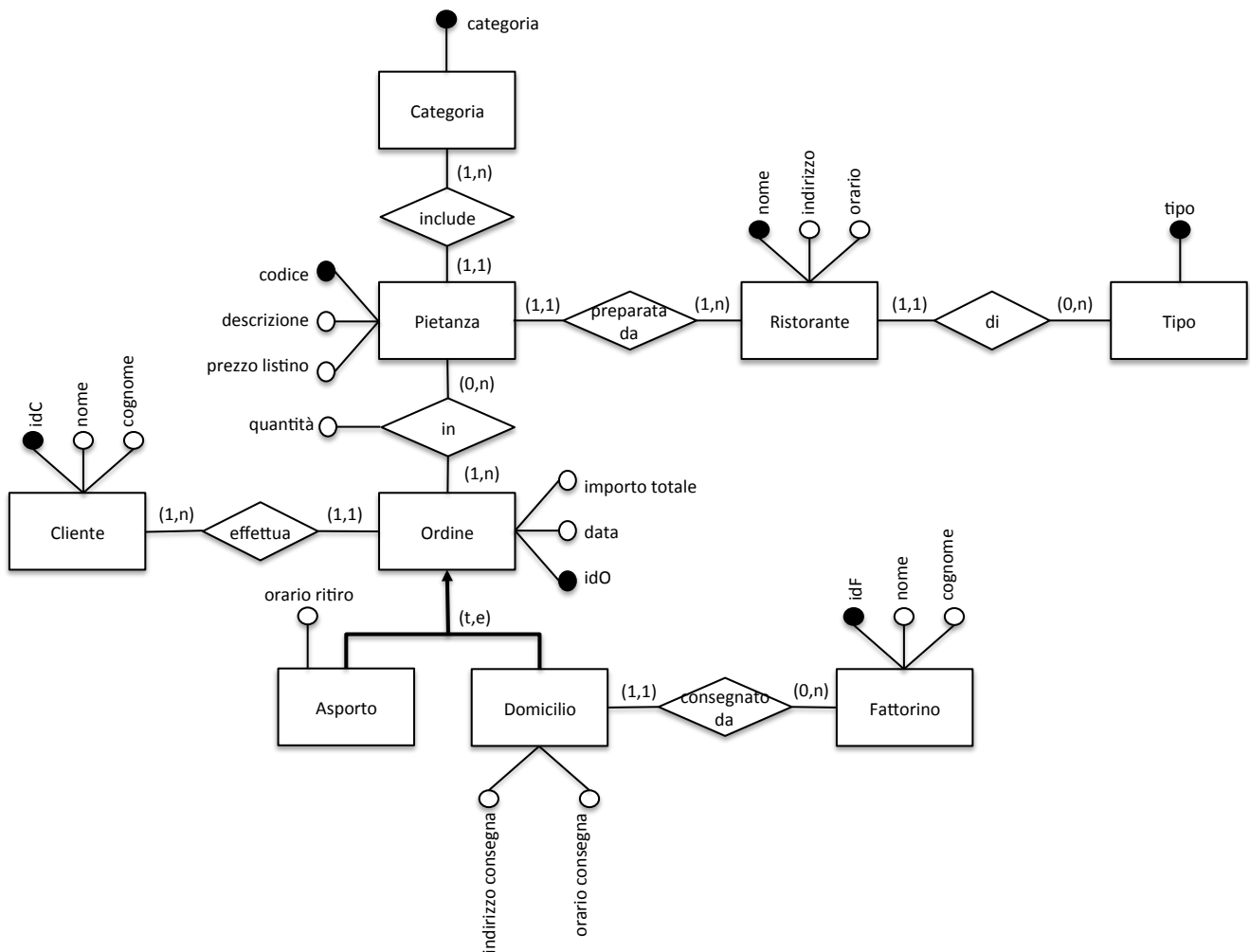
2) Dato lo schema di fatto in figura, si elenchino le viste candidate alla materializzazione per un carico di lavoro comprendente i soli group-by cf, ce e af (8/32 punti).



3) Illustrare brevemente le tre principali soluzioni per la progettazione logica delle gerarchie temporali (*slowly-changing dimension*), con particolare riferimento agli scenari temporali che ciascuna di esse supporta (6/32 punti).

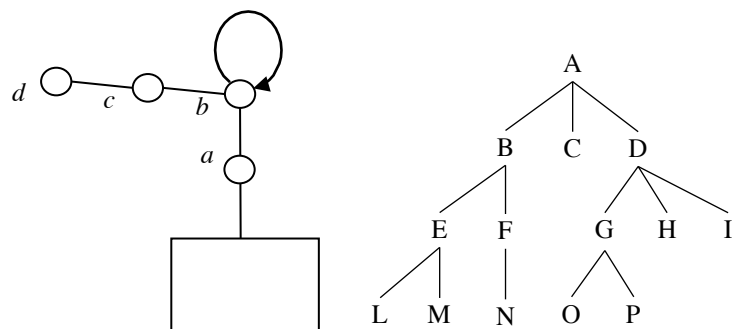
4) Descrivere i due principali approcci alla progettazione di data mart, ossia *data-driven* (guidato dai dati) e *requirement-driven* (guidato dai requisiti), discutendo brevemente per ciascuno i vantaggi, gli svantaggi e l'applicabilità (6/32 punti).

1) E' dato il seguente schema Entity/Relationship relativo agli ordini online di pranzi e cene presso un insieme di ristoranti convenzionati:



Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) Dato lo schema di fatto in figura, si individuino **due** possibili soluzioni per il progetto logico e se ne valuti l'occupazione di memoria, sapendo che gli attributi occupano 100 Byte e che le cardinalità di  $a$ ,  $c$  e  $d$  sono pari a 100, 10 e 2 rispettivamente. Per determinare la cardinalità delle tabelle si faccia riferimento alla rappresentazione estensionale della ricorsione riportata a destra, in cui A..P sono i valori assunti dall'attributo  $b$  (8/32 punti).



3) Quali sono i vantaggi metodologici che si conseguono prevedendo una fase di *progettazione concettuale*? (6/32 punti).

4) Illustrare la struttura e il funzionamento di un *bitmap index*, evidenziandone in particolare i pro e i contro rispetto a un B<sup>+</sup>-tree (6/32 punti).

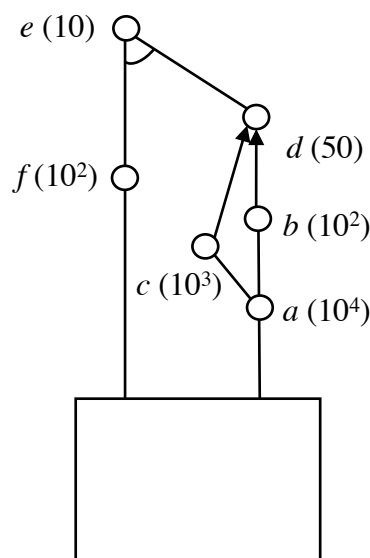
1) E' dato il seguente schema logico relazionale relativo a un'applicazione per la ricerca di un'anima gemella:

```

PROFILO (nickName, password, email, sesso, dataIscrizione, coloreOcchi, coloreCapelli,
         altezza, peso)
HOBBY (codHobby, nomeHobby)           // per esempio musica, musei, ecc.
PERSONA_HA_HOBBY (nickName, codHobby)
CUCINA (codCucina, nomeCucina)        // per esempio cinese, indiana, ecc.
PERSONA_AMA_CUCINA (nickName, codCucina)
LIKE (daNickName, aNickName, dataLike)
MESSAGGIO (codMSG, daNickName, aNickName, testo, data)
    
```

Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individuino **due** fatti di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

