Modelowanie danych to proces definiowania zawartości i struktury danych w określonym celu. Proces modelowania danych ma charakter ogólny i jest wykorzystywany zarówno w aplikacjach, jak i analizach.

Często dzieje się to w trzech częściach. Najpierw tworzony jest model koncepcyjny, który definiuje wymagania dotyczące danych wysokiego poziomu dla modelu danych. Jest to następnie tłumaczone na model logiczny, który dodaje więcej szczegółów i definiuje strukturę logiczną niezależną od miejsca docelowego. Wreszcie model logiczny jest tłumaczony na model fizyczny w źródle danych, takim jak hurtownia danych w chmurze np. Snowflake.

—---------------------------------------------------------------------------------------------------------------------

Kardynalność to termin używany w relacjach bazy danych do oznaczania wystąpień danych po obu stronach relacji.

Istnieją 3 podstawowe rodzaje kardynalności:

-Wysoka kardynalność danych:

Wartości kolumny danych są bardzo rzadkie.

-Normalna liczność danych:

Wartości kolumny danych są rzadkie, ale nie niepowtarzalne.

-Niska kardynalność danych:

-Wartości kolumny danych są bardzo typowe, np kolumny z wartościami logicznymi

Służy do określenia relacji między tabelami

—---------------------------------------------------------------------------------------------------------------------

Normalizacja to czynność reorganizacji danych w hurtowni danych w celu spełnienia dwóch podstawowych wymagań:

Usunięcie nadmiarowość danych, przechowywanych wszystkie dane w jednym miejscu

Zapewnienie odpowiedniej zależność danych, tj. wszystkie odpowiadające im elementy danych są gromadzone razem

Normalizacja ma kluczowe znaczenie z kilku powodów, ale przede wszystkim dlatego, że umożliwia magazynom danych zajmowanie jak najmniejszej ilości miejsca na dysku. Powoduje to lepszą wydajność.

Denormalizacja

Ta strategia hurtowni danych służy do zwiększania funkcjonalności infrastruktury bazy danych. Denormalizacja dodaje kontrolowane nadmiarowe dane do znormalizowanego magazynu danych, aby zminimalizować czas wykonywania określonych zapytań do bazy danych, które łączą dane z wielu tabel w jedną. Dzięki denormalizacji bazy unika się kosztownych operacji połączeń tabel

—---------------------------------------------------------------------------------------------------------------------

Data mart to tematyczna baza danych, która często stanowi podzielony na partycje segment hurtowni danych przedsiębiorstwa. Podzbiór danych przechowywanych w hurtowni danych zazwyczaj jest zgodny z konkretną jednostką biznesową, taką jak sprzedaż, finanse lub marketing. Magazyny danych przyspieszają procesy biznesowe, umożliwiając dostęp do odpowiednich informacji w hurtowni danych lub operacyjnym magazynie danych w ciągu kilku dni, a nie miesięcy lub dłużej. Ponieważ data mart zawiera tylko dane mające zastosowanie do określonego obszaru biznesowego, jest to opłacalny sposób na szybkie uzyskanie praktycznych informacji.

—---------------------------------------------------------------------------------------------------------------------

Kostka OLAP to struktura danych, która przezwycięża ograniczenia relacyjnych baz danych, zapewniając szybką analizę danych. Kostki mogą wyświetlać i sumować duże ilości danych, zapewniając jednocześnie użytkownikom dostęp z możliwością wyszukiwania do dowolnych punktów danych. W ten sposób dane mogą być gromadzone, dzielone i dzielone na kawałki w zależności od potrzeb w celu obsługi najszerszej gamy pytań, które są istotne dla obszaru zainteresowań użytkownika.

