# Zajęcia 14

**Zadanie 1**

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

**Proces modelowanie danych** – proces tworzenia struktury i organizacji danych w celu zapewnienia skutecznego zarządzania danymi w hurtowni danych. Obejmuje projektowanie schematu danych, definiowanie relacji między tabelami, określanie kluczy głównych i obcych i opracowanie zasad integrowania danych z różnych źródeł.

**Cardinality** – (pojęcie w modelowaniu danych) relacja między dwoma zestawami danych. Określa, jak wiele rekordów z jednego zestawu danych jest powiązanych z rekordami w drugim zestawie danych. Może to być one-to-one, one-to-many lub many-to-many. W ML kardynalność to ilość etykiet jakie przyjmuje zmienna jakościowa.

**Normalizacja i denormalizacja** – organizacja danych w bazie danych w celu minimalizacji redundancji i unikania anomalii. W celu zmniejszenia się powtarzania się informacji, dane dzieli się na tabele i tworzy się odpowiednie relacje między tymi tabelami. Denormalizacja to proces odwrotny (łączenie tych tabel, zwiększanie redundancji, zwiększanie powtarzalności) służący poprawie wydajności operacji odczytu.

**Co to jest Datamart** – wyspecjalizowana, tematyczna część hurtowni danych. Mniejszy zbiór danych skupiony na konkretnym obszarze biznesowym lub grupie użytkowników. Zawiera szczegółowe, przetworzone dane z hurtowni danych, które są zoptymalizowane pod kątem określonych analiz i raportów.

**Co to jest Lakehouse i jak różni się od Hurtowni** – Lakehouse to koncepcja łącząca cechy hurtowni danych i jeziora danych. Podejście do przechowywania i przetwarzania danych, które łączy elastyczność i skalowalność jeziora danych z kontrolą jakości i strukturą hurtowni danych. Umożliwia przechowywanie surowych danych, a jednocześnie zapewnia narzędzia do ich przekształcania, modelowania i analizy w sposób zbliżony do tradycyjnej hurtowni danych. Główne różnice między lakehouse a hurtownią to format przechowywania danych (surowe vs przetworzone), sposób przetwarzania danych (na bieżąco vs wcześniej), elastyczność (różne dane w różnych formatach, strukturalne, półstrukturalne, niestrukturalne vs tylko strukturalne) i koszt (elastyczne zasoby obliczeniowe i chmura).

**Zadanie 2**

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

**Kostka OLAP** – Online Analytical Processing Cube – wielowymiarowa struktura wykorzystywana w technologii OLAP. Wirtualna reprezentacja danych, która umożliwia analitykom biznesowym przeglądanie, analizowanie i generowanie raportów z dużych zbiorów danych w interaktywny i efektywny sposób. Zbudowana jest z wymiarów i miar. Dzięki strukturze wielowymiarowej i zastosowaniu odpowiednich algorytmów umożliwia wykonywanie szybkich operacji agregacyjnych i wiercenie w danych na różnych poziomach szczegółowości.

**DAX** – Data Analysis Expressions – język zapytań i formuł używany w Power BI i Microsoft Analysis Services. DAX jest stosowany do definiowania wzorców obliczeniowych, tworzenia metryk i definiowania zależności między danymi w kostkach OLAP. Pozwala na agregację danych, tworzenie kalkulacji, filtrowanie i manipulację danymi w celu generowania zaawansowanych analiz i raportów. Potężne narzędzie dla analityków biznesowych do ekstrakcji wartościowych informacji z danych zgromadzonych w kostkach OLAP.