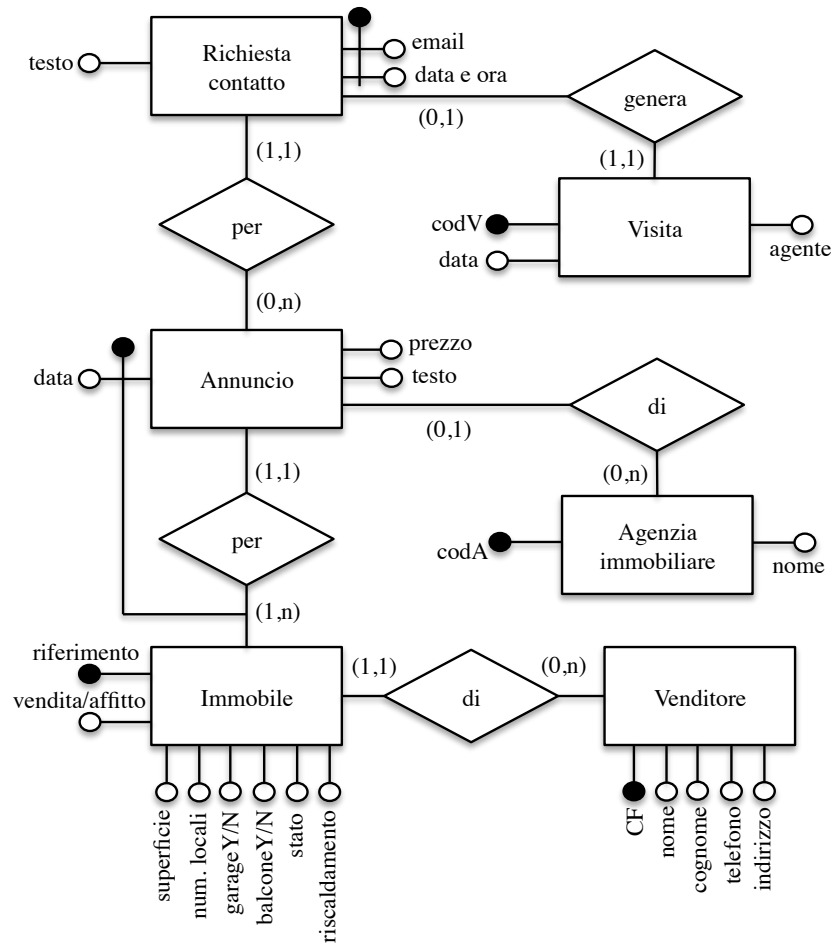
2) Date le gerarchie DFM in figura (tra parentesi le cardinalità di ciascun attributo), effettuarne la progettazione logica. Sapendo che ciascun attributo occupa 60 byte, mentre un surrogato ne richiede 4, stimare l'occupazione complessiva di memoria delle dimension table ed eventuali bridge table (8/32 punti).



3) Spiegare il ruolo della fase di *progettazione fisica* all'interno del ciclo di vita di un data mart (6/32 punti).

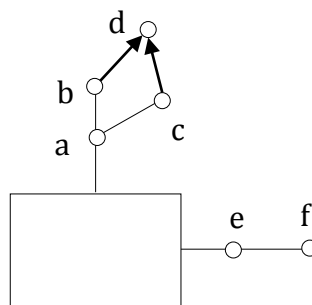
4) Discutere pro e contro degli approcci *top-down* e *bottom-up* alla progettazione di data warehouse (6/32 punti).

1) E' dato il seguente schema Entity/Relationship relativo a un portale di intermediazione immobiliare:



Il riscaldamento di un immobile può essere autonomo o centralizzato, mentre lo stato può essere nuovo, ristrutturato, buono, da ristrutturare; la superficie è espressa in metri quadri. Un immobile può avere più annunci in date diverse (con prezzi diversi), ma anche più annunci in contemporanea da diverse agenzie immobiliari. Alcuni annunci sono fatti dal venditore senza l'intermediazione di un'agenzia. Dopo avere effettuato la ricognizione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) Dato lo schema di fatto in figura, si elenchino le viste candidate alla materializzazione per un carico di lavoro comprendente i soli group-by bf, c e de (8/32 punti).



3) Definire i concetti di *data warehouse*, *data mart* e *operational data store* (6/32 punti).

4) Discutere pro e contro degli *schemi snowflake* rispetto agli schemi a stella (6/32 punti).

1) E' dato il seguente schema logico relazionale relativo all'applicazione che gestisce i rimborsi per le trasferte del personale all'Università di Bologna:

```
DIPENDENTI (codDip, nome, cognome, sesso, dataNascita, ruolo:RUOLI)
RUOLI (ruolo, massimaleSpesa)
RICHIESTE_AUTORIZZAZIONE (codRA, codDip:DIPENDENTI, dataRichiestaAutorizzazione,
    luogoTrasferta:CITTA', motivoTrasferta, dataInizioTrasferta, dataFineTrasferta,
    concessaYN)
CITTA' (codCittà, nomeCittà, nazione)
RICHIESTE_RIMBORSO (codRR, relativaARichiesta:RICHIESTE_AUTORIZZAZIONE,
    dataRichiestaRimborso)
DOCUMENTI_RIMBORSO (codDoc, codRR, importo, tipoSpesa, dataSpesa)
```

Il dipendente richiede l'autorizzazione prima della trasferta; se l'autorizzazione viene accordata, al ritorno chiede il rimborso producendo i documenti di spesa raccolti. L'attributo `ruolo` può assumere i seguenti valori: *tecnico*, *amministrativo*, *ricercatore*, *prof. associato*, *prof. ordinario*, mentre `tipoSpesa` può valere *aereo*, *treno*, *pasto*, *hotel*, *taxi*. Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) Uno schema di fatto (con cardinalità di eventi primari pari a 1M) include tre dimensioni degeneri: a (5 byte), b (20 byte) e c (50 byte), con cardinalità rispettivamente 100, 100 e 50. Individuare la soluzione di progettazione logica più conveniente dal punto di vista dello spazio occupato, assumendo che le chiavi surrogate occupino 4 byte (8/32 punti).

3) Spiegare la differenza tra *reportistica statica*, *OLAP* e *reportistica semi-statica* (6/32 punti).

4) Cos'è il *group-by set* di un'interrogazione OLAP e in che modo è implicitamente rappresentato nel DFM? (6/32 punti).

1) E' dato il seguente schema logico relazionale relativo all'accesso ai servizi (palestra, piscina, sauna, ping pong, sala business), registrati tramite una smartcard personale, da parte dei clienti di un residence:

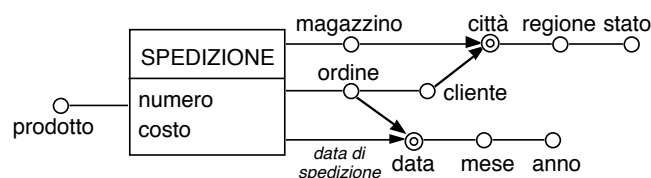
```

CLIENTE (CF, nome, cognome, numDocumento, sesso, indirizzo, CAP, città, nazione)
SOGGIORNO (idSoggiorno, CF:CLIENTE, daData, aData, numeroStanza, dataPrenotazione,
           stagione)
SERVIZIO (nomeServizio, piano, stanza)
ACCESSO (idSoggiorno:SOGGIORNO, dataAccesso, daOORMSS, aOORMSS,
          nomeServizio:SERVIZIO)
TARIFFARIO (nomeServizio:SERVIZIO, stagione, costoOrario)
    
```

Si noti che alcuni servizi sono gratuiti, altri a pagamento (la tariffa oraria varia a seconda della stagione alta, media o bassa). Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) Dato lo schema di fatto in figura, se ne effettui la progettazione logica (8/32 punti) supponendo che:

- con riferimento all'associazione tra cliente e città, gli utenti siano interessati agli scenari temporali di attualizzazione (oggi per ieri) e verità storica (oggi o ieri);
- con riferimento a tutte le altre associazioni, gli utenti siano interessati al solo scenario di verità storica;
- il progettista abbia deciso di creare due viste materializzate, una aggregata per {Mese dell'Ordine, Città del Magazzino}, una per {Prodotto, Mese di Spedizione}.



3) Si illustri, anche attraverso un esempio, la differenza tra *pulitura* e *trasformazione* dei dati (6/32 punti).

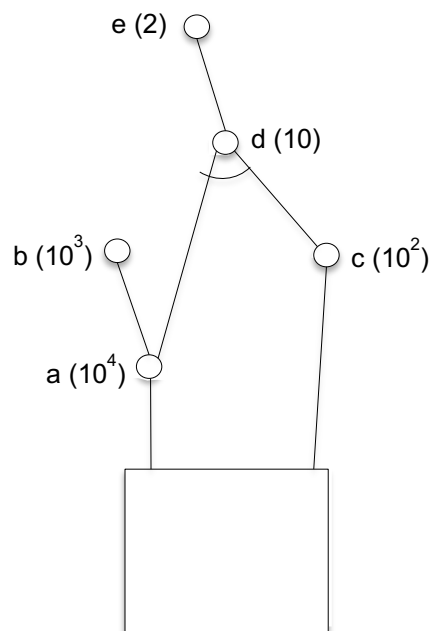
4) Si elenchino i fattori che incidono sulla *scelta dell'architettura* in un progetto di data warehouse (6/32 punti).

