**Proces modelowanie danych**

Modelowanie danych to kluczowy proces projektowania hurtowni danych, który obejmuje organizowanie i porządkowanie danych w celu wspierania efektywnej analizy danych i raportowania. Pomaga zdefiniować relacje między różnymi elementami danych i zapewnia strukturę danych zgodną z wymaganiami biznesowymi.

**Cardinality**

Liczność: Kardynalność odnosi się do relacji między jednostkami w modelu danych. Opisuje liczbę wystąpień jednej jednostki, które mogą być powiązane z inną jednostką. Typowe typy liczności to jeden do jednego (1:1), jeden do wielu (1: N) i wiele do wielu (N: N). Zrozumienie liczności ma kluczowe znaczenie dla projektowania wydajnych modeli danych i ustanawiania właściwych relacji między tabelami w hurtowni danych.

**Normalizacja i denormalizacja.**

Normalizacja i denormalizacja: Normalizacja to technika modelowania danych, która zmniejsza redundancję i poprawia integralność danych poprzez organizowanie danych w wiele tabel i eliminację powielania danych. Jest zgodny z zestawem reguł zwanych normalnymi formularzami, aby zapewnić wydajne przechowywanie danych. Z drugiej strony denormalizacja polega na łączeniu tabel lub duplikowaniu danych w celu optymalizacji wydajności zapytań. Może być stosowany w hurtowniach danych w celu zwiększenia wydajności odczytu zapytań analitycznych.

**Datamart**

DataMart: DataMart to podzbiór hurtowni danych, który koncentruje się na określonym obszarze tematycznym lub dziale w organizacji. Zawiera podzbiór danych z hurtowni danych, dostosowany do potrzeb raportowania i analizy określonej grupy użytkowników. Bazy danych są zwykle zaprojektowane do obsługi określonych funkcji biznesowych, takich jak sprzedaż, marketing lub finanse, dostarczając użytkownikom odpowiednich i zwięzłych informacji ułatwiających podejmowanie decyzji.

**Lakehouse vs Warehouse (hurtownia danych)**

Lakehouse vs. Warehouse: Hurtownia danych to scentralizowane repozytorium, które integruje dane z różnych źródeł, przekształca je w spójny format i zapewnia ustrukturyzowane środowisko do analizy danych i raportowania. Działa zgodnie z podejściem schematu przy zapisie, co oznacza, że ​​dane są strukturyzowane i definiowane przed załadowaniem do magazynu. Z drugiej strony, lakehouse łączy w sobie zalety DataLake i hurtowni danych. Wykorzystuje podejście oparte na schemacie podczas odczytu, w którym dane są przechowywane w surowej i nieprzetworzonej formie (podobnie jak w przypadku DataLake), ale można je również przeszukiwać w ustrukturyzowany sposób (jak hurtownia danych). Architektura Lakehouse ma na celu zapewnienie elastyczności i skalowalności DataLake przy jednoczesnym zapewnieniu niezawodności i wydajności tradycyjnych hurtowni danych.

**Kostka OLAP**

Kostka OLAP (Online Analytical Processing), często nazywana po prostu „kostką”, to wielowymiarowa struktura danych, która umożliwia wydajną analizę dużych ilości danych z różnych perspektyw. Zapewnia sposób organizowania i podsumowywania danych w formacie wielowymiarowym, umożliwiając tworzenie złożonych zapytań i agregacji na potrzeby analiz biznesowych i systemów wspomagania decyzji.

Kostki OLAP są przeznaczone do obsługi przetwarzania analitycznego online, które polega na wykonywaniu złożonych obliczeń analitycznych i agregacji na dużych zbiorach danych w czasie rzeczywistym. Struktura kostki składa się z wymiarów, miar i hierarchii. Wymiary reprezentują różne atrybuty lub perspektywy danych, takie jak czas, położenie geograficzne lub produkt. Miary, zwane też faktami, to analizowane wartości liczbowe, takie jak przychody ze sprzedaży czy sprzedana ilość. Hierarchie definiują relacje między różnymi poziomami wymiaru, umożliwiając operacje drążenia w dół lub zestawienia.

Kostki OLAP są zwykle tworzone przy użyciu specjalistycznych narzędzi i technologii. Jednym z popularnych języków używanych do wykonywania zapytań i manipulowania kostkami OLAP jest Data Analysis Expressions (DAX). DAX to język formuł, który umożliwia użytkownikom definiowanie obliczeń i zapytań na kostkach OLAP i innych tabelarycznych modelach danych. Zapewnia bogaty zestaw funkcji i operatorów do wykonywania obliczeń, filtrowania danych oraz tworzenia zaawansowanej logiki biznesowej.

Dzięki kostce OLAP i DAX użytkownicy mogą przeprowadzać szeroki zakres analiz, w tym ciąć i dzielić dane, drążyć szczegółowe informacje, tworzyć niestandardowe obliczenia i KPI (Key Performance Indicators) oraz generować interaktywne raporty i pulpity nawigacyjne. Struktura kostki i możliwości języka DAX umożliwiają użytkownikom szybkie i wydajne uzyskiwanie wglądu w duże zbiory danych, dzięki czemu jest to cenne narzędzie dla analityków biznesowych, analityków danych i decydentów w różnych branżach.