

Tecnologie e architetture per la gestione dei dati

Prova parziale — 13 giugno 2022

Tempo a disposizione: un'ora e trenta minuti.

Cognome _____ Nome _____ Matricola _____

Domanda 1 (75%) La base di dati seguente descrive le spedizioni di un corriere che lavora in Italia:

- SPEDIZIONI(Codice, DataSpedizione, DataConsegna, Mittente, Destinatario, Peso, Costo) con vincoli di riferimento da Mittente verso AZIENDE e da Destinatario ancora verso AZIENDE
- AZIENDE(Codice, Nome, Via, Città, Categoria) con vincolo di riferimento da Categoria verso CATEGORIE e da Città verso CITTÀ
- CITTÀ (Codice, Nome, Regione)
- CATEGORIA (Codice, Descrizione); ad esempio: “base”, “premium”, “top”

Con riferimento a tale base di dati, progettare uno schema dimensionale (specificando grana, misure e dimensioni) che permetta di calcolare

- numero di spedizioni
- costo complessivo
- numero medio di giorni impiegato (differenza fra data di spedizione e data di consegna)
- peso medio delle spedizioni

distinguendo sulla base:

- della data di spedizione
- del mittente o della sua categoria
- della città di partenza o di quella di destinazione
- del peso (distinguendo tre gruppi: “lettere,” fino a 1Kg; “pacchetti” fra 1Kg e 5Kg; “pacchi” oltre 5Kg; supporre che questa classificazione non vari mai)
- del tipo di percorso (“locale,” nella stessa città; “regionale,” fra città diverse della stessa regione; “interregionale”)

Si noti che:

- per (presunte) ragioni di privacy, non si vuole avere traccia del destinatario nello schema dimensionale (se non attraverso la città)
- la categoria di una azienda può variare nel tempo; la sede invece non cambia mai

Specificare la grana e gli schemi delle tabelle dei fatti e delle dimensioni, con brevi commenti se necessario

Illustrare sinteticamente ma in modo chiaro il procedimento di calcolo della tabella dei fatti e di quelle delle dimensioni (escludendo quelle ovvie) a partire dai dati disponibili. Continuare alla pagina seguente se necessario (vengono per comodità di lettura ripetute le specifiche)

Tecnologie e architetture per la gestione dei dati — 13 giugno 2022

Specifiche ripetute per comodità:

- SPEDIZIONI(Codice, DataSpedizione, DataConsegna, Mittente, Destinatario, Peso, Costo) con vincoli di riferimento da Mittente verso AZIENDE e da Destinatario ancora verso AZIENDE
- AZIENDE(Codice, Nome, Via, Città, Categoria) con vincolo di riferimento da Categoria verso CATEGORIE e da Città verso CITTÀ
- CITTÀ (Codice, Nome, Regione)
- CATEGORIA (Codice, Descrizione); ad esempio: “base”, “premium”, “top”

Con riferimento a tale base di dati, progettare uno schema dimensionale (specificando grana, misure e dimensioni) che permetta di calcolare

- numero di spedizioni
- costo complessivo
- numero medio di giorni impiegato (differenza fra data di spedizione e data di consegna)
- peso medio delle spedizioni

distinguendo sulla base:

- della data di spedizione
- del mittente o della sua categoria
- della città di partenza o di quella di destinazione
- del peso (distinguendo tre gruppi: “lettere,” fino a 1Kg; “pacchetti” fra 1Kg e 5Kg; “pacchi” oltre 5Kg; supporre che questa classificazione non vari mai)
- del tipo di percorso (“locale,” nella stessa città; “regionale,” fra città diverse della stessa regione; “interregionale”)

Si noti che:

- per (presunte) ragioni di privacy, non si vuole avere traccia del destinatario nello schema dimensionale (se non attraverso la città)
- la categoria di una azienda può variare nel tempo; la sede invece non cambia mai

Tecnologie e architetture per la gestione dei dati — 13 giugno 2022

Domanda 2 (25%) Si supponga di voler aggiungere anche informazioni sulle eventuali emergenze, relative al percorso di una certa spedizione, nella data in cui essa deve partire: si supponga di avere a disposizione una tabella che, per ogni coppia di città e per ogni data, rappresenti l'eventuale presenza di problemi (indicata con un aggettivo che descrive il livello del problema), ad esempio con la forma seguente:

Data	Partenza	Destinazione	Problema
15/03/2022	Roma	Roma	Lieve
16/03/2022	Roma	Milano	Grave
18/03/2022	Milano	Milano	Gravissimo
...

