**Proces modelowanie danych:**

1. Pozyskiwanie – pobieranie danych z systemów źródłowych.
2. Przekształcanie – transformacja danych.
3. Ładowanie – ładowanie przekształconych danych.
4. Dostarczanie wiedzy – udostępnianie wiedzy decyzyjnej.

**Cardinality:**

Kardynalność relacji to liczba rekordów encji (tabeli) nadrzędnej powiązanych z rekordami encji podrzędnej.

Przykłady: "jeden do jednego", "jeden do wielu", "wiele do wielu"

**Normalizacja i denormalizacja:**

Normalizacja jest techniką minimalizowania anomalii wstawiania, usuwania i aktualizacji poprzez eliminację zbędnych danych.

Denormalizacja jest odwrotnym procesem normalizacji, w którym redundancja jest dodawana do danych w celu poprawy wydajności konkretnej aplikacji i integralności danych.

Pierwsza postać normalna = 1PN (ang. 1NF = First normal form) – każdy atrybut w tabeli ma wartość elementarną (brak powtarzających się grup informacji), tabela posiada klucz;

Druga postać normalna = 2PN (ang. 2NF = Second normal form) – tabela powinna przechowywać dane dotyczące tylko konkretnej klasy obiektów;

Trzecia postać normalna = 3PN (ang. 3NF = Third normal form) – każdy niekluczowy atrybut jest bezpośrednio zależny tylko od klucza głównego.

**Co to jest Datamart.**

Data Mart to tabela/tabele zawierające tylko te informacje, które są analitykowi niezbędne do pracy. Są to dane zaciągnięte z różnych źródeł, przetworzone w sposób jednolity, udokumentowany i zoptymalizowany.

Data Marty często zawierają agregaty na poziomie klienta, np. średnia liczba transakcji z ostatnich 6 miesięcy, czy liczba pożyczek gotówkowych zaciągniętych przez klienta w ciągu ostatnich 12 miesięcy itp. Posiadając gotowe już wyliczone agregaty możemy sprawniej przygotować zestawienia i raporty.

Data Marty są budowane tylko raz, na początku procesu analitycznego, natomiast są cyklicznie i automatycznie aktualizowane, dzięki czemu zawierają wszystkie niezbędne informacje związane z klientami/produktami/transakcjami za dany okres czasu.

**Kostka OLAP (OLAP CUBE, Dax)**

Wielowymiarowa kostka OLAP (cube) jest podstawową strukturą danych w każdym systemie OLAP działającym w środowisku Hurtowni Danych.

Cube składa się z Miar (Measures), Wymiarów (Dimensions) i Poziomów (Levels) i jest zoptymalizowany pod kątem szybkiego i bezpiecznego dostępu do danych wielowymiarowych.

Miary to wskaźniki numeryczne, natomiast wymiary reprezentują dane opisowe. Wymiary są pogrupowane za pomocą poziomów, które odzwierciedlają hierarchię funkcjonującą w organizacji i pozwalają użytkownikom końcowym zwiększać lub zmniejszać poziom szczegółowości analizowanego wymiaru.

Z reguły w hurtowni danych jest zdefiniowanych co najmniej kilkanaście wymiarów, a najczęściej spotykanym i wymiarami są:

- Czas  
 - Klient  
- Produkt  
- Lokalizacja  
- Biuro Sprzedaży

Hierarchia każdego z wymiarów ustawiona jest za pomocą Poziomów.

Przykładowo, hierarchia poziomów może być ułożona w następujący sposób::

wymiar Czas: Rok -> Kwartał -> Miesiąc -> Tydzień -> Dzień  
Klient: Grupa klientów -> Nazwa klienta  
Produkt: Linia Produktu -> Grupa Produktu -> Produkt  
Lokalizacja: Obszar -> Region -> Kraj

Kategorie to elementy danych które opisują poziomy w wymiarach.

Przykładowo, dla wymiaru Lokalizacji, w hurtowni danych zostały ustawione poziomy obszaru, regionu i kraju. W tym przykładzie dla Polski kategoriami będą:

Obszar – Europa  
Region - Europa Środkowa  
Kraj - Polska

Typowe, najczęściej występujące Miary w hurtowniach danych to:

Przychód netto  
Przychód brutto  
Waga  
Ilość  
Koszt  
Upust