Anna Kaniowska – 407334 – Lab14

**Zadanie 1:**

* **Proces modelowania danych.**

Proces modelowania danych obejmuje projektowanie struktury danych w celu skutecznego przechowywania, zarządzania i przetwarzania informacji w systemach informatycznych. Podczas tego procesu określa się, jak dane będą zorganizowane, jakie będą miały typy, jak będą połączone i jak będą przechowywane w bazie danych. Modelowanie danych jest kluczowe dla skutecznego zarządzania danymi i zapewnienia, że są one dostępne w formie, która jest łatwa do zrozumienia i wykorzystania.

* **Cardinality.**

Cardinality odnosi się do liczby unikalnych wartości w kolumnie w bazie danych. Może być wykorzystywana do określania relacji między tabelami. Na przykład, relacja jeden-do-jednego oznacza, że w jednej tabeli każdy rekord ma powiązanie z dokładnie jednym rekordem w innej tabeli. Natomiast relacja jeden-do-wielu oznacza, że w jednej tabeli rekord może mieć powiązanie z wieloma rekordami w innej tabeli.

* **Normalizacja i denormalizacja.**

Normalizacja i denormalizacja są technikami stosowanymi w projektowaniu baz danych. Normalizacja polega na organizowaniu danych w sposób, który minimalizuje redundancję (redundancja - nadmiarowość w stosunku do tego, co konieczne lub zwykłe) i zapewnia spójność. Proces ten opiera się na zasadach normalizacji, takich jak eliminacja powtórzeń, przechowywanie danych w odpowiednich tabelach i podział danych na logiczne jednostki. Normalizacja pomaga uniknąć anomalii danych i utrzymać integralność bazy danych. Normalizować można do 3 postaci normalnych:

* + Pierwsza postać normalizacji (1NF - First Normal Form): W tej postaci normalizacji dane w bazie danych są atomowe, czyli pojedyncze wartości nie są powtarzalne. Każda kolumna w tabeli powinna zawierać tylko pojedynczą wartość. Powtarzające się grupy danych powinny być rozdzielone do innych tabel i połączone za pomocą kluczy obcych.
  + Druga postać normalizacji (2NF - Second Normal Form): W 2NF dane muszą spełniać warunki 1NF, a dodatkowo każda niekluczowa kolumna w tabeli powinna w pełni zależeć od klucza podstawowego. Oznacza to, że niepowtarzalne wartości w tabeli muszą być powiązane tylko z jednym kluczem, a nie z częścią klucza. W przeciwnym razie, dane muszą być przeniesione do innej tabeli.
  + Trzecia postać normalizacji (3NF - Third Normal Form): W 3NF dane muszą spełniać warunki 2NF, a dodatkowo każda niekluczowa kolumna w tabeli nie może zależeć od innych niekluczowych kolumn. Oznacza to, że dane, które zależą od innych danych niekluczowych, powinny zostać przeniesione do osobnej tabeli.

Z drugiej strony, denormalizacja polega na wprowadzeniu redundancji w celu zwiększenia wydajności operacji odczytu. Denormalizacja może być stosowana, gdy konieczne jest szybkie pobieranie danych z bazy danych kosztem pewnej redundancji.

* **Co to jest Datamart.**

Datamart to specjalnie zaprojektowana i zorganizowana część hurtowni danych, która koncentruje się na konkretnym obszarze tematycznym lub grupie użytkowników. Datamart zawiera agregowane, przefiltrowane i przetworzone dane, które są dostosowane do potrzeb konkretnych użytkowników lub działów organizacji. Datamarty są tworzone w celu ułatwienia analizy i raportowania, umożliwiając dostęp do potrzebnych informacji bez konieczności przeszukiwania całej hurtowni danych.

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

Lakehouse to nowoczesne podejście do zarządzania danymi, które łączy cechy hurtowni danych i data lake. Tradycyjnie, hurtownie danych i data lake były stosowane jako odrębne rozwiązania. Hurtownie danych były oparte na uśrednionym, znormalizowanym schemacie danych, które były zintegrowane i skonsolidowane w celu udostępnienia jednolitego źródła prawdy. Z kolei data lake był bardziej elastycznym magazynem danych, w którym gromadzone były różnorodne typy danych w ich pierwotnej postaci.

Lakehouse łączy te dwa podejścia, umożliwiając przechowywanie zarówno strukturalnych, jak i niestrukturalnych danych w jednym miejscu. Daje to możliwość przechowywania danych w pierwotnej postaci, co ułatwia elastyczność i skalowalność analizy danych. Lakehouse umożliwia również przetwarzanie w czasie rzeczywistym, co oznacza, że dane mogą być analizowane i wykorzystywane niemal natychmiast po ich pojawieniu się.

Różnica między lakehouse a tradycyjną hurtownią danych polega na tym, że lakehouse umożliwia przechowywanie danych w ich pierwotnej postaci, nie wymagając uprzedniego ich przetwarzania i normalizacji. W przeciwności do tego, tradycyjne hurtownie danych wymagają dokładnego zdefiniowania schematu danych i przetwarzania etapowego w celu załadowania danych do odpowiednich tabel i relacji.

**Zadanie 2 :**

* **Co to jest kostka OLAP (OLAP CUBE, Dax)**

Kostka OLAP (OLAP Cube): Kostka OLAP, znana również jako OLAP Cube, jest wielowymiarową strukturą danych służącą do analizy wielowymiarowych danych biznesowych. Kostka OLAP jest zbudowana na podstawie wymiarów i miar. Wymiary reprezentują różne aspekty analizowanych danych, takie jak czas, produkt, lokalizacja, klient itp. Miary to numeryczne wartości, które są obiektami analizy, takie jak sprzedaż, zysk, ilość itp.

Kostka OLAP zawiera agregowane dane z różnych kombinacji wymiarów i miar, tworząc strukturę hierarchiczną. Pozwala to na efektywne wykonywanie złożonych analiz i zagregowanych obliczeń, takich jak suma, średnia, maksimum, minimum itp. Kostki OLAP są wykorzystywane w narzędziach biznesowych do generowania raportów, analizy trendów, wykrywania wzorców i podejmowania decyzji biznesowych na podstawie wielowymiarowych danych.

DAX (Data Analysis Expressions): DAX jest językiem formuł używanym w narzędziach biznesowych firmy Microsoft, takich jak Power BI i Excel Power Pivot, do definiowania obliczeń i wyrażeń w raportach i analizach. DAX umożliwia manipulację danymi w kostkach OLAP oraz tworzenie wyrażeń i metryk biznesowych.

DAX oferuje szeroki zakres możliwości obliczeniowych i funkcji, które umożliwiają zaawansowaną analizę danych. Można go używać do tworzenia miar, obliczeń czasu, filtrowania danych, sortowania, grupowania, tworzenia warunkowych wyrażeń, a także do tworzenia zaawansowanych analiz, takich jak wskaźniki wzrostu, udziały procentowe, porównania okresowe, trendowanie danych i wiele innych.

DAX umożliwia również tworzenie złożonych relacji między tabelami oraz agregację i filtrację danych na podstawie różnych wymiarów. Jego składnia przypomina formuły języka Excel, co ułatwia zrozumienie i stosowanie dla osób, które są już zaznajomione z tym narzędziem.

Dzięki DAX można tworzyć interaktywne i dynamiczne raporty oraz wykresy, które umożliwiają użytkownikom analizę danych z różnych perspektyw i poziomów szczegółowości. DAX jest wszechstronnym narzędziem do analizy danych, które daje użytkownikom możliwość manipulowania, transformowania i prezentowania informacji biznesowych w sposób intuicyjny i skuteczny.