Discutere brevemente come modificare lo schema dimensionale e il suo contenuto, per tenere conto di questa esigenza

Tecnologie e architetture per la gestione dei dati

Prova parziale — 13 giugno 2022

Cenni sulle soluzioni

Tempo a disposizione: un'ora e trenta minuti.

Cognome _____ **Nome** _____ **Matricola** _____

Domanda 1 (75%) La base di dati seguente descrive le spedizioni di un corriere che lavora in Italia:

- SPEDIZIONI(Codice, DataSpedizione, DataConsegna, Mittente, Destinatario, Peso, Costo) con vincoli di riferimento da Mittente verso AZIENDE e da Destinatario ancora verso AZIENDE
- AZIENDE(Codice, Nome, Via, Città, Categoria) con vincolo di riferimento da Categoria verso CATEGORIE e da Città verso CITTÀ
- CITTÀ (Codice, Nome, Regione)
- CATEGORIA (Codice, Descrizione); ad esempio: “base”, “premium”, “top”

Con riferimento a tale base di dati, progettare uno schema dimensionale (specificando grana, misure e dimensioni) che permetta di calcolare

- numero di spedizioni
- costo complessivo
- numero medio di giorni impiegato (differenza fra data di spedizione e data di consegna)
- peso medio delle spedizioni

distinguendo sulla base:

- della data di spedizione
- del mittente o della sua categoria
- della città di partenza o di quella di destinazione
- del peso (distinguendo tre gruppi: “lettere,” fino a 1Kg; “pacchetti” fra 1Kg e 5Kg; “pacchi” oltre 5Kg; supporre che questa classificazione non vari mai)
- del tipo di percorso (“locale,” nella stessa città; “regionale,” fra città diverse della stessa regione; “interregionale”)

Si noti che:

- per (presunte) ragioni di privacy, non si vuole avere traccia del destinatario nello schema dimensionale (se non attraverso la città)
- la categoria di una azienda può variare nel tempo; la sede invece non cambia mai

Specificare la grana e gli schemi delle tabelle dei fatti e delle dimensioni, con brevi commenti se necessario

Grana: insieme di spedizioni relative a

- una data
- un mittente (che include la città di partenza)
- una città di destinazione
- una fascia di peso
- un tipo di percorso (dimensione secondaria, perché funzione delle città di partenza e destinazione)

Schema

- FATTISPEDIZIONI(KData, KMittente, KDestinazione, KFasciaPeso, KTipoPercorso, NumeroSpedizioni, CostoTotale, PesoMedio, NumeroGiorniMedio)
- DIMDATA(KData, ...)
- DIMMITTENTE(KVersioneMittente, KMittente, ..., CodiceCategoria, DescrizioneCategoria, CodiceCittàMittente, NomeCittàMittente, RegioneMittente, ...)
- DIMDESTINAZIONE(KDestinazione, CodiceCittàDestinazione, NomeCittàDestinazione, RegioneDestinazione...)
- DIMFASCIAPESO(KFasciaPeso, DescrizioneFasciaPeso, Minimo, Massimo...)
- DIMTIPOPERCORSO(KTipoPercorso, DescrizioneTipoPercorso)

Commenti:

- sono indicate chiavi ad hoc per le dimensioni e anche per il mittente (cosa forse non indispensabile)
- per il destinatario è rappresentata solo la città
- la dimensione del mittente varia lentamente, per il possibile cambio di categoria
- DimTipoPercorso è una dimensione secondaria, calcolabile sulla base del mittente e della città di destinazione
- il peso effettivo non può comparire nella dimensione della fascia di peso

Illustrare sinteticamente ma in modo chiaro il procedimento di calcolo della tabella dei fatti e di quelle delle dimensioni (escludendo quelle ovvie) a partire dai dati disponibili. Continuare alla pagina seguente se necessario (vengono per comodità di lettura ripetute le specifiche)

Una possibile procedura

- join di SPEDIZIONI con AZIENDE su Mittente e ancora con AZIENDE su Mittente (con opportune ridenominazioni) e poi anche due volte con CITTÀ (questi ultimi due join servono per avere le due regioni e ovviamente richiedono anch'essi ridenominazioni); il risultato di questi join è una relazione sugli attributi

