

Modelowanie wielowymiarowe hurtowni danych

18 listopada 2017

Opis pliku z zadaniami

Wszystkie zadania na zajęciach będą przekazywane w postaci plików `.pdf`, sformatowanych podobnie do tego dokumentu. Zadania będą różnego rodzaju. Za każdym razem będą one odpowiednio oznaczone:

- Zadania do wykonania na zajęciach oznaczone są symbolem $\triangle$ – nie są one punktowane, ale należy je wykonać w czasie zajęć.
- Punktowane zadania do wykonania na zajęciach oznaczone są symbolem $\diamond$ – należy je wykonać na zajęciach i zaprezentować prowadzącemu.
- Zadania do wykonania w domu oznaczone są symbolem $\star$ – są one punktowane, należy je dostarczyć w sposób podany przez prowadzącego i w wyznaczonym terminie (zwykle przed kolejnymi zajęciami).

1 Hurtownia danych dla sieci kin



Treść

Zaprojektuj hurtownię danych dla sieci kin. Kina wyświetlają filmy w trzech porach dnia (popołudniu, wieczorem, nocą). Kina posiadają adres (ulica, rejon miasta, miasto). Każde kino ma wiele sal. Kina posiadają różnego rodzaju bilety zależne od miejsca siedzenia, sali, długości filmu, pory dnia. Wyświetlane filmy posiadają tytuł, kategorię, czas trwania, studio, które zrealizowało film oraz wersję językową dźwięku i podpisów. Część widzów posiada karty lojalnościowe, dzięki czemu znana jest płeć, adres oraz wiek widza. Widzowie z kartami lojalnościowymi mogą ocenić film od 1 do 5 gwiazdek. Celem hurtowni danych jest analiza sprzedaży biletów oraz ocen udzielanych przez klientów. Hurtownia danych ma na celu odpowiedzieć na następujące pytania:

1. Które filmy przyniosły największe zyski?
2. Które filmy zostały najlepiej ocenione?
3. Podaj charakterystykę widza, który ocenia najwyżej oraz najniżej?
4. Dla każdej kategorii filmu podaj widza, który najlepiej ją ocenia.
5. Dla każdego filmu podaj, dla której pory dnia miał on najwięcej widzów.
6. Dla danego filmu podaj liczbę widzów na każdym seansie.
7. Dla każdego kina podaj salę, która przynosi największe zyski?
8. Podaj miesiące, w których filmy są najlepiej oceniane?

Zadania do wykonania

1. Narysuj schemat hurtowni danych w postaci gwiazdy. Wskaż tabelę faktów oraz tabele wymiarów. W każdej tabeli uwzględnij wszystkie konieczne atrybuty. W tabeli faktów wskaż atrybuty, które są miarami. Krótko opisz i uzasadnij wybór ziarna tabeli faktów. Krótko opisz tabele wymiarów.
2. Czy w zaproponowanym przez Ciebie modelu występują naturalne hierarchie wymiarów? Pokaż dwie takie hierarchie.
3. Przerysuj schemat normalizując jeden z wymiarów. Nie musisz przepisywać atrybutów pozostałych wymiarów.

4. Napisz zapytania SQL dla wszystkich pytań przedstawionych w opisie studium przypadku. Pamiętaj, że wyniki zapytań mają mieć postać raportu, więc muszą się one odwoływać do wszystkie potrzebnych danych w sposób czytelny.

2 Firma telekomunikacyjna



Treść

W problemie modelowania hurtowni danych dla firmy telekomunikacyjnej natrafiono na następujący problem. Niektórzy klienci często zmieniają numer telefonu lub posiadają ich wiele.

Zamodeluj hurtownię danych dla firmy telekomunikacyjnej tak, aby w łatwy sposób mieć dostęp do informacji o wartości klienta (np. liczba rozmów telefonicznych danego klienta, suma rachunków telefonicznych dla różnych numerów telefonów), jego miesięcznych fakturach i liczbie posiadanych numerów telefonów.

Narysuj schemat hurtowni danych, wskaż i zdefiniuj tablicę faktów oraz tablice wymiarów, zaznacz podkreśleniem klucze główne relacji oraz linią falowaną klucze obce, zaznacz miary.

Opisz zwięźle na czym polega rozwiązanie problemu w zaproponowanym modelu hurtowni danych.

3 Pensja pracownika



Treść

Rozważmy następującą tabelę faktów i tabele wymiarów:

- sprzedaż (tabela faktów),
- data,
- oddział,
- lokalizacja,
- produkt,
- sprzedawca.

W jaki sposób zamodelować informację o pensji sprzedawcy? Opisz różne sposoby, wskaż ich zalety oraz wady.

4 Serwis aukcyjny

25p.◇

Treść

Firma internetowa prowadząca serwis aukcyjny postanowiła wdrożyć hurtownię danych w celu analizy zawieranych transakcji i zysków przez nie generowanych. Serwis posiada dużą bazę użytkowników. Każdy użytkownik może być zarazem sprzedawcą, jak i kupującym. Każdy użytkownik musi podać dokładne dane personalne, aby mógł korzystać z serwisu. Sprzedawca chcąc wystawić przedmiot na sprzedaż musi go najpierw opisać podając jego nazwę, kategorię, oraz dodatkowe cechy i otworzyć dla niego aukcję. Zamknięcie aukcji odbywa się w momencie, gdy kupujący wyśle odpowiednie potwierdzenie dokonania zakupu. Firma zarabia na prowizji od sprzedaży. Funkcja wyliczająca procent prowizji zależna jest od dotychczasowych wyników sprzedawcy. Zadanie polega na zaprojektowaniu hurtowni danych, która będzie wspomagała decyzje kierownictwa w zakresie prowadzenia przedsiębiorstwa. Kierownictwo chce otrzymać przede wszystkim odpowiedzi na następujące pytania:

1. Ile aukcji zostało otwartych/zakończonych w podziale na jednostki czasu (np. w zestawieniu dziennym, tygodniowym lub miesięcznym)?
2. Ile aukcji nie zostało zakończonych sukcesem (tzn. zakupem)?
3. Jak wygląda ranking sprzedawców pod względem kwot sprzedaży?
4. W jakich miastach kupuje się sumarycznie najwięcej przedmiotów?
5. Jakie kategorie produktów są najchętniej kupowane rano, w południe, wieczorem?

Zadania do wykonania

1. Narysuj schemat hurtowni danych w postaci gwiazdy. Wskaż tabelę faktów oraz tabele wymiarów. W każdej tabeli uwzględnij wszystkie konieczne atrybuty. W tabeli faktów wskaż atrybuty, które są miarami. Krótko opisz i uzasadnij wybór ziarna tabeli faktów. Krótko opisz tabele wymiarów.
2. Napisz zapytania SQL dla wszystkich pytań przedstawionych w opisie studium przypadku. Pamiętaj, że wyniki zapytań mają mieć postać raportu, więc muszą się one odwoływać do wszystkie potrzebnych danych w sposób czytelny.

Wyniki należy zaprezentować w krótkim raporcie. Szablon raportu dostępny jest na stronie przedmiotu (`raport-2.tex`).