**Zestaw 14 Big Data Karol Śliwa**

**ZADANIE 1**

**Proces modelowania danych**: Proces modelowania danych to etap projektowania struktury danych w kontekście hurtowni danych lub systemu bazodanowego. Polega na opracowaniu odpowiednich modeli danych, które reprezentują strukturę danych, ich zależności, a także reguły przetwarzania i relacje między nimi. Proces ten obejmuje identyfikację wymagań biznesowych, analizę danych, projektowanie schematu, normalizację, denormalizację i inne techniki, które mają na celu efektywne przechowywanie i zarządzanie danymi.

**Cardinality**: Cardinality odnosi się do liczby unikalnych wartości w danym zbiorze danych lub w relacji między tabelami. Może opisywać relację jeden-do-jeden, jeden-do-wielu lub wiele-do-wielu między jednymi a drugimi danymi. Cardinality jest istotnym czynnikiem w projektowaniu baz danych, ponieważ wpływa na sposób organizacji danych, efektywność zapytań i optymalizację wydajności.

**Normalizacja i denormalizacja:** Normalizacja jest procesem projektowania struktury danych w taki sposób, aby uniknąć redundancji i anomalii danych. Polega na podziale danych na mniejsze i bardziej znormalizowane tabele, aby minimalizować duplikację informacji i utrzymać integralność danych. Zastosowanie normalizacji pozwala na efektywne zarządzanie danymi, ale może wymagać bardziej złożonych operacji łączenia tabel podczas wykonywania zapytań.

Denormalizacja jest procesem odwrotnym do normalizacji i polega na łączeniu danych z różnych tabel w celu zwiększenia wydajności zapytań. Często stosuje się denormalizację w przypadkach, gdy zapytania są bardziej skomplikowane i wymagają łączenia wielu tabel, co może prowadzić do spadku wydajności. Denormalizacja może być stosowana w celu optymalizacji wydajności hurtowni danych lub systemów analitycznych.

**Datamart:** Datamart to specjalnie zaprojektowana część hurtowni danych, która koncentruje się na konkretnej dziedzinie biznesowej lub grupie użytkowników. Jest to uproszczona wersja hurtowni danych, która zawiera wybrane dane i miary dla określonego obszaru biznesowego. Datamart jest często tworzony w celu zapewnienia szybkiego dostępu do danych, ułatwienia analizy i raportowania dla określonej grupy użytkowników.

**Lakehouse**: Lakehouse to pojęcie, które łączy cechy hurtowni danych i Data Lake w jednym środowisku. Jest to nowa koncepcja, która integruje zarówno przechowywanie dużych ilości danych (Data Lake) jak i przetwarzanie analityczne (Hurtownia danych) w jednym rozwiązaniu. Lakehouse umożliwia elastyczne przechowywanie danych w ich oryginalnej postaci, jednocześnie zapewniając strukturę, zarządzanie metadanymi i możliwość przetwarzania analitycznego w czasie rzeczywistym. Różni się od tradycyjnej hurtowni danych, która często ma ściśle określoną strukturę i ogranicza dostęp do danych, a także od Data Lake, który może być mniej uporządkowany i trudniejszy do przetwarzania. Lakehouse łączy w sobie elastyczność Data Lake i funkcjonalność hurtowni danych, co pozwala na efektywne zarządzanie dużymi zbiorami danych i wykonywanie zaawansowanych analiz.

**ZADANIE 2**

Kostka OLAP (Online Analytical Processing) to wielowymiarowa struktura danych, która umożliwia efektywne analizowanie dużych zbiorów danych z różnych perspektyw. Kostka OLAP jest stosowana w procesie analityki biznesowej do przeglądania, badania i analizy danych w sposób zorganizowany i zrozumiały.

Kostka OLAP jest zbudowana na podstawie danych z hurtowni danych lub innych źródeł danych, które są agregowane w różnych wymiarach. Wymiary reprezentują aspekty biznesowe, takie jak czas, lokalizacja, produkt, klient, itp. Natomiast miary to liczbowe wartości, które są obiektami analizy, takie jak sprzedaż, zysk, koszty, itp.

DAX (Data Analysis Expressions) to język programowania używany w kostkach OLAP, który umożliwia definiowanie wyrażeń i formuł do obliczania nowych miar, filtrowania danych, agregacji i innych operacji analitycznych. DAX jest używany w narzędziach analitycznych, takich jak Microsoft SQLServer Analysis Services (SSAS) i Microsoft Power BI.

Kostka OLAP i DAX są często wykorzystywane w procesie analizy biznesowej do tworzenia interaktywnych raportów, przeglądania danych na różnych poziomach szczegółowości, wykonywania zaawansowanych obliczeń i badania trendów. Dzięki możliwościom przekształcania i analizowania danych z różnych perspektyw, kostka OLAP i DAX stanowią ważne narzędzia wspierające podejmowanie decyzji i odkrywanie wiedzy biznesowej.