1) E' dato il seguente schema logico relazionale relativo a un ateneo multicampus, in cui uno studente può sostenere anche esami appartenenti a corsi di studio differenti da quello qui è iscritto in un certo anno accademico:

STUDENTE (CF, matricola, nome, cognome, sesso, via, numeroCivico, CAP, città, nazione, dataNascita)  
 DOCENTE (CF, codDocente, nome, cognome, sesso, dataNascita)  
 CORSODISTUDI (codCdS, nome, scuola, sedeCampus)  
 ISCRIZIONE (CF:STUDENTE, CodCdS:CORSODISTUDI, annoAccademico)  
 MATERIA (codMateria, nomeMateria, codCdS:CORSODISTUDI, settoreScientifico)  
 DOCENZA (codMateria, annoAccademico, codDocente)  
 ESAME (CF:STUDENTE, codMateria:MATERIA, data, annoAccademico, voto, lodeYN)

Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) Data la gerarchia DFM in figura (tra parentesi le cardinalità di ciascun attributo), ricavarne due soluzioni di progettazione logica. Sapendo che ciascun attributo occupa 100 byte (mentre un surrogato ne richiede 4), stimare l'occupazione complessiva di memoria delle tabelle coinvolte nelle due soluzioni (8/32 punti).



3) Illustrare il costrutto di *gerarchia condivisa* nel DFM, descrivendone le possibili implementazioni su piattaforma ROLAP (6/32 punti).

4) Definire la *business intelligence* e descrivere brevemente i livelli della piramide (6/32 punti).

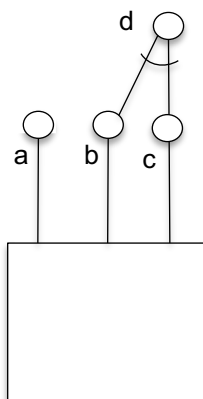


1) E' dato il seguente schema logico relazionale relativo ai preventivi compilati dai tecnici di un'azienda di posa impianti di climatizzazione:

```
TECNICI (CF, nome, cognome, sesso, CAP, città, nazione)
CLIENTI (CF, nome, cognome, via, numeroCivico, CAP, città, nazione)
PREVENTIVI (numPreventivo, anno, CFCliente:CLIENTE, dataPreventivo, importoTotale,
            scontoPercentuale, importoScontato, compilatoDa:TECNICI)
RIGHE ( (numPreventivo, anno) : PREVENTIVO, numRiga, codServizio:SERVIZI,
        codProdotto:PRODOTTI, quantità, costoUnitario, importo, IVAPercentuale,
        importoConIVA)
SERVIZI (codServizio, descrizione)
PRODOTTI (codProdotto, nome, marca, tipo, peso, potenza, inverterYN, rumorosità,
          classeEnergetica, colore, multiSplitYN, BTU)
```

Una riga di preventivo è relativa o a un servizio (per esempio montaggio) o a un prodotto (per esempio split). BTU è un campo numerico (valori possibili: 6000, 9000, 12000) che indica la capacità oraria di raffreddamento e riscaldamento del climatizzatore. Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) Dato lo schema di fatto in figura, se ne disegni il reticolo multidimensionale completo (8/32 punti).



3) Spiegare l'importanza dell'uso di *chiavi surrogate* nella progettazione logica di data mart (5/32 punti).

4) Illustrare il costrutto DFM di *arco multiplo*, spiegando in particolare i problemi che esso induce sull'aggregazione e come può essere tradotto in un'implementazione ROLAP (7/32 punti).

1) E' dato il seguente schema logico relazionale relativo a NextDoor, una app di "buon vicinato" attraverso la quale iscritti dello stesso quartiere pubblicano messaggi di varia natura:

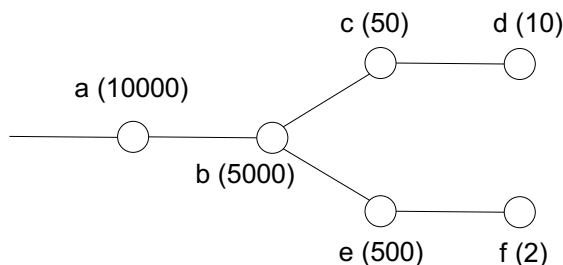
```

ISCRITTI (codIscritto, nome, cognome, sesso, via, civico, quartiere, comune, foto)
CATEGORIE_MESSAGGIO (codCat, tipo)
PAROLE_CHIAVE (parolaChiave)
INTERESSI (codIscritto:ISCRITTI, parolaChiave:PAROLE_CHIAVE)
MESSAGGI (codMessaggio, testo, pubblicatoDa:ISCRITTI, data, ora,
          categoriaMessaggio:CATEGORIE_MESSAGGI)
MESS_PC (codMessaggio:MESSAGGI, parolaChiave:PAROLE_CHIAVE)
RISPOSTE (codRisposta, inRispostaA:MESSAGGI, testo, pubblicatoDa:ISCRITTI, data, ora)
RINGRAZIAMENTI (codIscritto:ISCRITTI, risposta:RISPOSTE)

```

I quartieri sono una suddivisione dei comuni. Le categorie messaggio sono per esempio *offerta lavoro*, *richiesta lavoro*, *tempo libero*, *sport*. Un messaggio appartiene sempre a una categoria e può essere abbinato a parole chiave; esempi di parole chiave sono *passeggiate*, *palestra*, *musica*. Un iscritto può vedere solo i messaggi pubblicati da iscritti dello stesso quartiere, e può ad essi rispondere. Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

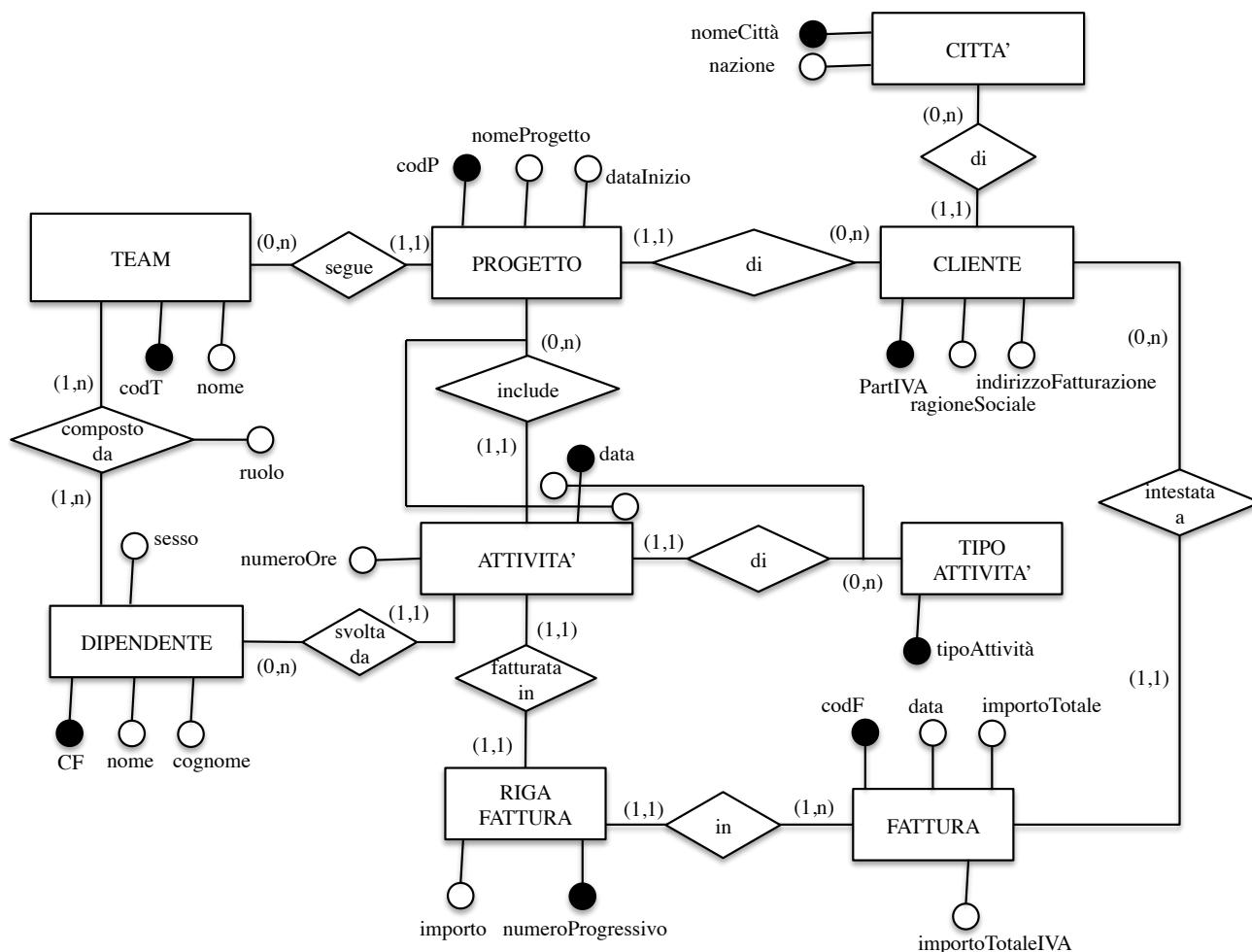
2) Dato il frammento di gerarchia in figura (tra parentesi le cardinalità di dominio degli attributi), determinare euristicamente i due punti di snowflaking potenzialmente più vantaggiosi dal punto di vista dello spazio di memoria occupato. Per ciascuno dei punti di snowflaking individuati disegnare lo schema logico risultante; supponendo che tutti gli attributi abbiano la stessa lunghezza, determinare tra i due schemi ottenuti quello che occupa meno spazio (8/32 punti).



