**Modelowanie danych pod hurtownię danych:**

Projektowanie hurtowni danych polega na stworzeniu modelu pojęciowego, logicznego i fizycznego hurtowni.

Modelowanie zachodzi na trzech poziomach:

* Model pojęciowy to opis struktury, zawartości i przeznaczenia hurtowni danych. Model pojęciowy może np. określić, że konieczne jest gromadzenie pewnych informacji o klientach, wskazuje też na biznesowe cele zaplanowanych analiz.
* Model logiczny to opis odwołujący się do elementów logicznych baz danych i procesów hurtowni, a więc kolumn, tablic, relacji itp. Jest to typowy projekt bazy danych wykonany np. w języku UML.
* Model fizyczny to opis parametrów mających na celu optymalizację działania hurtowni danych, takich jak indeksowanie, partycjonowanie, rozmieszczenie poszczególnych zasobów logicznych, itp.

**Cardinality**

W modelowaniu danych, mówimy o cardinality, gdy jedna tabela odnosi się do innej.

* 1-1 (jeden wiersz w tabeli A odnosi się do jednego wiersza w tabeli B)
* 1-Many (jeden wiersz w tabeli A odnosi się do wielu wierszy w tabeli B)
* Many-Many (wiele wierszy w tabeli A odnosi się do wielu wierszy w tabeli B)

**Normalizacja & Denormalizacja**

Normalizacja

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.

Istnieje kilka reguł normalizacji bazy danych. Każda reguła jest nazywana „postacią normalną”.

### Wpływ normalizacji na objętość i wydajność baz danych

* baza danych zajmuje mniej fizycznego miejsca na dysku
* możliwe jest łatwe przetwarzanie danych
* większa spójność danych

Denormalizacja

Polega na tworzeniu danych nadmiarowych przechowywanych w relacjach, co pozwala - podczas wykonywania zapytań analitycznych - zmniejszyć liczbę kosztownych czasowo operacji złączenia.

### Zalety denormalizacji baz danych

* radykalny zysk wydajności,
* obniżenie kosztów CPU/RAM,

**Data mart - tematyczna hurtownia danych**

Hurtownie tematyczne stanowią wydzielone i przetworzone podzbiory danych,przeznaczone do konkretnych rodzajów analiz. Przykładowe zastosowanie hurtowni tematycznych to OLAP

**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. Operacje które można wykonać na kostce to np. zwijanie (agregacja), projekcja (zmniejszenie liczby wymiarów), sortowanie (tworzenie rankingów).