# Zajęcia 14

**Zadanie 1**

Modelowanie danych pod hurtownię danych. Napisz notatkę co to jest:

Proces modelowanie danych.

Cardinality.

Normalizacja i denormalizacja.

Co to jest Datamart.

Co to jest Lakehouse i jak różni się od Hurtowni.

Proces modelowania danych:

Proces modelowania danych odnosi się do tworzenia struktury i schematu danych w celu

reprezentowania informacji w sposób logiczny i zrozumiały dla systemu informatycznego. W kontekście hurtowni danych, proces ten obejmuje identyfikację źródeł danych,

projektowanie modelu danych, tworzenie relacji między tabelami, określanie kluczy

głównych i obcych oraz definiowanie hierarchii danych. Celem jest stworzenie spójnego i

optymalnego środowiska, które umożliwia łatwe przetwarzanie i analizę danych.

Cardinality:

Cardinality (kardynalność) odnosi się do liczby unikalnych wartości, które mogą wystąpić w

danej kolumnie tabeli. Może być określona jako jednoznaczna (1:1), jeden do wielu (I:N) lub wiele do wielu (N:M), w zależności od relacji między danymi. Cardinality jest ważnym

czynnikiem w projektowaniu modelu danych, ponieważ wpływa na wydajność operacji

związanych z bazą danych.

Normalizacja i denormalizacja:

Normalizacja to proces organizowania danych w relacyjnej bazie danych w celu

wyeliminowania nadmiarowości, uniknięcia anomalii i zapewnienia spójności. Polega na

podziale danych na mniejsze tabele i tworzeniu relacji między nimi. Denormalizacja z kolei

polega na powiązaniu danych z różnych tabel w jedną tabelę w celu zoptymalizowania

wydajności zapytań. Oba procesy są stosowane w zależności od potrzeb i celów projektu.

Datamart:

Datamart to podzbiór hurtowni danych, który zawiera wyselekcjonowane dane z jednej lub

kilku tematycznych dziedzin. Jest to zdecentralizowana wersja hurtowni danych, której

celem jest dostarczenie uproszczonych, zoptymalizowanych i gotowych do analizy danych

dla określonych działów, zespołów lub procesów biznesowych. Datamarty są często

tworzone w celu zwiększenia dostępności i łatwości analizy danych przez użytkowników

końcowych.

Lakehouse:

Lakehouse to nowy model architektury danych, który łączy cechy tradycyjnej hurtowni

danych i data lake. Jest to rozwiązanie, które pozwala na przechowywanie, przetwarzanie i

analizę danych w czasie rzeczywistym. W przeciwieństwie do tradycyjnych hurtowni danych,

które wymagają przekształcenia i ładowania danych przed analizą, lakehouse umożliwia

bezpośredni dostęp do surowych danych przechowywanych w formacie zbioru danych (data

lake). Lakehouse wprowadza również dodatkowe funkcje, takie jak kontrola wersji danych,

zarządzanie metadanymi i zastosowanie struktury w czasie rzeczywistym, co różni go od

tradycyjnej hurtowni danych.

**Zadanie 2**

Znajdź informację i napisz krótką notatkę co to jest kostka OLAP (OLAP CUBE, Dax).

Zadanie 2:

Kostka OLAP (OLAP Cube):

Kostka OLAP (Online Analytical Processing) jest strukturą wielowymiarową służącą do

przechowywania i organizowania danych analitycznych. Jest to jeden z głównych

komponentów systemów OLAP, które umożliwiają interaktywną analizę dużych zbiorów

danych. Kostka OLAP jest zbudowana z wymiarów i miar. Wymiary to atrybuty, według

których dane są grupowane, natomiast miary to wartości numeryczne, które są analizowane i agregowane.

Kostka OLAP pozwala na wykonywanie złożonych operacji analitycznych, takich jak

przeglądanie, filtrowanie, sortowanie, grupowanie, obliczanie sum, średnich, itp. Jest to

efektywny sposób analizy danych, szczególnie w kontekście danych wielowymiarowych,

takich jak sprzedaż, magazyn, finanse, itp.

DAX (Data Analysis Expressions):

DAX to język programowania używany w systemach Microsoft Power Bl, Analysis Services i

Power Pivot. Służy do definiowania wrażeń i formuł analitycznych do przeprowadzania

obliczeń i transformacji danych w kontekście analizy danych biznesowych. DAX umożliwia

tworzenie zaawansowanych obliczeń, agregacji, obliczeń tabelarycznych, filtrowania i innych

operacji na danych.

DAX jest często używany w połączeniu z kostkami OLAP, ponieważ umożliwia tworzenie

bardziej zaawansowanych obliczeń i wyrażeń dla danych przechowywanych w kostkach.

DAX jest łatwy do nauczenia i używania, co czyni go popularnym narzędziem w dziedzinie

analizy danych.