3) Spiegare l'utilizzo del costrutto di gerarchia ricorsiva nel DFM, descrivendone le possibili implementazioni su piattaforma ROLAP (6/32 punti).

4) Elencare e descrivere brevemente le tecniche per l'estrazione incrementale dei dati dalle sorgenti (6/32 punti).

1) E' dato il seguente schema ER relativo a un'azienda di consulenza informatica.



Una fattura può includere righe relative ad attività svolte in progetti differenti, ma tutti dello stesso cliente (che coincide con il cliente cui la fattura è intestata). Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) Uno schema di fatto (con cardinalità di eventi primari pari a 1M) include tre dimensioni degeneri: a (10 byte), b (50 byte) e c (20 byte), con cardinalità rispettivamente 5, 50 e 20. Individuare la soluzione di progettazione logica più conveniente dal punto di vista dello spazio occupato, assumendo che le chiavi surrogate occupino 4 byte (8/32 punti).

3) Illustrare il *paradigma OLAP* per l'analisi dei dati multidimensionali (6/32 punti).

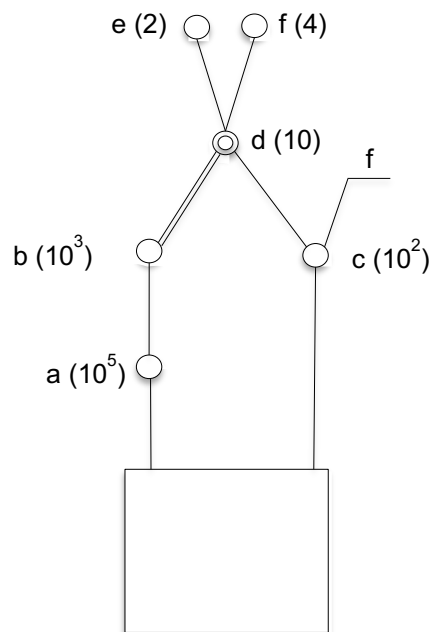
4) Illustrare la struttura di uno *star index*, spiegandone pro e contro (6/32 punti).

1) E' dato lo schema relazionale utilizzato dall'app ThePork per raccogliere prenotazioni a prezzi scontati presso ristoranti convenzionati:

```
RISTORANTI (codRistorante, nome, testoDescrittivo, via, civico, comune, nazione,
             tipoCucina, convenzionatoPuntiYN, punteggioTripAdvisor)
SLOT_ORARI (data, ora, codRistorante:RISTORANTI, sconto%, puntiFedeltà)
UTENTI (codUtente, nome, cognome, sesso, puntiFedeltàTotali, categoriaUtente)
PRENOTAZIONI (codUtente:UTENTI, (data, ora, codRistorante):SLOT_ORARI, numeroPersone,
              buonFineYN)
RECENSIONI (codUtente:UTENTI, data, codRistorante:RISTORANTI, testoRecensione, voto)
```

Ogni ristorante è disponibile in alcuni slot orari giornalieri, e in ciascuno con un determinato sconto (10%, 25%, 50%). Ogni prenotazione dà diritto a un certo numero di punti fedeltà (50, 100 o 200). L'attributo booleano `buonFineYN` indica se la prenotazione è stata onorata, ossia se l'utente si è effettivamente recato al ristorante, perché solo in caso affermativo verranno accreditati i corrispondenti punti fedeltà. Al raggiungimento dei 1000 punti fedeltà totali, l'utente ha diritto a uno sconto di 20 euro su una prenotazione effettuata presso uno dei ristoranti che accetta punti fedeltà (attributo booleano `convenzionatoPuntiYN`). Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) Data la gerarchia DFM in figura (tra parentesi le cardinalità di ciascun attributo), ricavarne una soluzione di progettazione logica. Sapendo che ciascun attributo occupa 100 byte (mentre un surrogato ne richiede 4), stimare l'occupazione complessiva di memoria delle tabelle coinvolte (8/32 punti).



3) Illustrare la differenza tra *operational data store* e *staging area* e i rispettivi ruoli architetturali (6/32 punti).

4) Spiegare le differenze tra *schemi snowflake* e *schemi a stella*, illustrando i rispettivi pro e contro (6/32 punti).

1) E' dato il seguente schema logico relazionale relativo a una palestra in cui si pratica Crossfit:

```

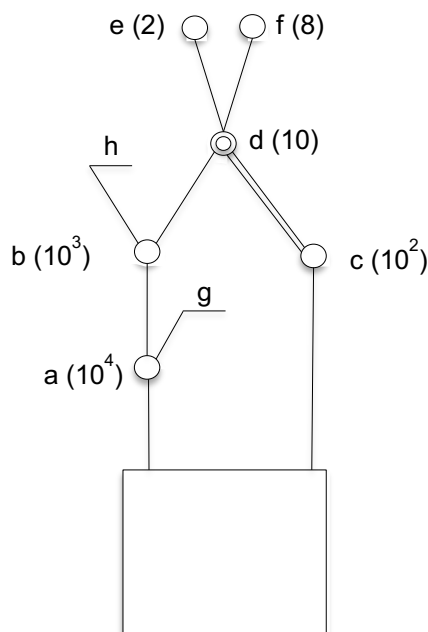
ISCRITTI (CF, nome, cognome, sesso, indirizzo, CAP, città, altezza, peso)
ESERCIZI (idEse, nome, tipoEse)
MASSIMALI (CF:ISCRITTI, idEse:ESERCIZI, peso, data)
WOD (data, nome, tipoWOD, numeroDiRound)
ES_IN_WOD (data:WOD, idEse:ESERCIZI, posizioneInSequenza, numeroRipetizioni)
ALLENAMENTI (data:WOD, ora, istruttore)
PRESENZA (CF:ISCRITTI, (data, ora):ALLENAMENTI)

```

Ogni giorno la palestra propone un Workout-Of-the-Day (WOD), che consiste di una sequenza di esercizi, ciascuno eseguito un certo numero di volte (*numeroRipetizioni*); l'intera sequenza deve essere ripetuta una o più volte (*numeroDiRound*). Per esempio, un WOD può essere costituito dalla sequenza *OverheadSquat* x 10 + *MuscleUp* x 5 da ripetersi 20 volte. Per ciascun esercizio, gli iscritti possono registrare i loro massimali (ossia il peso massimo con cui riescono a effettuare l'esercizio) nel tempo. Il WOD giornaliero si compone di più allenamenti che si tengono durante il giorno, a orari diversi. Ai fini dell'analisi interessa principalmente monitorare le presenze degli iscritti in palestra, analizzandole anche per singolo esercizio svolto.

Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) Data la gerarchia DFM in figura (tra parentesi le cardinalità di ciascun attributo), ricavarne una soluzione di progettazione logica. Sapendo che ciascun attributo occupa 50 byte (mentre un surrogato ne richiede 4), stimare l'occupazione complessiva di memoria delle tabelle coinvolte (8/32 punti).



3) Si elenchino le differenti situazioni in cui può risultare vantaggioso effettuare lo *snowflaking* di una dimensioni (6/32 punti).

4) Si illustrino le possibili soluzioni di progettazione logica a supporto della *materializzazione di viste* (6/32 punti).

