**Zadanie 1**

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

**Proces modelowanie danych**

Modelowanie danych to proces analizowania oraz definiowania wszystkich typów danych zbieranych i tworzonych w firmie, a także relacji między tymi danymi. W toku modelowania danych wykorzystuje się tekst, symbole i wykresy, by tworzyć wizualne reprezentacje danych rejestrowanych, przechowywanych i wykorzystywanych w firmie. Firma może zrozumieć, jak i kiedy są wykorzystywane dane — w tym kontekście proces modelowania danych sprzyja zrozumieniu i wyjaśnieniu wymagań dotyczących tych danych.

**Cardinality**

Kardynalność relacji opisuje, ile liczby wierszy w jednej tabeli może się odnieść do liczby wierszy w innej tabeli. Na przykład jeden dział może mieć wiele pracowników, ale jeden pracownik może przynależeć do tylko jednego działu. Kardynalność najczęściej wyrażana jest jako relacja jeden-do-jednego, jeden-do-wielu oraz wiele-do-wielu.

**Normalizacja i denormalizacja**

Normalizacja to proces organizowania danych w bazie danych. Obejmuje to tworzenie tabel i ustanawianie relacji między tymi tabelami zgodnie z regułami opracowanymi w celu zarówno ochrony danych, jak i zapewnienia większej elastyczności bazy danych przez wyeliminowanie nadmiarowości i niespójnych zależności.

Denormalizacja bazy danych jest to wprowadzenie kontrolowanej nadmierności do bazy danych w celu przyspieszenia wykonywania na niej operacji (np. obsługiwania zapytań); dzięki denormalizacji bazy unika się kosztownych operacji połączeń tabel.

**Co to jest Datamart**

Składnica danych (data mart) to uproszczona postać hurtowni danych skupiająca się na jednym temacie lub jednej linii biznesowej. Składnica danych zapewnia szybszy dostęp do danych i informacji, ponieważ nie trzeba tu poświęcać czasu na przeszukiwanie bardziej złożonej hurtowni danych ani na ręczne agregowanie danych z różnych źródeł.

**Co to jest Lakehouse i jak różni się od Hurtowni**

Hurtownia danych to scentralizowane repozytorium danych używane przez organizację do przechowywania ogromnych ilości danych z wielu źródeł. Hurtownia danych działa jako jedyne źródło „prawdy danych” dla organizacji i jest niezbędna do raportowania i analizy biznesowej. Zazwyczaj hurtownie danych łączą relacyjne zestawy danych z kilku źródeł, takich jak dane aplikacyjne, biznesowe i transakcyjne, w celu przechowywania danych historycznych. Przed załadowaniem do systemu magazynowego dane są przekształcane i oczyszczane w hurtowniach danych, dzięki czemu mogą być wykorzystywane jako pojedyncze źródło prawdziwości danych.

Nowatorska architektura przechowywania dużych zbiorów danych, zwana „domem jezior danych”, łączy w sobie najlepsze aspekty jezior danych i hurtowni danych. Wszystkie dane, zarówno ustrukturyzowane, częściowo ustrukturyzowane, jak i nieustrukturyzowane, mogą być przechowywane w jednym miejscu dzięki najlepszym możliwym funkcjom uczenia maszynowego, analizy biznesowej i przesyłania strumieniowego dzięki Data Lakehouse. Jeziora danych wszelkiego rodzaju są często punktem wyjścia dla Data Lakehouses; następnie dane są przekształcane do formatu Delta Lake (warstwa pamięci masowej typu open source, która zapewnia niezawodność w jeziorach danych). Najnowsza architektura przechowywania danych, znana jako „data lakehouse”, łączy niezawodność i spójność hurtowni danych z przystępną ceną i możliwościami adaptacji jezior danych.

**Zadanie 2**

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

Kostka OLAP – struktura danych, która pozwala na szybką analizę danych. Przechowuje ona dane w sposób bardziej przypominający wielowymiarowe arkusze kalkulacyjne niż tradycyjną, relacyjną bazę danych. Każda kostka składa się z wymiarów i miar, czyli analizowanych wartości. Przykładowa kostka danych zbudowana w oparciu o następujące wymiary: produkt, sklep oraz czas, gdzie w poszczególnych komórkach kostki znajdują się wartości miar, np. liczba sprzedanych sztuk czy wartość sprzedaży dostarczy użytkownikowi końcowemu informacji, jaka była ilość sprzedanego towaru (lub grup) w danym sklepie lub sklepach w konkretnym przedziale czasu (można porównywać np. odpowiadające sobie okresy różnych lat). Miary to wskaźniki numeryczne (ile), wymiary reprezentują dane opisowe (kto, co). Wymiary mogą być podzielone na poziomy.