Cod, DataSped, DataCons, CodMitt, CatMitt, CodCittàM, RegioneM, CodCittàD, RegioneD, Peso, Costo

- Alla relazione ottenuta al passo precedente si possono aggiungere attributi (o codici) relativi a
 - fascia di peso: calcolabile anche in SQL con opportune interrogazioni condizionali; si può inserire un valore esplicativo (“lettera”, ...)
 - tipo di percorso: pure calcolabile in SQL a partire dalle due città e regioni; anche qui si può inserire un valore esplicativo
 - tempo impiegato (in giorni, come differenza fra le due date)

In questo modo, eliminando gli attributi che non servono più si può ottenere una relazione sugli attributi

Cod, DataSped, CodMitt, CatMitt, CodCittàD, FasciaPeso, TipoPercorso, Peso, Costo, NumeroGiorni

- A questo punto si può eseguire l'aggregazione (sulle dimensioni e anche sulla categoria, così si gestisce la variabilità), calcolando le misure, ottenendo una relazione su

DataSped, CodMitt, CatMitt, CodCittàD, FasciaPeso, TipoPerc, NumSped, PesoMedio, CostoTot, NumGiorniMed

- Infine, si sostituiscono le chiavi surrogate alle chiavi di produzione

Tecnologie e architetture per la gestione dei dati — 13 giugno 2022

Specifiche ripetute per comodità:

- SPEDIZIONI(Codice, DataSpedizione, DataConsegna, Mittente, Destinatario, Peso, Costo) con vincoli di riferimento da Mittente verso AZIENDE e da Destinatario ancora verso AZIENDE
- AZIENDE(Codice, Nome, Via, Città, Categoria) con vincolo di riferimento da Categoria verso CATEGORIE e da Città verso CITTÀ
- CITTÀ (Codice, Nome, Regione)
- CATEGORIA (Codice, Descrizione); ad esempio: “base”, “premium”, “top”

Con riferimento a tale base di dati, progettare uno schema dimensionale (specificando grana, misure e dimensioni) che permetta di calcolare

- numero di spedizioni
- costo complessivo
- numero medio di giorni impiegato (differenza fra data di spedizione e data di consegna)
- peso medio delle spedizioni

distinguendo sulla base:

- della data di spedizione
- del mittente o della sua categoria
- della città di partenza o di quella di destinazione
- del peso (distinguendo tre gruppi: “lettere,” fino a 1Kg; “pacchetti” fra 1Kg e 5Kg; “pacchi” oltre 5Kg; supporre che questa classificazione non vari mai)
- del tipo di percorso (“locale,” nella stessa città; “regionale,” fra città diverse della stessa regione; “interregionale”)

Si noti che:

- per (presunte) ragioni di privacy, non si vuole avere traccia del destinatario nello schema dimensionale (se non attraverso la città)
- la categoria di una azienda può variare nel tempo; la sede invece non cambia mai

Domanda 2 (25%) Si supponga di voler aggiungere anche informazioni sulle eventuali emergenze, relative al percorso di una certa spedizione, nella data in cui essa deve partire: si supponga di avere a disposizione una tabella che, per ogni coppia di città e per ogni data, rappresenti l'eventuale presenza di problemi (indicata con un aggettivo che descrive il livello del problema), ad esempio con la forma seguente:

Data	Partenza	Destinazione	Problema
15/03/2022	Roma	Roma	Lieve
16/03/2022	Roma	Milano	Grave
18/03/2022	Milano	Milano	Gravissimo
...

Discutere brevemente come modificare lo schema dimensionale e il suo contenuto, per tenere conto di questa esigenza

1. Si può introdurre una dimensione con una chiave surrogata e la descrizione dei problema
2. La tabella dei fatti (la cui grana non cambia) può essere modificata ripartendo dalla prima tabella preliminare, con entrambe le città (o loro codici, ma la sostituzione è semplice): con un join fra tale tabella preliminare e la tabella mostrata sopra nelle specifiche della domanda si può aggiungere la descrizione del problema alla tabella preliminare dei fatti, in cui poi si può sostituire la descrizione con la chiave surrogata (o qualcosa di simile)