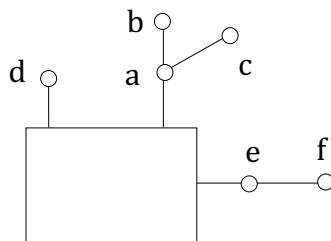
1) E' dato il seguente schema logico relazionale, relativo ai pagamenti dei condòmini degli stabili gestiti da un amministratore:

```

CONDOMINO (CodCondomino, cognome, nome, codStabile: STABILE, scala, piano, interno)
STABILE (codStabile, via, civico, cap, comune, regione, numeroDiPiani, numeroDiCondòmini)
STORICO_SALDO_CONTABILITA (CodCondomino: CONDOMINO, meseAnno, saldo)
RATA (CodCondomino: CONDOMINO, numeroRata, anno, dataScadenza, importo, pagataYN)
PAGAMENTO (CodPagamento, (CodCondomino, numeroRata, anno): RATA, importo, dataPagamento)
    
```

Si tenga presente che un CAP identifica un singolo comune, e che esistono comuni con lo stesso nome in regioni diverse; inoltre, sebbene l'importo del pagamento corrisponda in linea di massima a quello della relativa rata, in alcuni casi ci sono discrepanze dovute a errori. Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

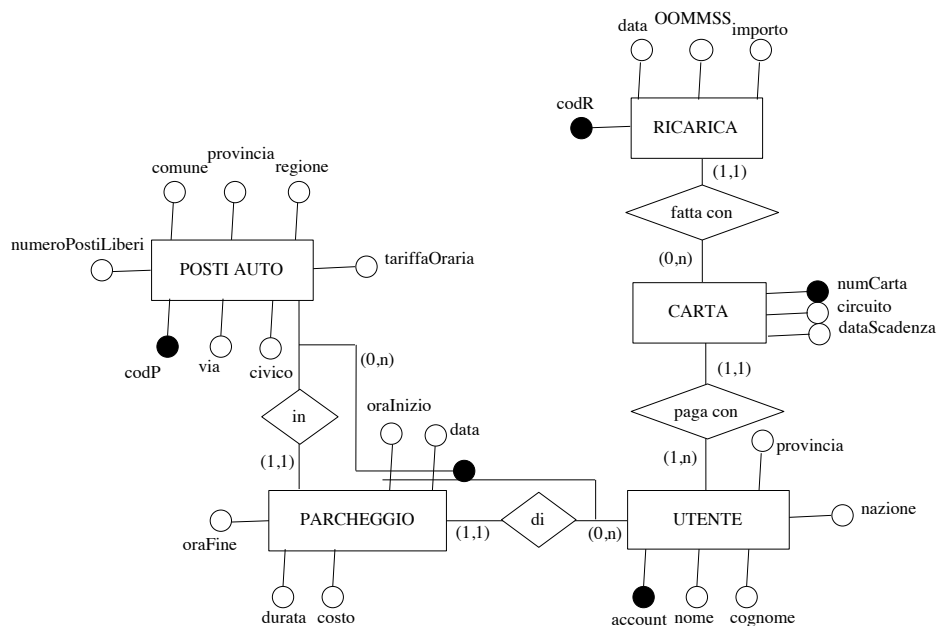
2) Dato lo schema di fatto in figura, si elenchino le viste candidate alla materializzazione per un carico di lavoro comprendente i soli group-by ad, ce e b (8/32 punti).



3) Illustrare le differenze tra *schemi snowflake* e *schemi a stella*, illustrando i rispettivi pro e contro (6/32 punti).

4) Spiegare cos'è la *business intelligence* (6/32).

1) E' dato il seguente diagramma E/R, su cui si appoggia un'app che permette di vedere i posti auto liberi sul territorio italiano e di pagare il parcheggio in base a una tariffa oraria:



(il costo del parcheggio è calcolato moltiplicando la tariffa oraria per la durata del parcheggio). Dopo avere effettuato la ricognizione e normalizzazione dello schema, **si individuino due fatti di interesse** e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) Ciascun cliente di un'azienda di e-commerce è seguito da un agente, che può cambiare nel tempo. Le tabelle mostrano gli assegnamenti degli agenti a tre clienti validi rispettivamente nel 2020 e nel 2021, e gli ordini effettuati da questi clienti in un periodo di tempo. Calcolare, sulla base dei dati disponibili, l'importo totale degli ordini per agente con riferimento a tre scenari temporali: attualizzazione, verità storica e retrodatazione al 31/12/2020 (8/32 punti).

1/1/2020		ORDINI		
cliente	agente	cliente	data	importo
Bianchi	Ferrari	Bianchi	15/11/2020	1000
Rossi	Esposito	Neri	3/12/2020	2000
Neri	Esposito	Neri	21/12/2020	150
1/1/2021		Neri	19/1/2021	100
cliente	agente	Bianchi	21/1/2021	300
Bianchi	Ferrari	Neri	4/2/2021	1500
Rossi	Esposito	Rossi	4/2/2021	100
Neri	Ferrari	Rossi	17/3/2021	2000

3) Cosa si intende per *architettura federata* in ambito data warehouse e in quali situazioni può risultare utile? (6/32 punti).

4) Si discutano pro e contro dei tre principali approcci all'analisi dei dati multidimensionali: *reportistica statica*, *OLAP* e *reportistica semi-statica* (6/32 punti).

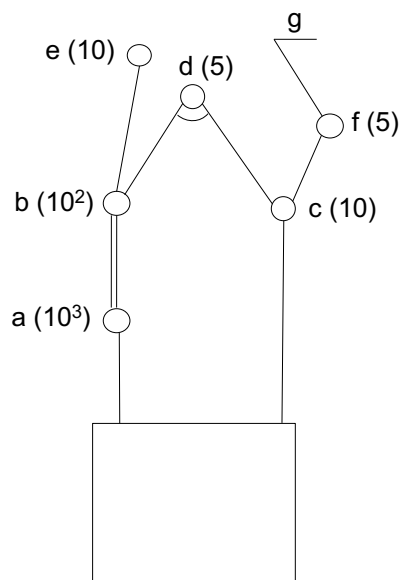
1) E' dato il seguente schema logico relazionale, relativo al database delle contravvenzioni elevate da vigili e ausiliari:

```

VIGILE(matricola, CF, nome, cognome, dataNascita, sesso, anzianitàServizio,
        comando:COMANDO_VIGILE)
COMANDO_VIGILE(codCV, via, civico, CAP, comune, provincia, regione)
AUSILIARIO(CF, nome, cognome, dataNascita, sesso, dataAssunzione)
VEICOLO(targa, modello, marca, cilindrata, alimentazione, annoImmatricolazione,
         intestataA:PROPRIETARIO)
PROPRIETARIO(CF, nome, cognome, dataNascita, sesso, via, civico, CAP, comune,
              provincia, regione, dataConseguimentoPatente)
INFRAZIONE(codV, descrizione, tipoInfrazione:TIPO_INFRAZIONE,
           decurtazionePuntiYN)
TIPO_INFRAZIONE(codTI, descrizione)
CONTRAVVENZIONE(numero, data, OOMM, elevataDaVigile:VIGILE,
                 elevataDaAux:AUSILIARIO, aVeicolo: VEICOLO, infrazione:INFRAZIONE,
                 viaInfrazione, comuneInfrazione, testoDescrizione, importo, puntiDecurtati)
  
```

L'anzianità di servizio di un vigile è calcolata come il numero di anni trascorsi dall'assunzione; l'alimentazione di un'auto può essere benzina, diesel, elettrica o gpl; un modello appartiene a una sola marca ma può avere differenti cilindrata e alimentazioni; una contravvenzione è elevata o da un vigile o da un ausiliario. Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

2) Data la gerarchia DFM in figura (tra parentesi le cardinalità di ciascun attributo), ricavarne una soluzione di progettazione logica. Sapendo che ciascun attributo occupa 20 byte (mentre un surrogato ne richiede 4), stimare l'occupazione complessiva di memoria delle tabelle coinvolte (8/32 punti).



3) Illustrare brevemente le tre principali soluzioni per la progettazione logica delle gerarchie temporali (slowly-changing dimension), con particolare riferimento agli scenari temporali che ciascuna di esse supporta (6/32 punti).

4) Spiegare, eventualmente con l'ausilio di un diagramma, come si effettua il caricamento delle dimension table e delle fact table (6/32 punti).



1) E' dato il seguente schema logico relazionale, relativo a una piattaforma per la visione di contenuti video:

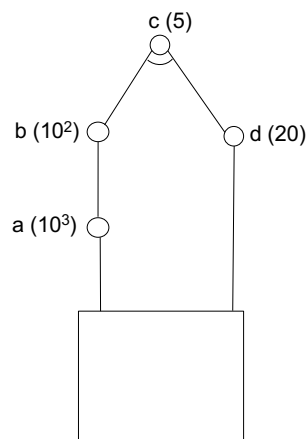
```

UTENTE (account, nome, adultoYN)
CONTENUTO (codC, titolo, regista, genere, trama)
ATTORE (nomeAttore)
ATTORE_IN_CONTENUTO (codC:CONTENUTO, nomeAttore:ATTORE)
FILM (codC:CONTENUTO, durata)
SERIE (codC:CONTENUTO, numeroStagioni, numeroTotaleEpisodi)
STAGIONE (codC:SERIE, numeroStagione)
EPISODIO ((codC, numeroStagione):STAGIONE, numeroEpisodio, titolo, durata)
VISIONE_FILM (account:UTENTE, codC:FILM, data, OOMMSinizio, durataVisione,
visioneCompletataYN, tipoDispositivo)
VISIONE_EPISODIO (account:UTENTE, (codC, numeroStagione, numeroEpisodio):EPISODIO,
data, OOMMSinizio, durataVisione, visioneCompletataYN, tipoDispositivo)

```

(tipoDispositivo può essere “TV”, “smartphone” o “PC”). Si tenga presente che sono di interesse anche analisi per attore. Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (12/32 punti).

