Zadanie 1

Modelowanie danych pod hurtownię danych.

Proces modelowanie danych polega na stworzeniu trzech modeli:

-modelu pojęciowego – opis struktur, zawartości i przeznaczenia hurtowni danych na poziomie pojęciowym

-modelu logicznego – opis odwołujący się do elementów logicznych baz danych i procesów hurtowni (kolumn, tablic, relacji, itp.)

-modelu fizycznego – opis parametrów mających na celu optymalizacje działania hurtowni danych (indeksowanie, partycjonowanie, kopiowanie danych, sprzęt komputerowy, sieć, itp.)

Na tych trzech poziomach modelowanie dotyczy wszystkich elementów hurtowni danych – centralnej hurtowni, procesów ETL, hurtowni tematycznych, itd.

Cardinality – wystąpienia danych po obu stronach relacji. Są trzy rodzaje:

-wysoka kardynalność danych – wartości danych w kolumnie są rzadkie

-normalna kardynalność danych – wartości są rzadkie, ale nigdy unikalne

-niska kardynalność danych – wartości są powszechne.

Normalizacja – czynność reorganizacji danych w hurtowni w celu spełnienia dwóch podstawowych warunków:

-usunięcie nadmiarowości danych, przechowując wszystkie dane w jednym miejscu

-zapewnienie zależności danych – wszystkie odpowiadające im elementy danych są gromadzone razem

Ma kluczowe znaczenie przede wszystkich dlatego, że umożliwia hurtowniom zajmowanie jak najmniejszej ilości miejsca na dysku, co powoduje lepszą wydajność.

Denormalizacja – służy do zwiększenia funkcjonalności infrastruktury hurtowni danych. Wywołuje nadmiarowe dane do znormalizowanej hurtowni, aby zminimalizować czas wykonywania określonych zapytań, które łączą dane z wielu tabel w jedną.

Datamart – wyselekcjonowany podzbiór danych, często generowany dla użytkowników analityki i biznesowych. Datamarts są często tworzone jako repozytorium istotnych informacji dla podgrupy pracowników lub konkretnego przypadku użycia.

Zadanie 2

OLAP (Online analytical processing) - technologia, która organizuje duże biznesowe bazy danych i obsługuje złożone analizy. Może służyć do wykonywania złożonych zapytań analitycznych bez negatywnego wpływu na systemy transakcyjne.

Bazy danych, których firma używa do przechowywania wszystkich swoich transakcji i rekordów, nazywane są bazami danych przetwarzania transakcji online (OLTP). Te bazy danych zwykle zawierają rekordy, które są wprowadzane pojedynczo. Bazy danych używane przez OLTP nie zostały jednak zaprojektowane do analizy. Dlatego pobieranie odpowiedzi z tych baz danych jest kosztowne pod względem czasu i wysiłku. Systemy OLAP zostały zaprojektowane tak, aby pomóc w wyodrębnianiu z danych tych informacji z analizy biznesowej w wysoce wydajny sposób. Dzieje się tak, ponieważ bazy danych OLAP są zoptymalizowane pod kątem dużych obciążeń odczytu i niskiego zapisu.