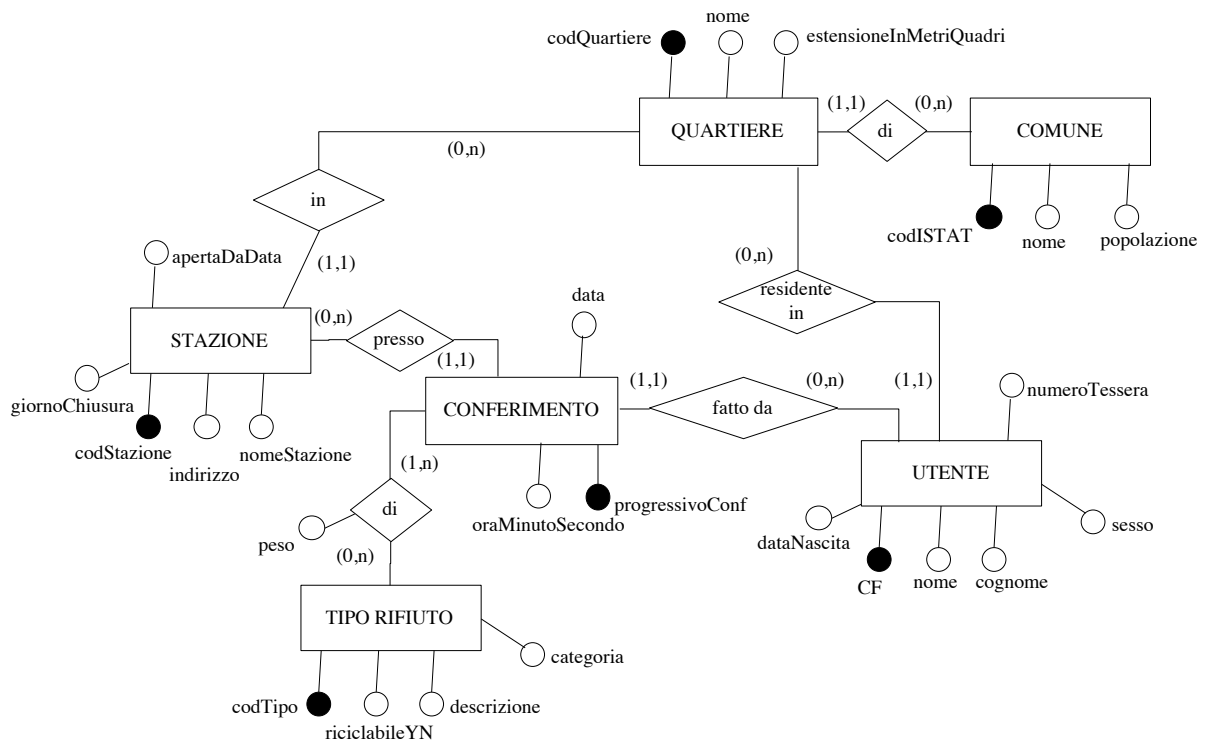
2) E' dato lo schema di fatto in figura, con cardinalità di eventi pari a 1M. Ciascun attributo occupa 50 byte. Individuare due possibili soluzioni di progettazione logica per l'attributo cross-dimensionale c, calcolando per ciascuna lo spazio occupato (i surrogati occupano 4 byte) (8/32 punti).



3) Descrivere sinteticamente le tre architetture “data mart indipendenti”, “data mart bus” e “hub-and-spoke”, indicando pregi e difetti di ciascuna di esse (6/32 punti).

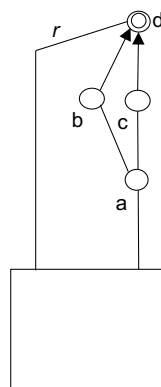
4) Descrivere la tecnica di *materializzazione delle viste* spiegandone pro e contro, e discutere in particolare il ruolo del reticolo multidimensionale (6/32 punti).

1) E' dato il seguente diagramma E/R, relativo ai conferimenti di rifiuti da parte di cittadini muniti di apposita tessera presso le stazioni ecologiche del territorio:



Si noti che, per i piccoli comuni, non esiste una suddivisione in quartieri; in questo caso, nel database operativo, gli utenti e le stazioni sono legati a un'istanza fittizia di quartiere con lo stesso nome del comune di appartenenza. Dopo avere effettuato la ricognizione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (10/31 punti).

2) Dato lo schema di fatto in figura, se ne disegni il reticolo multidimensionale completo (6/31 punti).



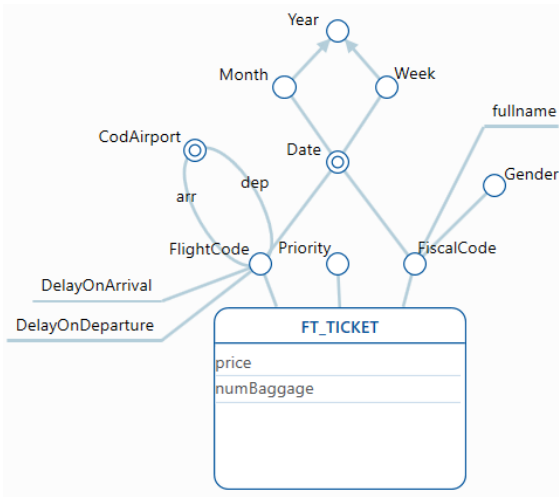
3) Spiegare punti in comune e differenze tra i costrutti di *gerarchia incompleta* e *gerarchia ricorsiva* nel DFM (5/31 punti).

4) Illustrare la struttura di uno *star index*, spiegandone pro e contro (4/31 punti).

**5) [OLAP]** Si acceda con Tableau Desktop al database *biexam* di PostgreSQL (137.204.78.85) con le seguenti credenziali.

- Username: biexam
- Password: 202201flights

Sullo schema *exam202201* è presente un'implementazione ROLAP del data warehouse relativo ad una compagnia aerea:



```
FT_TICKET(idCustomer:DT_CUSTOMER,  
idFlight:DT_FLIGHT,priority,price,numBaggage)
```

```
DT_FLIGHT(id,flightCode,  
idDateOfFlight:DT_DATE,  
delayOnDeparture,delayOnArrival,  
codAirportDeparture,codAirportArrival)
```

```
DT_CUSTOMER(id,fiscalCode,fullname,gender,  
idDateOfBirth:DT_DATE)
```

```
DT_DATE(id,date,week,month,year)
```

Si esegua il setup del cubo su Tableau Desktop e si esegua la seguente query.

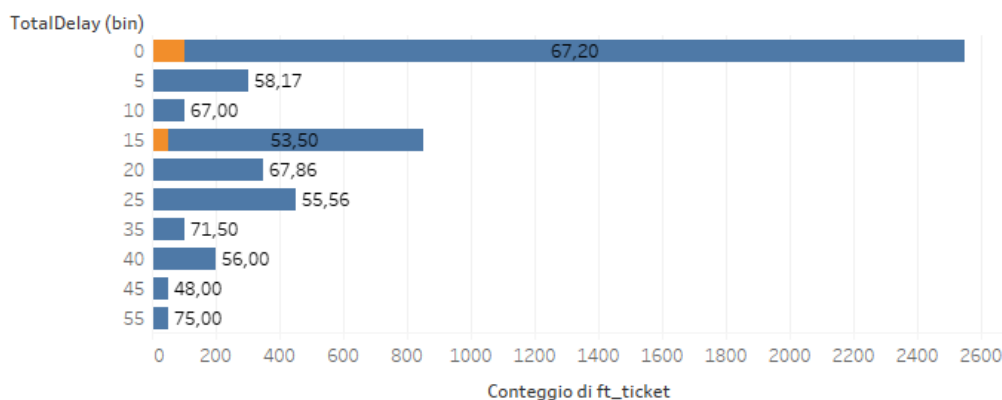
Group by set: Priority, TotalDelay (bin)

Filtri: -

Misure: avg(Price), count(\*)

dove TotalDelay è calcolato come somma di delayOnDeparture e delayOnArrival.

Si produca quindi un risultato simile al seguente.



Si consegna su EOL il file prodotto con Tableau Desktop (6/31 punti).

1) E' dato il seguente schema logico relazionale, relativo alle visite effettuate da medici di base e alle patologie riscontrate:

```
VISITA (idVisita, dataVisita, oraMinuto, durata, medico:MEDICO, paziente:PAZIENTE,
        dataPrenotazione, ambulatorioVia, ambulatorioCivico, ambulatorioComune,
        ambulatorioProvincia)
MEDICO (CFM, nome, cognome, sesso, dataNascita, numTelefono)
SPECIALITA' (CFM, specialità, dataConseguimento)
ORARI_DI_VISITA (CFM, giornoDellaSettimana, oraInizio, oraFine)
PAZIENTE (CFP, nome, cognome, sesso, dataNascita, numeroTesserSanitaria, residenzaVia,
          residenzaCivico, residenzaComune, residenzaProvincia)
ESENZIONE (CFP, esenzione, dataScadenza)
PATOLOGIA (codP, nome, descrizione, categoria)
PAT_IN_VISITA (codP:PATOLOGIA, idVisita:VISITA, gradoGravità)
```

(gradoGravità è un intero compreso tra 1 e 5). Si tenga presente che sono di interesse anche analisi per specialità. Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (10/32 punti).

2) Uno schema di fatto (con cardinalità di eventi primari pari a 5M) include tre dimensioni degeneri: a (4 byte), b (10 byte) e c (50 byte), con cardinalità rispettivamente 2, 100 e 20. Individuare la soluzione di progettazione logica più conveniente dal punto di vista dello spazio occupato; a tale scopo si valutino anche (seguendo un criterio euristico) soluzioni con junk table che includano un sottoinsieme delle dimensioni degeneri. Considerare che le chiavi surrogate occupino 4 byte (6/32 punti).

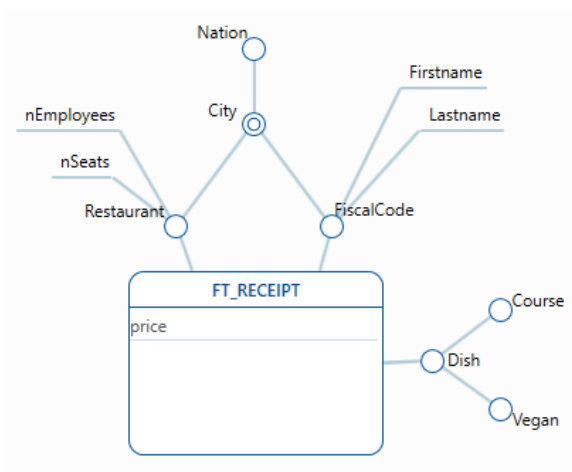
3) Discutere pro e contro degli approcci *top-down* e *bottom-up* alla progettazione di data warehouse (4/32 punti).

4) Si spieghi il ruolo della *tabella di look-up* nell'ETL, con particolare riferimento all'alimentazione di una dimension table gestita come slowly-changing dimension di tipo 1 (5/32 punti).

5) Si acceda con Tableau Desktop al database *biexam* di PostgreSQL (137.204.78.85) con le seguenti credenziali.

- Username: biexam
- Password: 202202receipts

Sullo schema *exam202202* è presente un'implementazione ROLAP del data warehouse relativo ad una catena di ristoranti:



```
FT_RECEIPT (idRestaurant:DT_RESTAURANT,
            idCustomer:DT_CUSTOMER, idDish:DT_DISH, price)
```

```
DT_RESTAURANT (id, idCity:DT_CITY, name,
               nEmployees, nSeats)
```

```
DT_CUSTOMER (id, idCity:DT_CITY, fiscalCode,
             firstname, lastname)
```

```
DT_DISH (id, name, course, vegan)
```

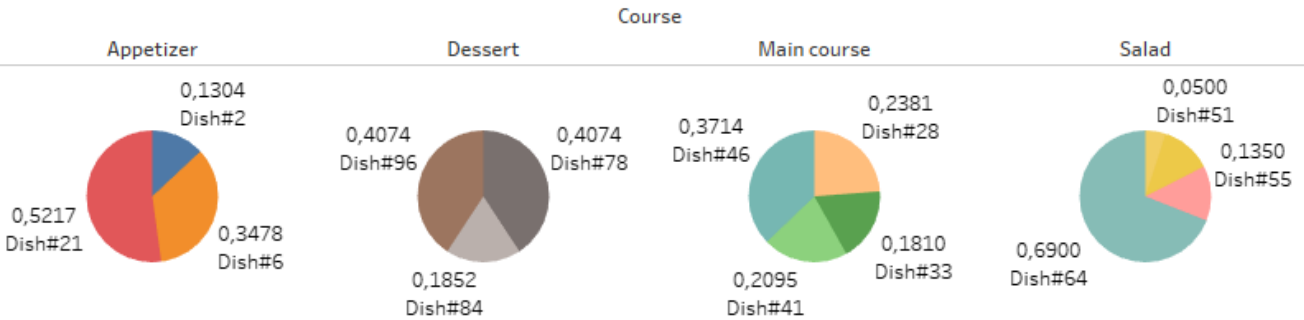
```
DT_CITY (id, city, nation)
```

Si esegua il setup del cubo su Tableau Desktop e si esegua la seguente query.

```
Group by set: Course, Dish
Filtri: Vegan = "Y", FiscalCode IN (Top10Customers)
Misure: PercentageOfPriceByCourse
```

Dove Top10Customers è l'insieme dei 10 clienti con la miglior media di Price, mentre PercentageOfPriceByCourse è il valore percentuale di Price del Dish rispetto al totale per Course.

Il risultato deve essere mostrato attraverso grafici a torta che mostrano la distribuzione dei piatti vegani rispetto al valore di PercentageOfPriceByCourse (calcolato sui soli piatti vegani dei Top10Customers) come segue.



Si consegna il file prodotto con Tableau Desktop su EOL.

```
=====
FT(a,idB,idC) 12x5x1000000+14x100+54x20
JT(idJ,a,b,c) 4x5x1000000+68x4000
JT(idJ,a,c) 8x5x1000000+14x100+58x40
```

## Compito di Business Intelligence - 10/6/2022      Tempo concesso : 120 minuti

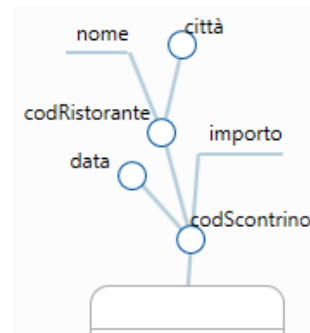
1) Si acceda con Tableau Prep al database *bi* di PostgreSQL (137.204.78.85) con le stesse credenziali utilizzate in laboratorio. Ogni utente ha accesso in lettura alle tabelle dello schema *exam202206*, in cui è implementata una parte del seguente schema logico relazionale, relativo ad una catena di fast-food:

```
RISTORANTE (codRistorante, nome, numCoperti, numDipendenti, indirizzo, città)
SCONTRINO (codScontrino, codRistorante:RISTORANTE, data, importo)
DETTAGLIO_SCONTRINO (codScontrino:SCONTRINO, codProdotto:PRODOTTO, importo)
PRODOTTO (codProdotto, nome, nomeAbbreviato, prezzo, categoria)
```

Si consideri la dimensione dello scontrino (DFM in figura) con il seguente schema logico:

```
DT_SCONTRINO (idScontrino codScontrino, data, importo,
codRistorante, nome, città)
LOOKUP_DT_SCONTRINO (idScontrino, codScontrino)
```

Si richiede di implementare (ed eseguire) su Tableau Prep i flussi per mantenere la tabella di lookup e per gestire l'alimentazione incrementale della dimension table. Si consegna il file prodotto con Tableau Prep via email a [enrico.gallinucci@unibo.it](mailto:enrico.gallinucci@unibo.it) (6/32 punti).

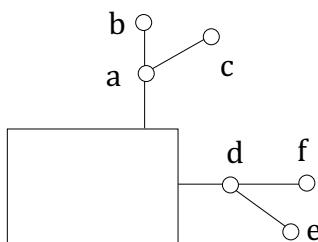


2) E' dato il seguente schema logico relazionale, relativo alle lauree presso un'università:

```
CITTA' (codC, nomeCittà, provincia, regione)
STUDENTE (matricola, CF, nome, cognome, dataNascita, sesso, indirizzo, cittàNascita:CITTA')
DOCENTE (matricola, CF, nome, cognome, dataNascita, sesso, indirizzo,
cittàNascita:CITTA', ruolo, dataPresadServizio)
SCUOLA (codS, nomeScuola, area)
CORSO_DI_STUDI (codC, denominazione, scuolaDiRiferimento:SCUOLA,
scuolaSecondaria:SCUOLA, coordinatore:DOCENTE, sedePrincipale:CITTA')
MATERIA (codM, nomeMateria, CFU, corso:CORSO_DI_STUDI, titolare:DOCENTE)
LAUREA (studente:STUDENTE, corsoDiStudi:CORSO_DI_STUDI, dataLaurea, mediaEsami,
votoLaurea, titoloTesi, materia:MATERIA, annoAccademico, correlatore:DOCENTE)
```

Si tenga presente che un corso di studi (per esempio, Ingegneria e Scienze Informatiche) ha sempre una scuola di riferimento (per esempio, Ingegneria) e può avere una scuola secondaria (per esempio, Informatica); il correlatore di tesi è opzionale. La materia di laurea (per esempio, Business Intelligence) deve appartenere allo stesso corso di studi in cui lo studente si laurea. Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individui un fatto di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (10/32 punti).

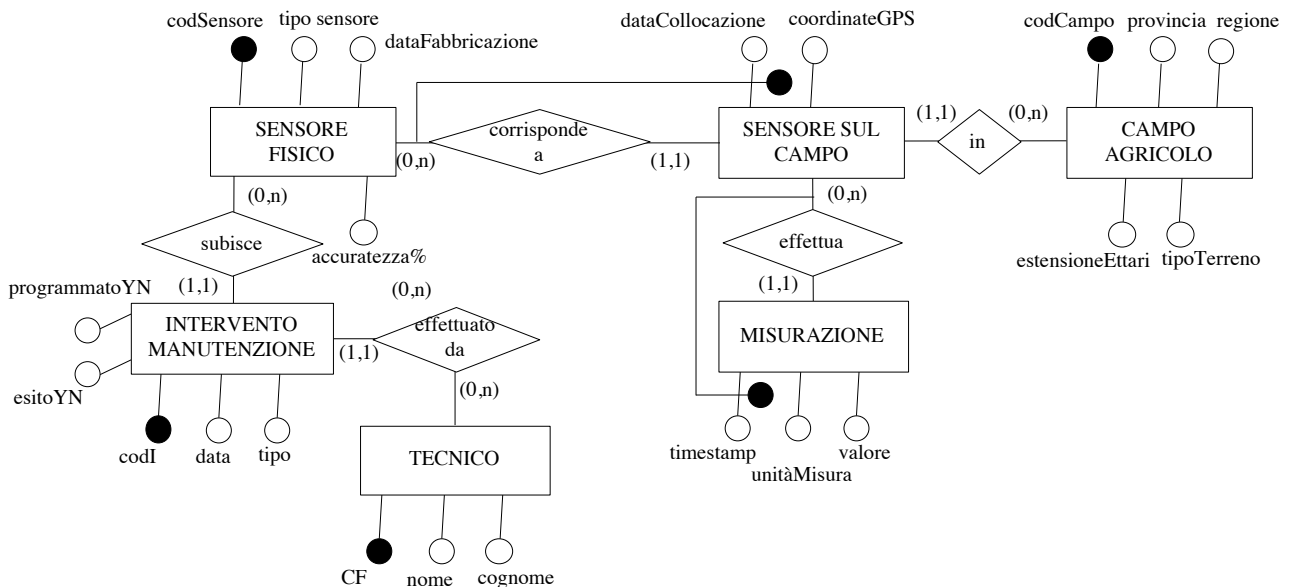
3) Dato lo schema di fatto in figura, si elenchino le viste candidate alla materializzazione per un carico di lavoro comprendente i soli group-by *af*, *ce* e *b* (6/32 punti).



4) Discutere pro e contro delle diverse tecniche per l'estrazione incrementale dei dati dalle sorgenti (5/32 punti).

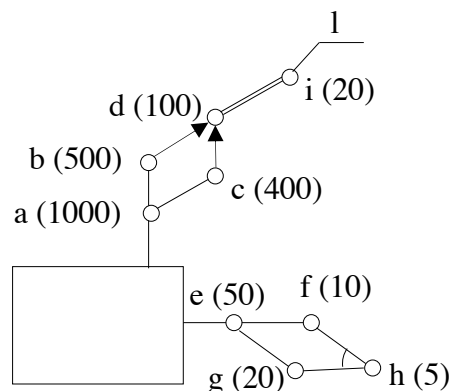
**5)** Illustrare la struttura e il funzionamento di un *bitmap index*, evidenziandone in particolare i pro e i contro rispetto a un B<sup>+</sup>-tree (4/32 punti).

1) E' dato il seguente diagramma ER, relativo ai sensori (per esempio, di umidità e temperatura) collocati in terreni agricoli:



In momenti diversi, uno stesso sensore fisico può essere collocato in posizioni differenti. Dopo avere effettuato la ricognizione e normalizzazione dello schema, si individuino **due** fatti di interesse e se ne disegni lo schema concettuale secondo il DFM e lo schema logico per un'implementazione ROLAP (10/31 punti).

2) Date le gerarchie DFM in figura (tra parentesi le cardinalità di ciascun attributo), ricavarne una soluzione di progettazione logica. Sapendo che ciascun attributo occupa 10 byte (mentre un surrogato ne richiede 4) e che ciascun membro di  $d$  corrisponde in media a 5 membri di  $i$  attraverso l'acrop multiplo, stimare l'occupazione complessiva di memoria delle tabelle coinvolte (6/31 punti).



3) Si elenchino le differenti situazioni in cui può risultare vantaggioso effettuare lo *snowflaking* di una dimensione (5/31 punti).

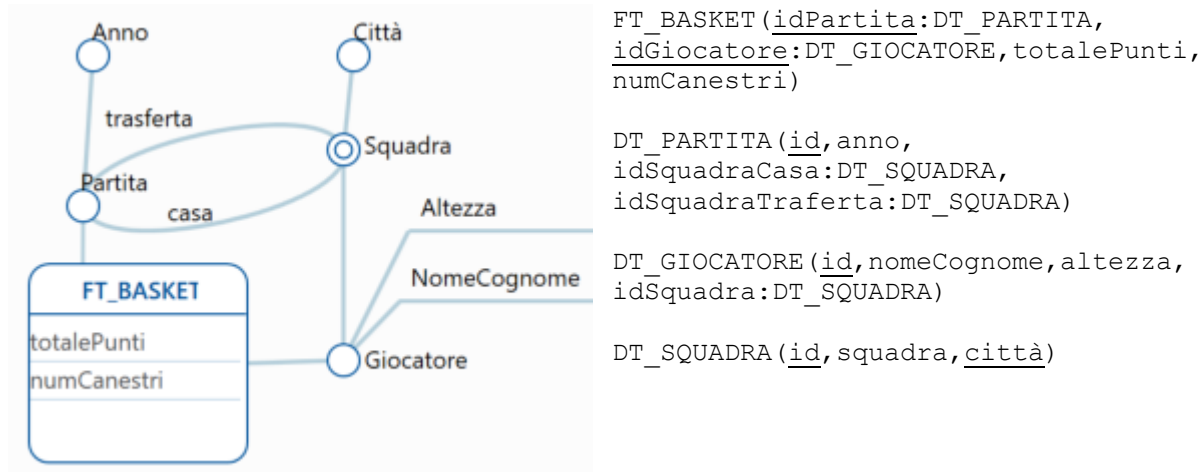
4) Spiegare il ruolo della fase di *progettazione fisica* all'interno del ciclo di vita di un data mart (4/31 punti).



5) Si acceda con Tableau Desktop al database *biexam* di PostgreSQL (137.204.78.85) con le seguenti credenziali.

- Username: biexam
- Password: 202207basket

Sullo schema *exam202207* è presente un'implementazione ROLAP del data warehouse relativo ad una lega di basket:



Si esegua il setup del cubo su Tableau Desktop e si esegua la seguente query, per vedere – anno dopo anno – il totale dei punti realizzati dai giocatori in base alle loro altezze.

Group by set: AltezzaBin, Anno  
 Filtri: [\*]  
 Misure: SUM(totalePunti)

AltezzaBin è l'Altezza raggruppata in bin di dimensione 10.

[\*] Come filtro vanno considerate le sole squadre il cui totalePunti (calcolato senza considerare i singoli eventi in cui il numCanestri è  $\leq 1$ ) è  $\geq 10.000$  (attenzione alla dipendenza tra i due vincoli).

Il risultato deve essere mostrato attraverso una semplice matrice come segue.

Anno	Altezza (bin)				
	170	180	190	200	210
2010	5.093	5.570	4.404	4.900	6.213
2011	6.233	5.604	4.918	6.007	6.689
2012	5.712	5.480	4.408	5.256	5.320
2013	6.713	5.968	5.286	6.365	6.271
2014	6.185	6.167	5.249	5.762	5.901

Si consegna il file prodotto con Tableau Desktop su EOL via email a [enrico.gallinucci@unibo.it](mailto:enrico.gallinucci@unibo.it) (6/31 punti).